Appendix A
Error Bound for Reduction

Given a vector $\mathbf{x} \in \mathbb{R}^{1\times n}$, let $\mathbf{x}' \in \mathbb{R}^{1\times n}$ be its perturbed version. Define $\mathbf{y} = \mathbf{x}\mathbf{A}$ and $\mathbf{y}' = \mathbf{x}'\mathbf{A}$. The condition number $\kappa(\mathbf{A})$ provides an upper bound on the amplification induced by linear transformation:

$$\frac{\|\mathbf{y}' - \mathbf{y}\|}{\|\mathbf{y}\|} \le \kappa(\mathbf{A}) \cdot \frac{\|\mathbf{x}' - \mathbf{x}\|}{\|\mathbf{x}\|}. \tag{14}$$

Recall that the transmitted message of node $t \in \mathcal{T}$ on its outgoing links is given by

$$\mathbf{s}_t = \mathbf{x}_t((\overline{\mathbf{F}}_t^{-1})^T + \mathbf{E}_t^{(1)}),$$

where

$$\mathbf{E}_t^{(1)} = (\mathbf{E}_t^{(rep)})^T + \mathbf{E}_t^{(mul)}.$$

After transmitted by t, $\mathbf{s}_t$ is recoded by the intermediate nodes and finally decoded at ρ. The whole process can be ideally modeled as $\mathbf{s}_t\overline{\mathbf{F}}_t^T$. The decoded message then becomes

$$\mathbf{x}_t^{(0)} = \mathbf{s}_t\overline{\mathbf{F}}_t^T.$$

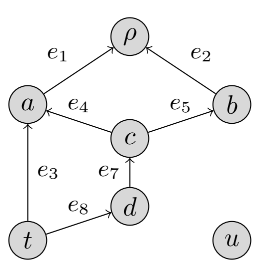
The inequality (14) suggests that the error between $\mathbf{x}_t^{(0)}$ and the original message $\mathbf{x}_t = [\mathbf{x}_t(\overline{\mathbf{F}}_t^{-1})^T](\overline{\mathbf{F}}_t^T)$ is lower bounded by:

$$\begin{aligned}\frac{\|\mathbf{x}_t^{(0)} - \mathbf{x}_t\|}{\|\mathbf{x}_t\|} &\le \kappa(\overline{\mathbf{F}}_t) \cdot \frac{\|\mathbf{s}_t - \mathbf{x}_t(\overline{\mathbf{F}}_t^{-1})^T\|}{\|\mathbf{x}_t(\overline{\mathbf{F}}_t^{-1})^T\|} \\ &\le \kappa^2(\overline{\mathbf{F}}_t) \cdot \frac{\|\mathbf{E}_t^{(1)}\|_{op}}{\|\overline{\mathbf{F}}_t^{-1}\|_{op}}. \end{aligned}\tag{15}$$

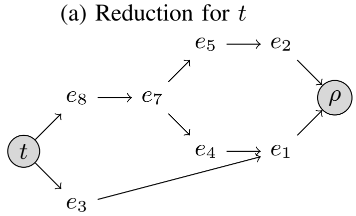
Additionally, the following inequality holds:

$$\begin{aligned}\frac{\|\mathbf{x}_t^{(0)}\|}{\|\mathbf{x}_t\|} &\le \frac{\|\mathbf{x}_t^{(0)} - \mathbf{x}_t\| + \|\mathbf{x}_t\|}{\|\mathbf{x}_t\|} \\ &\le 1 + \kappa^2(\overline{\mathbf{F}}_t) \cdot \frac{\|\mathbf{E}_t^{(1)}\|_{op}}{\|\overline{\mathbf{F}}_t^{-1}\|_{op}}. \end{aligned}\tag{16}$$

Fig. 5 illustrates the reduction process in a butterfly network.



(a) Reduction for t



(b) Paths from t to ρ

Fig. 5: Reduction process

In the butterfly network, three paths connect t to ρ, which are not mutually disjoint. Some paths may bifurcate into multiple distinct routes midway, while in other instances multiple paths may converge into a single route. We can use vector-matrix multiplication to represent the process of coding and decoding:

$$\begin{aligned}&\begin{bmatrix}\mathbf{x}_{t,1}^{(0)} & \mathbf{x}_{t,2}^{(0)}\end{bmatrix} = \\ &\begin{bmatrix}\mathbf{s}_{t,1} & \mathbf{s}_{t,2}\end{bmatrix}\begin{bmatrix}\beta_{e_7,e_5} & \beta_{e_7,e_4} & 0 \\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 \\ 0 & 1 \\ 0 & 1\end{bmatrix}\begin{bmatrix}a_{11} & a_{21} \\ a_{12} & a_{22}\end{bmatrix}, \end{aligned}\tag{17}$$

where $\begin{bmatrix}a_{11} & a_{21} \\ a_{12} & a_{22}\end{bmatrix}$ is a transpose of $\mathbf{A}$.

In practice, the round-off error introduced by this sequence of vector-matrix multiplication can't be avoided. The real decoded message $\mathbf{x}_\mathbf{t}^{(\mathbf{1})}$ is modeled as:

$$\mathbf{x}_t^{(1)} = \mathbf{s}_t(\overline{\mathbf{F}}_t^T + \mathbf{E}_t^{(2)}).$$

In this model, we define an error matrix $\mathbf{E}_t^{(2)} = (e_{ij} : 1 \le i \le r, 1 \le j \le r)$ to represent error introduced by floating-point number arithmetic operations. $\mathbf{E}_t^{(2)}$ is a random matrix depending on the input $\mathbf{s}_t$ and the subsequent vector-matrix multiplication. It can be derived that:

$$\frac{\left\|\mathbf{x}_t^{(1)} - \mathbf{x}_t^{(0)}\right\|}{\left\|\mathbf{x}_t^{(0)}\right\|} \le \kappa(\overline{\mathbf{F}}_t) \cdot \frac{\left\|\mathbf{E}_t^{(2)}\right\|_{op}}{\left\|\overline{\mathbf{F}}_t\right\|_{op}}. \tag{18}$$

By combining inequalities (18) and (16), we derive the upper bound for the distortion rate $\frac{\left\|\mathbf{x}_t^{(1)} - \mathbf{x}_t\right\|}{\|\mathbf{x}_t\|}$ as follows:

$$\begin{aligned}\frac{\left\|\mathbf{x}_t^{(1)} - \mathbf{x}_t\right\|}{\|\mathbf{x}_t\|} &\le \frac{\left\|\mathbf{x}_t^{(1)} - \mathbf{x}_t^{(0)}\right\| + \|\mathbf{x}_t^{(0)} - \mathbf{x}_t\|}{\|\mathbf{x}_t\|} \\ &\le \frac{\left\|\mathbf{x}_t^{(1)} - \mathbf{x}_t^{(0)}\right\|}{\left\|\mathbf{x}_t^{(0)}\right\|} \cdot \frac{\left\|\mathbf{x}_t^{(0)}\right\|}{\|\mathbf{x}_t\|} + \frac{\|\mathbf{x}_t^{(0)} - \mathbf{x}_t\|}{\|\mathbf{x}_t\|} \\ &\le \kappa(\overline{\mathbf{F}}_t) \cdot \frac{\left\|\mathbf{E}_t^{(2)}\right\|_{op}}{\left\|\overline{\mathbf{F}}_t\right\|_{op}} \cdot \left(1 + \kappa^2(\overline{\mathbf{F}}_t) \cdot \frac{\|\mathbf{E}_t^{(1)}\|_{op}}{\|\overline{\mathbf{F}}_t^{-1}\|_{op}}\right) \\ &\quad + \kappa^2(\overline{\mathbf{F}}_t) \cdot \frac{\|\mathbf{E}_t^{(1)}\|_{op}}{\|\overline{\mathbf{F}}_t^{-1}\|_{op}} \\ &\le \kappa^3(\overline{\mathbf{F}}_t) \cdot \frac{\|\mathbf{E}_t^{(1)}\|_{op} \cdot \|\mathbf{E}_t^{(2)}\|_{op}}{\|\overline{\mathbf{F}}_t^{-1}\|_{op} \cdot \|\overline{\mathbf{F}}_t\|_{op}} \\ &\quad + \kappa^2(\overline{\mathbf{F}}_t) \cdot \frac{\|\mathbf{E}_t^{(1)}\|_{op}}{\|\overline{\mathbf{F}}_t^{-1}\|_{op}} + \kappa(\overline{\mathbf{F}}_t) \cdot \frac{\|\mathbf{E}_t^{(2)}\|_{op}}{\|\overline{\mathbf{F}}_t\|_{op}}. \end{aligned}\tag{19}$$

Appendix B
Example of Network Codes Construction

Appendix C
Proof of Lemma 3

Note that $\phi(\boldsymbol{\alpha}) = \prod_{i=1}^{m}(\boldsymbol{\alpha}^T\boldsymbol{\alpha} + \gamma_i^2)^{r/2} >= \prod_{i=1}^{m}\gamma_i^r$, $f(\boldsymbol{\alpha}) = \prod_{i=1}^{m}|\boldsymbol{b}_i^T\boldsymbol{\alpha}|$ and usually $r >= 2$. We denote the domain of $\frac{\phi(\boldsymbol{\alpha})}{f(\boldsymbol{\alpha})}$ as $D = \{\boldsymbol{\alpha} : f(\boldsymbol{\alpha}) \neq 0\}$, then randomly select a point $\boldsymbol{\alpha}' \in D$. As $\|\boldsymbol{\alpha}\|$ approaches 0, $f(\boldsymbol{\alpha})$ approaches 0 while $\phi(\boldsymbol{\alpha}) > \prod_{i=1}^{m} \gamma_i^r$. Thus $\frac{\phi(\boldsymbol{\alpha})}{f(\boldsymbol{\alpha})}$ tends to positive infinity. There exists $l \in \mathbb{R}$ such that

APPENDIX D
PROOF OF THEOREM 1

This section we verify that coding scheme based on floating-point number exists. Assume that $\widetilde{\mathbf{F}}_t^{i-1}$ is full rank. The updating in the i-th step of the algorithm should guarantee that $\widetilde{\mathbf{F}}_t^{i}$ is full rank in the i-th step.

Let $U := \langle \mathbf{c}_{e'} : e' \in \mathrm{In}(v) \rangle$ and $V_t := \langle \widetilde{\mathbf{c}}_{t,k}^{i-1}, 0 \le k < r, k \neq j \rangle$. Then $\mathbf{c}_e$ must satisfy $\mathbf{c}_e \in U \setminus \bigcup_{t \in \mathcal{T}_e} (U \cap V_t)$, where $\mathcal{T}_e$ is the set of sink nodes with e on one of its paths from the source node.

Let's view this problem from the perspective of linear subspaces. For any $t \in \mathcal{T}$ with e on its path, there exists $\dim(U \cap V_t) \le \dim(U) - 1$. Since $|\mathcal{T}_e|$ is finite, nearly all of the points in linear space U lie in the set $U \setminus \bigcup_{t \in \mathcal{T}_e} (U \cap V_t)$, which implies that, when based on floating-point number, randomly choosing the $\beta_{e',e}$ for each $e' \in \mathrm{In}(v)$ is sufficient to derive a full-rank matrix $\widetilde{\mathbf{F}}_t^{i}$. Similar randomized approach can be utilized to construct the matrix $\mathbf{A}$.

As the algorithm stops, we see that $\widetilde{\mathbf{F}}_t = \overline{\mathbf{F}}_t$ since, in each step i of the algorithm, $\widetilde{\mathbf{F}}_{t,j}$ always takes the value of $\mathbf{c}_e$ where e is the most downstream edge in the truncation in G^i of the j-th edge-disjoint path from the source node ρ to sink node t in $\mathcal{T}$.

APPENDIX E
PROOF OF THEOREM 2

Assume that point $\boldsymbol{\alpha}_1$ minimizes function $\frac{\phi}{f}$ globally and $\boldsymbol{\alpha}_0$ is the maximizer for f as it constrained on the unit sphere, then the point $\frac{\boldsymbol{\alpha}_1}{\|\boldsymbol{\alpha}_1\|} \in \{x : \boldsymbol{\alpha}^T \boldsymbol{\alpha} = 1\}$ satisfying $f(\frac{\boldsymbol{\alpha}_1}{\|\boldsymbol{\alpha}_1\|}) \le f(\boldsymbol{\alpha}_0)$. Since $f(\boldsymbol{\alpha}_1) = \|\boldsymbol{\alpha}_1\|^m \cdot f(\frac{\boldsymbol{\alpha}_1}{\|\boldsymbol{\alpha}_1\|})$ and $f(\|\boldsymbol{\alpha}_1\| \cdot \boldsymbol{\alpha}_0) = \|\boldsymbol{\alpha}_1\|^m \cdot f(\boldsymbol{\alpha}_0)$ there exists $f(\boldsymbol{\alpha}_1) \le f(\|\boldsymbol{\alpha}_1\| \cdot \boldsymbol{\alpha}_0)$. Simultaneously, $\phi(\boldsymbol{\alpha}_1^T \boldsymbol{\alpha}_1) = \phi(\ (\|\boldsymbol{\alpha}_1\| \cdot \boldsymbol{\alpha}_0^T)(\|\boldsymbol{\alpha}_1\| \cdot \boldsymbol{\alpha}_0)\ )$ because $\boldsymbol{\alpha}_0^T \boldsymbol{\alpha}_0 = 1$. Thus $\|\boldsymbol{\alpha}_1\| \cdot \boldsymbol{\alpha}_0$ minimizes our objective function.

APPENDIX F
CONSTRUCTION OF GENERATOR MATRIX

To optimize the initial generator matrix $\mathbf{A}$, we propose a heuristic algorithm to reduce $\prod\limits_{t \in \mathcal{T}} \kappa(\widetilde{\mathbf{F}}_t^0)$.

First, initialize $\mathbf{A}$ by sampling each entry uniformly from $[0, 2]$. Note that each column in $\mathbf{A}$ corresponds to a coefficient vector $\mathbf{c}_e$ such that $e \in \mathrm{Out}(\rho)$. And each transfer matrix $\widetilde{\mathbf{F}}_t^0$ consists of a subset of $\{\mathbf{c}_e : e \in \mathrm{Out}(\rho)\}$. We replace the columns of $\mathbf{A}$ appearing in only one transfer matrix $\widetilde{\mathbf{F}}_t^0$ with standard basis vectors $\{\hat{\mathbf{e}}_i : i = 1, \ldots r\}$ while preserving the full-rank property of $\widetilde{\mathbf{F}}_t^0$.

Next, shared columns are optimized. For a column $\mathbf{c}_{\mathbf{e}}$, let $\mathcal{T}_e \subset \mathcal{T}$ denote the set of sink nodes where $\mathbf{c}_e$ appears in their transfer matrices. Using our optimization framework, we compute coefficients $\{\beta_i\}_{i=1}^r$ to reconstruct $\mathbf{c}_e = \sum_{i=1}^r \beta_i \hat{\mathbf{e}}_i$. This process is iterated for all shared columns, with multiple refinement cycles possible to improve results.