

# Faster-than-Nyquist Signaling for MIMO Communications

Zichao Zhang, *Student Member, IEEE*, Melda Yuksel, *Senior Member, IEEE*,  
and Halim Yanikomeroglu, *Fellow, IEEE*

**Abstract**—Faster-than-Nyquist (FTN) signaling is a non-orthogonal transmission technique, which has the potential to provide significant spectral efficiency improvement. This paper studies the capacity of FTN signaling for both frequency flat and for frequency selective (FS) multiple-input multiple-output (MIMO) channels. As FTN is another reason of frequency selectivity, we find that precoding in time (or equivalently spectrum shaping in frequency) and waterfilling in spatial domain is capacity achieving for frequency flat MIMO channels with FTN. Meanwhile, waterfilling in both spatial domain and spectrum domain, followed by spectrum shaping, is capacity achieving for FS MIMO channels with FTN.

**Index Terms**—Capacity, faster-than-Nyquist (FTN), multiple-input multiple-output (MIMO).

## I. INTRODUCTION

In recent years, there has been an increasing interest in faster-than-Nyquist (FTN) signaling, as researchers are seeking for new breakthrough technologies to improve transmission rates in modern wireless communication systems. FTN signaling happens to be a promising solution for this search [1] because of the higher spectral efficiency it offers.

In classical digital communications, orthogonal pulses or Nyquist pulses are used so that there is no inter-symbol interference (ISI) at the demodulator output at sampling time instants. This maximum signaling rate above which orthogonality between consecutive symbols no longer exist is called the Nyquist limit. FTN, on the other hand, intentionally operates beyond this Nyquist limit. It introduces intentional ISI using the same pulse shape within the same bandwidth to improve spectral efficiency. In [2], Mazo studied a binary linear modulation scheme utilizing sinc pulses, where the Nyquist limit is  $\frac{1}{T}$ . He showed that it is possible to decrease the symbol interval down to  $0.802T$  with no performance degradation. This rate is called the Mazo's limit, it is the limiting rate beyond which performance starts to deteriorate for the same signal-to-noise ratio (SNR) for the same bandwidth.

A reasonable question at this point is: What is the maximum transmission rate over an FTN channel? For orthogonal

communication over additive white Gaussian noise (AWGN) channels, the capacity expression for a bandpass channel is the one obtained by Shannon [3], which states that

$$C_{orth} = \frac{1}{2} \log_2 \left( 1 + \frac{P}{WN_0} \right) \text{ bit/s/Hz}, \quad (1)$$

when  $P$  is the power constraint,  $N_0/2$  is the AWGN power spectral density,  $W = \frac{1}{2T}$  is the bandwidth, and  $\frac{1}{T}$  is the signaling rate. In the literature, there are multiple papers that seek for a similar capacity expression for a single-input single-output (SISO) FTN system. In [4], [5], the authors find an achievable rate for a SISO FTN system assuming that transmitted symbols are independent. The optimal input symbol distribution, on the other hand, is not necessarily independent, and some correlation can help reduce ISI. In [6] and [7], correlation among input symbols is explored and a power allocation scheme based on waterfilling is suggested. However, because of the influence of other symbols, the power constraint of FTN signaling differs from that of orthogonal signaling, and the power constraints in [6] or [7] do not account for the effect of ISI. In [8], [9] the authors include the impact of ISI on the power constraint and show that the optimal power spectral density (PSD) of the transmitted symbols is related to the spectrum of the pulse used for signaling. Additionally, [10] argues that for the capacity expression to be valid, the resultant FTN signal must be well-localized in time, meaning that it should contain a significant portion of its energy in the corresponding time interval. In addition to the above, precoding at the transmitter side is a viable method to mitigate FTN induced ISI [11]. Similar to [11], [12], [13] and [14] suggest an eigen-decomposition based precoding for FTN signaling.

It is well known that using multiple antennas at both transmitter and receiver sides in wireless communications can significantly improve the channel capacity [15]. Therefore, the multiple-input multiple-output (MIMO) technology has been an integral part of 3GPP standards since Release 7 [16]. In order to further improve communication rates, it is then reasonable to employ FTN signaling in MIMO systems. Rusek [17] studied the Mazo limit in MIMO communication systems. Modenini, et al. [18] demonstrated performance improvements for FTN in large-scale antenna systems with a simple receiver structure. Yuhas, et al. [19] explored MIMO FTN capacity with independent inputs as in [4], but their assumption is not general enough as the optimal input distribution is not necessarily independent. The ergodic capacity of MIMO FTN over triply-selective Rayleigh fading channels is studied in

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, NSERC, under a Discovery Grant, in part by the Scientific and Technological Research Council of Turkey, TUBITAK, under Grant 122E248, and in part by Nokia-Bell Labs, under the Corporate University Donations Program.

Z. Zhang and H. Yanikomeroglu are with the Department of Systems and Computer Engineering at Carleton University, Ottawa, ON, K1S 5B6, Canada e-mail: zichaozhang@cmail.carleton.ca, halim@sce.carleton.ca.

M. Yuksel is with the Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, 06800, Turkey, e-mail: ymelda@metu.edu.tr.

[20] with the assumption that channel state information is not available at the transmitter. However, the capacity of a MIMO system with FTN signaling with channel state information at both the transmitter and the receiver is still unknown.

Frequency selectivity is another important phenomenon in fading channels [21], and many results on frequency selective (FS) MIMO channels have appeared in the literature. In [22], the authors achieved the capacity of FS MIMO channels with joint waterfilling in spatial and frequency domains. In [23], the authors proposed an algorithm to evaluate the capacity-achieving covariance matrix for FS MIMO channels based on an iterative waterfilling scheme. In [24], a method is proposed to derive the average capacity of FS MIMO channels for arbitrary finite number of transmit and receive antennas. There are also efforts on exploring the performance of FTN in FS channels. The authors of [13] obtain the maximized mutual information for a FS SISO channel with FTN as well as its optimal input distribution. In [25], a time-space transformation approach is proposed to increase the achievable rate in FS MIMO channels with FTN. Finally, [26] suggests an FTN transmission and detection approach for FS SISO systems. However, to the best of our knowledge there is no result on the capacity of FS MIMO channels with FTN. In this paper we study the channel capacity of MIMO FTN signaling both for frequency flat and for FS fading channels.

The organization of the paper is as follows. For MIMO FTN, we define the system model in Section II. We develop the mutual information expression in Section III-A, find the related power constraint in Section III-B, and solve for the capacity expression in Section III-C. In Section IV, we generalize the results for MIMO FTN to FS channels. In Section V, we calculate the same capacity via frequency domain analysis. We present our numerical results in Section VI. Finally, in Section VII we conclude the paper.

## II. SYSTEM MODEL

In MIMO FTN, each transmitting antenna transmits symbols simultaneously using FTN signaling. We assume the transmitter has  $L$ , and the receiver has  $K$  antennas. The MIMO channel matrix is denoted with  $\tilde{\mathbf{H}} \in \mathbb{C}^{K \times L}$ . The  $(k, l)$ th entry,  $h_{kl}$ , in  $\tilde{\mathbf{H}}$  is the complex channel gain from transmit antenna  $l$  to receive antenna  $k$ ,  $l \in \{1, \dots, L\}$ ,  $k \in \{1, \dots, K\}$ . The sequence of data symbols transmitted from the  $l$ th transmitting antenna  $a_l[n]$  passes through the pulse shaping filter  $p(t)$ . We assume that the pulse shaping filter has unit energy; i.e.,  $\int_{-\infty}^{+\infty} |p(t)|^2 dt = 1$ . Moreover, the transmitter has a power constraint  $P$ . The transmitted signal from this antenna for a length of  $N$  symbols is written as

$$x_l(t) = \sum_{m=1}^N a_l[m] p(t - m\delta T), \quad (2)$$

where  $T$  denotes the signaling period, and  $\delta \in (0, 1]$  is the acceleration (compression) factor in FTN signaling. In FTN, pulses are transmitted every  $\delta T$  seconds, creating intentional inter-symbol interference.

We assume an additive white Gaussian noise (AWGN) channel and denote the noise at the  $k$ th receiving antenna

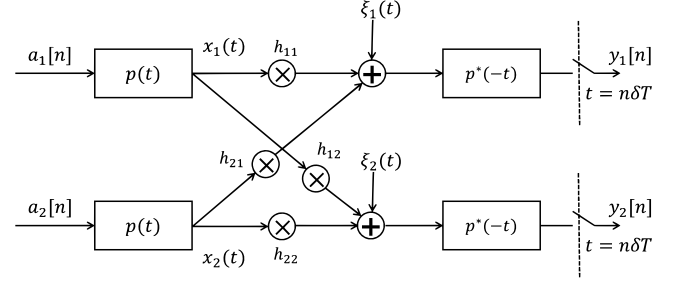


Fig. 1.  $2 \times 2$  MIMO FTN system model, where the matched filter outputs are sampled every  $\delta T$  seconds.

with  $\xi_k(t)$ ,  $k = 1, 2, \dots, K$ . This noise process has zero mean and a constant power spectral density equal to  $\sigma_0^2$ . We assume noise at each receiving antenna are independent and identically distributed and are independent of data symbols transmitted. In order to maximize the signal-to-noise ratio (SNR), matched filter  $p^*(-t)$ , where  $*$  denotes the complex conjugate, is used at the receiver side. Then, the output of the matched filter at the  $k$ th receive antenna can be written as

$$y_k(t) = \left[ \sum_{l=1}^L h_{kl} \left( \sum_{m=1}^N a_l[m] p(t - m\delta T) \right) + \xi_k(t) \right] \star p^*(-t), \quad (3)$$

$$= \sum_{l=1}^L h_{kl} \left( \sum_{m=1}^N a_l[m] g(t - m\delta T) \right) + \eta_k(t), \quad (4)$$

where  $g(t) \triangleq p(t) \star p^*(-t)$ ,  $\eta_k(t) \triangleq \xi_k(t) \star p^*(-t)$  and  $\star$  denotes the convolution operation. This matched filter output at each receive antenna  $k$  is then sampled every  $\delta T$  seconds. Fig. 1 shows the structure of a  $2 \times 2$  MIMO FTN system as a simple example.

The noise process  $\eta_k(t)$  is a correlated Gaussian process, as the white noise process  $\xi_k(t)$  is filtered through  $p^*(-t)$ . As  $\eta_k(m\delta T) = \int_{-\infty}^{\infty} \xi_k(\tau) p^*(\tau - m\delta T) d\tau$ , we have

$$\begin{aligned} \mathbb{E}[\eta_k(n\delta T) \eta_k^*(m\delta T)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{E}[\xi_k(t) \xi_k^*(\tau)] p^*(t - n\delta T) p(\tau - m\delta T) dt d\tau \\ &= \sigma_0^2 \int_{-\infty}^{\infty} p(t) p^*(t - (n - m)\delta T) dt \end{aligned} \quad (5)$$

$$= \sigma_0^2 \int_{-\infty}^{\infty} p(t) p^*(t - (n - m)\delta T) dt \quad (6)$$

$$= \sigma_0^2 g((n - m)\delta T). \quad (7)$$

Here  $\mathbb{E}$  denotes the expectation operation. After sampling, the output of the sampler becomes

$$y_k[n] = \sum_{l=1}^L h_{kl} \left( \sum_{m=1}^N a_l[m] g((n - m)\delta T) \right) + \eta_k[n]. \quad (8)$$

The input-output relationship in (8) can also be written in a matrix multiplication form as

$$\mathbf{y}_k = \sum_{l=1}^L h_{kl} \mathbf{G} \mathbf{a}_l + \boldsymbol{\eta}_k, \quad (9)$$

where  $\mathbf{y}_k \in \mathbb{C}^{N \times 1}$ ,  $\mathbf{a}_l \in \mathbb{C}^{N \times 1}$ ,  $\boldsymbol{\eta}_k \in \mathbb{C}^{N \times 1}$  and  $\mathbf{G} \in \mathbb{C}^{N \times N}$ . The  $n$ th entry of the vectors are  $(\mathbf{y}_k)_n = y_k[n]$ ,  $(\mathbf{a}_l)_n = a_l[n]$

and  $(\eta_k)_n = \eta_k[n]$ , and the  $(n, m)$ th entry of the matrix  $\mathbf{G}$  is given as  $(\mathbf{G})_{n,m} = g[n - m] \triangleq g((n - m)\delta T)$ . Note that  $\mathbf{G}$  is a Toeplitz matrix [27]. The matrix  $\mathbf{G}$  is also Hermitian, as  $g(-t) = p(-t) \star p^*(t) = g^*(t)$ , and  $(\mathbf{G})_{n,m} = g[n - m] = g^*[m - n] = (\mathbf{G})_{m,n}^*$ . From this we observe that  $\mathbf{G}$  is a Gramian matrix as well as a Toeplitz matrix. Using the cross correlation calculated in (7), the covariance matrix for  $\eta_k, k = \{1, 2, \dots, K\}$ , becomes  $\Sigma_{\eta_k} = \mathbb{E}[\eta\eta^\dagger] = \sigma_0^2 \mathbf{G}$ , where  $\dagger$  denotes Hermitian conjugation. Manipulating (9) further, we can write the MIMO FTN signaling model as

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_K \end{bmatrix} = \begin{bmatrix} h_{11}\mathbf{G} & \dots & h_{1L}\mathbf{G} \\ \vdots & \ddots & \vdots \\ h_{K1}\mathbf{G} & \dots & h_{KL}\mathbf{G} \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_L \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_K \end{bmatrix} \quad (10)$$

$$\mathbf{Y} = (\tilde{\mathbf{H}} \otimes \mathbf{G})\mathbf{A} + \Omega \quad (11)$$

$$\mathbf{Y} = \mathbf{H}\mathbf{A} + \Omega, \quad (12)$$

where  $\mathbf{y}_k \in \mathbb{C}^{N \times 1}$ ,  $\mathbf{a}_l \in \mathbb{C}^{N \times 1}$ ,  $\eta_k \in \mathbb{C}^{N \times 1}$ ,  $\mathbf{Y} = [\mathbf{y}_1^T, \dots, \mathbf{y}_K^T]^T$ ,  $\mathbf{A} = [\mathbf{a}_1^T, \dots, \mathbf{a}_L^T]^T$ ,  $\Omega = [\eta_1^T, \dots, \eta_K^T]^T$ ,  $\mathbf{H} = \tilde{\mathbf{H}} \otimes \mathbf{G}$  and  $\otimes$  is the Kronecker product.

Before we proceed with the next section, we have the following lemmas for the positive definiteness of  $\mathbf{G}$ .

*Lemma 1:* [28] As the matrix  $\mathbf{G}$  is Gramian, it is always positive semi-definite.

*Proof 1:* The proof is provided in [28].

*Lemma 2:* [29, Proposition 5.1.1] If  $g[0] > 0$  and  $g[N] \rightarrow 0$  as  $N \rightarrow \infty$  then the covariance matrix  $\Sigma_{\eta_k} = \sigma_0^2 g[i - j]_{i,j=1,\dots,N}$  of  $(\eta_k[1], \dots, \eta_k[N])^T$  is non-singular for every  $n$ , for all  $k$ .

*Proof 2:* The proof is provided in [29, Proposition 5.1.1].

If root raised cosine pulses with roll-off factor  $\beta$  are used for  $p(t)$ , then  $g(t)$  becomes a raised cosine pulse with the frequency response

$$G(f) = \begin{cases} T, & |f| \leq \frac{1-\beta}{2T} \\ \frac{T}{2} \left[ 1 + \cos\left(\frac{\pi T}{\beta} \left(|f| - \frac{1-\beta}{2T}\right)\right) \right], & \frac{1-\beta}{2T} < |f| \leq \frac{1+\beta}{2T} \\ 0, & \text{otherwise} \end{cases}$$

which satisfies Lemma 2 [5]. Thus, we also have the following lemma.

*Lemma 3:* [5], [10] For a raised cosine pulse  $g(t)$  with roll-off factor  $\beta$ , the matrix  $\mathbf{G}$  is positive definite. However, when  $\delta(1 + \beta) < 1$ , calculating  $\mathbf{G}^{-1}$  is not numerically stable.

*Proof 3:* The proofs are provided in [5], [10], [13].

### III. MIMO FTN CHANNEL CAPACITY

In this section, we find the capacity expression for MIMO FTN transmission in frequency flat channels by using the equivalent discrete time model given in (12). In the following subsection, we derive the related mutual information expression for MIMO FTN. Then, we will determine the power constraint and the capacity.

#### A. Mutual Information Expressions

Given the discrete time channel model in (12), we can write the channel capacity as [3]

$$C = \lim_{N \rightarrow \infty} \max_{p(\mathbf{A}), \mathcal{P}} \frac{1}{N} I(\mathbf{Y}; \mathbf{A}), \quad (13)$$

$$= \lim_{N \rightarrow \infty} \max_{p(\mathbf{A}), \mathcal{P}} \frac{1}{N} [h(\mathbf{Y}) - h(\Omega)], \quad (14)$$

where  $I(\mathbf{Y}; \mathbf{A})$  denotes the mutual information between two random vectors  $\mathbf{Y}$  and  $\mathbf{A}$ . The terms  $h(\mathbf{Y})$  and  $h(\Omega)$  respectively denote the differential entropy for  $\mathbf{Y}$  and  $\Omega$ , and  $\mathcal{P}$  is the power constraint, which we will explain in detail in Section III-B.

Because the circularly symmetric complex Gaussian noise process on each receiving antenna  $\xi_k(t)$  is independent of each other and also independent of input data, samples of  $\eta_i(t)$  and  $\eta_j(t)$  are also independent; i.e.,  $\mathbb{E}[\eta_i[n]\eta_j^*[k]] = 0, i \neq j, \forall k, n \in \mathbb{Z}$ . We can then write the covariance matrix and the differential entropy for  $\Omega$  as

$$\Sigma_\Omega = \sigma_0^2 (\mathbf{I}_K \otimes \mathbf{G}), \quad (15)$$

$$h(\Omega) = \log_2 \left[ (2\pi e)^{KN} (\det(\Sigma_\Omega)) \right], \quad (16)$$

where  $e$  is Euler's number,  $\mathbf{I}_K$  is the  $K$  by  $K$  identity matrix and  $\det(\cdot)$  stands for the determinant of a matrix.

*Lemma 4:* [30] Let  $\tilde{\mathbf{A}} \in \mathbb{C}^{A \times A}$  and  $\tilde{\mathbf{B}} \in \mathbb{C}^{B \times B}$  be two Hermitian matrices respectively with the eigenvalues  $\eta_1, \dots, \eta_A$  and  $\mu_1, \dots, \mu_B$ . Then the matrix  $\tilde{\mathbf{A}} \otimes \tilde{\mathbf{B}}$  has the eigenvalues  $\eta_i \mu_j$ , for  $i = 1, 2, \dots, A$  and  $j = 1, 2, \dots, B$ .

*Proof 4:* The proof is provided in [30].

*Remark 1:* The eigenvalues of the identity matrix  $\mathbf{I}_K$  are all 1s. Then, according to Lemma 4, the eigenvalues of  $\Sigma_\Omega$  are equal to the eigenvalues of  $\mathbf{G}$  repeated  $K$  times each. As a result, if  $\mathbf{G}$  is positive definite, so is  $\Sigma_\Omega$ .

When the noise process is a stationary Gaussian process, the optimal input is also a stationary Gaussian process [3]. It is also optimal to assume the input has zero mean. Then, the optimal input data sequence is a zero mean, stationary Gaussian random process. As both input and noise are Gaussian, the output  $\mathbf{Y}$  is also Gaussian with the covariance matrix  $\Sigma_Y$ , leading to the differential entropy

$$h(\mathbf{Y}) = \log_2 \left[ (2\pi e)^{KN} (\det(\Sigma_Y)) \right]. \quad (17)$$

By the definition of covariance, we can write

$$\begin{aligned} \Sigma_Y &= \mathbb{E}[(\mathbf{H}\mathbf{A} + \Omega)(\mathbf{H}\mathbf{A} + \Omega)^\dagger] \\ &= (\tilde{\mathbf{H}} \otimes \mathbf{G})\Sigma_A(\tilde{\mathbf{H}} \otimes \mathbf{G})^\dagger + \sigma_0^2 (\mathbf{I}_L \otimes \mathbf{G}), \end{aligned} \quad (18)$$

where the  $LN \times LN$  covariance matrix  $\Sigma_A$  is defined as

$$\Sigma_A = \begin{bmatrix} \mathbf{K}_{a,11} & \mathbf{K}_{a,12} & \dots & \mathbf{K}_{a,1L} \\ \mathbf{K}_{a,21} & \mathbf{K}_{a,22} & \dots & \mathbf{K}_{a,2L} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_{a,L1} & \mathbf{K}_{a,L2} & \dots & \mathbf{K}_{a,LL} \end{bmatrix}, \quad (19)$$

with  $\mathbf{K}_{a,ij} = \mathbb{E}[\mathbf{a}_i \mathbf{a}_j^\dagger]$ ,  $i, j = 1, 2, \dots, L$ . Then, the capacity expression in (14) becomes equal to

$$C = \lim_{N \rightarrow \infty} \max_{p(\mathbf{A}), \mathcal{P}} \frac{1}{N} \log_2 \frac{\det(\Sigma_Y)}{\det(\Sigma_\Omega)}. \quad (20)$$

$$C = \lim_{N \rightarrow \infty} \max_{p(\mathbf{A}), \mathcal{P}} \frac{1}{N} \log_2 \frac{\det(\mathbf{\Sigma}_Y)}{\det(\mathbf{\Sigma}_\Omega)} \quad (21)$$

$$= \lim_{N \rightarrow \infty} \max_{\mathbf{\Sigma}_A, \mathcal{P}} \frac{1}{N} \log_2 \det \left[ \left( \mathbf{\Sigma}_\Omega + (\tilde{\mathbf{H}} \otimes \mathbf{G}) \mathbf{\Sigma}_A (\tilde{\mathbf{H}} \otimes \mathbf{G})^\dagger \right) \mathbf{\Sigma}_\Omega^{-1} \right] \quad (22)$$

$$\stackrel{(a)}{=} \lim_{N \rightarrow \infty} \max_{\mathbf{\Sigma}_A, \mathcal{P}} \frac{1}{N} \log_2 \det \left[ \mathbf{I}_{LN} + (\tilde{\mathbf{H}} \otimes \mathbf{G}) \mathbf{\Sigma}_A (\tilde{\mathbf{H}}^\dagger \otimes \mathbf{I}_N) (\mathbf{I}_K \otimes \mathbf{G}^\dagger) \mathbf{\Sigma}_\Omega^{-1} \right] \quad (23)$$

$$\stackrel{(b)}{=} \lim_{N \rightarrow \infty} \max_{\mathbf{\Sigma}_A, \mathcal{P}} \frac{1}{N} \log_2 \det \left[ \mathbf{I}_{LN} + \mathbf{\Sigma}_A (\tilde{\mathbf{H}}^\dagger \otimes \mathbf{I}_N) (\mathbf{I}_K \otimes \mathbf{G}^\dagger) \mathbf{\Sigma}_\Omega^{-1} (\tilde{\mathbf{H}} \otimes \mathbf{G}) \right] \quad (24)$$

$$= \lim_{N \rightarrow \infty} \max_{\mathbf{\Sigma}_A, \mathcal{P}} \frac{1}{N} \log_2 \det \left[ \mathbf{I}_{LN} + \mathbf{\Sigma}_A (\tilde{\mathbf{H}}^\dagger \otimes \mathbf{I}_N) (\mathbf{I}_K \otimes \mathbf{G}^\dagger) (\mathbf{I}_K \otimes \mathbf{G}^{-1}) \sigma_0^{-2} (\tilde{\mathbf{H}} \otimes \mathbf{G}) \right] \quad (25)$$

$$\stackrel{(c)}{=} \lim_{N \rightarrow \infty} \max_{\mathbf{\Sigma}_A, \mathcal{P}} \frac{1}{N} \log_2 \det \left[ \mathbf{I}_{LN} + \sigma_0^{-2} \mathbf{\Sigma}_A (\tilde{\mathbf{H}}^\dagger \otimes \mathbf{I}_N) (\tilde{\mathbf{H}} \otimes \mathbf{G}) \right] \quad (26)$$

$$\stackrel{(d)}{=} \lim_{N \rightarrow \infty} \max_{\mathbf{\Sigma}_A, \mathcal{P}} \frac{1}{N} \log_2 \det \left[ \mathbf{I}_{LN} + \sigma_0^{-2} \mathbf{\Sigma}_A (\tilde{\mathbf{H}}^\dagger \otimes \mathbf{I}_N) (\tilde{\mathbf{H}} \otimes \mathbf{I}_N) (\mathbf{I}_L \otimes \mathbf{G}) \right] \quad (27)$$

$$\stackrel{(e)}{=} \lim_{N \rightarrow \infty} \max_{\mathbf{\Sigma}_A, \mathcal{P}} \frac{1}{N} \log_2 \det \left[ \mathbf{I}_{LN} + \sigma_0^{-2} \mathbf{\Sigma}_A (\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}} \otimes \mathbf{I}_N) (\mathbf{I}_L \otimes \mathbf{G}) \right] \quad (28)$$

$$\stackrel{(f)}{=} \lim_{N \rightarrow \infty} \max_{\mathbf{\Sigma}_A, \mathcal{P}} \frac{1}{N} \log_2 \det \left[ \mathbf{I}_{LN} + \sigma_0^{-2} (\mathbf{I}_L \otimes \mathbf{G}) \mathbf{\Sigma}_A (\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}} \otimes \mathbf{I}_N) \right] \quad (29)$$

The derivation of MIMO FTN capacity expression is shown in (21)-(29). The steps (a), (c), (d) and (e) are because of the mixed product property of the Kronecker product, which states that an  $A \times A$  matrix  $\mathbf{A}$  and an  $B \times B$  matrix  $\mathbf{B}$  has the Kronecker product  $(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{A} \otimes \mathbf{I}_B)(\mathbf{I}_A \otimes \mathbf{B})$  [28]. In (a), we also make use of the property  $(\tilde{\mathbf{H}} \otimes \mathbf{G})^\dagger = (\tilde{\mathbf{H}}^\dagger \otimes \mathbf{G}^\dagger)$ . The steps (b) and (f) are because of Sylvester's determinant identity [30]. Moreover, notice that in (c),  $(\mathbf{I}_K \otimes \mathbf{G}^\dagger)(\mathbf{I}_K \otimes \mathbf{G}^{-1}) = (\mathbf{I}_K \mathbf{I}_K \otimes \mathbf{G}^\dagger \mathbf{G}^{-1}) = (\mathbf{I}_K \otimes \mathbf{I}_N) = \mathbf{I}_{KN}$ . To make sure that the matrix inverse  $\mathbf{\Sigma}_\Omega^{-1} = \sigma_0^{-2}(\mathbf{I}_K \otimes \mathbf{G}^{-1})$  in (23) exists, we require  $\mathbf{G}$  to be invertible. In other words, (29) is valid only if  $\mathbf{G}$  is positive definite or all its eigenvalues are strictly positive.

and the total power  $P_x$  equals the summation of all the  $P_{x_l}$ 's, namely,

$$P_x = \sum_{l=1}^L R_{x_l}(0) = \frac{1}{\delta T} \sum_{l=1}^L \sum_{m=-\infty}^{+\infty} R_{a_l}[m]g[-m]. \quad (32)$$

If we express the total power constraint in a matrix multiplication form, we observe that the following limit is also equal to  $P_x$ , as

### B. Power Constraint

In FTN signaling, the signal transmitted from the  $l$ th antenna,  $x_l(t)$  has the following autocorrelation function  $\hat{R}_{x_l}(t, t - \tau)$

$$\begin{aligned} & \hat{R}_{x_l}(t, t - \tau) \\ &= \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} R_{a_l}[m - n] p(t - m\delta T) p^*(t - \tau - n\delta T), \end{aligned} \quad (30)$$

where  $R_{a_l}[n]$  is the autocorrelation function of the data sequence  $a_l[n]$ . One can observe that  $\hat{R}_{x_l}(t, t - \tau)$  is a cyclostationary process. For cyclostationary processes, we need to find the average of the autocorrelation function over one period and evaluate its value at 0 to find the average power. Then, the power of  $x_l(t)$  becomes

$$P_{x_l} = R_{x_l}(0) = \mathbb{E}[x_l(t)x_l^*(t)] = \frac{1}{\delta T} \sum_{m=-\infty}^{+\infty} R_{a_l}[m]g[-m], \quad (31)$$

$$\lim_{N \rightarrow \infty} \frac{1}{N\delta T} \text{tr} \left( \begin{bmatrix} \mathbf{G} & 0 & \cdots & 0 \\ 0 & \mathbf{G} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{G} \end{bmatrix} \mathbf{\Sigma}_A \right) \quad (33)$$

$$= \lim_{N \rightarrow \infty} \frac{1}{N\delta T} \text{tr} ((\mathbf{I}_L \otimes \mathbf{G}) \mathbf{\Sigma}_A) \quad (34)$$

$$= \lim_{N \rightarrow \infty} \frac{1}{\delta T} \sum_{l=0}^{L-1} \sum_{m=-(N-1)}^{N-1} R_{a_l}^*[m]g^*[-m] \frac{N - |m|}{N} \quad (35)$$

$$= \frac{1}{\delta T} \sum_{l=0}^{L-1} \sum_{m=-\infty}^{+\infty} (R_{a_l}[m]g[-m])^* \quad (36)$$

$$= P_x. \quad (37)$$

In orthogonal transmission schemes, the input power constraint is simply a constraint on the trace of  $\mathbf{\Sigma}_A$ . However, in FTN the intentional ISI generated at the transmitter changes the power constraint and we observe that the power constraint is now on the trace of the product of  $\mathbf{\Sigma}_A$  and  $(\mathbf{I}_L \otimes \mathbf{G})$ . Finally, the power constraint  $\mathcal{P}$  in (13) becomes  $P_x \leq P$ .

### C. Capacity Calculation

Combining (29) and (33), the capacity optimization problem for MIMO FTN becomes

$$C = \lim_{N \rightarrow \infty} \max_{\Sigma_A} \frac{1}{N} \log_2 \det \left( \mathbf{I}_{LN} + \sigma_0^{-2} (\mathbf{I}_L \otimes \mathbf{G}) \Sigma_A (\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}} \otimes \mathbf{I}_N) \right) \quad (38a)$$

$$s.t. \quad \lim_{N \rightarrow \infty} \frac{1}{N \delta T} \text{tr}((\mathbf{I}_L \otimes \mathbf{G}) \Sigma_A) \leq P. \quad (38b)$$

Define

$$\mathbf{K} \triangleq (\mathbf{I}_L \otimes \mathbf{G}) \Sigma_A \quad (39)$$

$$\mathbf{H}_{eq} \triangleq (\tilde{\mathbf{H}} \otimes \mathbf{I}_N) \quad (40)$$

$$\begin{aligned} \mathbf{W} &\triangleq \mathbf{H}_{eq}^\dagger \mathbf{H}_{eq} \\ &= (\tilde{\mathbf{H}} \otimes \mathbf{I}_N)^\dagger (\tilde{\mathbf{H}} \otimes \mathbf{I}_N) \\ &= (\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}}) \otimes \mathbf{I}_N. \end{aligned} \quad (41)$$

Then the problem in (38) can be written as an optimization problem about the matrix  $\mathbf{K}$  for the extended equivalent channel  $\mathbf{H}_{eq}$  as

$$C = \lim_{N \rightarrow \infty} C_N = \lim_{N \rightarrow \infty} \max_{\mathbf{K}} \frac{1}{N} \log_2 \det \left( \mathbf{I}_{LN} + \sigma_0^{-2} \mathbf{K} \mathbf{W} \right) \quad (42a)$$

$$s.t. \quad \lim_{N \rightarrow \infty} \frac{1}{N \delta T} \text{tr}(\mathbf{K}) \leq P. \quad (42b)$$

In order to solve this optimization problem, we should solve it for each  $N$ , and then take the limit as  $N \rightarrow \infty$ . For a specific  $N$ , the above optimization is equivalent to the well-known waterfilling problem for a regular MIMO channel with the channel gain matrix  $\mathbf{H}_{eq}$  [15]. Thus, we will first solve for (42) for a fixed  $N$  to obtain  $C_N$  in (42a) and then take  $N$  to infinity. If  $C_N$  converges to a constant, the limit will then be the capacity  $C$ .

To find the solution for a fixed  $N$ , we write the eigen-decomposition for the composite covariance matrix  $\mathbf{K}$  as  $\mathbf{K} = \mathbf{P}_K \mathbf{\Lambda}_K \mathbf{P}_K^{-1}$ , where  $\mathbf{P}_K$  is composed of the eigenvectors of  $\mathbf{K}$  and  $\mathbf{\Lambda}_K$  is a diagonal matrix with the eigenvalues of  $\mathbf{K}$ ,  $\alpha_1, \alpha_2, \dots, \alpha_{NL}$ , on the main diagonal. We also perform eigen-decomposition on  $\mathbf{W}$  and write  $\mathbf{W} = \mathbf{V}_W \mathbf{\Lambda}_W \mathbf{V}_W^\dagger$ . Because  $\mathbf{W}$  is Hermitian,  $\mathbf{V}_W$  is a unitary matrix and it is composed of the eigenvectors of  $\mathbf{W}$ . Moreover,  $\mathbf{\Lambda}_W$  is a diagonal matrix with the eigenvalues of  $\mathbf{W}$  on the main diagonal. We define

$$\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}} \triangleq \mathbf{Z} = \mathbf{V}_Z \mathbf{\Lambda}_Z \mathbf{V}_Z^\dagger, \quad (43)$$

where the unitary matrix  $\mathbf{V}_Z$  is composed of the eigenvectors of  $\mathbf{Z}$ . Then, for a fixed  $N$ , we can write and bound the determinant in (42a) using Hadamard's inequality as [3]

$$C_N = \max_{\mathbf{K}} \frac{1}{N} \log_2 \det \left( \mathbf{I}_{LN} + \sigma_0^{-2} \mathbf{P}_K \mathbf{\Lambda}_K \mathbf{P}_K^{-1} \mathbf{V}_W \mathbf{\Lambda}_W \mathbf{V}_W^\dagger \right) \quad (44)$$

$$\leq \max_{\mathbf{\Lambda}_K} \frac{1}{N} \log_2 \det \left( \mathbf{I}_{LN} + \sigma_0^{-2} \mathbf{\Lambda}_K \mathbf{\Lambda}_W \right). \quad (45)$$

The above inequality is achieved with equality when  $\mathbf{K} \mathbf{W}$  becomes equal to the diagonal matrix  $\mathbf{\Lambda}_K \mathbf{\Lambda}_W$ . Thus, we

choose  $\mathbf{P}_K = \mathbf{V}_W$ . Note that, as  $\mathbf{V}_W$  is unitary, this choice also makes the matrix  $\mathbf{K}$  Hermitian.

Assuming the product matrix  $\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}}$  has eigenvalues  $\tau_1, \dots, \tau_L$ , we can write the eigenvalues of  $\mathbf{W}$  in (41) due to Lemma 4. As the identity matrix has all of its eigenvalues equal to 1, the eigenvalues of  $\mathbf{W}$  repeat themselves  $N$  times and we write  $\text{eig}(\mathbf{W})_{j*N+i} = \tau_i$  for all  $j = 0, 1, \dots, L-1, i = 1, 2, \dots, N$ . Then, we can write the optimization problem in the form of a summation of logarithms as

$$C_N = \max_{\alpha_1, \alpha_2, \dots, \alpha_{NL}} \frac{1}{N} \sum_{i=1}^{NL} \log_2 (1 + \sigma_0^{-2} \alpha_i \tau_{\lceil i/N \rceil}) \quad (46a)$$

$$s.t. \quad \lim_{N \rightarrow \infty} \frac{1}{N \delta T} \sum_{i=1}^{NL} \alpha_i \leq P, \quad (46b)$$

where  $\lceil i/N \rceil$  takes the ceiling of fraction  $i/N$ . The solution to this problem is obtained via waterfilling as

$$\bar{\alpha}_i = \sigma_0^2 \left( \frac{1}{\mu} - \frac{1}{\tau_{\lceil i/N \rceil}} \right)^+, \quad (47)$$

where  $i = 1, 2, \dots, NL$ , and  $(a)^+ = \max(a, 0)$ . Then,  $\mu$  is solved by

$$\sigma_0^2 \sum_{i=1}^{NL} \left( \frac{1}{\mu} - \frac{1}{\tau_{\lceil i/N \rceil}} \right)^+ = N P \delta T, \quad (48)$$

and the optimal covariance matrix becomes

$$\bar{\mathbf{K}} = \mathbf{V}_W \bar{\mathbf{\Lambda}}_K \mathbf{V}_W^\dagger, \quad (49)$$

where  $\bar{\mathbf{\Lambda}}_K = \text{diag}\{\bar{\alpha}_1, \bar{\alpha}_2, \dots, \bar{\alpha}_{NL}\}$ . Using this optimal covariance matrix  $\bar{\mathbf{K}}$  in (49) and (39), we can find the optimal input covariance matrix  $\bar{\Sigma}_A$  as

$$\bar{\Sigma}_A = (\mathbf{I}_L \otimes \mathbf{G})^{-1} \bar{\mathbf{K}}. \quad (50)$$

At this stage, we have to guarantee that  $\bar{\Sigma}_A$  is a valid covariance matrix, so it has to be both positive semidefinite and Hermitian.

*Lemma 5:* The matrix  $\bar{\Sigma}_A$  is always Hermitian and is positive definite.

*Proof 5:* The proof is provided in Appendix A.

Overall, combining (46a), (47), (50) and Lemma 5, we write that the maximum mutual information for a specific  $N$  is equal to

$$C_N = \sum_{i=1}^L \log_2 (1 + \sigma_0^{-2} \bar{\alpha}_{(i-1)N+1} \tau_i), \quad \text{when } \delta(1 + \beta) \geq 1. \quad (51)$$

As (51) does not depend on  $N$ , we conclude that this solution applies to any  $N$  and the MIMO FTN capacity is given in Theorem 1.

*Remark 2:* Note that Lemma 5 is valid for all  $\delta$ . However, due to Lemma 3, to provide numerical stability in the calculation of  $\mathbf{G}^{-1}$ , we need  $\delta(1 + \beta) \geq 1$  in (51).

*Theorem 1:* When a raised cosine pulse with roll-off factor  $\beta$  is used as  $g(t)$ , the capacity of MIMO FTN signaling with acceleration factor  $\delta$  becomes

$$C = \sum_{i=1}^L \log_2 (1 + \sigma_0^{-2} \bar{\alpha}_{(i-1)N+1} \tau_i), \quad \text{when } \delta(1 + \beta) \geq 1 \quad (52)$$

$$C = \lim_{N \rightarrow \infty} \frac{1}{N} \log_2 \det \left[ \mathbf{I}_{LN} + \sigma_0^{-2} (\mathbf{I}_L \otimes \mathbf{G}) \left( (\mathbf{V}_Z \mathbf{\Lambda}_{\bar{\alpha}} \mathbf{V}_Z^\dagger) \otimes \mathbf{G}^{-1} \right) (\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}} \otimes \mathbf{I}_N) \right] \quad (58)$$

$$= \lim_{N \rightarrow \infty} \frac{1}{N} \log_2 \det \left[ \mathbf{I}_{LN} + \sigma_0^{-2} \left( (\mathbf{V}_Z \mathbf{\Lambda}_{\bar{\alpha}} \mathbf{V}_Z^\dagger) \otimes \mathbf{I}_N \right) (\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}} \otimes \mathbf{I}_N) \right] \quad (59)$$

$$= \lim_{N \rightarrow \infty} \frac{1}{N} \log_2 \det \left[ \mathbf{I}_L \otimes \mathbf{I}_N + \sigma_0^{-2} \left( (\mathbf{V}_Z \mathbf{\Lambda}_{\bar{\alpha}} \mathbf{V}_Z^\dagger) \mathbf{Z} \otimes \mathbf{I}_N \right) \right] \quad (60)$$

$$= \lim_{N \rightarrow \infty} \frac{1}{N} \log_2 \det \left[ (\mathbf{I}_L + \sigma_0^{-2} (\mathbf{V}_Z \mathbf{\Lambda}_{\bar{\alpha}} \mathbf{V}_Z^\dagger) \mathbf{Z}) \otimes \mathbf{I}_N \right] \quad (61)$$

$$= \lim_{N \rightarrow \infty} \frac{1}{N} N \log_2 \det \left[ \mathbf{I}_L + \sigma_0^{-2} (\mathbf{V}_Z \mathbf{\Lambda}_{\bar{\alpha}} \mathbf{\Lambda}_Z \mathbf{V}_Z^\dagger) \right] \quad (62)$$

$$= \log_2 \det (\mathbf{I}_L + \sigma_0^{-2} \mathbf{\Lambda}_{\bar{\alpha}} \mathbf{\Lambda}_Z) \quad (63)$$

in bits/channel use. As we transmit  $\frac{W}{\delta}$  symbols per second and the bandwidth is  $(1 + \beta)W$  Hz, we can rewrite (52) to obtain

$$\bar{C} = \frac{1}{\delta(1 + \beta)} \sum_{i=1}^L \log_2 (1 + \sigma_0^{-2} \bar{\alpha}_{(i-1)N+1} \tau_i), \quad \text{when } \delta(1 + \beta) \leq 1 \quad (53)$$

in bit/s/Hz.

*Remark 3:* In (52),  $\bar{\alpha}_i$  of (47) depends on  $\delta$ , because of the power constraint in (48). In other words, the acceleration factor is implicitly included in the result.

*Remark 4:* The optimal transmission scheme is equivalent to precoding in time and waterfilling in space. Precoding in time results in perfect ISI mitigation at the transmitter side.

This remark is justified as follows. First, note that in (47) the value for  $\tau_{[i/N]}$  is the same for all groups of  $i = \{1, \dots, N\}$ ,  $i = \{N+1, \dots, 2N\}$ , until  $i = \{(N-1)L+1, \dots, NL\}$ , thus there are only  $L$  distinct  $\alpha_i$  values. The same  $\bar{\alpha}_i$  values repeat themselves  $N$  times consecutively, with  $L$  groups in total in  $\bar{\mathbf{\Lambda}}_K$  in (49), and we can rewrite  $\bar{\mathbf{\Lambda}}_K$  as  $\bar{\mathbf{\Lambda}}_K = \mathbf{\Lambda}_{\bar{\alpha}} \otimes \mathbf{I}_N$ , where  $\mathbf{\Lambda}_{\bar{\alpha}} = \text{diag}\{\bar{\alpha}_1, \bar{\alpha}_{N+1}, \dots, \bar{\alpha}_{N(L-1)+1}\}$  is the diagonal matrix composed of the distinct  $\bar{\alpha}_i$  values.

As the details are shown in Appendix A in (118), we can write the optimal covariance matrix  $\bar{\mathbf{\Sigma}}_A$  in (50) as

$$\bar{\mathbf{\Sigma}}_A = (\mathbf{V}_Z \otimes \mathbf{V}_G) (\mathbf{I}_L \otimes \mathbf{\Lambda}_G^{-1}) \bar{\mathbf{\Lambda}}_K (\mathbf{V}_Z \otimes \mathbf{V}_G)^\dagger, \quad (54)$$

where  $\mathbf{V}_G \mathbf{\Lambda}_G^{-1} \mathbf{V}_G^\dagger$  is defined as the eigenvalue decomposition for  $\mathbf{G}^{-1}$ . Then, inserting  $\bar{\mathbf{\Lambda}}_K = \mathbf{\Lambda}_{\bar{\alpha}} \otimes \mathbf{I}_N$  in the above equation, we obtain

$$\bar{\mathbf{\Sigma}}_A = (\mathbf{V}_Z \otimes \mathbf{V}_G) (\mathbf{I}_L \otimes \mathbf{\Lambda}_G^{-1}) (\mathbf{\Lambda}_{\bar{\alpha}} \otimes \mathbf{I}_N) (\mathbf{V}_Z \otimes \mathbf{V}_G)^\dagger \quad (55)$$

$$= (\mathbf{V}_Z \otimes \mathbf{V}_G) (\mathbf{\Lambda}_{\bar{\alpha}} \otimes \mathbf{\Lambda}_G^{-1}) (\mathbf{V}_Z \otimes \mathbf{V}_G)^\dagger \quad (56)$$

$$= (\mathbf{V}_Z \mathbf{\Lambda}_{\bar{\alpha}} \mathbf{V}_Z^\dagger) \otimes \mathbf{G}^{-1}, \quad (57)$$

as a valid covariance matrix due to Lemma 5. Next, we insert the optimal input covariance matrix (57) into the MIMO FTN capacity expression (38a) to obtain (58). Using the properties for the Kronecker product, we can simplify (58) as in (59)-(62) to obtain the final form for the MIMO FTN capacity expression as in (63).

This result in (63) clearly shows that the best achievable scheme is waterfilling for the MIMO channel. The best  $\bar{\alpha}_i$ 's, are solely determined by the MIMO channel  $\tilde{\mathbf{H}}$  in (47).

Comparing (58) with (59), we also observe that the effect of ISI is completely mitigated, the original input data  $\mathbf{A}$  is precoded for ISI removal via  $\mathbf{G}^{-1}$  in (57), and the matrix  $\mathbf{G}$  no longer appears in (59)-(63). We also express that in MIMO FTN the symbols transmitted from each antenna have the same autocorrelation in time but only differ in their power levels, determined according to the MIMO channel  $\tilde{\mathbf{H}}$ .

*Remark 5:* In addition to the numerical instability of  $\mathbf{G}^{-1}$  mentioned in Lemma 3, there will be other hurdles when implementing FTN in practice. For example, practically ideal root raised cosine pulses are not possible, and truncation in time is performed. This truncation generally broadens the spectrum [31] more than it does in orthogonal signaling. Therefore, overall spectral efficiency calculations in FTN should take spectrum broadening into account as well.

Finally, note that a complete analysis which accounts for  $\delta(1 + \beta) < 1$ , is a very complex problem. In [32], the authors prove that binary signaling is asymptotically optimal and achieves the capacity of FTN signaling with Gaussian inputs. In addition [9], claims a constant capacity for  $\delta(1 + \beta) < 1$ , which is equal to the capacity achieved for  $\delta = \frac{1}{1+\beta}$  for SISO FTN. This paper is a work in progress, and this issue of  $\delta(1 + \beta) < 1$  is one of our top priorities for future work.

#### IV. CAPACITY OF MIMO FTN SIGNALING IN FREQUENCY SELECTIVE CHANNELS

FS fading occurs when the channel spectrum is not constant over the transmission bandwidth [21]. Frequency selectivity distorts the transmitted signal. It presents itself in time domain as multipath. Signal arrives at the receiver from different paths, each with a different delay and gain. This effect can be modeled by a tapped-delay-filter with  $J$  taps. In this section we solve the capacity of MIMO FTN signaling in FS channels.

Similar to the SISO model in [13], in MIMO, the signal component  $x_{kl}(t)$  at the output of the matched filter at the  $k$ th receive antenna, due to the  $l$ th transmit antenna has the expression

$$x_{kl}(t) = \sum_{n=0}^{N-1} \sum_{j=0}^{J-1} h_{kl}^j a_l[n] g(t - (j + n)\delta T), \quad (64)$$

where  $h_{kl}^j$  is the channel gain from the  $l$ th transmit antenna to the  $k$ th receive antenna for the  $j$ th tap. Note that multiple taps result in ISI, in addition to the ISI due to FTN signaling. We

$$\mathbf{x}_{kl} = \begin{bmatrix} x_{kl}[0] \\ x_{kl}[1] \\ \vdots \\ x_{kl}[N-1] \end{bmatrix} = \begin{bmatrix} \sum_{j=0}^{J-1} h_{kl}^j g[-j] & \sum_{j=0}^{J-1} h_{kl}^j g[-1-j] & \cdots & \sum_{j=0}^{J-1} h_{kl}^j g[1-N-j] \\ \sum_{j=0}^{J-1} h_{kl}^j g[1-j] & \sum_{j=0}^{J-1} h_{kl}^j g[-j] & \cdots & \sum_{j=0}^{J-1} h_{kl}^j g[2-N-j] \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=0}^{J-1} h_{kl}^j g[N-1-j] & \sum_{j=0}^{J-1} h_{kl}^j g[N-2-j] & \cdots & \sum_{j=0}^{J-1} h_{kl}^j g[-j] \end{bmatrix} \begin{bmatrix} a_l[0] \\ a_l[1] \\ \vdots \\ a_l[N-1] \end{bmatrix} \quad (65)$$

would like to express that, every channel from one transmit antenna to a receive antenna does not need to have equal number of taps. The number of taps  $J$  is simply the maximum number of taps in all channels. For those channels with the number of taps less than  $J$ , we can assume the remaining channel coefficients are zero, and every channel has  $J$  taps. If we sample this signal at  $\delta T$  intervals, as in [13], the resulting  $N$  samples can be written in vector form as (65). We can also write  $\mathbf{x}_{kl}$  as

$$\mathbf{x}_{kl} = \tilde{\mathbf{G}}_{kl} \mathbf{a}_l. \quad (66)$$

When we define  $\mathbf{G}^j$ , which is a  $j$ -shifted version of  $\mathbf{G}$ , as

$$\mathbf{G}^j \triangleq \begin{bmatrix} g[-j] & g[-1-j] & \cdots & g[1-N-j] \\ g[1-j] & g[-j] & \cdots & g[2-N-j] \\ \vdots & \vdots & \ddots & \vdots \\ g[N-1-j] & g[N-2-j] & \cdots & g[-j] \end{bmatrix}, \quad (67)$$

with  $\mathbf{G}^0 = \mathbf{G}$ , we have  $\tilde{\mathbf{G}}_{kl}$  in (66) as

$$\tilde{\mathbf{G}}_{kl} = \sum_{j=0}^{J-1} h_{kl}^j \mathbf{G}^j. \quad (68)$$

Thus, the samples at the output of the matched filters at all receive antennas become

$$\mathbf{Y} = \begin{bmatrix} \tilde{\mathbf{G}}_{11} & \tilde{\mathbf{G}}_{12} & \cdots & \tilde{\mathbf{G}}_{1L} \\ \tilde{\mathbf{G}}_{21} & \tilde{\mathbf{G}}_{22} & \cdots & \tilde{\mathbf{G}}_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{G}}_{K1} & \tilde{\mathbf{G}}_{K2} & \cdots & \tilde{\mathbf{G}}_{KL} \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_L \end{bmatrix} + \mathbf{\Omega} \quad (69)$$

$$= \left[ \sum_{j=0}^{J-1} (\tilde{\mathbf{H}}^j \otimes \mathbf{G}^j) \right] \mathbf{A} + \mathbf{\Omega}, \quad (70)$$

$$\triangleq \mathbf{G}_H \mathbf{A} + \mathbf{\Omega} \quad (71)$$

where  $(\tilde{\mathbf{H}}^j)_{k,l} = h_{kl}^j$ . As  $\sum_{j=0}^{J-1} (\tilde{\mathbf{H}}^j \otimes \mathbf{G}^j) = \mathbf{G}_H$  in (71), we derive the capacity expression in FS fading channels as in (72)-(74), where  $\Phi \triangleq (\mathbf{I}_K \otimes \mathbf{G}^{-\frac{1}{2}}) \mathbf{G}_H$ . We also assume that  $\mathbf{G}$  is invertible in (72). Because  $\Phi^\dagger \Phi$  is a Hermitian matrix, we can write its eigenvalue decomposition as  $\Phi^\dagger \Phi = \mathbf{U}_\Phi \mathbf{\Lambda}_\Phi \mathbf{U}_\Phi^\dagger$ . Following the same approach as in [13], we first solve the

optimization problem for a fixed  $N$  and then send  $N$  to infinity, resulting in the capacity if the limit exists. Applying Hadamard's inequality [3] to get an upper bound on  $C_N^{\text{FS}}$ , we write

$$C_N^{\text{FS}} \leq \frac{1}{N} \log_2 \det \left( \mathbf{I}_{LN} + \sigma_0^{-2} \mathbf{\Lambda}_\Sigma \mathbf{\Phi} \right), \quad (75)$$

$$= \frac{1}{N} \sum_{i=0}^{LN-1} \log_2 (1 + \sigma_0^{-2} \lambda_i \phi_i). \quad (76)$$

where the upper bound is achieved when  $\mathbf{\Sigma}_A$  can be decomposed into  $\mathbf{U}_\Phi \mathbf{\Lambda}_\Sigma \mathbf{U}_\Phi^\dagger$ ,  $\lambda_i$  and  $\phi_i$  are eigenvalues of  $\mathbf{\Sigma}_A$  and  $\mathbf{\Phi}$  respectively.

The power constraint for FS channels remains the same as in (38b). When (75) is achieved with equality, we can manipulate the power constraint as follows

$$\lim_{N \rightarrow \infty} \frac{1}{N \delta T} \text{tr}((\mathbf{I}_L \otimes \mathbf{G}) \mathbf{\Sigma}_A) = \lim_{N \rightarrow \infty} \frac{1}{N \delta T} \text{tr}(\mathbf{U}_\Phi^\dagger (\mathbf{I}_L \otimes \mathbf{G}) \mathbf{U}_\Phi \mathbf{\Lambda}_\Sigma) \quad (77)$$

$$= \lim_{N \rightarrow \infty} \frac{1}{N \delta T} \text{tr}(\mathbf{\Psi} \mathbf{\Lambda}_\Sigma) \quad (78)$$

$$= \lim_{N \rightarrow \infty} \frac{1}{N \delta T} \sum_{i=0}^{LN-1} \psi_i \lambda_i \leq P, \quad (79)$$

where  $\mathbf{\Psi} \triangleq \mathbf{U}_\Phi^\dagger (\mathbf{I}_L \otimes \mathbf{G}) \mathbf{U}_\Phi$ , and  $\psi_i$ 's are the diagonal entries of matrix the  $\mathbf{\Psi}$ . Note that  $\mathbf{\Psi}$  is not necessarily a diagonal matrix. The optimization problem then becomes

$$C_N^{\text{FS}} = \max_{\lambda_i} \frac{1}{N} \sum_{i=0}^{LN-1} \log_2 (1 + \sigma_0^{-2} \lambda_i \phi_i) \quad (80)$$

$$\text{s.t.} \quad \frac{1}{N \delta T} \sum_{i=0}^{LN-1} \psi_i \lambda_i \leq P. \quad (81)$$

Similar to the SISO case in [13], we solve this problem by the Lagrange multiplier method, forming the Lagrangian function and by using Kuhn-Tucker conditions to find the optimal  $\lambda_i^*$ 's as

$$\lambda_i^* = \max \left( \frac{\delta T \ln 2}{\mu \psi_i} - \frac{\sigma_0^2}{\phi_i}, 0 \right). \quad (82)$$

$$C^{\text{FS}} = \lim_{N \rightarrow \infty} \max_{\mathbf{\Sigma}_A, \mathcal{P}} \frac{1}{N} \log_2 \det \left[ \mathbf{I}_{KN} + \sigma_0^{-2} \mathbf{G}_H \mathbf{\Sigma}_A \mathbf{G}_H^\dagger (\mathbf{I}_K \otimes \mathbf{G}^{-1}) \right] \quad (92)$$

$$= \lim_{N \rightarrow \infty} \max_{\mathbf{\Sigma}_A, \mathcal{P}} \frac{1}{N} \log_2 \det \left[ \mathbf{I}_{LN} + \sigma_0^{-2} \mathbf{\Sigma}_A \mathbf{G}_H^\dagger (\mathbf{I}_K \otimes \mathbf{G}^{-\frac{1}{2}}) (\mathbf{I}_K \otimes \mathbf{G}^{-\frac{1}{2}}) \mathbf{G}_H \right] \quad (93)$$

$$= \lim_{N \rightarrow \infty} \max_{\mathbf{\Sigma}_A, \mathcal{P}} \frac{1}{N} \log_2 \det \left[ \mathbf{I}_{LN} + \sigma_0^{-2} \mathbf{\Sigma}_A \Phi^\dagger \Phi \right] \quad (94)$$

As a result, assuming the limit exists, the capacity for FTN signaling in FS MIMO channels become

$$C^{\text{FS}} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^{LN-1} \log_2(1 + \sigma_0^{-2} \lambda_i^* \phi_i). \quad (83)$$

The parameters  $\phi_i$ 's and  $\psi_i$ 's are determined by the tap coefficients of the MIMO channel as well as the  $\mathbf{G}$  matrix. In other words, all factors in space and frequency have an effect on the optimal transmission power level. In FS MIMO FTN, in addition to the effect of multiple antennas, both the frequency spectrum of the channel and the frequency spectrum due to FTN ISI, have an influence on the transmitted signal. The optimal transmission scheme includes joint waterfilling in spatial and frequency domains, this is illustrated in more details in the following section.

## V. ANALYSIS IN FREQUENCY DOMAIN

We now analyse the above channel capacities in the frequency domain. In the following discussion we still assume  $\delta(1 + \beta) \geq 1$ .

### A. Spectrum Analysis of FS MIMO Channels with FTN

We start this subsection with the following lemma.

*Lemma 6:* [33, Theorem 3] The generalized Szegő's theorem states that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{l=1}^{NL} F[\lambda_l(\mathbf{R})] = \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{j=1}^L F[\lambda_j(\mathbf{R}(f_n))] df_n, \quad (84)$$

where  $F[\cdot]$  is any function continuous on the range of  $f_n$ ,  $\lambda_l(\mathbf{R})$  is the  $l$ th eigenvalue of block Toeplitz matrix  $\mathbf{R}$  and  $\lambda_j(\mathbf{R}(f_n))$  is the  $j$ th eigenvalue of matrix  $\mathbf{R}(f_n)$ . The matrix  $\mathbf{R}$  is of size  $NL \times NL$  and has the form

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \dots & \mathbf{R}_{1L} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \dots & \mathbf{R}_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{L1} & \mathbf{R}_{L2} & \dots & \mathbf{R}_{LL} \end{bmatrix}, \quad (85)$$

where  $\mathbf{R}_{ij}$ 's are  $N \times N$  Toeplitz matrices individually. The matrix  $\mathbf{R}(f_n)$  is called the generating matrix of  $\mathbf{R}$ , since it is composed of the generating functions of  $\mathbf{R}_{ij}$ 's, namely,  $(\mathbf{R}(f_n))_{ij} = \mathcal{G}_{ij}(f_n)$ , where  $\mathcal{G}_{ij}(f_n)$  is the generating function of  $\mathbf{R}_{ij}$ ,  $i, j = 1, \dots, L$ . For an arbitrary Toeplitz matrix  $\mathbf{T}_N$ , with  $(\mathbf{T}_N)_{ij} = t_{i-j}$ , its generating function is defined as [33, Theorem 3]

$$\mathcal{G}(\mathbf{T}_N) = \sum_{k=-\infty}^{\infty} t_k e^{jk f_n}, \quad f_n \in \left[-\frac{1}{2}, \frac{1}{2}\right]. \quad (86)$$

In order to express the mutual information in (72) in the frequency domain, we first need to verify that Lemma 6 can be applied, in other words, the matrix inside the determinant in (72) is a block Toeplitz matrix. First of all, it is easy to see that  $\mathbf{G}_H$ ,  $(\mathbf{I}_K \otimes \mathbf{G}^{-1})$  and  $\Sigma_A$  are block Toeplitz matrices when  $N \rightarrow \infty$ <sup>1</sup>. According to [34], we have the following lemma on the product of block Toeplitz matrices

*Lemma 7:* [34, Theorem 2] The product of block Toeplitz matrices is asymptotically a block Toeplitz matrix.

After manipulating (72) we can obtain

$$C^{\text{FS}} = \lim_{N \rightarrow \infty} \max_{\Sigma_A, \mathbf{P}} \frac{1}{N} \log_2 \det \left[ \mathbf{I}_{LN} + \sigma_0^{-2} \Sigma_A \mathbf{G}_H^\dagger (\mathbf{I}_K \otimes \mathbf{G}^{-1}) \mathbf{G}_H \right]. \quad (87)$$

In this equation, the matrix  $\mathbf{G}_H^\dagger (\mathbf{I}_K \otimes \mathbf{G}^{-1}) \mathbf{G}_H$  can be written as (88)-(89), where  $\mathbf{G}^j$  was defined in (67) and  $i, j = 0, \dots, J-1$ . Note that,  $\mathbf{G}_H^\dagger (\mathbf{I}_K \otimes \mathbf{G}^{-1}) \mathbf{G}_H$  is an asymptotically block Toeplitz matrix. In order to find its generating matrix, we need to find the generating function of  $\mathbf{G}^{j\dagger} \mathbf{G}^{-1} \mathbf{G}^i$ . According to [27], this can be found by multiplying the generating functions of each of the three matrices. To find the generating function of  $\mathbf{G}$ , let's define the discrete time Fourier transform of  $g[n]$  as

$$G_d(f_n) \triangleq \sum_{p=-\infty}^{+\infty} g[p] e^{-j2\pi f_n p} = \frac{1}{\delta T} \sum_{m=-\infty}^{+\infty} G\left(\frac{f_n - m}{\delta T}\right), \quad (90)$$

where  $G(f)$  is the continuous time Fourier transform of  $g(t)$ . It is easy to see that  $G_d(-f_n)$  is the generating function of matrix  $\mathbf{G}$  according to the definition of generating function, but since  $G_d(f_n)$  is an even function  $G_d(-f_n) = G_d(f_n)$ . The folded spectrum  $G_d(f_n)$  is periodic in  $f_n$  with period 1. Thus,

$$\mathcal{G}(\mathbf{G}^{j\dagger} \mathbf{G}^{-1} \mathbf{G}^i) = \mathcal{G}(\mathbf{G}^{j\dagger}) \mathcal{G}(\mathbf{G}^{-1}) \mathcal{G}(\mathbf{G}^i) \quad (91)$$

$$= \left( \sum_n g(-(n+j)\delta T) e^{j2\pi f_n n} \right) \times \left( \sum_m g((m-i)\delta T) e^{j2\pi f_n m} \right) G_d^{-1}(f_n) \quad (92)$$

$$= G_d(f_n) e^{-j2\pi f_n j} G_d(f_n) e^{j2\pi f_n i} G_d^{-1}(f_n) \quad (93)$$

$$= G_d(f_n) e^{-j2\pi f_n (j-i)}, \quad (94)$$

where we let  $\mathcal{G}$  represent both finding the generating function and finding the generating matrix operation. With the derivation above we can easily obtain the generating matrix of  $\mathbf{G}_H^\dagger (\mathbf{I}_K \otimes \mathbf{G}^{-1}) \mathbf{G}_H$ , whose  $(n, m)$ th entry is

$$(\mathcal{G}(\mathbf{G}_H^\dagger (\mathbf{I}_K \otimes \mathbf{G}^{-1}) \mathbf{G}_H))_{n,m} = \sum_{i,j} \left( \sum_{k=1}^K h_{kn}^{j*} h_{km}^i \right) \mathcal{G}(\mathbf{G}^{j\dagger} \mathbf{G}^{-1} \mathbf{G}^i) \quad (95)$$

$$= \sum_{k=1}^K \left[ \left( \sum_{i=0}^{J-1} h_{km}^i e^{j2\pi f_n i} \right) \left( \sum_{j=0}^{J-1} h_{kn}^{j*} e^{-j2\pi f_n j} \right) G_d(f_n) \right] \quad (96)$$

$$= \sum_{k=1}^K H_{kn}^*(-f_n) H_{km}(-f_n) G_d(f_n). \quad (97)$$

Here  $H_{km}(-f_n)$  is defined as  $H_{km}(-f_n) = \sum_{i=0}^{J-1} h_{km}^i e^{j2\pi f_n i}$ , which is the discrete time Fourier

<sup>1</sup>According to [27], since the inverse of a Toeplitz matrix is asymptotically Toeplitz,  $\mathbf{I}_K \otimes \mathbf{G}^{-1}$  is also an asymptotically block Toeplitz matrix.



$$\mathbf{G}_H^\dagger (\mathbf{I}_K \otimes \mathbf{G}^{-1}) \mathbf{G}_H = \left( \sum_{j=0}^{J-1} \left( \tilde{\mathbf{H}}^{j\dagger} \otimes \mathbf{G}^{j\dagger} \mathbf{G}^{-1} \right) \right) \left( \sum_{i=0}^{J-1} \left( \tilde{\mathbf{H}}^i \otimes \mathbf{G}^i \right) \right) = \sum_{j=0}^{J-1} \sum_{i=0}^{J-1} \left( \tilde{\mathbf{H}}^{j\dagger} \tilde{\mathbf{H}}^i \otimes \mathbf{G}^{j\dagger} \mathbf{G}^{-1} \mathbf{G}^i \right) \quad (88)$$

$$= \begin{bmatrix} \sum_{i,j} \left( \sum_{k=1}^K h_{k1}^{j*} h_{k1}^i \right) \mathbf{G}^{j\dagger} \mathbf{G}^{-1} \mathbf{G}^i & \sum_{i,j} \left( \sum_{k=1}^K h_{k1}^{j*} h_{k2}^i \right) \mathbf{G}^{j\dagger} \mathbf{G}^{-1} \mathbf{G}^i & \cdots & \sum_{i,j} \left( \sum_{k=1}^K h_{k1}^{j*} h_{kL}^i \right) \mathbf{G}^{j\dagger} \mathbf{G}^{-1} \mathbf{G}^i \\ \sum_{i,j} \left( \sum_{k=1}^K h_{k2}^{j*} h_{k1}^i \right) \mathbf{G}^{j\dagger} \mathbf{G}^{-1} \mathbf{G}^i & \sum_{i,j} \left( \sum_{k=1}^K h_{k2}^{j*} h_{k2}^i \right) \mathbf{G}^{j\dagger} \mathbf{G}^{-1} \mathbf{G}^i & \cdots & \sum_{i,j} \left( \sum_{k=1}^K h_{k2}^{j*} h_{kL}^i \right) \mathbf{G}^{j\dagger} \mathbf{G}^{-1} \mathbf{G}^i \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i,j} \left( \sum_{k=1}^K h_{kL}^{j*} h_{k1}^i \right) \mathbf{G}^{j\dagger} \mathbf{G}^{-1} \mathbf{G}^i & \sum_{i,j} \left( \sum_{k=1}^K h_{kL}^{j*} h_{k2}^i \right) \mathbf{G}^{j\dagger} \mathbf{G}^{-1} \mathbf{G}^i & \cdots & \sum_{i,j} \left( \sum_{k=1}^K h_{kL}^{j*} h_{kL}^i \right) \mathbf{G}^{j\dagger} \mathbf{G}^{-1} \mathbf{G}^i \end{bmatrix} \quad (89)$$

transform of the coefficients of the  $(n, m)$ th antenna link. We call this the link spectrum. Now we define the matrix

$$\tilde{\mathbf{H}}(-f_n) = \begin{bmatrix} H_{11}(-f_n) & H_{12}(-f_n) & \cdots & H_{1L}(-f_n) \\ H_{21}(-f_n) & H_{22}(-f_n) & \cdots & H_{2L}(-f_n) \\ \vdots & \vdots & \ddots & \vdots \\ H_{K1}(-f_n) & H_{K2}(-f_n) & \cdots & H_{KL}(-f_n) \end{bmatrix} \quad (98)$$

as the channel spectrum matrix, which is composed of the link spectrums. Then it is straightforward to write

$$\mathcal{G}(\mathbf{G}_H^\dagger (\mathbf{I}_K \otimes \mathbf{G}^{-1}) \mathbf{G}_H) = G_d(f_n) \tilde{\mathbf{H}}(-f_n)^\dagger \tilde{\mathbf{H}}(-f_n) = G_d(f_n) \tilde{\mathbf{Z}}(-f_n), \quad (99)$$

where  $\tilde{\mathbf{Z}}(-f_n) \triangleq \tilde{\mathbf{H}}(-f_n)^\dagger \tilde{\mathbf{H}}(-f_n)$ .

We also need to find the generating matrix for  $\Sigma_A$  to obtain  $\mathcal{G}(\Sigma_A \mathbf{G}_H^\dagger (\mathbf{I}_K \otimes \mathbf{G}^{-1}) \mathbf{G}_H)$ , which is in the resulting expression after applying Szegő's theorem on (87). As  $\Sigma_A$  is also a block Toeplitz matrix, we define

$$\tilde{\mathbf{S}}_a(f_n) = \mathcal{G}(\Sigma_A) = \begin{bmatrix} S_{a,11}(f_n) & S_{a,12}(f_n) & \cdots & S_{a,1L}(f_n) \\ S_{a,21}(f_n) & S_{a,22}(f_n) & \cdots & S_{a,2L}(f_n) \\ \vdots & \vdots & \ddots & \vdots \\ S_{a,L1}(f_n) & S_{a,L2}(f_n) & \cdots & S_{a,LL}(f_n) \end{bmatrix}, \quad (100)$$

where  $S_{a,ij}(f_n) = \mathcal{G}(\mathbf{K}_{a,ij})$ ,  $i, j = 1, \dots, L$ . We call  $\tilde{\mathbf{S}}_a(f_n)$  as the input generating matrix, the diagonal elements of which are the PSDs of input data sequences  $a_i[n]$ ,  $i = 1, \dots, L$ , and the off-diagonal elements are the cross PSDs of input data sequences  $a_i[n]$  and  $a_j[n]$ ,  $i, j = 1, \dots, L$ ,  $i \neq j$ . By applying (84) of Lemma 6 on (87) and on the power constraint (77), the optimization problem becomes

$$C^{FS} = \max_{\tilde{\mathbf{S}}_a(f_n)} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log_2 \det \left[ \mathbf{I}_L + \frac{G_d(f_n)}{\sigma_0^2} \tilde{\mathbf{S}}_a(f_n) \tilde{\mathbf{Z}}(f_n) \right] df_n \quad (101)$$

$$s.t. \quad \frac{1}{\delta T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \text{tr} [G_d(f_n) \tilde{\mathbf{S}}_a(f_n)] df_n \leq P. \quad (102)$$

Before we solve the above optimization problem, let's define  $\tilde{\mathbf{S}}(f_n) \triangleq G_d(f_n) \tilde{\mathbf{S}}_a(f_n)$ . We call  $\tilde{\mathbf{S}}(f_n)$  as the power generating matrix, since the integration over the sum of its diagonal elements is the transmit power. Note that, (101) contains both the influence from the FS channel  $\tilde{\mathbf{Z}}(f_n)$  and the spectrum of the pulse shaping filter  $G_d(f_n)$ . This is because the ISI

caused by FTN and the channel-induced ISI jointly affect the system. It can be considered that the channel has additional frequency selectivity with spectrum  $G_d(f_n)$ , which is due to FTN signaling.

We would like to emphasize that  $\tilde{\mathbf{S}}(f_n)$  and  $\tilde{\mathbf{Z}}(f_n)$  are always Hermitian matrices regardless of  $f_n$ . Therefore, we can use eigenvalue decomposition to diagonalize them and write

$$\tilde{\mathbf{S}}(f_n) = G_d(f_n) \tilde{\mathbf{S}}_a(f_n) = \tilde{\mathbf{U}}(f_n) \tilde{\mathbf{\Phi}}(f_n) \tilde{\mathbf{U}}(f_n)^\dagger \quad (103)$$

$$\tilde{\mathbf{Z}}(f_n) = \tilde{\mathbf{V}}(f_n) \tilde{\mathbf{T}}(f_n) \tilde{\mathbf{V}}(f_n)^\dagger, \quad (104)$$

where the diagonal matrices  $\tilde{\mathbf{\Phi}}(f_n)$  and  $\tilde{\mathbf{T}}(f_n)$  have the structure  $\tilde{\mathbf{\Phi}}(f_n) = \text{diag}\{\phi_1(f_n), \phi_2(f_n), \dots, \phi_L(f_n)\}$  and  $\tilde{\mathbf{T}}(f_n) = \text{diag}\{\tau_1(f_n), \tau_2(f_n), \dots, \tau_L(f_n)\}$ . The channel matrix  $\mathbf{Z}(f_n)$  is decomposed into eigenchannels,  $\tau_i(f_n)$ 's are the eigenmodes at frequency  $f_n$ , and  $\phi_i(f_n)$ 's are the power allocated to the eigenchannels, which we call the eigenspectrum.

In the next step, we can upper bound (101) using Hadamard inequality as

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \log_2 \det \left[ \mathbf{I}_L + \sigma_0^{-2} \tilde{\mathbf{S}}(f_n) \tilde{\mathbf{Z}}(f_n) \right] df_n \leq \quad (105)$$

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \log_2 \det \left[ \mathbf{I}_L + \sigma_0^{-2} \tilde{\mathbf{\Phi}}(f_n) \tilde{\mathbf{T}}(f_n) \right] df_n, \quad (106)$$

and the upper bound is achieved when  $\tilde{\mathbf{U}}(f_n) = \tilde{\mathbf{V}}(f_n)$ ,  $\forall f_n$ . Then the problem in (101)-(102) becomes

$$C^{FS} = \max_{\phi_1(f_n), \dots, \phi_L(f_n)} \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{i=1}^L \log_2 (1 + \sigma_0^{-2} \phi_i(f_n) \tau_i(f_n)) df_n \quad (107)$$

$$s.t. \quad \frac{1}{\delta T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{i=1}^L \phi_i(f_n) df_n \leq P. \quad (108)$$

By applying the Lagrange multiplier method, we get the optimal solution for  $\phi_i(f_n)$  as

$$\phi_i(f_n)^* = \left( \frac{1}{\mu} - \frac{1}{\tau_i(f_n)} \right)^+, \quad i = 1, \dots, L, \quad f_n \in \left[ -\frac{1}{2}, \frac{1}{2} \right], \quad (109)$$

where superscript  $*$  means optimal. The Lagrange multiplier  $\mu$  can be found by solving

$$\frac{1}{\delta T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{i=1}^L \left( \frac{1}{\mu} - \frac{1}{\tau_i(f_n)} \right)^+ df_n = P. \quad (110)$$

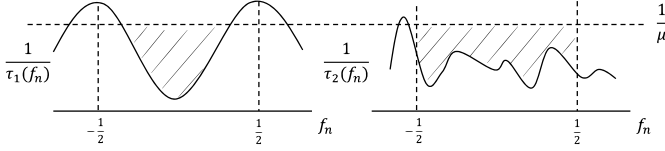


Fig. 2. Waterfilling over a FS  $2 \times 2$  MIMO channel with FTN. According to the inverted eigenmodes, the water level is the same for both eigenchannels.

The optimal eigenspectrum for each eigenchannel is in the form of joint waterfilling in both spatial and spectral domains. There are  $L$  eigenchannels in total, and waterfilling is done according to the inverted eigenmodes  $1/\tau_i(f_n)$ . All eigenchannels have the same water level  $\frac{1}{\mu}$  such that the integrated power over all frequencies and all channels satisfies (108). This is illustrated in Fig. 2.

Because we are applying FTN signaling, the resulting optimal generating matrix  $\tilde{\mathbf{S}}_a^*(f_n)$  can be obtained according to (103) as  $\tilde{\mathbf{S}}_a^*(f_n) = \frac{\tilde{\mathbf{S}}^*(f_n)}{G_d(f_n)}$ . This is equivalent to dividing  $\phi_i(f_n)$ 's of  $\tilde{\Phi}(f_n)$  in (103) by  $G_d(f_n)$ , and we call this shaped power allocation  $\tilde{\phi}_i(f_n) = \frac{\phi_i(f_n)}{G_d(f_n)}$  as the input eigenspectrum. This also shows the direct influence of FTN on the power allocation. When we design the input distribution, the power spectrum for each eigenchannel is shaped by  $\frac{1}{G_d(f_n)}$ .

*Remark 6:* The influence of FTN on the power allocation is separate from the influence of the channel. The only difference FTN makes is further shaping the power spectrum by  $\frac{1}{G_d(f_n)}$ . Moreover, for orthogonal signaling with pulses following the Nyquist ISI criterion, the optimal input generating matrix is just  $\tilde{\mathbf{S}}^*(f_n)$  since  $G_d(f_n) = 1$ .

### B. Spectrum Analysis of Frequency Flat MIMO Channels with FTN

In this section, we obtain the power spectrum of the frequency flat MIMO FTN channel as a special case of the above discussion. In frequency flat fading, the channel response has only one tap,  $h_{kl}^j = h_{kl}\delta[j]$ , where  $h_{kl}$  is the channel coefficient for link  $(k, l)$  and  $\delta[j]$  is the discrete Dirac delta function. Then (101) and (102) become

$$C = \max_{\tilde{\mathbf{S}}(f_n)} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log_2 \det \left[ \mathbf{I}_L + \sigma_0^{-2} \tilde{\mathbf{S}}(f_n) \tilde{\mathbf{Z}}(f_n) \right] df_n \quad (111)$$

$$\text{s.t.} \quad \frac{1}{\delta T} \int_{-\frac{1}{2}}^{\frac{1}{2}} \text{tr}[\tilde{\mathbf{S}}(f_n)] df_n \leq P. \quad (112)$$

This optimization is solved by the same spectrum waterfilling strategy of the previous section, only that the eigenmodes  $\tau_i(f_n)$ 's have a constant spectrum on  $[-\frac{1}{2}, \frac{1}{2}]$  with amplitude  $\tau_i$ 's. The resulting optimal  $\phi_i^*(f_n)$ 's also have constant spectrum according to (109), and the amplitudes are identical to the solution in (47). It is easy to see that the optimal power generating matrix in (103) becomes

$$\tilde{\mathbf{S}}(f_n) = \mathbf{V}(f_n) \mathbf{\Lambda}_{\tilde{\alpha}} \mathbf{V}^\dagger(f_n) = \mathbf{V}_{\mathbf{Z}} \mathbf{\Lambda}_{\tilde{\alpha}} \mathbf{V}_{\mathbf{Z}}^\dagger. \quad (113)$$

However, the optimal input generating matrix is not constant, it is equal to

$$\tilde{\mathbf{S}}_a(f_n) = \frac{\mathbf{V}_{\mathbf{Z}} \mathbf{\Lambda}_{\tilde{\alpha}} \mathbf{V}_{\mathbf{Z}}^\dagger}{G_d(f_n)}. \quad (114)$$

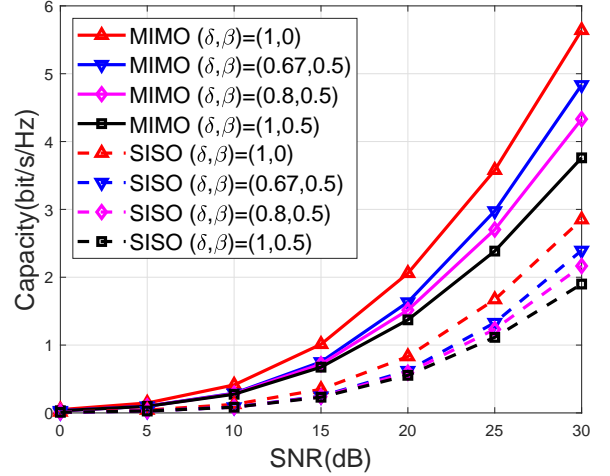


Fig. 3. Frequency flat,  $2 \times 2$  MIMO FTN and SISO FTN capacity results for different  $\delta$  values for the roll-off factor  $\beta = 0.5$ . The pairs  $(\delta, \beta) = (1, 0)$  and  $(1, 0.5)$  are respectively equivalent to Nyquist signaling with an ideal sinc pulse, and to Nyquist signaling with  $\beta = 0.5$ .

We note that this matrix is exactly the same as the generating matrix of the optimal input covariance matrix in (57). The optimal input eigenspectrum in this case will be  $\frac{\tilde{\alpha}_i}{G_d(f_n)}$ .

*Remark 7:* Due to (114), we can see that in flat fading MIMO channels, the influence of FTN on the spectrum can be separated from the spatial effects of MIMO. This is also a natural consequence of Remark 6 for FS MIMO FTN channels.

*Remark 8:* We can consider frequency flat SISO channel with FTN as a special case of the above result and easily obtain the result in [9].

## VI. NUMERICAL RESULTS

In this section, we display FTN capacity for different  $(\delta, \beta)$  pairs for both frequency flat and FS MIMO channels. We assume  $g(t)$  is a raised cosine pulse and  $T = 0.01$ . The SNR is computed as  $\text{SNR} = P/\sigma_0^2$ . For frequency flat fading cases in Figs. 3 - 6, channel coefficients follow a circularly symmetric complex Gaussian distribution with zero mean and unit variance. All the simulation results are averaged over 1000 random channel realizations, unless otherwise stated.

In Fig. 3 we compare the capacity of a frequency flat,  $2 \times 2$  MIMO FTN channel with the capacity of a frequency flat SISO FTN channel, for different acceleration factors  $\delta$ . The roll-off factor  $\beta$  is set to 0.5. We observe that for a given  $\beta$ , as we decrease  $\delta$  capacity becomes larger for both MIMO and SISO. This spectral efficiency improvement clearly demonstrates the superiority of FTN for practical pulse shapes. FTN signaling utilizes the bandwidth more efficiently and outperforms orthogonal signaling. Despite the fact that FTN introduces ISI, increasing signaling rate or decreasing  $\delta$  while  $\delta(1 + \beta) \geq 1$  satisfied, increases capacity. Note that,  $(\delta, \beta) = (1, 0)$ ; i.e. orthogonal signaling with ideal sinc pulses, is a performance upper bound for both SISO and MIMO FTN.

In Fig. 4, we compare different  $(\delta, \beta)$  pairs for the  $2 \times 2$  MIMO FTN channel with frequency flat fading. We observe that the capacity increases as  $\beta$  decreases, and for each  $\beta$

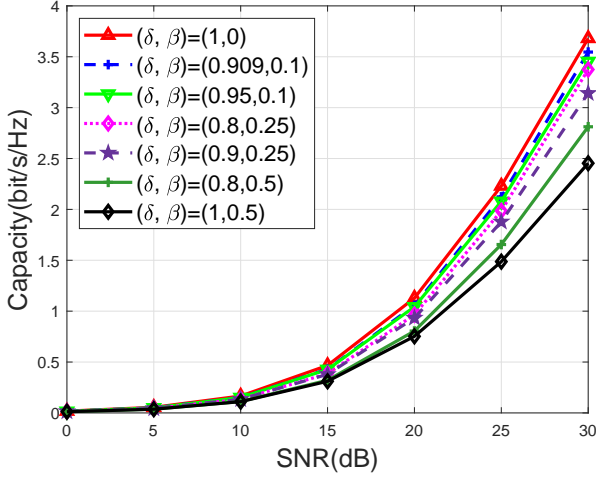


Fig. 4. Frequency flat,  $2 \times 2$  MIMO FTN capacity for different  $(\delta, \beta)$  values, displaying the effect of  $\beta$  approaching 0. The pairs  $(\delta, \beta) = (1, 0)$  and  $(1, 0.5)$  are respectively equivalent to Nyquist signaling with an ideal sinc pulse, and to Nyquist signaling with  $\beta = 0.5$ .

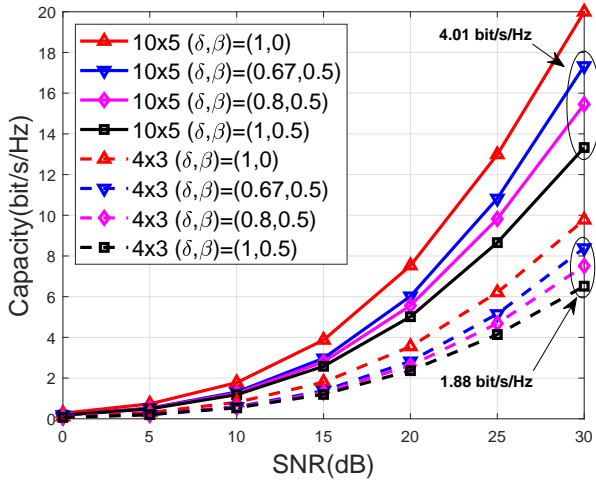


Fig. 5. Frequency flat,  $4 \times 3$  and  $10 \times 5$  MIMO FTN capacity results for different  $\delta$  values for the roll-off factor  $\beta = 0.5$ . The pairs  $(\delta, \beta) = (1, 0)$  and  $(1, 0.5)$  are respectively equivalent to Nyquist signaling with an ideal sinc pulse, and to Nyquist signaling with  $\beta = 0.5$ .

the capacity increases as  $\delta$  decreases while  $\delta \geq 1/(1 + \beta)$ . Similar to the SISO case, we observe that for each  $\beta$ , the best  $\delta = 1/(1 + \beta)$ , and overall the best  $(\delta, \beta)$  pair is  $(1, 0)$ . Although orthogonal signaling with sinc pulses achieve the best performance, practical systems operate at non-zero  $\beta$  values and the acceleration factor should be chosen as  $1/(1 + \beta)$  to achieve the highest spectral efficiency possible for in MIMO FTN as well.

We compare  $4 \times 3$  and  $10 \times 5$  MIMO FTN for frequency flat fading channels in Fig. 5. We observe that there is a larger gain of FTN against Nyquist transmission when we use larger sized antenna arrays. For example, in Fig 5, for  $\beta = 0.5$ , changing  $\delta$  from 1 to 0.67 brings in an extra 4.01 bits/s/Hz in  $10 \times 5$  MIMO FTN, while it offers only 1.88 bits/s/Hz in  $4 \times 3$  MIMO FTN signaling. Thus, the potential of FTN signaling increases as we increase the number of antennas.

We plot the spectrum domain power allocation for a frequency flat,  $2 \times 2$  MIMO channel with FTN for one particular channel realization in Fig. 6. We assume  $\delta = 0.9$ ,  $\beta = 0.25$  and SNR is 10dB. In the figure, the two eigenmodes  $\tau_i$ ,  $i = 1, 2$ , power allocation  $\bar{\alpha}_i$  (or equivalently  $\phi_i(f_n)$ ) and the optimal input eigenspectrum  $\frac{\bar{\alpha}_i}{G_d(f_n)}$  (equivalently  $\frac{\phi_i(f_n)}{G_d(f_n)}$ ) are shown. As we can see, the power allocation is composed of two independent parts, waterfilling in spatial domain and spectrum shaping in frequency domain as discussed in Remark 6 in Section V-A. Amplitude of  $\phi_i(f_n)$  is bigger when eigenmode  $\tau_i$  is stronger, which is a natural consequence of waterfilling.

We plot the MIMO FTN spectral efficiency (SE) under FS fading for different  $(\delta, \beta)$  pairs in Fig. 7. We also compare the performance of equal power allocation with that of optimal power allocation. For equal power allocation, we assume  $\Sigma_A = cI_{LN}$ , where  $c$  is a scaling factor that satisfies the power constraint (79) with equality. As in (64), we assume the gap between taps is  $\delta T$ , and all channel coefficients  $h_{kl}^j$  are independent and identically distributed according to the complex Gaussian distribution  $\mathcal{CN}(0, 1/(LJ))$ . We set the number of taps to  $J = 20$  and  $N = 1000$ . As can be seen from the figure, FTN with optimum power allocation performs better than Nyquist signaling for the same  $\beta$ . The performance of Nyquist signaling with  $(\delta, \beta) = (1, 0)$  is again an upper bound on system performance. Moreover, when we compare optimal and equal power allocation for FTN signaling with  $(\delta, \beta) = (0.8, 0.5)$ , we observe that optimal power allocation can achieve 0.381bit/s/Hz more rate with respect to equal power allocation.

We plot Fig. 8 with the same assumptions as in Fig. 7 to further study the effect of frequency selectivity. All curves are for optimized power allocation. We can see that FTN improves the performance for both flat fading and FS channels, and the benefit of FTN is independent from the channel, whether it is FS or flat fading. Note that the superiority of FS channel capacity against flat fading is due to additional diversity gain [21]. Moreover, in support of Remark 6, the figure shows that FTN does not change the fact that frequency selectivity increases performance.

In Fig. 9, we plot the spectrum domain power allocation of FS  $2 \times 2$  MIMO channel with FTN to show the joint influence of both the FS channel and FTN. We assume all channel coefficients follow i.i.d. Gaussian distribution with zero mean and variance  $\frac{1}{LJ}$  where  $J = 5$ , SNR is 5dB,  $\delta = 0.9$ , and  $\beta = 0.25$ . The figures are plotted for one random channel realization only. All the spectrum in the subfigures are periodic with period 1. As we can see, the water-filling result  $\phi_1(f_n)$  and  $\phi_2(f_n)$  of (109) have bigger amplitude where the eigenmodes  $\tau_1(f_n)$  and  $\tau_2(f_n)$  from (104) are strong. This is reasonable since it is not desirable to invest in weak channel conditions. The optimal input eigenspectrum  $\frac{\phi_i(f_n)}{G_d(f_n)}$ ,  $i = 1, 2$ , is amplified at where  $G_d(f_n)$  has a dent, so that it will cancel the influence brought by FTN.

## VII. CONCLUSION

In this paper we investigate the capacity of FTN signaling for both frequency flat and frequency selective MIMO

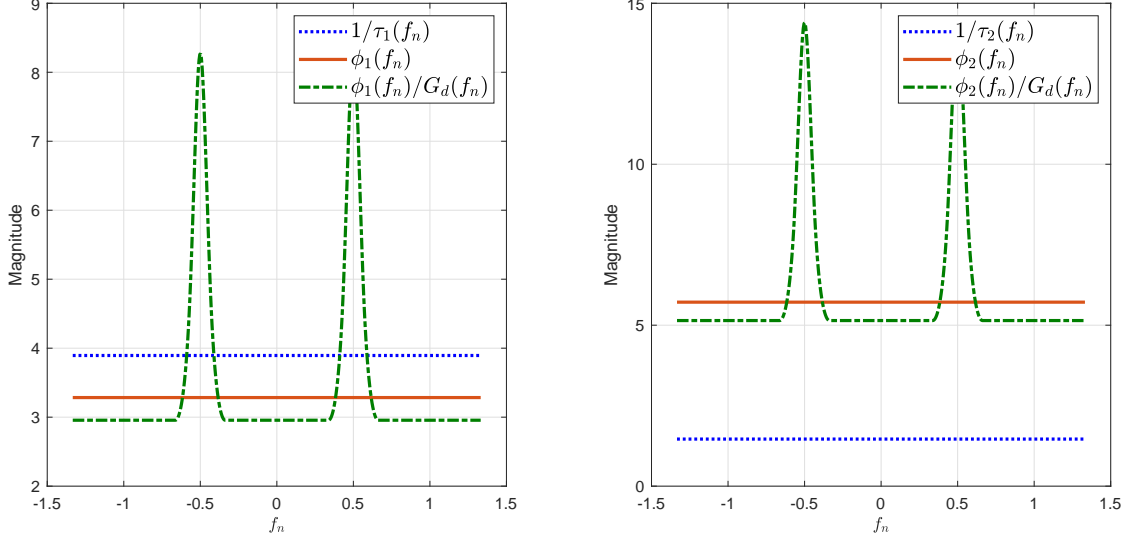


Fig. 6. Frequency domain power allocation for a  $2 \times 2$  MIMO FTN in frequency flat channel.

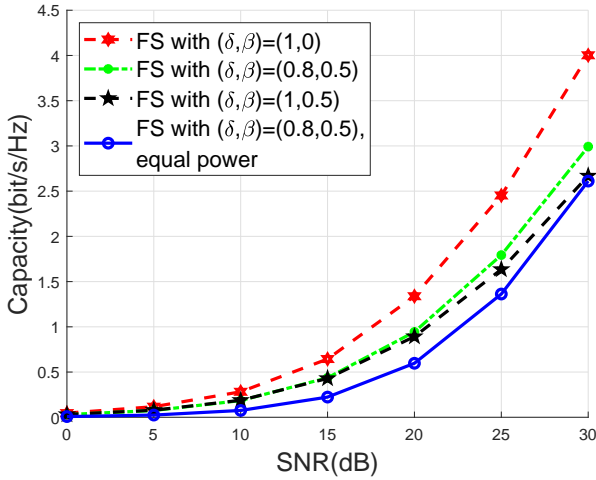


Fig. 7.  $2 \times 2$  MIMO FTN spectral efficiency results under frequency selective fading for different  $\delta$  values for the roll-off factor  $\beta = 0.5$ . The pairs  $(\delta, \beta) = (1, 0)$  and  $(1, 0.5)$  are respectively equivalent to Nyquist signaling with an ideal sinc pulse, and to Nyquist signaling with  $\beta = 0.5$ . The blue curve with circle marks display the performance for equal power allocation.

channels. The transmit power constraint takes the effect of inter-symbol interference into account, and is expressed in a matrix multiplication form. This expression simplifies the capacity optimization problem, and leads to the optimal joint precoding and waterfilling solution for MIMO FTN in flat fading channels. The optimal transmission scheme removes inter-symbol interference at the transmitter side and allocates power optimally over transmit antennas via waterfilling. We also study the capacity and the optimal transmission scheme for frequency selective fading channels. The capacity of MIMO FTN signaling in frequency selective channels can be achieved by water-filling in both spatial and spectrum domains together with spectrum shaping. Moreover, we prove that the effects of FTN and frequency selectivity on the

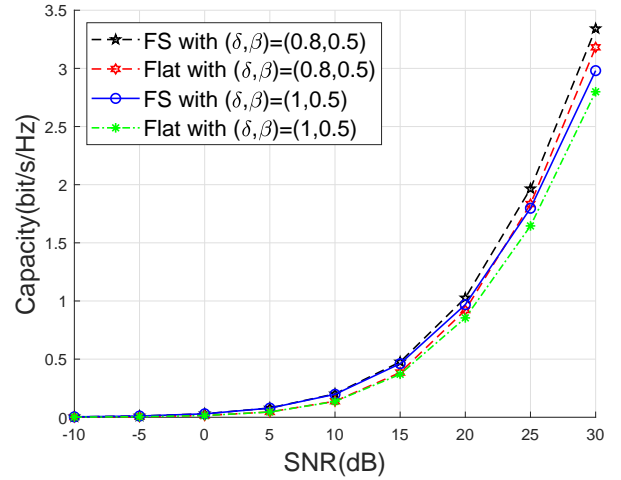


Fig. 8.  $2 \times 2$  MIMO FTN capacity results under frequency flat and frequency selective fading channels for different  $\delta$  values for the roll-off factor  $\beta = 0.5$ .

optimized power spectrum are separable. The optimal input eigenspectrum functions are determined jointly by the channel and the influence of FTN; however, the MIMO channel only affects the eigenspectrum functions, and the pulse shaping filter of FTN provides spectrum shaping. Overall, we conclude that MIMO FTN significantly improves spectral efficiency with respect to orthogonal signaling for practical root raised cosine pulses with roll-off factor  $\beta > 0$  for both frequency flat and frequency selective channels. We believe that a complete analysis which accounts for  $\delta(1 + \beta) < 1$ , is a very complex problem, and it will be considered as future work.

#### APPENDIX A

In this appendix, we prove that the matrix  $\bar{\Sigma}_A$  defined in (50) as  $(I_L \otimes G)^{-1} \bar{K}$  is a valid covariance matrix or equivalently it is both positive semidefinite and Hermitian.

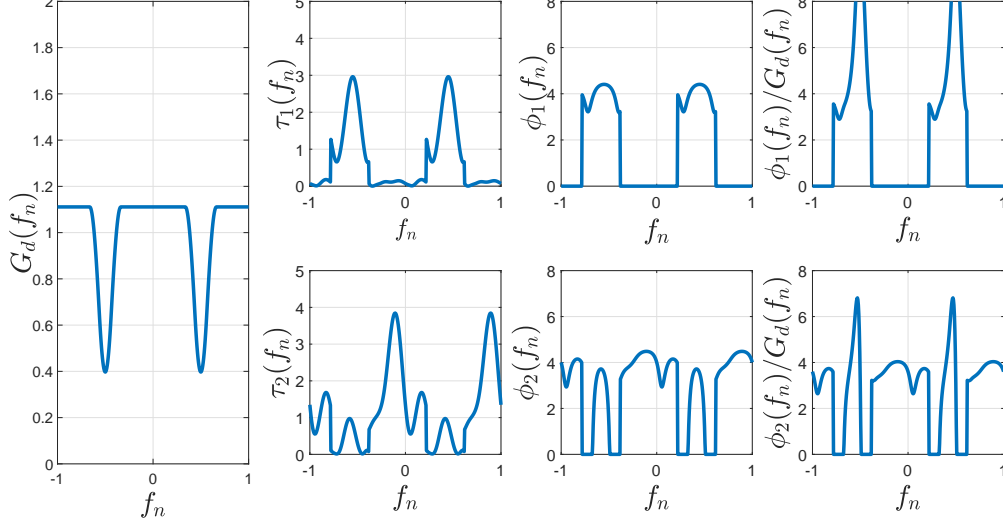


Fig. 9. Spectrum domain analysis of FTN signaling in  $2 \times 2$  MIMO FS channel.

We know that matrix  $\bar{\mathbf{K}}$  is a Hermitian matrix. Moreover,  $\bar{\mathbf{K}}$  is positive semidefinite, and its eigenvalues  $\bar{\alpha}_i \geq 0, \forall i \in \{1, \dots, NL\}$ .

The matrix  $(\mathbf{I}_L \otimes \mathbf{G})$  is also Hermitian. Moreover, due to Lemma 1,  $\mathbf{G}$  is a positive semidefinite matrix. According to the property of the Kronecker product we have  $(\mathbf{I} \otimes \mathbf{G})^{-1} = \mathbf{I} \otimes \mathbf{G}^{-1}$ , which can be decomposed into  $(\mathbf{Q} \otimes \mathbf{V}_G)(\mathbf{I}_L \otimes \Lambda_G^{-1})(\mathbf{Q} \otimes \mathbf{V}_G)^\dagger$ , where  $\mathbf{Q} \in \mathbb{C}^{L \times L}$  is a unitary matrix composed of the eigenvectors of  $\mathbf{I}_L$  [30]. Furthermore, we can diagonalize the matrix  $\mathbf{W} = (\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{H}}) \otimes \mathbf{I}_N$  of (41) as  $(\mathbf{V}_Z \otimes \mathbf{U})(\Lambda_Z \otimes \mathbf{I}_N)(\mathbf{V}_Z \otimes \mathbf{U})^\dagger$ , where  $\mathbf{U} \in \mathbb{C}^{N \times N}$  is a unitary matrix composed of the eigenvectors of  $\mathbf{I}_N$ . As  $\mathbf{Q}$  and  $\mathbf{U}$  are composed of the eigenvectors of the identity matrix, they can be chosen as any unitary matrix. Therefore, they can also be chosen as  $\mathbf{Q} = \mathbf{V}_Z$  and  $\mathbf{U} = \mathbf{V}_G$ . Thus, the optimal  $\bar{\mathbf{K}}$  becomes

$$\bar{\mathbf{K}} = \mathbf{V}_W \bar{\Lambda}_K \mathbf{V}_W^\dagger = (\mathbf{V}_Z \otimes \mathbf{V}_G) \bar{\Lambda}_K (\mathbf{V}_Z \otimes \mathbf{V}_G)^\dagger. \quad (115)$$

We can also write the matrix  $\mathbf{I} \otimes \mathbf{G}^{-1}$  as

$$(\mathbf{I} \otimes \mathbf{G}^{-1}) = (\mathbf{V}_Z \otimes \mathbf{V}_G)(\mathbf{I}_L \otimes \Lambda_G^{-1})(\mathbf{V}_Z \otimes \mathbf{V}_G)^\dagger. \quad (116)$$

Then, the optimal solution for  $\bar{\Sigma}_A$  in (50) can be simplified as

$$\bar{\Sigma}_A = (\mathbf{I}_L \otimes \mathbf{G})^{-1} \bar{\mathbf{K}} \quad (117)$$

$$= (\mathbf{V}_Z \otimes \mathbf{V}_G)(\mathbf{I}_L \otimes \Lambda_G^{-1}) \bar{\Lambda}_K (\mathbf{V}_Z \otimes \mathbf{V}_G)^\dagger. \quad (118)$$

As  $(\mathbf{I}_L \otimes \Lambda_G^{-1}) \bar{\Lambda}_K$  is a diagonal matrix, we first observe that  $\bar{\Sigma}_A$  is a Hermitian matrix. Moreover, its eigenvalues are the diagonal elements of  $(\mathbf{I}_L \otimes \Lambda_G^{-1}) \bar{\Lambda}_K$ . As the elements of  $(\mathbf{I}_L \otimes \Lambda_G^{-1})$  are strictly greater than zero, all entries of  $(\mathbf{I}_L \otimes \Lambda_G^{-1}) \bar{\Lambda}_K$  are also greater than or equal to zero. Thus,  $\bar{\Sigma}_A$  is also a positive semidefinite matrix. We conclude that  $\bar{\Sigma}_A$  is covariance matrix and the optimal solution found by (50) is a qualified solution.

#### ACKNOWLEDGMENT

The authors would like to thank the editor and all three anonymous reviewers for all their valuable comments that improved the quality of the paper.

#### REFERENCES

- [1] T. Nakamura, "5G evolution and 6G," in *IEEE Symposium on VLSI Technology*, 2020, pp. 1–5.
- [2] J. E. Mazo, "Faster-than-Nyquist signaling," *The Bell System Technical Journal*, vol. 54, no. 8, pp. 1451–1462, 1975.
- [3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 2006.
- [4] F. Rusek and J. B. Anderson, "Constrained capacities for faster-than-Nyquist signaling," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 764–775, 2009.
- [5] Y. J. Daniel Kim, "Properties of faster-than-Nyquist channel matrices and folded-spectrum, and their applications," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2016, pp. 1–7.
- [6] Y. J. D. Kim and J. Bajcsy, "Information rates of cyclostationary faster-than-Nyquist signaling," in *12th Canadian Workshop on Information Theory (CWIT)*, 2011, pp. 1–4.
- [7] D. Kapetanovic and F. Rusek, "The effect of signaling rate on information rate for single carrier linear transmission systems," *IEEE Transactions on Communications*, vol. 60, no. 2, pp. 421–428, 2012.
- [8] M. E. Hefnawy, G. Dietl, and G. Kramer, "Spectral shaping for faster-than-Nyquist signaling," in *11th International Symposium on Wireless Communications Systems (ISWCS)*, 2014, pp. 496–500.
- [9] M. Ganji, X. Zou, and H. Jafarkhani, "On the capacity of faster than Nyquist signaling," *IEEE Communications Letters*, vol. 24, no. 6, pp. 1197–1201, 2020.
- [10] A. Gattami, E. Ringh, and J. Karlsson, "Time localization and capacity of faster-than-Nyquist signaling," in *IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–7.
- [11] H. Wang, A. Liu, X. Liang, S. Peng, and K. Wang, "Linear precoding for faster-than-Nyquist signaling," in *3rd IEEE International Conference on Computer and Communications (ICCC)*, 2017, pp. 52–56.
- [12] T. Ishihara and S. Sugiura, "SVD-precoded faster-than-Nyquist signaling with optimal and truncated power allocation," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5909–5923, 2019.
- [13] —, "Eigendecomposition-precoded faster-than-nyquist signaling with optimal power allocation in frequency-selective fading channels," *IEEE Transactions on Wireless Communications*, vol. 21, no. 3, pp. 1681–1693, 2022.

- [14] P. Chaki, T. Ishihara, and S. Sugiura, "Eigenvalue decomposition pre-coded faster-than-Nyquist transmission of index modulated symbols," in *IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 3279–3284.
- [15] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
- [16] 3GPP, TR 25.814, "Physical layer aspects for evolved universal terrestrial radio access (Release 7)."
- [17] F. Rusek, "On the existence of the MIMO limit on MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1118–1121, 2009.
- [18] A. Modenini, F. Rusek, and G. Colavolpe, "Faster-than-Nyquist signaling for next generation communication architectures," in *22nd European Signal Processing Conference (EUSIPCO)*, 2014, pp. 1856–1860.
- [19] M. Yuhas, Y. Feng, and J. Bajcsy, "On the capacity of faster-than-Nyquist MIMO transmission with CSI at the receiver," in *IEEE Globecom Workshops (GC Wkshps)*, 2015, pp. 1–6.
- [20] S. Wen, G. Liu, F. Xu, L. Zhang, C. Liu, and M. A. Imran, "Ergodic capacity of mimo faster-than-nyquist transmission over triply-selective rayleigh fading channels," *IEEE Transactions on Communications*, vol. 70, no. 8, pp. 5046–5058, 2022.
- [21] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
- [22] P. Almers, F. Tufvesson, O. Edfors, and A. Molisch, "Measured capacity gain using water filling in frequency selective MIMO channels," in *13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2002, vol. 3, pp. 1347–1351.
- [23] F. Dupuy and P. Loubaton, "On the capacity achieving covariance matrix for frequency selective MIMO channels using the asymptotic approach," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 5737–5753, 2011.
- [24] A. Scaglione, "Statistical analysis of the capacity of MIMO frequency selective Rayleigh fading channels with arbitrary number of inputs and outputs," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, 2002, p. 278.
- [25] J. T. Wang, "Time-space mimo system with faster-than-Nyquist signaling in frequency-selective fading channel," *Multidimensional Syst. Signal Process.*, vol. 32, no. 3, pp. 1027–1039, Jul 2021. [Online]. Available: <https://doi.org/10.1007/s11045-021-00771-2>
- [26] T. E. Bogale, L. B. Le, X. Wang, and L. Vandendorpe, "Multipath multiplexing for capacity enhancement in SIMO wireless systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6895–6911, 2017.
- [27] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends® in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006. [Online]. Available: <http://dx.doi.org/10.1561/01000000006>
- [28] F. R. Gantmakler, *The Theory of Matrices*. New York: Chelsea Publishing Co., 1960.
- [29] P. J. Brockwell, *Time Series: Theory and Methods*, 2nd ed., ser. Springer Series in Statistics. New York: Springer-Verlag, 1991.
- [30] R. A. Horn, *Matrix Analysis*, 2nd ed. New York: Cambridge University Press, 2013.
- [31] Y. J. D. Kim and J. Bajcsy, "On spectrum broadening of pre-coded faster-than-Nyquist signaling," in *IEEE 72nd Vehicular Technology Conference - Fall*, 2010, pp. 1–5.
- [32] Y. G. Yoo and J. H. Cho, "Asymptotic optimality of binary faster-than-Nyquist signaling," *IEEE Communications Letters*, vol. 14, no. 9, pp. 788–790, 2010.
- [33] M. Ganji and H. Jafarkhani, "Novel time asynchronous NOMA schemes for downlink transmissions," 2018. [Online]. Available: <https://arxiv.org/abs/1808.08665>
- [34] J. Gutierrez-Gutierrez and P. M. Crespo, "Asymptotically equivalent sequences of matrices and Hermitian block Toeplitz matrices with continuous symbols: Applications to MIMO systems," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5671–5680, 2008.