Multi-hop Cooperative Wireless Networks: Diversity Multiplexing Tradeoff and Optimal Code Design

K. Sreeram
Department of ECE
Indian Institute of Science
Bangalore, India
Email: sreeramkannan@ece.iisc.ernet.in

S. Birenjith
Department of ECE
Indian Institute of Science
Bangalore, India
Email: biren@ece.iisc.ernet.in

P. Vijay Kumar
Department of ECE
Indian Institute of Science
Bangalore, India
Email: vijay@ece.iisc.ernet.in

***Abstract*— In this paper, we consider single-source, single-sink (ss-ss) multi-hop relay networks, with slow-fading Rayleigh links and single-antenna relay nodes operating under the half-duplex constraint. We present protocols and codes to achieve the optimal diversity-multiplexing tradeoff (DMT) of two classes of networks. Networks belonging to the first class can be viewed as the union of K node-disjoint parallel paths, each of length > 1, labeled here as KPP networks. The results are extended to variants including KPP(I) networks which permit causal interference between paths and KPP(D) networks which posses a direct link from source to sink. The second class is comprised of layered networks in which each layer is fully connected.**

We also draw some results for more general networks. For an arbitrary network with multiple flows, we show that the maximum achievable diversity gain for each flow is equal to the corresponding min-cut and present a simple amplify-and-forward (AF) scheme for achieving the same. For arbitrary ss-ss directed networks with full-duplex relays, we prove that a linear tradeoff between maximum diversity and maximum multiplexing gain is achievable using an AF protocol.

Explicit codes with short block-lengths based on cyclic division algebras are given for all the proposed protocols.

Two key implications of the results in the paper are that the half-duplex constraint does not necessarily entail rate loss by a factor of two as previously believed and that simple AF protocols are often sufficient to attain the best possible DMT.

I. Introduction

The diversity-multiplexing gain tradeoff (DMT) has been proposed as a tool to study cooperative relay networks, because it is simple enough to be analytically tractable and powerful enough to compare different protocols [10], [5], [1], [6]. In this paper, we study multi-hop slow-fading relay networks from a DMT viewpoint.

A. Channel Model

All nodes in the network are assumed to have a single antenna, unless otherwise mentioned. We adopt a discrete-time, complex-baseband viewpoint of each channel in the network.

1) The time duration required to communicate a message is short enough to invoke the quasi-static assumption.
2) All channels are assumed to experience Rayleigh fading and hence all fade coefficients are i.i.d., circularly-symmetric complex Gaussian $\mathbb{CN}(0,1)$ random variables.
3) The additive noise at each receiver is also modeled as possessing an i.i.d., circularly-symmetric complex Gaussian $\mathbb{CN}(0,1)$ distribution.
4) Each receiver (but not at the transmitter)is assumed to have perfect channel state information of all the upstream channels in the network. [1]

An AF protocol $\wp$ i.e., a protocol $\wp$ in which each node in the network operates in an amplify-and-forward fashion, induces the following linear channel model between source and sink:

$$\underline{y} \ = \ H(\wp)\underline{x} + \underline{w} \ , \tag{1}$$

where $\underline{y} \in \mathbb{C}^m$ denotes the signal received at the sink, $\underline{w}$ is the noise vector, $H(\wp)$ is the $(m \times n)$ induced channel matrix and $\underline{x} \in \mathbb{C}^n$ is the vector transmitted by the source. The components of the n-tuple $\underline{x}$ are the n symbols transmitted by the source and similarly, the components of the m-tuple $\underline{y}$ represent the symbols received at the sink. Typically m equals n. We impose the following energy constraint on the transmitted vector $\underline{x}$

$$\text{Tr}(\Sigma_x) \ := \ \text{Tr}(\mathbb{E}\{\underline{x}\underline{x}^\dagger\}) \quad \leq \quad n\rho$$

and we will regard ρ as representing the SNR on the network where Tr denote the trace operator. We will assume a symmetric power constraint on the relays and the source. However the exact power constraint is immaterial in the scale of interest. We consider both half and full-duplex operation at the relay nodes.

B. Diversity Multiplexing Gain Tradeoff

Let R denote the rate of communication across the network in bits per network use. Let $\wp$ denote the protocol used across the network. Let r denote the multiplexing gain associated to rate R defined by

$$R \quad = \quad r\log(\rho).$$

The probability of outage for the network operating under protocol $\wp$, i.e., the probability of the induced channel in (1)

[1]However, for the protocols proposed in this paper, the CSIR is utilized only at the sink, since all the relay nodes are required to simply amplify and forward the received signal.

is then given by

$$P_{\text{out}}(\wp, R) = \inf_{\Sigma_x \geq 0, \ \text{Tr}(\Sigma_x) \leq n\rho} \Pr(I(\underline{x}; \underline{y}) | H(\wp)) \leq nR).$$

Let the outage exponent $d_{\text{out}}(\wp, r)$ be defined by

$$d_{\text{out}}(\wp, r) = -\lim_{\rho \to \infty} \frac{P_{\text{out}}(\wp, R)}{\log(\rho)}$$

and we will indicate this by writing

$$\rho^{-d_{\text{out}}(\wp,r)} \doteq P_{\text{out}}(\wp, R).$$

The symbols $\dot{\geq}$, $\dot{\leq}$ are similarly defined.

The outage $d_{\text{out}}(r)$ of the network associated to multiplexing gain r is then defined as the supremum of the outages taken over all possible protocols, i.e.,

$$d_{\text{out}}(r) = \sup_{\wp} d_{\text{out}}(\wp, r).$$

A distributed space-time code (more simply a code) operating under a protocol $\wp$ is said to achieve a diversity gain $d(\wp, r)$ if

$$P_e(\wp, \rho) \doteq \rho^{-d(\wp,r)} ,$$

where $P_e(\rho)$ is the average error probability of the code $C(\rho)$ under maximum likelihood decoding. Using Fano's inequality, it can be shown (see [3]) that for a given protocol,

$$d(\wp, r) \leq d_{\text{out}}(\wp, r).$$

We will refer to the outage exponent $d_{\text{out}}(r)$ as the DMT $d(r)$ of the corresponding channel since for every protocol discussed in this paper we shall identify a corresponding optimal coding strategy in Section IV-E.

For each of the protocols described in this paper, we can get an upper bound on the DMT, based on the cut-set upper bound on mutual information [15]. This was formalized in [10] as follows:

Lemma 1.1: Given a cut $\mathcal{C}_i, i = 1, 2, .., M$ between any source and destination, let $r^{(\mathcal{C}_i)} \log(\rho)$ be the rate of information flow across the cut. Given a cut, there is a H matrix connecting the input terminals of the cut to the output terminals. Let us call the DMT corresponding to this H matrix as the DMT of the cut, $d_{\mathcal{C}_i}(r^{(\mathcal{C}_i)})$. Then the DMT between the source and the destination is upper bounded by

$$d(r) \leq \min_i \{d_{\mathcal{C}_i}(r^{(\mathcal{C}_i)})\}.$$

Definition 1: Given a random matrix H of size $m \times n$, we define the *DMT of the matrix* as the DMT of the associated channel $\mathbf{Y} = \mathbf{HX} + \mathbf{W}$ where $\mathbf{Y}$ is a m length received column vector, $\mathbf{X}$ is a n length transmitted column vector and $\mathbf{W}$ is a $\mathcal{CN}(0, I)$ column vector.

II. Taxonomy of Multi-hop Networks

Any wireless network can be associated with a directed graph, with vertices representing nodes in the network and edges representing connectivity between nodes. If an edge is bidirectional, we will represent it by two edges one pointing in either direction. An edge in a directed graph is said to be *live* at a particular time instant if the node at the head of the edge is transmitting at that instant. An edge in a directed graph is said to be *active* at a particular time instant if the node at the head of the edge is transmitting and the tail of the edge is receiving at that instant.

A wireless network is characterized by broadcast and interference constraints. Under the *broadcast* constraint, all edges connected to a transmitting node are simultaneously live and transmit the same information. Under the *interference* constraint, the symbol received by a receiving end is equal to the sum of the symbols transmitted on all incoming live edges. We say a protocol avoids interference if only one incoming edge is live for all receiving nodes.

In wireless networks, the relay nodes operate in either half or full-duplex mode. In case of half duplex operation, a node cannot simultaneously listen and transmit, i.e., an incoming edge and an outgoing edge of a node cannot be simultaneously active.

Definition 2: A set of edges $(v_1, v_2), (v_2, v_3), \ldots, (v_{n-1}, v_n)$ connecting the vertices v_1 to v_n is called a path. The length of a path is the number of edges in the path. The K-parallel path (KPP) network is defined as a network containing a single source and a single sink, that can be expressed as the union of K vertex-disjoint paths, each of length greater than one, connecting the source to the sink. Each of the node-disjoint paths is called a relaying path. All edges in a KPP network are bidirectional.

Definition 3: Consider a KPP network. Let v_1, v_2, v_3, v_4 be four consecutive vertices lying on one of the K paths leading from source to sink. Let v_1 and v_3 transmit, thereby causing the edges (v_1, v_2) and (v_3, v_4) to be active. Due to the broadcast and interference constraints, transmission from v_3 interferes with the reception at v_2. This is termed as back-flow.

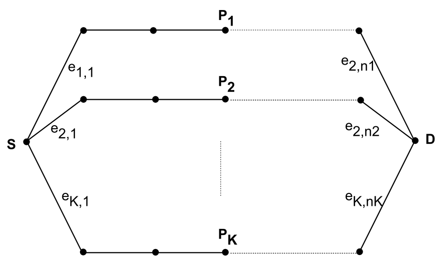


Fig. 1. The KPP network

The Definition 2 of KPP networks precludes the possibility of either having a direct link between the source and the destination, or of the existence of links connecting nodes lying on distinct node-disjoint paths. We now expand the definition of KPP networks to include both possibilities.

Definition 4: If a network graph is a union of a KPP network and an edge between the source vertex and sink vertex, then it is called a KPP network with direct link, denoted by KPP(D). If a given network graph is a union of a KPP network and edges interconnecting the vertex-disjoint paths of the KPP network, then the network is called a KPP network

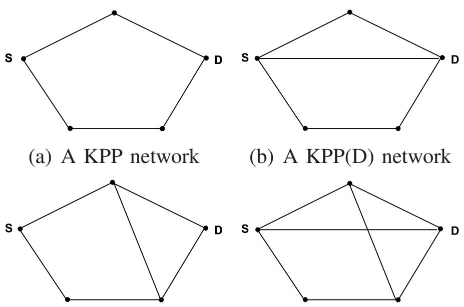
TABLE I

PRINCIPAL RESULTS SUMMARY

Network	No of sources/ sinks	No of antennas in nodes	Full/half duplex	Direct Link present?	Upper bound on Diversity/DMT $d_{\text{bound}}(r)$	Achievable Diversity/DMT $d_{\text{achieved}}(r)$	Is upper bound achieved?	Reference
Arbitrary	Multiple	Multiple	FD/HD	Either	$d(0) = M$	$d(0) = M$	✓ (d_{max} achieved)	Theorem 4.1
Arbitrary Directed Acyclic Networks	Single	Single	FD	Either	Concave in general	$M(1-r)^+$	A linear DMT between d_{max} and r_{max} is achieved	Theorem 4.3
KPP(D)(K $\geq$ 3)	Single	Single	HD	✓	$(K+1)(1-r)^+$	$(K+1)(1-r)^+$	✓	Theorem 4.7
(K, L) Regular	Single	Single	HD	×	$K(1-r)^+$	$K(1-r)^+$	✓	Theorem 4.10
KPP(I)(K $\geq$ 3)	Single	Single	HD	×	$K(1-r)^+$	$K(1-r)^+$	✓	Theorem 4.7
Fully-Connected Layered	Single	Single	HD	×	Concave in general	$M(1-r)^+$	A linear DMT between d_{max} and r_{max} is achieved	Theorem 4.16
Layered (with conditions in Lemma 4.14)	Single	Single	HD	×	Concave in general	$M(1-r)^+$	A linear DMT between d_{max} and r_{max} is achieved	Lemma 4.14

with Interference and is denoted by KPP(I). If a given network graph is a union of a KPP network, a direct edge and edges interconnecting nodes in various paths, then the network is called a KPP network with interference and direct path, and denoted by KPP(I, D).

Figure 2(a) below provides examples of all four variants of KPP networks.



(a) A KPP network (b) A KPP(D) network

(c) A KPP(I) network (d) A KPP(I, D) network

Fig. 2. Examples of KPP networks with K = 2

For a KPP(D), KPP(I) or a KPP(I, D) network, we consider the union of the K node disjoint paths as the backbone KPP network. While there may be many choices for the K node disjoint paths, we can choose any one such choice and call that the backbone KPP network. These K relaying paths in these networks are referred to as the K *backbone* paths. A *start node* and *end node* of a backbone path are the first and the last relays respectively in the path.

In a general KPP network, let $P_i, i = 1, 2, ..., K$ be the K backbone paths. Let P_i have n_i edges. The j-th edge on the i-th path P_i will be denoted by e_{ij} and the fading coefficient on that edge be denoted as g_{ij}.

Definition 5: Consider a single source single destination single antenna bidirectional network. A network is said to be a layered network if there exists a a partition of the vertex set V into subsets $V_0, V_1, ..., V_L, V_{L+1}$, such that

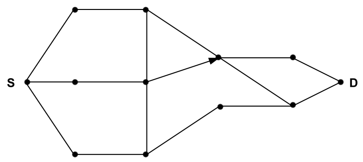
- V_0, V_{L+1} denote the singleton sets corresponding to the source and sink respectively.
- If there is an edge between a node in vertex set V_i and a node in V_j, then $|i - j| \leq 1$.

We call $V_1, ..., V_L$ as the relaying layers of the network. A layered network is said to be fully connected if for any i, $v_1 \in V_i$ and $v_2 \in V_{i+1}$, then the (v_1, v_2) is an edge in the network.

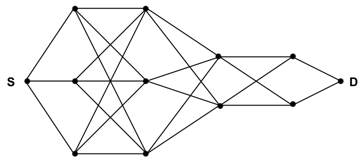
In Fig.3, examples of layered networks are given. Layered networks were also considered in [8] and [11]. In particular, [11] considered layered networks with equal number of relays on all layers. We call such layered networks as regular networks.

Definition 6: The (K, L) Regular network is defined as a KPP(I) network which is also a layered network with L layers of relays.

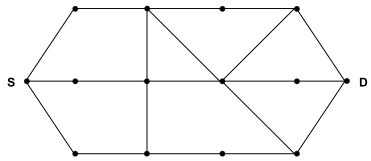
A simple example of a (K, L) Regular network is a $(K, 1)$ network which comprises of a single source, one layer of K relays and a sink node. This is the well studied two-hop network without a direct link.



(a) A Layered network with with 4 relaying layers



(b) A fully connected layered network



(c) A (3,4) regular network

Fig. 3. Examples of Layered and Regular networks

III. Lower Bounds on the DMT of Certain Derived MIMO Channels

A. White in the Scale of Interest

Lemma 3.1: Consider a channel of the form $Y = HX + Z$. Let H, $F_i, i = 1, 2, .., L$ be $n \times n$ independent random matrices, with each entry of each matrix having a finite mean and variance. Let $G_i, i = 1, 2, .., M$ comprise of finite products of various matrices from the set of F_i. Let $z = z_0 + \sum_{i=1}^{M} G_i z_i$. Let $\{z_i\}$ be i.i.d. circularly symmetric n-dimensional complex Gaussian $\mathbb{C}\mathcal{N}(\underline{0}, I)$ random vectors. Then,

1) $\lambda_i \doteq \rho^0 \quad \forall i$ with probability one, where λ_i are eigenvalues of the noise covariance matrix Σ.
2) $\log\det(I + \rho H H^\dagger \Sigma^{-1}) \doteq \log\det(I + \rho H H^\dagger)$ with probability one.
3) $Pr(\log\det(I + \rho H H^\dagger \Sigma^{-1}) \leq r \log\rho) \doteq Pr(\log\det(I + \rho H H^\dagger) \leq r \log\rho)$

Whenever a noise vector satisfies the conditions of this Lemma, we will say the noise is *white in the scale of interest*.

Lemma 3.2: [3] For any channel that is of the form $Y = HX + N$, iid gaussian inputs are sufficient to attain the best possible outage exponent of the channel, with N being white gaussian noise.

B. DMT Lower Bound for Triangular H

Definition 7: Consider a set of matrices $A_{ij}, j = 1, 2, ..., N, i \geq j$, of size $N_i \times N_j$. Let A be a matrix comprised of the block matrices A_{ij} in the (i, j)-th position, i.e.,

$$A = \begin{bmatrix} A_{11} & 0 & \dots & 0 \\ A_{21} & A_{22} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ A_{N1} & A_{N2} & \dots & A_{NN} \end{bmatrix}.$$

We will call A as a block lower-triangular matrix.

Define the l-th sub-diagonal matrix, A_l of a block lower triangular matrix A as the block lower triangular matrix comprising of entries $A_{l1}, A_{(l+1)2}, ..., A_{(l+N-1)N}$ and zeros elsewhere.

$$(A_l)_{ij} = A_{ij} \text{ if } i - j = l - 1, \text{ else } (A_l)_{ij} = 0_{N_i \times N_j}.$$

The last sub-diagonal matrix of A is defined as the sub-diagonal matrix A_l of A, with the highest l such that A_l is a *non-zero* matrix.

Theorem 3.3: Consider a block lower-triangular matrix H made of matrices H_{ij} of size $N_i \times N_j$. Let $M := \sum_{i=1}^{N} N_i$ be the size of the square matrix H. Consider a channel of the form $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{W}$, where $\mathbf{H}$ is a $M \times M$ block lower-triangular random matrix, X, Y, W are $M \times 1$ vectors. Let W be a noise vector, which is white in the scale of interest. Let X_i, Y_i, W_i be vectors of length N_i such that $X = [X_1, X_2, ..., X_N]^T$, $Y = [Y_1, Y_2, ..., Y_N]^T$ and $W = [W_1, W_2, ..., W_N]^T$. Let H_d be the block-diagonal part of the matrix H and H_ℓ denote the last sub-diagonal matrix of H. Then

1) $d_H(r) \geq d_{H_d}(r)$.
2) $d_H(r) \geq d_{H_\ell}(r)$.
3) In addition, if the entries of H_ℓ are independent of the entries in H_d, then $d_H(r) \geq d_{H_d}(r) + d_{H_\ell}(r)$

Proof: The channel is given by $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{W}$. Since the noise is white in the scale of interest, we have by Lemma 3.1 that the DMT of this channel is the same as the DMT of a channel with $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{V}$, where V is distributed as $\mathcal{CN}(0, I)$. Therefore, we can assume without loss of generality that W is distributed as $\mathcal{CN}(0, I)$.

The outage probability exponent [3] is given by

$$\rho^{-d(r)} \doteq \inf_{\Sigma_X : Tr\Sigma_X \leq \mathbb{P}} Pr\{I(\mathbf{X}; \mathbf{Y} : \mathbf{H} = H) \leq r \log\rho\}$$

In order to evaluate this exponent, we first evaluate the mutual information. Let us assume that the input X is distributed as $\mathcal{CN}(0, I)$. By Lemma 3.2, it can be seen that this input distribution is indeed DMT optimal. We will compute the mutual information terms under this assumption that the inputs are iid gaussian. Now, we proceed to find a lower bound on the DMT of the matrix. Consider the following series of inequalities for all $i = 1, ..., N$.

$$
\begin{aligned}
Z_i &:= I(\mathbf{X_i};\mathbf{Y}|\mathbf{H}=H,\mathbf{X_1^{i-1}}) \\
&\geq I(\mathbf{X_i};\mathbf{Y_i}|\mathbf{H}=H,\mathbf{X_1^{i-1}}) \\
&= I(\mathbf{X_i};\sum_{k=i-\ell}^{i}\mathbf{H_{ik}X_k}+\mathbf{W_i}|\mathbf{H}=H,\mathbf{X_1^{i-1}}) \\
&= I(\mathbf{X_i};\sum_{k=i-\ell}^{i}H_{ik}\mathbf{X_k}+\mathbf{W_i}|\mathbf{H}=H,\mathbf{X_1^{i-1}}) \\
&= I(\mathbf{X_i};H_{ii}\mathbf{X_i}+\mathbf{W_i}|\mathbf{H}=H,\mathbf{X_1^{i-1}}) \\
&= I(\mathbf{X_i};H_{ii}\mathbf{X_i}+\mathbf{W_i}|\mathbf{H}=H) \\
&= I(\mathbf{X_i};H_{ii}\mathbf{X_i}+\mathbf{W_i})
\end{aligned}
$$

$$
\begin{aligned}
I(\mathbf{X};\mathbf{Y}|\mathbf{H}=H) &= \sum_{i=1}^{M} I(\mathbf{X_i};\mathbf{Y}|\mathbf{H}=H,\mathbf{X_1^{i-1}}) \\
&\geq \sum_{i=1}^{M} I(\mathbf{X_i};H_{ii}\mathbf{X_i}+\mathbf{W_i}) \\
&= I(\mathbf{X};\mathbf{H_d}\mathbf{X}+\mathbf{W}|\mathbf{H_d}=H_d)
\end{aligned}
$$

$$
\begin{aligned}
\rho^{-d_H(r)} &= Pr\{I(\mathbf{X};\mathbf{Y}|\mathbf{H}=H)\leq r\log\rho\} \\
&\leq Pr\{I(\mathbf{X};\mathbf{H_d}\mathbf{X}+\mathbf{W}|\mathbf{H_d}=H_d)\leq r\log\rho\} \\
&= \rho^{-d_{H_d}(r)} \\
d_H(r) &\geq d_{H_d}(r) \qquad (2)
\end{aligned}
$$

We have a second series of inequalities for $i=1,...,M-1$.

$$
\begin{aligned}
Y_i &:= I(\mathbf{X_{i-\ell}};\mathbf{Y}|\mathbf{H}=H,\mathbf{X_{i-\ell+1}^N}) \\
&\geq I(\mathbf{X_{i-\ell}};\mathbf{Y_i}|\mathbf{H}=H,\mathbf{X_{i-\ell+1}^N}) \\
&= I(\mathbf{X_{i-\ell}};\sum_{k=i-\ell}^{i}\mathbf{H_{ik}X_k}+\mathbf{W_i}|\mathbf{H}=H,\mathbf{X_{i-\ell+1}^N}) \\
&= I(\mathbf{X_{i-\ell}};\sum_{k=i-\ell}^{i}H_{ik}\mathbf{X_k}+\mathbf{W_i}|\mathbf{H}=H,\mathbf{X_{i-\ell+1}^N}) \\
&= I(\mathbf{X_{i-\ell}};H_{i(i-\ell)}\mathbf{X_{i-\ell}}+\mathbf{W_i}|\mathbf{H}=H,\mathbf{X_{i-\ell+1}^N}) \\
&= I(\mathbf{X_{i-\ell}};H_{i(i-\ell)}\mathbf{X_{i-\ell}}+\mathbf{W_i}|\mathbf{H}=H) \\
&= I(\mathbf{X_{i-\ell}};H_{i(i-\ell)}\mathbf{X_{i-\ell}}+\mathbf{W_i})
\end{aligned}
$$

$$
\begin{aligned}
I(\mathbf{X};\mathbf{Y}|\mathbf{H}=H) &= \sum_{i=N}^{1} I(\mathbf{X_i};\mathbf{Y}|\mathbf{H}=H,\mathbf{X_{i+1}^N}) \\
&\geq \sum_{i=N}^{\ell+1} I(\mathbf{X_{i-\ell}};\mathbf{Y}|\mathbf{H}=H,\mathbf{X_{i-\ell+1}^N}) \\
&\geq \sum_{i=N}^{\ell+1} I(\mathbf{X_{i-\ell}};H_{i(i-\ell)}\mathbf{X_{i-\ell}}+\mathbf{W_i}) \\
&= I(\mathbf{X};\mathbf{H_\ell X}+\mathbf{W}|\mathbf{H_\ell}=H_\ell)
\end{aligned}
$$

Mapping the mutual information inequality to outage exponent, we get,

$$d_H(r) \quad \geq \quad d_{H_\ell}(r) \qquad (3)$$

Therefore,

$$
\begin{aligned}
I(\mathbf{X};\mathbf{Y}|\mathbf{H}=H) \quad \geq \quad & \max(I(\mathbf{X};\mathbf{H_d X}+\mathbf{W}|\mathbf{H_d}=H_d), \\
& I(\mathbf{X};\mathbf{H_\ell X}+\mathbf{W}|\mathbf{H_\ell}=H_\ell)) \qquad (4)
\end{aligned}
$$

Now by Eq.(4),

$$
\begin{aligned}
\rho^{-d_H(r)} &= Pr\{I(\mathbf{X};\mathbf{Y}:\mathbf{H}=H)\leq r\log\rho\} \qquad (5) \\
&\leq Pr\{I(\mathbf{X};\mathbf{H_d X}+\mathbf{W}|\mathbf{H_d}=H_d)\leq r\log\rho, \\
&\quad I(\mathbf{X};\mathbf{H_\ell X}+\mathbf{W}|\mathbf{H_\ell}=H_\ell))\leq r\log\rho\} \\
&= Pr\{I(\mathbf{X};\mathbf{H_d X}+\mathbf{W}|\mathbf{H_d}=H_d)\leq r\log\rho\} \\
&\quad \times Pr\{I(\mathbf{X};\mathbf{H_\ell X}+\mathbf{W}|\mathbf{H_\ell}=H_\ell)\leq r\log\rho\} \\
&= \rho^{-d_{H_d}(r)}\rho^{-d_{H_\ell}(r)} \\
&= \rho^{-d_{H_d}(r)+d_{H_\ell}(r)} \\
d_H(r) &\geq d_{H_d}(r)+d_{H_\ell}(r) \qquad (6)
\end{aligned}
$$

where the first step comes about because of the independence of the processes H_s and H_i, which is indeed the case because of the assumption that all the fading coefficients in the system are independent. The second step is because iid complex gaussian inputs are optimal in the scale of interest. ■

Corollary 3.4: Theorem. 3.3 holds for the case when the matrix H is block upper-triangular instead of block lower-triangular

Remark 1: Theorem 3.3 yields lower bounds on the DMT of various existing AF protocols including the NAF protocol [5], the SAF protocol [7], and the MIMO AF protocol [12]. The theorem also settles the conjecture on the DMT of a network with direct link in [13].

IV. RESULTS

A tabulation of the principal results of this paper is given in Table I.

A. *For an Arbitrary Multi-hop Network*

1) Min-Cut Equals Max Diversity:

Theorem 4.1: Consider a multi-terminal fading network with nodes having multiple antennas with each edge having iid rayleigh fading coefficients. The maximum diversity achievable for any flow is equal to the min-cut between the source and the sink corresponding to the flow. Each flow can achieve its maximum diversity simultaneously.

Proof: We will distinguish between two cases.

Case I: Network with single antenna nodes

Choose a source S_i and sink D_j. Let $\mathbb{C}_{ij}$ denote the set of all cuts between S_i and D_j.

Lemma 4.2: If a matrix H has exactly m independent Rayleigh fading coefficients in m arbitrary locations, the diversity for that matrix is exactly equal to m.

From the cut-set bound on DMT 1.1,

$$
\begin{aligned}
d(r) &\leq \min_{C\in\mathbb{C}_{ij}} d_C(r) \\
\Rightarrow d(0) &\leq \min_{C\in\mathbb{C}_{ij}} d_C(0)
\end{aligned}
$$

Given a cut C, there is a matrix corresponding to the cut H_C. If the cut has m_C edges crossing it, then this matrix has exactly m independent fading coefficients, and therefore has diversity equal to m by Lemma 4.2.

$$d(0) \leq \min_{C \in \mathbb{C}_{ij}} m_C =: m \text{ say}$$

It is now sufficient to prove that diversity order of m is achievable. Let us first consider the case when there is only one flow.

By Menger's Theorem [16], the number of edges in the min-cut is equal to the maximum number of edge disjoint paths between source and the destination. Schedule the network in such a way that each edge in a given edge disjoint path is activated one by one. Same is repeated for all the edge disjoint paths. Thus, the same data symbol is transmitted through all the edge disjoint paths from S_i to D_j. Let the number of edges in the i-th edge disjoint path be n_i. The jth edge in the the ith edge disjoint path is denoted by e_{ij} and the associated fading coefficient be h_{ij}. The activation schedule can be represented as follows: Activate each of the following edge individually in successive time instants: $e_{11}, e_{12}, \cdots, e_{1n_1}, e_{21}, \cdots, e_{2n_2}, \cdots, e_{m1}, e_{m2}, \cdots, e_{mn_m}$. Now define $h_i := \prod_{j=1}^{n_i} h_{ij}$ be the product fading coefficient on the i-th path. Let the total number of time slots required be $N = \Sigma_{i=1}^{m} n_i$.

With this protocol in place, the equivalent channel seen by a symbol is

$$H = \begin{bmatrix} h_1 & 0 & \dots & 0 \\ 0 & h_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & h_m \end{bmatrix}$$

Let u_{ij} be defined as: $\rho^{-u_{ij}} := |h_{ij}|^2$.
If $d_H(r)$ is the outage exponent for this channel,

$$\begin{aligned} \rho^{-d_H(r)} &\doteq Pr\{\Sigma_{i=1}^{m} \log(1 + |h_i|^2) \leq rlog\rho\} \\ &= Pr\{\Sigma_{i=1}^{m} \log(1 + \prod_{j=1}^{n_i} |h_{ij}|^2) \leq r \log \rho\} \\ &\doteq Pr\{\Sigma_{i=1}^{m} \log(1 + \rho^{(1-\Sigma_{j=1}^{n_i} u_{ij})}) \leq r \log \rho\} \\ &\doteq Pr\{\Sigma_{i=1}^{m} \log(1 + \rho^{(1-\Sigma_{j=1}^{n_i} u_{ij})}) \leq \log \rho^r\} \\ &\doteq Pr\{\prod_{i=1}^{m} (\rho^{(1-\Sigma_{j=1}^{n_i} u_{ij})^+}) \leq \rho^r\} \end{aligned}$$

Following the same lines of arguments as in [3],

$$d_H(r) = \inf_{\mathcal{A}} \Sigma_{i=1}^{m} \Sigma_{j=1}^{n_i} u_{ij} \tag{7}$$

where

$$\mathcal{A} = \{u_{ij} : \Sigma_{i=1}^{m} (1 - \Sigma_{j=1}^{n_i} u_{ij})^+ \leq r\} \tag{8}$$

Let $\Sigma_{j=1}^{n_i} u_{ij} = u_i$. Then,

$$\begin{aligned} d(r) &= \inf_{\mathcal{A}'} \Sigma_{i=1}^{m} u_i \\ \text{where } \mathcal{A}' &= \{u_i : \Sigma_{i=1}^{m} (1 - u_i)^+ \leq r\} \\ \Rightarrow d_H(r) &= m - r \end{aligned}$$

Since we use N channel uses, the effective outage exponent is given by,

$$\begin{aligned} d(r) &= d_H(Nr) \\ &= m - Nr \end{aligned} \tag{9}$$

Thus the maximum diversity of m can be achieved. If there are multiple flows in the network, schedule the data of all the flows in a time division manner. This will entail further rate loss - however, since we are interested only in the diversity, we can still achieve each flow's maximum diversity simultaneously.

Case II: Network with multiples antenna nodes

In the multiple antenna case, we regard any link between a n_t transmit and n_r receive antenna as being composed of $n_t n_r$ links, with one link between each transmit and each receive antenna. Note that it is possible to selectively activate precisely one of the $n_t n_r$ Tx-antenna-Rx-antenna pairs by appropriately transmitting from just one antenna and listening at just one Rx antenna.

We create a modified network from the original network by doing the following: We will replace all nodes by a super node which is comprised of small nodes on either sides. Let us say a node i has n_i antennas. Add n_i small nodes to the left side (receive side) of the node and n_i small nodes to the right (transmit side). We have thus converted a multiple antenna network into a virtual single network with super nodes and small nodes. If we evaluate the min-cut on this network, with the understanding that any cut should partition only the super-nodes, then we will get the maximum diversity on this network. The same strategy as in the single antenna case can then be used in order to achieve this diversity in the network.

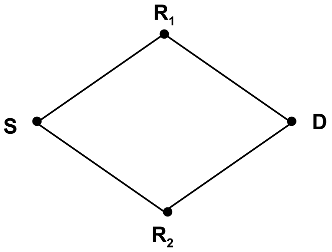
Figure 4 illustrates this conversion for the case of a single source S, two relays R_1 and R_2 and a destination D. Having thus converted the multiple antenna network into one with single antenna nodes, the theorem now follows from *Case I.* ■

2) Lower Bound to the DMT of Certain Full-Duplex Networks: In this section, all networks considered will have full-duplex relay nodes.

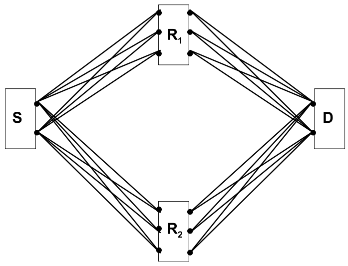
Definition 8: Consider a network N and a path P from source to destination. This path P is said to have an intermediate direct path if there is a direct link in N connecting two non-consecutive nodes in P.

Theorem 4.3: Consider a ss-ss full-duplex network with single antenna nodes. Let the min-cut of the network be M. Then a linear DMT between the maximum multiplexing gain and a maximum diversity gain, $M(1 - r)^+$, is achievable if *either* of the two conditions are satisfied:

- *None* of the M edge disjoint paths between source and destination have intermediate direct paths, or
- The directed graph representing the network has no directed cycles.



(a) Original network with multiple antenna nodes



(b) Equivalent network with single antenna nodes

Fig. 4. Illustration: $n_S = n_D = 2, n_1 = n_2 = 3$

B. DMT of KPP networks

In this section, we present AF protocols achieving the cut-set bound on DMT for KPP networks, $d(r) = K(1-r)^+$, with half-duplex constraint on relay nodes.

1) Orthogonal Protocols:

Definition 9: An AF protocol assuming half duplex operation at relay nodes is said to be an orthogonal protocol if at any node, only one of the incoming or outgoing edges is active at a given time instant. The rate, R of an orthogonal protocol is defined as the ratio of the number of symbols transmitted by the source to the total number of time slots.

We consider orthogonal protocols with periodic activation sets, i.e., there exists an integer N such that the set of edges activated at any particular time t is equal to the set of edges activated at $t + N$. Let $C = \{c_1, c_2, \ldots, c_N\}$ be a set of N colors. An edge coloring is a map $\psi : E \rightarrow \mathcal{P}_C$ which takes e_{ij} to A_{ij}. The subset of colors assigned to the edge e_{ij} will be denoted by A_{ij}. Each color in A_{ij} represents the time instants during which the edge e_{ij} is active. Every orthogonal protocol can be described as an edge coloring of the network satisfying the following constraints. Similarly, every edge coloring satisfying the following constraints describes an orthogonal protocol.

$$A_{i1} \cap A_{j1} = \phi, i \neq j. \tag{10}$$

$$A_{in_i} \cap A_{jn_j} = \phi, i \neq j. \tag{11}$$

$$A_{ij} \cap A_{ij+1} = \phi, j = 1, 2, ..., n_i - 1. \tag{12}$$

$$|A_{ij}| = m_i, j = 1, 2, ..., n_i. \tag{13}$$

Remark 2: An orthogonal protocol avoids back-flow if the corresponding coloring satisfies the following condition:

$$A_{ij} \cap A_{ij+2} = \phi, j = 1, 2, ..., n_i - 2.$$

By Remark 2, it is evident that any three adjacent edges $e_{ij}, e_{i(j+1)}$, and $e_{i(j+2)}$ will map to disjoint sets of colors when the coloring scheme corresponds to an orthogonal protocol avoiding back-flow. Moreover, it remains consistent with the constraints to repeat the same set of colors in every third edge. This suggests an easy way of describing the edge coloring. For a given path in the network, we will have three sets of colors in order and they are cyclically associated to edges starting from source to destination. For reasons that will become apparent later, the last edge (edge connected to the sink) in the given path may get associated to a different set of colors. So, to describe an orthogonal protocol, we define a tuple of sets $G_i = [G_{i0}, G_{i1}, G_{i2}]$ and a set F^i for all i such that,

$$A_{ij} = \begin{cases} G_{i(j \bmod 3)}, & j \neq n_i \\ F^i, & j = n_i \end{cases} \tag{14}$$

Hereafter, we will use G^i and F^i for $i = 1, 2, ..., K$ to completely describe an orthogonal protocol. Here, G^i specifies the colors that are repeated cyclically on the edges of the path P_i and F^i specifies the color on the last edge e_{in_i} of path P_i.

2) Protocols achieving MISO bound :

Lemma 4.4: Consider a KPP network. An orthogonal protocol satisfying the following constraints achieves the MISO bound, i.e.,

$$d(r) = K(1-r) \ , \quad 0 \leq r \ \leq \ 1 \ , \tag{15}$$

1) In every cooperation frame, the destination receives equal number of symbols from each one of the K parallel paths.
2) The protocol avoids back-flow.

Theorem 4.5: When $K \geq 3$, there exists a protocol achieving MISO bound.

Proof: (Outline) For $K \geq 4$, the orthogonal protocol achieving MISO bound is described by a coloring using the set of colors $C = \{c_1, c_2, ..., c_K\}$. Whenever we refer to color c_i assume $c_0 = c_K$ and for $i > K$, $c_i = c_{(i \bmod K)}$. The protocol is given by,

$G^i = [\{c_i\}, \{c_{i+1}\}, \{c_{i+2}\}]$

$F^i = \{c_{i+3}\}$

By Lemma 4.4, this protocol achieves the MISO bound.

The orthogonal protocol achieving the MISO bound for $K = 3$ is also of cycle length K, but it is tricky, and hence not given for the purpose of brevity. ■

Theorem 4.6: For $K = 2$ and $n_i > 1$, the maximum achievable rate for any orthogonal protocol is given by

$$R_{max} \leq \begin{cases} 1, & n_1 + n_2 = 0 \mod 2 \\ \frac{2n_2 - 1}{2n_2}, & n_1 + n_2 = 1 \mod 2 \end{cases} \tag{16}$$

where $n_1 \leq n_2$. A linear DMT between maximum diversity and R_{max} is achievable.

C. DMT of KPP(D) and KPP(I) networks

We consider KPP(D) and KPP(I) networks with relay nodes operating under half-duplex constraint.

1) Optimal DMT for KPP(D) networks:

Theorem 4.7: For KPP(D) networks with half-duplex relays, the optimal DMT $(K+1)(1-r)^+$ is achievable if the optimal DMT of the backbone KPP network is achievable by an orthogonal protocol.

2) Causal interference in KPP(I) networks: By Theorem 3.3, if the interference in a KPP(I) network is such that input-output relation can be written in the following form, then a cut-set bound DMT of $d(r) = K(1-r)^+$ is achievable.

$$\begin{bmatrix} \underline{y_1} \\ \underline{y_2} \\ \vdots \\ \underline{y_K} \end{bmatrix} = \begin{bmatrix} g_1 & & & \\ * & g_2 & & \\ * & * & \ddots & \\ * & * & * & g_K \end{bmatrix} \begin{bmatrix} \underline{x_1} \\ \underline{x_2} \\ \vdots \\ \underline{x_K} \end{bmatrix} + \underline{n} \quad (17)$$

where $g_i = \prod_{j=1}^{n_i} g_{ij}$ and $*$ denotes any entry, either zero or non-zero. This is formalized in the following Lemma.

Lemma 4.8: If the interference in a KPP(I) network, running on a particular protocol, satisfy the following two conditions, then the matrix connecting the output and a permuted version of the input will be lower triangular with K independent coefficients along its diagonal repeated periodically (except maybe the first D time instants).

For each backbone path, P_i,

1) The delay experienced by the data traveling on any other path from the first node of P_i to the sink should be no lesser than the delay on P_i from the first node to the sink.
2) The unique shortest delay from the first node of P_i to the last node of P_i is through the backbone path from the first node to the sink.

Definition 10: An orthogonal protocol for a KPP(I) network is said to have continuous activation at a relay node if the node transmits whatever it receives from the incoming edge in the last instant in the immediately next time instant.

Definition 11: An orthogonal protocol for a KPP(I) network is said to have almost continuous activation if the protocol has continuous activation at all relay nodes except possibly at the first node on each parallel path.

The conditions in Lemma 4.8 depend on the actual delays experienced by the message symbol traveling through various paths. However, the actual delays depend on the protocol used, and it becomes difficult to account for delays experienced by messages in each path. The class of protocols with almost continuous activation, which is defined above, provides us with a means to reduce the condition delays to a condition on the actual path lengths, which are simpler parameters of the network. Also, it can be proved that there exists a protocol with almost continuous activation whenever $K \geq 3$. For $K = 2$, protocols with almost continuous activation exists when paths lengths are same modulo 2.

Lemma 4.9: The optimal DMT of $d(r) = K(1-r)^+$ can be achieved in a KPP(I) network with $K \geq 3$, if the following conditions are satisfied:

For each backbone path, P_i,

1) The length of any other path from the first node of P_i to the sink should be no lesser than the length of the backbone path, P_i, from the first node to the sink.
2) The unique shortest path from the first node of P_i to the last node on that path is through the actual path from that node to the sink.

Theorem 4.10: Optimal DMT of (K, L) Regular networks is achievable.

Proof: (K, L) Regular networks satisfy the two conditions of the Lemma 4.8 with the orthogonal protocol for the backbone KPP network. ■

Remark 3: In particular Theorem 4.10 also implies that for the two hop network without direct link, the optimal DMT is achievable. This was also observed in an independent work [14]. The protocol used in this paper and in [14] is essentially the SAF protocol used in a network without direct link [7]. However, the proof techniques are different.

3) Optimal DMT for KPP(I) networks: In section IV-C.2, we gave a sufficient condition to establish when a network can be used along with a protocol with almost continuous activation in order to achieve the optimal DMT. In this section, we prove that *all* KPP(I) networks, with $K \geq 3$ can be made to achieve the optimal DMT.

Suppose the network does not meet the sufficient condition given in Lemma 4.9. It is possible that the protocol can be modified to make the network meet the sufficient condition of Lemma 4.8. We do so here by adding delays to internal nodes of the network such that, even though the path lengths do not satisfy the constraints, the delays do. By appropriately choosing a protocol and adding delays, we can make the network and the protocol jointly satisfy the conditions of Lemma 4.8. This leads us to the following Theorem.

Theorem 4.11: Consider a KPP(I) network with $K = 3$. There exists a set of delays which when added appropriately to various nodes in the networks, and when used along with the protocol with almost continuous activation, satisfies the conditions of Lemma 4.8.

Theorem 4.12: Consider a KPP(I) network with $K \geq 3$. The cut-set bound on the DMT $d(r) = K(1-r)^+$ is achievable.

Proof: For $K = 3$, it follows from Theorem 4.11.

Now, we will consider the case when $K > 3$. Consider a 3 parallel path sub-network of the original network. By Theorem 4.11, we can get a matrix with these three product coefficients along the diagonal. There are now ${}^KC_3$ possible 3PP subnetworks. If each of these subnetworks is activated in succession, it would yield a lower triangular matrix with all the K product coefficient g_i repeated thrice K choose 3 times on the diagonal. By Theorem 3.3, the DMT of this matrix is better than that of the diagonal matrix alone. The diagonal matrix has a DMT equal to $K(1-r)^+$. Hence proved. ■

D. DMT of Fully-Connected Layered Networks

In section, we present AF protocols to achieve the optimal DMT of fully connected layered networks. To begin with, we have the following Lemma to compute the DMT of the parallel channel with product coefficients occurring in some structure.

Lemma 4.13: Let $\mathcal{H} \subset \{h_{11}, h_{12}, ..., h_{1M_1}\} \times \{h_{21}, h_{22}, ..., h_{2M_2}\} \times ... \times \{h_{K1}, h_{K2}, ..., h_{KM_K}\}$ such that each h_{ij} appear in N_i of the elements in $\mathcal{H}$ irrespective of j. Let $\mathcal{H} = \{h^1, h^2, \ldots, h^N\}$. Then $N_i M_i = N$. Let $N_{\max} := \max_{i=1}^N N_i$ and $M_{\min} := \min_{i=1}^K M_i$.

Let $\psi : \mathcal{H} \to G$ be a map such that $\psi((a_1, a_2, ..., a_K)) = \Pi_{j=1}^K a_i$. Now let $g_i = \psi(h^i),\ i = 1, 2, ..., N$. Then each g_i is of the form $\Pi_{k=1}^K h_{kl(i,k)}$, where $l(i, k)$ is a map from $[N] \to [M_k]$ for a fixed $k \in [K]$.

Let H be a $N \times N$ diagonal matrix with the diagonal elements given by $H_{ii} = G_i$. Then the parallel channel H has a linear DMT between a diversity of $\frac{N}{N_{\max}}$ and a multiplexing gain of N:

$$d(r) = \frac{(N-r)^+}{N_{\max}} \qquad (18)$$

Definition 12: Let $\mathcal{P}$ denote a set of paths from the source to the sink in a layered network. The bipartite graph corresponding to $\mathcal{P}$ is defined as follows.

- Each path $P_i \in \mathcal{P}$ form a vertex on the left as well as the right of the bipartite graph.
- A vertex P_i on the left is connected to the vertex P_j on the right if the corresponding paths are node disjoint.

Lemma 4.14: Consider the set of all paths from source to sink, $\mathcal{P} = \{P_1, P_2, \ldots, P_N\}$, of a given layered network. Let the product of the fading coefficients on the i-th edge disjoint path P_i be g_i. Construct the bipartite graph corresponding to $\mathcal{P}$. If there exists a complete matching in this bipartite graph, then these paths can be activated orthogonally in such a way that the DMT of this protocol is greater than or equal to the DMT of a parallel channel with fading coefficients $g_i,\ i = 1, 2, ..., N$ with the rate reduced by a factor of N, i.e., $d(r) \geq d_{H_d}(Nr)$, where $H_d = diag(g_1, g_2, ..., g_N)$

Lemma 4.15: Let $\mathcal{P} = \{P_1, P_2, \ldots, P_N\}$ be set of all paths from the source to the sink in a fully connected layered network having L layers. Then the bipartite graph corresponding to $\mathcal{P}$ has a complete matching.

Theorem 4.16: For a fully-connected layered network, a linear DMT between maximum diversity and maximum multiplexing gain of 1 is achievable.

Proof: Consider a fully connected layered network with L relaying layers. Let there be R_i relays in the i-th layer for $i = 1, 2, .., L$. Let $R_0 := 1$ and $R_{L+1} := 1$ and $M_i = R_{i-1} R_i, i = 1, 2, ..., L+1$ be the number of fading coefficients in the cut between $(i-1)$-th and i-th layers. Let $h_{ij}, j = 1, 2, .., M_i$ be the corresponding fading coefficients. Let N be the total number of paths from source to sink, and $P_i, i \in [N]$ denote these paths. Let $\mathcal{P}$ denote the set of all paths. Then $N = \Pi_{i=1}^L R_i$. Let g_i be the product fading coefficient on path P_i. Let $M_{\min} = \min_{i=1}^{L+1} M_i$, and $N_{\max} = \max_{i=1}^{L+1} M_i$. Then $d_{\max} = M_{\min}$

By Lemma 4.15, the bipartite graph corresponding to $\mathcal{P}$ has a complete matching. Hence by Lemma 4.14 we can obtain a DMT of $d(r) \geq d_{H_d}(Nr)$. Now, we need to compute $d_{H_d}(r)$. To that effect, we make the following observations, which will enable us utilize Lemma 4.13.

A given path P_i can be alternately represented as the set of fading coefficients on that path $h^i = (h_{1l(i,1)}, h_{2l(i,2)}, ..., h_{(L+1)l(i,(L+1))})$. Consider the set of all h^i, i.e., $\mathcal{H} = \{h^i, i \in [N]\}$. Now let $g_k, k \in [N]$ be the product of fading coefficients on path P_i.

Now clearly $\mathcal{H} \subset \{h_{11}, h_{12}, ..., h_{1M_1}\} \times \{h_{21}, h_{22}, ..., h_{2M_2}\} \times ... \times \{h_{K1}, h_{K2}, ..., h_{KM_K}\}$. Also, each h_{ij} appears in the same number N_i of times irrespective of j. This means that $N_i = \frac{N}{M_i}$.

If ψ is defined as in Lemma 4.13, then $g_i = \psi(h^i)$. Now we have satisfied all the conditions of Lemma 4.13 and therefore, $d_{H_d}(r) = \frac{N-r}{N_1}$. And we get $d(r) \geq d_{\max}(1-r)^+$ ■

A sufficient condition that guarantees that a linear DMT between the maximum diversity and multiplexing gain on a general layered network is given in Lemma 4.17

Lemma 4.17: For a general layered network, a linear diversity multiplexing tradeoff of $d(r) = d_{\max}(1-r)^+$ between the maximum diversity gain $d_{\max}$ and the maximum multiplexing gain 1 is achievable whenever the bipartite graph corresponding to the set of edge disjoint paths $e_i,\ i = 1, 2, ..., d_{\max}$ from the source to the destination has a complete matching.

E. Code Design to Achieve Optimal DMT

Consider any network and protocol described above, and let us say the protocol operates for M slots. The induced channel is given by $Y = HX + W$ where X, Y, W is a $M \times 1$ vector and H is a $M \times M$ matrix. However, to design an optimal code for this channel, we need to use a space time code matrix X. In order to obtain an induced channel with X being a $M \times T$ matrix, we do the following. Instead of transmitting a single symbol, each node transmits a row vector comprising of T symbols during each activation. Then the induced channel matrix takes the form: $Y = HX + W$, with X, Y, W being $M \times T$ matrices and H the same $M \times M$ matrix as earlier.

So there are totally MT symbols transmitted. In the matrix X, let us call the row vector of T symbols in slot i as x_i. To address a specific symbol: the j-th symbol in slot i, we use the notation x_{ij}. Let us use similar notation for the output matrix Y.

Now from [4], we know that if we use an approximately universal code for X, then it will achieve the optimal DMT of the channel matrix H irrespective of the statistics of the channel. Explicit minimal delay approximately universal codes for the case when $T = M$ are given in [2], constructed based on appropriate cyclic division algebras [9].

1) Short DMT Optimal Code Design: The code construction provided above affords a code length of $TM = M^2$. Also we need M very large for the initial delay overhead to be minimal. This entails a very large block length, and indeed very high decoding complexity. Now a natural question is whether optimal DMT performance can be achieved with shorter block lengths. We answer this question for KPP networks by constructing DMT optimal codes that have $T = K$ and a block length of K^2. We also provide a DMT optimal decoding strategy that also requires only decoding a $K \times K$ matrix at a time. This is a constant which does not depend on M and therefore, even if we make M large, the delay and decoding complexity are unaffected. This code construction can be easily extended to other networks considered in this paper as well.

Let us assume that after D time instants the KPP network comes to steady state (i.e., all K paths start delivering symbols to the destination). Consider the first K inputs after attaining steady state $x_{D+1}, x_{D+2}, ..., x_{D+K}$. If the channel matrix is restricted to these K time slots alone, then channel matrix would be a lower triangular matrix with the K independent coefficients g_i, $i = 1, 2, .., K$. The DMT of this matrix alone is $d_K(r) = K(1-r)^+$. So if we use a $K \times K$ DMT optimal matrix as the input (this can be done by setting $T = K$ and using a $K \times K$ approximately universal CDA based code for the input), we will be able to obtain a DMT of $d_K(r)$ for this subset of the data. This means that the probability of error for this vector comprising of T input symbols will be of exponential order $P_e \doteq \rho^{-d_K(r)}$ if an ML decoder is used to decode the $K \times K$ matrix.

Let us assume that the first K symbols has been decoded independently. Let us now focus on the next K received symbols $y_{D+K+1}, y_{D+K+2}, ..., y_{D+K+K}$. These symbols potentially depend on the previous block of K symbols and it is optimal to decode all of these together. However we show that a Successive Interference Cancellation (SIC) based method is DMT optimal as well. After the first block of K symbols are decoded, its effect will be subtracted out from the remaining symbols, and then the next block of K symbols decoded independently. For the third block, the effect of the first two blocks each of length K will be subtracted out and the third block decoded independently and so on.

Let us evaluate the probability of error when this SIC based method is used. Let us find the probability of error for B blocks after the initial D instants of silence. Let E_i denote the event that there is an error in any of the first i blocks, F_i denote the event that there is an error in decoding the i-th block. Proceeding by induction on the i-th statement $P(E_i) = \rho^{-d_K(r)}$, we get

$$\begin{aligned}
P(F_i) &= P(F_i/E_{i-1})P(E_{i-1}) + P(F_i/\overline{E_{i-1}})P(\overline{E_{i-1}}) \\
&\leq P(E_{i-1}) + P(F_i/\overline{E_{i-1}}) \\
&\doteq \rho^{-d_K(r)} + \rho^{-d_K(r)} \\
&\doteq \rho^{-d_K(r)} \\
\Rightarrow P(E_i) &= P(\bigcup_{j=1}^{i} F_j) \leq \sum_{j=1}^{i} P(F_j) \\
&\doteq \sum_{j=1}^{i} \rho^{-d_K(r)} \doteq \rho^{-d_K(r)}
\end{aligned}$$

Therefore, we have that the entire probability of error is of the exponential order of $\rho^{-d_K(r)}$ and the scheme achieves the optimal DMT of the H matrix.

ACKNOWLEDGMENT

This work is supported in part by NSF-ITR Grant Grant CCR-0326628 and in part by the DRDO-IISc Program on Advanced Mathematical Engineering. Thanks are due to K. Vinodh and M. Anand for useful discussions.

REFERENCES

[1] P. Elia, K. Vinodh, M. Anand, and P. Vijay Kumar, "D-MG Tradeoff and Optimal Codes for a Class of AF and DF Cooperative Communication Protocols," submitted to *IEEE Trans. Inform. Theory*, Nov 2006. Available:http://arxiv.org/abs/cs/0611156

[2] P. Elia, K. Raj Kumar, S. A. Pawar, P. Vijay Kumar, and H-F. Lu, "Explicit, Minimum-Delay Space-Time Codes Achieving The Diversity-Multiplexing Gain Tradeoff," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3869–3884, Sept 2006.

[3] L. Zheng and D. Tse, "Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[4] S. Tavildar and P. Viswanath, "Approximately Universal Codes Over Slow-Fading Channels" *IEEE Trans. Inform. Theory*, vol. 52, no. 7, pp. 3233–3258, July 2006.

[5] K. Azarian, H. El Gamal, and P. Schniter, "On the Achievable Diversity-Multiplexing Tradeoff in Half-Duplex Cooperative Channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4152–4172, Dec. 2005.

[6] J. N. Laneman and G. W. Wornell, "Distributed Space-Time-Coded Protocols for Exploiting Cooperative Diversity in Wireless Networks," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[7] S. Yang and J.-C. Belfiore, "Towards the Optimal Amplify-and-Forward Cooperative Diversity Scheme," IEEE Transactions on Information Theory, vol. 53, Issue 9, pp 3114-3126, Sept. 2007.

[8] S. Yang and J.-C. Belfiore, "Diversity of MIMO Multihop Relay Channels," *submitted to IEEE Trans. on Inform. Theory*, Available Online: http://arxiv.org/abs/0708.0386

[9] B. A. Sethuraman, B. Sundar Rajan, and V. Shashidhar,"Full-Diversity, High-Rate, Space–Time Block Codes From Division Algebras," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2596–2616, Oct. 2003.

[10] M. Yuksel and E. Erkip, "Multiple-Antenna Cooperative Wireless Systems: A DiversityMultiplexing Tradeoff Perspective," *IEEE Trans. on Inform. Theory, Special Issue on Relaying and Cooperation in Communication Networks* , vol 53, no.10, pp 3371-3393, Oct.2007.

[11] S. Borade, L. Zheng and R. Gallager, "Amplify and Forward in Wireless Relay Networks: Rate, Diversity and Network Size," *IEEE Trans. on Inform. Theory, Special Issue on Relaying and Cooperation in Communication Networks*,vol 53, no.10, pp 3302-3318, Oct. 2007.

[12] S. Yang and J.-C. Belfiore, "Optimal space-time codes for the MIMO Amplify-and-Forward cooperative channel," IEEE Transactions on Information Theory, vol. 53, Issue 2, pp 647-663, Feb. 2007

[13] C. Rao and B. Hassibi, "Diversity-Multiplexing Gain Trade-Off of a MIMO System with Relays," *IEEE Information Theory Workshop*, Norway, July 2007

[14] SO. Gharan, A. Bayesteh, and AK. Khandani, "Optimum Diversity-Multiplexing Tradeoff in the Multiple Relays Network," *Available Online: http://arxiv.org/abs/0709.4506*.

[15] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, 2nd Edition, John Wiley and Sons, 2006.

[16] D. B. West, *Introduction to Graph Theory*, Prentice Hall, New Jersey, 2001.