STRUCTURED SPATIAL INTERFERENCE REJECTION COMBINING

George Jöngren

Signal Processing
Royal Institute of Technology
SE-100 44 Stockholm, Sweden
George.Jongren@s3.kth.se

David Astély

Nokia Networks
Radio Access Systems
P.O. Box 1070
SE-164 25 Kista, Sweden
David.Astely@nokia.com

Björn Ottersten

Signal Processing
Royal Institute of Technology
SE-100 44 Stockholm, Sweden
Bjorn.Ottersten@s3.kth.se

ABSTRACT

Modern communication systems are often interference limited. By modeling the co-channel interference as spatially colored, temporally white Gaussian noise, it is straightforward to incorporate interference rejection in the metric of a sequence estimator. In general, estimates of both the channels and the spatial color of the co-channel interference and the noise are needed. In this work, a structured model for the spatial noise covariance matrix is proposed and maximum likelihood estimates of the parameters are derived. The choice of model order is also addressed. Simulation results show large gains due to the use of these structured estimates compared with the conventional, unstructured, approach.

1 INTRODUCTION

Mobile communication systems have found widespread use during the last decade. With the rapid growth expected in the future, demands for high capacity and reliable services continue to increase. Among the challenges in designing a system meeting these demands, particularly when the number of users is large, is interference from other users. One way of reducing this problem is to utilize the spatial dimension by means of an antenna array at the receiver.

This work considers detecting the transmitted data in the presence of inter symbol interference (ISI) and co-channel interference (CCI) using a maximum likelihood sequence estimator (MLSE). Similar to, e.g. [1–3], the MLSE is modified by modeling the interference together with the noise as a spatially colored, temporally white complex Gaussian process. In this way, the ISI is handled while also rejecting the CCI.

Recently, there has been an interest in using a low-rank model for the single-input multi-output (SIMO) channel between a transmitter and a receiving antenna array [4]. However, in the present work, a low-rank model is instead used for the CCI contribution to the spatial covariance matrix. This imposes a structure on the covariance matrix that reduces the number of parameters to be estimated, which leads to reduced estimation errors and hopefully improved sequence estimation. The problem with an unstructured covariance matrix becomes clear if we count the number of parameters and equations. Assuming an unstructured covariance matrix and m antennas, the number of parameters is proportional to m^2. For a given training sequence length, the number of equations that determine the estimates grows linearly with m. Therefore, the quality of an estimate of the unstructured covariance matrix is in this case expected to be low if the number of antennas is large.

In this work, we derive the joint maximum likelihood (ML) estimates of the SIMO channel and the structured noise covariance matrix. The estimate of the channel turns out to be given by the solution of a least squares problem whereas an estimate of the low-rank CCI contribution and noise variance is obtained by an eigenvalue decomposition (EVD) of the residual sample covariance matrix. By utilizing the minimum description length (MDL) criteria [5], the model order is determined as well. The performance is investigated when the estimated parameters are used for MLSE in the presence of CCI. Simulation results illustrate significant gains for an EDGE scenario with eight receiving antenna elements.

2 DATA MODEL

The model used to describe the symbol sampled baseband equivalent signal received by an array with m elements is

$$\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n) + \mathbf{e}(n)\,,$$

where $\mathbf{x}(n)$ is an $m \times 1$ vector representing the array output and

$$\mathbf{H} = [\mathbf{h}_0\ \mathbf{h}_1\ \ldots \mathbf{h}_L]$$

is an $m \times (L+1)$ matrix modeling the channel between the transmitter and the receiving array. Oversampling with respect to the symbol period may be modeled by simply increasing the number of channels to mq, where q is the oversampling factor. The symbol sequence transmitted from the user of interest, $s(n)$, is used to construct $\mathbf{s}(n)$ as

$$\mathbf{s}(n) = [s(n)\ s(n-1)\ \ldots\ s(n-L)]^T\,.$$

The term $\mathbf{e}(n)$ represents noise and interference. As in [1–3], this term is modeled as a zero-mean temporally white complex Gaussian process with second order statistics given by

$$\mathrm{E}\left[\mathbf{e}(n)\mathbf{e}^*(k)\right] = \mathbf{Q}\delta_{n-k}\,,$$

where $\mathbf{Q}$ is the spatial covariance matrix and δ_n denotes the Kronecker delta. This is clearly suboptimal as the CCI has the same properties as the signal of interest, finite alphabet and, in the time-dispersive case, some temporal correlation. The Gaussian modeling assumption leads to a structure that takes only the second order moments of the CCI into account. As compared to a joint detection approach, this applies to a larger class of interfering signals.

2.1 A Structured Covariance Model

In this work, the spatial covariance matrix $\mathbf{Q}$ is assumed to consist of a low-rank signal part in addition to spatially white noise. Such a model is commonly assumed in the field of subspace based direction of arrival estimation.

Suppose that d co-channel interferers are present and that the time-dispersion introduced by the multi-path propagation is small in comparison with the inverse bandwidth of the signals, i.e. the channels are frequency non-selective. A reasonable model of the covariance matrix is then

$$\mathbf{Q} = \mathbf{\Psi} + \sigma^2\mathbf{I}_m\,, \tag{1}$$

where $\mathbf{\Psi}$ is a Hermitian positive semidefinite matrix with rank $d \in \{0, \cdots, m-1\}$. Note that this model is appropriate even in the case of frequency selective channels. However, the rank d is then typically larger than the number of interfering users.

3 ESTIMATORS

This section describes how the transmitted signal is detected and how to estimate the necessary parameters in our structured model. Detection of the rank of $\mathbf{\Psi}$ is also considered.

3.1 Sequence Detection

Let us for a moment assume that all parameters are known. Sequence estimation is considered, although other detection schemes are also applicable. With the Gaussian assumption for the noise and CCI, the maximum likelihood sequence estimate is given by

$$\{\hat{s}(n)\} = \arg\min_{\{s(n)\}} \sum_n \|\mathbf{x}(n) - \mathbf{H}\mathbf{s}(n)\|^2_{\mathbf{Q}^{-1}}\,,$$

where $\|\mathbf{z}\|^2_{\mathbf{W}} = \mathbf{z}^*\mathbf{W}\mathbf{z}$. Note that the metric increment, i.e. each term in the sum, is a function of $\mathbf{s}(n)$, and that the search over allowed sequences may be implemented with the Viterbi Algorithm with a memory of L symbols.

3.2 A Structured Estimator

It is obvious from the previous section that the sequence detection algorithm needs to know both the channel and the covariance matrix. For optimal detection, these parameters and the transmitted data sequence should be estimated jointly. However, this is often considered too complex in practice. The transmitted signal is assumed to be divided into bursts, where each burst consist of a training sequence and a data sequence. The approach taken in this work is then to first estimate the unknown parameters during the training period of the signal and then use these estimates in the sequence detector. A quasi-stationary scenario is assumed where the signal is stationary during a burst but may alter its characteristics from one burst to another. This means that the signal of interest and the interference must be roughly burst synchronized. An example of such a system might be a GSM system with synchronized base stations and not too large cells so that the synchronism is reasonably accurate.

The maximum likelihood estimates of the parameters $\mathbf{H}$ and $\mathbf{Q}$ that are needed in the sequence detector are now derived, assuming a known training sequence. The transmitted signal $s(n)$ is known during the training period $n = n_0, \cdots, n_0 + N$ which means $\mathbf{s}(n)$ may be formed for $n = n_0 + L, \cdots, n_0 + N$. Based on the complex Gaussian assumption, the scaled negative log-likelihood function is then given by [6]

$$l\left(\mathbf{H}, \mathbf{Q}\right) = \log\det(\mathbf{Q}) + \mathrm{tr}\left(\mathbf{Q}^{-1}\mathbf{C}\left(\mathbf{H}\right)\right)\,,$$

where parameter independent terms have been neglected, $\mathrm{tr}(\cdot)$ denotes the trace operator and where

$$\mathbf{C}\left(\mathbf{H}\right) = \hat{\mathbf{R}}_{\mathrm{xx}} - \hat{\mathbf{R}}_{\mathrm{xs}}\mathbf{H}^* - \mathbf{H}\hat{\mathbf{R}}^*_{\mathrm{xs}} + \mathbf{H}\hat{\mathbf{R}}_{\mathrm{ss}}\mathbf{H}^*$$

$$\hat{\mathbf{R}}_{\mathrm{xx}} = \frac{1}{N-L+1}\sum_{k=n_0+L}^{n_0+N}\mathbf{x}(k)\mathbf{x}^*(k)\,,$$

with $\hat{\mathbf{R}}_{\mathrm{xs}}$ and $\hat{\mathbf{R}}_{\mathrm{ss}}$ formed as $\hat{\mathbf{R}}_{\mathrm{ss}}$. It is now straightforward to verify that the ML estimate of the channel is given by

$$\hat{\mathbf{H}} = \hat{\mathbf{R}}_{\mathrm{xx}}\hat{\mathbf{R}}^{-1}_{\mathrm{ss}}\,,$$

which is also the least-squares estimate. The concentrated likelihood function then becomes

$$l\left(\mathbf{Q}\right) = \log\det(\mathbf{Q}) + \mathrm{tr}(\mathbf{Q}^{-1}\hat{\mathbf{R}}_{\mathrm{ee}})\,, \tag{2}$$

where $\hat{\mathbf{R}}_{\mathrm{ee}} = \hat{\mathbf{R}}_{\mathrm{xx}} - \hat{\mathbf{R}}_{\mathrm{xs}}\hat{\mathbf{R}}^{-1}_{\mathrm{ss}}\hat{\mathbf{R}}^*_{\mathrm{xs}}$ is seen to correspond to the residual covariance. We now wish to minimize $l\left(\mathbf{\Psi} + \sigma^2\mathbf{I}_m\right)$ with respect to $\mathbf{\Psi}$ and σ^2, while taking the constraints described in Section 2.1 into account. Since $\mathbf{\Psi}$ is low-rank, the approach taken in this work is to rewrite the covariance matrix in terms of its EVD,

$$\mathbf{Q} = \mathbf{\Psi} + \sigma^2\mathbf{I}_m = \mathbf{E}_{\mathrm{s}}\mathbf{\Lambda}_{\mathrm{s}}\mathbf{E}^*_{\mathrm{s}} + \sigma^2\mathbf{E}_{\mathrm{n}}\mathbf{E}^*_{\mathrm{n}}\,, \tag{3}$$

and then formulate an equivalent optimization problem with respect to the new parameters $\mathbf{E}_\mathrm{s}$, $\mathbf{\Lambda}_\mathrm{s}$, $\mathbf{E}_\mathrm{n}$ and σ^2. In the expression above, $\mathbf{\Lambda}_\mathrm{s}$ is a diagonal matrix containing, in descending order, the d largest eigenvalues of $\mathbf{Q}$ and $\mathbf{E}_\mathrm{s}$ is composed of the corresponding eigenvectors. The remaining $m-d$ eigenvalues are all equal to σ^2 with the columns of $\mathbf{E}_\mathrm{n}$ representing the associated eigenvectors. The eigenvectors are here normalized to unit norm. In order to simplify the derivation, the eigenvalues described by $\mathbf{\Lambda}_\mathrm{s}$ are assumed to be distinct. This is not a restrictive assumption in practice.

Clearly, the parameterization given by the EVD is not unique. For example, the eigenvectors belonging to the signal part, described by the columns of $\mathbf{E}_\mathrm{s}$, are only unique up to an arbitrary phase factor. It is also obvious that $\mathbf{Q}$ is unaffected if $\mathbf{E}_\mathrm{n}$ is multiplied from the right by a unitary matrix. However, as is shown below, the use of a non-unique parameterization does not introduce any serious difficulties.

In our case, an important property of the parameterization is that the set of possible covariance matrices $\mathbf{Q}$ is the same under both parameterizations. This fact may be expressed as

$$\{\mathbf{\Psi}+\sigma^2\mathbf{I}_m\} = \{\mathbf{E}_\mathrm{s}\mathbf{\Lambda}_\mathrm{s}\mathbf{E}_\mathrm{s}^* + \sigma^2\mathbf{E}_\mathrm{n}\mathbf{E}_\mathrm{n}^*\}\,,$$

where the constraints on the parameters have been omitted for notational convenience. This implies that it is possible to minimize $l\,(\mathbf{Q})$ by first optimizing with respect to the parameters of the EVD and then mapping the result back to the original parameters. The resulting optimization problem can therefore be formulated as

$$\{\hat{\mathbf{E}}_\mathrm{s}, \hat{\mathbf{E}}_\mathrm{n}, \hat{\mathbf{\Lambda}}_\mathrm{s}, \hat{\sigma}^2\} = \arg \min_{\substack{\mathbf{E}_\mathrm{s},\mathbf{E}_\mathrm{n},\mathbf{\Lambda}_\mathrm{s},\sigma^2:\\ \mathbf{E}_\mathrm{s}\mathbf{E}_\mathrm{s}^*+\mathbf{E}_\mathrm{n}\mathbf{E}_\mathrm{n}^*=\mathbf{I}_m\\ \mathbf{\Lambda}_\mathrm{s}\succ\sigma^2,\,\sigma^2\geq 0}} l\left(\mathbf{E}_\mathrm{s}\mathbf{\Lambda}_\mathrm{s}\mathbf{E}_\mathrm{s}^* + \sigma^2\mathbf{E}_\mathrm{n}\mathbf{E}_\mathrm{n}^*\right)$$

where $\mathbf{\Lambda}_\mathrm{s}\succ\sigma^2$ means that the diagonal elements of $\mathbf{\Lambda}_\mathrm{s}$ are larger than σ^2. It is easily verified that the above optimization problem is the same as when computing the ML estimate of the EVD of $\mathbf{Q}$, using a sample covariance matrix equal to $\hat{\mathbf{R}}_\mathrm{ee}$. As a consequence of this, a derivation similar to as in [7, p. 131], but generalized to complex valued matrices, provides the solution to our problem. It turns out that $l\left(\mathbf{E}_\mathrm{s}\mathbf{\Lambda}_\mathrm{s}\mathbf{E}_\mathrm{s}^* + \sigma^2\mathbf{E}_\mathrm{n}\mathbf{E}_\mathrm{n}^*\right)$ is minimized if

$$\begin{aligned}\hat{\mathbf{E}}_\mathrm{s} &= \tilde{\mathbf{E}}_\mathrm{s}, \quad \hat{\mathbf{E}}_\mathrm{n} = \tilde{\mathbf{E}}_\mathrm{n}\\ \hat{\mathbf{\Lambda}}_\mathrm{s} &= \tilde{\mathbf{\Lambda}}_\mathrm{s}, \quad \hat{\sigma}^2 = \frac{1}{m-d}\operatorname{tr}\tilde{\mathbf{\Lambda}}_\mathrm{n}\,,\end{aligned}$$

where $\tilde{\mathbf{E}}_\mathrm{s}$, $\tilde{\mathbf{E}}_\mathrm{n}$, $\tilde{\mathbf{\Lambda}}_\mathrm{s}$ and $\tilde{\mathbf{\Lambda}}_\mathrm{n}$ represent the EVD of the residual covariance matrix. Formulating this in mathematical terms, $\hat{\mathbf{R}}_\mathrm{ee} = \tilde{\mathbf{E}}_\mathrm{s}\tilde{\mathbf{\Lambda}}_\mathrm{s}\tilde{\mathbf{E}}_\mathrm{s}^* + \tilde{\mathbf{E}}_\mathrm{n}\tilde{\mathbf{\Lambda}}_\mathrm{n}\tilde{\mathbf{E}}_\mathrm{n}^*$, where the individual matrices are defined similarly to as in (3). An explicit expression for the ML estimate of the low-rank signal part of $\mathbf{Q}$ is finally obtained as $\hat{\mathbf{\Psi}} = \hat{\mathbf{E}}_\mathrm{s}(\hat{\mathbf{\Lambda}}_\mathrm{s} - \hat{\sigma}^2\mathbf{I}_d)\hat{\mathbf{E}}_\mathrm{s}^*$. The derivation in this section thus shows that the joint ML estimates of the channel and the structured covariance matrix are given by

$$\begin{aligned}\hat{\mathbf{Q}} &= \hat{\mathbf{E}}_\mathrm{s}(\hat{\mathbf{\Lambda}}_\mathrm{s} - \hat{\sigma}^2\mathbf{I}_d)\hat{\mathbf{E}}_\mathrm{s}^* + \hat{\sigma}^2\mathbf{I}_m = \hat{\mathbf{E}}_\mathrm{s}\hat{\mathbf{\Lambda}}_\mathrm{s}\hat{\mathbf{E}}_\mathrm{s}^* + \hat{\sigma}^2\hat{\mathbf{E}}_\mathrm{n}\hat{\mathbf{E}}_\mathrm{n}^*\\ \hat{\mathbf{H}} &= \hat{\mathbf{R}}_\mathrm{xx}\hat{\mathbf{R}}_\mathrm{ss}^{-1}\,.\end{aligned} \tag{4}$$

Note that it is well known that if an unstructured parameterization of $\mathbf{Q}$ is used, the concentrated likelihood function given in (2) is minimized with respect to $\mathbf{Q}$ when $\hat{\mathbf{Q}} = \hat{\mathbf{R}}_\mathrm{ee}$ [6]. Using this estimate in the sequence detector would then correspond to the interference rejection schemes found in e.g. [1–3].

3.3 Rank Detection

We have assumed so far that the rank d, of the matrix $\mathbf{\Psi}$, is known. In practice, this rank also needs to be estimated. The model in (1) suggests that the rank d could be determined from the multiplicity of the smallest eigenvalue of $\mathbf{Q}$. However, in the finite sample case, the eigenvalues will all be different with probability one. The rank is therefore difficult to determine using this method.

There are several other detection schemes applicable to this problem. We arbitrarily consider the MDL criterion [5] for detecting the rank. This scheme is based on information theoretic principles. It is clear from the previous section that the family of possible ranks imply a corresponding family of models for the observed data. Assuming the model is used for encoding the received signal, the MDL criterion selects the model which results in the minimum code length, in an asymptotic sense.

Let $\mathbf{X} = \{\mathbf{x}(n_0+L), \cdots, \mathbf{x}(n_0+N)\}$ represent the set of observations and let

$$\mathbf{\Theta}_k = \left\{\mathbf{H}, \mathbf{\Psi}, \sigma^2\right\}$$

represent the unknown parameters with k degrees of freedom. As an alternative, the equivalent parameterization based on the EVD may be used in place of $\mathbf{\Psi}$. The asymptotic form of the criterion function is now given by

$$\mathrm{MDL(k)} = -\log p(\mathbf{X}|\hat{\mathbf{\Theta}}_k) + \frac{1}{2}k\log(N-L+1)\,,$$

where $p(\mathbf{X}|\hat{\mathbf{\Theta}}_k)$ is the probability density function of the observed data conditioned on the ML estimate $\hat{\mathbf{\Theta}}_k$ of the parameters. The first term is thus the minimum negative log-likelihood function and the second term penalizes over modeling.

From the definition of $\mathbf{\Theta}_k$ it is clear that the channel and the noise variance are described by a total of $2m(L+1)+1$ unconstrained real valued parameters. Determining the number of degrees of freedom due to $\mathbf{\Psi}$ is complicated by the rank constraint imposed on the matrix. A more detailed analysis shows that $\mathbf{\Psi}$ contributes with $2md - d^2$ degrees of freedom[1].

[1] This may be seen by counting parameters in a factorization of the form $\mathbf{\Psi} = \mathbf{F}\mathbf{F}^*$, where $\mathbf{F}$ is an $m \times d$ lower triangular matrix with non-negative real valued diagonal elements.

The total number of degrees of freedom is therefore $k = 2m(L+1) + 2md - d^2 + 1$ and the rank d is thus determined as

$$\hat{d} = \arg \min_{d \in \{0,\cdots,m-1\}} \text{MDL}\left(2m(L+1) + 2md - d^2 + 1\right) .$$

4 NUMERICAL SIMULATIONS

In order to illustrate the gains of using a structured covariance model, this section presents the results of simulations assuming a system similar to EDGE, which is one of the standardized modes in the evolution of the GSM system. The simulations have been conducted for an uplink scenario with synchronized base stations having eight antennas. The raw symbol rate is approximately 270 ksymb/s and 8-PSK modulation with a linearized GMSK pulse is used. Each burst consists of a training sequence containing 26 symbols and a total of 116 data symbols. The channel is assumed constant during a burst and independently fading from one burst to another, corresponding to ideal frequency hopping. The fading is modeled as independent from antenna to antenna and each impulse response is generated according to the 12-tap power delay profile of the GSM typical urban (TU) channel [8].

At the receiver, the baseband equivalent output from each antenna is filtered using a fourth order Butterworth filter with a 3 dB bandwidth of 200 kHz and then sampled at the symbol rate prior to the sequence estimation. Simulations are performed for two cases, corresponding to an MLSE based on the structured and unstructured covariance matrix, respectively. As previously mentioned, the structured approach uses the MDL criterion to determine the appropriate rank for each individual burst and then forms the parameter estimates as in (4). The unstructured receiver uses $\hat{\mathbf{R}}_{\text{ee}}$ as the estimate of the covariance matrix. The length of the filtered channel was assumed to correspond to 5 symbols, i.e. $L = 4$. This includes the effects of transmitter and receiver filters. A scenario with one interferer was studied and the signal-to-noise ratio (SNR), measured at the input of the receiver filters, was set to 15 dB while the carrier-to-interference ratio (C/I) was varied. Channel realizations for the interferer were independently drawn from the same fading model as the user of interest. The resulting uncoded bit error rate (BER) is shown in Figure 1. At a BER of 1% the gain over the unstructured receiver is seen to be about 10 dB. This example thus demonstrates that there are significant gains to be made by taking the structure of the covariance matrix into account.

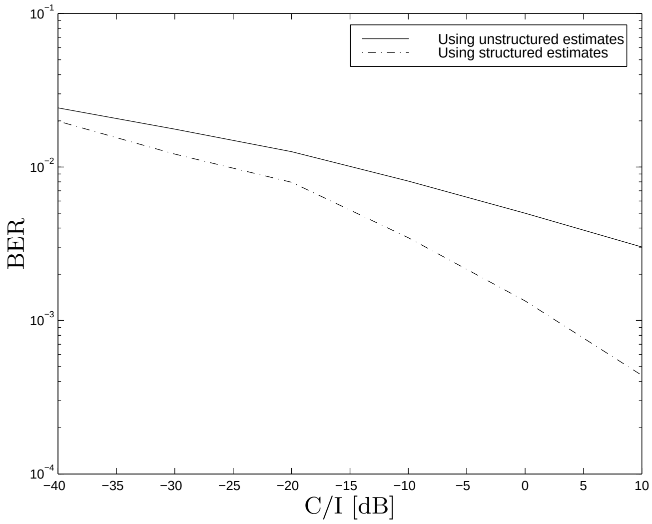

Figure 1: Typical urban scenario, one interferer, eight antennas, SNR=15 dB.

5 SUMMARY

Structured estimation of parameters used in an MLSE was considered. Interference was modeled as a low-rank contribution to the spatial covariance matrix of the observed signal. Maximum likelihood estimates of both the channel and the structured covariance matrix were derived. An MDL criterion for estimating the rank of the low-rank contribution was formulated. A receiver using the structured estimate was then simulated and compared with a receiver utilizing an unstructured covariance matrix. Significant gains were shown for an EDGE scenario using eight receiver antennas.

6 ACKNOWLEDGEMENT

The authors wish to thank Dr. Peter Muszynski at Nokia for inspiring them to study the problem considered in this work.

REFERENCES

[1] D. Asztély, "On Antenna Arrays in Mobile Communication Systems, Fast Fading and GSM Base Station Receiver Algorithms" Master's thesis, Royal Institute of Technology, Feb. 1995.

[2] G. E. Bottomley and K. Jamal, "Adaptive Arrays and MLSE Equalization" in *Proc. IEEE Vehicular Technology Conference*, pp. 50–54, July 1995.

[3] S. Mayrargue, "Practical Implementation of a Multisensor Array Receiver Structure for Wireless Communications" in *Proc. SPAWC*, pp. 201–204, Apr. 1997.

[4] E. Lindskog and C. Tidestav, "Reduced Rank Channel Estimation" in *Proc. IEEE Vehicular Technology Conference*, pp. 1126–1130, May. 1999.

[5] J. Rissanen, "Modeling by Shortest Data Description" *Automatica*, vol. 14, pp. 465–471, 1978.

[6] N. R. Goodman, "Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution" *Annals of Mathematical Statistics*, vol. 34, pp. 152–177, Mar. 1963.

[7] T. W. Anderson, "Asymptotic theory for principal component analysis" *Annals of Mathematical Statistics*, vol. 34, pp. 122–148, Feb. 1963.

[8] European Telecommunication Standards Institute (ETSI), "Digital Cellular Telecommunications System (Phase 2+); Radio Transmission and Reception (GSM 05.05 version 5.2.0)" Jul. 1996.