Something Ancient and Something Recent

Raymond W. Yeung
Institute of Network Coding, CUHK

Something Ancient

PROCEEDINGS

ITW '93
1993 IEEE Information Theory Workshop

June 4 (FRI) – June 8 (TUE), 1993

at

Fuji Institute of Education and Training

Susono-shi, Shizuoka, Japan

Sponsored by:
IEEE Information Theory Society

In cooperation with:
IEICE Information Society Group
The Society of Information Theory
and Its Applications

Cosponsored by:
International Communications Foundation
Inoue Foundation for Science
The Telecommunications Advancement Foundation
ADVANCE COMPANY. LTD.
Canon Inc.
IBM Japan, Ltd.
Mitsubishi Electric Corporation
NEC Corporation
Nippon Telegraph and Telephone Corporation
NTT Mobile Communications Network Inc.
OKI Electric Industry Company Ltd.
Toshiba Corporation

Diversified Coding with One Distortion Criterion

Raymond W. Yeung
Department of Information Engineering
The Chinese University of Hong Kong
email: whyeung@ie.cuhk.hk

1 Introduction

In a *Diversified Coding System* (DCS), an information source is encoded by a number of encoders. There are a number of decoders, each of which can access a certain subset of the encoders. Each decoder is to reconstruct the source either perfectly or subject to a distortion criterion. The problem is to determine the coding rate region for a particular configuration of a DCS subject to certain distortion criteria.

Diversified coding has wide application in distributed information storage (e.g. [3]), fault-tolerant communication network (e.g. [4]), and secret sharing (e.g. [5]). Most of these works are application of the pioneering work of Singleton [1] on maximum distance error-correcting codes.

Diversified coding from the rate-distortion point of view is discussed in the work of El Gamal and Cover [2] on the multiple descriptions problem. In their work, each decoder makes it best effort to reconstruct the source with no reference to the reconstructions by other decoders. By contrast, in our problem, the decoders are divided into classes, and it is required that the reconstructions of the source by decoders within the same class are identical. This is a natural requirement for many applications. For example, if the users of decoders within the same class are to discuss the information they receive subsequently, it would be critical that the information they receive are identical.

IEEE TRANSACTIONS ON

INFORMATION THEORY

A Journal Devoted to the Theoretical and Experimental Aspects of Information Transmission, Processing, and Utilization

MARCH 1995 VOLUME 41 NUMBER 2 IETTAW (ISSN 0018-9448)

PAPERS

for all $j \in B$. Now each DC decoder can reconstruct the output of the source encoder with probability of error less than ϵ''. Thus the source decoder can reconstruct the source with average distortion of less than

$$(1-\epsilon'')(D+\epsilon') + \epsilon'' d_{\max}$$

where $d_{\max}$ is the largest possible value of the distortion function (we assume that $d_{\max}$ is finite). For any $\epsilon > 0$, by taking $\epsilon', \epsilon'' > 0$ satisfying

$$\epsilon' + \epsilon'' < \epsilon$$

and

$$(1-\epsilon'')(D+\epsilon') + \epsilon'' d_{\max} < D + \epsilon,$$

we have shown the existence of a diversity coding scheme with coding rates satisfying

$$\sum_{i \in G_j} R_i < R(D) + \epsilon$$

for all $j \in B$, where each decoder can reconstruct the source with average distortion of less than $D+\epsilon$. This proves that the inequalities in (2) are necessary and sufficient conditions for the coding rates in the usual Shannon sense. This observation appears to be new.

B. Subtlety of Multilevel Diversity Coding

The class of MDCS's we treat in this paper, which will be described in the next section, have two levels of decoders. In this subsection, we first illustrate the subtlety of such problems by means of a simple example. Consider the i.i.d. source $\{(X_k, Y_k)\}$ where $\{X_k\}$ and $\{Y_k\}$ are independent bit streams with rates 1 bit per unit time. The MDCS we consider is shown in Fig. 3, where it is required that Decoder 1 can reconstruct $\{X_k\}$ with zero error, and Decoder i, $i = 2, 3, 4$ can reconstruct $\{(X_k, Y_k)\}$ with zero error; $\{X_k\}$ can be regarded as a degraded version of $\{(X_k, Y_k)\}$. Now $\{X_k\}$ has to be delivered to all decoders, while $\{Y_k\}$ has to be delivered to Decoders 2, 3, and 4. Since $\{X_k\}$ and $\{Y_k\}$ are independent bit streams, one may expect that optimality can be achieved by

$(1, 1, 1)$ is achievable

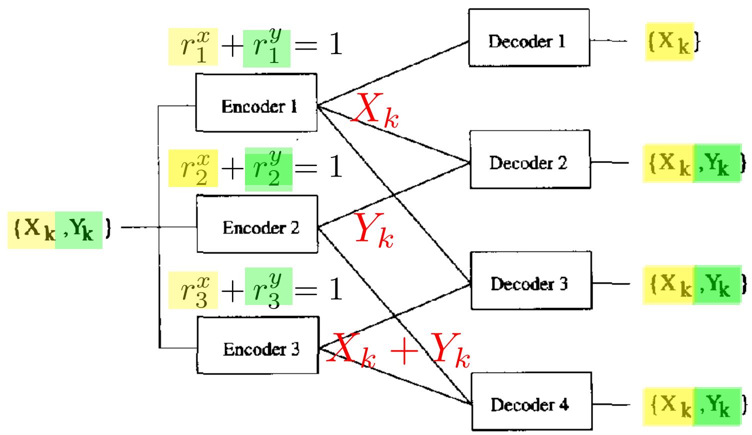


Fig. 3. An MDCS for which the principle of superposition fails.

coding $\{X_k\}$ and $\{Y_k\}$ separately, as for the case of point-to-point communication. This will be referred to as the *principle of superposition*. The argument is that as $\{X_k\}$ and $\{Y_k\}$ are independent bit streams, the coding rates contributing to the coding of $\{X_k\}$ do not contribute to the coding of $\{Y_k\}$. At this point, however, it is not clear whether the principle of superposition actually applies.

We now define the coding rate region corresponding to coding $\{X_k\}$ and $\{Y_k\}$ separately, $\mathcal{R}_{\text{sup}}$, as the set containing (R_1, R_2, R_3) such that for $i = 1, 2, 3$

$$1\ \cancel{R_i} = r_i^x + r_i^y \tag{3}$$

where $r_i^x, r_i^y \ge 0$, and

$$r_1^x \ge 1 \tag{4}$$

$$r_1^x + r_2^x \ge 1 \tag{5}$$

$$r_2^x + r_3^x \ge 1 \tag{6}$$

$$r_1^x + r_3^x \ge 1 \tag{7}$$

and

$$\cancel{r_1^y \ge 1} \quad \cancel{(8)}$$

$$r_1^y + r_2^y \ge 1 \tag{9}$$

$$r_2^y + r_3^y \ge 1 \tag{10}$$

$$r_1^y + r_3^y \ge 1. \tag{11}$$

$(1,1,1)$ is achievable

for all $j \in B$. Now each DC decoder can reconstruct the output of the source encoder with probability of error less than ϵ''. Thus the source decoder can reconstruct the source with average distortion of less than

$$(1-\epsilon'')(D+\epsilon') + \epsilon'' d_{\max}$$

where $d_{\max}$ is the largest possible value of the distortion function (we assume that $d_{\max}$ is finite). For any $\epsilon > 0$, by taking $\epsilon', \epsilon'' > 0$ satisfying

$$\epsilon' + \epsilon'' < \epsilon$$

and

$$(1-\epsilon'')(D+\epsilon') + \epsilon'' d_{\max} < D + \epsilon,$$

we have shown the existence of a diversity coding scheme with coding rates satisfying

$$\sum_{i \in G_j} R_i < R(D) + \epsilon$$

for all $j \in B$, where each decoder can reconstruct the source with average distortion of less than $D + \epsilon$. This proves that the inequalities in (2) are necessary and sufficient conditions for the coding rates in the usual Shannon sense. This observation appears to be new.

B. Subtlety of Multilevel Diversity Coding

The class of MDCS's we treat in this paper, which will be described in the next section, have two levels of decoders. In this subsection, we first illustrate the subtlety of such problems by means of a simple example. Consider the i.i.d. source $\{(X_k, Y_k)\}$ where $\{X_k\}$ and $\{Y_k\}$ are independent bit streams with rates 1 bit per unit time. The MDCS we consider is shown in Fig. 3, where it is required that Decoder 1 can reconstruct $\{X_k\}$ with zero error, and Decoder i, $i = 2, 3, 4$ can reconstruct $\{(X_k, Y_k)\}$ with zero error; $\{X_k\}$ can be regarded as a degraded version of $\{(X_k, Y_k)\}$. Now $\{X_k\}$ has to be delivered to all decoders, while $\{Y_k\}$ has to be delivered to Decoders 2, 3, and 4. Since $\{X_k\}$ and $\{Y_k\}$ are independent bit streams, one may expect that optimality can be achieved by

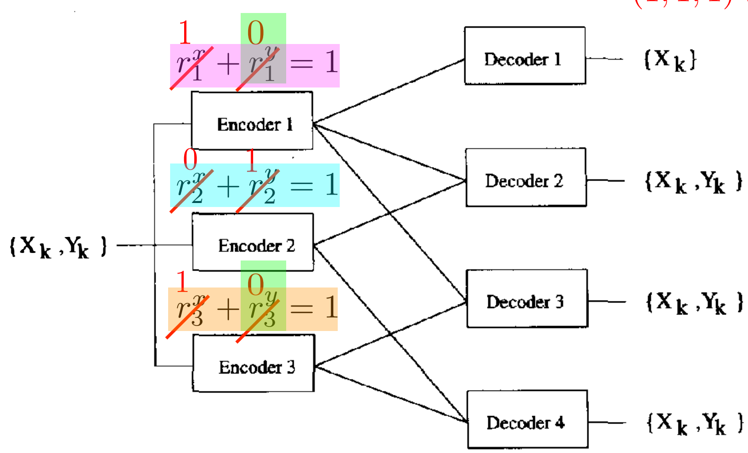


Fig. 3. An MDCS for which the principle of superposition fails.

coding $\{X_k\}$ and $\{Y_k\}$ separately, as for the case of point-to-point communication. This will be referred to as the *principle of superposition*. The argument is that as $\{X_k\}$ and $\{Y_k\}$ are independent bit streams, the coding rates contributing to the coding of $\{X_k\}$ do not contribute to the coding of $\{Y_k\}$. At this point, however, it is not clear whether the principle of superposition actually applies.

We now define the coding rate region corresponding to coding $\{X_k\}$ and $\{Y_k\}$ separately, $\mathcal{R}_{\text{sup}}$, as the set containing (R_1, R_2, R_3) such that for $i = 1, 2, 3$

$$R_i = r_i^x + r_i^y \tag{3}$$

where $r_i^x, r_i^y \ge 0$, and

$$r_1^x \ge 1 \tag{4}$$

$$r_1^x + r_2^x \ge 1 \tag{5}$$

$$\overset{0}{\cancel{r_2^x}} + \overset{1}{\cancel{r_3^x}} \ge 1 \tag{6}$$

$$r_1^x + r_3^x \ge 1 \tag{7}$$

and

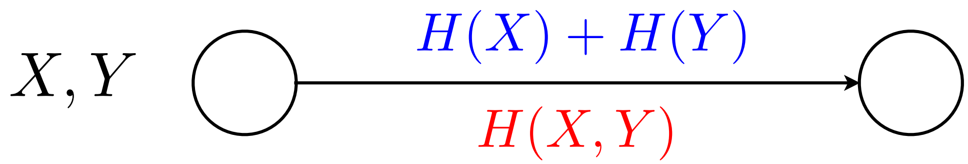
$$\cancel{r_1^y \ge 1} \quad \cancel{(8)}$$

$$\overset{0}{\cancel{r_1^y}} + \overset{1}{\cancel{r_2^y}} \ge 1 \tag{9}$$

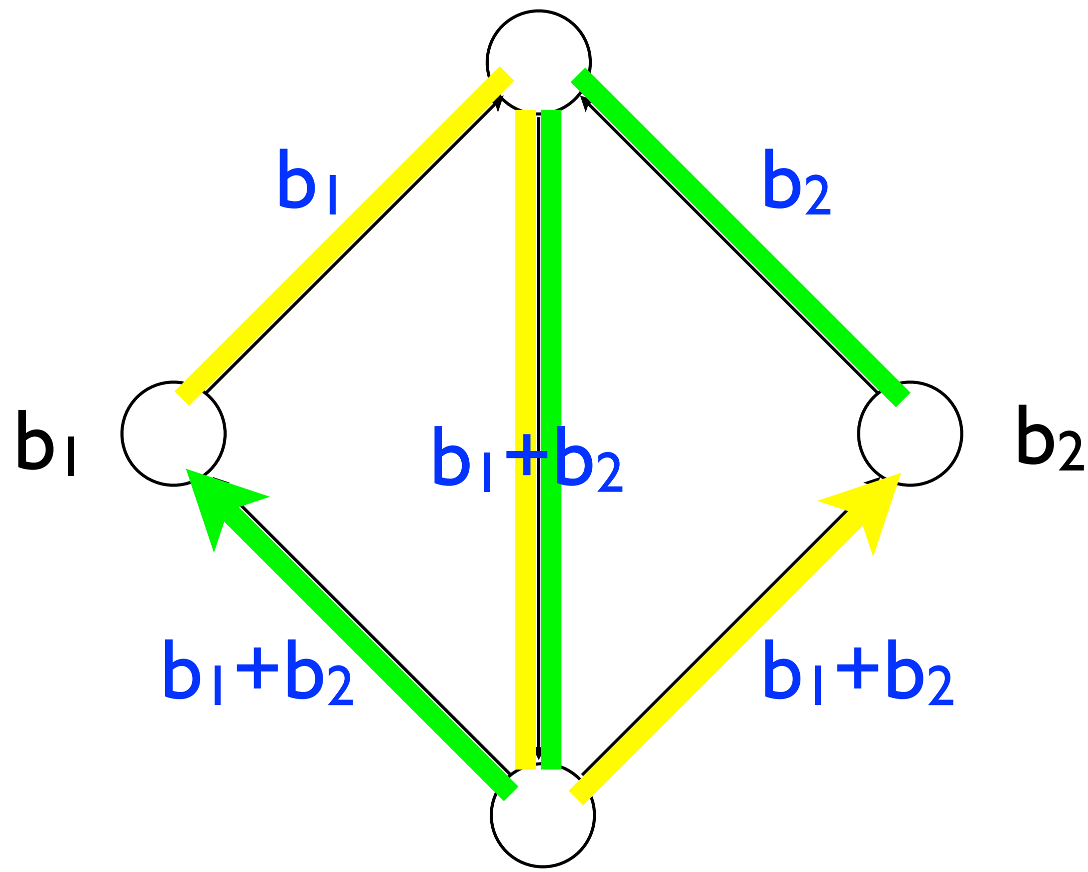
$$r_2^y + r_3^y \ge 1 \tag{10}$$

$$r_1^y + r_3^y \ge 1. \tag{11}$$

Why is this interesting?



- When X and Y are independent, $H(X) + H(Y) = H(X, Y)$.
- From classical information theory, we know that there is no difference between compressing X and Y separately or together.
- Therefore, in classical information theory, there is no distinction between single-source or multi-source data compression.
- The last example shows that the behavior of information in a network deviates from what we would expect from classical information theory.
- Transmission of well-compressed information sources in a network is not a commodity flow.
- A gold mine ahead $\rightarrow\rightarrow\rightarrow$ Network coding



b1
b2
b1
b1+b2
b2
b1+b2
b1+b2

Other Works Leading to Network Coding

- K. P. Hau, "Multilevel diversity coding with independent data streams," MPhil thesis, CUHK, 1995.
- J. R. Roche, R. W. Yeung and K. P. Hau, "Symmetrical multilevel diversity coding," 1997.
- R. W. Yeung and Z. Zhang, "On symmetrical multilevel diversity coding," 1999.
- R. W. Yeung and Z. Zhang, "Distributed source coding for satellite communications," 1999. "Multi-source network coding on two-tier networks"
- R. Ahlswede, N. Cai, S.-Y. R. Li and R. W. Yeung, "Network information flow," 2000. "Single-source network coding on general networks"

Network Coding and Entropy Function

Network Coding | Entropy Function

1992

1993

1994

This problem is either trivial or great. (1995)

1995

1996

1997 — Y97, ZY97

1998 — ZY98

1999 — YZ99

2000 — ACLY2000

LYC03, KM03, JSCEEJT04, . . .

DFZ05, SYC06, DFZ07, CG08, YYZ12, . . .

Something about Network Coding

A very unique class of multi-user information theory problems:

- Non-trivial even with very simplistic assumptions
  - independent sources
  - individual sources well compressed
  - no distortion consideration
- Exists a unifying implicit single-letter characterisation of the capacity region in terms of the entropy function region Γ^*

Something Recent

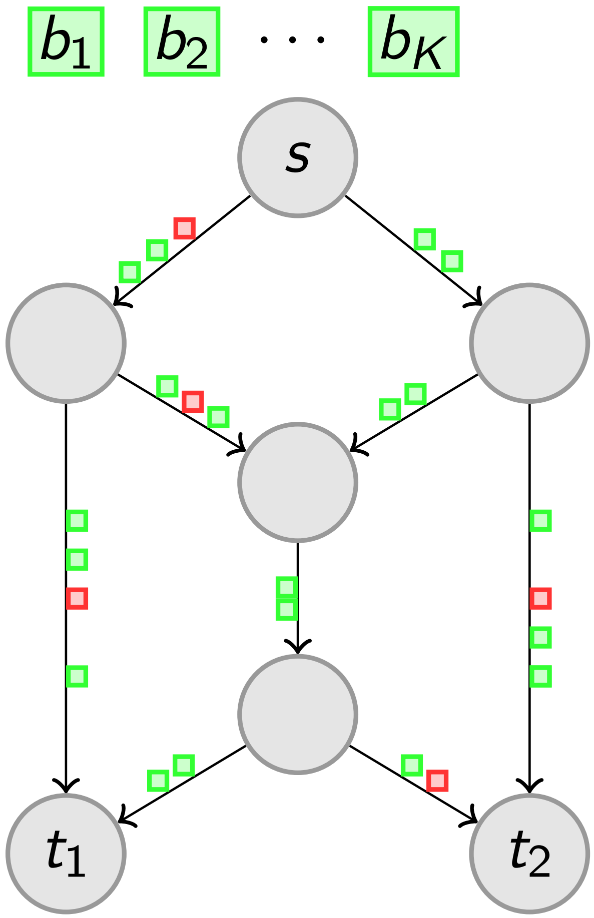
BATched Sparse (BATS) Code

- S. Yang and R. W. Yeung, "Coding for a network coded fountain," 2011 ISIT.
- S. Yang and R. W. Yeung, "Batched sparse code," IEEE IT, 2014.

Shenghao Yang
CUHK (Shenzhen)

Transmission through Packet Networks (Erasure Networks)

One 20MB file ≈ 20,000 packets

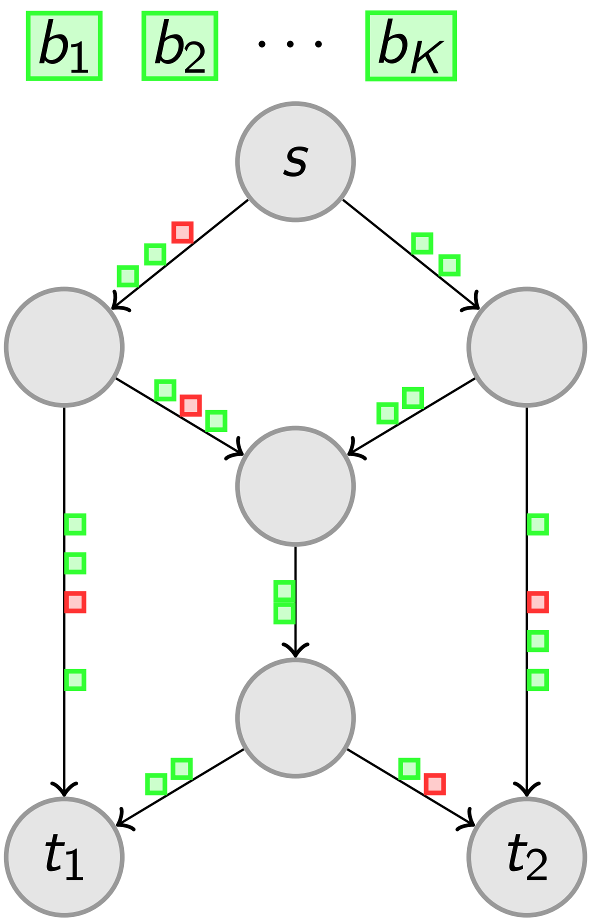


Transmission through Packet Networks (Erasure Networks)

One 20MB file ≈ 20,000 packets

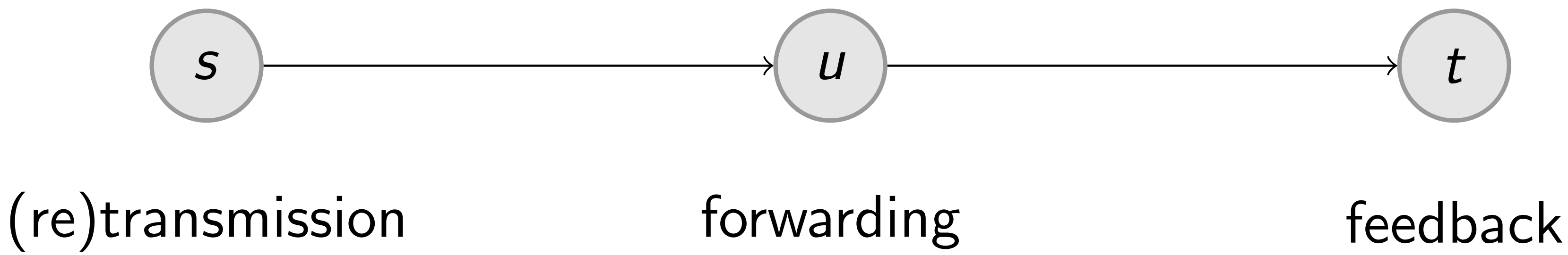
A practical solution

- low computational and storage costs
- high transmission rate
- small protocol overhead



Retransmission

- Example: TCP
- Not scalable for multicast
- Cost of feedback



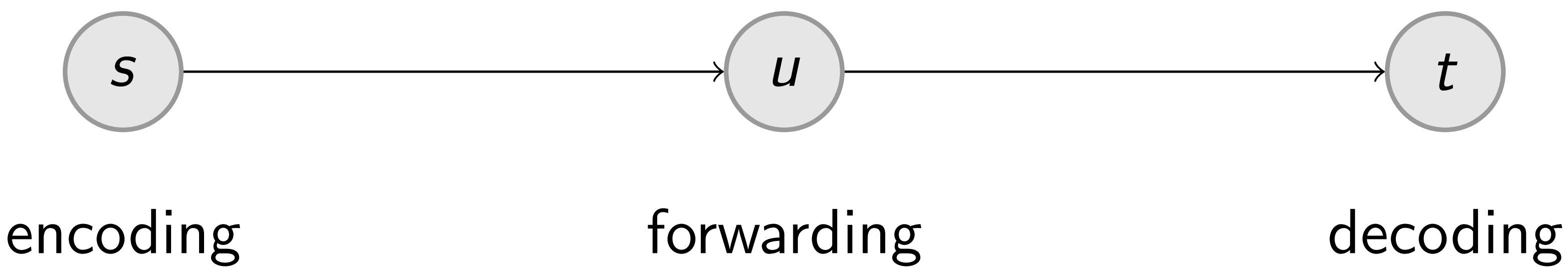
Routing Networks

Retransmission

- Example: TCP
- Not scalable for multicast
- Cost of feedback

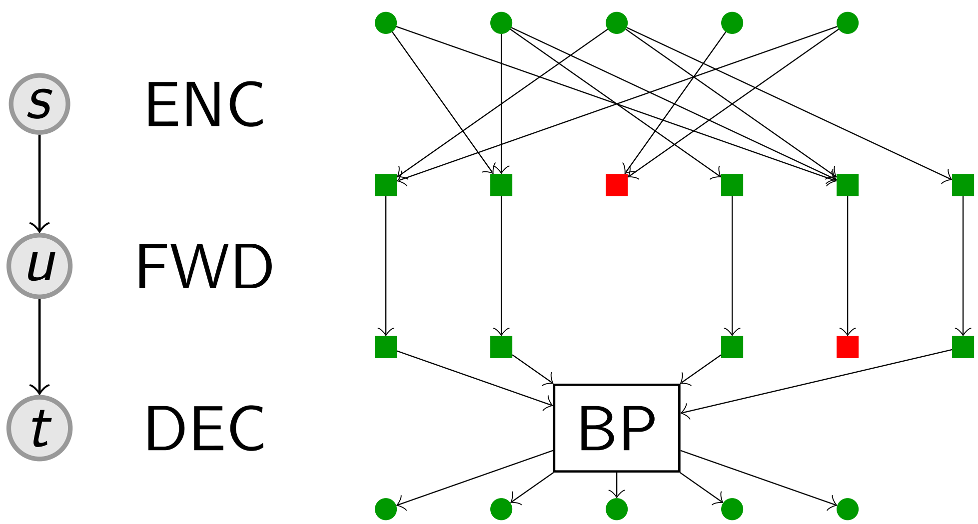
Forward error correction

- Example: fountain codes
- Scalable for multicast
- Neglectable feedback cost



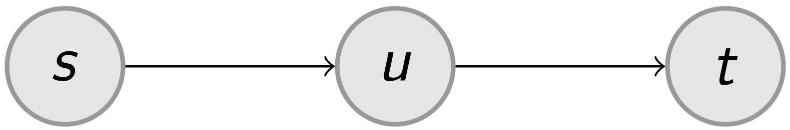
Complexity of Fountain Codes with Routing

- K packets, T symbols in a packet.
- Encoding: $\mathcal{O}(T)$ per packet.
- Decoding: $\mathcal{O}(T)$ per packet.
- Routing: $\mathcal{O}(1)$ per packet and fixed buffer size.



[Luby02] M. Luby, "LT codes," in Proc. 43rd Ann. IEEE Symp. on Foundations of Computer Science, Nov. 2002.
[Shokr06] A. Shokrollahi, "Raptor codes," IEEE Trans. Inform. Theory, vol. 52, no. 6, pp. 2551-2567, Jun 2006.

Achievable Rates

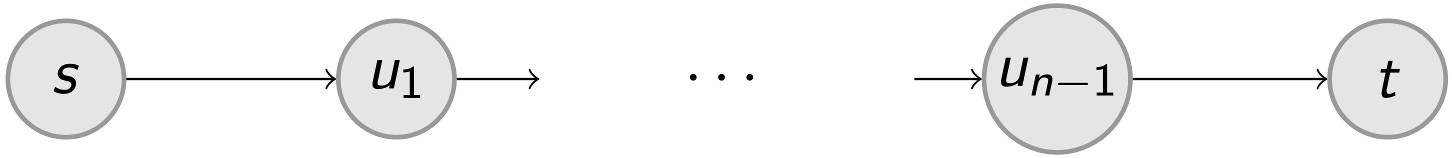


Both links have a packet loss rate 0.2.

The capacity of this network is 0.8.

Intermediate	End-to-End	Maximum Rate
forwarding	retransmission	0.64
forwarding	fountain codes	0.64
network coding	random linear codes	0.8

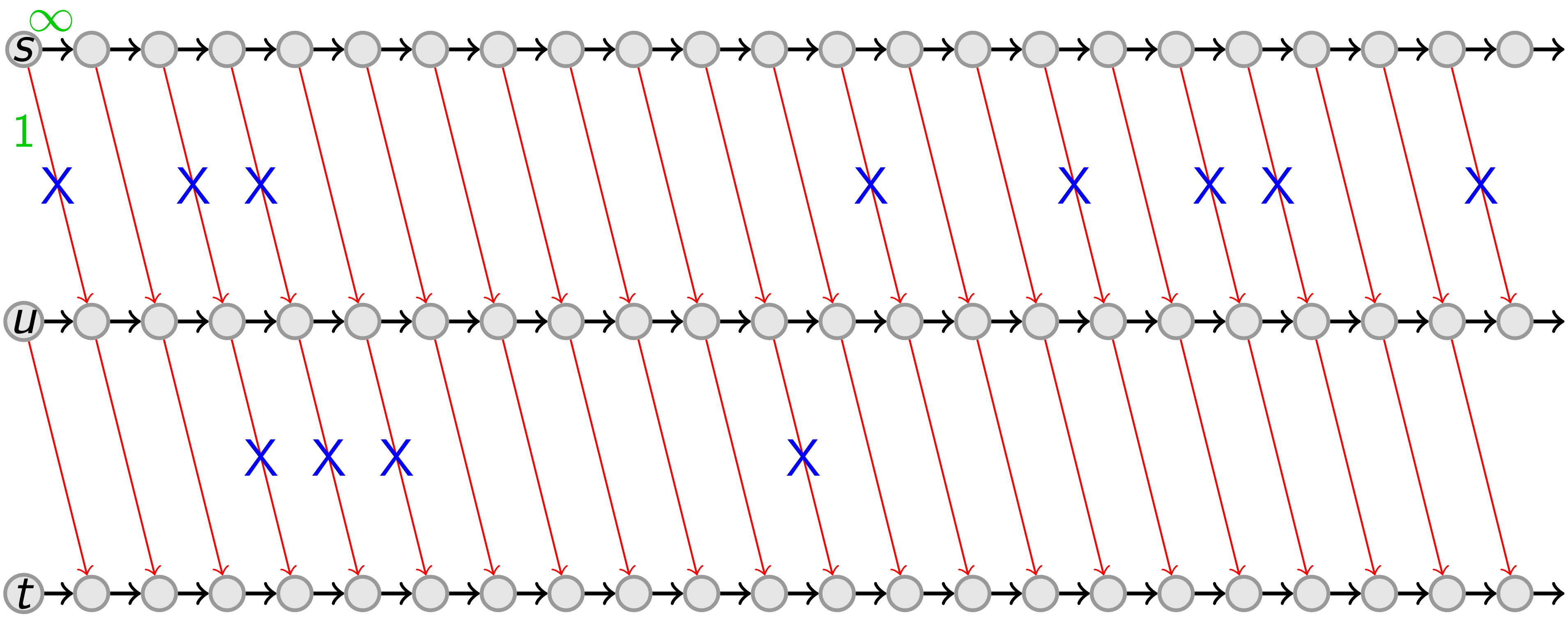
Achievable Rates: n hops



All links have a packet loss rate 0.2.

Intermediate Operation	Maximum Rate
forwarding	$0.8^n \to 0,\ n \to \infty$
network coding	0.8

An Explanation



Multicast capacity of erasure networks

Theorem

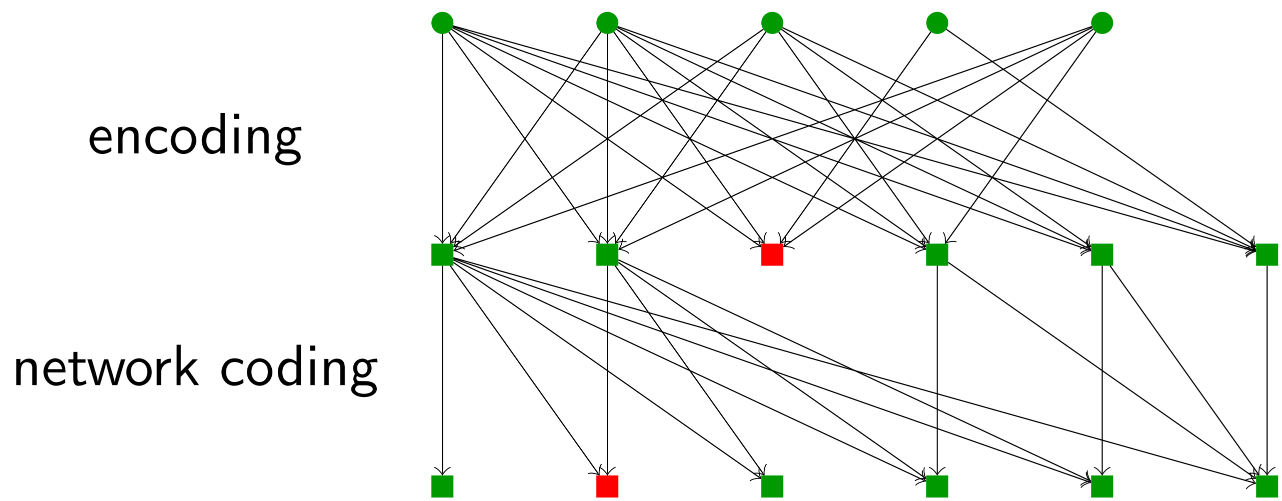
Random linear network codes achieve the capacity of a large range of multicast erasure networks.

[Wu06] Y. Wu, "A trellis connectivity analysis of random linear network coding with buffering," in Proc. IEEE ISIT 06, Seattle, USA, Jul. 2006.

LMKE08] D. S. Lun, M. Médard, R. Koetter, and M. Effros, "On coding for reliable communication over packet networks," Physical Communication, vol. 1, no. 1, pp. 320, 2008.

Complexity of Linear Network Coding

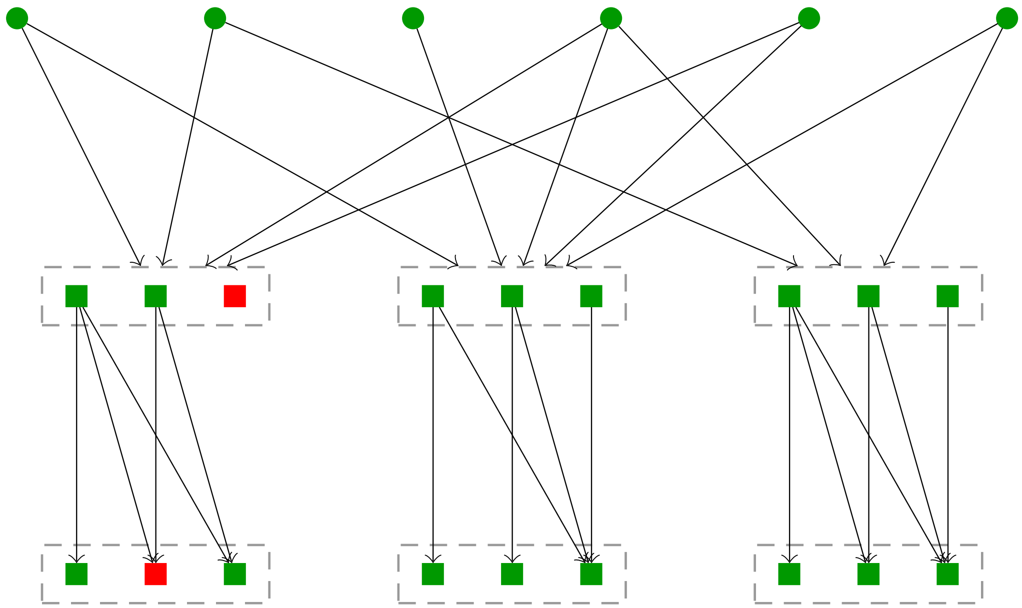
- Encoding: $\mathcal{O}(TK)$ per packet.
- Decoding: $\mathcal{O}(K^2 + TK)$ per packet.
- Network coding: $\mathcal{O}(TK)$ per packet. Buffer K packets.



Batched Sparse (BATS) Codes

outer code
(matrix fountain code)

inner code
(network code)

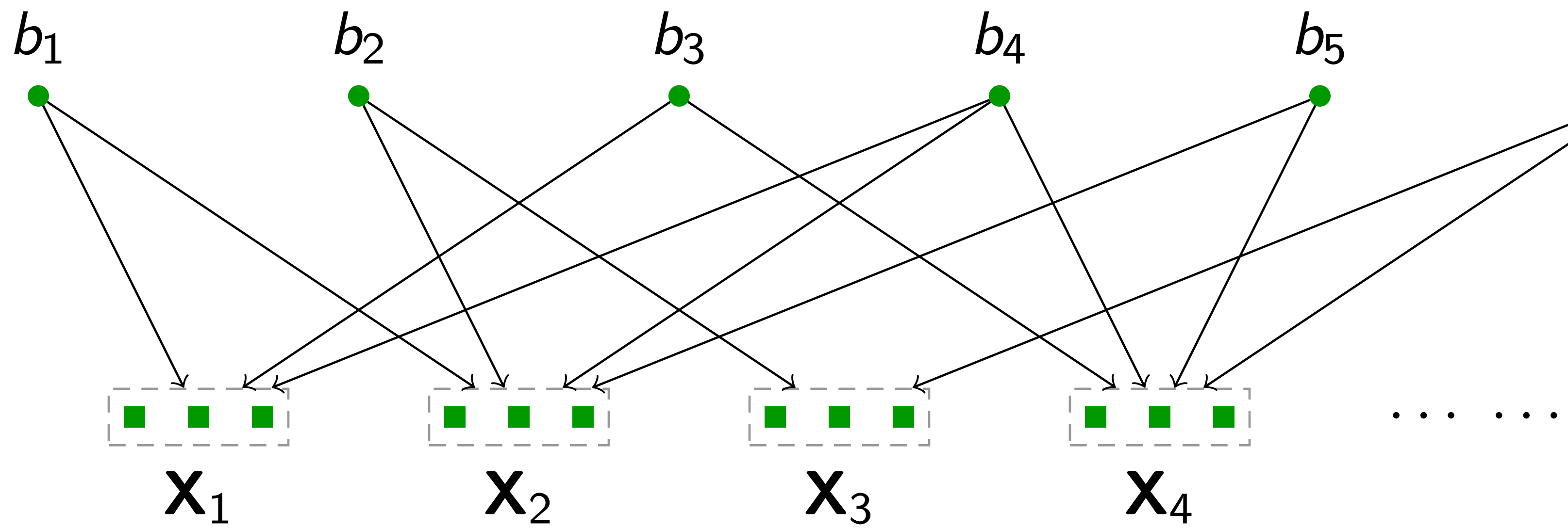


[YY11] S. Yang and R. W. Yeung. Coding for a network coded fountain. ISIT 2011, Saint Petersburg, Russia, 2011.

[YY14] S. Yang and R. W. Yeung. Batched sparse codes. Information Theory, IEEE Transactions on, vol. 60, no. 9, pp. 53225346, Sep. 2014

Encoding of BATS Code: Outer Code

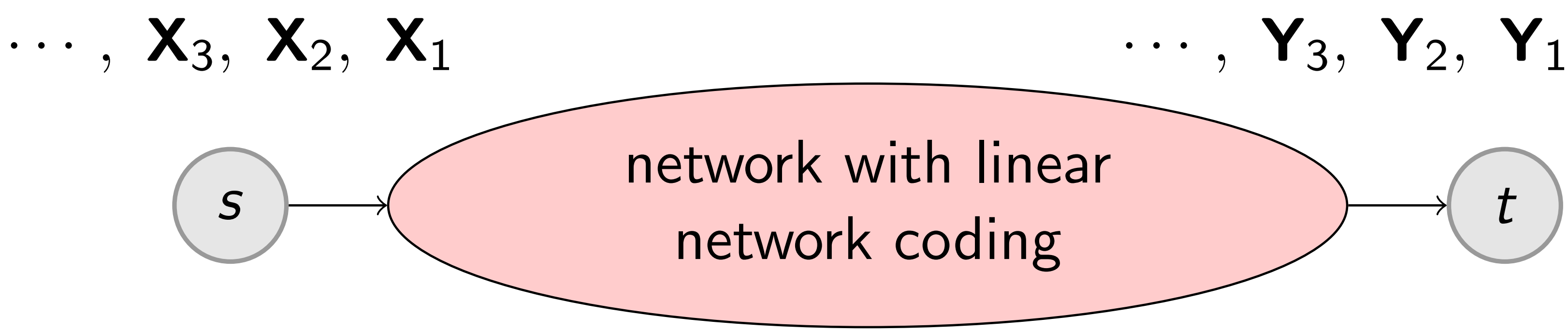
- Apply a "matrix fountain code" at the source node:
  1. Obtain a degree d by sampling a degree distribution Ψ.
  2. Pick d distinct input packets randomly.
  3. Generate a batch of M coded packets using the d packets.
- Transmit the batches sequentially.



$$\mathbf{X}_i = \begin{bmatrix} b_{i1} & b_{i2} & \cdots & b_{id_i} \end{bmatrix} \mathbf{G}_i = \mathbf{B}_i \mathbf{G}_i.$$

Encoding of BATS Code: Inner Code

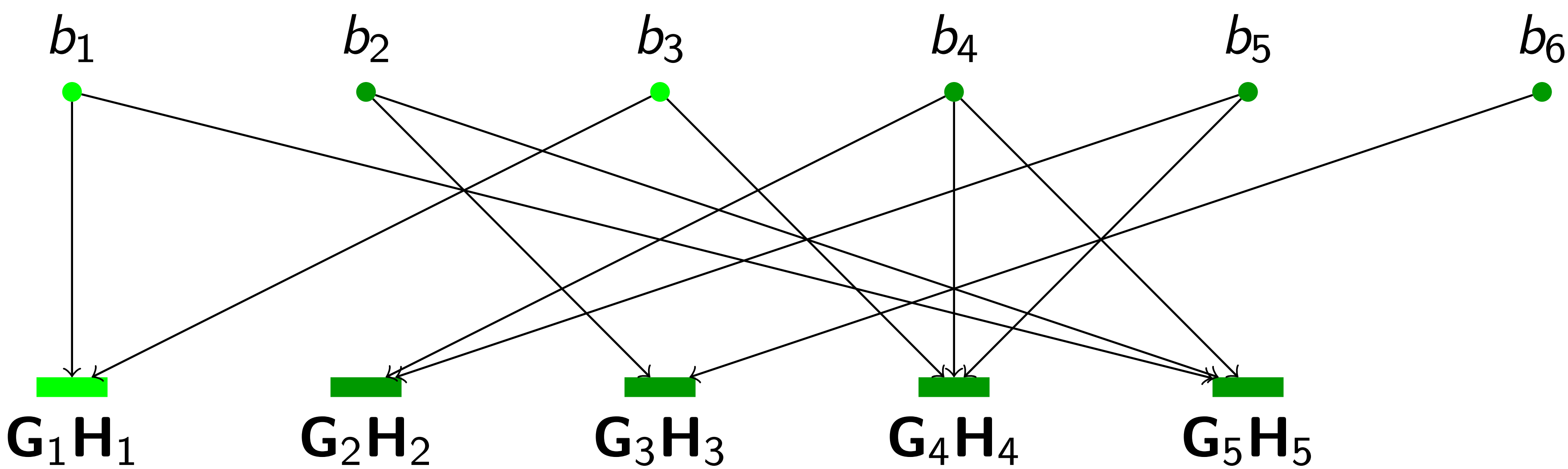
- The batches traverse the network.
- Encoding at the intermediate nodes forms the inner code.
- Linear network coding is applied in a causal manner within a batch.



$$\mathbf{Y}_i = \mathbf{X}_i \mathbf{H}_i,\ i = 1, 2, \ldots.$$

Belief Propagation Decoding

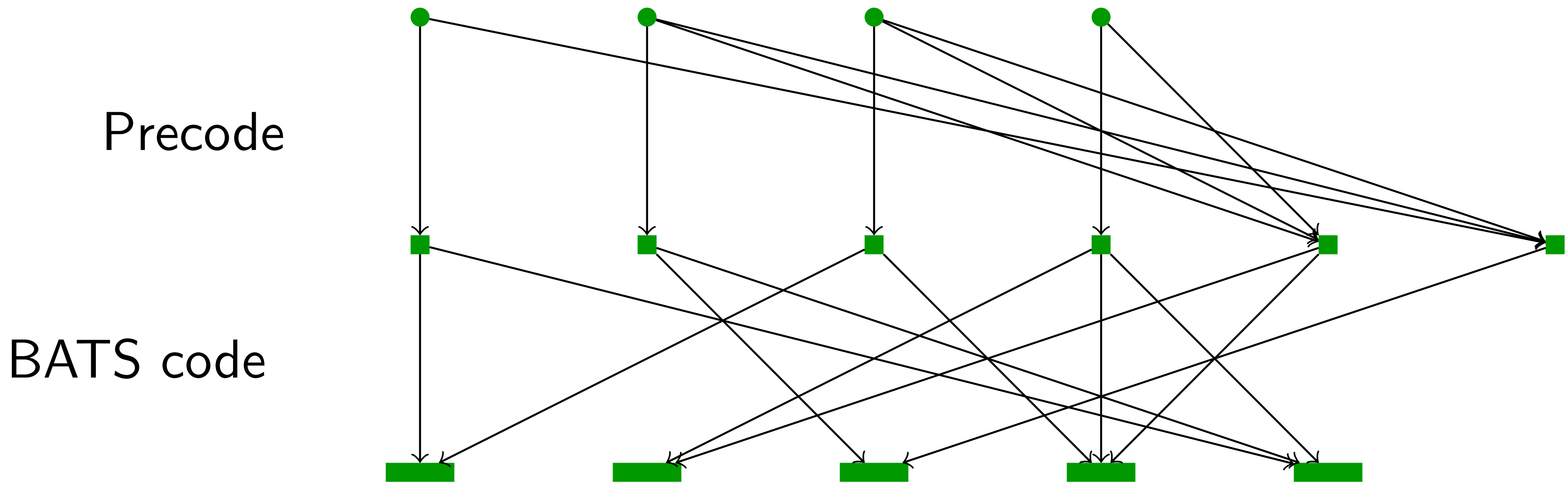
1. Find a check node i with $\text{degree}_i = \text{rank}(\mathbf{G}_i\mathbf{H}_i)$.
2. Decode the ith batch.
3. Update the decoding graph. Repeat 1).



The linear equation associated with a check node: $\mathbf{Y}_i = \mathbf{B}_i\mathbf{G}_i\mathbf{H}_i$.

Precoding

- Precoding by a fixed-rate erasure correction code.
- The BATS code recovers $(1-\eta)$ of its input packets.



[Shokr06] A. Shokrollahi, Raptor codes, IEEE Trans. Inform. Theory, vol. 52, no. 6, pp. 25512567, Jun. 2006.

Complexity of Sequential Scheduling

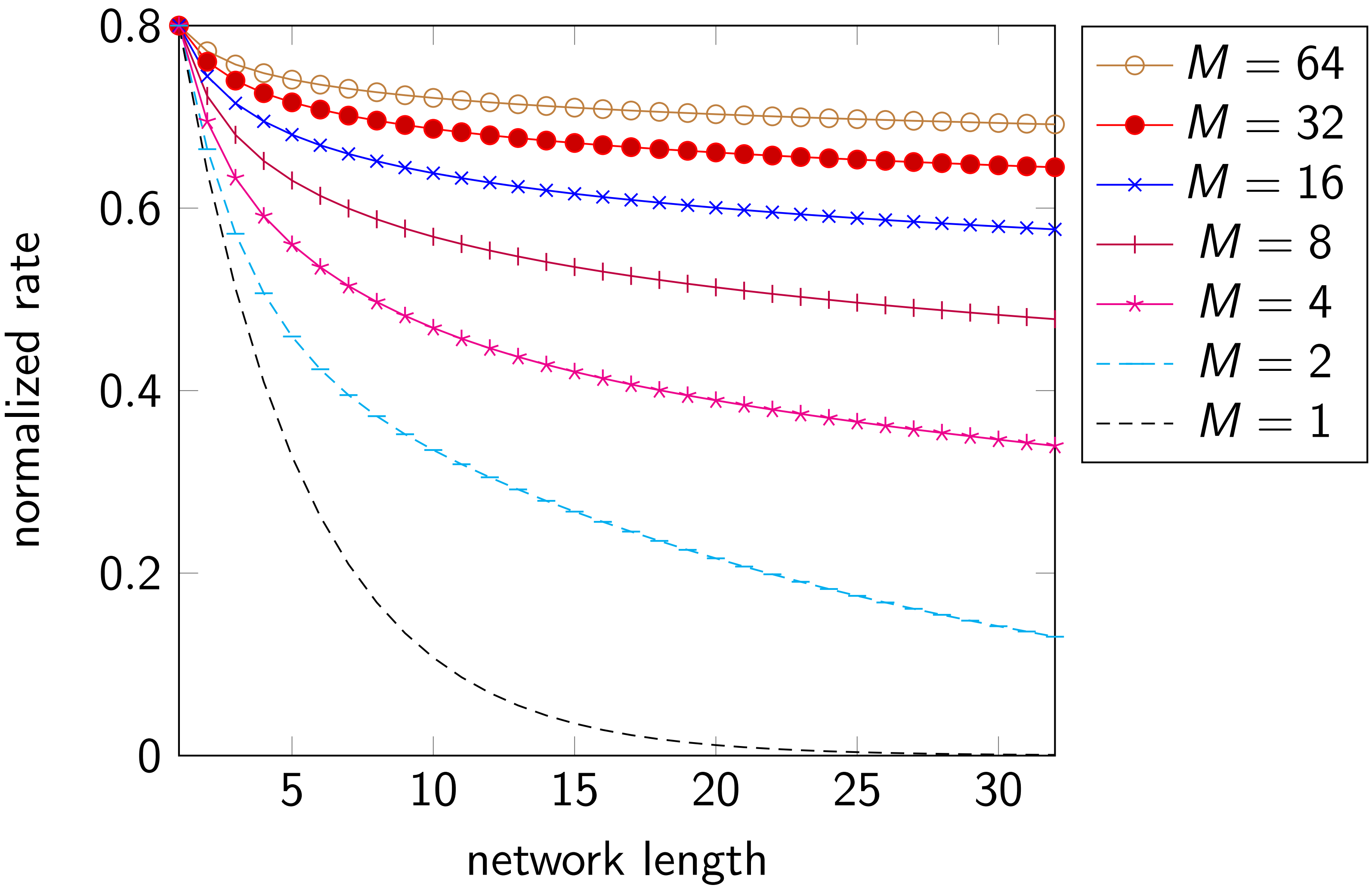
Source node encoding		$\mathcal{O}(TM)$ per packet
Destination node decoding		$\mathcal{O}(M^2 + TM)$ per packet
Intermediate Node	buffer	$\mathcal{O}(TM)$
	network coding	$\mathcal{O}(TM)$ per packet

T: length of a packet

K: number of packets

M: batch size

Achievable Rates for Line Networks



WiFi Experiments

Potential applications

- 5G mobile network
- Wireless mesh network
- Vehicular ad-hoc network
- Mobile ad-hoc network
- Satellite network
- Content delivery network (CDN)
- Internet of Things (IoT)
-
-

Productization

- An all-software prototype running BATS code was recently built.
- Source node, relay nodes, and receiving nodes are all notebook computers.
- A notebook with Intel i7 CPU was employed for decoding.
- A transmission rate > 500 Mb/s was achieved.
- Will collaborate with P2MT to implement BATS code in mesh network products (802.11).

Summary

- BATS codes provide a digital fountain solution with linear network coding:
  - Outer code at the source node is a matrix fountain code.
  - Linear network coding at the intermediate nodes forms the inner code.
  - Prevents BOTH packet loss and delay from accumulating along the way.
- The more hops between the source node and the sink node, the larger the benefit.
- Future work:
  - Finite-length analysis
  - Proof of (nearly) capacity achieving
  - Design of intermediate operations to maximize the throughput and minimize the buffer size

[NY13] T. C. Ng and S. Yang, Finite length analysis of BATS codes, NetCod 2013.