Spatial T-H Precoding for Packet Data Systems with Scheduling

Jing Jiang, R. Michael Buehrer and William H. Tranter

Mobile and Portable Radio Research Group (MPRG)
432 Durham Hall, Mail Stop 0350
Bradley Dept. of Elec. & Comp. Engineering, Virginia Polytechnic Institute and State University
Blacksburg, VA, USA, 24061, Phone: (540)231-2927, FAX: (540)231-2968
E-mail: {jjiang, buehrer, btranter}@vt.edu

***Abstract*–This paper combines spatial Tomlinson-Harashima (T-H) precoding with packet scheduling at the transmitter to improve the system-level performance of packet data systems. Under the assumptions of instantaneous and perfect channel state information (CSI) at the transmitter and independent and identically distributed (i.i.d.) user channels with flat fading, the spatially equalized and decoupled multiuser channels are shown to be independently faded with distinct orders of diversity. The potential power gain obtained by multiuser scheduling indicates that at medium to high SNR, the average total system throughput can be greatly improved with high spatial multiplexing as compared to a single-antenna system.**

Keywords-multiple antennas; spatial multiplexing; dirty paper coding; precoding; scheduling

I. INTRODUCTION

Built on Costa's work [1], recent developments in information theory have indicated that in an interference-limited fading environment, so-called dirty paper coding (DPC) achieves the sum capacity of a multiple-antenna broadcast channel when the transmitter has full knowledge of the channel state information (CSI) to all the receivers [2], [3]. Theoretically, DPC can immunize a link against interference when the interference is from the same transmitter, (hence known to the transmitter but not to the receiver,) by making use of large dimensional lattices, signal space tessellation and system properties of modular arithmetic [4]. This concept of modulo-lattice precoding establishes the connection between DPC and Tomlinson-Harashima (T-H) precoding [5], [6], which can be viewed as a one-dimensional suboptimal implementation of DPC [7]. T-H precoding has long been applied in DSL systems for inter-symbol interference cancellation, and recently for multiuser crosstalk cancellation [8]. Due to the strict causality constraint, the T-H precoder suffers a shaping loss defined by the capacity deficiency of about 0.509 bits per complex symbol (or 1.53 dB in SNR) at high SNR, and a larger gap at low SNR. This paper extends the basic spatial Tomlinson-Harashima precoding for decentralized receivers [9] to a packet data network with packet scheduling. The goal is to evaluate the total throughput gain from T-H precoding with scheduling.

We focus on the downlink transmission with a single transmitter and distributed receivers which are not cooperating with each other at reception. We assume that user channels experience independent and identically distributed (i.i.d.) flat fading, and all the CSI is known to the transmitter perfectly and instantaneously. We also assume that the transmission time is divided into consecutive and equal time slots, with the duration of each slot being less than the fading coherence time, and much less than the possible delay constraint of data services.

II. SPATIAL T-H PRECODING

A. Spatial Implementation

Figure 1 shows a spatial zero-forcing T-H precoding (ZF-THP) system equivalent to the model in [9]. The model actually forms a distributed multiple-input and multiple-output (MIMO) system with an n_t -element antenna at the transmitter, and K receive antennas distributed across users with each user having a single-element antenna. At each time slot, the base station sends independent and synchronized data packets to n_s users simultaneously, with $n_s \le n_t$ constrained by the n_t degrees of freedom at transmitter. Let $\mathbf{H}(\mathcal{S})$ be the $n_s \times n_t$ channel matrix between the transmit antenna array and the selected n_s receive antennas defined by set $\mathcal{S}$, with its entries being i.i.d. fading coefficients of unit variance. The Hermitian matrix $\mathbf{A} = \mathbf{H}(\mathcal{S}) \cdot \mathbf{H}(\mathcal{S})^H$ is almost surely positive definite, and has a unique Cholesky factorization as $\mathbf{A} = \mathbf{L} \cdot \mathbf{L}^H$, where the $n_s \times n_s$ lower triangular matrix $\mathbf{L}$ has real positive diagonal entries. The superscript ^{H} denotes conjugate transposition. The $n_t \times n_s$ transmit matrix $\mathbf{F} = (\mathbf{L}^{-1} \cdot \mathbf{H}(\mathcal{S}))^H$ has orthonormal columns. The matrix $\mathbf{L}$ is decomposed into $\mathbf{L} = \mathbf{G} \cdot (\mathbf{B} + \mathbf{I}_{n_s})$, where $\mathbf{G}$ is a diagonal matrix with the main diagonal entries g_i of $\mathbf{L}$, $i = 1,...,n_s$, the matrix $\mathbf{B}$ is lower triangular with zero main diagonal entries, and $\mathbf{I}_{n_s}$ is the identity matrix of size n_s. At the selected receiver inputs, the spatially and temporally white complex Gaussian noises of zero mean and unit variance, $\tilde{n}_i$, $i = 1,...,n_s$, are added to signals.

The zero-forcing T-H precoder in Figure 1 uses a

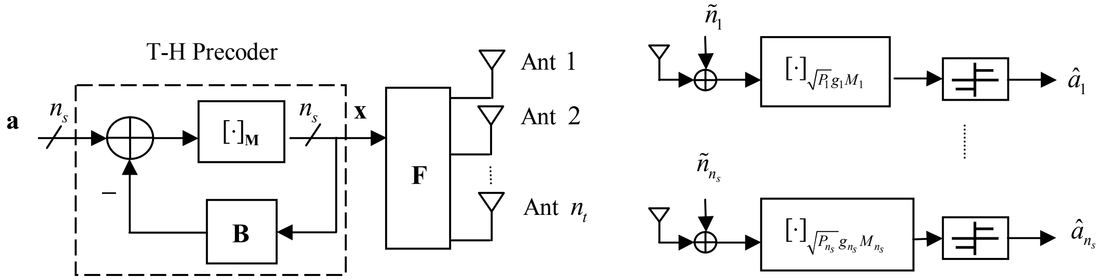


Figure 1. The block diagram of a downlink system with spatial zero-forcing T-H precoding (ZF-THP).

symmetric vector modulo operator $[\cdot]_{\mathbf{M}}$ for data streams, with the modulus vector $\mathbf{M}=[M_1,...,M_{n_s}]^T$ being set equal to the positive Voronoi boundary values [10] of input signal constellations of unit average symbol energy, possibly adaptive over each slot. For simplicity, we ignore the modulo loss [10] and assume that over each slot, the total transmit power P_T is equally split over the scheduled users with $P_i = P_T / n_s$, $i=1,...,n_s$. At scheduled receivers, the moduli M_i, $i=1,...,n_s$, of the symmetric modulo operators are weighted by $\sqrt{P_T/n_s}\cdot g_i$, $i=1,...,n_s$, which are assumed perfectly known from the transmitter. Each receiver simply performs a maximum-likelihood detection at its modulo output. In the following, we set $n_s = n_t$ for maximum spatial multiplexing gain with ZF-THP. For channel matrix $\mathbf{H}$ having i.i.d. circular symmetric complex Gaussian entries of zero mean and unit variance, Bartlett decomposition indicates that g_i^2 are central chi-square random variables with $2(n_t - i + 1)$, $i=1,...,n_t$, degrees of freedom [11].

B. Achievable Rate

It can be shown that through the ZF-THP above, the downlink channels are decoupled into n_t independently faded channels with SNRs $\rho_i = P_T g_i^2 / n_t$, $i=1,...,n_t$, and the decoupled signals at receive modulo operator outputs are

$$\begin{aligned}\tilde{r}_i &= \left[\sqrt{P_T/n_t}\, g_i \cdot a_i + \tilde{n}_i\right]_{\sqrt{P_T/n_t}\cdot g_i M_i} \\ &= [a_i + \tilde{n}_i /(\sqrt{P_T/n_t}\, g_i)]_{M_i},\ i=1,...,n_t, \qquad (1)\end{aligned}$$

Following the analysis in [12], it is not hard to derive the achievable rates for the user channels in (1) as

$$\begin{aligned}R_i^{\text{zfthp}}(P_T/n_t) &= 2\log_2(2M_i) - \\ & h\left([\tilde{n}_i/(\sqrt{P_T/n_t}\, g_i)]_{M_i}\right),\ i=1,...,n_t, \qquad (2)\end{aligned}$$

in bps/Hz. For uniformly distributed signals over a square Voronoi region $(-M_i, M_i]\times(-M_i, M_i]$ of a unit second-order moment, we have $M_i = \sqrt{3/2}$, $i=1,...,n_t$. Therefore when ignoring modulo loss and with equal transmit power allocation, the achievable rates of ZF-THP become

$$R_i^{\text{zfthp}}(P_T/n_t) = \log_2 6 - h\left([\tilde{n}_i/(\sqrt{P_T/n_t}\, g_i)]_{\sqrt{3/2}}\right),\ i=1,...,n_t, \quad (3)$$

where $\tilde{n}_i/(\sqrt{P_T/n_t}\, g_i)$, $i=1,...,n_t$ are i.i.d. zero-mean and circular symmetric complex Gaussian noise with variance $n_t/\left(P_T g_i^2\right)$, the reciprocal of the SNRs at the receive modulo inputs. It can be easily shown that at high SNR, (3) incurs a rate loss of 0.509 bits per complex symbol from that of an AWGN channel, whereas at low SNR, the loss is larger. Results on DPC indicate that with the knowledge of input SNR, THP can do better at low SNR with a noise cooling factor $\alpha = (P_T/n_t)/(1+P_T/n_t)$ [1], [4]. In Figure 1, this corresponds to replacing the matrix $\mathbf{B}$ with matrix $\tilde{\mathbf{B}} = \left((P_T/n_s)/(1+P_T/n_s)\right)\cdot\mathbf{B}$, and multiplying each receive modulo input with $(P_T/n_s)/(1+P_T/n_s)$. Correspondingly, the achievable data rates with noise cooling are

$$\begin{aligned}&\tilde{R}_i^{\text{zfthp}}(P_T/n_t) = \log_2 6 - \\ &2h\left([\tilde{\alpha}_i \cdot n/(\sqrt{P_T/n_t}\, g_i) + (1-\tilde{\alpha}_i)\cdot x]_{\sqrt{3/2}}\right),\ i=1,...,n_t, \qquad (4)\end{aligned}$$

in bps/Hz, where $\tilde{\alpha}_i = \rho_i/(1+\rho_i)$, $i=1,...,n_t$, n is a zero-mean real Gaussian random variable of variance 0.5, and x is a real random variable uniformly distributed over the interval $(-\sqrt{3/2}, \sqrt{3/2}]$. Numerical integration can be used to evaluate the differential entropies of modulo random variables in (3) and (4). Figure 2 plots the average user achievable rates in (3) and (4). As we can see in the figure, at high SNR, THP has a SNR loss of about 1.53 dB, and noise cooling increases the rate at low SNR.

With CSI at transmitter and noise cooling, the achievable sum rate of THP is then

$$\begin{aligned}\tilde{R}_{\text{sum}}^{\text{zfthp}} &= \sum_{i=1}^{n_t} \tilde{R}_i^{\text{zfthp}}(P_T/n_t) \\ &= \sum_{i=1}^{n_t}\left[\log_2 6 - 2h\left([\tilde{\alpha}_i\cdot n/(\sqrt{P_T/n_t}\, g_i) + (1-\tilde{\alpha}_i)\cdot x]_{\sqrt{3/2}}\right)\right], \qquad (5)\end{aligned}$$

for equal power allocation across users. For a network with a large number of users so that $K > n_t$, the maximum sum rate can be achieved through multiuser selection as

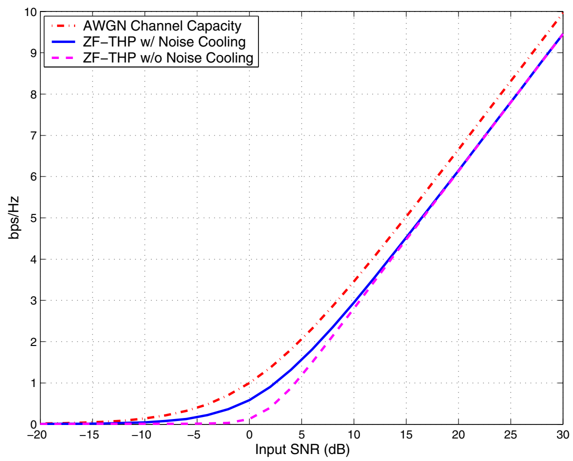


Figure 2. User achievable rate of THP.

$$\tilde{R}_{\text{sum}}^{\text{zfthp-max}} = \max_{\mathcal{S}} \tilde{R}_{\text{sum}}^{\text{zfthp}} \quad (6)$$

over all *ordered* user subsets $\mathcal{S}$ with cardinality $|\mathcal{S}| = n_t$ to obtain multiuser diversity gain. For comparison, we also give the maximum achievable sum rates of zero-forcing DPC (ZF-DPC) and zero-forcing transmit linear beamforming (ZF-LBF) as specified in [13]. For ZF-DPC with optimal transmit power allocation,

$$R_{\text{sum}}^{\text{zfdpc-max}} = \max_{\mathcal{S}} \sum_{i=1}^{n_t} \left[\log_2(\xi \cdot \rho_i^{\text{zfdpc}})\right]_+ , \quad (7)$$

where $\rho_i^{\text{zfdpc}} = g_i^2,\ i = 1,\ldots,n_t$, are the channel gains of scheduled users, and ξ is the waterfilling solution to

$$\sum_{i=1}^{n_t} \left[\xi - 1/\rho_i^{\text{zfdpc}}\right]_+ = P_T ,$$

over each slot, and the operator $[x]_+$ is defined as $[x]_+ = \max(x,0)$. The maximization in (7) is over all *ordered* user subsets $\mathcal{S}$ with cardinality $|\mathcal{S}| = n_t$. For ZF-LBF with equal power allocation,

$$R_{\text{sum}}^{\text{zflbf-max}} = \max_{\mathcal{S}} \sum_{i=1}^{|\mathcal{S}|} \log_2(1+\rho_i^{\text{zflbf}}) , \quad (8)$$

where $\rho_i^{\text{zflbf}} = (P_T/n_t)/\left[\left(\mathbf{H}(\mathcal{S})\mathbf{H}(\mathcal{S})^H\right)^{-1}\right]_{i,i}$, $i = 1,\ldots,n_t$, are the received SNRs of scheduled users, and the maximization is over all *unordered* user subsets $\mathcal{S}$ with cardinality $|\mathcal{S}| \le n_t$. For $K \ge n_t$, the total number of unordered subsets is $\sum_{i=1}^{n_t} \mathrm{C}_K^i$, with C_m^n being the number combinations of n out of m. When the channel matrix $\mathbf{H}$ has i.i.d. circular symmetric complex Gaussian entries of zero mean and unit variance, the SNRs ρ_i^{zflbf} are weighted central chi-square random variables with $2(n_t - |\mathcal{S}|+1)$ degrees of freedom [14].

III. MULTIUSER PACKET SCHEDULING

We assume that user subset selection is performed by a packet scheduler at the transmitter. We call the scheduler which maximizes the sum rate a greedy scheduler whose objective is defined as in (6), (7) or (8) for different transmission schemes. The greedy scheduler ignores fairness of time-slot allocation to users, thus a proportional fair (PF) scheduler was introduced for a balanced tradeoff between multiuser diversity and fair allocation of time slots [15]. The PF scheduler can be extended for multiuser transmission, which assigns the slot t to the user subset $\mathcal{S}^*$ satisfying

$$\mathcal{S}^* = \arg \max_{\mathcal{S}} \sum_{i=1}^{n_s} \frac{R_i(t)}{T_i(t)} \quad (9)$$

among all active users (users requesting for services simultaneously), where $R_i(t)$ is the achievable rate of user i over slot t, and $T_i(t)$ is its average throughput in a window of length T_c ending at slot t and is updated slot-wise as

$$T_i(t) = \begin{cases} (1-1/T_c)T_i(t-1) + R_i(t)/T_c, & i \in \mathcal{S}^* \\ (1-1/T_c)T_i(t-1), & \text{otherwise} \end{cases} .$$

For ZF-THP and ZF-DPC with $K \ge n_t$, a direct maximization of the sum rate has $\mathrm{P}_K^{n_t}$ total number of subsets for scheduling, i.e., the number of permutations of n_t out of K. This is a large number even for moderate values of n_t and K. Fortunately, we know from [16] that the computation of g_i only depends on users $j,\ j \le i$, henceforth we can use a suboptimal scheduler as follows. At step $k,\ k = 1,\ldots,n_t$, we select the user k^* subject to

$$k^* = \arg \max_{k \in \{1^*,\ldots,(k-1)^*\}^c} \left[\sum_{i=1^*}^{(k-1)^*} \mu_i R_i(t) + \mu_k R_k(t)\right], \quad (10)$$

where $\{\mathcal{A}\}^c$ denotes the complement of set $\mathcal{A}$, and μ_i are 1 and $1/T_i(t)$ for greedy and PF scheduler, respectively. This reduces the total number of subsets under scheduling to

$$N_1 = \sum_{i=0}^{n_t-1} \mathrm{C}_{K-i}^1 = n_t(2K - n_t + 1)/2 . \quad (11)$$

For $K >> n_t$, we have N_1 in the order of $O(K)$. This is analogous to the Viterbi algorithm, since the suboptimal scheduler approximates the global optimum through a local optimization. This principle of optimality was also noted in

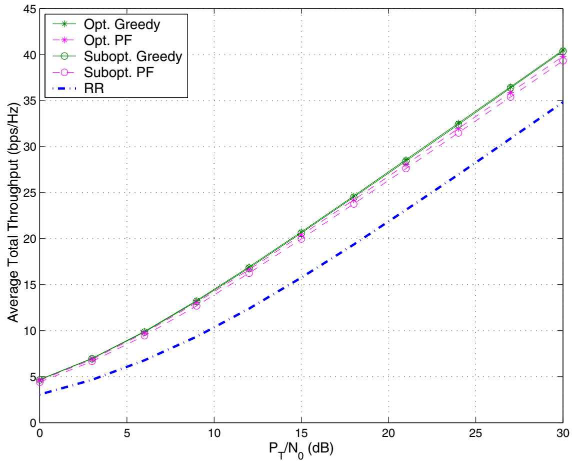


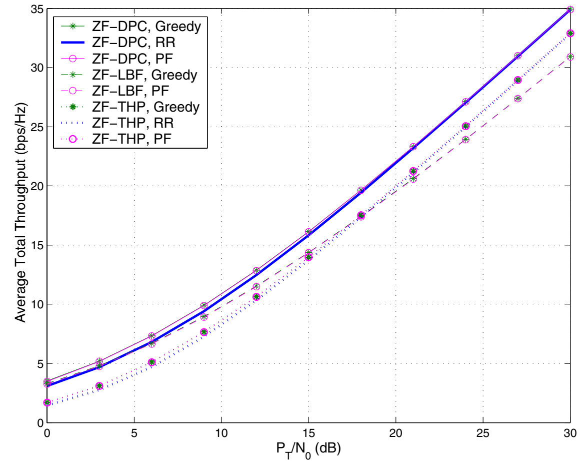
Figure 3. Average maximum sum rate of ZF-DPC for $n_t = 4$ and $K = 16$ under different packet scheduling.

the V-BLAST algorithm [17], which was proved equivalent to a generalized decision feedback equalizer [18]. Figure 3 plots the average maximum sum rate in bps/Hz for ZF-DPC with optimal power allocation, using both optimal (7) and suboptimal (10) schedulers, for $n_t = 4$ and $K = 16$ with i.i.d. flat Rayleigh fading of 5.6 Hz Doppler over 2000 time slots of 16.7 ms each. For the PF scheduler, we set $T_c = 1.67$ s. For comparison, Figure 3 also shows the average total throughput with a round-robin (RR) scheduler which assigns time slots to every subset of n_t alternately. It can be seen that the suboptimal scheduler approximates the optimum very closely over a fairly large input SNR range.

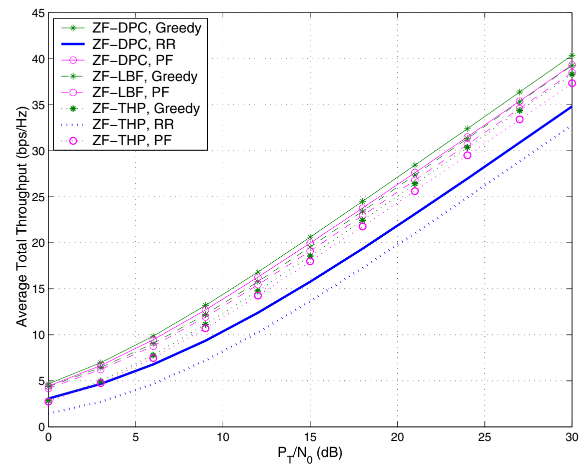
IV. Numerical Results

We present Monte Carlo simulation results for i.i.d. flat fading channels. We consider two flat fading scenarios, one is Rayleigh with 5.6 Hz Doppler as used in Figure 3, and the other is Rician with 1.5 Hz Doppler and a 10 dB Rice factor. The time slot width for both cases is set to 16.7 ms. Figures 4 and 5 depict the average maximum total throughputs of ZF-DPC, ZF-THP, and ZF-LBF vs. P_T / N_0 for $n_t = 4$ and $K = 4 \text{ and } 16$, where N_0 is the complex noise sample variance normalized to 1. The ZF-DPC with optimal power allocation is used as an upper bound to both ZF-THP and ZF-LBF with equal power allocations. Throughputs with RR scheduling provide baselines for comparison in these figures. Both ZF-DPC and ZF-THP use the suboptimal scheduler in (10), whereas ZF-LBF uses the extensive scheduler in (8). The optimal noise cooling is employed for ZF-THP. The simulation runs for 2000 slots for Rayleigh channels and 8000 slots for Rician channels. For the PF scheduler, we set $T_c = 1.67$ s.

As can be seen from these plots, on the average, more than 16 bps/Hz spectral efficiency is achievable with both ZF-LBF and ZF-THP under PF scheduling at $P_T / N_0 = 15$ dB and $K = 16$, and for $K = 16$, the average total throughputs benefit from the multiuser diversity through the channel-aware



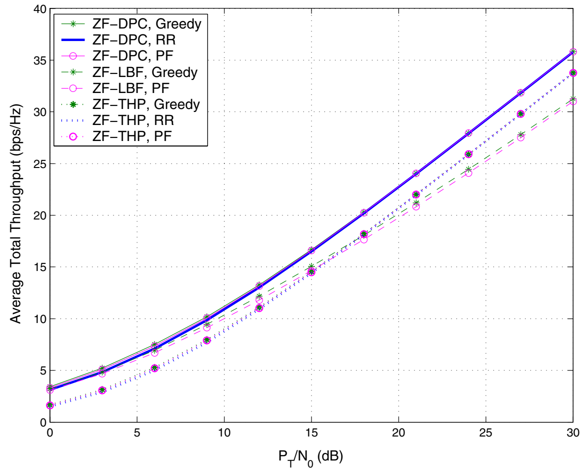
(a) $K = 4$.



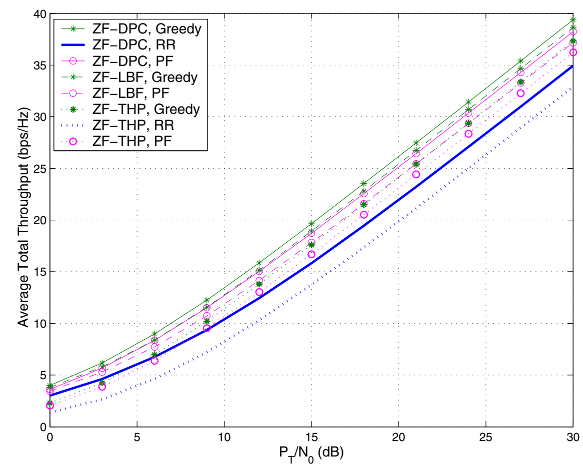
(b) $K = 16$.

Figure 4. Average maximum sum rate for $n_t = 4$ in flat Rayleigh fading.

scheduling (greedy and PF schedulers) as compared to the non-channel- aware scheduling (RR scheduler). We note that the multiuser diversity gain is more pronounced in channels with large fading variations in comparing parts (b) in Figures 4 and 5, and at high SNR, the ZF-LBF enjoys about 1 bps/Hz higher average spectral efficiency at $K = 16$. In all cases, the rate deficiency of ZF-THP, even with the optimal noise cooling, is apparent at low SNR due to the shaping loss. However, at moderate to high SNR, ZF-THP is a competitive alternative to ZF-LBF, while the latter normally requires a large transmit peak-to-rms ratio and an extensive user subset search. The advantage of ZF-THP over ZF-LBF becomes more apparent with $K = n_t$. Finally, it should be mentioned that in addition to providing spatial multiplexing, multiple transmit antennas also help to boost the multiuser diversity gain even in a slow fading environment with limited channel variations as shown in Figure 5. This is in contrast to the dumb antennas proposed in [19] without full CSI at the transmitter.



(a) $K = 4$.



(b) $K = 16$.

Figure 5. Average maximum sum rate for $n_t = 4$ in flat Rician fading.

V. CONCLUSION

We compared the average maximum total throughputs of ZF-LBF and ZF-THP with equal power allocated to scheduled users over each time slot. As a one-dimensional suboptimal implementation of ZF-DPC, ZF-THP is a competitive solution at medium to high SNR for providing high spatial multiplexing with multiuser diversity. A suboptimal scheduler was proposed for ZF-DPC and ZF-THP, which approximates the optimal scheduler closely and has a complexity which increases linearly with the number of active users. Unlike ZF-LBF, ZF-THP has the transmit outputs uniformly distributed over the Voronoi regions of signal constellations, hence a restricted transmit peak-to-rms ratio. From a system point of view, we demonstrated the performance advantage of joint optimization of physical-layer signaling and MAC-layer packet scheduling.

REFERENCES

[1] M. H. N. Costa, “Writing on dirty paper,” *IEEE Trans. on Inform. Theory*, vol. 29, pp. 439-441, May 1983.

[2] S. Vishwanath, N. Jindal, and A. Goldsmith, “On the capacity of multiple input multiple output broadcast channels,” *Proc. Int.Conf. Commun.*, Apr. 2002, pp. 1444-1450.

[3] P. Viswanath and D. Tse, “Sum capacity of the multiple antenna Gaussian broadcast channel and uplink-downlink duality,” to appear in *IEEE Trans. on Inform. Theory* (Aug. 2003). See also: *IEEE Intern. Symp. Inform. Theory, ISIT 2002*, p. 497, Lausanne, Switzerland, Jun./Jul. 2002.

[4] G. J. Foschini and A. H. Diaz, “Dirty paper coding: Perturbing off the infinite dimensional lattice limit,” to appear in *Multi-Antenna Channels: Capacity, Coding and Signal Processing*, Edited by G. J. Foschini and S. Verdu, *DIMACS Series in Discrete Mathematics and Computer Science*, AMS, Providence, RI, 2003.

[5] M. Tomlinson,”New automatic equalizer employing modulo arithmetic,” *Electron. Lett.*, vol. 7, pp. 138-139, Mar. 1971.

[6] H. Harashima and H. Miyakawa, “Matched-transmission technique for channels with intersymbol interference,” *IEEE Trans. Commun.*, vol. COM-20, pp. 774-780, Aug. 1972.

[7] U. Erez, S. Shamai, and R. Zamir, “Capacity and lattice-strategies for canceling known interference,” submitted to *IEEE Trans. on Inform. Theory*. See also *ISITA 2000*, Honolulu, Hawaii (USA), Nov. 2000, pp. 681-684.

[8] G. Ginis and J. M. Cioffi, “A multiuser precoding scheme achieving crosstalk cancellation with application to DSL systems” in *34th Asilomar Conf. on Signals, Systems and Computers*, vol. 2, Pacific-Grove, CA, USA, Nov. 2000, pp. 1627-1631.

[9] R. F. H. Fischer, et al., “MIMO precoding for decentralized receivers,” in *Proc. of ISIT’02*, Lausanne, Switzerland, Jun./Jul. 2002, p. 496.

[10] G. D. Forney and M. V. Eyuboglu, “Combined equalization and coding using precoding,” *IEEE Commun. Mag.*, pp. 25-34, Dec. 1991.

[11] R. J. Muirhead, Aspects of Multivariate Statistical Theory, New York: Wiley, 1982.

[12] R. D. Wesel and J. M. Cioffi, “Achievable rates for Tomlinson-Harashima precoding,” *IEEE Trans. on Inform. Theory*, vol. 44, no. 2, pp. 824-831, Mar. 1998.

[13] G. Caire and S. Shamai, “On the achievable throughput of a multi-antenna Gaussian broadcast channel,” to appear in *IEEE Trans. Inform. Theory*, preprint in Nov. 2002.

[14] D. A. Gore, R. W. Heath, and A. J. Paulraj, “Transmit selection in spatial multiplexing systems,” *IEEE Commun. Lett.*, vol. 6, no. 11, pp. 491-493, Nov. 2002.

[15] A. Jalali, R. Padovani, and R. Pankaj, “Data throughput of CDMA-HDR: A high efficiency-high data rate personal communication wireless system,” *IEEE 51st Vehicular Technology Conference*, Tokyo, Japan, Vol. 3, May 2000, pp. 1854-1858.

[16] G. H. Golub, C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore: John Hopkins Univ. Press, 1996.

[17] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, “V-BLAST: An architecture for achieving very high data rates over the rich-scattering wireless channel,” in *Proc. ISSSE-98*, Pisa, Italy, Sept. 1998, pp. 295-300.

[18] G. Ginis and J. M. Cioffi, “On the relation between V-BLAST and the GDFE,” *IEEE Commun. Lett.*, vol. 5, pp. 364-366, Sept. 2001.

[19] P. Viswanath, D. N. C. Tse and R. Laroia, “Opportunistic beamforming using dumb antennas,” *IEEE Trans. Inform. Theory*, vol. 48, pp.1277-1294, June 2002.