MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Joint Pilot and Data Loading Technique for Mimo Systems Operating with Covariance Feedback

Digham, F.; Mehta, N.; Molisch, A.; Zhang, J.

TR2004-105   October 2004

Abstract

We consider a multiple input multiple output (MIMO) system wherein channel correlation as well as noisy channel estimation are both taken into account. For a given block fading duration, pilot-assisted channel estimation is carried out with the aid of a minimum mean square error estimator. The pilot sequences and data are jointly optimized to maximize the ergodic capacity while fulfilling a total transmit energy constraint. Using the optimization results, we study the effect of various system parameters and different types of channel knowledge. We show that systems with channel covariance knowledge and imperfect channel estimates can even outperform those with perfect channel estimation at the receiver, but without any channel knowledge at the transmitter. Further, we show that the signal to noise ratio and the block fading duration determine the qualitative behavior of ergodic capacity as a function of antenna spacing or angular spread. Finally, we find that the optimum pilot-to-data power ratio is always greater than one. It increases as the total transmit power decreases or as the block fading length increases.

3G 2004

JOINT PILOT AND DATA LOADING TECHNIQUE FOR MIMO SYSTEMS OPERATING WITH COVARIANCE FEEDBACK

Fadel F. Digham[†,*], Neelesh B. Mehta[‡], Andreas F. Molisch[‡,+], and Jin Zhang[‡]
[†]Univ. of Minnesota, Minneapolis, MN, USA. fadel.digham@ieee.org
[‡]Mitsubishi Electric Research Labs, 201 Broadway, Cambridge, MA, USA. {mehta, molisch, jzhang}@merl.com

Index Terms— **MIMO systems, Antenna arrays, Fading channels, Training, Information Rates**

Abstract— **We consider a multiple input multiple output (MIMO) system wherein channel correlation as well as noisy channel estimation are both taken into account. For a given block fading duration, pilot-assisted channel estimation is carried out with the aid of a minimum mean square error estimator. The pilot sequences and data are jointly optimized to maximize the ergodic capacity while fulfilling a total transmit energy constraint. Using the optimization results, we study the effect of various system parameters and different types of channel knowledge. We show that systems with channel covariance knowledge and imperfect channel estimates can even outperform those with perfect channel estimation at the receiver, but without any channel knowledge at the transmitter. Further, we show that the signal to noise ratio and the block fading duration determine the qualitative behavior of ergodic capacity as a function of antenna spacing or angular spread. Finally, we find that the optimum pilot-to-data power ratio is always greater than one. It increases as the total transmit power decreases or as the block fading length increases.**

I. INTRODUCTION

While second generation cellular systems were intended for speech communications as well as some rudimentary data services, third-generation (3G) systems distinguish themselves by providing data communications with high rates. The original specifications of 3GPP (Third-Generation Partnership Project) called for data rates up to 2 Mbps, which is easily achievable in the allocated 5 MHz bandwidth. However, higher system capacity and higher data rates are now desired. Given the scarce system bandwidth, these can only be achieved by transmission schemes with high spectral efficiency such as Multiple input multiple output (MIMO) [5], [13], [14]. Furthermore, algorithms, testbeds, and prototypes have demonstrated the practical viability of MIMO [2], [6]. Therefore, MIMO technology is now an active work item in 3GPP standardization efforts [1].

An important aspect of MIMO systems is the use of channel state information (CSI) at the transmitter (CSIT). In the frequency division duplex (FDD) mode of operation, instantaneous CSIT has to be obtained by feedback from the receiver. While only one transmit-receive antenna link estimate needs to be fed back in conventional non-MIMO systems, the same needs to be done for all transmit-receive antenna pair links in MIMO. Therefore, the feedback burden is considerable in MIMO systems, and, if not optimized, can adversely affect the overall spectral efficiency of the system. Even in the time division duplex (TDD) mode of operation, which does not require explicit feedback, the instantaneous CSIT acquired from reverse link transmissions may be outdated during the transmission phase, especially when the mobile station moves fast. For this reason, a great deal of interest has been paid to systems that exploit statistical or covariance knowledge [8], [12] at the transmitter – denoted henceforth as CovKT. Such systems have the advantage of requiring a much lower rate of feedback, or can eliminate it altogether.

While theoretical investigations of MIMO systems almost always assume that the CSI at the receiver (CSIR) is perfect, this assumption cannot be fulfilled in practical 3G systems. The noise present during the channel estimation phase will render the receiver CSI (CSIR) imperfect, and can appreciably decrease the capacity. Most previous studies that took imperfect CSIR into account were restricted to the case of spatially white channels. For example, [7] first derived a lower bound to a MIMO system capacity. This paper considered pilot-aided channel estimation for a block fading spatially white wireless channel and derived the optimum training sequence, pilot-to-data power ratio, and training duration. One paper that studied data transmission for correlated MIMO channels with estimation error is [15]. However, it modeled the estimation error in an ad hoc manner.

An obvious way to improve CSIR is to increase the power of the pilots or the training duration. However, for systems with a fixed power budget and a given channel coherence time, doing so reduces both the power and time allowed for data transmission. The trade-off between pilots and data is thus of great practical interest.

In this paper, we discuss the use of optimum pilot and data sequences, pilot duration, and pilot-to-data power ratio in MIMO systems operating in a spatially correlated channel when the transmitter has access to only covariance information. The receiver obtains its CSIR from noisy pilot sequences by means

*The author was with Mitsubishi Electric Research Labs during the course of this work.
+A. F. Molisch is also at the Department of Electroscience, Lund University, Lund, Sweden.

of a simple minimum mean square error (MMSE) channel estimator. The main focus in the current paper is to highlight the implications on the link performance and study the effect of different system parameters.

This paper is organized as follows. Section II sets up the system model and the notation used in the paper. Section III gives a concise description of the system and lists the key theoretical results, the details of which can be found in [4], [11]. The main results of the paper can be found in Sec. IV, which provides numerical examples and discusses the impact on the system performance of different system parameters such as block fading duration, the channel correlation, transmit power, etc. Concluding remarks are given in Section V.

II. System Model

We consider an $N_t \times N_r$ MIMO system with N_t transmit and N_r receive antennas. A block fading frequency-flat channel model [10], in which the channel remains constant for T time instants (symbol durations) and decorrelates thereafter is assumed. T corresponds to the coherence interval of the channel. Of the T time instants, T_p are used for transmitting pilots, and the remaining $T_d = T - T_p$ for data. P_p and P_d denote the power allocated to pilots and data, respectively. We shall use the subscripts p and d for variables related to pilots and data, respectively. Lower and upper case boldface letters shall be used to denote vectors and matrices, respectively. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix, $(.)^\dagger$ the Hermitian transpose, $\text{Tr}\{.\}$ the trace, and $\mathbb{E}_X$ is the expectation over X.

A. Channel Model

The $N_r \times N_t$ matrix $\mathbf{H} = [h_{ij}]$ denotes the instantaneous channel state, where the element h_{ij} is the complex fading gain from transmit antenna j to receive antenna i. Experimental results have demonstrated the validity of the Kronecker model for several typical channels [9]. $\mathbf{H}$ is then given by

$$\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t^{1/2}, \tag{1}$$

where $\mathbf{R}_t$ and $\mathbf{R}_r$ are the transmit and receive antenna correlation matrices, respectively. $\mathbf{R}_t$ ($\mathbf{R}_r$) depends on the mean angle of departure (arrival), θ, the rms angular spread, σ_θ, and the relative antenna spacing, d_t, [3]. The matrix $\mathbf{H}_w$ is spatially white, i. e., its entries are zero-mean, independent, complex Gaussian random variables with unit variance. Furthermore, we assume that the receiver is in a rich scattering environment, as is typically the case in the downlink of a cellular system. Therefore, $\mathbf{R}_r = \mathbf{I}_n$. $\mathbf{R}_t$ is full ranked.

B. Training and Data Transmission Phases

The signal received during the entire training phase of duration T_p instants, is an $N_r \times T_p$ matrix, $\mathbf{Y}_p = [y_{ij}]$, where y_{ij} is the signal received by receive antenna i at time instant j. $\mathbf{Y}_p$ is given by

$$\mathbf{Y}_p = \mathbf{H}\mathbf{X}_p + \mathbf{W}_p, \tag{2}$$

where $\mathbf{X}_p = [x_{ij}]$ is the transmitted pilot matrix of size $N_r \times T_p$ and is known a priori at the receiver. Here, x_{ij} is the signal transmitted from antenna i at time j. $\mathbf{W}_p$ is the spatially and temporally white noise matrix, defined in a similar manner; with its entries having variance σ_w^2. $\mathbf{Q}_p = \mathbf{X}_p \mathbf{X}_p^\dagger$ denotes the pilot covariance matrix.[1]

The received vector, $\mathbf{y}_d$, at any given time instant is related to the transmitted signal vector, $\mathbf{x}_d$, by

$$\mathbf{y}_d = \mathbf{H}\mathbf{x}_d + \mathbf{w}_d, \tag{3}$$

where $\mathbf{w}_d$ is the spatially white and temporally uncorrelated noise vector. The vectors $\mathbf{y}_d$, $\mathbf{x}_d$, and $\mathbf{w}_d$ are of dimensions $N_r \times 1$, $N_t \times 1$, and $N_r \times 1$, respectively. $\mathbf{Q}_d = \mathbb{E}_{\mathbf{x}_d}[\mathbf{x}_d \mathbf{x}_d^\dagger]$ denotes the data covariance matrix.

C. MMSE Channel Estimator

Given the second-order statistics of the channel and the pilot sequence $\mathbf{X}_p$, the MMSE channel estimator generates the channel estimate, $\widehat{\mathbf{H}}$, by passing the received $\mathbf{Y}_p$ through a deterministic matrix filter. For $\mathbf{R}_r = \mathbf{I}_{N_r}$, it can be shown that [4]

$$\widehat{\mathbf{H}} = \widetilde{\mathbf{H}}_w \widetilde{\mathbf{R}}_t^{1/2}, \tag{4}$$

where $\widetilde{\mathbf{H}}_w$ is spatially white with its entries having unit variance and $\widetilde{\mathbf{R}}_t = \mathbf{R}_t \mathbf{X}_p (\mathbf{X}_p^\dagger \mathbf{R}_t \mathbf{X}_p + \sigma_w^2 \mathbf{I}_{T_p})^{-1} \mathbf{X}_p^\dagger \mathbf{R}_t$.

The channel estimation error is defined as $\boldsymbol{\Delta} = \mathbf{H} - \widehat{\mathbf{H}}$. From (3), it follows that the data transmission phase is governed by $\mathbf{y}_d = \widehat{\mathbf{H}}\mathbf{x}_d + \boldsymbol{\Delta}\mathbf{x}_d + \mathbf{w}_d$. A lower bound, C_L, on the capacity, based on a receiver that treats the term $\boldsymbol{\Delta}\mathbf{x}_d + \mathbf{w}_d$ as noise, can be shown to be [4]

$$C_L = \left(1 - \frac{T_p}{T}\right) \mathbb{E}_{\widetilde{\mathbf{H}}_w} \log_2 \left| \mathbf{I}_{N_t} + \frac{\widetilde{\mathbf{H}}_w^\dagger \widetilde{\mathbf{H}}_w \mathbf{Q}_d}{\sigma_w^2 + \sigma_l^2} \right|, \tag{5}$$

where $\sigma_l^2 = \text{Tr}\left\{\mathbf{Q}_d(\mathbf{R}_t - \widetilde{\mathbf{R}}_t)\right\}$ is the noise term due to imperfect estimation. Here, $\left(1 - \frac{T_p}{T}\right)$ is the penalty factor due to training because no data transmission occurs then.

III. Optimal Joint Pilot and Data Loading

The optimum pilot and data design that maximizes C_L, subject to a total energy constraint $P_p T_p + P_d T_d = PT$, satisfies the following properties [4]:

1) The eigenspaces of both pilot and data covariance matrices, $\mathbf{Q}_d$ and $\mathbf{Q}_p$, should be the same as that of the channel transmit covariance matrix, $\mathbf{R}_t$.
2) The ranks of both pilot and data covariance matrices should match. They shall be denoted by k.
3) The optimum training interval, T_p, also equals k. This result also implies that $T_p \leq \min(N_t, N_r)$.

[1]Given that the pilot $\mathbf{X}_p$ is a deterministic matrix, no expectation operator is used for defining $\mathbf{Q}_p$.

Here, $\text{Tr}\{\mathbf{Q}_d\} = P_d$, $\text{Tr}\{\mathbf{Q}_p\} = P_pT_p$, and P is the total power budget available at the transmitter. With the above results, the capacity maximization problem reduces to a numerical search over the eigenvalues of $\mathbf{Q}_p$ and $\mathbf{Q}_d$. A considerably simpler power allocation strategy that distributes the pilot power uniformly among the k non-zero eigenmodes of $\mathbf{Q}_p$ and the data power uniformly among the k non-zero eigenmodes of $\mathbf{Q}_d$ results in near-optimal performance [4].

IV. Numerical Examples and Discussions

Numerical results presented in this section are for uniform linear arrays at the transmitter and the receiver. The mean angle of departure is $\theta = 45°$ and the noise variance is $\sigma_w^2 = 1$. We also assume a Gaussian shape for the angular spectrum, which is characterized by its rms angular spread, σ_θ.

We first focus on the $N_t = N_r = 4$ system.

A. Impact of Block Fading Duration (T)

Figure 1 plots the ergodic capacity as a function of the block fading duration T for two angular spreads, $\sigma_\theta = 10°$ and $45°$ and for $P = 5$ dB and 15 dB ($d_t = 0.5$). For both angular spreads, C_L increases as T increases. This is because the impact of the training penalty $\left(1 - \frac{T_p}{T}\right)$ diminishes as T increases.

B. Impact of Spatial Correlation (σ_θ and d_t)

Changing the angular spread affects the spatial correlations between the transmit antenna elements. As σ_θ increases, the correlations between the transmit antenna elements decrease and the MIMO channel changes from highly correlated to spatially white. Increasing the relative antenna spacing, d_t, also has a similar effect [3]. Figure 2 plots C_L as a function of σ_θ for different total transmit power values.[2] It is clear that the power budget, P, affects the qualitative behavior of C_L with respect to σ_θ. While C_L increases monotonically with σ_θ for $P = 20$ dB and 30 dB, it decreases monotonically with σ_θ for $P = 0$ dB. For an intermediate value of $P = 10$ dB, C_L even reaches a maximum value at $\sigma_\theta = 15°$.

We delve into this further in the equivalent Figures 3 and 4 that plot C_L as a function of d_t for $P = 5$ dB and 15 dB, $\sigma_\theta = 10°$ and $45°$, and $T = 10$ and 100. The behavior of the capacity as a function of the antenna spacing is governed by several effects. As the antenna spacing increases: (i) the number of used eigenmodes and, thus, the number of data streams increases, (ii) the eigenvalue spread decreases, which increases the strength of the unused (weaker) *channel* eigenmodes and decreases the strength of the used channel eigenmodes. This can decrease capacity. At the same time, a smaller eigenvalue spread also increases capacity as it leads to a better distribution of power among the used eigenmodes.

[2]As the noise and channel coefficients are assumed to have unit variance, the total transmit power is equivalent to the average received signal to noise ratio (SNR).

The relative strength of the different effects determines the different observed behaviors of capacity as a function of the antenna spacing. For smaller P, only the largest eigenmode ($k = 1$) is used [8]. As mentioned, decreasing d_t increases the largest eigenmode and improves capacity. On the other hand, multiple eigenmodes are used for larger P. Therefore, as d_t increases, the eigen spread decreases and the capacity increases. Notice that T affects the number of eigenmodes used, and thus, impacts the threshold values of P at which the above two different trends occur.

C. Optimal Pilot and Data Power Allocation

Figure 5 plots the optimal allocation of power between the pilots and data, $\alpha = P_p/P_d$, as a function of P for $\sigma_\theta = 10°$ and $45°$. The figure contains two sub-plots corresponding to $T = 10$ and $T = 100$. It can be seen that α decreases as P increases. However, α is always greater than 1, which means that the transmitted (pilot) power during training must be greater than the power transmitted during data communication. Note that this does not imply any ordering on the pilot and data energies. The discontinuities in α occur when the number of eigenmodes used for transmission, k, increases. It can be seen that α is significantly greater for $T = 100$ than for $T = 10$. This can be explained as follows. Even when the total energy budget, PT, increases as T increases, T_p cannot exceed N_t. Therefore, P_p increases instead. In summary, k and T are the principal parameters that affect α; given these, it is mostly insensitive to σ_θ and P.

D. Combined Impact of CSIT and CSIR

We now proceed to evaluate the impact of the CSIR and CSIT assumptions on the ergodic capacity of MIMO systems. We compare C_L, the capacity achieved by the system that jointly designs the pilot and data given CovKT and pilot-aided imperfect estimation at the receiver, with systems in which channel knowledge at the transmitter is not exploited or is unavailable and with perfect or imperfect CSIR. Note that if perfect CSIR is assumed, no resources need to be wasted on pilots ($P_p = 0$). Without any CSIT, T_p needs to be N_t [7]. Specifically, we compare C_L with the ergodic capacity of the following systems: no CSIT and imperfect pilot-aided estimation at the receiver ($T_p = N_t$), no CSIT and perfect CSIR ($T_p = N_t$), CovKT and perfect CSIR ($T_p = 0$), perfect instantaneous CSIT and perfect CSIR ($T_p = 0$). The last scheme above is the classic water-filling over space solution. For better understanding the role of the training penalty, a system with CovKT and perfect CSIR (with $P_p = 0$) that still trains for a duration equal to the number of modes used for transmission is also considered.

Figure 6 compares the above for a 2×4 system for $\sigma_\theta = 10°$, $d_t = 0.5$, and $T = 10$. When the CSIR is perfect and $T_p = 0$, we observe that CovKT performs as well as instantaneous CSIT. For imperfect CSIR, we find that at higher P values, when all the eigenmodes are in operation, the capacities with CovKT and no CSIT become the same.

Figure 7 compares the same for a 4×1 transmit diversity system. The capacity with instantaneous CSIT is better than that with CovKT by 0.5 bits/sec/Hz for all P. This is unlike the 2×4 system. Figure 8 [11] deals with a 4×4 system. There is one key difference between the 2×4 case and the 4×1 and 4×4 cases – the capacity with imperfect CSIR (but, with CovKT) exceeds that with perfect CSIR (without CSIT). As expected, in all the three cases, for the same CSIT assumptions, the capacity with perfect CSIR is always greater that that with imperfect CSIR.

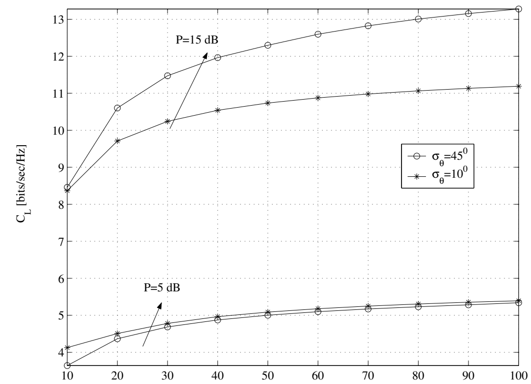

Fig. 1: The impact of the block fading length, T, on C_L for $P = 5$ dB and 15 dB ($N_t = N_r = 4$, $d_t = 0.5$).

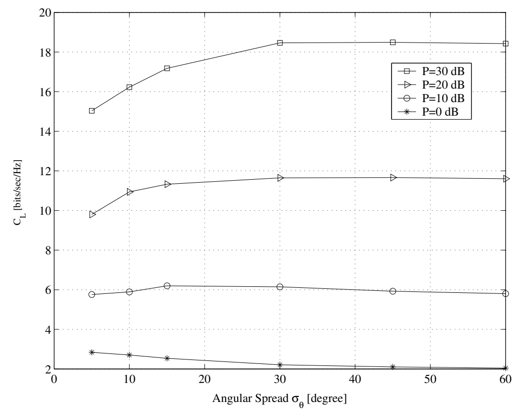

Fig. 2: Impact of the angular spread, σ_θ, on C_L ($N_t = N_r = 4$, $d_t = 0.5$, $T = 10$).

V. Conclusions

In this paper, we considered a MIMO system with channel covariance knowledge at the transmitter and noisy CSI at the receiver, and operating in a spatially correlated channel. The channel state is estimated at the receiver via an MMSE estimation filter using a priori defined pilot sequences. Using jointly optimized pilot and data sequences, we studied the effect of different system parameters on the ergodic capacity. We compared the ergodic capacity of systems with different types of CSIT (instantaneous, covariance, and no CSIT) and either perfect or imperfect CSIR. We showed that systems with covariance knowledge and imperfect CSIR can achieve higher ergodic capacity than systems with no CSIT, even if they have perfect CSIR. The capacity always increased with the (block) fading duration. The capacity, as a function of the antenna spacing, can show a monotonic increase, a monotonic decrease, or a local maxima. We also found that the training duration should be kept small, while the power assigned to the pilot symbols should be higher than that assigned to the data symbols; the optimum power values also depend on the number of used eigenmodes and the block fading duration. Our results thus motivate exploiting covariance knowledge at a multiple antenna transmitter, and provide fundamental guidelines for the optimized design of 3G MIMO systems.

References

[1] "Multiple input multiple output (MIMO) antennae in UTRA," TR 25.876, 3rd Generation Partnership Project, 2004.

[2] A. Adjoudani *et al.*, "Prototype experience for MIMO BLAST over third-generation wireless system," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 440–451, Apr. 2003.

[3] D. Asztely, "On antenna arrays in mobile communication systems: Fast fading and GSM base station receiver algorithms," Tech. Rep. IR-S3-SB-9611, Royal Institute of Technology, 1996.

[4] F. F. Digham, N. B. Mehta, A. F. Molisch, and J. Zhang, "Joint pilot and data loading for MIMO systems with covariance knowledge," Submitted to *IEEE Trans. Wireless Commun.*, 2004.

[5] G. J. Foschini and M. J. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, 1998.

[6] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the V-BLAST space-time communication architecture," *Electronics Letters*, pp. 14–15, Jan. 1999.

[7] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, pp. 951–963, 2003.

[8] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," To appear in *IEEE Trans. Wireless Commun.*, 2004.

[9] J. P. Kermoal *et al.*, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Select. Areas Commun.*, pp. 1211–1226, 2002.

[10] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, January 1999.

[11] N. B. Mehta, F. F. Digham, A. F. Molisch, and J. Zhang, "Rate of MIMO systems with CSI at transmitter and receiver from pilot-aided estimation," To appear in *Proc. VTC*, Sept. 2004.

[12] S. H. Simon and A. L. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 406–417, April 2003.

[13] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, pp. 585–595, 1999.

[14] J. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE Trans. Commun.*, vol. 5, pp. 871–878, Jun. 1987.

[15] T. Yoo, E. Yoon, and A. Goldsmith, "MIMO capacity with channel uncertainty: Does feedback help?," Submitted to *Globecom*, 2004.

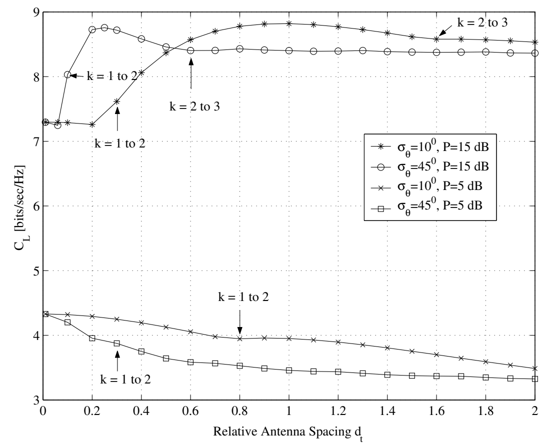

Fig. 3: Impact of the relative antenna spacing, d_t, on C_L for $T = 10$ ($N_t = N_r = 4$, $P = 15$ dB and 5 dB).

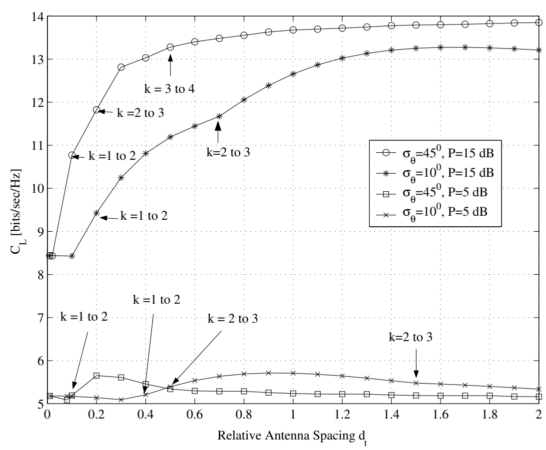

Fig. 4: Impact of the relative antenna spacing, d_t, on C_L for $T = 100$ ($N_t = N_r = 4$, $P = 15$ dB and 5 dB).

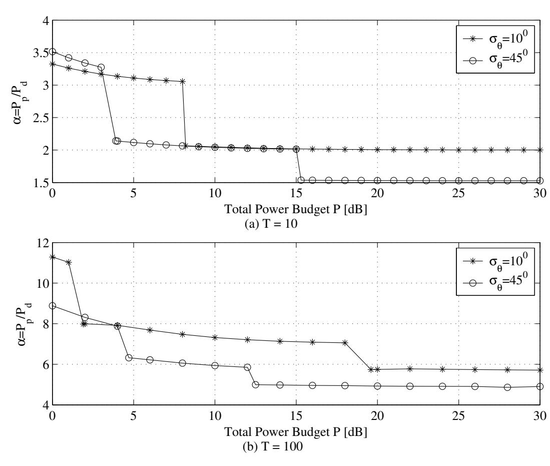

Fig. 5: Power allocation ratio, α, versus total power budget, P, for $T = 10$ and 100 ($N_t = N_r = 4$).

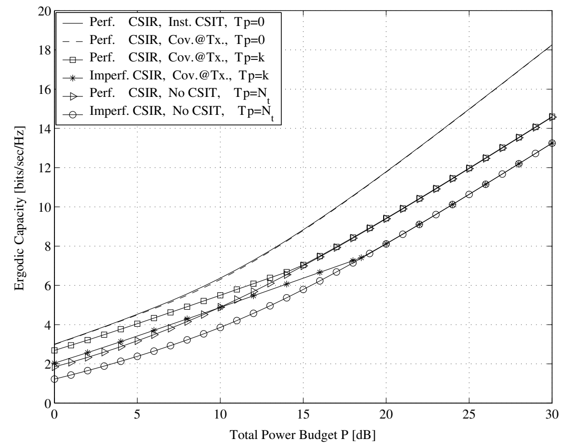

Fig. 6: Ergodic capacity comparison of systems with different CSIT and CSIR assumptions for $N_t = 2$ and $N_r = 4$. ($T = 10$, $\sigma_\theta = 10°$).

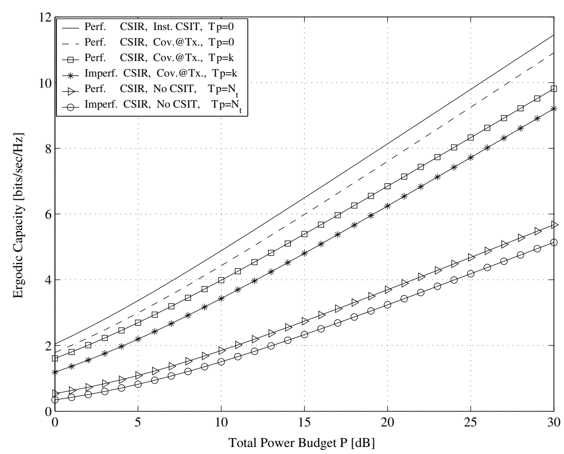

Fig. 7: Ergodic capacity comparison of systems with different CSIT and CSIR assumptions for $N_t = 4$ and $N_r = 1$. ($T = 10$, $\sigma_\theta = 10°$).

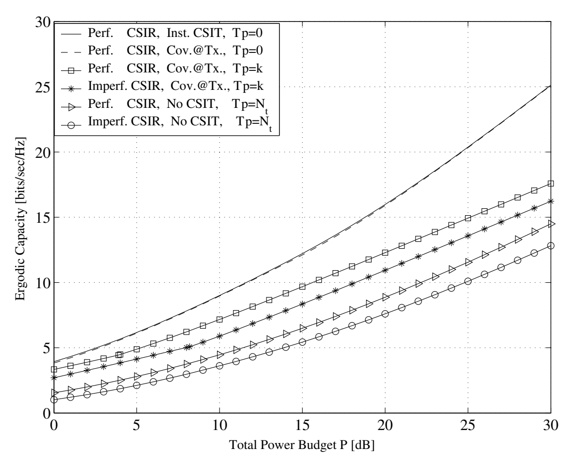

Fig. 8: Ergodic capacity comparison of systems with different CSIT and CSIR assumptions for $N_t = 4$ and $N_r = 4$, ($T = 10$, $\sigma_\theta = 10°$) [11].