# A BAYESIAN NONPARAMETRIC APPROACH FOR BLIND MULTIUSER CHANNEL ESTIMATION

*Isabel Valera*[1], *Francisco J. R. Ruiz*[2], *Lennart Svensson*[3], *Fernando Perez-Cruz*[24]

[1] Max Planck Institute for Software Systems
Kaiserslautern, Germany
ivalera@mpi-sws.org

[2] Universidad Carlos III de Madrid
Department of Signal Processing and Communications
Leganés, Spain
{franrruiz, fernando}@tsc.uc3m.es

[3] Chalmers University of Technology
Department of Signals and Systems
Gothenburg, Sweden
lennart.svensson@chalmers.se

[4] Technical Staff at Bell Labs
Alcatel-Lucent, NJ

## ABSTRACT

In many modern multiuser communication systems, users are allowed to enter and leave the system at any given time. Thus, the number of active users is an unknown and time-varying parameter, and the performance of the system depends on how accurately this parameter is estimated over time. We address the problem of blind joint channel parameter and data estimation in a multiuser communication channel in which the number of transmitters is not known. For that purpose, we develop a Bayesian nonparametric model based on the Markov Indian buffet process and an inference algorithm that makes use of slice sampling and particle Gibbs with ancestor sampling. Our experimental results show that the proposed approach can effectively recover the data-generating process for a wide range of scenarios.

***Index Terms—*** Bayesian nonparametric, factorial HMM, multiuser communication, machine-to-machine.

## 1. INTRODUCTION

One of the trends in wireless communication networks (WCNs) is the increase of heterogeneity. It is not new that users of WCNs are no longer only humans talking, but they also include machine-to-machine (M2M) communications, involving communication between a sensor/actuator and a corresponding application server in the network. Moreover, although there are millions of M2M cellular devices currently operating in WCNs, the industry expects this number to increase ten-fold in the years to come [1]. However, unlike consumer traffic, which is characterized by a small number of long lived sessions, the M2M traffic involves a large number of short-lived sessions with transactions of a few hundred bytes [1]. This results in a change of the traffic in WCNs, leading to multiuser communication systems in which a large numbers of users may aim to enter or leave the system (i.e., start or stop transmitting) at any given time. In this context, establishing dedicated bearers for data transmission may be highly inefficient [1]. Thus, the first question that arises is how to allow the users access the system in a way that the signaling overhead is reduced. Previous works have found that transmitting small pieces of information in the random access request itself is more efficient [2].

In this paper, we focus on the problem of determining the number of users transmitting in a memoryless communication system jointly with the channel estimation and the detection of the transmitted data. This problem appears in several specific applications. For instance, in the context of wireless sensor networks, where the communication nodes can often switch on and off asynchronously during operation. It also appears in massive multiple-input multiple-output (MIMO) multiuser communication systems [3], in which the base station has a very large number of antennas and the mobile devices use a single antenna to communicate within the network. In a code-division multiple access (CDMA) context, a set of terminals randomly access the channel to communicate with a common access point, which receives the superposition of signals from the active terminals only [4]. In the context of CDMA systems, several recent papers addressing the problem of user activity and identification can be found in the literature. The authors in [5] propose a method to identify the number and identity of the active users in a direct sequence CDMA (DS-CDMA) system, by using a set of training data. Therefore, no symbol detection is performed in this stage. In [6], a Bayesian approach, restricted to the case where the channel has been previously estimated, is presented. More recently, in [4], the authors solve the user identification problem while performing joint channel estimation and data detection. A characteristic shared by all these methods is the assumption of an explicit upper bound for the number of transmitters, which makes sense in a DS-CDMA system but may represent

[1] Francisco and Fernando are also affiliated to the Gregorio Marañón Health Research Institute. Francisco J. R. Ruiz is supported by an FPU fellowship from the Spanish Ministry of Education (AP2010-5333). Isabel Valera acknowledges the support of *Plan Regional-Programas I+D* of *Comunidad de Madrid* (AGES-CM S2010/BMD-2422). This work is also partially supported by Ministerio de Economía of Spain (projects 'COMONSENS', id. CSD2008-00010, and 'ALCIT', id. TEC2012-38800-C03-01), by Comunidad de Madrid (project 'CASI-CAM-CM', id. S2013/ICE-2845), and by the European Union 7th Framework Programme through the Marie Curie Initial Training Network 'Machine Learning for Personalized Medicine' (MLPM2012, Grant No. 316861).

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

a limitation in other scenarios.

In this paper, we propose a Bayesian nonparametric (BNP) model which, due to its nonparametric nature, becomes flexible enough to account for any number of transmitters, without the need of additional previous knowledge or bounds. Moreover, the proposed model allows us to solve this problem in a fully unsupervised way, without requiring signaling data. In particular, we assume a potentially infinite number of transmitters that might start transmitting short bursts of symbols at any time, such that only a finite subset of the transmitters become active during an observation period while the remaining (infinite) transmitters remain in an idle state (i.e., they do not transmit). Our approach consists in modeling all transmitters as an unbounded number of independent chains in an infinite factorial hidden Markov model (iFHMM) [7], in which each chain representing a transmitter has high probability of remaining in its current state (either active or idle). Under this model, the symbols sent by each transmitter can be viewed as a hidden sequence that the receiver needs to reconstruct from the observations (i.e., the received sequence). Our experimental results show that the proposed approach efficiently solves user identification, channel estimation and data detection in a jointly and fully blind way and, as a consequence, they shed light on the suitability of BNPs applied to signal processing for communications.

## 2. PROPOSED SYSTEM MODEL

Let us assume a multiuser digital memoryless communication system with $M$ transmitters (users) and $R$ receiving antennas. Each receiving antenna observes a linear combination of all the transmitted data sequences, corrupted by additive white Gaussian noise (AWGN).[1] Specifically, the $R$-dimensional observation vector compound of the observations at all the receiving antennas at time instant $t$ can be written as

$$\mathbf{y}_t = \sum_{m=1}^{M} \mathbf{h}_m x_{tm} + \mathbf{n}_t, \tag{1}$$

where $x_{tm}$ is the symbol sent by transmitter $m$ at time instant $t$, $\mathbf{h}_m$ is an $R$-vector containing the channel coefficients between the $m$-th transmitter and all the receiving antennas, and $\mathbf{n}_t$ denotes the additive noise. Figure 1 shows a diagram of such digital communication system, where a set of users or transmitters ('Tx') send their messages to a unique receiver ('Rx'), equipped with multiple antennas.

Transmitters are allowed to start or stop transmitting at any given time and, while active, the $m$-th transmitter sends symbols $x_{tm}$ that belong to a complex constellation $\mathcal{A}$, with cardinality $|\mathcal{A}|$. While idle, we can assume that $x_{tm} = 0$ and, therefore, each symbol $x_{tm} \in \mathcal{A} \bigcup \{0\}$. Moreover, according to (1), the observation $\mathbf{y}_t$ only depends on the transmitted

[1]Our model is also applicable for non-Gaussian noise models.

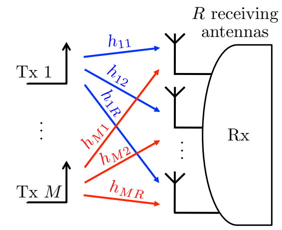



**Fig. 1**: Illustration of a digital communication system.

symbols by the active transmitters at time $t$. Hence, we can assume an infinite number of transmitters in the system ($M \to \infty$) by ensuring that only a finite subset $M_+$ of them become active during the observation period $\{1, \ldots, T\}$, while the rest do not influence the observations.

In a realistic scenario, symbols tend to be transmitted as bursts. We model this effect with a first-order HMM that favors high self-transition probabilities of the active and inactive states. We introduce the auxiliary binary variable $s_{tm}$ to indicate whether the $m$-th transmitter is active at time $t$, such that $x_{tm} = 0$ if $s_{tm} = 0$ and $x_{tm} \sim \mathcal{U}(\mathcal{A})$ if $s_{tm} = 1$, being $\mathcal{U}(\mathcal{A})$ the uniform distribution over the set $\mathcal{A}$. We additionally introduce the variables $a_m$ to denote the self-transition probability of the inactive state for the $m$-th transmitter, such that $a_m = p(s_{tm} = 0 | s_{(t-1)m} = 0)$, and the variables $b_m$ to denote the transition probability from active to inactive, i.e., $b_m = p(s_{tm} = 0 | s_{(t-1)m} = 1)$, and we assume a dummy initial state 0 for all transmitters at time $t = 0$. Figure 2 shows an example of our factorial HMM.

In order to complete the description of our model, we need to specify the prior distribution over all the hidden variables. To allow for an infinite number of transmitters, we rely on the iFHMM in [7] and place a Markov Indian buffet process (mIBP) prior over the $T \times M$ matrix $\mathbf{S}$ that contains all variables $s_{tm}$, i.e., $\mathbf{S} \sim \text{mIBP}(\alpha, \gamma_1, \gamma_2)$. The mIBP is a prior over binary matrices with an infinite number of columns (transmitters), in which each column follows an HMM. The mIBP ensures that only a finite number of the columns become active for any finite number of rows (observations) $T$. This prior distribution is obtained by placing the priors

$$a_m \sim \text{Beta}\left(1, \frac{\alpha}{M}\right), \quad b_m \sim \text{Beta}(\gamma_1, \gamma_2), \tag{2}$$

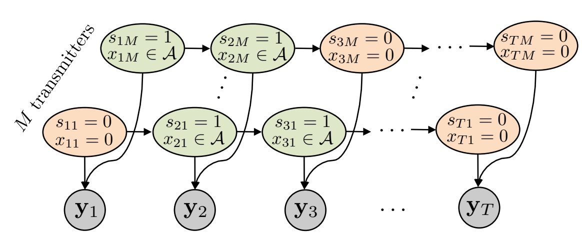



**Fig. 2**: Illustration of an example of our factorial model.

and taking the limit as $M \to \infty$ [7]. Regarding the hyper-parameters of the model, in our experiments we set $\alpha = 1$, $\gamma_1 = 0.1$ and $\gamma_2 = 2$. Note that these values of $\gamma_1$ and $\gamma_2$ favor the state persistence of the active state.

We finally place a circularly symmetric complex Gaussian prior distribution with independent elements over the channel coefficients $\mathbf{h}_m$ and the noise $\mathbf{n}_t$ of the form

$$\mathbf{h}_m \sim \mathcal{CN}(\mathbf{0}, \sigma_h^2 \mathbf{I}), \quad \mathbf{n}_t \sim \mathcal{CN}(\mathbf{0}, \sigma_y^2 \mathbf{I}). \tag{3}$$

This corresponds to a Rayleigh fading AWGN channel. We place an InvGamma$(2, 3)$ prior over the variable $\sigma_h^2$.

## 3. INFERENCE

We now focus on the joint estimation of the number of active users, the symbols they transmit and the channel coefficients given the sequence of observations $\{\mathbf{y}_1, \dots, \mathbf{y}_T\}$. To this end, we derive a Markov chain Monte Carlo (MCMC) algorithm which sequentially samples from the posterior probability of the unknown variables as follows:

- **Step 1:** Add $M_{\text{new}}$ new inactive transmitters. Note that at this step we restrict the considered number of transmitters in the system to $M_+ + M_{\text{new}}$.
- **Step 2:** Jointly sample the states $s_{tm}$ and the symbols $x_{tm}$ sent by the considered transmitters in the observation period. Remove those transmitters that remain inactive in the whole observation period (updating $M_+$).
- **Step 3:** For each active transmitter $m = 1, \dots, M_+$, sample the transition probabilities ($a_m$ and $b_m$) and the channel coefficients $\mathbf{h}_m$. Then, sample $\sigma_h^2$.

In **Step 1**, we make use of a slice sampling algorithm that resorts to the stick-breaking construction of the mIBP to add a set of new inactive transmitters, sampling their self-transition probabilities of the inactive state $a_m$ [8]. For these new inactive users, the probabilities $b_m$ and the channel coefficients $\mathbf{h}_m$ are drawn from the prior. **Step 2** consists in a particle Gibbs with ancestors sampling (PGAS) algorithm [9]. We make use of this algorithm because it allows us to jointly sample all the states $s_{tm}$ and transmitted symbols $x_{tm}$ of all the considered transmitters for each time instant $t$. This leads to a significant improvement in the mixing properties of the algorithm in comparison to the forward-filtering backward-sampling in [7], which samples the sequence of states and symbols sequentially for each transmitter (and conditioning on the symbols of the remaining transmitters). Afterwards, we remove the transmitters that remain inactive during the whole observation period, consequently updating $M_+$. In **Step 3**, we sample the parameters of the active users (the transition probabilities $a_m$ and $b_m$ and the channel coefficients $\mathbf{h}_m$) and the variance $\sigma_h^2$ from their posterior distributions.

## 4. EXPERIMENTS

We now run a battery of experiments to illustrate the performance of the proposed approach. To this end, we simulate different scenarios of a multiuser communication system, considering different values for the number of transmitters, the number of receiving antennas and the signal-to-noise ratio (SNR).

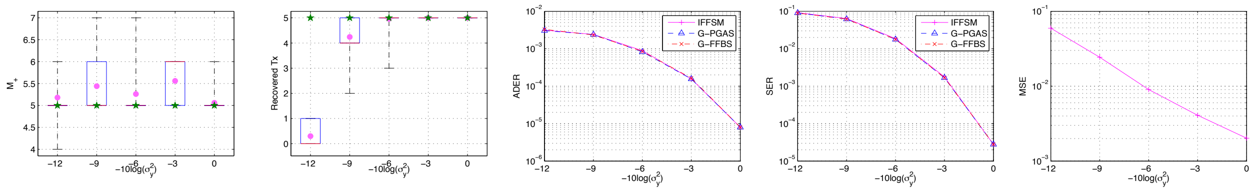
To generate the observations, we assume that each of the transmitters sends a burst of symbols during the observation period of length $T = 1000$. Transmitters use quadrature amplitude modulation (QAM) with cardinality $|\mathcal{A}|$, being the symbols in the constellation normalized to yield unit energy. Each transmitter becomes active at a random instant, uniformly sampled in the interval $[1, T/2]$, being the burst duration $T/2$. This ensures that the bursts of different transmitters overlap. A Rayleigh AWGN channel is assumed, i.e., the channel coefficients and the noise are circularly symmetric complex Gaussian distributed with zero mean, being the covariances matrices $\mathbf{I}$ and $\sigma_y^2\mathbf{I}$, respectively, where $\sigma_y^2$ depends on the considered SNR, which we define as

$$\text{SNR (dB)} = -10 \log_{10}(\sigma_y^2). \tag{4}$$

We choose the following parameters to run our experiments: 5 transmitters in the system, 20 receiving antennas, quadrature phase-shift keying (QPSK) modulation (i.e., $|\mathcal{A}| = 4$ symbols in the constellation), and SNR $= -3$ dB. Using this base configuration, we vary one of the parameters while holding the rest fixed.

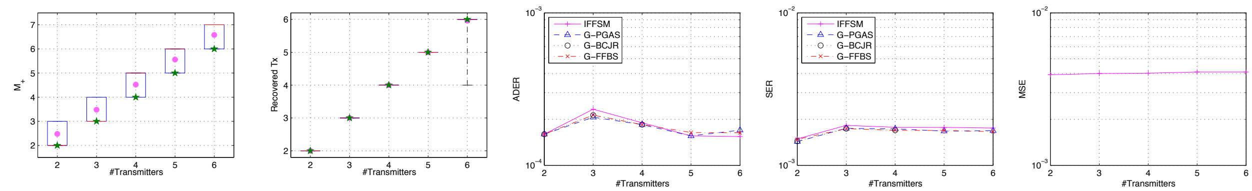
We observe in our experiments that the mixing properties of the proposed inference algorithm degrades as we increase the SNR. This is due to the fact that the posterior distribution of the channel coefficients becomes narrow as the noise variance decreases and, as a consequence, an inference algorithm based on random exploration needs a large number of iterations to find the peaks of the posterior distribution. In order to improve the performance of the inference algorithm, we propose a solution based on an heuristic to artificially widen the posterior distribution by adding artificial noise to the observations. In more detail, at each iteration of the algorithm, we linearly increase the SNR by reducing the variance of the artificial noise, and we repeat this procedure until we reach the actual value of the SNR. In our experiments, we initialize the inference algorithm with SNR $= -12$ dB, increasing this number by $0.002$ dB at each iteration of the algorithm.

In order to evaluate the performance of the proposed approach, we consider that the inference algorithm has recovered a transmitter if the symbol error rate (SER) for that transmitter is below a threshold of $0.1$. For the recovered transmitters, we evaluate the performance in terms of the activity detection error rate (ADER), the SER, and the mean square error (MSE) of the channel coefficient estimates. The ADER is the probability of detecting activity (inactivity) in a transmitter while that transmitter is actually inactive (active). When




(a) Inferred $M_+$. (b) Recovered transmitters. (c) ADER. (d) SER. (e) MSE.

**Fig. 3**: Results for different SNR's.




(a) Inferred $M_+$. (b) Recovered transmitters. (c) ADER. (d) SER. (e) MSE.

**Fig. 4**: Results for different number of transmitters.

computing the SER, an error is computed at time $t$ whenever the estimated symbol for a transmitter differs from the actual transmitted symbol, considering that the transmitted symbol while inactive is $x_{tm} = 0$. We compute the MSE as

$$\text{MSE} = \frac{1}{RM_+} \sum_{m=0}^{M_+} \|\mathbf{h}_m - \hat{\mathbf{h}}_m\|^2, \tag{5}$$

where $\hat{\mathbf{h}}_m$ is the $R$-dimensional vector containing the inferred channel coefficients corresponding to the $m$-th transmitter.

We compare our approach (denoted by IFFSM in the plots) with three genie-aided methods which have perfect knowledge of the true number of transmitters and channel coefficients.[2] In particular, we run:

(i) The PGAS algorithm that we use in Step 2 of our inference algorithm (referred as G-PGAS).

(ii) The FFBS algorithm over the equivalent factorial HMM with state space cardinality $|\mathcal{A} \bigcup \{0\}|$ (referred as G-FFBS).

(iii) The optimum BCJR algorithm [10], over an equivalent single HMM with a number of states equal to $|\mathcal{A} \bigcup \{0\}|^{N_t}$, being $N_t$ the true number of transmitters (referred as G-BCJR).

For each considered scenario, we run $50$ independent simulations, each with different simulated data. We run $20,000$ iterations of our inference algorithm, finally obtaining the inferred symbols $\hat{x}_{tm}$ as the component-wise *maximum a posteriori* (MAP) solution, only considering the last $2,000$ iterations of the sampler. The estimates of the channel coefficients $\hat{\mathbf{h}}_m$ are then obtained as the MAP solution, conditioned on the data and the inferred symbols $\hat{x}_{tm}$. For the BCJR algorithm, we obtain the symbol estimates according to the component-wise MAP solution for each transmitter $m$ and each instant $t$. For the genie-aided PGAS and FFBS methods, we run the algorithms for $10,000$ iterations and considering the last $2,000$ samples to obtain the symbol estimates.

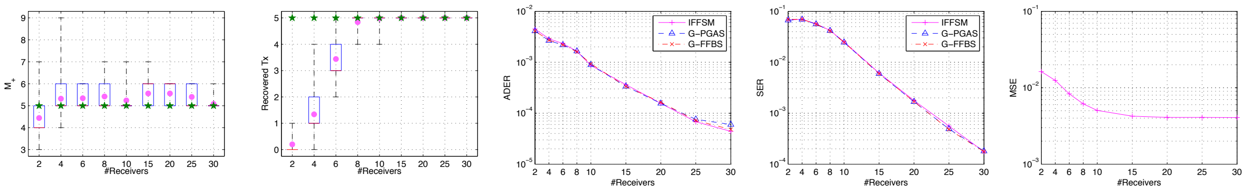
Figure 3 shows the results when the SNR varies from $-12$ dB to $0$ dB. Specifically, we show the box-plot representation[3] of the inferred number of transmitters $M_+$ and the number of recovered transmitters as well as the ADER, the SER, the MSE. As expected, we obtain a better performance as the SNR increases. For low values of SNR, transmitters are more likely to be masked by the noise and, therefore, our algorithm cannot recover the transmitters with a SER below $0.1$, although it still finds five hidden chains. We also observe that the performance (in terms of ADER and SER) of the proposed iFHMM reaches similar values to the genie-aided methods.

Figure 4 shows the results when the true number of transmitters changes from 2 to 6. As expected, the performance degrades as the number of transmitters increases, because more parameters need to be estimated. Nevertheless, we can observe that the iFHMM recovers the true number of transmitters, with similar SER and ADER values than the genie-aided approaches.

Figure 5 shows the results when the number of receiving antennas varies from 2 to 30. In this figure, we observe that, as

[2] For the genie-aided methods, we use $a^m = 0.998$ and $b^m = 0.002$.

[3] We depict the 25-th, 50-th and 75-th percentiles in the standard format, as well as the most extreme values. Moreover, the mean value is represented with a pink circle, and the true number of transmitters $M_+$ is represented with a green star.




(a) Inferred $M_+$. (b) Recovered transmitters. (c) ADER. (d) SER. (e) MSE.

**Fig. 5**: Results for different number of receiving antennas.

the number of receiving antennas increases, our algorithm is able to recover more transmitters, with a similar performance (in terms of SER and ADER) than the genie-aided methods. We also observe that the MSE decreases when the number of receivers increases, but after approximately 15 receivers, the curve flattens as it reaches the threshold imposed by the noise level.

## 5. CONCLUSIONS

We have proposed a blind approach for joint channel estimation and detection of the transmitted data when the number of transmitters is unknown. Our approach is based on a BNP model (the iFHMM), for which we have derived an efficient inference algorithm that exploits the properties of MCMC and sequential Monte Carlo techniques. Our experiments on different communication scenarios show that the proposed approach efficiently solves the problem of joint user identification, channel estimation and data detection in a fully blind way. These results are promising for the suitability of BNPs applied to signal processing for communications. However, there is still a lot of work to do in this area. Further research lines may focus on the adaptation of this model to frequency-selective or time-varying channels, or on deriving an online inference algorithm suitable for a real communication system.

## REFERENCES


[1] H. S. Dhillon, H. C. Huang, H. Viswanathan, and Valenzuela R. A., “Fundamentals of throughput maximization with random arrivals for M2M communications,” *CoRR*, vol. abs/1307.0585, 2013.

[2] Y. Chen and W. Wang, “Machine-to-machine communication in LTE-A,” in *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, Sept 2010, pp. 1–4.

[3] J. Hoydis, S. ten Brink, and M. Debbah, “Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160 – 171, Feb. 2013.

[4] M. A. Vázquez and J. Míguez, “User activity tracking in DS-CDMA systems,” *Vehicular Technology, IEEE Transactions on*, vol. 62, no. 7, pp. 3188–3203, Sept 2013.

[5] D. Angelosante, E. Grossi, G. B. Giannakis, and M. Lops, “Sparsity-aware estimation of CDMA system parameters,” in *Signal Processing Advances in Wireless Communications, 2009. SPAWC ’09. IEEE 10th Workshop on*, June 2009, pp. 697–701.

[6] H. Zhu and G. B. Giannakis, “Exploiting sparse user activity in multiuser detection,” *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 454–465, Feb. 2011.

[7] J. Van Gael, Y. W. Teh, and Z. Ghahramani, “The infinite factorial hidden Markov model,” in *Advances in Neural Information Processing Systems (NIPS)*, 2009, vol. 21.

[8] Y. W. Teh, D. Görür, and Z. Ghahramani, “Stick-breaking construction for the Indian buffet process,” in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2007, vol. 11.

[9] F. Lindsten, M. I Jordan, and T. B. Schön, “Particle Gibbs with ancestor sampling,” *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 2145–2184, 2014.

[10] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate,” *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.