Optimal Single-Bit Relaying Strategies with Multi-Relay Diversity

Matthew Bliss*, Chih-Chun Wang*, and David J. Love*

***Abstract*—Many emerging applications require multi-hop wireless relaying, for which reliability requirements increase packet retransmissions, amplifying latency across hops. Existing works mostly focus on the simple two-hop, single-relay setting and ignore spatial diversity that enables a destination to receive independent noisy copies of data from multiple relays in parallel to improve error performance. In this paper, we consider a single-bit source message and construct learning-rate-optimal, delay-constrained multi-hop schemes by jointly designing time-varying, distributed relay mapping functions with destination decoding strategies. The learning-rate-optimal scheme, however, requires channel and relaying knowledge, which limits practical implementation. With the aim of practical implementation, we have also considered several low-complexity relay and destination strategies and analyzed their performances under different combinations. Numerical comparisons show that none of the alternatives universally dominate, and the system designer thus has to carefully opt for the best suitable schemes depending on the channel conditions. Finally, we show that carefully coordinating many low-quality relay channels in parallel can vastly outperform having only one high-quality relay channel, which demonstrates the spatial diversity gains for the first time in the learning over parallel-relay setting.**

***Index Terms*—Multi-hop, relay channels, low latency, high reliability, transcoding, teaching and learning**

I. Introduction

The relay channel [1] is a long-standing information-theoretic topic growing (again) in recent interest due to its vast applicability to emerging wireless systems such as machine-to-machine relaying in Internet of Things (IoT) networks [2], integrated terrestrial-air-space communications [3]–[5], and multi-hop wireless backhauling for rural networks and small cells [6], [7]. While the general relay channel capacity is unknown, many real systems can be modeled as a specialized case known as the *separated relay channel*, which is the concatenation of two point-to-point (single-hop) channels where the source has no direct connection to the destination. The capacity of this specialized relay channel is the bottleneck capacity of the two concatenated point-to-point channels, i.e., $C = \min\{C_1, C_2\}$, and is achieved with the decode-and-forward (DF) protocol in the absence of delay constraints [1].

However, recent interest in low-latency and high-reliability communication requires a better understanding of the finite blocklength regime's throughput-delay and delay-reliability tradeoffs [8]–[10]. This regime differs from the asymptotic Shannon capacity and is relevant to many applications, i.e., healthcare, robotics, autonomous vehicles, and augmented reality, that need latencies of less than a millisecond and packet error rates below 10^{-5} [11]. For such applications, ultra-reliable and low-latency communication (URLLC) is essential, as specified by the Third Generation Partnership Project [11]. Low-latency solutions for single-hop channels include shortened transmission time intervals and frame structures, hybrid automatic repeat request (HARQ) protocols, edge caching, computing, and slicing [8]. High-reliability techniques over single-hop channels include adapting coding rates and exploiting time, frequency, and spatial diversity [8].

Despite existing solutions for single-hop channels, little is known about designing low-latency and high-reliability multi-hop networks [9], [10]. With strict reliability requirements, packets are more likely to require multiple retransmissions across the hops to guarantee reliable delivery at the destination, amplifying end-to-end latency. Conversely, strict delay requirements entail that relays in the multi-hop network can not afford the time to queue packets needed for retransmissions, decreasing the end-to-end reliability [12].

Furthermore, unlike single-hop communication, multi-hop networks have an additional design component: the *relaying protocol*. Relays typically use either amplify-and-forward (AF) or decode-and-forward (DF), which skew performance to the extremes of feasible delay-reliability regions [9]. AF relays transmit amplified versions of their received signals without processing, resulting in minimal end-to-end latency, but often unreliable transmission, due to error accumulation in intermediate nodes. DF relays wait to receive the entire block before decoding, re-encoding, and transmitting to the next hop. The DF protocol has maximal end-to-end latency but often improves reliability because relays perform error correction [1].

A. Related Work

In [9], the *transcoding* principle was introduced for the separated relay channel as an alternative to the latency extremes of AF and DF. Transcoding relays perform (potentially) time-varying mapping functions from the partial sequence of symbols received up to time k to the next-hop input at time k. For example, [9] introduced a concatenated coding-inspired method, in which the block transmitted by the source is processed and transmitted as a sequence of independent sub-blocks at the relays. Each sub-block uses inner codes, and error patterns outside of a selected decoding radius propagate between hops to the destination, where the destination handles errors by decoding with the entire block.

This work is in part supported by the Office of Naval Research under ONR Grant N00014-21-1-2472 and the National Science Foundation under NSF grants CCF1816013, EEC1941529, CNS2212565, and CNS2225577.

*School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN.

Transcoding for the separated relay channel demonstrates untapped potential to outperform the AF and DF relaying schemes when restricted by latency constraints [9]. However, the proposed framework neither directly maximizes throughput nor minimizes the probability of decoding error at the destination. Instead, it uses the error exponent and the finite-length approximation formulae as proxies of the traditional communication performance metrics. In this paper, we are explicitly interested in 1) analytically determining error performance bounds on transcoding in the delay-constrained setting, and 2) constructing joint time-varying relaying and destination decoding strategies to achieve the optimal error performance.

Recent research has explored the problem of teaching and learning in uncertainty, with works [13], [14] studying a single separated relay with independent binary symmetric channels (BSCs). Additionally, [15] extended this framework to include general binary-input discrete memoryless channels. The goal of [13]–[15] was to maximize the asymptotic learning rate by designing joint time-varying relaying functions and destination decoding strategies. However, multi-bit transmission remained an open problem until [16] addressed the transmission of an arbitrary number of source messages over a single separated relay with discrete memoryless channels (DMCs). In these studies, the teacher (relay) receives n (blocklength) repeated noisy observations of an unknown message, θ, from a source and conveys this information to the student (destination). The teaching and learning framework seeks to maximize the asymptotic learning rate with explicitly designed time-varying relaying functions and destination decoding strategies. Denoting $\hat{\theta}$ as the message estimate at the destination, the asymptotic learning rate is defined as

$$\mathcal{L}\left(\mathtt{P},\mathtt{Q}\right)=\lim_{n\to\infty}\left\{-\frac{1}{n}\log\mathbb{P}\left(\hat{\theta}\neq\theta\right)\right\},$$

where $\mathtt{P}$ and $\mathtt{Q}$ are the transition laws of the source-to-relay and relay-to-destination links, respectively. For a single-bit source message over BSCs (with source-to-relay and relay-to-destination crossover probabilities $p,q\in[0,1/2)$, respectively), the learning rate is denoted as $\mathcal{L}(p,q)$ and a converse bound on the learning rate is proposed as

$$\mathcal{L}(p,q)\leq D_{\mathrm{kl}}(0.5||\max p,q),$$

where $D_{\mathrm{kl}}(a||b)=a\log\frac{a}{b}+(1-a)\log\frac{1-a}{1-b}$ is the Kullback-Leibler divergence of independent Bernoulli random variables parameterized by a and b.

In [13], [14] (see Sec. III), the proposed strategies fall short of the converse bound. However, [15] introduces a relaying protocol over BSCs (see Sec. III) that processes and transmits independent sub-blocks of $\delta_T=o(n)$ bits (sub-linear with respect to the blocklength). The relay transmits each received sub-block as a sorted sequence of k_0 zeros followed by k_1 ones (with $\delta_T=k_0+k_1$), where the number of zeros and ones depends on the fraction $\upsilon\in[0,1]$ of zeros in the sub-block. With maximum likelihood (ML) destination decoding, the independent sub-block-structured strategy of [15] asymptotically achieves the optimal learning rate over BSCs.

Wireless systems exploit spatial diversity to enable high reliability [8], [17]. For example, multi-antenna systems transmit signals through parallel paths, so that the receiver obtains independently-faded replicas of data [18]. Distributed array systems use virtual elements of a multi-antenna architecture, where adding or dropping nodes varies system performance and energy efficiency [19]. In wireless sensor and mesh networks, spatial diversity is exploited by opportunistically selecting multi-hop routing paths [20]–[24]. Our work studies the effect of spatial diversity using the teaching and learning or transcoding framework, developing relaying strategies over time and space. This approach resembles the distributed detection of signals over wireless channels or distributed estimation in wireless sensor networks [25], [26]. In this framework, each receiving antenna or node makes a multi-bit decision with its observations, and a fusion center combines the decisions to produce a final decoding decision [25].

B. Contributions

This paper exploits spatial diversity to achieve low-latency and high-reliability relaying over parallel separated relay channels. We extend the teaching and learning framework [13]–[15] to $M\geq 2$ parallel relays receiving n repeated observations of an unknown single-bit state from a source, corrupted by independent BSC noise. The destination receives signals from all relays through independent BSCs to make a final decoding decision and learn the state θ. We derive analytical error performance-related bounds for the multi-relay setting and develop distributed coding schemes across the M relays jointly designed with destination decoding schemes. Next, we outline the contributions of this work.

Contribution 1: We derive an upper bound on the asymptotic learning rate of the final decoding decision at the destination.
Contribution 2: We show that an M parallel-relay generalization of the independent sub-block-structured strategy from [15], paired with ML destination decoding, achieves the optimal learning rate on BSCs.

The optimal strategy, however, has implementation shortcomings that we address. First, each relay needs to know its source-to-relay BSC channel parameter. Likewise, ML decoding at the destination requires knowledge of channel parameters and relaying functions. Moreover, for a small tolerable end-to-end delay (often the case in practice), the error probability of the optimal strategy may suffer, despite the asymptotic learning rate of the scheme being provably optimal.
Contribution 3: To address implementation shortcomings of the optimal scheme, we consider alternative relaying protocols, generalized from the single relay analyses in [13], [14]. Specifically, the relays perform either one of

- *Simple forwarding:* Each relay forwards its newest received bit to the destination in the next channel use;
- *Sequential best guessing:* In each channel use, the relay takes a majority vote of all bits received up to that time and transmits the majority decision to the destination in the next channel use.

The above relaying alternatives do not require knowledge of the source-to-relay channel parameters and may exhibit

improved error performance when operating with a small tolerable end-to-end delay.

Additionally, we design the following alternative destination decoding strategies to pair with the *simple forwarding* and *sequential best guessing* relaying schemes.

- *Full matrix fusion:* The destination takes a majority vote over all Mn received bits from all relays to determine the final decoding decision;
- *Weighted fusion:* The destination quantizes each received relay signal via sub-decoders into a single-bit decision. The M sub-decoder outputs are combined as a weighted, signed summation and compared to a threshold to produce the final decoding decision;
- *Optimally-weighted fusion:* The destination uses knowledge of the channels and relaying functions to determine the optimal weights that provide the maximal learning rate of all weighted fusion strategies.

We derive the exact asymptotic learning rates for all combinations of *simple forwarding* and *sequential best guessing* relaying paired with *full matrix*, *weighted*, and *optimally-weighted* fusion (six distinct combinations). We show that no joint strategy universally outperforms the others.

Contribution 4: We characterize the learning rates of all schemes, as the number of relays grows asymptotically, to answer a fundamental question, "Is it better to have high spatial diversity with many low-quality channels or low spatial diversity with one or two high-quality channels?" We characterize the asymptotic behavior of the M parallel-relay learning rate

$$\lim_{M\to\infty} M\mathcal{L}(p(M), q(M)),$$

where $\mathcal{L}(p(M), q(M))$ represents the single-relay learning rate, and $p(M)$ and $q(M)$ are the crossover probabilities of the source-to-relay and relay-to-destination links, respectively, that scale with M as $M \to \infty$. In particular, we consider

$$p(M) = \frac{1}{2} - \frac{\frac{1}{2} - p}{M^\alpha} \to \frac{1}{2}, \text{ and } q(M) = \frac{1}{2} - \frac{\frac{1}{2} - q}{M^\alpha} \to \frac{1}{2}$$

as $M \to \infty$. The growth rate $\alpha > 0$ dictates how rapidly the multi-relay crossovers degrade to $1/2$. We then show that all schemes exhibit similar scaling behavior, namely, the learning rates either 1) converge to zero, 2) converge to a constant depending on the channel parameters, or 3) grow unbounded, depending on *how quickly* the channels degrade, i.e., depending on α. For both the optimal strategy and strategies employing sequential best guessing at the relays, we observe increased resilience against channel degradation when compared to simple forwarding.

Contribution 5: We characterize the influence of the channel parameters on the learning rates of all schemes to answer a fundamental question, "Is it better to direct resources to improve the reliability of the source-to-relay links or the reliability of the relay-to-destination links?" We show that, given a total amount of noise present in the combined source-to-relay and relay-to-destination links, the reliability tradeoff between the two link types depends on the scheme used. For the optimal scheme, both link types exert equal influence on the learning rate. For schemes employing simple forwarding at the relays, both link types also exert equal influence. However, with sequential best guessing at the relays, a range of behavior is observed, depending on 1) the total level of noise present and 2) the destination decoding strategy used.

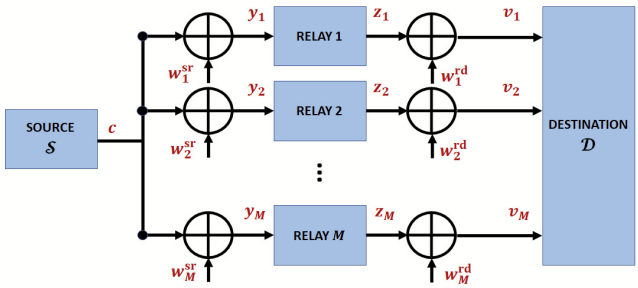


Fig. 1: System model of source, relays, and destination.

In Sec. II, we develop the system model. In Sec. III, we discuss existing single-relay results [13]–[15]. In Sec. IV, we upper bound the learning rate with M relays and show that the M-relay generalization of the independent sub-block-structured strategy [15] achieves the optimal learning rate. In Sec. V, we apply alternative strategies to address implementation shortcomings of the optimal scheme. In Sec. VI, we derive the learning rates of alternative schemes using full matrix fusion. In Sec. VII, we derive the learning rates of alternative schemes using weighted fusion. In Sec. VIII, we characterize the learning rate behavior as the number of relays grows asymptotically. In Sec. IX, we compare all learning rates and investigate the influence of the channel parameters. In Sec. X, we make concluding remarks.

II. System Model

A. Communication Model

The system model is depicted in Fig. 1. One of two equally likely binary messages, $\{0, 1\}$, are transmitted by a source $\mathcal{S}$ to a destination $\mathcal{D}$ by sequentially transmitting a codeword of blocklength n. The source $\mathcal{S}$, however, does not have a direct link to the destination $\mathcal{D}$. To facilitate communication, a collection of M parallel relays are used as intermediate hops to transmit the message over the source-to-relay ($\mathcal{S}\to\mathcal{R}$) and relay-to-destination ($\mathcal{R}\to\mathcal{D}$) links. Both the $\mathcal{S}\to\mathcal{R}$ and $\mathcal{R}\to\mathcal{D}$ links are independent binary symmetric channels (BSCs).

The source message $\theta \in \{0, 1\}$ is encoded, respectively, as the all-zeros or all-ones codewords, denoted $\{\mathbf{0}_n, \mathbf{1}_n\}$. Note that the two codewords have maximum distance. The particular maximum distance codewords are chosen without loss of generality, as any combination of source, relay, and destination mappings can be complemented to give the all-zeros and all-ones codewords. The source codeword with blocklength n is

$$\mathbf{c} = [c_0, c_1, \ldots, c_{n-1}] \in \{\mathbf{0}_n, \mathbf{1}_n\}, \tag{1}$$

and is transmitted sequentially to all relays simultaneously at a rate of one bit per channel use.

The received signal at relay $i \in [1, M] \triangleq \{1, 2, \ldots, M\}$ is

$$\mathbf{y}_i = [y_{i,0}, y_{i,1}, \ldots, y_{i,n-1}], \tag{2}$$

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2023.3292356

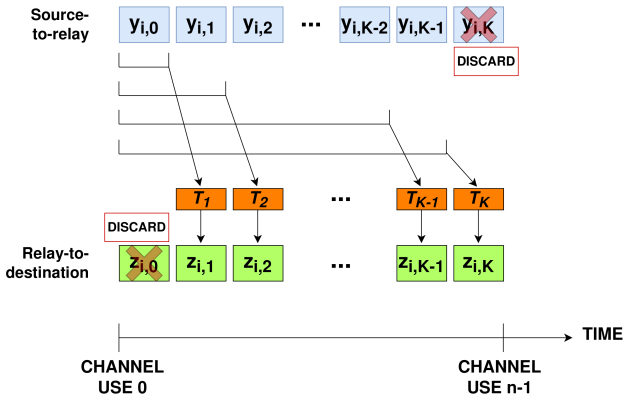


Fig. 2: Depiction of the strictly causal transcoding relaying framework.

whose components are $y_{i,t} = c_t \oplus w_{i,t}^{\mathrm{sr}}$, where $\oplus$ is the binary XOR operation, and $t \in [0, n-1]$ denotes the channel use. Each $w_{i,t}^{\mathrm{sr}} \in \{0,1\}$ is i.i.d. Bernoulli $\mathrm{Ber}(p_i)$, with $p_i \in [0, 0.5)$. The signal transmitted by relay i to the destination is

$$\mathbf{z}_i = [z_{i,0}, z_{i,1}, \ldots, z_{i,n-1}], \tag{3}$$

and the signal received at the destination from relay i is

$$\mathbf{v}_i = [v_{i,0}, v_{i,1}, \ldots, v_{i,n-1}], \tag{4}$$

where $v_{i,t} = z_{i,t} \oplus w_{i,t}^{\mathrm{rd}}$, and each $w_{i,t}^{\mathrm{rd}} \in \{0,1\}$ is i.i.d. Bernoulli $\mathrm{Ber}(q_i)$, with $q_i \in [0, 0, 5)$.

B. Relaying Description

To facilitate the design of time-varying relaying schemes, we introduce a transcoding framework, which is parameterized by a scalar integer parameter δ_T. The value of δ_T can range from $[1, \lfloor 0.5n \rfloor]$. When $\delta_T = 1$, the transcoding scheme is the most general and includes any possible designs as a special case; when $\delta_T = \lfloor 0.5n \rfloor$, the transcoding scheme is the most restricted, since it places the most stringent structure in the scheme design.

The transcoding framework is depicted in Fig. 2. For a transcoding scheme to perform time-varying mapping functions, the relays process and transmit $K+1$ successive sub-blocks of $\delta_T = \frac{n}{K+1}$ bits (assuming $K+1$ divides the blocklength n). The received signal at relay i and the signal transmitted by relay i are re-expressed with sub-blocks as

$$\mathbf{y}_i = [\mathbf{y}_{i,0}, \ldots, \mathbf{y}_{i,K}], \quad \mathbf{z}_i = [\mathbf{z}_{i,0}, \ldots, \mathbf{z}_{i,K}], \tag{5}$$

respectively. The k-th sub-blocks, for $k \in [0, K]$, are

$$\mathbf{y}_{i,k} = \left[y_{i,k\delta_T}, \ldots, y_{i,(k+1)\delta_T - 1}\right], \tag{6}$$

$$\mathbf{z}_{i,k} = \left[z_{i,k\delta_T}, \ldots, z_{i,(k+1)\delta_T - 1}\right]. \tag{7}$$

Each relay performs *strictly causal* encoding between its input and output signals, as depicted in Fig. 2. The first sub-block, $\mathbf{z}_{i,0}$, transmitted by relay i is discarded/ignored. Then, for $k \in [1, K]$, the sub-blocks transmitted by relay i are determined by strictly causal mapping functions

$$\mathbf{z}_{i,k} = \mathcal{T}_{i,k}\left(\mathbf{y}_{i,0}, \ldots, \mathbf{y}_{i,k-1}\right), \ k \in [1, K], \tag{8}$$

and we use

$$\tilde{\mathbf{z}}_i = [\mathbf{z}_{i,1}, \ldots, \mathbf{z}_{i,K}] \tag{9}$$

to denote the part of the signal transmitted by relay i that excludes the discarded sub-block $\mathbf{z}_{i,0}$. Furthermore, the strictly causal encoding structure entails that the last sub-block received at relay i, i.e., $\mathbf{y}_{i,K}$, is discarded and never participates in the relay's transmission of $\tilde{\mathbf{z}}_i$ (Fig. 2).

C. Destination Decoding

The signal received at the destination from relay i is re-expressed with sub-blocks as

$$\mathbf{v}_i = [\mathbf{v}_{i,0}, \mathbf{v}_{i,1}, \ldots, \mathbf{v}_{i,K}], \tag{10}$$

whose k-th sub-block, for $k \in [0, K-1]$, is

$$\mathbf{v}_{i,k} = \left[v_{i,k\delta_T}, \ldots, v_{i,(k+1)\delta_T - 1}\right]. \tag{11}$$

Since the first sub-blocks transmitted by all relays are discarded due to strictly causal mapping functions, the destination accumulates the received signals from all relays and discards the first sub-blocks received from the relays. Let

$$\tilde{\mathbf{v}}_i = [\mathbf{v}_{i,1}, \ldots, \mathbf{v}_{i,K}] \tag{12}$$

be the signal received from relay i excluding the discarded sub-block $\mathbf{v}_{i,0}$. We define

$$\mathbf{V}_K = \left[\tilde{\mathbf{v}}_1^\top, \tilde{\mathbf{v}}_2^\top, \ldots, \tilde{\mathbf{v}}_M^\top\right]^\top \tag{13}$$

as the $M \times \delta_T K$ received signal matrix at the destination, which excludes the discarded sub-blocks. The destination then applies a fusion function

$$g_K : \{0,1\}^{M \times \delta_T K} \times [0, 0.5)^M \times [0, 0.5)^M \to \{0,1\} \tag{14}$$

to produce a final decoding decision $\hat{\theta}_K$. For clarity, we note that the decoding function in (14) is indexed by K to show the dependence of decoding on the number of sub-blocks (which dictates the sub-block size δ_T). The final decoding decision is

$$\hat{\theta}_K = g_K\left(\mathbf{V}_K, \mathcal{P}, \mathcal{Q}\right), \tag{15}$$

where we use

$$\mathcal{P} = (p_1, p_2, \ldots, p_M), \quad \mathcal{Q} = (q_1, q_2, \ldots, q_M), \tag{16}$$

to compactly represent the crossover probabilities of the $\mathcal{S} \to \mathcal{R}$ and $\mathcal{R} \to \mathcal{D}$ links, respectively. In its general form, the decoding function uses the received signal matrix $\mathbf{V}_K$, as well as knowledge of the channel parameters. However, there are many functional decoding schemes that do not require knowledge of the channel parameters. The dependence will be made clear when describing specific decoding schemes.

III. Problem Statement and Existing Results

For given channel parameters $\mathcal{P}$ and $\mathcal{Q}$, we are interested in designing strategies to *maximize* the asymptotic learning rate of the final decoding decision at the destination, namely,

$$\mathcal{L}^{\max}(\mathcal{P}, \mathcal{Q}) = \sup_{\texttt{protocols}} \lim_{n \to \infty} \left\{ -\frac{1}{n} \log \mathbb{P}\left(\hat{\theta}_K \neq \theta\right) \right\}.$$

The `protocols` represent any joint design of the number of sub-blocks K, the destination decoding function g_K, and the relay mapping functions $\cup_{i=1}^{M} \{\mathcal{T}_{i,k}\}_{k=1}^{K}$.

A. Spatial Diversity

Additionally, we seek to quantify the spatial diversity gains achieved with M relays. Formally, let $\mathcal{L}(\mathcal{P},\mathcal{Q})$ and $\mathcal{L}(p,q)$ denote the learning rates achieved by performing a protocol with M relays and a single relay, respectively. When $p_i = p$ and $q_i = q$, for all $i \in [1, M]$, we define the spatial diversity gain as

$$\Upsilon(p,q,M) = \frac{\mathcal{L}(\mathcal{P},\mathcal{Q})}{\mathcal{L}(p,q)}. \tag{17}$$

The spatial diversity gain measures the ratio of the learning rate of an M-relay system to that of a single-relay system. Note that the M-relay system with n channel uses differs fundamentally from that of a single-relay system with nM channel uses. Specifically, in the M-relay system, there is no collaboration among the relays. Furthermore, compared to the M-relay system, the causality constraints of the single-relay system are less strict, in the sense that the single-relay gains access to the entire sequence of nM channel uses. This allows the single relay to fully comprehend the totality of the transmitted information. This knowledge is leveraged within the transcoding framework, enabling the single-relay to potentially make more informed decisions for its transmitted bits.

Because of these differences, it is critical to analyze various relaying schemes and to compare their performance under the more realistic collaboration and causality constraints of the multi-relay setup. In the subsequent sections, we demonstrate that for relaying schemes employing an independent sub-block structure, e.g., simple forwarding and the optimal scheme proposed by [15] (detailed in the next subsection), the error probability of an M-relay system with n channel uses is equivalent to that of a single-relay system with nM channel uses when $p_i = p$ and $q_i = q$ for all $i \in [1, M]$. Furthermore, the benefits of using multiple relays extend from the source's ability to broadcast its codeword bits to multiple relays in each channel use. When characterizing the optimal performance of M-relay system, if all the channel conditions are identical, i.e., $p_i = p$ and $q_i = q$ for all $i \in [1, M]$, then we can compare the learning rate to those of the single-relay results, since the M-relay system has more restrictive collaboration and causality constraints.

B. Existing Results

Single-relay analyses of the learning rate for various relaying and destination decoding strategies can be found in [13]–[15]. In our model, the single separated relay channel corresponds to $M = 1$, with $\mathcal{S} \to \mathcal{R}$ and $\mathcal{R} \to \mathcal{D}$ crossover probabilities p and q, respectively, and a learning rate expressed with the lower-case parameters as $\mathcal{L}(p,q)$. To simplify discussion of existing single-relay results in this section, we drop the relay subscripts and denote $\mathbf{y} = [\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_K]$, $\tilde{\mathbf{z}} = [\mathbf{z}_1, \ldots, \mathbf{z}_K]$, and $\tilde{\mathbf{v}} = [\mathbf{v}_1, \ldots, \mathbf{v}_K]$ as the received signal at the relay, the transmitted relay signal, and the received signal at the destination, respectively.

The converse (upper) bound on the learning rate for the single-relay case is [13]–[15]

$$\mathcal{L}(p,q) \le \mathcal{L}^*(p,q) = D_{\mathrm{kl}}\left(0.5 || \max\{p,q\}\right), \tag{18}$$

by recognizing that the learning rate is upper bounded by

$$\mathcal{L}(p,q) \le \mathcal{L}^*(p,0) = D_{\mathrm{kl}}(0.5||p), \tag{19}$$

$$\mathcal{L}(p,q) \le \mathcal{L}^*(0,q) = D_{\mathrm{kl}}(0.5||q), \tag{20}$$

which are the optimal learning rates achievable when one of the two $\mathcal{S} \to \mathcal{R}$ and $\mathcal{R} \to \mathcal{D}$ crossover probabilities are held constant and the other is set to zero, reducing the separated relay into a single point-to-point channel. Considering both bounds (19)–(20) together, the upper bound (18) is given.

In [13], the following relaying strategies were considered:

- *Simple forwarding*: This is a special case of the transcoding scheme when setting the number of bits per sub-block to $\delta_T = 1$, hence $K = n-1$. Following strictly causal encoding, the relay transmits its $(k-1)$-th bit observation to the destination at time k, i.e.,

$$\mathbf{z}_k = \mathcal{T}_k(\mathbf{y}_0, \ldots, \mathbf{y}_{k-1}) = \mathbf{y}_{k-1},\ \forall\, k \in [1, K]. \tag{21}$$

- *Sequential best guessing*: The number of bits per sub-block is $\delta_T = 1$, with $K = n-1$. The k-th sub-block transmitted by the relay is the outcome of a majority vote over all sub-blocks received at the relay thus far, i.e.,

$$\begin{aligned}\mathbf{z}_k &= \mathcal{T}_k\left(\mathbf{y}_0, \ldots, \mathbf{y}_{k-1}\right) \\ &= \mathrm{Maj}\left([\mathbf{y}_0, \ldots, \mathbf{y}_{k-1}]\right),\ \forall\, k \in [1, K],\end{aligned} \tag{22}$$

  where $\mathrm{Maj}(\cdot) \in \{0,1\}$ denotes the majority vote taken over an input binary vector.

In conjunction with the above relaying protocols, [13] considered several destination decoding strategies and analyzed all combinations of joint relaying and destination decoding schemes. The destination decoding schemes are as follows:

- *Majority voting*: The final decoding decision $\hat{\theta}_K$ is a majority vote of the received signal at the destination, i.e., $\hat{\theta}_K = \mathrm{Maj}\left(\tilde{\mathbf{v}}\right)$.
- *ϵ-Majority voting*: The decoding decision is given by $\hat{\theta}_K = \mathrm{Maj}\left(\tilde{\mathbf{v}}^{\epsilon}\right)$, with superscript ϵ indicating the last $\lfloor \epsilon(n-1) \rfloor$ bits of $\tilde{\mathbf{v}}$. The intuition behind only processing the last $\lfloor \epsilon(n-1) \rfloor$ bits is that, when paired with well-designed relaying strategies, one hopes that later bits in the received signal at the destination are more reliable, due to the relay correcting errors in the received signal using more information.

The work [13] showed that no combination of these joint relaying and destination decoding strategies achieves the optimal learning rate $\mathcal{L}^*(p,q)$. Furthermore, none of the exact asymptotic learning rates of these schemes universally dominate the others for all channel parameters.

The work [14] considered the simple forwarding and sequential best guessing relaying schemes, as well as a protocol that updates the majority vote every $\delta_T = \sqrt{n}$ bits received at the relay. The updated vote is then transmitted to the destination for the next length-$\sqrt{n}$ sub-block. This relaying strategy was analyzed jointly with maximum likelihood (ML)

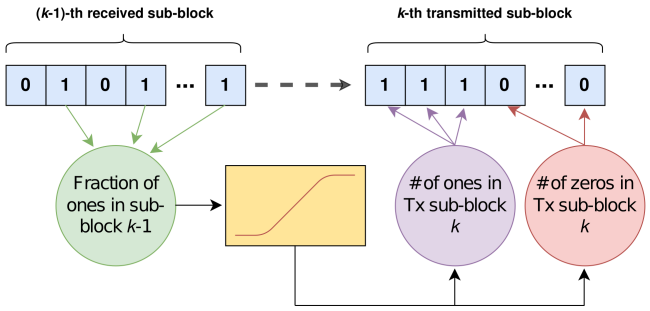


Fig. 3: Independent sub-block-structured relaying protocol developed in [15].

destination decoding, but the learning rate falls short of meeting the converse $\mathcal{L}^*(p,q)$. However, the learning rate is within a factor of $3/4$ for the case of $p=q$ and $p\to 0.5$ (p in the vicinity of 0.5). It was conjectured in [14] that more sophisticated schemes could get arbitrarily close to the converse $\mathcal{L}^*(p,q)$.

The work [15] devised an independent sub-block-structured relaying strategy for the single relay that *achieves* the converse $\mathcal{L}^*(p,q)$ when the $\mathcal{S}\to\mathcal{R}$ and $\mathcal{R}\to\mathcal{D}$ links are BSCs, hence proving the conjecture in [14]. The relay processes a sequence sub-blocks, where the sub-block size satisfies $\delta_T<n$, $\delta_T\to\infty$ and $n/\delta_T\to\infty$ as $n\to\infty$, i.e., the sub-block size is $\delta_T=o(n)$. The sub-blocks are processed independently, meaning that each sub-block $\mathbf{z}_k$ transmitted by the relay is only a function of the received sub-block $\mathbf{y}_{k-1}$ at the relay. As depicted in Fig. 3, by letting $\upsilon_{k-1}\in[0,1]$ be the fraction of ones in the $(k-1)$-th received sub-block at the relay, the k-th sub-block transmitted by the relay is comprised of $\lfloor\delta_T W(\upsilon_{k-1})\rfloor$ ones followed by $\delta_T-\lfloor\delta_T W(\upsilon_{k-1})\rfloor$ zeros, where the function $W:[0,1]\to[0,1]$ is designed in [15] to be

$$W(\upsilon)=\begin{cases}0, & \upsilon\in[0,p),\\ \frac{D_{\mathrm{kl}}(\upsilon||p)}{2D_{\mathrm{kl}}(0.5||p)}, & \upsilon\in[p,0.5],\\ 1-W(1-\upsilon), & \upsilon\in(0.5,1-p),\\ 1, & \upsilon\in[1-p,1],\end{cases}\tag{23}$$

which is monotonically increasing. Implementing the function $W(\cdot)$ also requires the relay to know the $\mathcal{S}\to\mathcal{R}$ crossover probability p. The destination applies optimal ML decoding.

The scheme in [15] achieves the optimal learning rate over the two-hop separated relay channel with $\mathcal{S}\to\mathcal{R}$ and $\mathcal{R}\to\mathcal{D}$ crossover probabilities p and q, respectively. With respect to the learning rate, it is equivalent to transmitting the codeword $\mathbf{c}$ over a single point-to-point BSC whose crossover probability is the bottleneck of the two relay links, i.e., $\max\{p,q\}$, and performing ML destination decoding. Achieving the optimal learning rate requires that the sub-block size be sub-linearly growing in terms of the blocklength. However, the first sub-block of the $\mathcal{R}\to\mathcal{D}$ hop and the last sub-block of the $\mathcal{S}\to\mathcal{R}$ hop are discarded completely. For a small and finite tolerable end-to-end delay, which is usually the case in practice, the error probability of the scheme [15] may suffer since a non-negligible portion of the channel uses is discarded, even though its (asymptotic) learning rate is provably optimal.

IV. Optimal Rate for the Parallel Relay Channel

In this section, we upper bound the learning rate of the problem and then show that the upper bound can be attained by an M-relay generalization of the strategy in [15].

A. Upper Bound on the Learning Rate

Theorem 1. *For any joint relaying and destination decoding strategy, the learning rate, $\mathcal{L}(\mathcal{P},\mathcal{Q})$, of the final decoding decision at the destination is upper bounded by*

$$\mathcal{L}(\mathcal{P},\mathcal{Q})\le\sum_{i=1}^{M}D_{\mathrm{kl}}(0.5||\max\{p_i,q_i\}).\tag{24}$$

Proof. See Appendix A. □

B. The Generalized M-Relay Scheme

For M relays, we propose generalizing the scheme developed in [15] by having each relay perform the same protocol in parallel using its independently received sub-blocks. Each relay processes a sequence of $K+1$ sub-blocks, with the sub-block length satisfying $\delta_T=o(n)$.

For non-discarded sub-blocks $k\in[1,K]$ of relay $i\in[1,M]$, the sub-block $\mathbf{z}_{i,k}$ transmitted by the relay is a function of the received sub-block $\mathbf{y}_{i,k-1}$ at that relay. Letting $\upsilon_{i,k-1}$ be the fraction of ones in the received sub-block $\mathbf{y}_{i,k-1}$, i.e., $\upsilon_{i,k-1}=\mathrm{wt}\left(\mathbf{y}_{i,k-1}\right)/\delta_T$, where $\mathrm{wt}(\cdot)$ is the Hamming weight, the transmitted sub-block $\mathbf{z}_{i,k}$ is a sorted sequence of $\lfloor\delta_T W(\upsilon_{i,k-1})\rfloor$ ones followed by $\delta_T-\lfloor\delta_T W(\upsilon_{i,k-1})\rfloor$ zeros. As [15] remarks, $\delta_T W(\upsilon_{i,k-1})$ is not always an integer. However, it is shown for the single-relay case that performing the learning rate analysis with $\delta_T W(\upsilon_{i,k-1})$ ones followed by $\delta_T(1-W(\upsilon_{i,k-1}))$ zeros helps with analytical tractability and does not change the asymptotic error analysis. We adopt this modification in our M-relay analysis.

The destination applies an ML fusion strategy. That is, since each relay channel is independent, we define the log-likelihood ratio (LLR) as

$$\mathrm{LLR}_{\mathrm{opt}}\left(\mathbf{V}_K,\mathcal{P},\mathcal{Q}\right)=\sum_{i=1}^{M}\sum_{k=1}^{K}\log\frac{\mathbb{P}\left(\mathbf{v}_{i,k}|\theta=0\right)}{\mathbb{P}\left(\mathbf{v}_{i,k}|\theta=1\right)},\tag{25}$$

and choose $\hat{\theta}_K$ with the LLR test

$$\hat{\theta}_K=\begin{cases}0, & \mathrm{LLR}_{\mathrm{opt}}\left(\mathbf{V}_K,\mathcal{P},\mathcal{Q}\right)\ge 0,\\ 1, & \mathrm{LLR}_{\mathrm{opt}}\left(\mathbf{V}_K,\mathcal{P},\mathcal{Q}\right)<0.\end{cases}\tag{26}$$

C. Statistical Closeness Metric

In this subsection, we review properties of the Bhattacharyya coefficient required to analyze the learning rate of the joint strategy from Sec. IV-B. We define the Bhattacharyya coefficient between two discrete random variables A and B defined on a sample space Ω as

$$\rho\left(A,B\right)=\sum_{x\in\Omega}\sqrt{\mathbb{P}\left(A=x\right)\mathbb{P}\left(B=x\right)},\tag{27}$$

where $\rho(A,B)=0$ if and only if the supports of A and B are disjoint. Similarly, $\rho(A,B)=1$ if and only if A and B

have identical distributions. We now list three standard properties of the Bhattacharyya coefficient that will be useful for characterizing the learning rate of the M-relay generalization of the independent sub-block-structured protocol.

Property 1: If a random variable X is known to be distributed according to A or B, then there exists a strategy, when X is drawn in a random trial, to obtain a classification error probability of at most $\rho(A,B)$. This error probability can be achieved using the ML test. For $X=x$, A is decided if $\mathbb{P}(A=x)>\mathbb{P}(B=x)$, and B otherwise. The probability of incorrectly choosing distribution A when B is true is [15]

$$\begin{aligned}\sum_{x\in\Omega}\mathbb{P}(B=x)\mathbb{I}\left(\mathbb{P}(A=x)>\mathbb{P}(B=x)\right)&\\ \leq\sum_{x\in\Omega}\sqrt{\mathbb{P}(A=x)\mathbb{P}(B=x)},&\\ =\rho(A,B).&\end{aligned}\tag{28}$$

The same applies for incorrectly choosing B when A is true, giving the error bound

$$\mathbb{P}\left(\text{error}\right)\leq\rho(A,B).\tag{29}$$

Property 2: Let $\mathbf{x}=(x_1,x_2)\in\Omega^2$. For A_1,A_2 independent and B_1,B_2 independent, with supports Ω, we have

$$\begin{aligned}&\rho\left((A_1,A_2),(B_1,B_2)\right)\\ &\quad=\sum_{\mathbf{x}\in\Omega^2}\sqrt{\mathbb{P}\left(A_1{=}x_1,A_2{=}x_2\right)\mathbb{P}\left(B_1{=}x_1,B_2{=}x_2\right)}\\ &\quad=\sum_{\mathbf{x}\in\Omega^2}\sqrt{\mathbb{P}\left(A_1{=}x_1\right)\mathbb{P}\left(B_1{=}x_1\right)\mathbb{P}\left(A_2{=}x_2\right)\mathbb{P}\left(B_2{=}x_2\right)}\\ &\quad=\rho(A_1,B_1)\rho(A_2,B_2),\end{aligned}\tag{30}$$

by expanding the terms and rearranging.

Property 3: For $\mathbf{A}=(A_1,A_2,\ldots,A_K)$ and $\mathbf{B}=(B_1,B_2,\ldots,B_K)$, where $\{A_k\}_{k=1}^K$ and $\{B_k\}_{k=1}^K$ are i.i.d. following the distributions of A and B, respectively, then $\rho\left(\mathbf{A},\mathbf{B}\right)=\rho\left(A,B\right)^K$ by following the same procedure that establishes Property 2 and noting that, in addition to being independent, the same-letter random variables are also *identically* distributed.

Keeping with (29), we assume for the case of Property 3, i.e., observing K i.i.d. observations following the distribution of A or B, that the ML decoding rule is adopted, resulting in the multi-variate upper bound on the decision error [15], given by

$$\mathbb{P}\left(\text{error}\right)\leq\rho(A,B)^K.\tag{31}$$

D. Lower Bound on the Learning Rate

Since sub-blocks received over the same relay channel are i.i.d., we use B_i as shorthand notation to represent the non-discarded sub-blocks $\{\mathbf{v}_{i,k}\}_{k=1}^K$ received on the i-th parallel channel at the destination. Moreover, $B_{i|0}$ and $B_{i|1}$ denote conditioning of B_i on the source message $\theta=0$ and $\theta=1$, respectively. Using Properties 1–3, the error probability is upper bounded by

$$\mathbb{P}\left(\hat{\theta}_K\neq\theta\right)\leq\prod_{i=1}^{M}\rho\left(B_{i|0},B_{i|1}\right)^K.\tag{32}$$

For a given sub-block size δ_T, the learning rate, denoted $\mathcal{L}_{\text{opt}}\left(\mathcal{P},\mathcal{Q}\right)$ for the generalized independent sub-block-structured scheme, is lower bounded as the blocklength n grows large by

$$\mathcal{L}_{\text{opt}}\left(\mathcal{P},\mathcal{Q}\right)\geq-\frac{1}{\delta_T}\sum_{i=1}^{M}\log\rho\left(B_{i|0},B_{i|1}\right).\tag{33}$$

As [15] remarks, the distributions $B_{i|0}$ and $B_{i|1}$ depend on δ_T, and while a fixed δ_T provides a lower bound to the learning rate, this also holds true when δ_T grows large with $\delta_T=o(n)$, which requires the additional limiting operation $\lim_{\delta_T\to\infty}$.

Since $\mathcal{L}_{\text{opt}}(\mathcal{P},\mathcal{Q})$ is bounded by the sum of terms that are functions of $\{\rho\left(B_{i|0},B_{i|1}\right)\}_{i=1}^M$, a closed-form bound on the learning rate can be achieved by determining bounds for each individual component of the sum in (33), which has been done in [15]. For a single relay channel characterized by crossover probabilities p and q, it was established in [15, Thm. 1] that

$$\mathcal{L}_{\text{opt}}(p,q)\geq D_{\text{kl}}(0.5||\max\{p,q\}).\tag{34}$$

Specializing the result (34) for each relay $i\in[1,M]$, we obtain

$$\mathcal{L}_{\text{opt}}(\mathcal{P},\mathcal{Q})\geq\sum_{i=1}^{M}D_{\text{kl}}(0.5||\max\{p_i,q_i\})\tag{35}$$

as $n\to\infty$, $\delta_T\to\infty$, and $\delta_T=o(n)$.

Recalling the upper bound, $\mathcal{L}^*(\mathcal{P},\mathcal{Q})$, on the learning rate of the M-relay setting (Theorem 1), we see that the lower bound (35) exactly coincides with $\mathcal{L}_{\text{opt}}(\mathcal{P},\mathcal{Q})=\mathcal{L}^*(\mathcal{P},\mathcal{Q})$, and hence the joint relaying and destination decoding strategy attains the optimal learning rate. In the special case where the $\mathcal{S}\to\mathcal{R}$ and $\mathcal{R}\to\mathcal{D}$ crossover probabilities satisfy $p_i=p$ and $q_i=q$ for all $i\in[1,M]$ and any $p,q\in[0,0.5)$, a direct comparison can be made to the single-relay scheme in [15]. In this case, we denote $\Upsilon_{\text{opt}}(p,q,M)$ as the spatial diversity gain of the optimal strategy, defined as the ratio of the learning rate for the M-relay case to the single-relay one, i.e.,

$$\Upsilon_{\text{opt}}(p,q,M)=\frac{\mathcal{L}_{\text{opt}}\left(\mathcal{P},\mathcal{Q}\right)}{\mathcal{L}^*(p,q)}=M,\tag{36}$$

hence the optimal strategy *fully utilizes* the spatial diversity gain provided by M relays.

V. Alternative Strategies

In this section, we apply alternative relaying and destination decoding strategies to address the implementation shortcomings of the (asymptotically) optimal scheme.

A. Alternative Relaying Strategies

We consider two relaying schemes, namely, simple forwarding and sequential best guessing, which were defined for the single-relay scenario (in (21) and (22), respectively). We extend these schemes to M relays, where they are applied independently and in parallel for each relay. Unlike the asymptotically optimal scheme, neither alternative relaying protocol must know the $\mathcal{S}\to\mathcal{R}$ channel parameters. Moreover, the alternative relaying protocols only discard the first channel use of the relay transmit signals, which may lead to improved error performance for small tolerable end-to-end delay.

B. Alternative Destination Decoding Strategies

In wireless networks, a destination node may receive parallel transmissions from multiple relays. Two practical scenarios are: (1) orthogonal multiplexing over frequency or space at a single fusion center, and making a final decoding decision using all n bits from each of the M relays (full matrix fusion) [18]; (2) a collection of distributed sub-decoders that make single-bit decoding decisions (weighted fusion) [25]. The latter scenario enables low-power and low-complexity sub-decoders, i.e., oblivious to knowledge of the channels and relaying functions, but requires a more sophisticated fusion center to make the final decoding decision [26]. In the full matrix fusion scheme, all nM bits are decoded *jointly* when the destination lacks knowledge of the relaying functions and the channel information. In the weighted fusion scheme, an inner-outer decoding approach is taken, where each of the M sub-decoders makes a hard decoding decision, and the sub-decoder outcomes are weighted to produce the final decoding decision. The weights can be selected to maximize the learning rate given channel information.

For the sub-decoders, majority voting is a natural choice in the absence of channel or relaying knowledge. However, a more sophisticated fusion center can address these sub-decoders limitations [27]. The fusion center leverages asymmetric channel knowledge to generate decisions based on a global view of the network and assigns sub-decoder weights using knowledge of the relaying functions, channel qualities, and sub-decoding functions to improve decoding performance. Next, we formalize the full matrix and weighted fusion schemes.

Full matrix fusion: A final decoding decision $\hat{\theta}_K$ is generated by a majority vote of the received signal matrix $\mathbf{V}_K$, i.e.,

$$\hat{\theta}_K = \mathrm{Maj}\left(\mathrm{vec}\left(\mathbf{V}_K\right)\right), \tag{37}$$

where $\mathrm{vec}\left(\mathbf{V}_K\right) = [\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_M]$ is the vectorization of the M rows of $\mathbf{V}_K$ into a single row vector. The destination does not need channel or relaying function knowledge.

Weighted fusion: A final decoding decision is generated by combining M single-bit sub-decoder decisions. Precisely, the i-th single-bit-quantized message estimate is given by

$$\hat{\theta}_{i,K} = \mathrm{Maj}\left(\tilde{\mathbf{v}}_i\right), \ \forall i \in [1, M]. \tag{38}$$

The M sub-decoder outcomes, defined with a vector as

$$\hat{\boldsymbol{\theta}}_K = \left[\hat{\theta}_{1,K}, \hat{\theta}_{2,K}, \ldots, \hat{\theta}_{M,K}\right] \in \{0,1\}^M, \tag{39}$$

are then fed into a fusion function that combines the M outputs to produce the final decoding decision $\hat{\theta}_K$. The weighted fusion strategy implements a linear threshold function (LTF) [28] characterized by a vector $\Pi = [\pi_0, \pi_1, \ldots, \pi_M] \in \mathbb{R}_+^{M+1}$, where $\mathbb{R}_+$ denotes the non-negative real numbers [28]. Using Π, the final decoding decision is generated as

$$\hat{\theta}_K = \begin{cases} 0, & \pi_0 + \sum_{i=1}^{M} (-1)^{\hat{\theta}_{i,K}} \pi_i \geq 0, \\ 1, & \pi_0 + \sum_{i=1}^{M} (-1)^{\hat{\theta}_{i,K}} \pi_i < 0. \end{cases} \tag{40}$$

Optimally-weighted fusion: With simple forwarding and sequential best guessing at the relays, we are interested in analytically finding the optimal weights $\Pi_{SF}^*(\mathcal{P}, \mathcal{Q})$ and $\Pi_{SG}^*(\mathcal{P}, \mathcal{Q})$, respectively, that maximize the learning rate with weighted fusion.

TABLE I: Equation numbers for the learning rates of alternative relaying and destination decoding strategies.

	Full matrix	Weighted	Optimally-weighted
Simple forwarding	Eqn. (68)	Eqn. (87)	Eqn. (89)
Seq. best guessing	Eqn. (50)	Eqn. (75)	Eqn. (88)

Lemma 1. *For simple forwarding and sequential best guessing at the relays with weighted fusion at the destination, respectively, the following weights are asymptotically optimal:*

$$\Pi_{SF}^*(\mathcal{P}, \mathcal{Q}) = \left[0, \ell_1^{SF}(p_1, q_1), \ldots, \ell_M^{SF}(p_M, q_M)\right], \tag{41}$$

$$\Pi_{SG}^*(\mathcal{P}, \mathcal{Q}) = \left[0, \ell_1^{SG}(p_1, q_1), \ldots, \ell_M^{SG}(p_M, q_M)\right], \tag{42}$$

where $\ell_i^{SF}(p_i, q_i)$ and $\ell_i^{SG}(p_i, q_i)$ are the learning rates of simple forwarding and sequential best guessing of the i-th branches, respectively. These learning rates determine the weights $\{\pi_i\}_{i=1}^M$ of the signed summation in (40)*. Specifically, the offset for both relaying schemes is $\pi_0 = 0$, and the weights of the signed summation are equal to the learning rates of their respective sub-decoders. We provide explicit definitions of these learning rates in equations* (74) *and* (66)*, which can be found in subsequent sections.*

Proof. See Appendix B. □

Thus far, we have considered two relaying strategies and three destination decoding strategies. Totally, there are six different combinations that can be deployed in a practical system. In the following sections, we characterize the learning rates of all combinations. For the weighted fusion schemes at the destination, we assume the weights are arbitrary. Since we analytically characterize the learning rates for arbitrary weights, we can further specialize to the optimal weights, provided in Lemma 1, and the results with optimal weights are also reported. Table I depicts the equation numbers where the learning rates are located throughout the paper.

VI. Learning Rate of Full Matrix Fusion

In this section, we derive the learning rates when simple forwarding and sequential best guessing at the relays are paired with full matrix fusion at the destination.

A. Preliminary Calculations

In order to determine the learning rates with full matrix fusion, we first develop large deviation properties of the relay transmit signals as well mixtures of independent Bernoulli random variables in Lemmas 2 and 3, respectively. Lemma 2 is a restatement of [13, Thm. 1] using our notation (which differs significantly from [13]), whereas Lemma 3 is derived directly in this work. To begin, let the transmitted relay signals be re-written as $\mathbf{z}_i = \mathbf{c} \oplus \mathbf{e}_i$, where

$$\mathbf{e}_i = [e_{i,0}, e_{i,1}, \ldots, e_{i,n-1}] \in \{0,1\}^n \tag{43}$$

is defined as the distortion vector of relay i. The Hamming weight, $\mathrm{wt}(\mathbf{e}_i)$, encapsulates the *distortion* between the transmitted relay signal $\mathbf{z}_i$ and the transmitted source codeword.

Lemma 2 ([13, Thm. 1])**.** *Let $\delta \in [0,1]$. With sequential best guessing at relay $i \in [1,M]$, the relay distortion vector weight satisfies*

$$\lim_{n\to\infty} \left\{ -\frac{1}{n} \log \mathbb{P}\left(\mathrm{wt}\left(\mathbf{e}_i\right) \geq \delta n\right) \right\} = \delta D_{\mathrm{kl}}(0.5||p_i). \quad (44)$$

Readers can refer to [13, Thm. 1] for verification of the result.[1] The result in Lemma 2 is obtained by recasting the received signal at relay i as as a p_i-biased random walk where the bit values $\{0,1\}$ are modulated as $\{+1,-1\}$, respectively, so that the state of the random walk relative to the zero state determines the identity of bits transmitted by the relay at each time. The weight $\mathrm{wt}\left(\mathbf{e}_i\right)$ of the relay distortion vector is viewed as the number of times the relay transmit signal is incorrect. With this perspective, the distribution of $\mathbb{P}\left(\mathrm{wt}\left(\mathbf{e}_i\right) \geq \delta n\right)$ is broken down into simpler cases by conditioning on various aspects of the random walk, such as the time of the final visit and number of returns to the zero state. The result (44) is then determined for large values of n by applying the Gärtner-Ellis theorem [29].

Next, we develop a fundamental result for mixtures of independent Bernoulli random variables. For analytical convenience, we characterize sums of Bernoulli random variables and normalize by the total number of variables, rather than writing Hamming weights. Note that the rate function I of a general sequence of random variables $\{A_n\}$ is such that for any closed set $F \subseteq \mathbb{R}$ and open set $G \subseteq \mathbb{R}$, the inequalities

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(A_n \in F\right) \leq -\inf_{w\in F} I(w),$$
$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(A_n \in G\right) \geq -\inf_{w\in G} I(w),$$

are satisfied [13], [29].

Lemma 3. *Let $\rho_i \in [0,1]$ and $q_i \in [0,0.5)$ for $i \in [1,M]$. For each i, consider sequences of i.i.d. $\mathrm{Ber}(1-q_i)$ random variables $\{U_{i,j}\}_{j=1}^{n-\lfloor \rho_i n \rfloor}$ and i.i.d. $\mathrm{Ber}(q_i)$ random variables $\{V_{i,j}\}_{j=1}^{\lfloor \rho_i n \rfloor}$, where the $U_{i,j}$ variables are independent of the $V_{i,j}$ variables. Next, define the sample mean*

$$\bar{W} = \frac{\sum_{i=1}^{M} \left(\sum_{j=1}^{n-\lfloor \rho_i n \rfloor} U_{i,j} + \sum_{j=1}^{\lfloor \rho_i n \rfloor} V_{i,j} \right)}{Mn}, \quad (45)$$

where Mn is the total number of random variables in the sample mean. The random variable $\bar{W}$ satisfies the large deviation principle and has a rate function

$$I_{\boldsymbol{\rho}}\left(w, \mathcal{Q}\right) = \sup_{\lambda} \left\{ \lambda w - \sum_{i=1}^{M} \left(\bar{\rho}_i \log\left(e^{\lambda/M} \bar{q}_i + q_i \right) + \rho_i \log\left(e^{\lambda/M} q_i + \bar{q}_i \right) \right) \right\}, \quad (46)$$

[1]Note that Lemma 2 pertains to sequential best guessing, for which $\delta_T{=}1$ (likewise for simple forwarding). Thus, the first bits of the transmitted relay signals are discarded. Asymptotically, it makes no difference whether we use n or $n{-}1$. For compactness, we use n to describe the results in Table I.

where $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_M)$, $\bar{\rho}_i = 1-\rho_i$, and $\bar{q}_i = 1-q_i$. The supremum is achieved when

$$w = \frac{1}{M} \left[\sum_{i=1}^{M} \left(\bar{\rho}_i \frac{e^{\lambda/M} \bar{q}_i}{e^{\lambda/M} \bar{q}_i + q_i} + \rho_i \frac{e^{\lambda/M} q_i}{e^{\lambda/M} q_i + \bar{q}_i} \right) \right] \quad (47)$$

is satisfied.

Proof. See Appendix C. □

The solution to (47) is a polynomial equation in e^{λ} that can be solved numerically when the number of channels is $M \geq 2$. In [13], a specialization of Lemma 3 is solved for the single-relay case $M = 1$. In this case, (47) reduces to a quadratic equation in e^{λ}, and the rate function was derived explicitly in [13, Lem. 4] for a crossover probability q and mixture fraction ρ as

$$I_{\rho}(w,q) = w\log(\eta) - \bar{\rho}\log(\bar{q}\eta + q) - \rho\log(q\eta + \bar{q}), \quad (48)$$

with

$$\eta = \frac{-\tau + \sqrt{\tau^2 + 4w\bar{w}}}{2\bar{w}},\ \tau = \frac{\bar{q}}{q}(\bar{\rho} - w) + \frac{q}{\bar{q}}(\rho - w), \quad (49)$$

where we have defined $\bar{w} = 1-w$, $\bar{\rho} = 1-\rho$, and $\bar{q} = 1-q$.

B. Statement of the Results

The next theorem derives the learning rate of sequential best guessing at the relays with full matrix fusion at the destination. The results readily extend to simple forwarding at the relays after.

Theorem 2. *Using sequential best guessing at the relays with full matrix fusion at the destination, the learning rate of the final decoding decision, denoted $\mathcal{L}_{SG}^{F}(\mathcal{P},\mathcal{Q})$, is*

$$\mathcal{L}_{SG}^{F}(\mathcal{P},\mathcal{Q}) = \inf_{\boldsymbol{\rho}\in[0,1]^M} \left\{ \sum_{i=1}^{M} \rho_i D_{\mathrm{kl}}(0.5||p_i) + \hat{I}\left(\boldsymbol{\rho},\mathcal{Q}\right) \right\}, \quad (50)$$

where

$$\hat{I}\left(\boldsymbol{\rho},\mathcal{Q}\right) = \inf_{w\in[0,0.5]} I_{\boldsymbol{\rho}}\left(w,\mathcal{Q}\right), \quad (51)$$

and $I_{\boldsymbol{\rho}}\left(w,\mathcal{Q}\right)$ is the rate function defined in (46)*.*

Proof. Let ε_K denote the error event that $\hat{\theta}_K \neq \theta$ and consider an integer N, to be specified later. First, we divide the interval $(0,n]$ into intervals $\{F_j\}_{j=0}^{N-1}$, where

$$F_j = \left(\frac{n(N-j-1)}{N}, \frac{n(N-j)}{N} \right]. \quad (52)$$

Note that, for $i \in [1,M]$,

$$\mathbb{P}\left(\mathrm{wt}(\mathbf{e}_i) \in F_j\right) = \mathbb{P}\left(\mathrm{wt}(\mathbf{e}_i) > \frac{n(N-j-1)}{N} \right) - \mathbb{P}\left(\mathrm{wt}(\mathbf{e}_i) > \frac{n(N-j)}{N} \right). \quad (53)$$

Defining $\mathbf{j} = (j_1, j_2, \ldots, j_M) \in [0, N-1]^M = \mathcal{J}$, the error probability with full matrix fusion is

$$\mathbb{P}(\varepsilon_K) = \sum_{\mathbf{j}\in\mathcal{J}} \mathbb{P}(\varepsilon_K \mid \mathrm{wt}(\mathbf{e}_1) \in F_{j_1}, \ldots, \mathrm{wt}(\mathbf{e}_M) \in F_{j_M}) \cdot \left(\prod_{i=1}^{M} \mathbb{P}(\mathrm{wt}(\mathbf{e}_i) \in F_{j_i}) \right). \quad (54)$$

Let $\epsilon_1 > 0$ be an arbitrarily small constant. Lemma 2 dictates that for all $j_i \in [0, N-1]$, for each $i \in [1, M]$, and for fixed N, the following inequality is satisfied for n sufficiently large

$$\left| -\frac{1}{n} \log \mathbb{P}(\mathrm{wt}(\mathbf{e}_i) \in F_{j_i}) - \frac{N-j_i-1}{N} D_{\mathrm{kl}}(0.5||p_i) \right| < \epsilon_1, \quad (55)$$

because the first term of (53) dominates for fixed N and n sufficiently large. Moreover, letting

$$\xi_{j_i,\mathrm{fl}} = \left\lfloor \frac{n(N-j_i-1)}{N} \right\rfloor, \quad \xi_{j_i,\mathrm{cl}} = \left\lceil \frac{n(N-j_i)}{N} \right\rceil \quad (56)$$

be used for compactness, upper and lower bounds on $\mathbb{P}(\varepsilon_K)$ are formed when conditioning on the weight of the relay distortion vectors of all channels simultaneously as

$$\begin{aligned} &\mathbb{P}(\varepsilon_K | \mathrm{wt}(\mathbf{e}_1) = \xi_{j_1,\mathrm{fl}}, ..., \mathrm{wt}(\mathbf{e}_M) = \xi_{j_M,\mathrm{fl}}) \\ &\le \mathbb{P}(\varepsilon_K \mid \mathrm{wt}(\mathbf{e}_1) \in F_{j_1}, \ldots, \mathrm{wt}(\mathbf{e}_M) \in F_{j_M}) \\ &\le \mathbb{P}(\varepsilon_K | \mathrm{wt}(\mathbf{e}_1) = \xi_{j_1,\mathrm{cl}}, \ldots, \mathrm{wt}(\mathbf{e}_M) = \xi_{j_M,\mathrm{cl}}). \end{aligned} \quad (57)$$

To verify (57), note that the error event ε_K corresponds to $\hat{\theta}_K \neq \theta$, where $\hat{\theta}_K$ is a majority vote of bits in the received signal matrix $\mathbf{V}_K$. Conditioned on the Hamming weights $\{\mathrm{wt}(\mathbf{e}_i)\}_{i=1}^M$ of the relay distortion vectors, the weight of all bits in $\mathbf{V}_K$ follows a Poisson-binomial distribution [30], [31]. The bounds (57) are readily confirmed using well-established results on the stochastic orderings of Poisson-binomial distributions (see, for example, [31, Thm. 3.2]).

Let $\epsilon_2 > 0$ be an arbitrarily small constant. Using the large deviation result of Lemma 3, for sufficiently large n and each $\mathbf{j} \in \mathcal{J}$, the following bounds hold:

$$\left| -\frac{1}{n} \log \mathbb{P}(\varepsilon_K | \mathrm{wt}(\mathbf{e}_1) = \xi_{j_1,\mathrm{cl}}, \ldots, \mathrm{wt}(\mathbf{e}_M) = \xi_{j_M,\mathrm{cl}}) - \hat{I}\left(\frac{N-j_1}{N}, \ldots, \frac{N-j_M}{N}, \mathcal{Q} \right) \right| < \epsilon_2, \quad (58)$$

$$\left| -\frac{1}{n} \log \mathbb{P}(\varepsilon_K | \mathrm{wt}(\mathbf{e}_1) = \xi_{j_1,\mathrm{fl}}, \ldots, \mathrm{wt}(\mathbf{e}_M) = \xi_{j_M,\mathrm{fl}}) - \hat{I}\left(\frac{N-j_1-1}{N}, \ldots, \frac{N-j_M-1}{N}, \mathcal{Q} \right) \right| < \epsilon_2, \quad (59)$$

because conditioned on the fraction of incorrect relay distortion bits $\mathrm{wt}(\mathbf{e}_i)/n = (N-j_i)/N$, the distribution of bits received at the destination behaves as a Bernoulli mixture with $\rho_i = 1 - j_i/N$. Moreover, the conditional error probability is the infimum of the rate function over the interval $[0, 0.5]$, corresponding to full matrix fusion taking a majority vote over the received signal matrix.

For sufficiently large n, we have for each $\mathbf{j} \in \mathcal{J}$ that

$$\begin{aligned} &e^{-n\left(\hat{I}\left(\frac{N-j_1-1}{N}, \ldots, \frac{N-j_M-1}{N}, \mathcal{Q}\right) + \epsilon_2\right)} \\ &< \mathbb{P}(\varepsilon_K \mid \mathrm{wt}(\mathbf{e}_1) \in F_{j_1}, \ldots, \mathrm{wt}(\mathbf{e}_M) \in F_{j_M}) \\ &< e^{-n\left(\hat{I}\left(\frac{N-j_1}{N}, \ldots, \frac{N-j_M}{N}, \mathcal{Q}\right) - \epsilon_2\right)}. \end{aligned} \quad (60)$$

Combining (60) with the bounds (55), the probability of error is bounded above and below as

$$\mathbb{P}(\varepsilon_K) \le \sum_{\mathbf{j}\in\mathcal{J}} e^{-n f_{\mathrm{ub}}(\mathbf{j})}, \quad \mathbb{P}(\varepsilon_K) \ge \sum_{\mathbf{j}\in\mathcal{J}} e^{-n f_{\mathrm{lb}}(\mathbf{j})}, \quad (61)$$

where

$$f_{\mathrm{ub}}(\mathbf{j}) = \sum_{i=1}^{M} \left(\frac{N-j_i-1}{N} \right) D_{\mathrm{kl}}(0.5||p_i) + \hat{I}\left(\frac{N-j_1}{N}, \ldots, \frac{N-j_M}{N}, \mathcal{Q} \right) - \epsilon_1 - \epsilon_2, \quad (62)$$

$$f_{\mathrm{lb}}(\mathbf{j}) = \sum_{i=1}^{M} \left(\frac{N-j_i-1}{N} \right) D_{\mathrm{kl}}(0.5||p_i) + \hat{I}\left(\frac{N-j_1-1}{N}, \ldots, \frac{N-j_M-1}{N}, \mathcal{Q} \right) + \epsilon_1 + \epsilon_2. \quad (63)$$

Let $\epsilon_3 > 0$ be an arbitrarily small constant. Let $\mathbf{x} = (x_1, x_2, \ldots, x_M) \in [0,1]^M$. We define three quantities:

$$a = \inf_{\mathbf{x}\in[0,1]^M} \left\{ \sum_{i=1}^{M} x_i D_{\mathrm{kl}}(0.5||p_i) + \hat{I}(x_1, \ldots, x_M, \mathcal{Q}) \right\},$$

$$a^* = \inf_{\mathbf{j}\in\mathcal{J}} \left\{ \sum_{i=1}^{M} \frac{N-j_i-1}{N} D_{\mathrm{kl}}(0.5||p_i) + \hat{I}\left(\frac{N-j_1}{N}, \ldots, \frac{N-j_M}{N}, \mathcal{Q} \right) \right\},$$

$$a_* = \inf_{\mathbf{j}\in\mathcal{J}} \left\{ \sum_{i=1}^{M} \frac{N-j_i-1}{N} D_{\mathrm{kl}}(0.5||p_i) + \hat{I}\left(\frac{N-j_1-1}{N}, \ldots, \frac{N-j_M-1}{N}, \mathcal{Q} \right) \right\}.$$

Due to the continuity of $\hat{I}(\cdot,\cdot)$, the value of N can be selected depending on ϵ_3 and $\mathcal{P}$ so that

$$\max\{|a-a^*|, |a-a_*|\} < \epsilon_3. \quad (64)$$

When n is sufficiently large, the following bounds hold

$$\mathbb{P}(\varepsilon_K) \le NMe^{-n(a-\epsilon_1-\epsilon_2-\epsilon_3)}, \ \mathbb{P}(\varepsilon_K) \ge e^{-n(a+\epsilon_1+\epsilon_2+\epsilon_3)}.$$

Dividing by $-n$ then taking the logarithm and limit gives

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}(\varepsilon_K) = a, \quad (65)$$

which completes the proof. □

Next, we present a corollary that immediately follows from Theorem 2 to provide the sub-decoder learning rates when sequential best guessing is used at the relays. This result will

aide in characterizing the learning rates with the weighted fusion strategy in the next section.

Corollary 1 ([13, Thm. 2])**.** *For source message θ, sub-decoder i commits an error if $\hat{\theta}_{i,K} = \mathrm{Maj}\left(\tilde{\mathbf{v}}_i\right) \neq \theta$. The sub-decoder learning rate with sequential best guessing at the relays is*

$$\ell_i^{SG}(p_i, q_i) = \inf_{\rho\in[0,1]} \left\{\rho D_{\mathrm{kl}}(0.5||p_i) + \hat{I}(\rho, q_i)\right\}, \tag{66}$$

where $\hat{I}(\rho, q_i) = \inf_{w\in[0,0.5]} I_\rho(w, q_i)$, and $I_\rho(w, q_i)$ is defined in (48)*.*

Proof. The result follows immediately from Theorem 2 when $M = 1$. The single $\mathcal{S}\to\mathcal{R}$ and $\mathcal{R}\to\mathcal{D}$ crossover probabilities are p_i and q_i, respectively. The Bernoulli mixture rate function is specialized to the single-relay form in (48), which is also derived in [13, Lem. 4]. □

Note that Corollary 1 is proven in [13, Thm. 2], which only studied the single relay. In our work, we derive the corollary as a *specialization* of the general M-relay setting.

For simple forwarding at the relays with full matrix fusion at the destination, the learning rate is a straightforward application of Lemma 3. With simple forwarding, bits arriving to the destination from relay i are i.i.d. with crossover probability $p_i * q_i = p_i(1-q_i) + q_i(1-p_i)$, since the relay is forwarding its received bits with no processing. Letting

$$\mathcal{P} * \mathcal{Q} = (p_1 * q_1, p_2 * q_2, \ldots, p_M * q_M) \tag{67}$$

be the crossover probabilities of the M parallel channels when the links from the relays to the destination are viewed as superchannels, the learning rate, denoted by $\mathcal{L}_{SF}^F(\mathcal{P}, \mathcal{Q})$, is

$$\mathcal{L}_{SF}^F(\mathcal{P}, \mathcal{Q}) = \hat{I}\left(\mathbf{0}, \mathcal{P} * \mathcal{Q}\right) = \inf_{w\in[0,0.5]} I_{\mathbf{0}}\left(w, \mathcal{P} * \mathcal{Q}\right). \tag{68}$$

C. Spatial Diversity Gains

When $p_i = p$ and $q_i = q$ for all $i \in [1, M]$, the learning rate for sequential best guessing and simple forwarding at the relays with full matrix fusion at the destination simplifies due to the symmetry of the M identical relaying channels. For the sequential best guessing case, we have

$$\mathcal{L}_{SG}^F(\mathcal{P}, \mathcal{Q}) = M \inf_{\rho\in[0,1]} \left\{\rho D_{\mathrm{kl}}(0.5||p) + \hat{I}\left(\rho, q\right)\right\}, \tag{69}$$

i.e., the learning rate of a single sub-decoder times the number of relays. For simple forwarding, the learning rate becomes

$$\mathcal{L}_{SF}^F(\mathcal{P}, \mathcal{Q}) = M\hat{I}(0, p*q) = MD_{\mathrm{kl}}(0.5||p*q). \tag{70}$$

Denoting $\Upsilon_{SG}^F(p, q, M)$ and $\Upsilon_{SF}^F(p, q, M)$ as the diversity gains of the alternative relaying schemes paired with full matrix fusion, both schemes attain the maximum spatial diversity

$$\Upsilon_{SG}^F(p, q, M) = \Upsilon_{SF}^F(p, q, M) = M. \tag{71}$$

VII. Learning Rate of Weighted Fusion

In this section, we derive the exact asymptotic learning rates of both simple forwarding and sequential best guessing at the relays paired with the weighted fusion strategies at the destination.

A. Statement of the Results

The LTF function (see (40)) used for weighted fusion is characterized by a vector $\Pi \in \mathbb{R}^{M+1}$. Without loss of generality, let ε_K be the error event that $\hat{\theta}_K = 1$, assuming $\theta = 0$ was sent. Let $\varepsilon_{i,K}$ be the error event $\hat{\theta}_{i,K} \neq \theta$ of sub-decoder i. Let $\varphi_{i,K} = \mathbb{P}\left(\varepsilon_{i,K}\right)$ be the error probability of sub-decoder i. The final decoding decision error probability is

$$\mathbb{P}\left(\varepsilon_K\right) = \sum_{k_1=0}^{M} \left(\sum_{A\in T_{k_1}} \left(\prod_{k_2\in A} \varphi_{k_2,K} \prod_{k_3\in A^c} \left(1-\varphi_{k_3,K}\right) \right) \cdot \chi\left(A, \Pi\right) \right), \tag{72}$$

where

$$\chi\left(A, \Pi\right) = \mathbb{I}\left(\pi_0 + \sum_{k'\in A^c} \pi_{k'} - \sum_{k\in A} \pi_k < 0\right), \tag{73}$$

T_k is defined as the set of all subsets of k integers than can be chosen from $[1, M]$, $A^c = [1, M] \setminus A$ is the complement of set A, and $\mathbb{I}\left(\cdot\right)$ is the indicator function.

The error probability (72) takes the sum over all possible sub-decoder outcomes $\hat{\boldsymbol{\theta}}_K \in \{0, 1\}^M$, weighting each probability term by a 0 or 1, depending on whether the LTF is less than 0 (when the decoder outputs the incorrect message estimate $\hat{\theta}_K = 1$ given that $\theta = 0$ was sent). In the following theorem, we assume that sequential best guessing is used at the relays, thus the sub-decoder learning rates are $\{\ell_i^{SG}(p_i, q_i)\}_{i=1}^M$, which are derived in Corollary 1. The results in Theorem 3 are readily extended to simple forwarding at the relays by substituting the sub-decoder learning rates with $\{\ell_i^{SF}(p_i, q_i)\}_{i=1}^M$, where the sub-decoder learning rates are

$$\ell_i^{SF}(p_i, q_i) = D_{\mathrm{kl}}(0.5||p_i * q_i). \tag{74}$$

Theorem 3. *Let $\{\ell_i^{SG}(p_i, q_i)\}_{i=1}^M$ be the learning rates of the M sub-decoders when using sequential best guessing at the relays, and let Π be the vector defining the weighted fusion function. The learning rate of the final decoding decision, denoted $\mathcal{L}_{SG}^{\Pi}(\mathcal{P}, \mathcal{Q})$, is*

$$\mathcal{L}_{SG}^{\Pi}(\mathcal{P}, \mathcal{Q}) = \min\left\{ \sum_{k\in A} \ell_k^{SG}(p_k, q_k) \,\middle|\, A \in T_i, \; i \in [1, M], \, \chi(A, \Pi) = 1 \right\}, \tag{75}$$

where $\chi\left(A, \Pi\right)$ is defined in (73)*.*

Proof. Let N_i divide the interval $(0, n]$ into subintervals, denoted by $\{F_j\}_{j=0}^{N_i-1}$ as defined in (52), for each $i \in [1, M]$.

The i-th parallel sub-decoder's learning rate and upper and lower bounds are reiterated as follows:

$$a_i = \inf_{x\in[0,1]} \left\{ x D_{\mathrm{kl}}(0.5||p_i) + \hat{I}(x, q_i) \right\},$$

$$a^{i,*} = \inf_{j\in[0,N_i-1]} \left\{ \frac{N_i-j-1}{N_i} D_{\mathrm{kl}}(0.5||p_i) + \hat{I}\left(\frac{N_i-j}{N_i}, q_i\right) \right\},$$

$$a_{i,*} = \inf_{j\in[0,N_i-1]} \left\{ \frac{N_i-j-1}{N_i} D_{\mathrm{kl}}(0.5||p_i) + \hat{I}\left(\frac{N_i-j-1}{N_i}, q_i\right) \right\}.$$

Given an arbitrary constant $\epsilon_3 > 0$ and p_i, we choose $N_i = N_i(\epsilon_3, p_i)$ so that

$$\max\left\{ \left|a_i - a^{i,*}\right|, |a_i - a_{i,*}| \right\} < \epsilon_3 \tag{76}$$

is satisfied. For arbitrary $\epsilon_1, \epsilon_2 > 0$, by selecting $N = \max\{N_1, N_2, \ldots, N_M\}$, then there exists some sufficiently large $n^* \in \mathbb{N}$ such that

$$\varphi_{i,K} = \mathbb{P}(\varepsilon_{i,K}) \le N e^{-n(a_i - \epsilon_1 - \epsilon_2 - \epsilon_3)}, \tag{77}$$

$$\varphi_{i,K} = \mathbb{P}(\varepsilon_{i,K}) \ge e^{-n(a_i + \epsilon_1 + \epsilon_2 + \epsilon_3)}, \tag{78}$$

for all $n \ge n^*$ and each $i \in [1, M]$.

Applying the upper and lower bounds in (77)–(78), the error probability is upper bounded by

$$\begin{aligned}\mathbb{P}(\varepsilon_K) &\le \sum_{k_1=0}^{M} \left(\sum_{A\in T_{k_1}} \left(\prod_{k_2\in A} \varsigma_{k_2,K} \prod_{k_3\in A^c} (1-\varrho_{k_3,K}) \right) \cdot \chi(A,\Pi) \right) \\ &= P_{\mathrm{UB},K},\end{aligned} \tag{79}$$

where

$$\varsigma_{k,K} = N e^{-n(a_k - \epsilon_1 - \epsilon_2 - \epsilon_3)}, \quad \varrho_{k,K} = e^{-n(a_k + \epsilon_1 + \epsilon_2 + \epsilon_3)},$$

and is similarly lower bounded by

$$\begin{aligned}\mathbb{P}(\varepsilon_K) &\ge \sum_{k_1=0}^{M} \left(\sum_{A\in T_{k_1}} \left(\prod_{k_2\in A} \varrho_{k_2,K} \prod_{k_3\in A^c} (1-\varsigma_{k_3,K}) \right) \cdot \chi(A,\Pi) \right) \\ &= P_{\mathrm{LB},K}.\end{aligned} \tag{80}$$

For functions of the form

$$\tilde{f}(n) = \sum_{u=0}^{U} C_u e^{-d_u n} \left(\prod_{u'\in A(u)} \left(1 - e^{-m_{u'} n}\right) \right), \tag{81}$$

where $A(u)$ is any finite set of positive integers determined by u, C_u and $m_{u'}$ are positive constants for all u and u', and $d_0 = 0$, we have that by taking the logarithm, dividing by $-n$, and taking the limit as $n \to \infty$, the smallest exponent of the $\{d_u\}_{u=1}^{U}$ terms is extracted, i.e.,

$$\lim_{n\to\infty} -\frac{1}{n} \log \tilde{f}(n) = \min\{d_1, d_2, \ldots, d_U\}. \tag{82}$$

With the bounds on the probability of decoding error, by taking the logarithm, dividing by $-n$, and taking the limit, the property (82) can be applied to obtain

$$\begin{aligned}&\lim_{n\to\infty} -\frac{1}{n} \log P_{\mathrm{UB},K} \\ &\quad \ge \min\left\{ \sum_{k\in A} a_k \ : \ A \in T_i,\, i \in [1,M],\, \chi(A,\Pi) = 1 \right\} \\ &\qquad - (\epsilon_1 + \epsilon_2 + \epsilon_3)\, \Psi,\end{aligned} \tag{83}$$

where $\Psi \in \mathbb{N}$ is a finite positive integer, dependent on the behavior of the LTF. Similarly,

$$\begin{aligned}&\lim_{n\to\infty} -\frac{1}{n} \log P_{\mathrm{LB},K} \\ &\quad \le \min\left\{ \sum_{k\in A} a_k \ : \ A \in T_i,\, i \in [1,M],\, \chi(A,\Pi) = 1 \right\} \\ &\qquad + (\epsilon_1 + \epsilon_2 + \epsilon_3)\, \Psi.\end{aligned} \tag{84}$$

As a consequence of (83) and (84), the learning rate is bounded by

$$\left| -\frac{1}{n} \log \mathbb{P}(\varepsilon_K) - \mathcal{L}_{SG}^{\Pi}(\mathcal{P}, \mathcal{Q}) \right| \le (\epsilon_1 + \epsilon_2 + \epsilon_3)\Psi, \tag{85}$$

for sufficiently large n. This entails that

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}(\varepsilon_K) = \mathcal{L}_{SG}^{\Pi}(\mathcal{P}, \mathcal{Q}), \tag{86}$$

since ϵ_1, ϵ_2, and ϵ_3 are arbitrary. This completes the proof. □

Analogously, the learning rate using simple forwarding at the relays and weighted fusion at the destination, denoted $\mathcal{L}_{SF}^{\Pi}(\mathcal{P}, \mathcal{Q})$, is obtained by exchanging the learning rates $\ell_i^{SG}(p_i, q_i)$ with $\ell_i^{SF}(p_i, q_i)$ for all $i \in [1, M]$, i.e.,

$$\begin{aligned}\mathcal{L}_{SF}^{\Pi}(\mathcal{P}, \mathcal{Q}) = \min\Bigg\{ &\sum_{k\in A} \ell_k^{SF}(p_k, q_k) \Bigg| A \in T_i, \\ &i \in [1,M],\, \chi(A,\Pi) = 1 \Bigg\}.\end{aligned} \tag{87}$$

The learning rates with weighted fusion are dictated by a *subset* of sub-decoder outcomes $\hat{\boldsymbol{\theta}}_K \in \{0,1\}^M$ producing an incorrect final decoding decision, i.e., $\hat{\theta}_K \ne \theta$. For each incorrect sub-decoder outcome $\hat{\boldsymbol{\theta}}_K$, the learning rates of the individually incorrect sub-decoders are summed, and the learning rate with weighted fusion is the minimum of these sums. This aligns with the learning rate being governed by the worst-case behavior.

Substituting the optimal weights in (42), the learning rate of optimally-weighted fusion with sequential best guessing is

$$\begin{aligned}\mathcal{L}_{SG}^{\Pi^*}(\mathcal{P}, \mathcal{Q}) = \min\Bigg\{ &\sum_{k\in A} \ell_k^{SG}(p_k, q_k) \Bigg| A \in T_i, \\ &i \in [1,M],\, \Phi_{SG}(A) = 1 \Bigg\},\end{aligned} \tag{88}$$

where

$$\Phi_{SG}(A) = \mathbb{I}\left(\sum_{k'\in A^c} \ell_{k'}^{SG}(p_{k'}, q_{k'}) - \sum_{k\in A} \ell_k^{SG}(p_k, q_k) < 0 \right).$$

Substituting the optimal weights given in (41), the learning rate of optimally-weighted fusion with simple forwarding at the relays is

$$\mathcal{L}_{SF}^{\Pi^*}(\mathcal{P},\mathcal{Q}) = \min\left\{ \sum_{k\in A} \ell_k^{SF}(p_k,q_k) \middle| A\in T_i, \right. \\ \left. i\in[1,M],\, \Phi_{SF}(A)=1 \right\}, \tag{89}$$

where

$$\Phi_{SF}(A) = \mathbb{I}\left(\sum_{k'\in A^c} \ell_{k'}^{SF}(p_{k'},q_{k'}) - \sum_{k\in A} \ell_k^{SF}(p_k,q_k) < 0 \right).$$

B. Spatial Diversity Gains

When the crossover probabilities satisfy $p_i = p$ and $q_i = q$ for all $i \in [1, M]$, we denote the sub-decoder learning rates for sequential best guessing and simple forwarding as $\ell^{SG}(p,q)$ and $\ell^{SF}(p,q)$, respectively. This implies that $\ell_i^{SG}(p_i,q_i) = \ell^{SG}(p,q)$ and $\ell_i^{SF}(p_i,q_i) = \ell^{SF}(p,q)$ for all $i \in [1,M]$. In this case, a direct comparison can be made between M relays (using optimally-weighted fusion) and a single relay (using majority voting at the destination). Specifically, the optimally-weighted fusion strategy reduces to a majority vote over the M sub-decoder outputs. In this case, the learning rates $\mathcal{L}_{SG}^{\Pi^*}(\mathcal{P},\mathcal{Q})$ and $\mathcal{L}_{SF}^{\Pi^*}(\mathcal{P},\mathcal{Q})$ simplify to

$$\begin{aligned}
\mathcal{L}_{SG}^{\Pi^*}(\mathcal{P},\mathcal{Q}) &= \min\left\{ |A|\ell^{SG}(p,q) \middle| A\in T_i, i\in[1,M], \right. \\
&\qquad \left. \Phi_{SG}(A)=1 \right\} \\
&= \left\lceil \frac{M}{2} \right\rceil \ell^{SG}(p,q), \\
\mathcal{L}_{SF}^{\Pi^*}(\mathcal{P},\mathcal{Q}) &= \min\left\{ |A|\ell^{SF}(p,q) \middle| A\in T_i, i\in[1,M], \right. \\
&\qquad \left. \Phi_{SF}(A)=1 \right\} \\
&= \left\lceil \frac{M}{2} \right\rceil \ell^{SF}(p,q),
\end{aligned}$$

since the smallest sets A satisfying $\Phi_{SG}(A) = 1$ and $\Phi_{SF}(A) = 1$ are of size $|A| = \lceil M/2 \rceil$ when the sub-decoder learning rates across M relays are identical. Leveraging this fact, we denote the diversity gains of optimally-weighted fusion with sequential best guessing and simple forwarding, respectively, as $\Upsilon_{SG}^{\Pi^*}(p,q,M)$ and $\Upsilon_{SF}^{\Pi^*}(p,q,M)$, which equal

$$\Upsilon_{SG}^{\Pi^*}(p,q,M) = \Upsilon_{SF}^{\Pi^*}(p,q,M) = \left\lceil \frac{M}{2} \right\rceil, \tag{90}$$

Therefore, the full diversity gain of M is *not* achieved when the alternative relaying strategies are paired with optimally-weighted fusion at the destination.

VIII. Multi-Relay Scaling Behavior

In this section, we investigate the multi-relay scaling behavior. To formalize, let p and q be the $\mathcal{S}\to\mathcal{R}$ and $\mathcal{R}\to\mathcal{D}$ crossover probabilities, respectively, of the single-relay channels. We parameterize the M-relay crossover probabilities by

$$p_i\left(M,p,\alpha\right) = \frac{1}{2} - \frac{\frac{1}{2}-p}{M^\alpha}, \;\; q_i\left(M,q,\alpha\right) = \frac{1}{2} - \frac{\frac{1}{2}-q}{M^\alpha}, \tag{91}$$

for $i \in [1,M]$. We use $\mathcal{P}(M,p,\alpha)$ to denote the setting that all M $\mathcal{S}\to\mathcal{R}$ channels are of crossover probability $p_i(M,p,\alpha)$. Similarly, we define $\mathcal{Q}(M,q,\alpha)$. Two remarks are in order. First, the crossover probabilities of the same link types are equal. Second, parameterization as a function of M ensures that the quality of channels decreases as the number of relays grows. The parameter $\alpha > 0$ controls *how rapidly* the crossover probabilities become completely unreliable. We note that the parameterization of $p_i(M,p,\alpha)$ and $q_i(M,q,\alpha)$ in (91) using M^α is a modeling choice. However, this parameterization is extremely general and not restrictive. In fact, by analyzing the behavior of learning rates as $M\to\infty$ with different parameterizations of crossover probabilities (e.g., $M!$ or $\log M$ instead of M^α for various values of $\alpha > 0$), we can compare the learning rates to known convergent or divergent sequences of learning rates parameterized by M^α and quickly identify the asymptotic behavior of the learning rates.

Optimal Scheme: Since the maximizer of $p_i(M,p,\alpha)$ or $q_i(M,q,\alpha)$ is the same for any M, the learning rate of the optimal scheme becomes

$$\begin{aligned}
&\mathcal{L}_{\text{opt}}\left(\mathcal{P}(M,p,\alpha),\mathcal{Q}(M,q,\alpha)\right) \\
&\qquad = M D_{\text{kl}}\left(0.5 \middle|\middle| \frac{1}{2} - \frac{\frac{1}{2}-\max\{p,q\}}{M^\alpha}\right).
\end{aligned} \tag{92}$$

With the form of equation (92), we observe several interesting behaviors of the asymptotic learning rate in the next lemma.

Lemma 4. *For the learning rate in* (92)*, the behavior as the number of relays $M\to\infty$ is*

$$\begin{aligned}
&\lim_{M\to\infty} \mathcal{L}_{\text{opt}}(\mathcal{P}(M,p,\alpha),\mathcal{Q}(M,q,\alpha)) \\
&\qquad = \begin{cases} +\infty, & \alpha < 0.5, \\ 2\left(\frac{1}{2}-\max\{p,q\}\right)^2, & \alpha = 0.5, \\ 0, & \alpha > 0.5. \end{cases}
\end{aligned} \tag{93}$$

Proof. See Appendix D. □

The growth rate $\alpha = 0.5$ behaves as a critical point, and when $\alpha < 0.5$, the learning rate grows unbounded, indicating that spatial diversity gains outweigh the pace of channel degradation.

Simple Forwarding: For tractability, we restrict the crossover probabilities to $p_i(M,t,\alpha) = q_i(M,t,\alpha)$, for $t \in [0, 0.5)$. With optimally-weighted fusion, equal weights are assigned to each sub-decoder, and the learning rate with M relays is

$$\begin{aligned}
&\mathcal{L}_{SF}^{\Pi^*}\left(\mathcal{P}(M,t,\alpha),\mathcal{Q}(M,t,\alpha)\right) \\
&\qquad = \left\lceil \frac{M}{2} \right\rceil D_{\text{kl}}\left(0.5 \middle|\middle| \left(\frac{1}{2} - \frac{\frac{1}{2}-t}{M^\alpha}\right) * \left(\frac{1}{2} - \frac{\frac{1}{2}-t}{M^\alpha}\right)\right).
\end{aligned} \tag{94}$$

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2023.3292356

Lemma 5. *For simple forwarding with optimally-weighted fusion, where $p_i(M,t,\alpha) = q_i(M,t,\alpha)$ for some $t \in [0, 0.5)$, the learning rate as the number of relays $M \to \infty$ is*

$$\lim_{M\to\infty} \mathcal{L}_{SF}^{\Pi^*}\left(\mathcal{P}(M,t,\alpha), \mathcal{Q}(M,t,\alpha)\right) = \begin{cases} +\infty, & \alpha < 1/4, \\ 4\left(\frac{1}{2}-t\right)^4, & \alpha = 1/4 \\ 0, & \alpha > 1/4. \end{cases}$$

Proof. See Appendix E. □

For simple forwarding with optimally-weighted fusion, the critical point is $\alpha = 1/4$, which shows reduced resiliency against channel degradation when compared to the optimal scheme.

When $p_i(M,t,\alpha) = q_i(M,t,\alpha)$, the learning rate of simple forwarding with full matrix fusion simplifies to

$$\mathcal{L}_{SF}^{F}\left(\mathcal{P}(M,t,\alpha), \mathcal{Q}(M,t,\alpha)\right) = MD_{\text{kl}}\left(0.5 \middle|\middle| \left(\frac{1}{2} - \frac{\frac{1}{2}-t}{M^\alpha}\right) * \left(\frac{1}{2} - \frac{\frac{1}{2}-t}{M^\alpha}\right)\right),$$

which is the same form as (94), except the scaling by M rather than $\lceil M/2 \rceil$ due to the improved spatial diversity gain. Thus, Lemma 5 can be applied to the asymptotic behavior of simple forwarding with full matrix fusion as

$$\lim_{M\to\infty} \mathcal{L}_{SF}^{F}\left(\mathcal{P}(M,t,\alpha), \mathcal{Q}(M,t,\alpha)\right) = \begin{cases} +\infty, & \alpha < 1/4, \\ 8\left(\frac{1}{2}-t\right)^4, & \alpha = 1/4, \\ 0, & \alpha > 1/4. \end{cases}$$

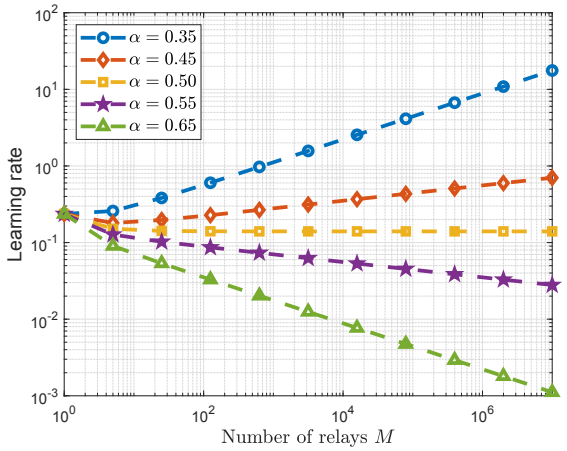
Sequential Best Guessing: The scaling behavior with sequential best guessing is difficult to characterize due to the complicated learning rate expressions. However, when $p_i(M,t,\alpha) = q_i(M,t,\alpha)$ for $t \in [0, 0.5)$, the learning rates simplify to scaled versions of the single-relay learning rates. Let $t(M,\alpha) = \frac{1}{2} - \frac{\frac{1}{2}-t}{M^\alpha}$ be the crossover probability of all links as a function of M. The learning rates become

$$\mathcal{L}_{SG}^{\Pi^*}\left(\mathcal{P}(M,t,\alpha), \mathcal{Q}(M,t,\alpha)\right) = \left\lceil \frac{M}{2} \right\rceil \inf_{\rho\in[0,1]} \left\{\rho D_{\text{kl}}(0.5||t(M,\alpha)) + \hat{I}(\rho, t(M,\alpha))\right\},$$

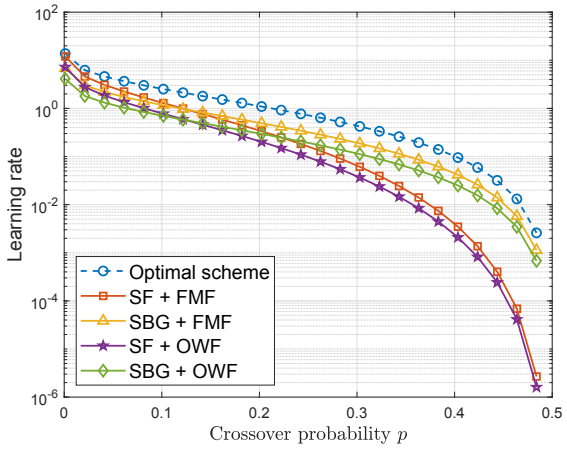
$$\mathcal{L}_{SG}^{F}\left(\mathcal{P}(M,t,\alpha), \mathcal{Q}(M,t,\alpha)\right) = M \inf_{\rho\in[0,1]} \left\{\rho D_{\text{kl}}(0.5||t(M,\alpha)) + \hat{I}(\rho, t(M,\alpha))\right\}.$$

Fig. 4a depicts the learning rate of sequential best guessing with full matrix fusion (denoted SBG + FMF in the figure) as the number of relays increases and demonstrates that $\alpha = 0.5$ is a critical point.[2] For $\alpha = 0.5$, the learning rate converges to a positive constant depending on the channel parameters. For $\alpha > 0.5$, it decays to zero. For $\alpha < 0.5$, the learning rate grows unbounded, showing resiliency to channel degradation on par with the optimal scheme.

[2]Note that the same behavior is observed with optimally-weighted fusion, as the only difference is the diversity gain as a scaling factor.



(a) SBG + FMF vs. M for various α values.



(b) Comparison of the optimal and alternative schemes.

Fig. 4: (a) Learning rate of SBG + FMF vs. M; (b) Learning rates of the optimal and alternative schemes for M=5, with p_i=p and q_i=q for all $i \in [1, M]$.

Lastly, we clarify that for Fig. 4a and all remaining figures, we use the following notation to improve readability: simple forwarding (SF), sequential best guessing (SBG), full matrix fusion (FMF), and optimally-weighted fusion (OWF).

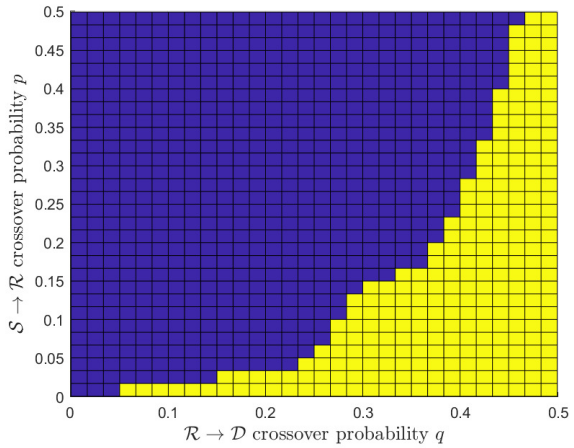
IX. Performance Comparison and Channel Influence

A. Comparison of Relaying and Decoding Strategies

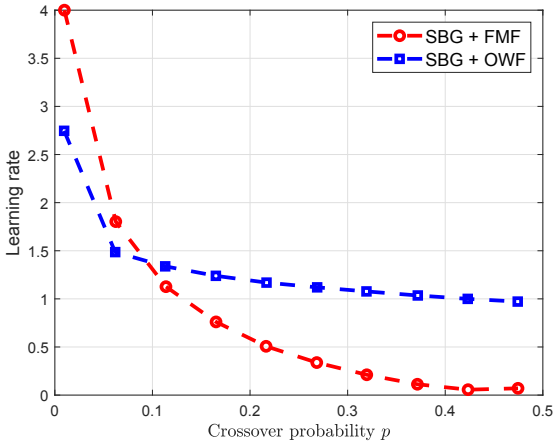
We first compare the learning rates of the various strategies. For $M = 5$ relays, we select $p_i = p$ and $q_i = p$ for all $i \in [1, M]$, and vary $p \in [0, 0.5)$. As Fig. 4b confirms, the optimal scheme always yields the largest learning rate of all the schemes. For both full matrix and optimally-weighted fusion at the destination, the learning rates with simple forwarding dominate sequential best guessing when the noise parameter is small. However, as p increases, sequential best guessing has the higher learning rate.

Next, we compare the learning rates of simple forwarding and sequential best guessing at the relays with optimally-

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2023.3292356



(a) Higher learning rate: blue = SF+OWF; yellow = SBG+OWF



(b) Comparison of SBG + FMF and SBG + OWF schemes.

Fig. 5: (a) Learning rate comparison of SF + OWF and SBG + OWF; (b) Learning rate comparison of SBG + FMF and SBG + OWF.

weighted fusion at the destination.[3] We set $M{=}5$, and parameterize the $\mathcal{S}{\to}\mathcal{R}$ and $\mathcal{R}{\to}\mathcal{D}$ channel parameters as $\mathcal{P} = [p, p/2, p/3, p/4, p/5]$ and $\mathcal{Q} = [q, q/2, q/3, q/4, q/5]$, respectively, so that the channel qualities among the M relays vary, hence the sub-decoders have different weights. Fig. 5a shows that the superior relaying protocol depends on the crossover probabilities. Optimally-weighted fusion applies more weight to more reliable sub-decoders. Simple forwarding dominates when the crossover probabilities are low (Fig. 5a for $p, q \in [0, 0.05]$). Thus, simple forwarding dominates over a wider range of channel parameters, due to assigning more weight to the superior sub-decoders.

In Fig. 5b, we compare full matrix and optimally-weighted fusion for $M{=}5$ relays employing sequential best guessing. The $\mathcal{S} \to \mathcal{R}$ crossover probabilities are $\mathcal{P} = [p, p, p, p, p/100]$, with $p \in [0, 0.5)$. We set $\mathcal{Q} = \mathcal{P}$, so that the reliability of the first four relaying channels are equal, and the fifth relay channel has links with $100\times$ higher reliability. For small p, the learning rate is higher with full matrix fusion, since a final decoding decision is made with all generally reliable bits of the received signal matrix. This outperforms optimally-weighted fusion, since the sub-decoders quantize information. As p increases, performance with full matrix fusion is throttled due to four unreliable relay channels outvoting the single reliable channel. Conversely, optimally-weighted fusion *adjusts* the sub-decoder weights to largely ignore unreliable outcomes, leading to a higher learning rate than full matrix fusion in the high-noise regime.

[3]We omit a comparison of the alternative relaying strategies paired with full matrix fusion due to the high degree of similarity seen with optimally-weighted fusion.

B. Influence of Channel Parameters

Consider the superchannel noise of relay channel i, given by $s_i = p_i * q_i = p_i(1-q_i) + q_i(1-p_i)$. For a fixed value of $s_i \in [0, 0.5)$, the set that gives the same superchannel noise is

$$\mathcal{S}(s_i) = \left\{ \left(r, \frac{s_i - r}{1 - 2r} \right) \middle| \, r \in [0, s_i] \right\}. \tag{95}$$

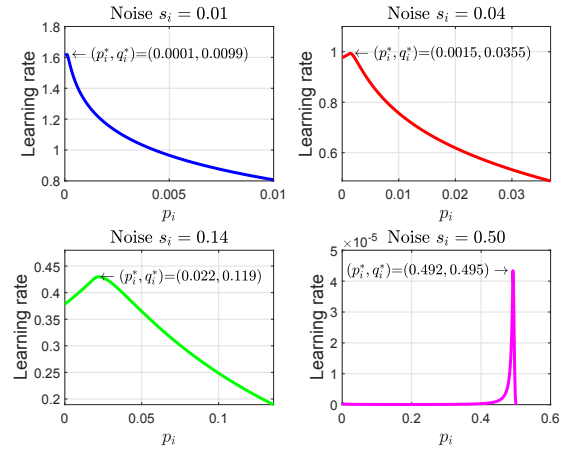
The superchannel noise s_i of relay channel i represents the effective single crossover probability between the source and destination if relay i employs simple forwarding. It allows for simple forwarding (the simplest relaying strategy) to act as a benchmark for comparing the influence of the channel parameters. Using simple forwarding, any combination of crossover probabilities belonging to $\mathcal{S}(s_i)$ will induce identical distributions of the received signal $\mathbf{v}_i$ at the destination. In contrast, any combination of crossover probabilities that is not a member of $\mathcal{S}(s_i)$ will result in a different distribution of $\mathbf{v}_i$ at the destination. Therefore, the set $\mathcal{S}(s_i)$ encompasses *all* possible crossover pairs that induce a particular distribution in $\mathbf{v}_i$ when simple forwarding is used, making it suitable for comparing the influence of the channel parameters. Given s_i and a joint relaying and destination decoding strategy, we wish to determine which allocation of noise, i.e., $(p_i^*, q_i^*) \in \mathcal{S}(s_i)$, maximizes the learning rate.

Optimal Scheme: For the optimal scheme, the noise allocation is in closed form. Recall the learning rate $\mathcal{L}_{\text{opt}}(\mathcal{P}, \mathcal{Q}) = \sum_{i=1}^{M} D_{\text{kl}}(0.5 || \max\{p_i, q_i\})$. For each $i \in [1, M]$, the optimal noise allocation is

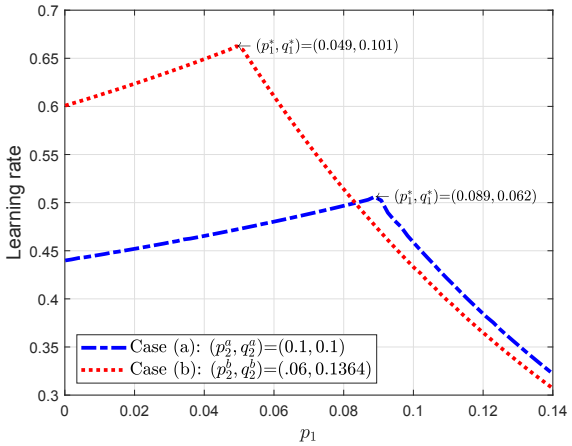
$$p_i^* = \min_{r \in [0, s_i]} \max \left\{ r, \frac{s_i - r}{1 - 2r} \right\}, \quad q_i^* = \frac{s_i - p_i^*}{1 - 2p_i^*}, \tag{96}$$

which simplifies to choosing $p_i^* = q_i^* = t_i^*$, where t_i^* satisfies $2t_i^*(1 - t_i^*) = s_i$, and $t_i^* \in [0, 0.5)$. Thus, when using the optimal scheme and given superchannel noises $\{s_i\}_{i=1}^{M}$, the learning rate is largest when the reliability of the i-th $\mathcal{S} \to \mathcal{R}$ and $\mathcal{R} \to \mathcal{D}$ links are equal. In other words, the noise should be equally distributed for a total superchannel noise.

Alternative Schemes: First, we consider optimally-weighted fusion with simple forwarding and sequential best guessing at the relays. For simple forwarding, the sub-decoder learning rates are $\ell_i^{SF}(p_i, q_i) = D_{\text{kl}}(0.5 || p_i * q_i)$, for each $i \in [1, M]$. Thus for a given superchannel noise $s_i = p_i * q_i$, the learning rate is unchanged by the specific allocation $(p_i, q_i) \in \mathcal{S}(s_i)$. With sequential best guessing, however, we observe in Fig. 6a, for a multitude of superchannel values s_i, that the learning rate of sub-decoder i is maximized when the $\mathcal{S} \to \mathcal{R}$ crossover probability p_i is smaller than the $\mathcal{R} \to \mathcal{D}$ crossover probability



(a) Sub-decoder learning rate of SBG vs. p_i.



(b) Learning rate of SBG + FMF.

Fig. 6: (a) Sub-decoder learning rate with SBG vs. p_i. Implicitly, $q_i = (s_i - p_i)/(1 - 2p_i)$. Each sub-plot indicates the crossover probabilities of the maximum learning rate. (b) Learning rate of SBG + FMF vs. p_1 of the first relay channel (for $M = 2$ relays with superchannel noises $s_1 = 0.14$ and $s_2 = 0.18$). Implicitly, $q_1 = (s_1 - p_1)/(1 - 2p_1)$. Shown for two cases of fixed channel parameters, (p_2^a, q_2^a) and (p_2^b, q_2^b), on the second relay channel.

q_i. In essence, given the superchannel value, higher reliability should be allocated to the $\mathcal{S} \to \mathcal{R}$ link.

Next, we consider full matrix fusion with the alternative relaying strategies. With simple forwarding at the relays, the learning rate with full matrix fusion is $\mathcal{L}_{SF}^{F}(\mathcal{P}, \mathcal{Q}) = \hat{I}(\mathbf{0}, \mathcal{P} * \mathcal{Q})$, which remains unchanged by the specific allocations $(p_i, q_i) \in \mathcal{S}(s_i)$. For full matrix fusion with sequential best guessing, however, the noise allocation is not definitive. Consider $M = 2$ relays and superchannel noise values $s_1 = 0.14$ and $s_2 = 0.18$. Fig. 6b shows how the learning rate changes with p_1 (and implicitly q_1) for two cases of channel parameters on the *second* relay channel, namely, $(p_2^a, q_2^a) = (0.1, 0.1)$ and $(p_2^b, q_2^b) = (0.06, 0.1364)$. For (p_2^a, q_2^a), the learning rate is maximized when the $\mathcal{R} \to \mathcal{D}$ link is more reliable than the $\mathcal{S} \to \mathcal{R}$ link on the first relay channel. For (p_2^b, q_2^b), learning rate is maximized when the $\mathcal{S} \to \mathcal{R}$ link is more reliable.

X. Conclusion

In this paper, we studied the joint design of distributed relaying and destination decoding strategies. Generalizing the teaching and learning framework, we upper bounded the learning rate of the final decoding decision at the destination and showed that an M parallel-relay generalization of the scheme in [15] achieves the optimal learning rate. To address implementation shortcomings of the asymptotically optimal scheme, we analyzed alternative strategies. Lastly, we compared the performance of all strategies, investigated the influence of the channel parameters, and explored multi-relay scaling behavior to quantify spatial diversity gains.

Appendix A
Proof of Theorem 1

For given channel parameters $\mathcal{P}$ and $\mathcal{Q}$, consider a subset given by

$$\mathcal{U}(\mathcal{P}, \mathcal{Q}) = \Big\{ (\mathbf{s}, \mathbf{r}) \in [0, 0.5)^M \times [0, 0.5)^M \,\Big|\, (s_i, r_i) \in \{(p_i, 0), (0, q_i)\}, \forall i \in [1, M] \Big\}. \tag{97}$$

We construct the subset by hard-wiring one of the $\mathcal{S} \to \mathcal{R}$ or $\mathcal{R} \to \mathcal{D}$ crossover probabilities to be zero for each relay channel $i \in [1, M]$. If $s_i = 0$ and $r_i = q_i$, the i-th $\mathcal{S} \to \mathcal{R}$ link is perfectly reliable, and relay i learns $\theta \in \{0, 1\}$ from the first channel use. Thus, relay i can perform simple forwarding, which is equivalent to transmitting the correct codeword $\mathbf{c}$ from relay i to the destination. If $s_i = p_i$ and $r_i = 0$, the i-th $\mathcal{R} \to \mathcal{D}$ link is perfectly reliable, and the signal $\mathbf{v}_i$ received at the destination equals $\mathbf{z}_i$. Thus, relay i does not need to correct bits in its received signal, and any additional processing used to form the transmitted signal $\mathbf{z}_i$ is unnecessary.

Thus for $(\mathbf{s}, \mathbf{r}) \in \mathcal{U}(\mathcal{P}, \mathcal{Q})$, the relays can perform simple forwarding with ML decoding at the destination to minimize the error probability [32]. Following this joint strategy, for any $(\mathbf{s}, \mathbf{r}) \in \mathcal{U}(\mathcal{P}, \mathcal{Q})$, the link between the source and the destination behaves as a set of M parallel and independent point-to-point BSCs, whose i-th crossover probability is $\max\{s_i, r_i\}$. In this equivalent M-output system, the learning rate is

$$\mathcal{L}^{UB}(\mathbf{s}, \mathbf{r}) = \sum_{i=1}^{M} D_{\mathrm{kl}}(0.5 || \max\{s_i, r_i\}), \tag{98}$$

which follows from the Gärtner-Ellis theorem [29]. Specifically, the Gärtner-Ellis theorem can be used to determine the learning rate of the i-th branch (namely, $D_{\mathrm{kl}}(0.5 || \max\{s_i, r_i\})$) of the equivalent M-output system. This is done by first determining the scaled cumulant generating function of a sequence of n i.i.d. Bernoulli random variables with crossover probability $\max\{s_i, r_i\}$ and taking the Legendre-Fenchel transform [29]. Since each of the M parallel branches are independent, the learning rates sum together, as in (98).[4]

[4]See the proof of Lemma 3 in Appendix C for an example of using the Gärtner-Ellis theorem on an M-output system, with each parallel branch consisting of a sequence of independent (but not necessarily identical) Bernoulli random variables.

Since any $(\mathbf{s},\mathbf{r}) \in \mathcal{U}(\mathcal{P},\mathcal{Q})$ either retains or improves the reliability of every $\mathcal{S}\to\mathcal{R}$ and $\mathcal{R}\to\mathcal{D}$ BSC characterized by channel parameters $\mathcal{P}$ and $\mathcal{Q}$, the learning rate of any joint relaying and destination decoding scheme with channel parameters $\mathcal{P}$ and $\mathcal{Q}$ satisfies

$$\mathcal{L}(\mathcal{P},\mathcal{Q}) \le \mathcal{L}^{UB}(\mathbf{s},\mathbf{r}),\ \forall(\mathbf{s},\mathbf{r}) \in \mathcal{U}(\mathcal{P},\mathcal{Q}). \tag{99}$$

Letting

$$\begin{aligned}\mathcal{L}^*(\mathcal{P},\mathcal{Q}) &= \min_{(\mathbf{s},\mathbf{r})\in\mathcal{U}(\mathcal{P},\mathcal{Q})} \mathcal{L}^{UB}(\mathbf{s},\mathbf{r}) \\ &= \sum_{i=1}^{M} D_{\mathrm{kl}}(0.5||\max\{p_i,q_i\}),\end{aligned} \tag{100}$$

then by (99), we have

$$\mathcal{L}(\mathcal{P},\mathcal{Q}) \le \mathcal{L}^*(\mathcal{P},\mathcal{Q}) = \sum_{i=1}^{M} D_{\mathrm{kl}}(0.5||\max\{p_i,q_i\}), \tag{101}$$

completing the proof.

Appendix B
Proof of Lemma 1

When the destination makes a final decoding decision with sub-decoder outcomes $\hat{\boldsymbol{\theta}}_K$, the error probability is minimized with a likelihood ratio test [32]. Since sub-decoder outcomes are independent when conditioned on θ, we have

$$\hat{\theta}_K = \begin{cases} 0, & \frac{\prod_{i=1}^{M}\mathbb{P}\left(\hat{\theta}_{i,K}\middle|\theta=0\right)}{\prod_{i=1}^{M}\mathbb{P}\left(\hat{\theta}_{i,K}\middle|\theta=1\right)} \ge 1, \\ 1, & \frac{\prod_{i=1}^{M}\mathbb{P}\left(\hat{\theta}_{i,K}\middle|\theta=0\right)}{\prod_{i=1}^{M}\mathbb{P}\left(\hat{\theta}_{i,K}\middle|\theta=1\right)} < 1. \end{cases} \tag{102}$$

Taking the logarithm of the likelihood ratios and dividing by $-n$, the test is re-expressed as

$$\hat{\theta}_K = \begin{cases} 0, & -\frac{1}{n}\sum_{i=1}^{M}\log\frac{\mathbb{P}\left(\hat{\theta}_{i,K}\middle|\,\theta=0\right)}{\mathbb{P}\left(\hat{\theta}_{i,K}\middle|\,\theta=1\right)} \le 0, \\ 1, & -\frac{1}{n}\sum_{i=1}^{M}\log\frac{\mathbb{P}\left(\hat{\theta}_{i,K}\middle|\,\theta=0\right)}{\mathbb{P}\left(\hat{\theta}_{i,K}\middle|\,\theta=1\right)} > 0. \end{cases} \tag{103}$$

Let $\{\ell_i\}_{i=1}^{M}$ be generic sub-decoder learning rates. As $n\to\infty$, the test reduces to

$$\hat{\theta}_K = \begin{cases} 0, & \sum_{i=1}^{M}(-1)^{\hat{\theta}_{i,K}}\ell_i \ge 0, \\ 1, & \sum_{i=1}^{M}(-1)^{\hat{\theta}_{i,K}}\ell_i < 0. \end{cases} \tag{104}$$

By substituting generic sub-decoder learning rates according to the relaying strategy used, the results (41) and (42) are obtained, completing the proof.

Appendix C
Proof of Lemma 3

Since the $U_{i,j}$ and $V_{i,j}$ terms are independent Bernoulli random variables, we can use the moment generating function (MGF) of the Bernoulli distribution to derive the scaled cumulant generating (SCGF) function of $\bar{W}$. The MGF of a generic Bernoulli random variable $X\in\{0,1\}$ with success probability $\omega\in[0,1]$ is $\mathbb{E}\left[e^{\lambda X}\right] = e^{\lambda}\omega + 1-\omega$. Therefore, the SGCF of $\bar{W}$ is

$$\begin{aligned}\Lambda(\lambda) &= \lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left[e^{n\lambda\bar{W}}\right] \\ &= \lim_{n\to\infty}\frac{\log\mathbb{E}\left[e^{\frac{\lambda}{M}\left\{\sum_{i=1}^{M}\left(\sum_{j=1}^{n-\lfloor\rho_i n\rfloor}U_{i,j}+\sum_{j=1}^{\lfloor\rho_i n\rfloor}V_{i,j}\right)\right\}}\right]}{n} \\ &\overset{(a)}{=} \sum_{i=1}^{M}\left(\bar{\rho}_i\log\left(e^{\frac{\lambda}{M}}\bar{q}_i+q_i\right)+\rho_i\log\left(e^{\frac{\lambda}{M}}q_i+\bar{q}_i\right)\right),\end{aligned} \tag{105}$$

where (a) follows from the MGFs of independent Bernoulli random variables and evaluating the limit. Since $\Lambda(\lambda)$ is differentiable for $\lambda\in\mathbb{R}$, then $\bar{W}$ satisfies the large deviation principle. Letting $\boldsymbol{\rho} = (\rho_1,\rho_2,\ldots,\rho_M)$, we denote the rate function of $\bar{W}$ as $I_{\boldsymbol{\rho}}(w,\mathcal{Q})$, which is determined via the Gärtner-Ellis theorem by taking the Legendre-Fenchel transform of $\Lambda(\lambda)$ [29], i.e.,

$$\begin{aligned}I_{\boldsymbol{\rho}}(w,\mathcal{Q}) &= \sup_{\lambda\in\mathbb{R}}\{\lambda w-\Lambda(\lambda)\} \\ &= \sup_{\lambda\in\mathbb{R}}\left\{\lambda w-\sum_{i=1}^{M}\left(\bar{\rho}_i\log\left(e^{\lambda/M}\bar{q}_i+q_i\right)\right.\right. \\ &\qquad \left.\left.+\rho_i\log\left(e^{\lambda/M}q_i+\bar{q}_i\right)\right)\right\}.\end{aligned} \tag{106}$$

Since $\Lambda(\lambda)$ is strictly convex with respect to $\lambda\in\mathbb{R}$, the unique supremum in (106) is achieved by differentiating $\lambda w-\Lambda(\lambda)$ with respect to λ, i.e.,

$$\begin{aligned}&\frac{\mathrm{d}\left[\lambda w-\Lambda(\lambda)\right]}{\mathrm{d}\lambda} \\ &= w-\frac{1}{M}\left[\sum_{i=1}^{M}\left(\bar{\rho}_i\frac{e^{\lambda/M}\bar{q}_i}{e^{\lambda/M}\bar{q}_i+q_i}+\rho_i\frac{e^{\lambda/M}q_i}{e^{\lambda/M}q_i+\bar{q}_i}\right)\right],\end{aligned} \tag{107}$$

setting $\mathrm{d}\left[\lambda w-\Lambda(\lambda)\right]/\mathrm{d}\lambda = 0$, and solving for λ. This completes the proof.

Appendix D
Proof of Lemma 4

The asymptotic learning rate of (92) can be expressed as

$$\begin{aligned}&\lim_{M\to\infty}\mathcal{L}_{\mathrm{opt}}(\mathcal{P}(M,p,\alpha),\mathcal{Q}(M,q,\alpha)) \\ &\overset{(a)}{=} \lim_{M\to\infty}-\frac{1}{2}\log\left[\left(1-\frac{4\left(\frac{1}{2}-\max\{p,q\}\right)^2}{M^{2\alpha}}\right)^{M^{2\alpha}}\right]^{\frac{M}{M^{2\alpha}}} \\ &\overset{(b)}{=} \lim_{M\to\infty}\frac{2\left(\frac{1}{2}-\max\{p,q\}\right)^2}{M^{2\alpha-1}},\end{aligned} \tag{108}$$

where (a) follows from algebraic manipulation, and (b) follows from applying the limit definition of the exponential function. The convergence of the limit depends on α as

$$\lim_{M\to\infty} \frac{2\left(\frac{1}{2}-\max\{p,q\}\right)^2}{M^{2\alpha-1}} = \begin{cases} +\infty, & \alpha<0.5, \\ 2\left(\frac{1}{2}-\max\{p,q\}\right)^2, & \alpha=0.5, \\ 0, & \alpha>0.5. \end{cases} \tag{109}$$

Appendix E
Proof of Lemma 5

The asymptotic learning rate of (94) can be expressed as

$$\begin{aligned} &\lim_{M\to\infty} \mathcal{L}_{SF}^{\Pi^*}\left(\mathcal{P}(M,t,\alpha),\mathcal{Q}(M,t,\alpha)\right) \\ &\overset{(a)}{=} \lim_{M\to\infty} -\frac{M}{4}\log\left(1-\frac{16\left(\frac{1}{2}-t\right)^4}{M^{4\alpha}}\right) \\ &\overset{(b)}{=} \lim_{M\to\infty} -\frac{1}{4}\log\left[\left(1-\frac{16\left(\frac{1}{2}-t\right)^4}{M^{4\alpha}}\right)^{M^{4\alpha}}\right]^{\frac{1}{M^{4\alpha-1}}} \\ &\overset{(c)}{=} \lim_{M\to\infty} \frac{4\left(\frac{1}{2}-t\right)^4}{M^{4\alpha-1}}, \end{aligned} \tag{110}$$

where parts (a)–(b) follow from algebraic manipulation. Part (c) follows from the limit definition of the exponential function. The convergence of the limit depends on α as

$$\lim_{M\to\infty} \frac{4\left(\frac{1}{2}-t\right)^4}{M^{4\alpha-1}} = \begin{cases} +\infty, & \alpha<1/4, \\ 4\left(\frac{1}{2}-t\right)^4, & \alpha=1/4, \\ 0, & \alpha>1/4. \end{cases} \tag{111}$$

References

[1] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[2] C.-Y. Chen, A. C.-S. Huang, S.-Y. Huang, and J.-Y. Chen, "Energy-Saving Scheduling in the 3GPP Narrowband Internet of Things (NB-IoT) Using Energy-Aware Machine-to-Machine Relays," in *2018 27th Wireless and Optical Communication Conference (WOCC)*, 2018, pp. 1–3, doi:10.1109/WOCC.2018.8373791.

[3] K. An and T. Liang, "Hybrid Satellite-Terrestrial Relay Networks with Adaptive Transmission," in *IEEE Transactions on Vehicular Technology*, vol. 68, no. 12, pp. 12 448–12 452, 2019, doi: 10.1109/TVT.2019.2944883.

[4] M. Bliss, F. J. Block, T. C. Royster, and D. J. Love, "Uplink NOMA for Heterogeneous NTNs with LEO Satellites and High-Altitude Platform Relays," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, 2022, pp. 172–177, doi: 10.1109/WCNC51071.2022.9771566.

[5] B. Keshavamurthy, M. A. Bliss, and N. Michelusi, "MAESTRO-X: Distributed Orchestration of Rotary-Wing UAV-Relay Swarms," in *IEEE Transactions on Cognitive Communications and Networking*, pp. 1–1, 2023, doi: 10.1109/TCCN.2023.3248859.

[6] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim, "Wireless Backhauling of 5G Small Cells: Challenges and Solution Approaches," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 22–31, 2015, doi: 10.1109/MWC.2015.7306534.

[7] Y. Zhang, D. J. Love, J. V. Krogmeier, C. R. Anderson, R. W. Heath, and D. R. Buckmaster, "Challenges and Opportunities of Future Rural Wireless Communications," *IEEE Communications Magazine*, vol. 59, no. 12, pp. 16–22, 2021, doi:10.1109/MCOM.001.2100280.

[8] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and Low-Latency Wireless Communication: Tail, Risk, and Scale," in *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, 2018, doi: 10.1109/JPROC.2018.2867029.

[9] C. Wang, D. J. Love, and D. Ogbe, "Transcoding: A New Strategy for Relay Channels," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2017, pp. 450–454, doi: 10.1109/ALLERTON.2017.8262772.

[10] D. Ogbe, C.-C. Wang, and D. J. Love, "On The Optimal Delay Growth Rate of Multi-hop Line Networks: Asymptotically Delay-Optimal Designs And The Corresponding Error Exponents," in *IEEE Transactions on Information Theory*, pp. 1–1, 2023, doi: 10.1109/TIT.2023.3283802.

[11] R. Ali, Y. B. Zikria, A. K. Bashir, S. Garg, and H. S. Kim, "URLLC for 5G and Beyond: Requirements, Enabling Incumbent Technologies and Network Intelligence," *IEEE Access*, vol. 9, pp. 67 064–67 095, 2021, doi: 10.1109/ACCESS.2021.3073806.

[12] S. Srinivasa and M. Haenggi, "Throughput-Delay-Reliability Tradeoffs in Multihop Networks with Random Access," in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2010, pp. 1117–1124, doi: 10.1109/ALLERTON.2010.5707035.

[13] V. Jog and P.-L. Loh, "Teaching and Learning in Uncertainty," in *IEEE Transactions on Information Theory*, vol. 67, no. 1, pp. 598–615, 2021, doi: 10.1109/TIT.2020.3030842.

[14] W. Huleihel, Y. Polyanskiy, and O. Shayevitz, "Relaying One Bit Across a Tandem of Binary-Symmetric Channels," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 2928–2932, doi: 10.1109/ISIT.2019.8849298.

[15] Y. H. Ling and J. Scarlett, "Optimal Rates of Teaching and Learning Under Uncertainty," in *IEEE Transactions on Information Theory*, vol. 67, no. 11, pp. 7067–7080, 2021, doi: 10.1109/TIT.2021.3107733.

[16] ——, "Multi-Bit Relaying Over a Tandem of Channels," in *IEEE Transactions on Information Theory*, vol. 69, no. 6, pp. 3511–3524, 2023, doi: 10.1109/TIT.2023.3247829.

[17] A. Paulraj, A. P. Rohit, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[18] Lizhong Zheng and Tse, D.N.C., "Diversity and Multiplexing: a Fundamental Tradeoff in Multiple-Antenna Channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, 2003, doi: 10.1109/TIT.2003.810646.

[19] J. A. Nanzer, R. L. Schmid, T. M. Comberiate, and J. E. Hodkin, "Open-Loop Coherent Distributed Arrays," in *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 5, pp. 1662–1672, 2017, doi: 10.1109/TMTT.2016.2637899.

[20] W.-C. Kuo and C.-C. Wang, "Robust and Optimal Opportunistic Scheduling for Downlink Two-Flow Network Coding With Varying Channel Quality and Rate Adaptation," in *IEEE/ACM Transactions on Networking*, vol. 25, no. 1, pp. 465–479, 2017, doi: 10.1109/TNET.2016.2583488.

[21] K. Zeng, W. Lou, and H. Zhai, "Capacity of Opportunistic Routing in Multi-Rate and Multi-Hop Wireless Networks," in *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5118–5128, 2008, doi: 10.1109/T-WC.2008.071239.

[22] F. S. Shaikh and R. Wismüller, "Routing in Multi-Hop Cellular Device-to-Device (D2D) Networks: A Survey," in *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2622–2657, 2018, doi: 10.1109/COMST.2018.2848108.

[23] D. Koutsonikolas, C.-C. Wang, and Y. C. Hu, "Efficient Network-Coding-Based Opportunistic Routing Through Cumulative Coded Acknowledgments," in *IEEE/ACM Transactions on Networking*, vol. 19, no. 5, pp. 1368–1381, 2011, doi: 10.1109/TNET.2011.2111382.

[24] W.-C. Kuo and C.-C. Wang, "Two-Flow Capacity Region of the COPE Principle for Wireless Butterfly Networks With Broadcast Erasure Channels," in *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7553–7575, 2013, doi: 10.1109/TIT.2013.2279161.

[25] R. Blum, "Distributed Detection for Diversity Reception of Fading Signals in Noise," in *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 158–164, 1999, doi: 10.1109/18.746782.

[26] J.-F. Chamberland and V. V. Veeravalli, "Wireless Sensors in Distributed Detection Applications," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 16–25, 2007, doi: 10.1109/MSP.2007.361598.

[27] X. Guo, Y. He, S. Atapattu, S. Dey, and J. S. Evans, "Power Allocation for Distributed Detection Systems in Wireless Sensor Networks With Limited Fusion Center Feedback," in *IEEE Transactions on Communications*, vol. 66, no. 10, pp. 4753–4766, 2018, doi: 10.1109/TCOMM.2018.2837101.

[28] R. O'Donnell, *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[29] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, 2011, vol. 902.

[30] Y. H. Wang, "On the Number of Successes in Independent Trials," *Statistica Sinica*, pp. 295–312, 1993.

[31] M. Xu and N. Balakrishnan, "On the Convolution of Heterogeneous Bernoulli Random Variables," *Journal of Applied Probability*, vol. 48, no. 3, pp. 877–884, 2011.
[32] S. M. Kay, *Fundamentals of Statistical Signal Processing: Practical Algorithm Development*. Pearson Education, 2013, vol. 3.

Matthew Bliss received the B.S. and M.S. degrees in electrical engineering from Purdue University, West Lafayette, IN, in 2018 and 2020, respectively, where he is currently pursuing the Ph.D. degree. His previous industrial experience includes a role as a Graduate Research Intern at the MIT Lincoln Laboratory, Lexington, MA. As a Graduate Research Assistant at Purdue University, he worked with the NSF Engineering Research Center (ERC) for the Internet of Things for Precision Agriculture (IoT4Ag). His current research interests include low-latency multi-hop wireless networks and waveform design for joint communication and sensing systems.

Chih-Chun Wang (S'01-M'05-SM'12) is a Professor of the School of Electrical and Computer Engineering of Purdue University. He received the B.E. degree in E.E. from National Taiwan University, Taipei, Taiwan in 1999, the M.S. degree in E.E., the Ph.D. degree in E.E. from Princeton University in 2002 and 2005, respectively. He worked in Comtrend Corporation, Taipei, Taiwan, as a design engineer in 2000 and spent the summer of 2004 with Flarion Technologies, New Jersey. In 2005, he held a post-doctoral researcher position in the Department of Electrical Engineering of Princeton University. He joined Purdue University in 2006, and became a Professor in 2017. He is currently a senior member of IEEE and served as an associate editor of IEEE Transactions on Information Theory during 2014 to 2017. He served as the technical co-chair of the 2017 IEEE Information Theory Workshop. His current research interests are in the latency minimization of 5G wireless networks and the corresponding protocol design, information theory for ultra-low latency communications, network coding, and cyber-physical systems. Other research interests of his fall in the general areas of networking, optimal control, information theory, detection theory, and coding theory. Dr. Wang received the National Science Foundation Faculty Early Career Development (CAREER) Award in 2009.

David J. Love (S'98 - M'05 - SM'09 - F'15) received the B.S. (with highest honors), M.S.E., and Ph.D. degrees in electrical engineering from the University of Texas at Austin in 2000, 2002, and 2004, respectively. Since 2004, he has been with the Elmore Family School of Electrical and Computer Engineering at Purdue University, where he is now the Nick Trbovich Professor of Electrical and Computer Engineering. He served as a Senior Editor for IEEE Signal Processing Magazine, Editor for the IEEE Transactions on Communications, Associate Editor for the IEEE Transactions on Signal Processing, and guest editor for special issues of the IEEE Journal on Selected Areas in Communications and the EURASIP Journal on Wireless Communications and Networking. He was a member of the Executive Committee for the National Spectrum Consortium. He holds 32 issued U.S. patents. His research interests are in the design and analysis of broadband wireless communication systems, beyond-5G wireless systems, multiple-input multiple-output (MIMO) communications, millimeter wave wireless, software defined radios and wireless networks, coding theory, and MIMO array processing.

Dr. Love is a Fellow of the American Association for the Advancement of Science (AAAS) and was named a Thomson Reuters Highly Cited Researcher (2014 and 2015). Along with his co-authors, he won best paper awards from the IEEE Communications Society (2016 Stephen O. Rice Prize and 2020 Fred W. Ellersick Prize), the IEEE Signal Processing Society (2015 IEEE Signal Processing Society Best Paper Award), and the IEEE Vehicular Technology Society (2010 Jack Neubauer Memorial Award).