Bounds on the Minimum Field Size of Network MDS Codes

Hengjia Wei and Moshe Schwartz, *Fellow, IEEE*

***Abstract*— We study network maximum distance separable (MDS) codes, which are a class of network error-correcting codes whose distance attains the Singleton-type bound. The minimum field size of a network MDS code is of particular interest, since it impacts the computing complexity at the network nodes. Previous constructions of network MDS codes, which are applicable to general single-source multicast networks, require large field sizes. In this paper, for two specific classes of network topologies, we derive upper and lower bounds on the minimum field size of the corresponding network MDS codes and present explicit constructions. The proposed upper bounds significantly improve upon the previous ones and differ from the lower bounds only by a small factor, which is asymptotically no more than 2. Additionally, we extend the concept of linear network error-correction coding from the scalar case to the vector case, and demonstrate a class of networks in which the minimum field size of the vector network MDS code is substantially smaller than that of the scalar case.**

***Index Terms*— Error-correcting codes, MDS codes, network coding.**

I. Introduction

IN THE network coding scheme, network nodes can encode their received messages and send the computed results to downstream nodes. Compared with simple message routing, network coding may achieve a higher information rate, and has attracted considerable attention in the past two decades. The idea of network coding can be traced back to Celebiler and Stette's work [6] on satellite communications. In 1999, Yeung and Zhang [33] investigated a general source-coding system which consists of multiple sources, multiple encoders, and multiple decoders. One year later, the concept of network coding was formally proposed by Ahlswede et al. [1], where it was shown that by using network coding, a source node can multicast messages to all the sink nodes at the maximum rate if the alphabet size tends to infinity. For a network-coding scheme, if the encoding function at each network node is linear, then it is called a *linear network coding*. Li et al. [19] examined the multicast problem and demonstrated that a linear network coding over a finite alphabet is sufficient to achieve the maximum information rate. Koetter and Médard [17] provided an algebraic formulation for linear network coding. Jaggi et al. [15] showed that there is a polynomial time algorithm to construct maximum-rate linear network codes, as long as the field size is no smaller than the number of sink nodes. For a given multicast network, determining the minimum field size over which there are linear network codes which achieve the maximum information rate is still open, and this problem is known to be NP-hard [18]. Nevertheless, several papers study the minimum field size required to solve some important families of network, e.g., see [27] and [28] and the recent [14]. Other papers study the similar problem for vector solutions, e.g., [3], [20], [26].

Network communications may suffer from various kinds of errors, including random errors caused by channel noise, erasure errors caused by traffic jams, malicious attacks by an adversary, and so on. Error correction in network communications is more challenging than that in the classical point-to-point communications, in the sense that even a single error that occurs in a link can propagate to all the downstream links and has the potential to corrupt all the messages received by a sink node. Cai and Yeung [4] combined network coding with error correction, and proposed a new kind of coding technique called *network error-correction coding*, which can combat errors by introducing redundancy in the space domain instead of in the time domain. In [5], [32], and [34], three well-known bounds in classical coding theory, i.e., the Hamming bound, the Gilbert-Varshamov bound, and the Singleton bound, are generalized to network error-correction coding. Linear network codes that attain the Singleton bound are called *network maximum distance separable (MDS) codes*. Different methods are proposed in [11], [12], [22], [32], and [36] to construct such codes.

In this paper, we study the minimum field size required by network MDS codes. Like their counterpart in classical coding theory, the field size of network MDS codes is of particular interest, as it impacts the computing complexity at the network nodes, and affects the practical implementation of network coding. The constructions of network MDS codes in [11], [12], [22], [32], [36] are applicable to general single-source multicast networks. However, they require quite large field sizes. In this paper, we study some specific networks and derive upper and lower bounds on the minimum field size of

Manuscript received 16 February 2023; revised 18 October 2023; accepted 20 June 2024. Date of publication 25 June 2024; date of current version 16 July 2024. The work of Hengjia Wei was supported in part by the Major Key Project of Peng Cheng Laboratory under Grant PCL2024AS103 and Grant PCL2023AS1-2 and in part by the National Natural Science Foundation of China under Grant 12371523. The work of Moshe Schwartz was supported in part by German Israeli Project Cooperation (DIP) Grant under Grant PE2398/1-1. *(Corresponding author: Hengjia Wei.)*

Hengjia Wei is with the Peng Cheng Laboratory, Shenzhen 518055, China, also with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China, and also with Pazhou Laboratory (Huangpu), Guangzhou 510555, China (e-mail: hjwei05@gmail.com).

Moshe Schwartz is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4L8, Canada, on leave from the School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Be'er Sheva, BGU 8410501, Israel (e-mail: schwartz.moshe@mcmaster.ca).

Communicated by A.-L. Horlemann-Trautmann, Associate Editor for Coding and Decoding.

Digital Object Identifier 10.1109/TIT.2024.3418471

the corresponding network MDS codes. The main contribution of this paper are:

1) For combination networks, we show that the existence of a network MDS code is equivalent to that of a classical MDS code of certain parameters. As a consequence, some known bounds for the latter can be used directly for the former. It is worth noting that in the error-free case a linear network coding of combination networks is also equivalent to a classical MDS code, but with different parameters, see [24].
2) For the Zosin-Khuller networks, we first give a lower bound on the minimum field size. Then we present code constructions in two parameter regions, the required field size of which only differs from the lower bound by a small factor, which is asymptotically no more than 2.
3) We extend the concept of network error-correction coding from the scalar case to the vector case, where the messages carried on the communication links are vectors and the coding coefficients are matrices. We demonstrate a class of networks in which the minimum field size of the vector network MDS code is substantially smaller than that of the scalar network MDS code.

II. Preliminaries

A. Network Coding

Let $G = (\mathcal{V}, \mathcal{E})$ be a finite directed acyclic graph with a vertex set $\mathcal{V}$ and an edge set $\mathcal{E}$, where multiple parallel edges are allowed between any two vertices. A *network* over G is denoted by $\mathcal{N} = (G, S, R)$, where $S \subset \mathcal{V}$ is a non-empty set of source nodes (transmitters) and $R \subset \mathcal{V}$ is a non-empty set of sink nodes (receivers) with $S \cap R = \emptyset$. An edge $e = (i, j) \in \mathcal{E}$ represents a link from node i to node j, and we use $\mathrm{tail}(e)$ and $\mathrm{head}(e)$ to denote the tail node i and the head node j, respectively. We denote by $\mathrm{In}(i)$ the set of incoming edges of node i, and by $\mathrm{Out}(i)$ the set of outgoing edges.

In this paper, we consider single-source multicast networks. In such a network, there is exactly one source, say σ, with h messages, and all the receivers request all these h messages. We assume that each link in the network has unit capacity and it transmits a packet to its head. We distinguish between two kinds of coding schemes, namely, *scalar network coding* and *vector network coding*: in the former case the messages $\{x_i : 1 \leqslant i \leqslant h\}$ and the packets $\{u_e : e \in \mathcal{E}\}$ carried by the links are symbols from a field $\mathbb{F}_q$, while in the latter case the messages $\{\mathbf{x}_i : 1 \leqslant i \leqslant h\}$ and the packets $\{\mathbf{u}_e : e \in \mathcal{E}\}$ are vectors of length t over a field $\mathbb{F}_{q'}$. In both cases, every packet carried by link e can be treated as a function of the packets carried by the links in $\mathrm{In}(\mathrm{tail}(e))$. The network coding is called *linear* if all these functions are linear.

In scalar linear network coding, the packet u_e carried by link e can be calculated as[1]

$$u_e = \sum_{d \in \mathrm{In}(\mathrm{tail}(e))} u_d k_{d,e},$$

where $k_{d,e} \in \mathbb{F}_q$ and $(k_{d,e} : d \in \mathrm{In}(\mathrm{tail}(e)))^\intercal$ is called the *local encoding vector* of e. Noting that u_e is a linear combination of the messages $x_1, x_2, \ldots, x_h$, there is a column vector $\mathbf{f}_e \in \mathbb{F}_q^{h \times 1}$ such that

$$u_e = (x_1, x_2, \ldots, x_h) \cdot \mathbf{f}_e,$$

where $\mathbf{f}_e$ is called the *global encoding vector* of e. We note that global encoding vectors can be determined by local encoding vectors, and a scalar network code can be described either by the local encoding vectors or by the global encoding vectors. For each receiver $\gamma \in R$, denote

$$G(\gamma) \triangleq (\mathbf{f}_e)_{e \in \mathrm{In}(\gamma)} \in \mathbb{F}_q^{h \times |\mathrm{In}(\gamma)|}.$$

Then, γ receives the packets $(u_e : e \in \mathrm{In}(\gamma)) = (x_1, x_2, \ldots, x_h) \cdot G(\gamma)$ and it can decode all the messages $x_1, x_2, \ldots, x_h$ if and only if $G(\gamma)$ has full row rank.

In vector linear network coding, all the messages and packets are *row* vectors of length t. Similarly to the scalar case, we have the local and global descriptions of a vector network code: the packet $\mathbf{u}_e$ carried by link e can be calculated as

$$\mathbf{u}_e = \sum_{d \in \mathrm{In}(\mathrm{tail}(e))} \mathbf{u}_d K_{d,e},$$

where $K_{d,e} \in \mathbb{F}_q^{t \times t}$, and the matrix $\left((K_{d,e}^\intercal)_{d \in \mathrm{In}(\mathrm{tail}(e))}\right)^\intercal$ is called the *local encoding matrix* of e; $\mathbf{u}_e$ can also be written as a linear combination of the messages $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_h$, namely,

$$\mathbf{u}_e = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_h) \cdot F_e,$$

where $F_e \in \mathbb{F}_q^{th \times t}$ is called the *global encoding matrix* of e. Again, a receiver γ can decode all the messages if and only if the matrix

$$G(\gamma) \triangleq (F_e)_{e \in \mathrm{In}(\gamma)} \in \mathbb{F}_q^{th \times t|\mathrm{In}(\gamma)|}$$

has full row rank. Here we use the same notation $G(\gamma)$ as in the scalar case, for simplicity of notation.

A linear network code $\mathscr{C}$ is called a *regular code* or a *solution* to the network $\mathcal{N}$ if $G(\gamma)$ has full row rank for every receiver $\gamma \in R$.

B. Linear Network Error-Correction Coding

This paper concerns linear network error-correction coding, which has been studied mainly in the scalar case in the literature, see [11], [12], [22], [32], and [36]. We revisit terminologies and bounds with respect to the scalar network error-correction coding in this subsection, and we shall generalize them to the vector case later.

If an error occurs in a link e, its head node $\mathrm{head}(e)$ receives a packet $\widetilde{u}_e = u_e + z_e$, where $u_e \in \mathbb{F}_q$ is the packet that is supposed to be transmitted by e and $z_e \in \mathbb{F}_q$ is the additive error. We treat z_e as a message, called the *error message*, and the vector $\mathbf{z} = (z_e)_{e \in \mathcal{E}}$ is referred to as the error message vector. An *error pattern* $\rho \subseteq \mathcal{E}$ is a set of links in which errors occur. We say an error message vector $\mathbf{z}$ matches an error pattern ρ, if $z_e = 0$ for all $e \notin \rho$, namely, $\mathrm{supp}(\mathbf{z}) \subseteq \rho$. The packet $\widetilde{u}_e$ can be calculated as

$$\widetilde{u}_e = \sum_{d \in \mathrm{In}(\mathrm{tail}(e))} k_{d,e} \widetilde{u}_d + z_e,$$

[1] If $\mathrm{tail}(e) = \sigma$, we add h imaginary links to the source σ and assume that each link carries a message x_i.

It can also be written as a linear combination of the messages $x_1, x_2, \ldots, x_h$ and error messages z_e with $e \in \mathcal{E}$, namely,

$$\widetilde{u}_e = (\mathbf{x}, \mathbf{z}) \cdot \widetilde{\mathbf{f}}_e$$

where the column vector $\widetilde{\mathbf{f}}_e \in \mathbb{F}_q^{(h+|\mathcal{E}|)\times 1}$ is called the *extended global encoding vector* of e. For each receiver $\gamma \in R$, denote

$$\widetilde{G}(\gamma) \triangleq (\widetilde{\mathbf{f}}_e)_{e\in \mathrm{In}(\gamma)} \in \mathbb{F}_q^{(h+|\mathcal{E}|)\times|\mathrm{In}(\gamma)|}.$$

Then $\widetilde{G}(\gamma)$ can be written as

$$\widetilde{G}(\gamma) = \begin{pmatrix} G(\gamma) \\ H(\gamma) \end{pmatrix},$$

where $G(\gamma) = (\mathbf{f}_e)_{e\in\mathrm{In}(\gamma)} \in \mathbb{F}_q^{h\times|\mathrm{In}(\gamma)|}$, and $H(\gamma) \in \mathbb{F}_q^{|\mathcal{E}|\times|\mathrm{In}(\gamma)|}$. We further denote

$$\Phi(\gamma) \triangleq \left\{\mathbf{x}\cdot G(\gamma) \ : \ \mathbf{x} \in \mathbb{F}_q^h\right\}, \tag{1}$$

$$\Delta(\gamma,\rho) \triangleq \left\{\mathbf{z}\cdot H(\gamma) \ : \ \mathbf{z} \in \mathbb{F}_q^{|\mathcal{E}|},\ \mathrm{supp}(\mathbf{z}) \subseteq \rho\right\}. \tag{2}$$

The minimum distance of a regular code $\mathscr{C}$ at receiver γ is defined by

$$d(\mathscr{C},\gamma) \triangleq \min\{|\rho| \ : \ \Phi(\gamma)\cap\Delta(\gamma,\rho) \neq \{\mathbf{0}\}\}.$$

Note that

$$\begin{aligned} &d(\mathscr{C},\gamma)\\ &= \min\{|\rho| \ : \ \Phi(\gamma)\cap\Delta(\gamma,\rho)\neq\{\mathbf{0}\}\}\\ &= \min\big\{|\rho| \ : \ \exists\ \mathbf{x},\mathbf{z} \text{ with } \mathbf{x}\neq\mathbf{0},\ \mathrm{supp}(\mathbf{z})\subseteq\rho, \text{ s.t.}\\ &\qquad\qquad \mathbf{x}\cdot G(\gamma) = \mathbf{z}\cdot H(\gamma)\big\}\\ &= \min\{|\rho| \ : \ \exists\ \mathbf{x},\mathbf{x}',\mathbf{z},\mathbf{z}' \text{ with } \mathbf{x}\neq\mathbf{x}',\ \mathrm{supp}(\mathbf{z}-\mathbf{z}')\subseteq\rho \text{ s.t.}\\ &\qquad\qquad (\mathbf{x},\mathbf{z})\cdot\widetilde{G}(\gamma) = (\mathbf{x}',\mathbf{z}')\cdot\widetilde{G}(\gamma)\}. \end{aligned}$$

Thus, the code $\mathscr{C}$ allows the receiver γ to correct up to $\lfloor (d(\mathscr{C},\gamma)-1)/2\rfloor$ link errors.

A cut between the source node σ and a receiver γ is a subset of vertices $U \subseteq \mathcal{V}$ such that $\sigma \in U$ but $\gamma \notin U$. We further denote by $\mathcal{E}_\gamma$ the set of all edges that are on a path from σ to γ. We can then talk about edges crossing the cut forward and backward,

$$\mathcal{E}_\gamma^{\rightarrow}(U) \triangleq \{e\in\mathcal{E}_\gamma \ : \ \mathrm{tail}(e)\in U, \mathrm{head}(e)\notin U\},$$
$$\mathcal{E}_\gamma^{\leftarrow}(U) \triangleq \{e\in\mathcal{E}_\gamma \ : \ \mathrm{tail}(e)\notin U, \mathrm{head}(e)\in U\}.$$

With these we can define the *minimum cut capacity* between σ and γ,

$$C_\gamma \triangleq \min_U\left\{\left|\mathcal{E}_\gamma^{\rightarrow}(U)\right| - \left|\mathcal{E}_\gamma^{\leftarrow}(U)\right|\right\}.$$

We have the following Singleton-type bound on the distance $d(\mathscr{C},\gamma)$.

Theorem 1 ([11, Theorem 2]): Let $\mathscr{C}$ be a scalar regular code. Then for each receiver $\gamma \in R$, we have that

$$d(\mathscr{C},\gamma) \leqslant C_\gamma - h + 1.$$

If a regular code has $d_{\min}(\mathscr{C},\gamma) = C_\gamma - h + 1$ for every receiver γ, it is called a *network maximum distance separable (MDS) code*. In [11], [12], [22], [32], and [36], different methods were proposed to construct network MDS codes. Among others, Guang and Yeung [12] obtained the best known upper bound on the minimum field size of network MDS codes by applying a graph-theoretic approach. This bound involves a series of new notions, which are not used anywhere else in this paper. For the reader's convenience, we simply review these notions. A *cut separating a sink $\gamma \in V$ from a subset of edges $\rho \subseteq \mathcal{E}$* is a set of edges such that if we remove these edges, then from any edge of ρ we cannot reach the sink γ. By definition, ρ is a cut separating ρ from γ. A cut separating γ from ρ is called a *minimum cut separating γ from ρ* if its capacity achieves the minimum. A minimum cut separating γ from ρ is *primary* if it separates γ from all the minimum cuts that separate γ from ρ. For a positive integer r, define

$$\mathscr{A}_\gamma(r) \triangleq \{\rho\subseteq\mathcal{E} \ : \ |\rho| = r \text{ and } \rho \text{ is primary}\}.$$

Theorem 2 ([12]): Let $\mathbb{F}_q$ be a finite field of order q. Let R be the set of receivers in the network $\mathcal{N}$ with $C_\gamma \geqslant h$ for every $\gamma \in R$. If

$$q > \sum_{\gamma\in R}|\mathscr{A}_\gamma(C_\gamma - h)|,$$

then there exists a network MDS code $\mathscr{C}$ over $\mathbb{F}_q$ in $\mathcal{N}$.

The size of $\mathscr{A}_\gamma(C_\gamma - h)$ can be bounded from below by an explicit form [12, Corollary 10], namely,

$$|\mathscr{A}_\gamma(C_\gamma - h)| \geqslant \binom{|\mathrm{In}(\gamma)|}{C_\gamma - h}.$$

To the extent of our knowledge, any of the known constructions of network MDS codes requires field size larger than

$$\sum_{\gamma\in R}\binom{|\mathrm{In}(\gamma)|}{C_\gamma - h}. \tag{3}$$

For a network $\mathcal{N}$, we denote the minimum field size of a network MDS code by $q^{\mathrm{MDS}}(\mathcal{N})$. In this paper, we shall study two classes of networks and determine $q^{\mathrm{MDS}}(\mathcal{N})$ up to a small constant factor. Our result shows that the value of $q^{\mathrm{MDS}}(\mathcal{N})$ is significantly smaller than the bound in (3). A summary of the bounds and constructions, and a comparison with the bound in (3), are given in Table I. Note that if h is fixed and the parameter of the network n or N tends to infinity, the upper and lower bounds are about a factor of 2 apart.

Some of our bounds for network MDS codes are derived from the ones for classical MDS codes. Let n, k be two positive integers with $2 \leqslant k < n$. For an $[n,k]_q$-MDS code, using the Griesmer bound, one can show that $n-k+1 \leqslant q$, see [25, pp. 340–341]. On the other hand, for every prime power q such that $q \geqslant n-1$, there is an $[n,k]_q$-MDS code (e.g., a Reed-Solomon code). We denote the minimum field size of an $[n,k]$-MDS code by $q^{\mathrm{MDS}}(n,k)$. Thus, for $n > k \geqslant 2$,

$$\psi(n-k+1) \leqslant q^{\mathrm{MDS}}(n,k) \leqslant \psi(n-1),$$

where $\psi(x)$ denotes the smallest prime power that is greater than or equal to x.

C. $\mathbb{F}_q$-Linear Codes Over $\mathbb{F}_q^t$

Let q be a prime power and n, t be positive integers. We consider vectors over $\mathbb{F}_q^t$ of length n. For a vector

TABLE I

LOWER AND UPPER BOUNDS ON THE MINIMUM FIELD SIZE OF NETWORK MDS CODES. THE NOTATION $\psi(x)$ OR $\psi_{\text{even}}(x)$ DENOTES THE SMALLEST PRIME POWER OR EVEN PRIME POWER, RESPECTIVELY, THAT IS GREATER THAN OR EQUAL TO x

Networks	Comment	Lower Bound	Ref.	Upper bound	Ref.	$\sum_{\gamma\in R}\binom{\lvert\text{In}(\gamma)\rvert}{C_\gamma-h}$
$\mathcal{N}_{h,n,\alpha}$		$\psi(n-h+1)$	Cor. 7	$\psi(n-1)$	Cor. 7	$\binom{n}{\alpha}\binom{\alpha}{h}$
$\mathcal{Z}_{h,m,N}$	$m=2$	$\psi(N-h)$	Lemma 11	$\psi_{\text{even}}(N)$	Thm. 13	$N\binom{N-1}{h}$
	$m>N/2$	$\psi(\binom{N-1}{m-1}-h+1)$		$\psi\left(2\binom{N-1}{m-1}-1\right)$	Thm. 14	$N\binom{\binom{N-1}{m-1}}{h}$

$\mathbf{x}=(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n)$ of $(\mathbb{F}_q^t)^n$ with each $\mathbf{x}_i\in\mathbb{F}_q^t$, its Hamming weight is the number of non-zero symbols $\mathbf{x}_i$,

$$\text{wt}_H(\mathbf{x})\triangleq|\{1\leqslant i\leqslant n\ :\ \mathbf{x}_i\neq\mathbf{0}\}|.$$

Similarly, for two vectors $\mathbf{x},\mathbf{y}$ of $(\mathbb{F}_q^t)^n$, their Hamming distance is measured with respect to symbols of $\mathbb{F}_q^t$, and is equal to the number of indices i where $\mathbf{x}_i\neq\mathbf{y}_i$, i.e.,

$$d_H(\mathbf{x},\mathbf{y})\triangleq\text{wt}_H(\mathbf{x}-\mathbf{y}).$$

A code $\mathcal{C}$ over $\mathbb{F}_q^t$ of length n is a subset of $(\mathbb{F}_q^t)^n$. We say $\mathcal{C}$ is *$\mathbb{F}_q$-linear* if it is a vector space over $\mathbb{F}_q$.

Lemma 1: Let $\mathcal{C}$ be an $\mathbb{F}_q$-linear code of $(\mathbb{F}_q^t)^n$ with a generator matrix G. Partition G into n blocks as $G=\left(B_1\ B_2\cdots B_n\right)$, where each B_i comprises of t columns. Then $\mathcal{C}$ has minimum Hamming distance at least d if and only if any $n-d+1$ blocks constitute a submatrix of full row rank.

Proof: Let K denote the dimension of $\mathcal{C}$. For the only if part, suppose to the contrary that there exists a submatrix G' that consists of $n-d+1$ blocks but has rank less than K. Then there is a non-zero row vector $\mathbf{x}\in\mathbb{F}_q^K$ such that $\mathbf{x}\cdot G'=\mathbf{0}$. Consider the codeword $\mathbf{x}\cdot G$. As a vector over $(\mathbb{F}_q)^t$ of length n, it has Hamming weight at most $n-(n-d+1)=d-1$, a contradiction.

For the if part, suppose to the contrary that there is a non-zero codeword $\mathbf{c}$ of $\mathcal{C}$ with Hamming weight less than d. Then there are $n-d+1$ indices $i_1,i_2,\ldots,i_{n-d+1}$ such that $\mathbf{c}$ has the zero symbol of $\mathbb{F}_q^t$ in the i_ℓ entry for each $1\leqslant\ell\leqslant n-d+1$. Let G' be the submatrix of G consisting of the $n-d+1$ blocks B_{i_ℓ}. According to the assumption, G has full rank. Let $\mathbf{x}$ be the non-zero vector of $\mathbb{F}_q^K$ such that $\mathbf{x}\cdot G=\mathbf{c}$. Then $\mathbf{x}\cdot G'=\mathbf{0}$, a contradiction. ∎

For a code over $\mathbb{F}_q^t$ of length n and minimum Hamming distance d, the Singleton bound shows that

$$d\leqslant n+1-\log_{q^t}|\mathcal{C}|.$$

Codes that attain this bound with equality are called MDS codes. Obviously, the dimension k' (over $\mathbb{F}_q$) of an $\mathbb{F}_q$-linear MDS code is $k'=\log_q|\mathcal{C}|$. Since the Singleton bound is attained with equality, $\log_{q^t}|\mathcal{C}|$ must be an integer, and so, k' should be a multiple of t. We refer to such a code as an $\mathbb{F}_q$-linear $[n,k]$ MDS code over $\mathbb{F}_q^t$, where $k=\log_{q^t}|\mathcal{C}|=\frac{k'}{t}$. Obviously, every linear code over $\mathbb{F}_{q^t}$ can be treated as an $\mathbb{F}_q$-linear code over $\mathbb{F}_q^t$. In this paper, we require the following $\mathbb{F}_q$-linear systematic MDS code.

Lemma 2: Let $n\leqslant q^t+1$. Then for any $0\leqslant k\leqslant n$, there is an $\mathbb{F}_q$-linear $[n,k]$ MDS code over $\mathbb{F}_q^t$ with a generator matrix $G=\left(B_1\ B_2\cdots B_n\right)$ such that $\left(B_1\ B_2\cdots B_k\right)$ is an identity matrix, and each matrix B_i is a $kt\times t$ matrix over $\mathbb{F}_q$.

Proof: Fix a basis $\{\beta_1,\beta_2,\ldots,\beta_t\}$ of $\mathbb{F}_{q^t}$ over $\mathbb{F}_q$. Let $\phi:\mathbb{F}_{q^t}\to\mathbb{F}_q^t$ be the bijection that sends $x\in\mathbb{F}_{q^t}$ to its coordinates $(c_1,c_2,\ldots,c_t)$. Thus, we can identify the elements of $\mathbb{F}_{q^t}$ with those of $\mathbb{F}_q^t$. We extend the definition of ϕ to apply to sequences over $\mathbb{F}_{q^t}$ to obtain sequences over $\mathbb{F}_q^t$. For each $1\leqslant i\leqslant tk$, let $\mathbf{e}_i$ be the ith unit vector of length tk. We treat $\mathbf{e}_i$ as a vector of $(\mathbb{F}_q^t)^k$ and denote $\mathbf{u}_i=\phi^{-1}(\mathbf{e}_i)$. Since $n\leqslant q^t+1$, there is an $[n,k]$ systematic MDS code $\mathcal{C}$ over $\mathbb{F}_{q^t}$. Let $\mathbf{c}_i$ be the codeword of $\mathcal{C}$ whose prefix is $\mathbf{u}_i$. Let G be the matrix consisting of the vectors $\{\phi(\mathbf{c}_i)\ :\ 1\leqslant i\leqslant tk\}$. Then G is the desired matrix, and it generates an $\mathbb{F}_q$-linear $[n,k]$ MDS code over $\mathbb{F}_q^t$. ∎

In Table II, we summarize the notations in this paper and give a brief explanation of them.

III. SCALAR ERROR-CORRECTING CODES FOR COMBINATION NETWORKS AND THEIR SUB-NETWORKS

A. Combination Networks

In this subsection, we study error correction in combination networks. The structure of such networks is simple, yet sufficiently rich to exhibit interesting behavior. Many known examples that demonstrate the benefits of network coding have an identical or similar structure as that of combination networks, see [21] and the references therein. Due to their importance in both theory and practice, combination networks have been studied in various topics in network coding [3], [10], [13], [21], [23], [30], and also in coded caching [8], [16], [29], [31], [35].

A combination network $\mathcal{N}_{h,n,\alpha}$, which is shown in Fig. 1, is a network with a single source σ multicasting h messages. The source σ is connected to n nodes in the middle layer, which are indexed by the elements of $[n]\triangleq\{1,2,\ldots,n\}$. We have $\binom{n}{\alpha}$ sink nodes, each of which is connected to a unique subset of α nodes from the middle layer and requests all the h messages. We may index the sink nodes by α-subsets of $[n]$. There is a total of $n+\alpha\binom{n}{\alpha}$ links in $\mathcal{N}_{h,n,\alpha}$ and we denote this number by L.

For each $i\in[n]$, let $\mathbf{v}_i\in\mathbb{F}_q^h$ be the local encoding vector at the link which connects σ and the middle node indexed by i. Noting that each node in the middle layer has exactly one incoming link, we may assume that each node in the middle layer simply forwards the packet that it

TABLE II
NOTATION SUMMARY

Notation	Remark
$G=(\mathcal{V},\mathcal{E})$	a finite directed acyclic graph with a vertex set $\mathcal{V}$ and an edge set $\mathcal{E}$
$\mathrm{In}(i),\mathrm{Out}(i)$	the set of incoming/outgoing edges of node i
$\mathcal{C}$	a block code
$\mathcal{N}$	a network
$\mathscr{C}$	a network code
$\mathbf{f}_e$	global encoding vector of edge e
F_e	global encoding matrix of edge e
γ	a sink node
$G(\gamma)$	for scalar code, $G(\gamma)=(\mathbf{f}_e)_{e\in\mathrm{In}(\gamma)}$; for vector code, $G(\gamma)=(F_e)_{e\in\mathrm{In}(\gamma)}$
$\widetilde{\mathbf{f}}_e$	extended global encoding vector of edge e
$\widetilde{F}_e$	extended global encoding matrix of edge e
$\widetilde{G}(\gamma)$	for scalar code, $\widetilde{G}(\gamma)=(\widetilde{\mathbf{f}}_e)_{e\in\mathrm{In}(\gamma)}$; for vector code, $\widetilde{G}(\gamma)=\left(\widetilde{F}_e\right)_{e\in\mathrm{In}(\gamma)}$
$H(\gamma)$	a matrix such that $\widetilde{G}(\gamma)^{\intercal}=(G(\gamma)^{\intercal}\ H(\gamma)^{\intercal})$
$\Phi(\gamma)$	$\Phi(\gamma)=\left\{\mathbf{x}\cdot G(\gamma)\ :\ \mathbf{x}\in\mathbb{F}_q^h\right\}$
$\Delta(\gamma,\rho)$	$\Delta(\gamma,\rho)=\left\{\mathbf{z}\cdot H(\gamma)\ :\ \mathbf{z}\in\mathbb{F}_q^{\lvert\mathcal{E}\rvert},\mathrm{supp}(\mathbf{z})\subseteq\rho\right\}$
$d(\mathscr{C},\gamma)$	the minimum distance of a regular code $\mathscr{C}$ at receiver γ which is defined by $d(\mathscr{C},\gamma)\triangleq\min\{\lvert\rho\rvert\ :\ \Phi(\gamma)\cap\Delta(\gamma,\rho)\neq\{\mathbf{0}\}\}$
$q^{\mathrm{MDS}}(n,k)$	the minimum field size of an $[n,k]$-MDS code
$q^{\mathrm{MDS}}(\mathcal{N})$	the minimum field size of a scalar network MDS code for a given network $\mathcal{N}$
$q_{\mathrm{v}}^{\mathrm{MDS}}(\mathcal{N})$	the minimum value of q^t such that there is a vector network MDS code over $\mathbb{F}_q^t$ for a given network $\mathcal{N}$
$\mathrm{gap}^{\mathrm{MDS}}(\mathcal{N})$	the gap of $\mathcal{N}$ which is defined by $\mathrm{gap}^{\mathrm{MDS}}(\mathcal{N})\triangleq q^{\mathrm{MDS}}(\mathcal{N})-q_{\mathrm{v}}^{\mathrm{MDS}}(\mathcal{N})$
$\psi(x)$	the smallest prime power that is greater than or equal to x
$\psi_{\mathrm{even}}(x)$	the smallest even prime power that is greater than or equal to x
$\mathrm{wt}_H(\mathbf{x})$	the (block-wise) Hamming weight of a vector a vector $\mathbf{x}=(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n)$ of $(\mathbb{F}_q^t)^n$
$d_H(\mathbf{x},\mathbf{y})$	the (block-wise) Hamming distance of two vectors $\mathbf{x},\mathbf{y}$ of $(\mathbb{F}_q^t)^n$

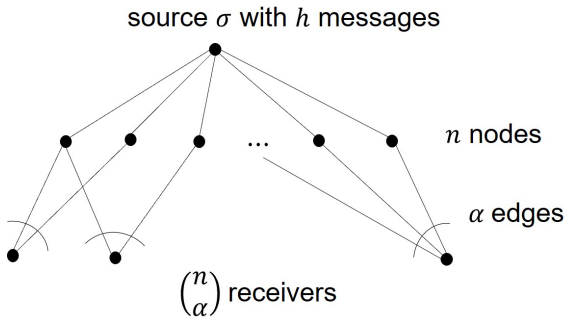

Fig. 1. A depiction of the combination network $\mathcal{N}_{h,n,\alpha}$.

receives to the receivers. Then for the receiver whose index is $\gamma=\{i_1,i_2,\ldots,i_\alpha\}$, the extended global encoding matrix is

$$\widetilde{G}(\gamma)=\begin{pmatrix}G(\gamma)\\H(\gamma)\end{pmatrix},$$

where $G(\gamma)=(\mathbf{v}_{i_1}\ \mathbf{v}_{i_2}\ \cdots\ \mathbf{v}_{i_\alpha})\in\mathbb{F}_q^{h\times\alpha}$, and $H(\gamma)\in\mathbb{F}_q^{L\times\alpha}$. The matrix $H(\gamma)$ has the following two properties:

(i) The rows indexed by the links of $\mathrm{In}(\gamma)$ constitute an identity matrix.

(ii) Each row of $H(\gamma)$ has at most one non-zero entry. This is because each link e in the combination network is connected to at most one incoming link of γ, and so the error z_e is propagated to at most one link in $\mathrm{In}(\gamma)$.

Now, let $\mathcal{C}(\gamma)$ be the linear code over $\mathbb{F}_q$ which is generated by the matrix $G(\gamma)$. The following result reveals the relation between the minimum Hamming distance of $\mathcal{C}(\gamma)$ and the minimum distance $d(\mathscr{C},\gamma)$.

Lemma 3: Let $\mathscr{C}$ be a scalar network code for the combination network $\mathcal{N}_{h,n,\alpha}$. Then

$$d(\mathscr{C},\gamma)=d_H(\mathcal{C}(\gamma)).$$

Proof: Recall from (2) that

$$\Delta(\gamma,\rho)=\left\{\mathbf{z}\cdot H(\gamma)\ :\ \mathbf{z}\in\mathbb{F}_q^{\lvert\mathcal{E}\rvert},\mathrm{supp}(\mathbf{z})\subseteq\rho\right\}.$$

Recalling from (1) that

$$\Phi(\gamma)=\left\{\mathbf{x}\cdot G(\gamma)\ :\ \mathbf{x}\in\mathbb{F}_q^h\right\},$$

let $\mathbf{u}_0$ be a non-zero vector of $\Phi(\gamma)$ which has the smallest Hamming weight. Since $H(\gamma)$ contains an identity matrix I_α as a submatrix, there are $\mathrm{wt}_H(\mathbf{u}_0)$ rows of $H(\gamma)$ which can generate a space containing $\mathbf{u}_0$. Thus, we have that

$$\begin{aligned}&\min\{\lvert\rho\rvert\ :\ \Phi(\gamma)\cap\Delta(\gamma,\rho)\neq\{\mathbf{0}\}\}\\&\leqslant\mathrm{wt}_H(\mathbf{u}_0)=\min\{\mathrm{wt}_H(\mathbf{u})\ :\ \mathbf{u}\in\Phi(\gamma)\setminus\{\mathbf{0}\}\}.\end{aligned}$$

On the other hand, let ρ_0 be the error pattern of $\{\rho\ :\ \Phi(\gamma)\cap\Delta(\gamma,\rho)\neq\{\mathbf{0}\}\}$ which has the smallest size, and let $\mathbf{v}_0$ be an arbitrary non-zero vector of $\Phi(\gamma)\cap\Delta(\gamma,\rho_0)$. Since each row of $H(\gamma)$ has at most one non-zero entry, we have that $\lvert\rho_0\rvert\geqslant\mathrm{wt}_H(\mathbf{v}_0)$, and so,

$$\begin{aligned}&\min\{\lvert\rho\rvert\ :\ \Phi(\gamma)\cap\Delta(\gamma,\rho)\neq\{\mathbf{0}\}\}\\&=\lvert\rho_0\rvert\geqslant\mathrm{wt}_H(\mathbf{v}_0)\geqslant\min\{\mathrm{wt}_H(\mathbf{u})\ :\ \mathbf{u}\in\Phi(\gamma)\setminus\{\mathbf{0}\}\}.\end{aligned}$$

Combining the two inequalities above, we have that

$$\begin{aligned}&\min\{|\rho| \ : \ \Phi(\gamma) \cap \Delta(\gamma, \rho) \neq \{\mathbf{0}\}\}\\ =&\min\{\mathrm{wt}_H(\mathbf{u}) \ : \ \mathbf{u} \in \Phi(\gamma) \setminus \{\mathbf{0}\}\}.\end{aligned}$$

Note that $\Phi(\gamma)$ consists of all the codewords of $\mathcal{C}(\gamma)$. Hence,

$$\begin{aligned}d(\mathscr{C}, \gamma) &= \min\{|\rho| \ : \ \Phi(\gamma) \cap \Delta(\gamma, \rho) \neq \{\mathbf{0}\}\}\\ &= \min\{\mathrm{wt}_H(\mathbf{u}) \ : \ \mathbf{u} \in \Phi(\gamma) \setminus \{\mathbf{0}\}\}\\ &= \min\{\mathrm{wt}_H(\mathbf{u}) \ : \ \mathbf{u} \in \mathcal{C}(\gamma) \setminus \{\mathbf{0}\}\}\\ &= d_H(\mathcal{C}(\gamma)).\end{aligned}$$

■

Now, let $\mathcal{C}$ be the linear code generated by the matrix

$$G \triangleq (\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n),$$

where $\mathbf{v}_i$ is the global encoding vector for the edge (σ, i).

Theorem 3: Let $\mathscr{C}$ be a scalar network code over $\mathbb{F}_q$ for the combination network $\mathcal{N}_{h,n,\alpha}$, and assume $1 \leqslant d \leqslant \alpha$ is some integer. Then $\mathscr{C}$ is regular and $d(\mathscr{C}, \gamma) \geqslant d$ for every receiver $\gamma \in R$ if and only if the code $\mathcal{C}$ is an $[n, h, \geqslant n - \alpha + d]_q$ code.

Proof: For the only if part, let $\mathbf{v}_{i_1}, \mathbf{v}_{i_2}, \ldots, \mathbf{v}_{i_{\alpha-d+1}}$ be arbitrary $\alpha - d + 1$ columns of G. Then there is a $\gamma \in R$ such that $G(\gamma)$ contains these columns. Since $\mathscr{C}$ is regular, the rank of $G(\gamma)$ is h. Additionally, $d(\mathscr{C}, \gamma) \geqslant d$, and so according to Lemma 3, $\mathcal{C}(\gamma)$ has minimum distance at at least d. Hence, $\mathcal{C}(\gamma)$ is an $[\alpha, h, \geqslant d]_q$ code. Due to the minimum distance of $\mathcal{C}(\gamma)$, it follows that the column-space of the $\alpha - d + 1$ columns $\mathbf{v}_{i_1}, \mathbf{v}_{i_2}, \ldots, \mathbf{v}_{i_{\alpha-d+1}}$ is the entire space $\mathbb{F}_q^h$. Because these columns are chosen arbitrarily, $\mathcal{C}$ has Hamming distance at least $n - (\alpha - d + 1) + 1 = n - \alpha + d$.

For the if part, since $\mathcal{C}$ is an $[n, h, \geqslant n - \alpha + d]_q$ code, any $\alpha - d + 1$ columns of G can span the whole space $\mathbb{F}_q^h$. Noting that for each $\gamma \in R$ the generator matrix $G(\gamma)$ is a submatrix of G, any $\alpha - d + 1$ columns of $G(\gamma)$ also generate the space $\mathbb{F}_q^h$. Thus $\mathscr{C}$ is regular and each $\mathcal{C}(\gamma)$ has minimum Hamming distance at least d. According to Lemma 3, we have that $d(\mathscr{C}, \gamma) \geqslant d$ for every receiver $\gamma \in R$. ■

Corollary 1: Let $\mathscr{C}$ be a scalar network code over $\mathbb{F}_q$ for the combination network $\mathcal{N}_{h,n,\alpha}$. Then $\mathscr{C}$ is MDS if and only if $\mathcal{C}$ is an $[n, h]_q$ MDS code. In particular, we have that

$$q^{\mathrm{MDS}}(\mathcal{N}_{h,n,\alpha}) = q^{\mathrm{MDS}}(n, h) \leqslant \psi(n - 1).$$

Proof: For each receiver of the combination network $\mathcal{N}_{h,n,\alpha}$, the minimum cut capacity is α. Thus the network code $\mathscr{C}$ is MDS if and only if $d(\mathscr{C}, \gamma) = \alpha - h + 1$ for every $\gamma \in R$. According to Theorem 3, this is equivalent to $\mathcal{C}$ being an $[n, h, n - h + 1]_q$ code, which is also MDS. ■

Let us illustrate the power of our result with the following example.

Example 1: According to the corollary above, we have a scalar MDS code for the combination network $\mathcal{N}_{h,n,\alpha}$ whenever $q \geqslant n - 1$. As an example, for $n = 6$ and $\alpha = 4$, a field of size 5 suffices for an MDS network code. In contrast, the method proposed in [11], which is used to construct a network MDS code for a general network topology, requires the field size to be at least $\psi(360) = 361$ (see [11, Example 1]).

B. Sub-Networks of the Combination Network

In this subsection, we consider sub-networks of the combination network obtained by removing some receivers as well as possibly some links connecting the middle layer to the remaining receivers. It has been shown in [3] that a class of such networks has the maximum gap in the field sizes between scalar and linear solutions among minimal multicast networks with two messages. In Section V of this paper, we shall use these networks to demonstrate that the field size of vector network MDS codes could be substantially smaller than that of the scalar network MDS codes.

Let $\mathcal{S}$ be a sub-network of $\mathcal{N}_{h,n,\alpha}$. Assume that $\mathcal{S}$ has m receivers. Let $H(\mathcal{S})$ be a hypergraph consisting of n vertices and m hyperedges, where the vertices are indexed by the n middle nodes of $\mathcal{S}$, and each hyperedge of $H(\mathcal{S})$ corresponds to a receiver γ and contains all the vertices that are indexed by the middle nodes connected to γ.

A *strong vertex coloring* of a hypergraph H is an assignment of colors to the vertices of H so that the vertices of each edge are assigned distinct colors. The minimum number of colors that allows a strong vertex coloring of H is called the *strong chromatic number* of H, denoted by $\chi(H)$.

Lemma 4: Let $\mathcal{S}$ be a sub-network of the combination network $\mathcal{N}_{h,n,\alpha}$. Assume that each receiver of $\mathcal{S}$ is connected to at least h middle nodes. Then

$$q^{\mathrm{MDS}}(\mathcal{S}) \leqslant q^{\mathrm{MDS}}(\chi(H(\mathcal{S})), h).$$

Proof: To avoid tedious notation, we denote $n' \triangleq \chi(H(\mathcal{S}))$ and $q' \triangleq q^{\mathrm{MDS}}(\chi(H(\mathcal{S})), h)$. Let $(\mathbf{g}_1 \ \mathbf{g}_2 \cdots \mathbf{g}_{n'})$ be the generator matrix of an $[n', h]_{q'}$ MDS code. Let $c : [n] \longrightarrow [n']$ be a strong coloring of $H(\mathcal{S})$. Let $\mathbf{v}_i = \mathbf{g}_{c(i)}$ be the encoding vector for the link from the source node to the middle node indexed by i. Since the middle nodes simply forward the packets, for each receiver γ, the matrix $G(\gamma)$ consists of C_γ distinct vectors of $\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_{n'}\}$, where C_γ is equal to the size of the hyperedge that is indexed by γ. Thus, the code generated by $G(\gamma)$ is an MDS code with minimum Hamming distance $C_\gamma - h + 1$. Hence, according to Lemma 3, the encoding vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ define a network MDS code over $\mathbb{F}_{q'}$ for $\mathcal{S}$. ■

IV. Scalar Error-Correcting Codes for Zosin-Khuller Networks

Let m, N be positive integers such that $2 \leqslant m \leqslant N$. Denote $\mathcal{A} \triangleq \binom{[N]}{m-1}$, the set of $(m-1)$-subsets of $[N]$, and similarly, denote $\mathcal{B} \triangleq \binom{[N]}{m}$. We consider a class of layered acyclic networks which are defined as follows: the networks consist of five layers and the source node σ transmits h messages to N receivers through three layers of nodes, which are called A-, B- and C-nodes. A-nodes are indexed by elements of $\mathcal{A}$, and B- and C-nodes are indexed by elements of $\mathcal{B}$. The source σ is connected to each of the A-nodes. An A-node is connected to a B-node if the index of A is a subset of the index of B. A B-node is connected to a C-node if their indices are the same. A receiver i is connected to a C-node if the index of C contains i. An example is shown in Figure 2. This class of networks, denoted as $\mathcal{Z}_{h,m,N}$, was originally described by Zosin and Khuller in [37] to demonstrate the

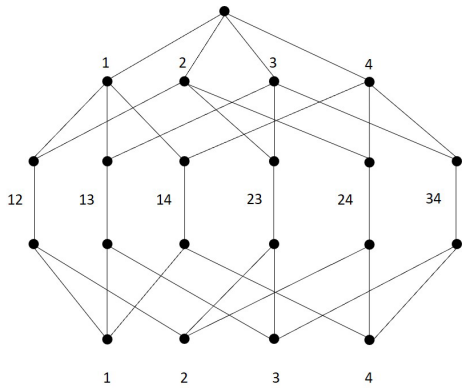

Fig. 2. A depiction of the network ZK(2,4).

integrality gap of a standard linear-programming formulation for the directed Steiner tree problem, and was also examined in [7] to demonstrate the throughput benefits that network coding offers compared with routing.

The following result shows that the capacity of the minimum cut at each receiver is $\binom{N-1}{m-1}$.

Proposition 1 ([7]): There are exactly $\binom{N-1}{m-1}$ edge disjoint paths between the source and each receiver.

It is worth noting that regular network codes for $\mathcal{Z}_{h,m,N}$ have been constructed in [7] over the binary field. In contrast, when error-correction is required, the following lower bound shows that a large field size is needed.

Lemma 5: Let m, N be positive integers such that $2 \leqslant m \leqslant N$. Denote $C = \binom{N-1}{m-1}$. Let h be a positive integer such that $h \leqslant C$. Then

$$q^{\mathrm{MDS}}(\mathcal{Z}_{h,m,N}) \geqslant q^{\mathrm{MDS}}(C,h).$$

Proof: Let $\mathscr{C}$ be an MDS code over $\mathbb{F}_q$ for the network $\mathcal{Z}_{h,m,N}$. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_C$ be global encoding vectors for the incoming links of the receiver indexed by 1. Since $d(\mathscr{C}, 1) = C - h + 1$, it is easy to check that the code generated by the matrix $(\mathbf{v}_1\ \mathbf{v}_2 \cdots \mathbf{v}_C)$ is a $[C,h]_q$ MDS code. ∎

In the following two subsections, we study the cases of $m = 2$ and $m > N/2$, and we present two classes of network MDS codes for $\mathcal{Z}_{h,m,N}$ with field size $q \leqslant 2N$ or $q \leqslant 2\binom{N-1}{m-1}$, respectively.

A. The Case of $m = 2$

In this case, $\binom{N-1}{m-1} = N - 1$ and $\mathcal{A} = [N]$. For each $1 \leqslant i \leqslant N$, let $\mathbf{v}_i \in \mathbb{F}_q^h$ be the local encoding vector of the link from σ to the A-node indexed by i. Each B-node has two incoming links. In our coding scheme, each B-node sums its two incoming packets and transmits the sum to the corresponding C-node, and the C-node simply forwards this sum to the receiver. We denote the network code described above as $\mathscr{C}$.

For each receiver i, the extended global encoding matrix is

$$\widetilde{G}(i) = \begin{pmatrix} G(i) \\ H(i) \end{pmatrix},$$

where $G(i) = (\mathbf{v}_i + \mathbf{v}_j)_{1 \leqslant j \leqslant N, j \neq i}$, and where $H(i)$ has the following two properties:

(i) The rows indexed by the links of $\mathrm{In}(i)$ constitute an identity matrix.

(ii) The row indexed by the link from σ to the A-node indexed by i is an all-one vector, and each of the other rows has at most one non-zero entry.

Now, let $\mathcal{C}(i)$ be the linear code over $\mathbb{F}_q$ which is generated by the matrix $G(i)$.

Lemma 6: If $d_H(\mathcal{C}(i)) \geqslant d$ and each non-zero element of $\mathbb{F}_q$ appears in each codeword of $\mathcal{C}(i)$ at most $N-d$ times, then

$$d(\mathscr{C}, i) \geqslant d.$$

Proof: Recall the definitions of $\Phi(i)$ and $\Delta(i,\rho)$ from (1), (2), and that $d(\mathscr{C}, i) = \min\{|\rho| \ : \ \Phi(i) \cap \Delta(i,\rho) \neq \{\mathbf{0}\}\}$. Assume to the contrary that there is an error pattern ρ with $|\rho| < d$ such that $\Phi(i) \cap \Delta(i,\rho) \neq \{\mathbf{0}\}$. Denote the link from σ to the A-node indexed by i as e_i. Since $d_H(\mathcal{C}(i)) \geqslant d$ and every row of $H(i)$ except the one indexed by e_i has at most one non-zero entry, necessarily $e_i \in \rho$. Therefore, every non-zero vector of $\Phi(i) \cap \Delta(i,\rho)$ is generated by the row indexed by e_i and some other $|\rho| - 1$ rows, and then contains a non-zero element of $\mathbb{F}_q$ occurring at least $N-1-(|\rho|-1) > N - d$ times, which contradicts the assumption. ∎

Now, let $q \geqslant N$ be an even prime power, and let $a_1, a_2, \ldots, a_N$ be N elements of $\mathbb{F}_q$. For each $1 \leqslant i \leqslant N$, let

$$\mathbf{v}_i = (a_i, a_i^2, a_i^3, \ldots, a_i^h)^{\mathsf{T}}.$$

Without loss of generality, we consider the determinant of the matrix $(\mathbf{v}_1 + \mathbf{v}_i)_{2 \leqslant i \leqslant h+1}$. If $a_1 = 0$, then $a_i \neq 0$ for all $2 \leqslant i \leqslant h+1$ and

$$\begin{vmatrix} a_1+a_2 & a_1+a_3 & \cdots & a_1+a_{h+1} \\ a_1^2+a_2^2 & a_1^2+a_3^2 & \cdots & a_1^2+a_{h+1}^2 \\ a_1^3+a_2^3 & a_1^3+a_3^3 & \cdots & a_1^3+a_{h+1}^3 \\ \vdots & \vdots & & \vdots \\ a_1^h+a_2^h & a_1^h+a_3^h & \cdots & a_1^h+a_{h+1}^h \end{vmatrix} = \begin{vmatrix} a_2 & a_3 & \cdots & a_{h+1} \\ a_2^2 & a_3^2 & \cdots & a_{h+1}^2 \\ a_2^3 & a_3^3 & \cdots & a_{h+1}^3 \\ \vdots & \vdots & & \vdots \\ a_2^h & a_3^h & \cdots & a_{h+1}^h \end{vmatrix} \neq 0.$$

If $a_1 \neq 0$, we have that

$$\begin{vmatrix} a_1+a_2 & a_1+a_3 & \cdots & a_1+a_{h+1} \\ a_1^2+a_2^2 & a_1^2+a_3^2 & \cdots & a_1^2+a_{h+1}^2 \\ a_1^3+a_2^3 & a_1^3+a_3^3 & \cdots & a_1^3+a_{h+1}^3 \\ \vdots & \vdots & & \vdots \\ a_1^h+a_2^h & a_1^h+a_3^h & \cdots & a_1^h+a_{h+1}^h \end{vmatrix} = \begin{vmatrix} a_1+a_2 & a_1+a_3 & \cdots & a_1+a_{h+1} \\ a_2^2-a_2a_1 & a_3^2-a_3a_1 & \cdots & a_{h+1}^2-a_{h+1}a_1 \\ a_2^3-a_2^2a_1 & a_3^3-a_3^2a_1 & \cdots & a_{h+1}^3-a_{h+1}^2a_1 \\ \vdots & \vdots & & \vdots \\ a_2^h-a_2^{h-1}a_1 & a_3^h-a_3^{h-1}a_1 & \cdots & a_{h+1}^h-a_{h+1}^{h-1}a_1 \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 1 & \cdots & 1 \\ a_2 & a_3 & \cdots & a_{h+1} \\ a_2^2 & a_3^2 & \cdots & a_{h+1}^2 \\ \vdots & \vdots & & \vdots \\ a_2^{h-1} & a_3^{h-1} & \cdots & a_{h+1}^{h-1} \end{vmatrix} \cdot \left(\prod_{1<j\leqslant h+1} (a_j - a_1) \right)$$

$$= \prod_{1\leqslant i<j\leqslant h+1} (a_j - a_i) \neq 0.$$

Hence, each code $\mathcal{C}(i)$ generated by the matrix $G(i) = (\mathbf{v}_i + \mathbf{v}_j)_{1\leqslant j\leqslant N, j\neq i}$ is an $[N-1, h, N-h]$ MDS code. Furthermore, for each non-zero codeword $(c_1, c_2, \ldots, c_h)G(i)$ and each non-zero element $a \in \mathbb{F}_q$, since the polynomial $c_1(a_i + x) + c_2(a_i^2 + x^2) + \cdots c_h(a_i^h + x^h) = a$ has at most h roots, a appears in the non-zero codeword at most h times. According to Lemma 6, the network $\mathscr{C}$ has $d(\mathscr{C}, i) \geqslant N - h$, namely, $\mathscr{C}$ is an MDS code. Therefore, we have the following result.

Theorem 4: Let N, h be positive integers such that $h \leqslant N-1$. Then

$$q^{\mathrm{MDS}}(\mathcal{Z}_{h,2,N}) \leqslant \psi_{\mathrm{even}}(N),$$

where $\psi_{\mathrm{even}}(N)$ is the smallest even prime power that is greater than or equal to N.

To compare this result to the best known bound we refer to (3). While our result guarantees an MDS network code for $\mathcal{Z}_{h,2,N}$ with a field of size $\psi_{\mathrm{even}}(N)$, the known guarantee of (3), gives a field of size at least $\psi(N\binom{N-1}{h})$.

B. The Case of $m > N/2$

Since $m > N/2$, we have that $|\mathcal{A}| \geqslant |\mathcal{B}|$. Consider the induced subgraph of $\mathcal{Z}_{h,m,N}$ where the vertices are the A-nodes and B-nodes. This graph is a bipartite graph, and an A-node is connected to a B-node if its index is a subset of the index of the B-node. So, each A-node has degree $N-m+1$ and each B-node has degree m. Since $N-m+1 \leqslant m$, the condition of the Hall's marriage theorem is satisfied. Hence, there is a matching which covers all the B-nodes. Denote $\mathcal{A}'$ the set of A-nodes covered by this matching.

We consider the sub-network $\mathcal{S}$ which is obtained from $\mathcal{Z}_{h,m,N}$ by removing all the nodes indexed by the elements in $\mathcal{A}\setminus\mathcal{A}'$ and removing all the links connected to these nodes. In this sub-network, the numbers of A-, B- and C-nodes all are equal to $\binom{N}{m}$. There are $\binom{N}{m}$ edge disjoint paths connecting the source σ and the C-nodes and each receiver i is connected to $\binom{N-1}{m-1}$ C-nodes indexed by the elements of $\mathcal{B}$ containing i. Since there are still $\binom{N-1}{m-1}$ edge disjoint paths between the source and each receiver, we have that

$$q^{\mathrm{MDS}}(\mathcal{Z}_{h,m,N}) \leqslant q^{\mathrm{MDS}}(\mathcal{S}). \tag{4}$$

Now, let $n = \binom{N}{m}$ and $\alpha = \binom{N-1}{m-1}$. Consider a sub-network $\mathcal{S}'$ of the combination network $\mathcal{N}_{h,n,\alpha}$: In $\mathcal{S}'$, the middle nodes are indexed by the elements of $\mathcal{B}$, the receivers are indexed by the elements of $[N]$, and a receiver i is connected to the middle nodes that are indexed by the elements of $\mathcal{B}$ containing i. Thus, the network $\mathcal{S}$ can be obtained from $\mathcal{S}'$ by replacing each link between the source and the middle node with a path of length 3. It follows that

$$q^{\mathrm{MDS}}(\mathcal{S}) \leqslant q^{\mathrm{MDS}}(\mathcal{S}'). \tag{5}$$

Combining (4) and (5), together with Lemma 4, we have

$$q^{\mathrm{MDS}}(\mathcal{Z}_{h,m,N}) \leqslant \psi(\chi(H(S')) - 1).$$

Since $m > N/2$, in the hypergraph $H(S')$, every two vertices are contained in an edge. So, the strong chromatic number

$$\chi(H(S')) = n = \frac{N}{m}\binom{N-1}{m-1} < 2\binom{N-1}{m-1}.$$

Hence, we have the following result.

Theorem 5: Let h, m, N be positive integers such that $m > N/2$ and $h \leqslant \binom{N-1}{m-1}$. Then

$$q^{\mathrm{MDS}}(\mathcal{Z}_{h,m,N}) \leqslant \psi\left(2\binom{N-1}{m-1} - 1\right).$$

By contrast, according to (3), the previous methods require field size to be larger than $N\binom{\binom{N-1}{m-1}}{h}$.

We note that the method in this subsection doesn't work for the case of $m \leqslant N/2$. In this case, we have that $|\mathcal{A}| < |\mathcal{B}|$. In order to obtain a sub-network which can be related to combination networks, one has to remove some B-nodes, as well as some C-nodes. Then the capacity of minimum cut at some receivers decreases. So, even if there exists an MDS code for this sub-network, it cannot garantee that there exist an MDS code for $\mathcal{Z}_{h,m,N}$.

V. Vector Network Error-Correcting Codes

A. Definition and Singleton-Like Bound

In vector network coding, if there is an error in a link e, its head node receives the packet $\widetilde{\mathbf{u}}_e = \mathbf{u}_e + \mathbf{z}_e$, where $\mathbf{u}_e \in \mathbb{F}_q^t$ is the packet that is supposed to be transmitted by e and $\mathbf{z}_e \in \mathbb{F}_q^t$ is the error. The error message vector $\mathbf{z} = (\mathbf{z}_e)_{e\in\mathcal{E}}$ is then a vector of length $t|\mathcal{E}|$. The packet $\widetilde{\mathbf{u}}_e$ can be written as

$$\widetilde{\mathbf{u}}_e = (\mathbf{x}, \mathbf{z}) \cdot \widetilde{F}_e$$

where $\widetilde{F}_e \in \mathbb{F}_q^{(th+t|\mathcal{E}|)\times t}$.

For each receiver $\gamma \in R$, the extended global encoding matrix $\widetilde{G}(\gamma) \triangleq (\widetilde{F}_e)_{e\in\mathrm{In}(\gamma)}$ can be written as

$$\widetilde{G}(\gamma) = \begin{pmatrix} G(\gamma) \\ H(\gamma) \end{pmatrix},$$

where $G(\gamma) = (F_e)_{e\in\mathrm{In}(\gamma)} \in \mathbb{F}_q^{th\times t|\mathrm{In}(\gamma)|}$ and $H(\gamma) \in \mathbb{F}_q^{t|\mathcal{E}|\times t|\mathrm{In}(\gamma)|}$. Again, here we use the same notations $\widetilde{G}(\gamma)$ and $H(\gamma)$ as those of the scalar case for the simplicity.

Like the scalar case, the minimum distance of a regular vector code $\mathscr{C}$ at receiver γ is defined by

$$d(\mathscr{C}, \gamma) \triangleq \min\{|\rho| \ : \ \Phi(\gamma) \cap \Delta(\gamma, \rho) \neq \{\mathbf{0}\}\},$$

where

$$\Phi(\gamma) = \{\mathbf{x} \cdot G(\gamma) \ : \ \mathbf{x} \in \mathbb{F}_q^{th}\}$$

and

$$\Delta(\gamma, \rho) = \{\mathbf{z} \cdot H(\gamma) \ : \ \mathbf{z} \in \mathbb{F}_q^{t|\mathcal{E}|}, \ \mathrm{supp}(\mathbf{z}) \subseteq \rho\}.$$

Since

$$\begin{aligned} &d(\mathscr{C}, \gamma) \\ &= \min\{|\rho| \ : \ \Phi(\gamma) \cap \Delta(\gamma, \rho) \neq \{\mathbf{0}\}\} \\ &= \min\big\{|\rho| \ : \ \exists\ \mathbf{x}, \mathbf{z} \text{ with } \mathbf{x} \neq \mathbf{0}, \ \mathrm{supp}(\mathbf{z}) \subseteq \rho, \text{ s.t.} \\ &\qquad \mathbf{x} \cdot G(\gamma) = \mathbf{z} \cdot H(\gamma)\big\} \\ &= \min\{|\rho| \ : \ \exists\ \mathbf{x}, \mathbf{x}', \mathbf{z}, \mathbf{z}' \text{ with } \mathbf{x} \neq \mathbf{x}', \ \mathrm{supp}(\mathbf{z} - \mathbf{z}') \subseteq \rho \text{ s.t.} \\ &\qquad (\mathbf{x}, \mathbf{z}) \cdot \widetilde{G}(\gamma) = (\mathbf{x}', \mathbf{z}') \cdot \widetilde{G}(\gamma)\}, \end{aligned}$$

the vector network code $\mathscr{C}$ can correct up to $\lfloor (d(\mathscr{C},\gamma)-1)/2 \rfloor$ link errors.

We have the following Singleton-type bound, the proof of which is exactly the same as that of the scalar case described in [11, Lemma 1], which we bring here for completeness.

Lemma 7: Let $\mathscr{C}$ be a regular vector network code. Then for each receiver γ we have that

$$d(\mathscr{C},\gamma) \leqslant C_\gamma - h + 1.$$

Proof: Let $\{e_1, e_2, \ldots, e_{C_\gamma}\}$ be a minimum cut between σ and γ with an upstream-to-downstream order and let $\rho = \{e_h, e_{h+1}, \ldots, e_{C_\gamma}\}$ be an error pattern. We are going to show that there is a non-zero message vector $\mathbf{x} \in \mathbb{F}_q^{ht}$ and an error message vector $\mathbf{z} \in \mathbb{F}_q^{|\mathcal{E}|t}$ matching ρ such that $\mathbf{x}G(\gamma) = \mathbf{z}H(\gamma)$, or equivalently, $(\mathbf{x},\mathbf{0})\widetilde{G}(\gamma) = (\mathbf{0},\mathbf{z})\widetilde{G}(\gamma)$. Note that the condition $\mathscr{C}$ is regular ensures that $\mathbf{x}G(\gamma)$ is non-zero.

Noting that the matrix $\left(\widetilde{F}_{e_1}\ \widetilde{F}_{e_2} \cdots \widetilde{F}_{e_{h-1}}\right)$ has rank at most $(h-1)t$, there is a non-zero message vector $\mathbf{x} \in \mathbb{F}_q^{ht}$ such that $(\mathbf{x},\mathbf{0})\left(\widetilde{F}_{e_1}\ \widetilde{F}_{e_2} \cdots \widetilde{F}_{e_{h-1}}\right) = \mathbf{0}$. Denote $\widetilde{\mathbf{u}}_{e_i} \triangleq (\mathbf{x},\mathbf{0})\widetilde{F}_{e_i}$ for each $h \leqslant i \leqslant C_\gamma$. Now, we construct the error message vector $\mathbf{z}$. For $e \notin \{e_h, e_{h+1}, \ldots, e_{C_\gamma}\}$, set $\mathbf{z}_e = \mathbf{0}$; for each $h \leqslant i \leqslant C_\gamma$, we imagine that the source σ transmits the all-zero message vector and let $\mathbf{z}_{e_i} = \widetilde{\mathbf{u}}_{e_i} - \sum_{e \in \mathrm{In}(\mathrm{tail}(e_i))} K_{e,e_i}\widetilde{\mathbf{u}}'_e$, where $\widetilde{\mathbf{u}}'_e$ is the output of the channel e in this case. Thus, we have that $\mathbf{z} = (\mathbf{z}_e)_{e\in\mathcal{E}}$ matches ρ, and

$$(\mathbf{0},\mathbf{z})\widetilde{F}_{e_i} = \begin{cases} \mathbf{0} & \text{if } 1 \leqslant i \leqslant h-1, \\ \widetilde{\mathbf{u}}_{e_i} & \text{if } h \leqslant i \leqslant C_\gamma. \end{cases}$$

That is,

$$(\mathbf{x},\mathbf{0})(\widetilde{F}_e)_{e\in\{e_1,e_2,\ldots,e_{C_\gamma}\}} = (\mathbf{0},\mathbf{z})(\widetilde{F}_e)_{e\in\{e_1,e_2,\ldots,e_{C_\gamma}\}}.$$

Since $\{e_1, e_2, \ldots, e_{C_\gamma}\}$ is a cut between σ and γ, it follows that $(\mathbf{x},\mathbf{0})\widetilde{G}_\gamma = (\mathbf{0},\mathbf{z})\widetilde{G}_\gamma$. ∎

For a regular vector network code $\mathcal{C}$, if $d_{\min}(\mathcal{C},\gamma) = C_\gamma - h + 1$ for every receiver γ, it is called a vector network MDS code. For a network $\mathcal{N}$, let $q_{\mathrm{v}}^{\mathrm{MDS}}(\mathcal{N})$ denote the minimum value of q^t such that there is a vector network MDS code over $\mathbb{F}_q^t$. Note that this notation takes the vector length t into account: the logarithm $\log(q_{\mathrm{v}}^{\mathrm{MDS}}(\mathcal{N}))$ is equal to the logarithm of the field size multiplied by the vector length.

Recall that $q^{\mathrm{MDS}}(\mathcal{N})$ is the minimum q such that there is a scalar network MDS code over $\mathbb{F}_q$. For a network $\mathcal{N}$, suppose that the minimum field size of a scalar network MDS code is p^t, i.e., $q^{\mathrm{MDS}}(\mathcal{N}) = p^t$. A scalar network code over $\mathbb{F}_{p^t}$ can be recast as a vector network code over $\mathbb{F}_p$ with vector length t. Thus, we have $q_{\mathrm{v}}^{\mathrm{MDS}}(\mathcal{N}) \leqslant p^t = q^{\mathrm{MDS}}(\mathcal{N})$. Moreover, since we have more freedom to choose the coding coefficients in vector network coding than in scalar network coding, $q_{\mathrm{v}}^{\mathrm{MDS}}(\mathcal{N})$ might be strictly smaller than $q^{\mathrm{MDS}}(\mathcal{N})$. We define the *gap* of $\mathcal{N}$ as

$$\mathrm{gap}^{\mathrm{MDS}}(\mathcal{N}) \triangleq q^{\mathrm{MDS}}(\mathcal{N}) - q_{\mathrm{v}}^{\mathrm{MDS}}(\mathcal{N}).$$

We shall demonstrate a class of networks $\mathcal{N}$ with a large gap.

B. Combination Networks

Now, we consider the vector coding for combination networks. For each $1 \leqslant i \leqslant n$, let $V_i \in \mathbb{F}_q^{th\times t}$ be the local encoding matrix of the link from the source σ to the node indexed by i in the middle layer. Again, we may assume that the nodes in the middle layer simply forward the packet that it receives to the receivers. For the receiver $\gamma = \{i_1, i_2, \ldots, i_\alpha\}$, the extended global encoding matrix

$$\widetilde{G}(\gamma) = \begin{pmatrix} G(\gamma) \\ H(\gamma) \end{pmatrix},$$

where $G(\gamma) = \begin{pmatrix} V_{i_1} & V_{i_2} & \cdots & V_{i_\alpha} \end{pmatrix}$, and $H(\gamma)$ is a $tL \times t\alpha$ matrix. We write $H(\gamma)$ as $H(\gamma) = (H_{ij})_{i\in[L],j\in[\alpha]}$, where each H_{ij} is a $t \times t$ matrix. Similar to the scalar case, we have the following two observations:

(i) There is a subset $A \subseteq [L]$ with $|A| = \alpha$ such that $(H_{ij})_{i\in A, j\in[\alpha]}$ is an identity matrix.
(ii) For each $i \in [L]$, there is at most one H_{ij} which is not the all-zero matrix.

Let $\mathcal{C}(\gamma)$ be the $\mathbb{F}_q$-linear code over $\mathbb{F}_q^t$ which is generated by the matrix $G(\gamma)$ and $\mathcal{C}$ be the $\mathbb{F}_q$-linear code over $\mathbb{F}_q^t$ generated by the matrix $G \triangleq \begin{pmatrix} V_1 & V_2 & \cdots & V_n \end{pmatrix}$. The following results are analogs of Lemma 3 and Theorem 3. The proofs are the same and we omit them.

Lemma 8: Let $\mathscr{C}$ be a vector network code for the combination network. Then

$$d(\mathscr{C},\gamma) = d_H(\mathcal{C}(\gamma)).$$

Theorem 6: Let $\mathscr{C}$ be a network code over $\mathbb{F}_q^t$ for the combination network. Then $\mathscr{C}$ is regular and $d(\mathscr{C},\gamma) \geqslant d$ for every receiver $\gamma \in R$ if and only if the $\mathbb{F}_q$-linear code $\mathcal{C}$ has dimension th and minimum Hamming distance $n-\alpha+d$.

The following theorem shows the equivalence between the MDS code for the combination network $\mathcal{N}_{h,n,\alpha}$ and the regular code for the minimal combination network $\mathcal{N}_{h,n,h}$.

Theorem 7: There is an MDS code over $\mathbb{F}_q^t$ for the combination network $\mathcal{N}_{h,n,\alpha}$ if and only if there is a regular network code over the same alphabet for the minimal combination network $\mathcal{N}_{h,n,h}$.

Proof: Note that in both networks, $\mathcal{N}_{h,n,\alpha}$ and $\mathcal{N}_{h,n,h}$, there is one source node and n middle nodes. Let $V_1, V_2, \ldots, V_n$ be the local encoding matrices for the links from the source to the middle nodes. According to Theorem 6, the code $\mathscr{C}$ described by $V_1, V_2, \ldots, V_n$ for $\mathcal{N}_{h,n,\alpha}$ is MDS if and only if the $\mathbb{F}_q$-linear code $\mathcal{C}$ generated by the matrix $\begin{pmatrix} V_1 & V_2 & \cdots & V_n \end{pmatrix}$ has dimension th and minimum Hamming distance $n-h+1$. Then according to Lemma 1, this is equivalent to the property that any h matrices of $V_1, V_2, \ldots, V_n$ constitute a matrix of full rank. Hence, the network code defined by $V_1, V_2, \ldots, V_n$ for the network $\mathcal{N}_{h,n,h}$ is regular. ∎

Corollary 2: Let n, h, α be positive integers such that $n \geqslant \alpha > h$. Then

$$\mathrm{gap}^{\mathrm{MDS}}(\mathcal{N}_{h,n,\alpha}) = \mathrm{gap}(\mathcal{N}_{h,n,h}) \leqslant \psi(n-1) - \psi(n-h+1),$$

where the $\mathrm{gap}(\mathcal{N}_{h,n,h})$ is the gap between the minimum field size of scalar solution and that of vector solution to the network $\mathcal{N}_{h,n,h}$.

Proof: According to Theorem 7, we have that $\mathrm{gap}^{\mathrm{MDS}}(\mathcal{N}_{h,n,\alpha}) = \mathrm{gap}(\mathcal{N}_{h,n,h})$. Then the conclusion follows from the bound in [3, Theorem 31]. ∎

C. Sub-Networks of the Combination Network With Large Gap

Let t, h, α be positive integers with $\alpha > h$. Let q be a prime power. Denote

$$n \triangleq \begin{bmatrix} th \\ t \end{bmatrix}_q = \frac{\prod_{i=1}^{th}(q^i-1)}{\prod_{i=1}^{t}(q^i-1)\prod_{i=1}^{th-t}(q^i-1)},$$

the Gaussian coefficient which equals the number of t-dimensional vector subspaces of a given vector space of dimension th over $\mathbb{F}_q$. In this section, we consider a sub-network $\mathcal{S}_{q,h,t,\alpha}$ of $\mathcal{N}_{h,n,\alpha}$. In this sub-network, the nodes in the middle layer are indexed by the t-dimensional subspaces $S_1, S_2, \ldots, S_n$ of $\mathbb{F}_q^{th}$. Instead of connecting any α middle nodes to a receiver, we only connect those α nodes where any h of them can span the whole space, that is, if we have a set of α indices $\{i_1, i_2, \ldots, i_\alpha\}$ such that for any h-subset $\{i'_1, i'_2, \ldots, i'_h\} \subset \{i_1, i_2, \ldots, i_\alpha\}$ we have

$$S_{i'_1} + S_{i'_2} + \cdots + S_{i'_h} = \mathbb{F}_q^{th},$$

then the corresponding α middle nodes are connected to a unique receiver.

Lemma 9: There is an MDS code over $\mathbb{F}_q^t$ for the network $\mathcal{S}_{q,h,t,\alpha}$, namely,

$$q_{\mathsf{v}}^{\mathrm{MDS}}(\mathcal{S}_{q,h,t,\alpha}) \leqslant q^t.$$

Proof: For each subspace S_i, we fix a $th \times t$ matrix V_i such that its columns form a basis of S_i. We then construct a network code $\mathscr{C}$ where the source σ sends the packet $\mathbf{x} \cdot V_i$ to the middle node indexed by the subspace S_i, and the middle node simply forwards this packet. For each receiver γ, assume that it is connected to the middle nodes indexed by $S_{i_1}, S_{i_2}, \ldots, S_{i_\alpha}$. Then, according to the definition of the network $\mathcal{S}_{q,h,t,\alpha}$, any h matrices of $V_{i_1}, V_{i_2}, \ldots, V_{i_\alpha}$ constitute a matrix of rank th. Thus, according to Lemma 1 and Lemma 8, we have that $d(\mathscr{C}, \gamma) \geqslant \alpha - h + 1$ at every receiver γ. Then $\mathscr{C}$ is an MDS code since the minimum cut capacity at each receiver is α. ∎

Lemma 10: Assume that $\alpha \leqslant q^t + 1$. Then for any two middle nodes of $\mathcal{S}_{q,h,t,\alpha}$ that are indexed by two subspaces S_{i_1}, S_{i_2} intersecting trivially, i.e., $S_{i_1} \cap S_{i_2} = \{\mathbf{0}\}$, there is a receiver that is connected to both of them.

Proof: Let V_{i_1} and V_{i_2} be two matrices of $\mathbb{F}_q^{th \times t}$ such that their columns constitute a bases of S_{i_1} and S_{i_2}, respectively. Since S_{i_1} and S_{i_2} intersect trivially, then $\mathrm{rank}\begin{pmatrix} V_{i_1} & V_{i_2} \end{pmatrix} = 2t$. Thus we can find $h-2$ matrices $V_{i_j} \in \mathbb{F}_q^{th \times t}$ with $3 \leqslant j \leqslant h$ such that $\mathrm{rank}(V) = th$, where $V = \begin{pmatrix} V_{i_1} & V_{i_2} & \cdots & V_{i_h} \end{pmatrix}$. Since $\alpha \leqslant q^t + 1$, according to Lemma 2, there is an $\mathbb{F}_q$-linear systematic $[\alpha, h]$ MDS code $\mathcal{C}$ over $\mathbb{F}_q^t$ with a generator matrix G such that the first h symbols are the information symbols. Then $V \cdot G$ is another generator matrix of $\mathcal{C}$ with the first h blocks being $V_{i_1}, V_{i_2}, \ldots . V_{i_h}$. We write $V \cdot G = \begin{pmatrix} V_{i_1} V_{i_2} & \cdots V_{i_\alpha} \end{pmatrix}$. Let S_{i_j} be the space spanned by the columns of V_{i_j}. Since $\mathcal{C}$ is MDS, according to Lemma 1, any h matrices of $V_{i_1}, V_{i_2}, \cdots, V_{i_\alpha}$ constitute a matrix of rank th. Thus, any h of the subspaces $S_{i,1}, S_{i_2}, \ldots, S_{i_\alpha}$ can generate the space $\mathbb{F}_q^{th}$. It follows that there is a receiver which is connected to the nodes indexed by $S_{i_1}, S_{i_2}, \ldots, S_{i_\alpha}$. ∎

Lemma 11: Assume that $3 \leqslant h < \alpha \leqslant q^t + 1$. Then

$$q^{\mathrm{MDS}}(\mathcal{S}_{q,h,t,\alpha}) \geqslant \begin{cases} \psi\left(q^t + \frac{1}{h-1}q^{t-1}\right), & \text{if } t \geqslant h, \\ \psi\left(q^t + \frac{1}{(h-1)^2}q^{t-1}\right), & \text{otherwise.} \end{cases} \tag{6}$$

Proof: To avoid tedious notation, we denote $q_{\mathsf{s}} \triangleq q^{\mathrm{MDS}}(\mathcal{S}_{q,h,t,\alpha})$. Let $\mathscr{C}$ be an MDS code over $\mathbb{F}_{q_{\mathsf{s}}}$ for the network $\mathcal{S}_{q,h,t,\alpha}$. For each $i \in [n]$, let $\mathbf{v}_i$ be the encoding vector for the link between σ and the middle node indexed by S_i. We contend that for any two subspace S_{i_1} and S_{i_2} that intersect trivially, the corresponding vectors $\mathbf{v}_{i_1}$ and $\mathbf{v}_{i_2}$ should be linearly independent, namely, they come from different 1-dimensional subspaces. According to Lemma 10, these two nodes are both connected to a receiver. Assume that the other middle nodes connected to this receiver are indexed by $S_{i_3}, S_{i_4}, \ldots, S_{i_\alpha}$. Then according to Lemma 3, the code generated by the matrix $\begin{pmatrix} \mathbf{v}_{i_1} & \mathbf{v}_{i_2} \cdots \mathbf{v}_{i_\alpha} \end{pmatrix}$ is an MDS code. This is equivalent to that any h vectors of $\mathbf{v}_{i_1}, \mathbf{v}_{i_2}, \ldots, \mathbf{v}_{i_\alpha}$ should be linearly independent.

Now, we consider the q-analog of the Kneser graph, denoted by $qK_{th:t}$, whose vertices are the t-dimensional subspaces $S_1, S_2, \ldots, S_n$ and an undirected edge connects two vertices if and only if the corresponding subspaces intersect trivially. Using the code $\mathscr{C}$ defined above, we can give a coloring of this graph: the set of colors is the set of all 1-dimensional subspaces of $\mathbb{F}_{q_{\mathsf{s}}}^h$, and the vertex indexed by S_i is colored by the unique 1-dimensional subspace generated by $\mathbf{v}_i$. The discussion above shows that this coloring is proper, namely, any two adjacent vertices receive different colors. Since $h \geqslant 3$, the chromatic number of this graph, which can be found in [2] and [9], is

$$\chi(qK_{th:t}) = \frac{q^{(h-1)t+1}-1}{q-1}.$$

Thus we have that

$$\frac{q_{\mathsf{s}}^h - 1}{q_{\mathsf{s}} - 1} \geqslant \chi(qK_{th:t}) = \frac{q^{(h-1)t+1}-1}{q-1}.$$

Note that the same inequality appears in the proof of [3, Theorem 26]. With the same process, we can solve for q_{s} and obtain the lower bound in (6). ∎

Corollary 3: Let q be a prime power and t, h, α be positive integers such that $t \geqslant 2$ and $3 \leqslant h < \alpha \leqslant q^t + 1$. Then

$$\begin{aligned} &\mathrm{gap}^{\mathrm{MDS}}(\mathcal{S}_{q,h,t,\alpha}) \\ &\geqslant \begin{cases} \psi\left(q^t + \frac{1}{h-1}q^{t-1}\right) - q^t \geqslant \frac{1}{h-1}q^{t-1}, & \text{if } t \geqslant h, \\ \psi\left(q^t + \frac{1}{(h-1)^2}q^{t-1}\right) - q^t \geqslant \frac{1}{(h-1)^2}q^{t-1}, & \text{otherwise.} \end{cases} \end{aligned}$$

REFERENCES

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[2] A. Blokhuis et al., "A Hilton–Milner theorem for vector spaces," *Electron. J. Combinatorics*, vol. 17, no. 1, pp. 1–12, May 2010.

[3] H. Cai, J. Chrisnata, T. Etzion, M. Schwartz, and A. Wachter-Zeh, "Network-coding solutions for minimal combination networks and their sub-networks," *IEEE Trans. Inf. Theory*, vol. 66, no. 11, pp. 6786–6798, Nov. 2020.

[4] N. Cai and R. W. Yeung, "Network coding and error correction," in *Proc. IEEE Inf. Theory Workshop*, Oct. 2002, pp. 119–122.

[5] R. W. Yeung and N. Cai, "Network error correction, Part I: Basic concepts and upper bounds," *Commun. Inf. Syst.*, vol. 6, pp. 19–36, Aug. 2006.

[6] M. Celebiler and G. Stette, "On increasing the down-link capacity of a regenerative satellite repeater in point-to-point communications," *Proc. IEEE*, vol. 66, no. 1, pp. 98–100, Jan. 1978.

[7] C. Chekuri, C. Fragouli, and E. Soljanin, "On average throughput and alphabet size in network coding," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2410–2424, Jun. 2006.

[8] M. Cheng, Y. Li, X. Zhong, and R. Wei, "Improved constructions of coded caching schemes for combination networks," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 5965–5975, Oct. 2020.

[9] A. Chowdhury, C. Godsil, and G. Royle, "Colouring lines in projective space," *J. Combinat. Theory A*, vol. 113, no. 1, pp. 39–52, Jan. 2006.

[10] S. Gheorghiu, S. S. Bidokhti, C. Fragouli, and A. L. Toledo, "Degraded multicasting with network coding over the combination network," in *Proc. Int. Symp. Netw. Coding*, Beijing, China, Jul. 2011, pp. 1–6.

[11] X. Guang, F.-W. Fu, and Z. Zhang, "Construction of network error correction codes in packet networks," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 1030–1047, Feb. 2013.

[12] X. Guang and R. W. Yeung, "Linear network error correction coding: A revisit," 2021, *arXiv:2103.08081*.

[13] W. Guo, X. Shi, N. Cai, and M. Médard, "Localized dimension growth: A convolutional random network coding approach to managing memory and decoding delay," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3894–3905, Sep. 2013.

[14] C. Hojny, A. B. Kiliç, and A. Ravagnani, "The role of the alphabet in network coding: An optimization approach," in *Proc. IEEE Inf. Theory Workshop*, Apr. 2023, pp. 526–531.

[15] S. Jaggi et al., "Polynomial time algorithms for multicast network code construction," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1973–1982, Jun. 2005.

[16] M. Ji et al., "On the fundamental limits of caching in combination networks," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Stockholm, Sweden, Jun. 2015, pp. 695–699.

[17] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[18] A. R. Lehman and E. Lehman, "Complexity classification of network information flow problems," in *Proc. 15th Annual ACM-SIAM Symp. Discrete Algorithms*, Jan. 2004, pp. 142–150.

[19] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.

[20] H. Liu, H. Wei, S. Puchinger, A. Wachter-Zeh, and M. Schwartz, "On the gap between scalar and vector solutions of generalized combination networks," *IEEE Trans. Inf. Theory*, vol. 67, no. 8, pp. 5580–5591, Aug. 2021.

[21] S. Maheshwar, Z. Li, and B. Li, "Bounding the coding advantage of combination network coding in undirected networks," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 570–584, Feb. 2012.

[22] R. Matsumoto, "Construction algorithm for network error-correcting codes attaining the singleton bound," *IEICE Trans. Fundamentals Electron., Commun. Comput. Sci.*, vol. 90, no. 9, pp. 1729–1735, Sep. 2007.

[23] C. K. Ngai and R. W. Yeung, "Network coding gain of combination networks," in *Proc. Inf. Theory Workshop*, 2004, pp. 283–287.

[24] S. Riis and R. Ahlswede, "Problems in network coding and error correcting codes," in *General Theory of Information Transfer and Combinatorics* (Lecture Notes in Computer Science), vol. 4123. Cham, Switzerland: Springer, 2006, pp. 861–897.

[25] R. M. Roth, *Introduction to Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[26] Q. T. Sun, X. Yang, K. Long, X. Yin, and Z. Li, "On vector linear solvability of multicast networks," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 5096–5107, Dec. 2016.

[27] Q. T. Sun, S. R. Li, and Z. Li, "On base field of linear network coding," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7272–7282, Dec. 2016.

[28] Q. T. Sun, X. Yin, Z. Li, and K. Long, "Multicast network coding and field sizes," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 6182–6191, Nov. 2015.

[29] K. Wan, M. Jit, P. Piantanida, and D. Tuninetti, "On the benefits of asymmetric coded cache placement in combination networks with end-user caches," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1550–1554.

[30] M. Xiao, M. Médard, and T. Aulin, "A binary coding approach for combination networks and general erasure networks," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2007, pp. 786–790.

[31] Q. Yan, M. Wigger, and S. Yang, "Placement delivery array design for combination networks with edge caching," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1555–1559.

[32] S. Yang, R. W. Yeung, and C. K. Ngai, "Refined coding bounds and code constructions for coherent network error correction," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1409–1424, Mar. 2011.

[33] R. W. Yeung and Z. Zhang, "Distributed source coding for satellite communications," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1111–1120, May 1999.

[34] N. Cai, N. Cai, R. W. Yeung, and R. W. Yeung, "Network error correction, II: Lower bounds," *Commun. Inf. Syst.*, vol. 6, no. 1, pp. 37–54, 2006.

[35] A. A. Zewail and A. Yener, "Combination networks with or without secrecy constraints: The impact of caching relays," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 6, pp. 1140–1152, Jun. 2018.

[36] Z. Zhang, "Linear network error correction codes in packet networks," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 209–218, Jan. 2008.

[37] L. Zosin and S. Khuller, "On directed Steiner trees," in *Proc. 13th Annual ACM-SIAM Symposium Discrete Algorithms*, Jan. 2002, pp. 59–63.

Hengjia Wei received the Ph.D. degree in applied mathematics from Zhejiang University, Hangzhou, China, in 2014.

He was a Post-Doctoral Fellow with Capital Normal University, Beijing, China, a Research Fellow with the School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, and a Post-Doctoral Fellow with the School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel. He is currently an Associate Researcher with the Peng Cheng Laboratory, Shenzhen, China. He is also a Professor with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China. His research interests include combinatorial design theory, coding theory, and their intersections.

Dr. Wei received the 2017 Kirkman Medal from the Institute of Combinatorics and Its Applications.

Moshe Schwartz (Fellow, IEEE) received the B.A. (summa cum laude), M.Sc., and Ph.D. degrees from the Computer Science Department, Technion—Israel Institute of Technology, Haifa, Israel, in 1997, 1998, and 2004, respectively.

He was a Fulbright Post-Doctoral Researcher with the Department of Electrical and Computer Engineering, University of California San Diego, and a Post-Doctoral Researcher with the Department of Electrical Engineering, California Institute of Technology. While on sabbatical 2012–2014, he was a Visiting Scientist with Massachusetts Institute of Technology (MIT). He is currently a Professor with the Department of Electrical and Computer Engineering, McMaster University, Canada, on leave from the School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel. His research interests include algebraic coding, combinatorial structures, and digital sequences.

Prof. Schwartz has been an Editorial Board Member of *Journal of Combinatorial Theory Series A* since 2021. He received the 2009 IEEE Communications Society Best Paper Award in Signal Processing and Coding for Data Storage and the 2020 NVMW Persistent Impact Prize. He served as an Associate Editor for Coding Techniques and Coding Theory for IEEE TRANSACTIONS ON INFORMATION THEORY from 2014 to 2021. Since 2021, he has been serving as an Area Editor for Coding and Decoding for IEEE TRANSACTIONS ON INFORMATION THEORY.