Multilevel Diversity Coding with Secure Regeneration: Separate Coding Achieves the MBR Point

Shuo Shao [1,*] **, Tie Liu** [2] **, Chao Tian** [2] **and Cong Shen** [3]

1   Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
2   Department of Electrical and Computer Engineering, Texas A&M University,
     College Station, TX 77843, USA; tieliu@tamu.edu (T.L.); chao.tian@tamu.edu (C.T.)
3   Department of Electronic Engineering and Information Science, University of Science and Technology of
     China, Hefei 230026, China; congshen@ustc.edu.cn
*   Correspondence: shuoshao@sjtu.edu.cn

Abstract: The problem of multilevel diversity coding with secure regeneration (MDC-SR) is considered, which includes the problems of multilevel diversity coding with regeneration (MDC-R) and secure regenerating code (SRC) as special cases. Two outer bounds are established, showing that separate coding can achieve the minimum-bandwidth-regeneration (MBR) point of the achievable normalized storage-capacity repair-bandwidth trade-off regions for the general MDC-SR problem. The core of the new converse results is an exchange lemma, which can be established using Han's subset inequality.

1. Introduction

Diversity coding and node repair are two fundamental ingredients of reliable distributed storage systems. While the study of diversity coding has been in the literature for decades [1–6], systematic studies of node repair mechanisms were started only recently by Dimakis et al. in their pioneering work [7]. A particular model, which was first introduced in [7] and has since received a significant amount of attention in the literature [8–20], is the so-called *(exact-repair) regenerating code (RC)* problem.

More specifically, in an (n, k, d) RC problem, a file M of size B is to be encoded in a total of n distributed storage nodes, each of capacity α. The encoding needs to ensure that the file M can be perfectly recovered by having full access to any k out of the total n storage nodes. In addition, when node failures occur and there are only d remaining nodes in the system, it is required that the data originally stored in any failed node can be recovered by downloading data of size β from each one of the d remaining nodes. An interesting technical challenge is to characterize the optimal *trade-offs* between the node capacity α and the download bandwidth β in satisfying both the file-recovery and node-repair requirements, which was studied in [8–20]. However, despite intensive research efforts that have yielded many interesting and highly non-trivial partial results including a precise characterization of the *minimum-storage-regenerating (MSR)* and the *minimum-bandwidth-regenerating (MBR)* rate points, the optimal trade-offs between the node capacity α and the download bandwidth β have *not* been fully understood for the general RC problem.

More recently, two extensions of the RC problem, namely *multilevel diversity coding with regeneration (MDC-R)* and *secure regenerating code (SRC)*, have also been studied in the literature. The problem of MDC-R was first introduced by Tian and Liu [21]. In an (n, d) MDC-R problem, a total of d independent files $M_1, \ldots, M_d$ of size $B_1, \ldots, B_d$, respectively, are to be stored in n distributed storage nodes, each

of capacity α. The encoding needs to ensure that the file M_j can be perfectly recovered by having full access to any j out of the total n storage nodes for any $j \in \{1, \dots, d\}$. In addition, when node failures occur and there are only d remaining nodes in the system, it is required that the data originally stored in any failed node can be recovered by downloading data of size β from each one of the d remaining nodes.

Clearly, an (n, k, d) RC problem can be viewed as an (n, d) MDC-R problem with *degenerate* messages $(\mathsf{M}_j : j \neq k)$ (i.e., $B_j = 0$ for all $j \neq k$). Therefore, from the code construction perspective, it is natural to consider the so-called *separate coding* scheme, i.e., to construct a code for the (n, d) MDC-R problem, we can simply use an (n, j, d) RC to encode the file M_j for each $j \in \{1, \dots, d\}$, and the coded messages for each file remain separate when stored in the storage nodes and during the repair processes. However, despite being a natural scheme, it was shown in [21] that separate coding is in general *suboptimal* in achieving the optimal trade-offs between the normalized storage-capacity and repair-bandwidth. On the other hand, it has been shown that separate coding can, in fact, achieve both the MSR [21] and the MBR [22] points of the achievable normalized storage-capacity and repair-bandwidth trade-off region for the general MDC-R problem.

The problem of SRC is an extension of the RC problem that further requires security guarantees during the repair processes. More specifically, the (n, k, d, ℓ) SRC problem that we consider is the (n, k, d) RC problem [7–16], with the additional constraint that the file M needs to be kept *information-theoretically* secure against an eavesdropper, which can access the data downloaded to regenerate a total of ℓ different failed nodes under all possible repair groups. Obviously, this is only possible when $\ell < k$. Furthermore, when $\ell = 0$, the secrecy requirement degenerates, and the (n, k, d, ℓ) SRC problem reduces to the (n, k, d) RC problem *without* any repair secrecy requirement.

Under the additional require secrecy requirement ($\ell \geq 1$), the optimal trade-offs between the node capacity α and repair bandwidth β have been studied in [23–30]. In particular, Shah, Rashmi and Kumar [25] showed that a particular trade-off point (referred to as the *SRK* point as the three first letters of the authors' names) can be achieved by extending an MBR code based on the product-matrix construction proposed in [8]. Later, it was shown [30] that, for any given (k, d) pair, there is a lower bound on ℓ, denoted by $\ell^*(k, d)$, such that, when $\ell \geq \ell^*(k, d)$, the SRK point is the *only* corner point of the trade-off region for the (n, k, d, ℓ) SRC problem. On the other hand, when $1 \leq \ell < \ell^*(k, d)$, it is possible that the trade-off region features *multiple* corner points, even though a precise characterization of the trade-off region, including both the MSR and the MBR points, remains missing in general.

In this paper, we introduce the problem of *multilevel diversity coding with secure regeneration (MDC-SR)* (The problem of secure multilevel diversity coding *without* any node regeneration requirement has been considered in [6,31].), which includes the problems of MDC-R and SRC as two special cases. In this model, multiple files are to be stored distributed in several storage nodes, like what in the *Multilevel Diversity Coding* problem. The system requires that, if a user can fully access some of the nodes, then the user can recover the corresponding part of the original files. Meanwhile, if any storage node failed, it can be regenerated by downloading messages from other nodes within a certain bandwidth limit. Additionally, if some nodes and repairing messages are leaked to an eavesdropper, the original files can still be information that is theoretically secure. The detailed definition of this model can be found in the next section. Similar to the MDC-R problem, it is natural to consider the separate coding scheme for the MDC-SR problem as well. Our main contribution consisted of three parts. Firstly, we established two nontrivial outer bounds for the MDC-SR problem. The secrecy constraint in the MDC-SR problem makes the outer bounding its trade-off region, not a simple extension of the bounding technic of the MDC-R problem in [22]. Secondly, we addressed a coding scheme with a separate coding structure that can achieve the intersection of the two outer bounds that we established, hence we can show that the optimality of separate coding in terms of achieving the MBR point of the achievable normalized storage-capacity and repair-bandwidth trade-off region extends more generally from the MDC-R problem to the MDC-SR problem. Last but not the least, during the process of establishing the two outer bounds, we proposed a lemma called *Exchange*

Lemma, which we believe can be used widely in other similar or even more generalized problems. We need to mention that our system model and main results can be degenerated to some unknown results. For example, when specialized to the SRC problem, our result shows that the SRK point [25] is, in fact, the MBR point of the achievable normalized storage-capacity and repair-bandwidth trade-off region, *regardless* of the number of corner points of the trade-off region.

From the technical viewpoint, this is mainly accomplished by establishing two outer bounds (one of them must be "horizontal", i.e., on the normalized repair-bandwidth *only*) on the achievable normalized storage-capacity and repair-bandwidth trade-off region, which intersect precisely at the superposition of the SRK points. The core of the new converse results is an *exchange* lemma, which we establish by exploiting the built-in symmetry of the problem via Han's subset inequality [32]. The meaning of "exchange" will be clear from the statement of the lemma. The lemma *only* relies on the functional dependencies for the repair processes and might be useful for solving some other related problems as well.

The rest of the paper is organized as follows. In Section 2, we formally introduce the problem of MDC-SR and the separate coding scheme. The main results of the paper are then presented in Section 3. In Section 4, we introduce the exchange lemma and use it to establish the main results of the paper. Finally, we conclude the paper in Section 5.

Notation and Remarks. Sets and random variables will be written in calligraphic and sans-serif fonts respectively, to differentiate from the real numbers written in normal math fonts. For any two integers $t \leq t'$, we shall denote the set of consecutive integers $\{t, t+1, \ldots, t'\}$ by $[t : t']$. The use of the brackets will be supressed otherwise.

Though many remarkable previous works are mentioned in this introduction, some of them, in fact, are more related to our work, such as [15,25,29]. We list them for the best convenience of our readers.

2. The MDC-SR Problem

Let $(n, d, N_1, \ldots, N_d, K, T, S)$ be a tuple of positive integers such that $d < n$. Formally, an $(n, d, N_1, \ldots, N_d, K, T, S)$ code consists of:

- for each $i \in [1 : n]$, a *message-encoding* function $f_i : \left(\prod_{j=1}^d [1 : N_j] \right) \times [1 : K] \to [1 : T]$;

- for each $\mathcal{A} \subseteq [1 : n] : |\mathcal{A}| \in [1 : d]$, a *message-decoding* function $g_{\mathcal{A}} : [1 : T]^{|\mathcal{A}|} \to [1 : N_{|\mathcal{A}|}]$;

- for each $\mathcal{B} \subseteq [1 : n] : |\mathcal{B}| = d$, $i' \in \mathcal{B}$, and $i \in [1 : n] \setminus \mathcal{B}$, a *repair-encoding* function $f_{i' \to i}^{\mathcal{B}} : [1 : T] \to [1 : S]$;

- for each $\mathcal{B} \subseteq [1 : n] : |\mathcal{B}| = d$ and $i \in [1 : n] \setminus \mathcal{B}$, a *repair-decoding* function $g_i^{\mathcal{B}} : [1 : S]^d \to [1 : T]$.

For each $j \in [1 : d]$, let M_j be a message that is uniformly distributed over $[1 : N_j]$. The messages $\mathsf{M}_1, \ldots, \mathsf{M}_d$ are assumed to be mutually independent. Let K be a random key that is uniformly distributed over $[1 : K]$ and independent of the messages $(\mathsf{M}_1, \ldots, \mathsf{M}_d)$. For each $i \in [1 : n]$, let $\mathsf{W}_i = f_i(\mathsf{M}_1, \ldots, \mathsf{M}_d, \mathsf{K})$ be the data stored at the ith storage node, and for each $\mathcal{B} \subseteq [1 : n] : |\mathcal{B}| = d$, $i' \in \mathcal{B}$, and $i \in [1 : n] \setminus \mathcal{B}$, let $\mathsf{S}_{i' \to i}^{\mathcal{B}} = f_{i' \to i}^{\mathcal{B}}(\mathsf{W}_{i'})$ be the data downloaded from the i'th storage node in order to regenerate the data originally stored at the ith storage node under the context of repair group $\mathcal{B}$. Obviously,

$$(B_j = \log N_j : j \in [1 : d]), \quad \alpha = \log T, \quad \text{and} \quad \beta = \log S$$

represent the message sizes, storage capacity, and repair bandwidth, respectively.

A normalized message-rate storage-capacity repair-bandwidth tuple $(\bar{B}_{\ell+1}, \ldots, \bar{B}_d, \bar{\alpha}, \bar{\beta})$ is said to be *achievable* for the (n, d, ℓ) MDC-SR problem if an $(n, d, 1, \ldots, 1, N_{\ell+1}, \ldots, N_d, K, T, S)$ code (i.e., $N_j = 1$ for all $j \in [1 : \ell]$) can be found such that:

- (rate normalization)

$$\frac{\alpha}{\sum_{t=\ell+1}^{d} B_t} = \bar{\alpha}, \quad \frac{\beta}{\sum_{t=\ell+1}^{d} B_t} = \bar{\beta}, \quad \frac{B_j}{\sum_{t=\ell+1}^{d} B_t} = \bar{B}_j \tag{1}$$

for any $j \in [\ell + 1 : d]$;

- (message recovery)

$$\mathsf{M}_{|\mathcal{A}|} = g_{\mathcal{A}}(\mathsf{W}_i : i \in \mathcal{A}) \tag{2}$$

for any $\mathcal{A} \subseteq [1 : n] : |\mathcal{A}| \in [\ell + 1 : d]$;

- (node regeneration)

$$\mathsf{W}_i = g_i^{\mathcal{B}}(\mathsf{S}_{i' \to i}^{\mathcal{B}} : i' \in \mathcal{B}) \tag{3}$$

for any $\mathcal{B} \subseteq [1 : n] : |\mathcal{B}| = d$ and $i \in [1 : n] \setminus \mathcal{B}$;

- (repair secrecy)

$$I((\mathsf{M}_{\ell+1}, \ldots, \mathsf{M}_d); (\mathsf{S}_{\to i} : i \in \mathcal{E})) = 0 \tag{4}$$

for any $\mathcal{E} \subseteq [1 : n]$ such that $|\mathcal{E}| = \ell$, where $\mathsf{S}_{\to i} := (\mathsf{S}_{i' \to i}^{\mathcal{B}} : \mathcal{B} \subseteq [1 : n], |\mathcal{B}| = d, \mathcal{B} \not\ni i, i' \in \mathcal{B})$ is the collection of data that can be downloaded from the other nodes to regenerate node i.

The closure of all achievable $(\bar{B}_{\ell+1}, \ldots, \bar{B}_d, \bar{\alpha}, \bar{\beta})$ tuples is the *achievable normalized message-rate storage-capacity repair-bandwidth trade-off region* $\mathcal{R}_{n,d,\ell}$ for the (n, d, ℓ) MDC-SR problem. For a fixed normalized message-rate tuple $(\bar{B}_{\ell+1}, \ldots, \bar{B}_d)$, the *achievable normalized storage-capacity repair-bandwidth trade-off region* is the collection of all normalized storage-capacity repair-bandwidth pairs $(\bar{\alpha}, \bar{\beta})$ such that $(\bar{B}_{\ell+1}, \ldots, \bar{B}_d, \bar{\alpha}, \bar{\beta}) \in \mathcal{R}_{n,d,\ell}$ and is denoted by $\mathcal{R}_{n,d,\ell}(\bar{B}_{\ell+1}, \ldots, \bar{B}_d)$.

Based on the above problem formulation, it should be clear that the MDC-SR problem includes several open problems of recent interest:

(1) the achievable normalized storage-capacity repair-bandwidth trade-off region $\mathcal{R}_{n,d}(\bar{B}_1, \ldots, \bar{B}_d)$ of the (n, d) MDC-R problem is simply $\mathcal{R}_{n,d,0}(\bar{B}_1, \ldots, \bar{B}_d)$ for any given normalized message-rate tuple $(\bar{B}_1, \ldots, \bar{B}_d)$,

(2) the achievable normalized storage-capacity repair-bandwidth trade-off region $\mathcal{R}_{n,k,d,\ell}$ of the (n, k, d, ℓ) SRC problem is simply $\mathcal{R}_{n,d,\ell}(0, \ldots, 0, \bar{B}_k = 1, 0, \ldots, 0)$,

(3) the achievable normalized storage-capacity repair-bandwidth trade-off region $\mathcal{R}_{n,k,d}$ of the (n, k, d) RC problem is simply $\mathcal{R}_{n,d}(0, \ldots, 0, \bar{B}_k = 1, 0, \ldots, 0)$ or, equivalently, $\mathcal{R}_{n,k,d,0}$.

Given these connections, our problem formulation can be viewed as providing a unified framework to investigate these closely-related problems.

A simple and natural strategy for constructing a code for the (n, d, ℓ) MDC-SR problem is to use to an (n, j, d, ℓ) SRC to encode the message M_j separately for each $j \in [\ell + 1 : d]$. Since the coded data are kept separate during the encoding, decoding and repair processes, we have

$$K = \prod_{j=\ell+1}^{d} K_j, \quad T = \prod_{j=\ell+1}^{d} T_j, \quad \text{and} \quad S = \prod_{j=\ell+1}^{d} S_j.$$

Thus, for the general MDC-SR problem, the *separate coding normalized storage-capacity repair-bandwidth trade-off region* $\hat{\mathcal{R}}_{n,d,\ell}(\bar{B}_{\ell+1}, \ldots, \bar{B}_d)$ for a fixed normalized message-rate tuple $(\bar{B}_{\ell+1}, \ldots, \bar{B}_d)$ is given by:

$$\left(\left(\sum_{j=\ell+1}^{d} \bar{\alpha}_j \bar{B}_j, \sum_{j=\ell+1}^{d} \bar{\beta}_j \bar{B}_j \right) : (\bar{\alpha}_j, \bar{\beta}_j) \in \mathcal{R}_{n,j,d,\ell} \right). \tag{5}$$

As mentioned previously, when $\ell = 0$, the repair secrecy requirement (4) degenerates, and the (n, d, ℓ) MDC-SR problem reduces to the (n, d) MDC-R problem. In this case, it was shown in [22] that any achievable normalized message-rate storage-capacity repair-bandwidth tuple $(\bar{B}_1, \ldots, \bar{B}_d, \bar{\alpha}, \bar{\beta}) \in \mathcal{R}_{n,d}$ must satisfy:

$$\bar{\beta} \geq \sum_{j=1}^{d} T_{d,j}^{-1} \bar{B}_j, \tag{6}$$

$$\text{and} \quad \bar{\alpha} + \frac{d(d-1)}{2} \bar{\beta} \geq \frac{d(d+1)}{2} \sum_{j=1}^{d} T_{d,j}^{-1} \bar{B}_j, \tag{7}$$

where $T_{d,j} := \sum_{t=1}^{j} (d + 1 - t)$. When set as equalities, the intersection of (6) and (7) is given by:

$$(\bar{\alpha}, \bar{\beta}) = \left(d \sum_{j=1}^{d} T_{d,j}^{-1} \bar{B}_j, \sum_{j=1}^{d} T_{d,j}^{-1} \bar{B}_j \right).$$

For any $j \in [1 : d]$, the MBR point for the (n, j, d) RC problem can be written as [8]

$$\left(dT_{d,j}^{-1}, T_{d,j}^{-1} \right) \in \mathcal{R}_{n,j,d}. \tag{8}$$

We may thus conclude immediately from (5) (with $\ell = 0$) that separate coding can achieve the MBR point for the general MDC-R problem.

Figure 1 shows the optimal trade-off curve between the normalized storage-capacity and repair-bandwidth and the best possible trade-offs that can be achieved by separate coding for the $(4, 3)$ MDC-R problem with $(\bar{B}_1, \bar{B}_2, \bar{B}_3) = (0, 1/3, 2/3)$ [21]. Clearly, for this example, separate coding is strictly suboptimal when $\bar{\alpha} \in (5/12, 1/2)$. On the other hand, when $\bar{\alpha} \leq 5/12$ or $\bar{\alpha} \geq 1/2$, separate coding can, in fact, achieve the optimal trade-offs. In particular, separate encoding can achieve the MSR point $(7/18, 11/36)$ and the MBR point $(8/15, 8/45)$. In the same figure, the outer bounds (6) and (7) have also been plotted. As illustrated, they intersect precisely at the MBR point $(8/15, 8/45)$. Notice that, for this example at least, the outer bound (7) is tight *only* at the MBR point.
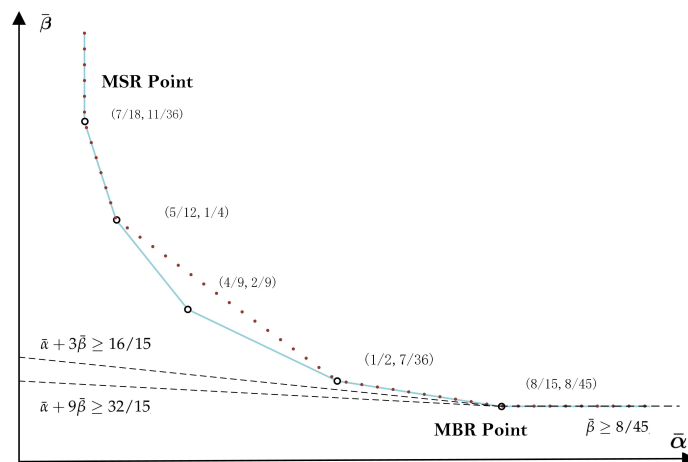
Figure 1. The optimal trade-off curve between the normalized storage-capacity $\bar{\alpha}$ and repair-bandwidth $\bar{\beta}$ (the solid line) and the best possible trade-offs that can be achieved by separate coding (dashed line) for the $(4,3)$ multilevel diversity coding with regeneration (MDC-R) problem with $(\bar{B}_1, \bar{B}_2, \bar{B}_3) = (0, 1/3, 2/3)$ (adapted from [21]). The outer bounds (6), (7) and (14) are evaluated as $\bar{\beta} \geq 8/45$, $\bar{\alpha} + 3\bar{\beta} \geq 16/15$, and $\bar{\alpha} + 9\bar{\beta} \geq 32/15$, respectively. When set as equalities, they intersect precisely at the MBR point $(8/15, 8/45)$.

3. Main Results

Our main result of the paper is to show that the optimality of separate coding in terms of achieving the MBR point of the normalized storage-capacity repair-bandwidth trade-off region extends more generally from the MDC-R problem to the MDC-SR problem. The results are summarized in the following theorem.

Theorem 1. *For the general MDC-SR problem, any achievable normalized message-rate storage-capacity repair-bandwidth tuple $(\bar{B}_{\ell+1}, \ldots, \bar{B}_d, \bar{\alpha}, \bar{\beta}) \in \mathcal{R}_{n,d,\ell}$ must satisfy:*

$$\bar{\beta} \geq \sum_{j=\ell+1}^{d} T_{d,j,\ell}^{-1} \bar{B}_j, \tag{9}$$

$$and \quad \bar{\alpha} + (d(d-\ell)-\ell)\bar{\beta} \geq (d-\ell)(d+1) \sum_{j=\ell+1}^{d} T_{d,j,\ell}^{-1} \bar{B}_j, \tag{10}$$

where $T_{d,k,\ell} := \sum_{t=\ell+1}^{k}(d+1-t)$. When set as equalities, the intersection of (9) and (10) is given by:

$$(\bar{\alpha}, \bar{\beta}) = \left(d \sum_{j=\ell+1}^{d} T_{d,j,\ell}^{-1} \bar{B}_j, \sum_{j=\ell+1}^{d} T_{d,j,\ell}^{-1} \bar{B}_j \right).$$

For any $j \in [\ell+1 : d]$, the SRK point for the (n, j, d, ℓ) SRC problem can be written as [25]:

$$(dT_{d,j,\ell}^{-1}, T_{d,j,\ell}^{-1}) \in \mathcal{R}_{n,j,d,\ell}. \tag{11}$$

We may thus conclude immediately from (5) that separate coding can achieve the MBR point for the general MDC-SR problem.

The following corollary follows immediately from Theorem 1 by setting $\bar{B}_j = 0$ for all $j \neq k$.

Corollary 1. *For the general SRC problem, any achievable normalized storage-capacity repair-bandwidth tuple* $(\bar{\alpha}, \bar{\beta}) \in \mathcal{R}_{n,k,d,\ell}$ *must satisfy:*

$$\bar{\beta} \geq T_{d,k,\ell}^{-1}, \tag{12}$$

$$and \quad \bar{\alpha} + (d(d - \ell) - \ell)\bar{\beta} \geq (d - \ell)(d + 1)T_{d,k,\ell}^{-1}. \tag{13}$$

When set as equalities, the intersection of (12) *and* (13) *is precisely the SRK point* (11) *(with* $j = k$*), showing that the SRK point is, in fact, the MBR point of the achievable normalized storage-capacity repair-bandwidth trade-off region for the general SRC problem.*

While the outer bound (12) is known [23,24,30], the outer bound (13) is new. Figure 2 shows the optimal trade-off curve between the normalized storage-capacity and repair-bandwidth for the $(7, 6, 6, 1)$ SRC problem. Notice that, for this example, the SRK point $(2/5, 1/15)$ is, in fact, the MBR point even though the trade-off region has two corner points. In the same figure, the outer bunds (12) and (13) have also been plotted. As illustrated, when set as equalities, they intersect precisely at the MBR/SRK point $(2/5, 1/15)$. Notice that for this example at least, the outer bound (13) is tight *only* at the MBR/SRK point.
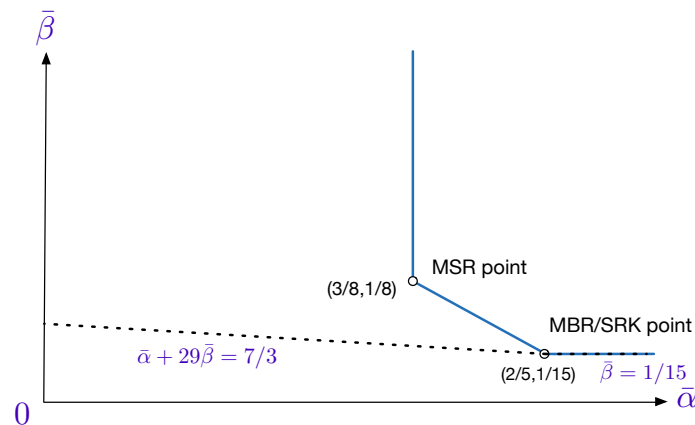

Figure 2. The optimal trade-off curve between the normalized storage-capacity $\bar{\alpha}$ and repair-bandwidth $\bar{\beta}$ for the $(7, 6, 6, 1)$ secure regenerating code (SRC) problem [30]. The outer bounds (12) and (13) are evaluated as $\bar{\beta} \geq 1/15$ and $\bar{\alpha} + 29\bar{\beta} \geq 7/3$, respectively. When set as equalities, they intersect precisely at the MBR/SRK point $(2/5, 1/15)$.

As a final remark, we mention here that when $\ell = 0$, the outer bound (9) is reduced to (6) for the (n, d) MDC-R problem by the fact that $T_{n,d,0} = T_{n,d}$. However, when $\ell = 0$, the outer bound (10) is reduced to:

$$\bar{\alpha} + d^2\bar{\beta} \geq d(d + 1) \sum_{j=1}^{d} T_{d,j}^{-1} \bar{B}_j, \tag{14}$$

which is *weaker* than the outer bound (7) by the fact that $d^2 > \frac{d(d-1)}{2}$. Figure 1 shows the outer bound (14) for the $(4, 3)$ MDC-R problem with $(\bar{B}_1, \bar{B}_2, \bar{B}_3) = (0, 1/3, 2/3)$. As illustrated, (14) is weaker than (7), and both are *only* tight at the MBR point $(8/15, 8/45)$.

4. Proof of the Main Results

Let us first outline the main ingredients for proving the outer bounds (9) and (10).

(1) *Total number of nodes.* To prove the outer bounds (9) and (10), let us first note that these bounds are *independent* of the total number of storage nodes n in the system. Therefore, in our proof, we only need to consider the cases where $n = d + 1$—for the cases where $n > d + 1$, since any

subsystem consisting of $d+1$ out of the total n storage nodes must give rise to a $(d+1, d, \ell)$ MDC-SR problem. Therefore, these *outer* bounds must apply as well. When $n = d+1$, any repair group $\mathcal{B}$ of size d is uniquely determined by the node j to be repaired, i.e., $\mathcal{B} = [1:n] \setminus \{j\}$, and hence can be dropped from the notation $S_{i \to j}^{\mathcal{B}}$ without causing any confusion.

(2) *Code symmetry.* Due to the built-in *symmetry* of the problem, to prove the outer bounds (9) and (10), we only need to consider the so-called *symmetrical* codes [10] for which the joint entropy of any subset of random variables from

$$((\mathsf{M}_1, \ldots, \mathsf{M}_d), \mathsf{K},$$
$$(\mathsf{W}_i : i \in [1:n]), (\mathsf{S}_{i \to j} : i, j \in [1:n], i \neq j))$$

remains *unchanged* under *any* permutation over the *storage-node* indices.

(3) *Key collections of random variables.* Focusing on the symmetrical $(n = d+1, d, N_1, \ldots, N_d, K, T, S)$ codes, the following collections of random variables play a key role in our proof:

$$\mathsf{M}_{\mathcal{A}} := (\mathsf{M}_i : i \in \mathcal{A}), \quad \mathcal{A} \subseteq [1:d],$$
$$\mathsf{M}^{(m)} := \mathsf{M}_{[1:m]}, \quad m \in [1:d],$$
$$\mathsf{W}_{\mathcal{A}} := (\mathsf{W}_i : i \in \mathcal{A}), \quad \mathcal{A} \subseteq [1:n],$$
$$\mathsf{S}_{i \to \mathcal{B}} := (\mathsf{S}_{i \to j} : j \in \mathcal{B}), \quad i \in [1:n], \mathcal{B} \subseteq [1:n] \setminus \{i\},$$
$$\mathsf{S}_{\mathcal{B} \to j} := (\mathsf{S}_{i \to j} : \quad i \in \mathcal{B}), j \in [1:n], \mathcal{B} \subseteq [1:n] \setminus \{j\},$$
$$\mathsf{S}_{\to j} := \mathsf{S}_{[1:j-1] \cup [j+1:n] \to j}, \quad j \in [1:n],$$
$$\mathsf{S}_{\to \mathcal{B}} := (\mathsf{S}_{\to j} : j \in \mathcal{B}), \quad \mathcal{B} \subseteq [1:n],$$
$$\underline{\mathsf{S}}_{\to j} := \mathsf{S}_{[1:j-1] \to j}, \quad j \in [1:n],$$
$$\underline{\mathsf{S}}_{\to \mathcal{B}} := (\underline{\mathsf{S}}_{\to j} : j \in \mathcal{B}), \quad \mathcal{B} \subseteq [1:n],$$
$$\overline{\mathsf{S}}_{\to j} := \mathsf{S}_{[j+1:n] \to j}, \quad j \in [1:n],$$
$$\overline{\mathsf{S}}_{\to \mathcal{B}} := (\overline{\mathsf{S}}_{\to j} : j \in \mathcal{B}), \quad \mathcal{B} \subseteq [1:n],$$
$$\mathsf{U}^{(t,s)} := (\mathsf{W}_{[1:t]}, \overline{\mathsf{S}}_{\to [t+1:s]}), \quad s \in [1:n], t \in [0:s],$$
$$\mathsf{U}^{(s)} := \mathsf{U}^{(0,s)}.$$

These collections of random variables have also been used in [22,30].

An important part of the proof is to understand the relations between the collections of random variables defined above, and to use them to derive the desired converse results. We shall discuss this next.

4.1. Technical Lemmas

Lemma 1. *For any* $(n = d+1, d, N_1, \ldots, N_d, K, T, S)$ *code that satisfies the node regeneration requirement* (3), $(\underline{\mathsf{S}}_{\to [t+1:s]}, \mathsf{W}_{[t+1:s]})$ *is a function of* $\mathsf{U}^{(t,s)}$ *for any* $s \in [1:n]$ *and* $t \in [0:s-1]$.

Proof of Lemma 1. Fix $s \in [1:n]$ and $t \in [0:s-1]$. Let us first note that $\underline{\mathsf{S}}_{\to t+1}$ is a function of $\mathsf{W}_{[1:t]}$. As a result, $\mathsf{S}_{\to t+1} = (\underline{\mathsf{S}}_{\to t+1}, \overline{\mathsf{S}}_{\to t+1})$ is a function of $\mathsf{U}^{(t,s)}$. It thus follows immediately from the node regeneration requirement (3) that W_{t+1} is a function of $\mathsf{U}^{(t,s)}$. Similarly and inductively, it can be shown that $(\underline{\mathsf{S}}_{\to j}, \mathsf{W}_j)$ is a function of $\mathsf{U}^{(t,s)}$ for all $j \in [t+2:s]$. This completes the proof of the lemma. $\square$

The above lemma demonstrates the "compactness" of $\mathsf{U}^{(t,s)}$ and has a number of direct consequences. For example, for any fixed $s \in [1:n]$, it is clear from Lemma 1 that $\mathsf{U}^{(t_2,s)}$ is a function of $\mathsf{U}^{(t_1,s)}$ and hence $H(\mathsf{U}^{(t_2,s)}) \le H(\mathsf{U}^{(t_1,s)})$ for any $0 \le t_1 \le t_2 \le s-1$.

The following lemma plays the key role in proving the outer bounds (6) and (7). The proof is rather long and is deferred to the Appendix to enhance the flow of the paper.

Lemma 2 (Exchange lemma). *For any symmetrical $(n = d + 1, d, N_1, \ldots, N_d, K, T, S)$ code that satisfies the node regeneration requirement (3), we have*

$$
\frac{d + 1 - j}{d - m} H(\mathsf{U}^{(i,m)}|\mathsf{M}^{(m)}) + H(\mathsf{U}^{(i',j)}|\mathsf{M}^{(m)})
$$
$$
\geq \frac{d + 1 - j}{d - m} H(\mathsf{U}^{(i,m+1)}|\mathsf{M}^{(m)}) + H(\mathsf{U}^{(i',j-1)}|\mathsf{M}^{(m)}) \tag{15}
$$

for any $m \in [1 : d - 1]$, $i \in [0 : m - 1]$, $i' \in [0 : i]$, and $j \in [i' + 1 : m - i + i' + 1]$.

Corollary 2. *For any symmetrical $(n = d + 1, d, N_1, \ldots, N_d, K, T, S)$ code that satisfies the node regeneration requirement (3), we have*

$$
T_{d,m,\ell}^{-1} H(\mathsf{U}^{(m)}|\mathsf{M}^{(m)}) \geq T_{d,m+1,\ell}^{-1} H(\mathsf{U}^{(m+1)}|\mathsf{M}^{(m)}) +
$$
$$
(T_{d,m,\ell}^{-1} - T_{d,m+1,\ell}^{-1}) H(\mathsf{U}^{(\ell)}|\mathsf{M}^{(m)}) \tag{16}
$$

for any $\ell \in [0 : d - 1]$ and $m \in [\ell + 1 : d - 1]$.

Proof of Corollary 2. Fix $\ell \in [0 : d - 1]$ and $m \in [\ell + 1 : d - 1]$. Setting $i = i' = 0$ in (15), we have

$$
\frac{d + 1 - j}{d - m} H(\mathsf{U}^{(m)}|\mathsf{M}^{(m)}) + H(\mathsf{U}^{(j)}|\mathsf{M}^{(m)})
$$
$$
\geq \frac{d + 1 - j}{d - m} H(\mathsf{U}^{(m+1)}|\mathsf{M}^{(m)}) + H(\mathsf{U}^{(j-1)}|\mathsf{M}^{(m)}) \tag{17}
$$

for any $j \in [1 : m + 1]$. Add the inequalities (17) for $j \in [\ell + 1 : m]$ and cancel the common term $\sum_{j=\ell+1}^{m-1} H(\mathsf{U}^{(j)}|\mathsf{M}^{(m)})$ from both sides. We have

$$
\frac{T_{d,m,\ell}}{d - m} H(\mathsf{U}^{(m)}|\mathsf{M}^{(m)}) + H(\mathsf{U}^{(m)}|\mathsf{M}^{(m)})
$$
$$
\geq \frac{T_{d,m,\ell}}{d - m} H(\mathsf{U}^{(m+1)}|\mathsf{M}^{(m)}) + H(\mathsf{U}^{(\ell)}|\mathsf{M}^{(m)}),
$$

which can be equivalently written as

$$
\frac{T_{d,m+1,\ell}}{d - m} H(\mathsf{U}^{(m)}|\mathsf{M}^{(m)})
$$
$$
\geq \frac{T_{d,m,\ell}}{d - m} H(\mathsf{U}^{(m+1)}|\mathsf{M}^{(m)}) + H(\mathsf{U}^{(\ell)}|\mathsf{M}^{(m)}) \tag{18}
$$

by the fact that $T_{d,m,\ell} + (d - m) = T_{d,m+1,\ell}$. Multiplying both sides of (18) by

$$
\frac{d - m}{T_{d,m+1,\ell} T_{d,m,\ell}} = T_{d,m,\ell}^{-1} - T_{d,m+1,\ell}^{-1}
$$

completes the proof of (16). □

Corollary 3. *For any symmetrical* $(n = d + 1, d, N_1, \ldots, N_d, K, T, S)$ *code that satisfies the node regeneration requirement* (3), *we have*

$$
\begin{aligned}
&H(\mathsf{U}^{(1,m)}|\mathsf{M}^{(m)}) + (d - m)T_{d,m,\ell}^{-1} H(\mathsf{U}^{(m)}|\mathsf{M}^{(m)}) \\
&\geq H(\mathsf{U}^{(1,m+1)}|\mathsf{M}^{(m)}) + (d - m)T_{d,m,\ell}^{-1} H(\mathsf{U}^{(\ell)}|\mathsf{M}^{(m)})
\end{aligned}
\tag{19}
$$

for any $\ell \in [0 : d - 1]$ *and* $m \in [\ell + 1 : d - 1]$.

Proof of Corollary 3. Fix $\ell \in [0 : d - 1]$ and $m \in [\ell + 1 : d - 1]$. Set $i = 1$ and $i' = 0$ in (15). We have

$$
\begin{aligned}
&\frac{d + 1 - j}{d - m} H(\mathsf{U}^{(1,m)}|\mathsf{M}^{(m)}) + H(\mathsf{U}^{(j)}|\mathsf{M}^{(m)}) \\
&\geq \frac{d + 1 - j}{d - m} H(\mathsf{U}^{(1,m+1)}|\mathsf{M}^{(m)}) + H(\mathsf{U}^{(j-1)}|\mathsf{M}^{(m)})
\end{aligned}
\tag{20}
$$

for any $j \in [1 : m]$. Add the inequalities (20) for $j \in [\ell + 1 : m]$ and cancel the common term $\sum_{j=\ell+1}^{m-1} H(\mathsf{U}^{(j)}|\mathsf{M}^{(m)})$ from both sides. We have

$$
\begin{aligned}
&\frac{T_{d,m,\ell}}{d - m} H(\mathsf{U}^{(1,m)}|\mathsf{M}^{(m)}) + H(\mathsf{U}^{(m)}|\mathsf{M}^{(m)}) \\
&\geq \frac{T_{d,m,\ell}}{d - m} H(\mathsf{U}^{(1,m+1)}|\mathsf{M}^{(m)}) + H(\mathsf{U}^{(\ell)}|\mathsf{M}^{(m)}).
\end{aligned}
\tag{21}
$$

Multiplying both sides of (21) by $(d - m)T_{d,m,\ell}^{-1}$ completes the proof of (19).　□

4.2. The Proof

Consider a symmetrical $(n = d + 1, d, 1, \ldots, 1, N_{\ell+1}, \ldots, N_d, K, T, S)$ regenerating code that satisfies the rate normalization requirement (1), the message recovery requirement (2), the node regeneration requirement (3), and the repair secrecy requirement (4). Let us first prove a few intermediate results. The outer bounds (9) and (10) will then follow immediately.

Proposition 1.

$$
\begin{aligned}
\frac{1}{d - \ell} H(\mathsf{U}^{(\ell+1)}) &\geq \sum_{j=\ell+1}^{m} T_{d,j,\ell}^{-1} B_j + \\
&T_{d,m,\ell}^{-1} H(\mathsf{U}^{(m)}|\mathsf{M}_{[\ell+1:m]}) + \left(\frac{1}{d - \ell} - T_{d,m,\ell}^{-1} \right) H(\mathsf{U}^{(\ell)})
\end{aligned}
\tag{22}
$$

for any $m \in [\ell + 1 : d]$. *Consequently,*

$$
\frac{1}{d - \ell} H(\mathsf{U}^{(\ell+1)}) \geq \sum_{j=\ell+1}^{d} T_{d,j,\ell}^{-1} B_j + \frac{1}{d - \ell} H(\mathsf{U}^{(\ell)}).
\tag{23}
$$

Proof of Proposition 1. To see (22), consider proof by induction. For the base case with $m = \ell + 1$, we have

$$\frac{1}{d-\ell} H(\mathsf{U}^{(\ell+1)}) \overset{(a)}{=} \frac{1}{d-\ell} H(\mathsf{U}^{(\ell+1)}, \mathsf{M}_{\ell+1})$$

$$\overset{(b)}{=} \frac{1}{d-\ell} \left(H(\mathsf{M}_{\ell+1}) + H(\mathsf{U}^{(\ell+1)} | \mathsf{M}_{\ell+1}) \right)$$

$$\overset{(c)}{=} \frac{1}{d-\ell} \left(B_{\ell+1} + H(\mathsf{U}^{(\ell+1)} | \mathsf{M}_{\ell+1}) \right)$$

$$\overset{(d)}{=} T_{d,\ell+1,\ell}^{-1} B_{\ell+1} + T_{d,\ell+1,\ell}^{-1} H(\mathsf{U}^{(\ell+1)} | \mathsf{M}_{\ell+1}),$$

where (a) follows from the fact that $\mathsf{M}_{\ell+1}$ is a function of $\mathsf{W}_{[1:\ell+1]}$, which is a function of $\mathsf{U}^{(\ell+1)}$ by Lemma 1; (b) follows from the chain rule for entropy; (c) follows from the fact that $H(\mathsf{M}_{\ell+1}) = B_{\ell+1}$; and (d) follows from the fact that $T_{d,\ell+1,\ell} = d - \ell$. Assuming that (22) holds for some $m \in [\ell + 1 : d - 1]$, we have

$$\frac{1}{d-\ell} H(\mathsf{U}^{(\ell+1)})$$

$$\overset{(a)}{\geq} \sum_{j=\ell+1}^{m} T_{d,j,\ell}^{-1} B_j + T_{d,m,\ell}^{-1} H(\mathsf{U}^{(m)} | \mathsf{M}_{[\ell+1:m]}) +$$
$$\left(\frac{1}{d-\ell} - T_{d,m,\ell}^{-1} \right) H(\mathsf{U}^{(\ell)})$$

$$\overset{(b)}{\geq} \sum_{j=\ell+1}^{m} T_{d,j,\ell}^{-1} B_j + T_{d,m+1,\ell}^{-1} H(\mathsf{U}^{(m+1)} | \mathsf{M}_{[\ell+1:m]}) +$$
$$\left(\frac{1}{d-\ell} - T_{d,m+1,\ell}^{-1} \right) H(\mathsf{U}^{(\ell)})$$

$$\overset{(c)}{\geq} \sum_{j=\ell+1}^{m} T_{d,j,\ell}^{-1} B_j + T_{d,m+1,\ell}^{-1} H(\mathsf{U}^{(m+1)}, \mathsf{M}_{m+1} | \mathsf{M}_{[\ell+1:m]}) +$$
$$\left(\frac{1}{d-\ell} - T_{d,m+1,\ell}^{-1} \right) H(\mathsf{U}^{(\ell)})$$

$$\overset{(d)}{=} \sum_{j=\ell+1}^{m} T_{d,j,\ell}^{-1} B_j + T_{d,m+1,\ell}^{-1} H(\mathsf{M}_{m+1} | \mathsf{M}_{[\ell+1:m]}) +$$
$$T_{d,m+1,\ell}^{-1} H(\mathsf{U}^{(m+1)} | \mathsf{M}_{[\ell+1:m+1]}) +$$
$$\left(\frac{1}{d-\ell} - T_{d,m+1,\ell}^{-1} \right) H(\mathsf{U}^{(\ell)})$$

$$\overset{(e)}{=} \sum_{j=\ell+1}^{m} T_{d,j,\ell}^{-1} B_j + T_{d,m+1,\ell}^{-1} B_{m+1} +$$
$$T_{d,m+1,\ell}^{-1} H(\mathsf{U}^{(m+1)} | \mathsf{M}_{[\ell+1:m+1]}) +$$
$$\left(\frac{1}{d-\ell} - T_{d,m+1,\ell}^{-1} \right) H(\mathsf{U}^{(\ell)})$$

$$= \sum_{j=\ell+1}^{m+1} T_{d,j,\ell}^{-1} + T_{d,m+1,\ell}^{-1} H(\mathsf{U}^{(m+1)} | \mathsf{M}_{[\ell+1:m+1]}) +$$
$$\left(\frac{1}{d-\ell} - T_{d,m+1,\ell}^{-1} \right) H(\mathsf{U}^{(\ell)}),$$

where (a) follows from the induction assumption; (b) follows from Corollary 2; (c) follows from the fact that M_{m+1} is a function of $\mathsf{W}_{[1:m+1]}$, which is a function of $\mathsf{U}^{(m+1)}$ by Lemma 1; (d) follows from

the chain rule for entropy; and (e) follows from the facts that M_{m+1} is independent of $M_{[\ell+1:m]}$ and that $H(M_{m+1}) = B_{m+1}$. This completes the induction step and hence the proof of (22).

To see (23), simply set $m = d$ in (22). We have

$$\frac{1}{d-\ell}H(U^{(\ell+1)}) \geq \sum_{j=\ell+1}^{d} T_{d,j,\ell}^{-1} B_j +$$
$$T_{d,d,\ell}^{-1} H(U^{(d)}|M_{[\ell+1:d]}) + \left(\frac{1}{d-\ell} - T_{d,d,\ell}^{-1}\right) H(U^{(\ell)}). \tag{24}$$

Note that

$$H(U^{(d)}|M_{[\ell+1:d]}) \geq H(U^{(\ell)}|M_{[\ell+1:d]}) = H(U^{(\ell)}) \tag{25}$$

where the last equality follows from the fact that $I(U^{(\ell)}; M_{[\ell+1:d]}) = 0$ by the repair secrecy requirement (4). Substituting (25) into (24) completes the proof of (23). □

Proposition 2.

$$H(S_{d+1\to[1:\ell]}) + (d(d-\ell) - \ell)\beta + dH(U^{(\ell)}) \geq dH(U^{(\ell+1)}). \tag{26}$$

Proof of Proposition 2. First note that, for any $m \in [1 : \ell]$, we have

$$H(S_{d+1\to[1:m]}) + H(U^{(\ell)})$$
$$\overset{(a)}{=} H(S_{d+1\to[1:m-1]\cup\{\ell+1\}}) + H(U^{(\ell)})$$
$$\overset{(b)}{\geq} H(S_{d+1\to[1:m-1]}) + H(U^{(\ell)}, S_{d+1\to\ell+1}), \tag{27}$$

where (a) follows from the fact that $H(S_{d+1\to[1:m]}) = H(S_{d+1\to[1:m-1]\cup\{\ell+1\}})$ due to the symmetrical code that we consider, and (b) follows from the submodularity of the entropy function. Add (27) over $m \in [1 : \ell]$ and cancel $\sum_{m=1}^{\ell-1} H(S_{d+1\to[1:m]})$ from both sides. We have

$$H(S_{d+1\to[1:\ell]}) + \ell H(U^{(\ell)}) \geq \ell H(U^{(\ell)}, S_{d+1\to\ell+1}). \tag{28}$$

It follows that

$$H(S_{d+1\to[1:\ell]}) + (d(d-\ell) - \ell)\beta + dH(U^{(\ell)})$$
$$= \left(H(S_{d+1\to[1:\ell]}) + \ell H(U^{(\ell)})\right) +$$
$$(d(d-\ell) - \ell)\beta + (d-\ell)H(U^{(\ell)})$$
$$\overset{(a)}{\geq} \ell H(U^{(\ell)}, S_{d+1\to\ell+1}) + (d(d-\ell) - \ell)\beta + (d-\ell)H(U^{(\ell)})$$
$$= \ell\left((d-\ell-1)\beta + H(U^{(\ell)}, S_{d+1\to\ell+1})\right) +$$
$$(d-\ell)\left((d-\ell)\beta + H(U^{(\ell)})\right)$$
$$\overset{(b)}{\geq} \ell\left(H(S_{[\ell+2:d]\to\ell+1}) + H(U^{(\ell)}, S_{d+1\to\ell+1})\right) +$$
$$(d-\ell)\left(H(\bar{S}_{\to\ell+1}) + H(U^{(\ell)})\right)$$
$$\overset{(c)}{\geq} \ell H(U^{(\ell+1)}) + (d-\ell)H(U^{(\ell+1)})$$
$$= dH(U^{(\ell+1)}),$$

where (*a*) follows from (28); (*b*) follows from the fact that $H(\mathsf{S}_{[\ell+2:d]\to\ell+1}) \leq (d-\ell-1)\beta$ and that $H(\overline{\mathsf{S}}_{\to\ell+1}) \leq (d-\ell)\beta$; and (*c*) follows from the fact that $H(\mathsf{S}_{[\ell+2:d]\to\ell+1}) + H(\mathsf{U}^{(\ell)}, \mathsf{S}_{d+1\to\ell+1}) \geq H(\mathsf{U}^{(\ell+1)})$ and that $H(\overline{\mathsf{S}}_{\to\ell+1}) + H(\mathsf{U}^{(\ell)}) \geq H(\mathsf{U}^{(\ell+1)})$ by the union bound on entropy. This completes the proof of the proposition. $\square$

Proposition 3.

$$H(\mathsf{U}^{(1,m)}) + \frac{d-m}{d-\ell} H(\mathsf{U}^{(\ell+1)})$$

$$\geq (d-m) \sum_{j=\ell+1}^{m} T_{d,j,\ell}^{-1} B_j + H(\mathsf{U}^{(1,m+1)}) + \frac{d-m}{d-\ell} H(\mathsf{U}^{(\ell)}) \tag{29}$$

for any $m \in [\ell+1, d-1]$. *Consequently,*

$$H(\mathsf{U}^{(1,\ell+1)}) + \frac{T_{d,d,\ell+1}}{d-\ell} H(\mathsf{U}^{(\ell+1)})$$

$$\geq T_{d,d,\ell} \sum_{j=\ell+1}^{d} T_{d,j,\ell}^{-1} B_j + \frac{T_{d,d,\ell}}{d-\ell} H(\mathsf{U}^{(\ell)}). \tag{30}$$

Proof of Proposition 3. To see (29), note that, for any $m \in [\ell+1, d-1]$, we have

$$H(\mathsf{U}^{(1,m)}|\mathsf{M}_{[\ell+1:m]}) + \frac{d-m}{d-\ell} H(\mathsf{U}^{(\ell+1)})$$

$$\overset{(a)}{\geq} H(\mathsf{U}^{(1,m)}|\mathsf{M}_{[\ell+1:m]}) + (d-m) \left(\sum_{j=\ell+1}^{m} T_{d,j,\ell}^{-1} B_j + \right.$$

$$\left. T_{d,m,\ell}^{-1} H(\mathsf{U}^{(m)}|\mathsf{M}_{[\ell+1:m]}) + \left(\frac{1}{d-\ell} - T_{d,m,\ell}^{-1} \right) H(\mathsf{U}^{(\ell)}) \right)$$

$$= H(\mathsf{U}^{(1,m)}|\mathsf{M}_{[\ell+1:m]}) + (d-m) T_{d,m,\ell}^{-1} H(\mathsf{U}^{(m)}|\mathsf{M}_{[\ell+1:m]}) +$$

$$(d-m) \left(\sum_{j=\ell+1}^{m} T_{d,j,\ell}^{-1} B_j + \left(\frac{1}{d-\ell} - T_{d,m,\ell}^{-1} \right) H(\mathsf{U}^{(\ell)}) \right)$$

$$\overset{(b)}{\geq} H(\mathsf{U}^{(1,m+1)}|\mathsf{M}_{[\ell+1:m]}) + (d-m) T_{d,m,\ell}^{-1} H(\mathsf{U}^{(\ell)}|\mathsf{M}_{[\ell+1:m]}) +$$

$$(d-m) \left(\sum_{j=\ell+1}^{m} T_{d,j,\ell}^{-1} B_j + \left(\frac{1}{d-\ell} - T_{d,m,\ell}^{-1} \right) H(\mathsf{U}^{(\ell)}) \right)$$

$$\overset{(c)}{=} H(\mathsf{U}^{(1,m+1)}|\mathsf{M}_{[\ell+1:m]}) + (d-m) T_{d,m,\ell}^{-1} H(\mathsf{U}^{(\ell)}) +$$

$$(d-m) \left(\sum_{j=\ell+1}^{m} T_{d,j,\ell}^{-1} B_j + \left(\frac{1}{d-\ell} - T_{d,m,\ell}^{-1} \right) H(\mathsf{U}^{(\ell)}) \right)$$

$$= H(\mathsf{U}^{(1,m+1)}|\mathsf{M}_{[\ell+1:m]}) + (d-m) \sum_{j=\ell+1}^{m} T_{d,j,\ell}^{-1} B_j +$$

$$\frac{d-m}{d-\ell} H(\mathsf{U}^{(\ell)}),$$

where (a) follows from (22) of Proposition 1; (b) follows from Corollary 3; and (c) follows from the fact that $I(\mathsf{U}^{(\ell)}; \mathsf{M}_{[\ell+1:m]}) = 0$ due to the repair secrecy requirement (4). Adding $H(\mathsf{M}_{[\ell+1:m]})$ to both sides and using the facts that

$$H(\mathsf{U}^{(1,m)}|\mathsf{M}_{[\ell+1:m]}) + H(\mathsf{M}_{[\ell+1:m]})$$
$$= H(\mathsf{U}^{(1,m)}, \mathsf{M}_{[\ell+1:m]}) \overset{(a)}{=} H(\mathsf{U}^{(1,m)})$$

and that

$$H(\mathsf{U}^{(1,m+1)}|\mathsf{M}_{[\ell+1:m]}) + H(\mathsf{M}_{[\ell+1:m]})$$
$$= H(\mathsf{U}^{(1,m+1)}, \mathsf{M}_{[\ell+1:m]}) \overset{(b)}{=} H(\mathsf{U}^{(1,m+1)})$$

complete the proof of (29). Here, (a) and (b) are due to the facts that $\mathsf{M}_{[\ell+1:m]}$ is a function of $\mathsf{W}_{[1:m]}$, which is a function of both $\mathsf{U}^{(1,m)}$ and $\mathsf{U}^{(1,m+1)}$ by Lemma 1.

To see (30), add (29) over $m \in [\ell+1 : d-1]$ and cancel $\sum_{m=\ell+2}^{d-1} H(\mathsf{U}^{(1,m)})$ from both sides of the inequality. We have

$$H(\mathsf{U}^{(1,\ell+1)}) + \frac{T_{d,d,\ell+1}}{d-\ell} H(\mathsf{U}^{(\ell+1)})$$
$$\geq \sum_{m=\ell+1}^{d-1} \left((d-m) \sum_{j=\ell+1}^{m} T_{d,j,\ell}^{-1} B_j \right) +$$
$$H(\mathsf{U}^{(1,d)}) + \frac{T_{d,d,\ell+1}}{d-\ell} H(\mathsf{U}^{(\ell)}). \tag{31}$$

Note that

$$\sum_{m=\ell+1}^{d-1} \left((d-m) \sum_{j=\ell+1}^{m} T_{d,j,\ell}^{-1} B_j \right)$$
$$= \sum_{j=\ell+1}^{d-1} T_{d,j,\ell}^{-1} B_j \left(\sum_{m=j}^{d-1} (d-m) \right) = \sum_{j=\ell+1}^{d-1} T_{d,j,\ell}^{-1} T_{d,d,j} B_j. \tag{32}$$

Furthermore,

$$H(\mathsf{U}^{(1,d)}) \overset{(a)}{=} H(\mathsf{U}^{(1,d)}, \mathsf{M}_{[\ell+1:d]})$$
$$\overset{(b)}{=} H(\mathsf{U}^{(1,d)}|\mathsf{M}_{[\ell+1:d]}) + H(\mathsf{M}_{[\ell+1:d]})$$
$$\overset{(c)}{=} H(\mathsf{U}^{(1,d)}|\mathsf{M}_{[\ell+1:d]}) + \sum_{j=\ell+1}^{d} B_j$$
$$\overset{(d)}{=} H(\mathsf{U}^{(1,d)}, \mathsf{S}_{1\to[2:d-1]}|\mathsf{M}_{[\ell+1:d]}) + \sum_{j=\ell+1}^{d} B_j$$
$$\overset{(e)}{=} H(\mathsf{U}^{(d-1)}, \mathsf{W}_{d+1}|\mathsf{M}_{[\ell+1:d]}) + \sum_{j=\ell+1}^{d} B_j$$
$$\geq H(\mathsf{U}^{(\ell)}|\mathsf{M}_{[\ell+1:d]}) + \sum_{j=\ell+1}^{d} B_j$$
$$\overset{(f)}{=} H(\mathsf{U}^{(\ell)}) + \sum_{j=\ell+1}^{d} B_j, \tag{33}$$

where (a) follows from the fact that $M_{[\ell+1:d]}$ is a function of $W_{[1:d]}$, which is a function of $U^{(1,d)}$ by Lemma 1; (b) follows from the chain rule for entropy; (c) follows from the fact that $H(M_{[\ell+1:d]}) = \sum_{j=\ell+1}^{d} B_j$; ($d$) follows from the fact that $S_{1\to[2:d-1]}$ is a function of W_1 and hence a function of $U^{(1,d)}$; (e) follows from the fact that $H(U^{(1,d)}, S_{1\to[2:d-1]}|M_{[\ell+1:d]}) = H(U^{(d-1)}, W_{d+1}|M_{[\ell+1:d]})$ due to the symmetrical code that we consider; and (f) follows from the fact that $I(U^{(\ell)}; M_{[\ell+1:d]}) = 0$ due to the repair secrecy requirement (4).

Substituting (32) and (33) into (31) gives:

$$H(U^{(1,\ell+1)}) + \frac{T_{d,d,\ell+1}}{d-\ell}H(U^{(\ell+1)})$$

$$\geq \sum_{j=\ell+1}^{d-1} T_{d,j,\ell}^{-1}T_{d,d,j}B_j + \sum_{j=\ell+1}^{d} B_j + \left(1 + \frac{T_{d,d,\ell+1}}{d-\ell}\right)H(U^{(\ell)})$$

$$= \sum_{j=\ell+1}^{d-1} T_{d,j,\ell}^{-1}(T_{d,d,j} + T_{d,j,\ell})B_j + B_d + \frac{T_{d,d,\ell}}{d-\ell}H(U^{(\ell)})$$

$$\overset{(a)}{=} T_{d,d,\ell}\sum_{j=\ell+1}^{d-1} T_{d,j,\ell}^{-1}B_j + B_d + \frac{T_{d,d,\ell}}{d-\ell}H(U^{(\ell)})$$

$$= T_{d,d,\ell}\sum_{j=\ell+1}^{d} T_{d,j,\ell}^{-1}B_j + \frac{T_{d,d,\ell}}{d-\ell}H(U^{(\ell)}),$$

where (a) follows from the fact that $T_{d,d,j} + T_{d,j,\ell} = T_{d,d,\ell}$. This completes the proof of the proposition. $\square$

Proof of Theorem 1. We are now ready to prove the outer bounds (9) and (10). To prove (9), note that

$$\beta + \frac{1}{d-\ell}H(U^{\ell}) \overset{(a)}{\geq} \frac{1}{d-\ell}\left(H(\overline{S}_{\to\ell+1}) + H(U^{(\ell)})\right)$$

$$\overset{(b)}{\geq} \frac{1}{d-\ell}H(U^{(\ell+1)})$$

$$\overset{(c)}{\geq} \sum_{j=\ell+1}^{d} T_{d,j,\ell}^{-1}B_j + \frac{1}{d-\ell}H(U^{(\ell)}),$$

where (a) follows from the fact that $H(\overline{S}_{\to\ell+1}) \leq (d-\ell)\beta$; ($b$) follows from the union bound on entropy; and (c) follows from (23) of Proposition 1. Cancelling $\frac{1}{d-\ell}H(U^{\ell})$ from both sides of the inequality and normalizing both sides by $\sum_{t=\ell+1}^{d} B_t$ complete the proof of (9).

To prove (10), note that

$$\alpha + (d(d-\ell)-\ell)\beta + (d+1)H(\mathsf{U}^{(\ell)})$$

$$\overset{(a)}{\geq} H(\mathsf{W}_{d+1}) + H(\mathsf{U}^{(\ell)}) + (d(d-\ell)-\ell)\beta + dH(\mathsf{U}^{(\ell)})$$

$$\overset{(b)}{=} H(\mathsf{W}_{d+1}, \mathsf{S}_{d+1\to[1:\ell]}) + H(\mathsf{U}^{(\ell)}) +$$
$$\quad (d(d-\ell)-\ell)\beta + dH(\mathsf{U}^{(\ell)})$$

$$\overset{(c)}{\geq} H(\mathsf{W}_{d+1}, \mathsf{U}^{(\ell)}) + H(\mathsf{S}_{d+1\to[1:\ell]}) +$$
$$\quad (d(d-\ell)-\ell)\beta + dH(\mathsf{U}^{(\ell)})$$

$$\overset{(d)}{\geq} H(\mathsf{W}_{d+1}, \mathsf{U}^{(\ell)}) + dH(\mathsf{U}^{(\ell+1)})$$

$$\overset{(e)}{=} H(\mathsf{U}^{(1,\ell+1)}, \mathsf{S}_{1\to[2:\ell+1]}) + dH(\mathsf{U}^{(\ell+1)})$$

$$\geq H(\mathsf{U}^{(1,\ell+1)}) + dH(\mathsf{U}^{(\ell+1)})$$

$$= H(\mathsf{U}^{(1,\ell+1)}) + \frac{T_{d,d,\ell+1}}{d-\ell}H(\mathsf{U}^{(\ell+1)}) +$$
$$\quad \left(d - \frac{T_{d,d,\ell+1}}{d-\ell}\right)H(\mathsf{U}^{(\ell+1)})$$

$$\overset{(f)}{\geq} T_{d,d,\ell}\left(\sum_{j=\ell+1}^{d} T_{d,j,\ell}^{-1}B_j + \frac{H(\mathsf{U}^{(\ell)})}{d-\ell}\right) +$$
$$\quad \left(d - \frac{T_{d,d,\ell+1}}{d-\ell}\right)\left((d-\ell)\sum_{j=\ell+1}^{d} T_{d,j,\ell}^{-1}B_j + H(\mathsf{U}^{(\ell)})\right)$$

$$= (T_{d,d,\ell} + d(d-\ell) - T_{d,d,\ell+1})\sum_{j=\ell+1}^{d} T_{d,j,\ell}^{-1}B_j +$$
$$\quad \left(\frac{T_{d,d,\ell}}{d-\ell} + d - \frac{T_{d,d,\ell+1}}{d-\ell}\right)H(\mathsf{U}^{(\ell)})$$

$$\overset{(g)}{=} (d+1)(d-\ell)\sum_{j=\ell+1}^{d} T_{d,j,\ell}^{-1}B_j + (d+1)H(\mathsf{U}^{(\ell)}),$$

where (a) follows from the fact that $H(\mathsf{W}_{d+1}) \leq \alpha$; (b) follows from the fact that $\mathsf{S}_{d+1\to[1:\ell]}$ is a function of W_{d+1}; (c) follows from the fact that $H(\mathsf{W}_{d+1}, \mathsf{S}_{d+1\to[1:\ell]}) + H(\mathsf{U}^{(\ell)}) \geq H(\mathsf{W}_{d+1}, \mathsf{U}^{(\ell)}) + H(\mathsf{S}_{d+1\to[1:\ell]})$ due to the submodularity of the entropy function; (d) follows from Proposition 2; (e) follows from the fact that $H(\mathsf{W}_{d+1}, \mathsf{U}^{(\ell)}) = H(\mathsf{U}^{(1,\ell+1)}, \mathsf{S}_{1\to[2:\ell+1]})$ due to the symmetrical code that we consider; (f) follows from (23) of Proposition 1 and (30) of Proposition 3; and (g) follows from the fact that $T_{d,d,\ell} - T_{d,d,\ell+1} = d - \ell$. Cancelling $(d+1)H(\mathsf{U}^{\ell})$ from both sides of the inequality and normalizing both sides by $\sum_{t=\ell+1}^{d} B_t$ complete the proof of (10). $\quad\square$

5. Conclusions

This paper considered the problem of MDC-SR, which includes the problems of MDC-R and SRC as special cases. Two outer bounds were established, showing that separate coding can achieve the MBR point of the achievable normalized storage-capacity repair-bandwidth trade-off regions for the general MDC-SR problem. When specialized to the SRC problem, it was shown that the SRK point [25] is the MBR point of the achievable normalized storage-capacity repair-bandwidth trade-off regions for the general SRC problem. The core of the new converse results is an exchange lemma, which we established by using Han's subset inequality [32]. The exchange lemma only relies on the functional

dependencies for the repair processes and might be useful for solving some other related problems as well.

Note that separate encoding can also achieve the MSR point of the achievable normalized storage-capacity repair-bandwidth trade-off regions for the general MDC-R problem [22]. We suspect that this also generalizes to the MDC-SR problem. To prove such this result, however, we shall need new converse results as well as new code constructions for the general SRC problem, both of which are currently under our investigations.

Author Contributions: T.L. and C.T. proposed the idea of this paper, S.S. was responsible for the technical proof of the paper and all for authors worked on writing, revising and editing this manuscript.

Appendix. Proof of the Exchange Lemma

Proof of the Exchange Lemma. This lemma is proved in an iterative way. The to be "exchanged" random variable sets are partitioned in a designed way, and every time after a small partition of the set is exchanged, we can establish an inequality. In our proof, we use not only the submodularity of the entropy function but also the properties of regeneration code, namely Lemma 1 as well.

Fix $m \in [1 : d - 1]$, $i \in [0 : m - 1]$, $i' \in [0 : i]$, and $j \in [i' + 1 : m - i + i' + 1]$. Let us first note that, if $j = m + 1$, we must have $i' = i$, and in this case the inequality (15) holds trivially with an equality. Therefore, for the remaining proof, we shall assume that $j \leq m$. Now that $d + 1 - j > d - m$, we may write $d + 1 - j = s(d - m) + r$ for some integer $s \geq 1$ and $r \in [1 : d - m]$. Furthermore, let

$$
a_t := \begin{cases}
t + i', & t \in [1 : i - i'], \\
t + j - 1, & t \in [i - i' + 1 : m - j + 1], \\
t + j, & t \in [m - j + 2 : d + 1 - j].
\end{cases}
$$

As illustrated in Figure A1, a_t is monotonically increasing with t. Finally, let $\tau_0 := \{a_t : t \in [1 : r]\}$ and

$$
\tau_q := \{a_t : t \in [r + 1 + (q - 1)(d - m) : r + q(d - m)]\}
$$

for any $q \in [1 : s]$. It is straightforward to verify that:

- $\tau_q \cap \tau_{q'} = \varnothing$ for any $q \neq q'$,
- $\bigcup_{q=0}^{s-1} \tau_q = [i' + 1 : i] \cup [i + j - i' : m]$,
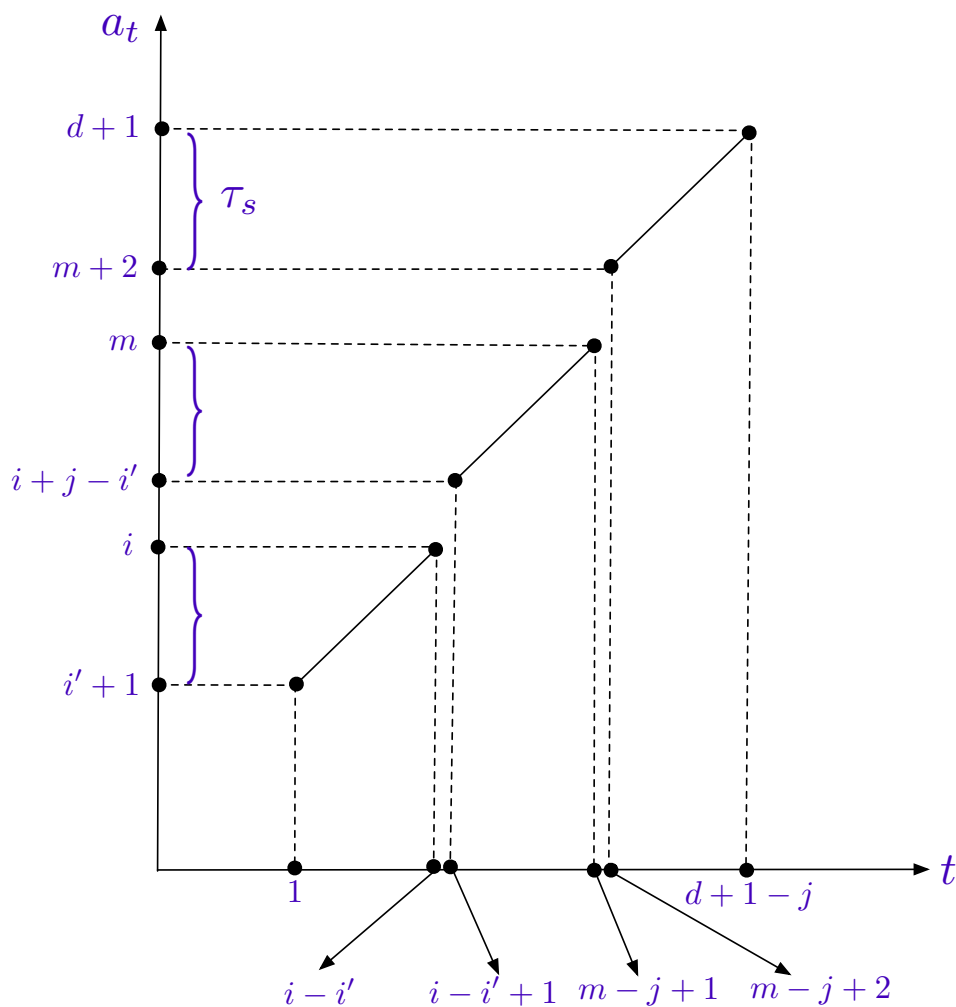- $\tau_s = [m + 2 : d + 1]$.

Figure A1. a_t as a function of t. The sets $(\tau_q : q \in [0 : s])$ form a partition of the set $[i' + 1 : i] \cup [i + j - i' : m] \cup [m + 2 : d + 1]$.

Consider a symmetrical $(n = d + 1, d, N_1, \ldots, N_d, T, S)$ code that satisfies the node regeneration requirement (3). Let us show by induction that for any $p \in [1 : s]$, we have

$$
\begin{aligned}
& pH(\mathsf{U}^{(i,m)}|\mathsf{M}^{(m)}) + H(\mathsf{U}^{(i',j)}|\mathsf{M}^{(m)}) \\
& \geq pH(\mathsf{U}^{(i,m+1)}|\mathsf{M}^{(m)}) + \\
& \quad H(\mathsf{W}_{[1:i']}, \mathsf{S}_{\rightarrow[i+1:i+j-i'-1]}, \mathsf{S}_{\cup_{q=0}^{s-p}\tau_q \rightarrow m+1}|\mathsf{M}^{(m)}).
\end{aligned}
\tag{A1}
$$

To prove the base case of $p = 1$, first note that

$$
\begin{aligned}
& H(\mathsf{U}^{(i,m)}|\mathsf{M}^{(m)}) \\
& \overset{(a)}{=} H(\mathsf{U}^{(i,m)}, \mathsf{W}_{[i+1,m]}, \underline{\mathsf{S}}_{\rightarrow[i+1:m]}|\mathsf{M}^{(m)}) \\
& = H(\mathsf{W}_{[1:m]}, \mathsf{S}_{\rightarrow[i+1:m]}|\mathsf{M}^{(m)}) \\
& \overset{(b)}{=} H(\mathsf{W}_{[1:m]}, \mathsf{S}_{\rightarrow[i+1:m]}, \mathsf{S}_{[1:m]\rightarrow m+1}|\mathsf{M}^{(m)})) \\
& \geq H(\mathsf{W}_{[1:i]}, \mathsf{S}_{\rightarrow[i+1:m]}, \mathsf{S}_{[1:m]\rightarrow m+1}|\mathsf{M}^{(m)})),
\end{aligned}
$$

where (a) follows from the fact that $(W_{[i+1,m]}, \underline{S}_{\to[i+1:m]})$ is a function of $U^{(i,m)}$ by Lemma 1, and (b) follows from the fact that $S_{[1:m]\to m+1}$ is a function of $W_{[1:m]}$. Furthermore,

$$
\begin{aligned}
&H(U^{(i',j)}|M^{(m)}) \\
&\overset{(a)}{=} H(U^{(i',j)}, \underline{S}_{\to[i'+1:j]}|M^{(m)}) \\
&= H(W_{[1:i']}, S_{\to[i'+1:j]}|M^{(m)}) \\
&\overset{(b)}{=} H(W_{[1:i']}, S_{\to[i+1:i+j-i']}|M^{(m)}) \\
&= H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, S_{\to i+j-i'}|M^{(m)}) \\
&\geq H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, S_{[i'+1:i]\to i+j-i'}, \\
&\qquad S_{[i+j-i'+1:d+1]\to i+j-i'}|M^{(m)}) \\
&\overset{(c)}{=} H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, S_{[i'+1:i]\to m+1}, \\
&\qquad S_{[i+j-i':m]\to m+1}, S_{[m+2:d+1]\to m+1}|M^{(m)}),
\end{aligned}
$$

where (a) follows from the fact that $\underline{S}_{\to[i'+1:j]}$ is a function of $U^{(i',j)}$ by Lemma 1, and (b) and (c) follow from the symmetrical code that we consider. It follows that

$$
\begin{aligned}
&H(U^{(i,m)}|M^{(m)}) + H(U^{i',j}|M^{(m)}) \\
&\geq H(W_{[1:i]}, S_{\to[i+1:m]}, S_{[1:m]\to m+1}|M^{(m)})) + \\
&\qquad H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, S_{[i'+1:i]\to m+1}, \\
&\qquad S_{[i+j-i':m]\to m+1}, S_{[m+2:d+1]\to m+1}|M^{(m)}) \\
&\overset{(a)}{\geq} H(W_{[1:i]}, S_{\to[i+1:m]}, S_{[1:m]\to m+1}, \\
&\qquad S_{[m+2:d+1]\to m+1}|M^{(m)}) + H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, \\
&\qquad S_{[i'+1:i]\to m+1}, S_{[i+j-i':m]\to m+1}|M^{(m)}) \\
&= H(U^{(i,m+1)}, \underline{S}_{\to[i+1:m+1]}|M^{(m)}) + \\
&\qquad H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, S_{\bigcup_{q=0}^{s-1}\tau_q\to m+1}|M^{(m)}) \\
&\geq H(U^{(i,m+1)}|M^{(m)}) + \\
&\qquad H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, S_{\bigcup_{q=0}^{s-1}\tau_q\to m+1}|M^{(m)}),
\end{aligned}
$$

where (a) follows from the submodularity of the entropy function. This completes the proof of the base case of $p = 1$.

Assume that (A1) holds for some $p \in [1:s-1]$. We have

$$
\begin{aligned}
&(p+1)H(U^{(i,m)}|M^{(m)}) + H(U^{(i',j)}|M^{(m)}) \\
&= H(U^{(i,m)}|M^{(m)}) + \left(pH(U^{(i,m)}|M^{(m)}) + H(U^{(i',j)}|M^{(m)})\right) \\
&\geq H(U^{(i,m)}|M^{(m)}) + pH(U^{(i,m+1)}|M^{(m)}) + \\
&\qquad H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, S_{\bigcup_{q=0}^{s-p}\tau_q\to m+1}|M^{(m)}). \tag{A2}
\end{aligned}
$$

Note that both $\underline{S}_{\to[i+1,i+j-i'-1]}$ and $S_{\bigcup_{q=0}^{s-(p+1)} \tau_q \to m+1}$ are functions of $W_{[1:m]}$, which is in turn a function of $U^{(i,m)}$ by Lemma 1. We thus have

$$H(U^{(i,m)}|M^{(m)})$$
$$= H(U^{(i,m)}, \underline{S}_{\to[i+1,i+j-i'-1]}, S_{\bigcup_{q=0}^{s-(p+1)} \tau_q \to m+1}|M^{(m)}).$$

Furthermore, by the symmetrical code that we consider, we have

$$H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, S_{\bigcup_{q=0}^{s-p} \tau_q \to m+1}|M^{(m)})$$
$$= H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]},$$
$$S_{\bigcup_{q=0}^{s-(p+1)} \tau_q \to m+1}, S_{[m+2:d+1] \to m+1}|M^{(m)}).$$

It follows that

$$H(U^{(i,m)}|M^{(m)})+$$
$$H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, S_{\bigcup_{q=0}^{s-p} \tau_q \to m+1}|M^{(m)})$$
$$= H(U^{(i,m)}, \underline{S}_{\to[i+1,i+j-i'-1]}, S_{\bigcup_{q=0}^{s-(p+1)} \tau_q \to m+1}|M^{(m)})+$$
$$H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]},$$
$$S_{\bigcup_{q=0}^{s-(p+1)} \tau_q \to m+1}, S_{[m+2:d+1] \to m+1}|M^{(m)})$$
$$\overset{(a)}{\geq} H(U^{(i,m)}, \underline{S}_{\to[i+1,i+j-i'-1]},$$
$$S_{\bigcup_{q=0}^{s-(p+1)} \tau_q \to m+1}, S_{[m+2:d+1] \to m+1}|M^{(m)})+$$
$$H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, S_{\bigcup_{q=0}^{s-(p+1)} \tau_q \to m+1}|M^{(m)})$$
$$\geq H(U^{(i,m+1)}|M^{(m)})+$$
$$H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, S_{\bigcup_{q=0}^{s-(p+1)} \tau_q \to m+1}|M^{(m)}), \tag{A3}$$

where (a) follows from the submodularity of the entropy function. Substituting (A3) into (A2) gives

$$(p+1)H(U^{(i,m)}|M^{(m)}) + H(U^{(i',j)}|M^{(m)})$$
$$\geq (p+1)H(U^{(i,m+1)}|M^{(m)})+$$
$$H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, S_{\bigcup_{q=0}^{s-(p+1)} \tau_q \to m+1}|M^{(m)}),$$

which completes the induction step and hence the proof of (A1).

Setting $p = s$ in (A1), we have

$$sH(U^{(i,m)}|M^{(m)}) + H(U^{(i',j)}|M^{(m)})$$
$$\geq sH(U^{(i,m+1)}|M^{(m)})+$$
$$H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, S_{\tau_0 \to m+1}|M^{(m)})$$
$$= sH(U^{(i,m+1)}|M^{(m)}) + H(W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}|M^{(m)})+$$
$$H(S_{\tau_0 \to m+1}|W_{[1:i']}, S_{\to[i+1:i+j-i'-1]}, M^{(m)}). \tag{A4}$$

By the symmetrical codes that we consider, we have

$$
\begin{aligned}
& H(\mathsf{W}_{[1:i']}, \mathsf{S}_{\to[i+1:i+j-i'-1]} | \mathsf{M}^{(m)}) \\
&= H(\mathsf{W}_{[1:i']}, \mathsf{S}_{\to[i'+1:j-1]} | \mathsf{M}^{(m)}) \\
&= H(\mathsf{U}^{(i',j-1)}, \underline{\mathsf{S}}_{\to[i'+1:j-1]} | \mathsf{M}^{(m)}) \\
&\geq H(\mathsf{U}^{(i',j-1)} | \mathsf{M}^{(m)})
\end{aligned}
\tag{A5}
$$

and

$$
\begin{aligned}
& H(\mathsf{S}_{\tau_0 \to m+1} | \mathsf{W}_{[1:i']}, \mathsf{S}_{\to[i+1:i+j-i'-1]}, \mathsf{M}^{(m)}) \\
&= H(\mathsf{S}_{\tau \to m+1} | \mathsf{W}_{[1:i']}, \mathsf{S}_{\to[i+1:i+j-i'-1]}, \mathsf{M}^{(m)}
\end{aligned}
$$

for *any* subset $\tau \subseteq [m+2:d+1]$ such that $|\tau| = r$. By Han's subset inequality [32], we have

$$
\begin{aligned}
& H(\mathsf{S}_{\tau_0 \to m+1} | \mathsf{W}_{[1:i']}, \mathsf{S}_{\to[i+1:i+j-i'-1]}, \mathsf{M}^{(m)}) \\
&\geq \frac{r}{d-m} H(\mathsf{S}_{[m+2:d+1] \to m+1} | \mathsf{W}_{[1:i']}, \\
& \qquad\qquad \mathsf{S}_{\to[i+1:i+j-i'-1]}, \mathsf{M}^{(m)}) \\
&\geq \frac{r}{d-m} H(\mathsf{S}_{[m+2:d+1] \to m+1} | \mathsf{W}_{[1:i']}, \\
& \qquad\qquad \mathsf{S}_{\to[i+1:i+j-i'-1]}, \mathsf{U}^{(i,m)}, \mathsf{M}^{(m)}) \\
&\overset{(a)}{=} \frac{r}{d-m} H(\mathsf{S}_{[m+2:d+1] \to m+1} | \mathsf{U}^{(i,m)}, \mathsf{M}^{(m)}) \\
&= \frac{r}{d-m} \Big(H(\mathsf{S}_{[m+2:d+1] \to m+1}, \mathsf{U}^{(i,m)} | \mathsf{M}^{(m)}) - \\
& \qquad\qquad H(\mathsf{U}^{(i,m)} | \mathsf{M}^{(m)}) \Big) \\
&= \frac{r}{d-m} \Big(H(\mathsf{U}^{(i,m+1)} | \mathsf{M}^{(m)}) - H(\mathsf{U}^{(i,m)} | \mathsf{M}^{(m)}) \Big),
\end{aligned}
\tag{A6}
$$

where (a) follows from the fact that $(\mathsf{W}_{[1:i']}, \underline{\mathsf{S}}_{\to[i+1:i+j-i'-1]})$ is a function of $\mathsf{U}^{(i,m)}$ by Lemma 1. Substituting (A5) and (A6) into (A4) gives:

$$
\begin{aligned}
& \left(s + \frac{r}{d-m}\right) H(\mathsf{U}^{(i,m)} | \mathsf{M}^{(m)}) + H(\mathsf{U}^{(i',j)} | \mathsf{M}^{(m)}) \\
&\geq \left(s + \frac{r}{d-m}\right) H(\mathsf{U}^{(i,m+1)} | \mathsf{M}^{(m)}) + H(\mathsf{U}^{(i',j-1)} | \mathsf{M}^{(m)}),
\end{aligned}
$$

which is equivalent to (15) by noting that

$$
s + \frac{r}{d-m} = \frac{s(d-m)+r}{d-m} = \frac{d+1-j}{d-m}.
$$

This completes the proof of the exchange lemma. □

References

1. Singleton, R.C. Maximum distance *q*-nary codes. *IEEE Trans. Inf. Theory* **1964**, *10*, 116–118. [CrossRef]
2. Roche, J.R. Distributed Information Storage. Ph.D. Dissertation, Stanford University, Stanford, CA, USA, 1992.
3. Roche, J.R.; Yeung, R.W.; Hau, K.P. Symmetrical multilevel diversity coding. *IEEE Trans. Inf. Theory* **1997**, *43*, 1059–1064. [CrossRef]
4. Yeung, R.W.; Zhang, Z. On symmetrical multilevel diversity coding. *IEEE Trans. Inf. Theory* **1999**, *45*, 609–621. [CrossRef]

5. Mohajer, S.; Tian, C.; Diggavi, S.N. Asymmetric multilevel diversity coding and asymmetric Gaussian multiple descriptions. *IEEE Trans. Inf. Theory* **2010**, *56*, 4367–4387. [CrossRef]

6. Jiang, J.; Marukala, N.; Liu, T. Symmetrical multilevel diversity coding and subset entropy inequalities. *IEEE Trans. Inf. Theory* **2014**, *60*, 84–103. [CrossRef]

7. Dimakis, A.G.; Godfrey, P.B.; Wu, Y.; Wainwright, M.; Ramchandran, K. Network coding for distributed storage systems. *IEEE Trans. Inf. Theory* **2010**, *56*, 4539–4551. [CrossRef]

8. Rashmi, K.V.; Shah, N.B.; Kumar, P.V. Optimal exact-regenerating codes for distributed storage at the MSR and MBR points via a product-matrix construction. *IEEE Trans. Inf. Theory* **2011**, *57*, 5227–5239. [CrossRef]

9. Cadambe, V.R.; Jafar, S.A.; Maleki, H.; Ramchandran, K.; Suh, C. Asymptotic interference alignment for optimal repair of MDS codes in distributed storage. *IEEE Trans. Inf. Theory* **2013**, *59*, 2974–2987. [CrossRef]

10. Tian, C. Characterizing the rate region of the $(4, 3, 3)$ exact-repair regenerating codes. *IEEE J. Sel. Areas Commun.* **2014**, *32*, 967–975. [CrossRef]

11. Goparaju, S.; El Rouayheb, S.; Calderbank, R. New codes and inner bounds for exact repair in distributed storage systems. In Proceedings of the 2014 IEEE International Symposium on Information Theory (ISIT), Honolulu, HI, USA, 29 June–4 July 2014; pp. 1036–1040.

12. Duursma, I.M. Outer bounds for exact repair codes. *Arxiv* **2014**, arxiv:1406.4852. [CrossRef]

13. Prakash, N.; Krishnan, M.N. The storage-repair-bandwidth trade-off of exact repair linear regenerating codes for the case $d = k = n - 1$. In Proceedings of the 2015 IEEE International Symposium on Information Theory (ISIT), Hong Kong, China, 14–19 June 2015; pp. 859–863.

14. Elyasi, M.; Mohajer, S.; Tandon, R. Linear exact repair rate region of $(k + 1, k, k)$ distributed storage systems: A new approach. In Proceedings of the 2015 IEEE International Symposium on Information Theory (ISIT), Hong Kong, China, 14–19 June 2015; pp. 2061–2065.

15. Tian, C.; Sasidharan, B.; Aggarwal, V.; Vaishampayan, V.A.; Kumar, P.V. Layered exact-repair regenerating codes via embedded error correction and block designs. *IEEE Trans. Inf. Theory* **2015**, *61*, 1933–1947. [CrossRef]

16. Ye, M.; Barg, A. Explicit constructions of high-rate MDS array codes with optimal repair bandwidth. *IEEE Trans. Inf. Theory* **2017**, *63*, 2001–2014. [CrossRef]

17. Kralevska, K.; Gligoroski, D. An Explicit Construction of Systematic MDS Codes with Small Sub-packetization for All-Node Repair. *Arxiv* **2018**, arXiv:1806.03103.

18. Goparaju, S.; Fazeli, A.; Vardy, A. Minimum storage regenerating codes for all parameters. *IEEE Trans. Inf. Theory* **2017**, *63*, 6318–6328. [CrossRef]

19. Kralevska, K.; Gligoroski, D.; Jensen, R.E.; Overby, H. Hashtag erasure codes: From theory to practice. *IEEE Trans. Big Data* **2017**, *1*. [CrossRef]

20. Kralevska, K.; Gligoroski, D.; Øverby, H. General Sub-packetized Access Optimal Regenerating Codes. *IEEE Commun. Lett.* **2016**, *20*, 1281–1284. [CrossRef]

21. Tian, C.; Liu, T. Multilevel diversity coding with regeneration. *IEEE Trans. Inf. Theory* **2016**, *62*, 4833–4847. [CrossRef]

22. Shao, S.; Liu, T.; Tian, C. Multilevel diversity coding with regeneration: Separate coding achieves the MBR point. In Proceedings of the 2016 Annual Conference on Information Science and Systems (CISS), Princeton, NJ, USA, 16–18 March 2016; pp. 602–607.

23. Pawar, S.; El Rouayheb, S.; Ramchandran, K. On secure distributed data storage under repair dynamics. In Proceedings of the 2010 IEEE International Symposium on Information Theory (ISIT), Austin, TX, USA, 13–18 June 2010; pp. 2543–2547.

24. Pawar, S.; El Rouayheb, S.; Ramchandran, K. Securing dynamic distributed storage systems against eavesdropping and adversarial Attacks. *IEEE Trans. Inf. Theory* **2011**, *57*, 6734–6753. [CrossRef]

25. Shah, N.B.; Rashmi, K.V.; Kumar, P.V. Information-theoretically secure regenerating codes for distributed storage. In Proceedings of the 2011 IEEE Global Telecommunications Conference (GLOBECOM), Kathmandu, Nepal, 5–9 December 2011; pp. 1–5.

26. Goparaju, S.; El Rouayheb, S.; Calderbank, R.; Poor, H.V. Data secrecy in distributed storage systems under exact repair. In Proceedings of the 2013 International Symposium on Network Coding (NetCod), Calgary, AB, Canada, 7–9 June 2013; pp. 1–6.

27. Rawat, A.S.; Koyluoglu, O.O.; Silberstein, N.; Vishwanath, S. Optimal locally repairable and secure codes for distributed storage systems. *IEEE Trans. Inf. Theory* **2014**, *60*, 212–236. [CrossRef]

28. Tandon, R.; Amuru, S.; Clancy, T.C.; Buehrer, R.M. Towards optimal secure distributed storage systems with exact repair. *IEEE Trans. Inf. Theory* **2016**, *62*, 3477–3492. [CrossRef]
29. Ye, F.; Shum, K.W.; Yeung, R.W. The rate region of secure exact-repair regenerating codes for 5 nodes. In Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016; pp. 1406–1410.
30. Shao, S.; Liu, T.; Tian, C.; Shen, C. On the trade-off region of secure exact-repair regenerating codes. *IEEE Trans. Inf. Theory* **2017**, *63*, 7253–7266. [CrossRef]
31. Balasubramanian, A.; Ly, H.D.; Li, S.; Liu, T.; Miller, S.L. Secure symmetrical multilevel diversity coding. *IEEE Trans. Inf. Theory* **2013**, *59*, 3572–3581. [CrossRef]
32. Han, T.S. Nonnegative entropy measures of multivariate symmetric correlations. *Inf. Control* **1978**, *36*, 133–156. [CrossRef]