



Andrieu, C., Piechocki, R. J., McGeehan, J. P., & Armour, S. M. (2003). Semi-blind identification of wideband MIMO channels via stochastic sampling. In *IEEE Vehicular Technology Conference* (2 ed., Vol. 57, pp. 994-997) <https://doi.org/10.1109/VETECS.2003.1207775>

Peer reviewed version

Link to published version (if available):  
[10.1109/VETECS.2003.1207775](https://doi.org/10.1109/VETECS.2003.1207775)

[Link to publication record on the Bristol Research Portal](#)  
PDF-document

## University of Bristol – Bristol Research Portal

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/brp-terms/>

# Semi-blind Identification of Wideband MIMO Channels via Stochastic Sampling

Christophe Andrieu<sup>1</sup>, Robert J. Piechocki<sup>2</sup>, Joe P. McGeehan<sup>2</sup> and Simon M. Armour<sup>2</sup>

<sup>1</sup>University of Bristol, Department of Mathematics  
University Walk, Bristol BS8 1TW, UK

<sup>2</sup>University of Bristol, Centre For Communications Research  
Woodland Road, MVB, BS8 1UB, Bristol, UK.

E-mail: {c.andrieu, r.j.piechocki, j.p.mcgeehan, simon.armour}@bristol.ac.uk

**Abstract**— In this paper we address the problem of wideband Multiple-Input Multiple-Output (MIMO) channel (multidimensional time invariant FIR filter) identification using Markov Chains Monte Carlo Methods. Towards this end we develop a novel stochastic sampling technique that produces a sequence of multidimensional channel samples. The method is semi-blind in the sense that it uses a very short training sequence. In such a framework the problem is no longer analytically tractable; hence we resort to stochastic sampling techniques. The developed technique samples the channel, the variance of the noise and the symbols in order to build an ergodic Markov chain whose equilibrium distribution is the distribution of interest. The estimates of the MIMO channel and the noise variance are inferred from marginal posterior distributions, which are by-products of the output of the algorithm.

**Index Terms**— MCMC, Stochastic Sampling, MIMO, Turbo principle

## I. INTRODUCTION

A plethora of space-time signal processing and coding techniques have recently emerged in pursuit of the vast capacities promised by the "generalised Shannon theorem". Various flavours of iterative (Turbo) and non-iterative MIMO signal processing schemes have been proposed for that purpose. More generally, detection of information transmitted over the wireless networks is an example of inference in latent variable models. Typically, a prime concern of the estimation process is with the actual data. The parameters characterising the wireless channel (channel transfer function, noise etc.) are just *nuisance parameters*. However, the detection problem complexity can be reduced if the channel transfer functions can be estimated first. Subsequently, the channel estimates can be plugged into a somewhat easier data detection problem.

In this paper we address the problem of Wideband MIMO channel identification from noisy and distorted observations using Markov chain Monte Carlo (MCMC) methods. MCMC techniques originate in statistical physics and were popularised in statistics and machine learning communities [1]. However, recently MCMC receive a great deal of interest amongst the communications research community [2] as well.

In our problem, we assume partial knowledge of the transmitted sequence to avoid identification ambiguities, which can-

not be resolved otherwise (semi-blind framework). In the case where all training sequences are known at the receiver, the channel estimates can be evaluated analytically (e.g. using LS), subject to sufficient length constraint. However, if the training sequence is known only partially, a closed form analytical solution is not possible. For that reason we develop a method based on statistical simulation. The method delivers estimates of the channel transfer functions as well as estimates of the additive Gaussian noise variance. This work can be viewed as extension of time-only Gibbs method presented in [3] to MIMO systems.

## II. MODEL DESCRIPTION AND AIMS

We consider the multiple source digital signalling problem over time dispersive channels. The binary stream of data  $\{d_t\}$  is first transformed by a channel encoder to obtain  $\{b_t\}$  - the encoded (redundant) sequence. Typically convolutional or LDPC (Low Density Parity Check) coding schemes are used for that purpose. The encoded sequence is then permuted  $\{b_{\pi(t)}\}$  and mapped to a sequence of digital modulation symbols  $\{s_t\}$ . To improve the spectral efficiency, the modulated sequence is divided into parallel streams that are transmitted simultaneously from  $M$  transmit antennas (sources).

### A. The equivalent channel

The signal is transmitted through a medium, introducing both noise, delays and attenuations. We consider here a model for the combination of the transmission channel and the pulse shaping. The mixing model is assumed to be a multidimensional time invariant FIR filter. More precisely we assume that the sources are mixed in the following manner, and corrupted by an additive Gaussian i.i.d. noise sequence: at the  $j^{th}$  sensor, and for  $j = 1, \dots, m$  (no constraint between  $m$  and  $n$ )

$$y_t^j = \sum_{i=1}^m h_{i,j}^T s_{t:L+1}^i + \eta_t^j \quad (1)$$

$$= h_j^T s_t + \eta_t^j, \quad (2)$$

where  $L$  is the length of the filters from sources to sensors, assumed to be independent of  $i, j$  and time invariant. The series

$\{\eta_t^j\}$  is a circular zero mean i.i.d. complex Gaussian sequence, i.e. for  $j = 1, \dots, m$  and  $t = 1, \dots$ ,

$$\eta_t^j \stackrel{i.i.d.}{\sim} \mathcal{N}_c(0, \sigma^2/2). \quad (3)$$

This can be reformulated in matrix form as (we stack the observations  $y_t^j$  in  $y_t$ ),

$$y_t = S_t h + \eta_t$$

and for  $T$  observed vectors, we will have

$$y = S h + \eta \quad (4)$$

It is important to emphasise the structure of (4). This is a rather unusual formulation of this problem, since it implies that it is "data that acts on channels" to produce observation.

In what follows we assume that  $h, \sigma^2$  and possibly  $s_{1:T}^{(1:m)}$  are unknown, and it is of interest to estimate  $h, \sigma^2$  (and  $s_{1:T}^{(1:m)}$  as a by-product) from the observations  $y_{1:T}$ .

### B. Bayesian model and prior distributions

We adopt here a Bayesian approach, whereby any prior knowledge concerning the quantities of interest can be incorporated in the inference process through *prior* distributions. Here we chose the following conjugate prior distributions

$$\begin{aligned} h &\sim \mathcal{N}_c(\mu_0, \sigma^2 \Sigma_0) \\ \sigma^2 &\sim \mathcal{IG}(\nu_0, \gamma_0) \end{aligned}$$

which can be made non-informative by taking  $\Sigma_0^{-1} = 0, \nu_0 \ll 1$  and  $\gamma_0 = 0$ . Other priors are also possible, and would not change the algorithms presented later. Now, applying Bayes' rule we can write the joint posterior distribution as

$$\begin{aligned} p(h, \sigma^2, s_{1:T}^{(1:m)} | y_{1:T}) &\propto \\ &\frac{1}{(\sigma^2)^{Tm}} \exp\left(\frac{-1}{\sigma^2} (y - S h)^\dagger (y - S h)\right) \\ &\times \frac{1}{|\sigma^2 \Sigma_0|} \exp\left(\frac{-1}{\sigma^2} (h - \mu_0)^\dagger \Sigma_0^{-1} (h - \mu_0)\right) \\ &\times \frac{\gamma_0^{\nu_0}}{\Gamma(\nu_0)} \frac{1}{(\sigma^2)^{\nu_0+1}} \exp\left(-\frac{\gamma_0}{\sigma^2}\right) \\ &\times p(s_{1:T}^{(1:m)}) \end{aligned}$$

where  $p(s_{1:T}^{(1:m)})$  is the prior distribution of the symbols, which contains the information related to the expected type of data transmitted, but also the structure of the code used for the transmission. We simplify this expression by considering the terms in the exponentials:

$$\begin{aligned} (y - S h)^\dagger (y - S h) + (h - \mu_0)^\dagger \Sigma_0^{-1} (h - \mu_0) + \gamma_0 &= \\ y^\dagger y + (S h)^\dagger (S h) - 2 \operatorname{Re}(y^\dagger S h) + h^\dagger \Sigma_0^{-1} h &+ \\ + \mu_0^\dagger \Sigma_0^{-1} \mu_0 - 2 \operatorname{Re}(\mu_0^\dagger \Sigma_0^{-1} h) + \gamma_0 &= \\ h^\dagger (S^\dagger S + \Sigma_0^{-1}) h - 2 \operatorname{Re}(h^\dagger (S^\dagger y + \Sigma_0^{-1} \mu_0)) + & \\ y^\dagger y + \mu_0^\dagger \Sigma_0^{-1} \mu_0 + \gamma_0 &= \\ (h - \mu)^\dagger \Sigma^{-1} (h - \mu) + \gamma_0 + \mu_0^\dagger \Sigma_0^{-1} \mu_0 + y^\dagger P y & \end{aligned}$$

where

$$\begin{aligned} \Sigma^{-1} &= S^\dagger S + \Sigma_0^{-1} \\ \mu &= \Sigma (S^\dagger y + \Sigma_0^{-1} \mu_0) \\ P &= I - \Sigma \Sigma^\dagger \end{aligned}$$

Consequently the joint posterior distribution can be rewritten as

$$\begin{aligned} p(h, \sigma^2, s_{1:T}^{(1:m)} | y_{1:T}) &\propto \\ &\frac{1}{(\sigma^2)^{Tm}} \exp\left(\frac{-1}{\sigma^2} (h - \mu)^\dagger \Sigma^{-1} (h - \mu)\right) \\ &\times \frac{\gamma_0^{\nu_0}}{\Gamma(\nu_0)} \frac{1}{(\sigma^2)^{Lm}} \frac{1}{(\sigma^2)^{\nu_0+1}} \exp\left(-\frac{\gamma_0 + \mu_0^\dagger \Sigma_0^{-1} \mu_0 + y^\dagger P y}{\sigma^2}\right) \\ &\times p(s_{1:T}^{(1:m)}) \end{aligned}$$

### III. ESTIMATION OBJECTIVES AND COMPUTATIONAL ISSUES

The receiver is built around the classical turbo concatenation with an innovation of a modified Gibbs sampler. The Gibbs sampler and the channel decoder exchange so-called extrinsic information. The extrinsic information is the incremental knowledge gleaned from the decoding process i.e.  $p(b_t^i | \sigma^2, b_{-t}^{(-i)}, y_{1:T})$  where  $b_{-t}^{(-i)}$  means  $b_{1:T}^{(1:m)}$  less  $b_t^{(i)}$ . The interleaved extrinsic information produced by the channel decoder serves as the prior distribution for the Gibbs sampler. The Gibbs sampler produces then a set of marginal posterior distributions for the modulation symbols. The marginal symbol posterior distributions are then transformed into marginal bit posterior distributions that are passed over to the channel decoder after the prior information has been removed and the distributions interleaved.

#### A. Estimation purposes

The turbo decoder requires the evaluation of the family of  $mT$  posterior marginal distributions of the symbols given the observations, in other words  $p(s_t^j | y_{1:T})$  for  $t = 1, \dots, T$  and  $j = 1, \dots, m$ . Whereas it is relatively easy to evaluation the joint posterior distribution of the channels, the variance of the observation noise and the symbols, as it is a simple by-product of the application of Bayes' rule, it is much more difficult in practice to estimate the marginal posterior distributions of the symbols. Indeed, for any  $t = 1, \dots, T$  and  $j = 1, \dots, m$ , the marginal posterior distribution is equal to

$$p(s_t^j | y_{1:T}) = \sum_{s_{t', t'' \neq t, j \neq i}} \int_{R^{Lmn}} p(h, \sigma^2, s_{1:T}^{(1:m)} | y_{1:T}) dh d\sigma^2.$$

To give an idea about the complexity involved with the evaluation of this quantity we consider a simple scenario. Let the modulation be a BPSK,  $T = 100, m = n = 2$  (i.e. two sources and two antennas). In this simple case the number of discrete terms in the sum is  $2^{2 \times 99}$ , i.e. a number of terms of order  $10^{60}$ ! Therefore, even in the case where the integrals over  $h$  and  $\sigma^2$

can be exactly evaluated or reasonably approximated, systematic evaluation of the discrete sum is impossible, and one has to resort to numerical approximations in order to beat the curse of dimensionality.

### B. The Monte Carlo method

Monte Carlo methods have proved to be very efficient at tackling such complex problems. They have been successfully applied in physics for 50 years [4], image processing for nearly 20 years [5] and statistics for over a decade where they have revolutionised Bayesian statistics. The basic principle of Monte Carlo methods consists of replacing the algebraic representation of  $\pi$  by a population based representation [1]. More precisely assume that we know how to produce  $N$  samples, the population, distributed according to  $\pi$ , then the probability of any region  $A$  of  $\mathcal{X}$ , i.e.  $\int_A \pi(x) dx$ , can be approximated by the number of samples that belong to  $A$ . Now if we wish to approximate an integral of the form

$$I(f) = \int_{\mathcal{X}} f(x) \pi(x) dx$$

(where here  $\int$  either means discrete or continuous sum), then a Monte Carlo estimator of  $I(f)$  is given by

$$\hat{I}(f) = \frac{1}{N} \sum_{i=1}^N f(x_i).$$

Intuitively this estimator ought to be efficient, as the samples  $\{x_i\}$  tend to concentrate on regions of high probability (i.e. where information is) and avoid regions of low probability, therefore making the most of the available computational power. This statement can be made mathematically rigorous, and it can be proved that under fairly general conditions, the rate of convergence of this estimator to the true value of the integral is of the order  $O\left(\frac{1}{\sqrt{N}}\right)$ , that is the rate of convergence is independent of the dimension of  $\mathcal{X}$ .

### C. MCMC methods

Sampling from such a distribution can be difficult in practice, however Markov chain Monte Carlo (MCMC) techniques have proved to be able to sample from potentially any type of posterior distribution. Their principle is the following: instead of producing independent and identically distributed (from  $\pi$ ) samples, it is generally easier to build a Markov chain whose invariant distribution is  $\pi$ , that is generate a series of samples that are correlated. Indeed, such Markovian schemes allows for "divide to conquer" techniques as partial update of the samples is possible. This is the approach that we follow here and to be precise we use a Gibbs sampler. The Gibbs sampler relies on the idea that although it might be difficult to sample from the joint posterior distribution  $p\left(h, \sigma^2, s_{1:T}^{(1:m)} | y_{1:T}\right)$ , it might be easier to sample from a subset of its conditional posterior distributions e.g.  $p\left(h | \sigma^2, s_{1:T}^{(1:m)}, y_{1:T}\right)$ ,  $p\left(\sigma^2 | h, s_{1:T}^{(1:m)}, y_{1:T}\right)$ ,  $p\left(s_t^i | h, \sigma^2, s_{-t}^{(-i)}, y_{1:T}\right)$  where  $s_{-t}^{(-i)}$  means  $s_{1:T}^{(1:m)}$  less  $s_t^{(i)}$  ... Intuitively it is likely that it will be easier to update some sub-blocks of the parameters (with the remaining parameters fixed) rather than the complete vector of parameters.

## IV. PROPOSED ALGORITHM

Implementation of the Gibbs sampler for our problem would be ( $\alpha_i$  below means the  $i$ th sample corresponding to parameter  $\alpha$  at iteration  $i$ ):

- Iteration  $i$ 
  - 1) Sample  $h_i \sim p\left(h | \sigma_{i-1}^2, s_{1:T,i-1}^{(1:m)}, y_{1:T}\right)$
  - 2) Sample  $\sigma_i^2 \sim p\left(\sigma^2 | h_i, s_{1:T,i-1}^{(1:m)}, y_{1:T}\right)$
  - 3) For  $t = 1, \dots, N$ , For  $j = 1, \dots, m$  sample  $s_{t,i}^j \sim p\left(s_t^j | h_i, \sigma_i^2, s_{-t}^{(-j)}, y_{1:T}\right)$ .

In the first iteration the unknown symbols are sampled from prior distributions. Since typically, we have no knowledge of these statistics, we initiate the symbols using uniform distributions. Once samples for the channels are generated, the noise variance is updated. With both samples of channels and noise variance updated, the remaining (un-known) symbols can be updated. In this part of the algorithm a number of schemes are possible. The simplest is to sequentially update all symbols "one at a time". In more elaborate versions, one can sample blocks of symbols at a time. We chose to use "one at a time" update with random order, which is different for each iteration.

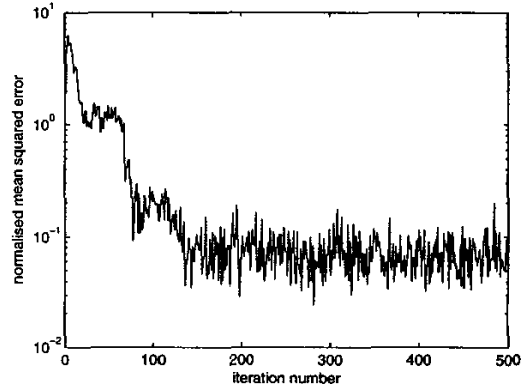


Fig. 1. Channel estimate convergence (semi-blind case): 2 Tx by 2 Rx,  $L = 3$  tap channel (all i.i.d  $\mathcal{N}\left(0, (2L)^{-1}\right)$ ) 8 symbol training sequence.

## V. SIMULATIONS

Figures 1 and 2 present an insight into properties of the proposed method. In both cases it is a 2Tx x 2Rx space time system. All channel taps ( $L=3$ ) are modelled as Gaussian  $\mathcal{N}\left(0, (2L)^{-1}\right)$ , the SNR = 10 dB per receive antenna. A short training sequence of 8 symbols is used to encourage the algorithm to find a "correct mode" of the posterior distribution. As can be seen in both cases the algorithm converges to acceptable levels of MSE very rapidly.

## VI. CONCLUSIONS

In this paper we present advanced Monte Carlo simulation techniques in order to solve the problem of wide-band MIMO

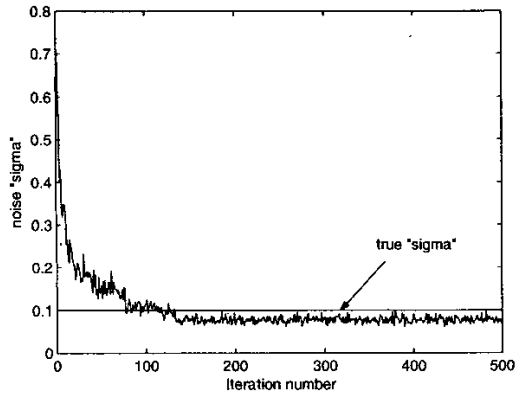


Fig. 2. Convergence of the noise variance estimate (semi-blind case): 2 Tx by 2 Rx,  $L = 3$  tap channel (all i.i.d)  $\mathcal{N}(0, (2L)^{-1})$  8 symbol training sequence.

channel identification. The algorithm relies on a Gibbs sampler procedure. Simulation confirm that our approach is sound.

#### ACKNOWLEDGMENTS

The authors wish to thank Toshiba TREL. Ltd. for sponsoring the work presented in this article.

#### REFERENCES

- [1] Christophe Andrieu, Nando de Freitas, Arnaud Doucet, and Michael I. Jordan, "An introduction to mcmc for machine learning," *Kluwer Machine Learning, Special Issue on MCMC.*, vol. 50, no. 1, pp. 5 – 43, 2003.
- [2] X Wang, R. Chen, and J. Liu, "Monte carlo bayesian signal processing for wireless communications," *Journal of VLSI Signal Processing*, vol. 30, pp. 89 – 105, 2002.
- [3] X. Wang and R. Chen, "Blind turbo equalization in gaussian and non-gaussian noise," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 4, pp. 1092 – 1105, 2001.
- [4] N. Metropolis and S. Ulam, "The monte carlo method," *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335 – 341, 1949.
- [5] S. Gelman and D. Gelman, "Stochastic relaxation, gibbs distributions and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721 – 741, 1984.