STATISTICAL MODELLING OF SPATIALLY CORRELATED MIMO CHANNELS

I. Viering, M. Reinhardt and T. Frey

Siemens AG
Lise-Meitner-Strasse 13
89081 Ulm, Germany
Phone:+49-731-9533-1132
Fax:+49-731-9533-1217
Ingo.Viering.GP@icn.siemens.de

Siemens AG
Lise-Meitner-Strasse 13
89081 Ulm, Germany
Phone:+49-731-9533-1264
Fax:+49-731-9533-1217
Markus.Reinhardt@icn.siemens.de

Siemens AG
Lise-Meitner-Strasse 13
89081 Ulm, Germany
Phone:+49-731-9533-1221
Fax:+49-731-9533-1217
Thomas.Frey@icn.siemens.de

ABSTRACT

A general statistical model for correlated MIMO channels is described. It will be valid for all kinds of algorithms using multiple antennas. The channel correlations are derived for the case of a flat fading environment. It is shown, that these correlations can be created in a statistical model by simply multiplying spatially uncorrelated channel matrices with the square roots of two covariance matrices. The result is generalized to the frequency selective case. In the following this model is used in Monte-Carlo simulations to compute MIMO capacities depending on the angular spread and the element spacing at the receive and transmit antenna array.

1. INTRODUCTION

The usage of multiple antennas improves the performance of wireless communication systems. To exploit this additional dimension, different methods are possible.
Diversity techniques [1] assume low antenna correlations to increase the diversity order, i.e. to steepen the BER curves.
In contrast, beamforming techniques [2] make use of spatial correlations. Beamforming generates an energy gain by coherently summing the antenna signals, and, with smart processing of the antenna signals, they allow for adaptive interference suppression, which indirectly increases the system capacity.
If multiple antennas are used on both sides, the Shannon capacity can be directly increased by transmitting different information streams over the antennas in parallel [3][4].
For all cases, the spatial channel correlations play a very important role. The aim of this paper is to develop a simple statistical model for Monte-Carlo simulations, that reproduce these correlations. It is organized as follows.
Section 2 briefly introduces the general MIMO system model. For the flat fading case, the spatial channel correlations are derived in section 3 and used for the development of the statistical model in section 4. In section 5, the result is generalized to the frequency selective case. Section 6 gives an example, where the model is employed to calculate MIMO capacities in Monte-Carlo simulations. Finally, the main aspects are summarized.
The notation will be as follows: matrices (e.g. $\mathbf{H}, \mathbf{L}$) are written in capital bold letters, vectors (e.g. $\vec{x}, \vec{y}$) in small letters with an arrow, $\mathbf{E}\{\}$ is the expectation, $\mathrm{diag}(\vec{x})$ is a diagonal matrix with $\vec{x}$ on the main diagonal, $\mathrm{vec}(\mathbf{H})$ is a vector with stacked columns of $\mathbf{H}$, and the superscripts ^{T} and ^{H} denote the transpose and the conjugate transpose of a vector or a matrix, respectively.

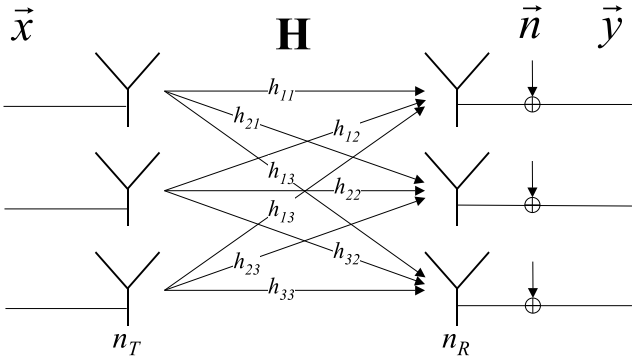

Figure 1: Structure of MIMO channel

2. SYSTEM MODEL

Figure 1 shows the general structure of a MIMO system. The transmitter uses n_T antennas to transmit the n_T×1 symbol vector $\vec{x}$. Let us first consider non-dispersive channels. We will extend the results to the frequency selective case in section 5. Each channel between the transmit and receive antennas is described by a complex coefficient h_{ij} where i denotes the receive antenna index and j denotes the transmit antenna index. The receiver consists of n_R antennas.

A reasonable assumption is, that the received signals are synchronous, as the symbol duration usually is very long compared with the delay differences, and provided perfect receiver synchronization.
In this case, the received n_R×1 symbol vector $\vec{y}$ can be expressed as

$$\vec{y} = \mathbf{H} \cdot \vec{x} + \vec{n} \tag{1}$$

where $\mathbf{H}$ is the n_R×n_T channel matrix containing the coefficients h_{ij} and $\vec{n}$ is an n_R×1 additive noise vector, which is not considered here.
Under the assumption of temporal stationarity [5], we may consider one temporal channel snapshot only. The obtained results are valid for correlated variations over the observation time. The temporal correlations have been widely investigated for many years [6] and therefore are not considered here.

3. CORRELATIONS IN FLAT MIMO CHANNELS

The channel coefficients h_{ij} are composed of a large number n_P of temporally unresolvable subpaths, which interfere at the receive antenna elements. Figure 2 depicts three of these subpaths for arbitrarily oriented receive and transmit antenna arrays.

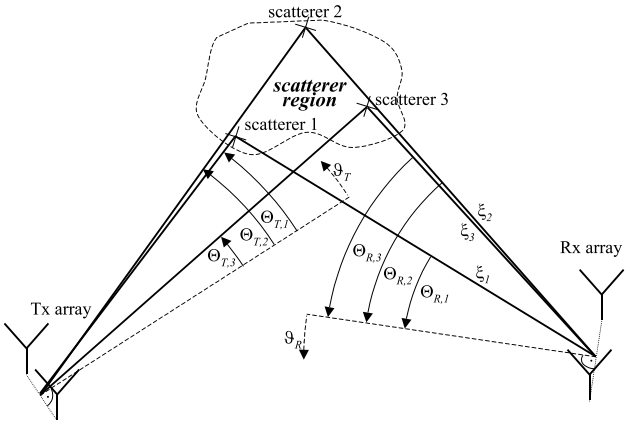

Figure 2: Subpath Model

Each subpath i is associated with a complex and random reference amplitude ξ_i, a direction of departure $\Theta_{T,i}$ (DOD) and a direction of arrival $\Theta_{R,i}$ (DOA).
The subpaths are caused by scatterers, which are concentrated in a scatterer region. Figure 2 shows, that a certain DOD $\Theta_{T,i}$ can cause any DOA $\Theta_{R,i}$ within an angular interval (angular spread). Therefore, we consider the DOAs and the DODs to be independent.

Assuming a large distance between the arrays and scattering objects compared with the array apertures (far-field assumption), and a small bandwidth compared with the carrier frequency (narrow-band assumption), the subpaths can be treated as planar waves. In this case, each subpath impinges at each antenna element under the same DOD or DOA, respectively, and all array elements receive the same complex subpath amplitudes, but with a fixed phase shift.
The channel matrix $\mathbf{H}$ can be written as [7]

$$\mathbf{H} = \mathbf{A}_R \cdot \mathrm{diag}\left(\vec{\xi}\right) \cdot \mathbf{A}_T^T \tag{2}$$

The n_P×1 vector $\vec{\xi}$ contains the complex and random amplitudes ξ_i. The entries in the n_T×n_P matrix $\mathbf{A}_T$ and the n_R×n_P matrix $\mathbf{A}_R$ represent the phase shifts corresponding to the DODs $\Theta_{T,i}$ and DOAs $\Theta_{R,i}$, respectively:

$$\begin{aligned} \mathbf{A}_T &= \left(\vec{a}_T\left(\Theta_{T,1}\right), \vec{a}_T\left(\Theta_{T,2}\right), ..., \vec{a}_T\left(\Theta_{T,n_P}\right)\right) \\ \mathbf{A}_R &= \left(\vec{a}_R\left(\Theta_{R,1}\right), \vec{a}_R\left(\Theta_{R,2}\right), ..., \vec{a}_R\left(\Theta_{R,n_P}\right)\right) \end{aligned} \tag{3}$$

where $\vec{a}_T(\Theta)$ and $\vec{a}_R(\Theta)$ denotes the array response vectors in look direction Θ at the transmitter or receiver, respectively.
Now, we need an expression for the statistical correlations between the entries in the channel matrix $\mathbf{H}$. A compact form of these correlations is obtained when the vec() operator is applied to $\mathbf{H}$, i.e. the columns of $\mathbf{H}$ are stacked. This yields the n_Rn_T×n_Rn_T covariance matrix:

$$\mathbf{R}_{HH}^{tot} = \mathrm{E}\left\{\mathrm{vec}\left(\mathbf{H}\right) \cdot \mathrm{vec}\left(\mathbf{H}\right)^H\right\} \tag{4}$$

If we apply the vec() operator to equation 2, we get by help of Kronecker Product law (27) (conf. appendix A):

$$\begin{aligned} \mathrm{vec}\left(\mathbf{H}\right) &= \mathrm{vec}\left(\mathbf{A}_R \cdot \mathrm{diag}\left(\vec{\xi}\right) \cdot \mathbf{A}_T^T\right) \\ &= \left(\mathbf{A}_T \otimes \mathbf{A}_R\right) \cdot \mathrm{vec}\left(\mathrm{diag}\left(\vec{\xi}\right)\right) \end{aligned} \tag{5}$$

This simplifies the covariance matrix to

$$\mathbf{R}_{HH}^{tot} = \left(\mathbf{A}_T \otimes \mathbf{A}_R\right) \cdot \mathbf{J} \cdot \left(\mathbf{A}_T^H \otimes \mathbf{A}_R^H\right) \tag{6}$$

where

$$\mathbf{J} = \mathrm{E}\left\{\mathrm{vec}\left(\mathrm{diag}\left(\vec{\xi}\right)\right) \cdot \mathrm{vec}\left(\mathrm{diag}\left(\vec{\xi}\right)\right)^H\right\} \tag{7}$$

With the subpath phases being uncorrelated and the subpath powers $(\sigma_1^2, \sigma_2^2, ..., \sigma_{n_P}^2) = \vec{p}_\sigma$, we can write

$$\mathrm{E}\left\{\vec{\xi} \cdot \vec{\xi}^H\right\} = \mathrm{diag}\left(\vec{p}_\sigma\right) \tag{8}$$

and therefore

$$\mathbf{J} = \mathrm{diag}\left(\mathrm{vec}\left(\mathrm{diag}\left(\vec{p}_\sigma\right)\right)\right) \tag{9}$$

It is obvious, that $\mathbf{J}$ is a $n_P^2 \times n_P^2$ sparse diagonal matrix with the main diagonal consisting of zeros and the elements of $\vec{p}_\sigma$.
It sets some columns in $(\mathbf{A}_T \otimes \mathbf{A}_R)$ to zero and weights the other ones with σ_p^2 (or rows in $(\mathbf{A}_T^H \otimes \mathbf{A}_R^H)$, respectively). Hence, with the abbreviations $\vec{a}_{T,p} = \vec{a}_T(\Theta_{T,p})$ and $\vec{a}_{R,p} = \vec{a}_R(\Theta_{R,p})$ and Kronecker Product law (25), (6) reduces to

$$\begin{aligned}\mathbf{R}_{HH}^{tot} &= \sum_{p=1}^{n_P} \sigma_p^2 \cdot (\vec{a}_{T,p} \otimes \vec{a}_{R,p}) \cdot \left(\vec{a}_{T,p}^H \otimes \vec{a}_{R,p}^H\right) \\ &= \sum_{p=1}^{n_P} \sigma_p^2 \cdot \left(\vec{a}_{T,p} \cdot \vec{a}_{T,p}^H\right) \otimes \left(\vec{a}_{R,p} \cdot \vec{a}_{R,p}^H\right) \quad (10)\end{aligned}$$

For a very large number of n_P ($n_P \to \infty$), the sum (10) becomes a 2-dimensional integral with the angular power density $P(\vartheta_T, \vartheta_R)$:

$$\begin{aligned}\mathbf{R}_{HH}^{tot} &= \int_0^{2\pi}\int_0^{2\pi} P(\vartheta_T, \vartheta_R) \cdot \left(\vec{a}_T(\vartheta_T) \cdot \vec{a}_T(\vartheta_T)^H\right) \\ &\quad \otimes \left(\vec{a}_R(\vartheta_R) \cdot \vec{a}_R(\vartheta_R)^H\right) d\vartheta_T d\vartheta_R \quad (11)\end{aligned}$$

This can be interpreted as an expectation operation with joint density $f_{\vartheta_T,\vartheta_R}(\vartheta_T, \vartheta_R) = \frac{1}{P_0} \cdot P(\vartheta_T, \vartheta_R)$, where

$$P_0 = \int_0^{2\pi}\int_0^{2\pi} P(\vartheta_T, \vartheta_R) d\vartheta_T d\vartheta_R \quad (12)$$

is the total subpath power.
We now use the assumption of the beginning of this section, that the DOAs are independent from the DODs. Hence, the joint density can be split:

$$f_{\vartheta_T\vartheta_R}(\vartheta_T, \vartheta_R) = f_{\vartheta_T}(\vartheta_T) \cdot f_{\vartheta_R}(\vartheta_R) \quad (13)$$

Generalizing the Kronecker Product law (26) to the continuous case, we separate the double integral in (11) into two integrals and finally get:

$$\begin{aligned}\mathbf{R}_{HH}^{tot} &= P_0 \cdot \int_0^{2\pi} \left(\vec{a}_T(\vartheta_T) \cdot \vec{a}_T(\vartheta_T)^H\right) \cdot f_{\vartheta_T}(\vartheta_T)\, d\vartheta_T \\ &\quad \otimes \int_0^{2\pi} \left(\vec{a}_R(\vartheta_R) \cdot \vec{a}_R(\vartheta_R)^H\right) \cdot f_{\vartheta_R}(\vartheta_R)\, d\vartheta_R \\ &= P_0 \cdot \mathbf{R}_{HH}^T \otimes \mathbf{R}_{HH}^R \quad (14)\end{aligned}$$

where $\mathbf{R}_{HH}^T$ and $\mathbf{R}_{HH}^R$ are the normalized covariance matrices in the case of only one array ([8]). These matrices have already been investigated for many algorithms.
Without loss of generality, we assume $P_0 = 1$ in the sequel.

Hence, we described the covariance matrix of a general MIMO system as a Kronecker Product of two known covariance matrices. In the following, we will show, how this result can be used in Monte-Carlo simulations.

4. SIMULATION OF CHANNEL CORRELATIONS

Filtering an uncorrelated random process z_{unc} usually leads to a correlated random process z_{cor} [9]. In matrix notation this can be expressed as a multiplication with a matrix $\mathbf{L}^H$:

$$\vec{z}_{cor} = \mathbf{L}^H \cdot \vec{z}_{unc} \quad (15)$$

The resulting random process has the covariance matrix:

$$\begin{aligned}\mathbf{R}_{zz} &= \mathrm{E}\left\{\vec{z}_{cor} \cdot \vec{z}_{cor}^H\right\} \\ &= \mathbf{L}^H \cdot \mathrm{E}\left\{\vec{z}_{unc} \cdot \vec{z}_{unc}^H\right\} \cdot \mathbf{L} \\ &= \mathbf{L}^H \cdot \mathbf{L} \quad (16)\end{aligned}$$

That is, a random variable with a desired covariance matrix $\mathbf{R}_{zz}$ can be generated by multiplying an uncorrelated process with any square root of this matrix. In the sequel we will use the Cholesky square roots [10], where $\mathbf{L}^H$ and $\mathbf{L}$ are the lower and upper triangular, respectively. However, the results are also valid for any square root.
Hence, we filter a vectorized uncorrelated MIMO channel matrix $\mathbf{H}_{unc}$ with the lower triangular $\mathbf{L}_{tot}^H$ according to (15). This yields the desired correlated channel matrix $\mathbf{H}_{unc}$ in vector form:

$$\mathrm{vec}(\mathbf{H}_{cor}) = L_{tot}^H \cdot \mathrm{vec}(\mathbf{H}_{unc}) \quad (17)$$

using the Cholesky decomposition[1] of the computed covariance matrix

$$\mathbf{R}_{HH}^{tot} = \mathbf{L}_{tot}^H \cdot \mathbf{L}_{tot} \quad (18)$$

With the Cholesky decompositions $\mathbf{R}_{HH}^R = \mathbf{L}_R^H \cdot \mathbf{L}_R$ and $\mathbf{R}_{HH}^T = \mathbf{L}_T^H \cdot \mathbf{L}_T$, and with Kronecker Product law (25), equation (14) can be written as:

$$\begin{aligned}\mathbf{R}_{HH}^{tot} &= \mathbf{R}_{HH}^T \otimes \mathbf{R}_{HH}^R \\ &= \left(\mathbf{L}_T^H \cdot \mathbf{L}_T\right) \otimes \left(\mathbf{L}_R^H \cdot \mathbf{L}_R\right) \\ &= \left(\mathbf{L}_T^H \otimes \mathbf{L}_R^H\right) \cdot \left(\mathbf{L}_T \otimes \mathbf{L}_R\right) \quad (19)\end{aligned}$$

Since the Cholesky decomposition is unique [10] and since the Kronecker Product of two lower triangular matrices is again lower triangular, we compare equations (18) and (19):

$$\mathbf{L}_{tot}^H = \mathbf{L}_T^H \otimes \mathbf{L}_R^H \quad (20)$$

[1] Spatial covariance matrices are usually positive definite. However, when the angular spreads tend to zero, the matrix gets singular (fully correlated), i.e. the Cholesky decomposition fails. Nevertheless, equation (10) shows, that we can simply state an upper triangular matrix containing only one non-zero column for this case. Therefore, we assume in the following, that the Cholesky decomposition succeeds.

This result is inserted in (17):

$$\begin{aligned} \mathrm{vec}\left(\mathbf{H}_{cor}\right) &= \left(\mathbf{L}_T^H \otimes \mathbf{L}_R^H\right) \cdot \mathrm{vec}\left(\mathbf{H}_{unc}\right) \\ &= \mathrm{vec}\left(\mathbf{L}_R^H \cdot \mathbf{H}_{unc} \cdot \mathbf{L}_T\right) \end{aligned} \quad (21)$$

As the dimensions of the matrices enclosed by the vec() operator are both $n_R \times n_T$, we omit the vec() operator :

$$\mathbf{H}_{cor} = \mathbf{L}_R^H \cdot \mathbf{H}_{unc} \cdot \mathbf{L}_T \quad (22)$$

This result is much more convenient than equation (17). It is applicable to all multiple antenna algorithms such as beamforming, diversity or MIMO techniques. If only a single antenna is used on one side, (22) simplifies to a form similar to equation (15), since one of the triangular matrices becomes a scalar.
In the following, this result for a flat environment is extended to the frequency selective case.

5. FREQUENCY SELECTIVE FADING

In the frequency selective case, the complex channel coefficients h_{ij} in figure 1 are replaced by complex channel impulse responses $h_{ij}(\tau)$. In mobile channels we usually assume that due to a limited receiver bandwidth, most of the energy in the impulse responses is concentrated in a small number n_τ of τ-values τ_l. Equivalent to section 2 we further assume, that all impulse responses $h_{ij}(\tau)$ experience the same delays τ_l. Together with the uncorrelated scattering assumption [5], we can express the correlated channel matrix impulse response $\mathbf{H}_{cor}$ containing the impulse responses $h_{ij}(\tau)$ as:

$$\mathbf{H}_{cor}\left(\tau\right) = \sum_{l=1}^{n_\tau} \mathbf{L}_{R,l}^H \cdot \mathbf{H}_{unc,l} \cdot \mathbf{L}_{T,l} \cdot \delta(\tau - \tau_l) \quad (23)$$

Each tap l is determined by a certain delay τ_l, by a spatially uncorrelated channel matrix $\mathbf{H}_{unc,l}$, by a covariance matrix at the transmitter side $\mathbf{R}_{xx,l}^T = \mathbf{L}_{T,l}^H \cdot \mathbf{L}_{T,l}$ and by a covariance matrix at the receiver side $\mathbf{R}_{xx,l}^R = \mathbf{L}_{R,l}^H \cdot \mathbf{L}_{R,l}$.

6. APPLICATION: MIMO CAPACITIES

With this model, MIMO capacities can very easily be computed in Monte-Carlo simulations. In [3], it was shown that the Shannon capacity of a MIMO system according to figure 1 is:

$$C_{MIMO} = \log_2 \left(\prod_{i=1}^{n_S} \left(1 + \frac{s_i^2 / n_T}{\sigma_N^2} \right) \right) \quad (24)$$

	Tx side	Rx side
array type	4-ULA	4-ULA
element spacing	$10\lambda/5\lambda/0.5\lambda$	0.3λ
size at 2GHz in[m]	4.5/2.25/0.225	0.135
mean direction	$0°$	$0°$
angular spread	$5°$	$360°$
angular distribution	uniform	uniform

Table 1: Simulation Parameters

where s_i are the n_S singular values of the channel matrix.
In figure 3 the channel matrix was computed according to (22), where the entries in $\mathbf{H}_{unc}$ are independent and complex gaussian distributed with variance 1. We assumed 4-element uniform linear arrays (4-ULA) on both sides with the parameters shown in table 1.
This reflects an urban downlink scenario with a low angular spread, but a large array aperture at the base station and vice versa at the mobile station.
The plots in figure 3 show cumulative distribution functions for the capacity C_{MIMO}, i.e. the percentage that the capacity is less than a certain value. For each curve 30000 snapshots have been evaluated. The signal-to-noise-ratio was set to 20dB, i.e. $\sigma_N^2 = 0.01$ in equation (24).

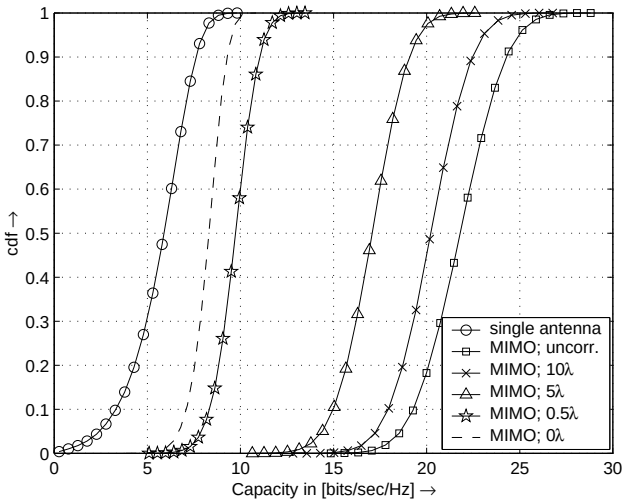

Figure 3: Shannon Capacity in uncorrelated and correlated MIMO systems for different element spacings at the transmitter

In the uncorrelated case, the capacity increases approximately by factor 4 compared with a single antenna. Uncorrelated corresponds to infinite array apertures on both sides. The covariance matrices as well as their cholesky decompositions are then identity matrices.

If we assume finite apertures, the curves shift to the left, as the channel matrix becomes correlated. This degradation is moderate for a 10λ-element spacing. It is particularly due to the correlations at the mobile station. However, with a smaller base station array, the degradation becomes critical. If we further reduce the array size, the dotted curve is approached, which is the Rx diversity bound, i.e. the case of one Tx-element or 0λ-spacing, respectively. Note that we do not see a downlink beamforming gain in these curve, since Shannon's formula (24) does not assume channel knowledge at the transmitter. With a $\lambda/2$-array the capacity gain over the Rx diversity scheme is rather low.

7. CONCLUSIONS

A statistical model for correlated MIMO channels has been proposed. For a flat fading environment, the spatial channel correlations have been described as one large covariance matrix. It was shown, that this covariance matrix can also be expressed as the Kronecker Product of the covariance matrices of the transmit and receive signals.
With this result a very simple way was found to filter an uncorrelated channel matrix.
Under the uncorrelated scattering assumption, the extension to the frequency selective case was done by superimposing a number of time-shifted flat fading taps.
A possible application of this model is the investigation of MIMO capacities. Results were shown for an urban downlink scenario with a large angular spread and small aperture at the mobile station and a large aperture and small angular spread at the base station. As expected, the capacity decreases with limited aperture and limited angular spread.

A. KRONECKER PRODUCT LAWS

The following Kronecker Product laws are used throughout this paper [11]:

$\mathbf{A} \in \mathcal{C}^{n_1 \times n_2}$,$\mathbf{B} \in \mathcal{C}^{n_2 \times n_3}$, $\mathbf{C} \in \mathcal{C}^{n_4 \times n_5}$ and $\mathbf{D} \in \mathcal{C}^{n_5 \times n_6}$:

$$(\mathbf{A} \otimes \mathbf{B}) \cdot (\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A} \cdot \mathbf{C}) \otimes (\mathbf{B} \cdot \mathbf{D}) \qquad (25)$$

$\mathbf{A} \in \mathcal{C}^{n_1 \times n_2}$ and $\mathbf{B}, \mathbf{C} \in \mathcal{C}^{n_3 \times n_4}$:

$$(\mathbf{A} \otimes \mathbf{B}) + (\mathbf{A} \otimes \mathbf{C}) = \mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) \qquad (26)$$

$\mathbf{A} \in \mathcal{C}^{n_1 \times n_2}$,$\mathbf{B} \in \mathcal{C}^{n_2 \times n_3}$ and $\mathbf{C} \in \mathcal{C}^{n_3 \times n_4}$:

$$\mathrm{vec}\,(\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{C}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) \cdot \mathrm{vec}\,(\mathbf{B}) \qquad (27)$$

REFERENCES

[1] V. Tarokh, N. Seshadri and A.R. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744-765, 1998

[2] J. C. Liberti jr. and T.S. Rappaport, *Smart Antennas for Wireless Communications*, Prentice-Hall, 1999

[3] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, 1999

[4] G. J. Foschini and M.J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communication*, vol. 6, pp. 311-335, 1998

[5] P. A. Bello, "Characterization of Randomly Time-Variant Linear Channels," *IEEE Transactions on Communications Systems*, vol. 11, pp. 360-393, 1963

[6] W. C. Jakes, *Microwave Mobile Communications*, IEEE Press, 1974

[7] R. R. Müller, "A random matrix model of communication via antenna arrays," *submitted to IEEE Transactions on Information Theory*, August 2000

[8] J. Salz, J. H. Winters, "Effect of Fading Correlation on Adaptive Arrays in Digital Wireless Communications," *IEEE International Conference on Communications ICC'93*, Geneva, Switzerland, pp. 1768-1774, 1993

[9] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York, 1984

[10] G. H. Golub and C.F. van Loan, *Matrix Computations*, John Hopkins University Press, 1996

[11] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, John Wiley & Sons, 1981