Joint Activity and Data Detection for Machine to Machine Communication via Bayes Risk Optimization

Fabian Monsees, Carsten Bockelmann and Armin Dekorsy
Department of Communications Engineering
University of Bremen, Germany
E-mail: {monsees,bockelmann,dekorsy}@ant.uni-bremen.de

Abstract—Performing joint detection of activity and data is a promising approach to reduce management overhead in Machine-to-Machine communication. However, erroneous activity detection has severe impacts on the system performance. Estimating an active node or user erroneously to be inactive results in a loss of data. To optimally balance activity and data detection, we derive a novel joint activity and data detector that bases on the minimization of the Bayes Risk. The Bayes Risk detector allows to control error rates with respect to the activity detection dynamically by a parameter that can be controlled by higher layers. In this paper we derive the Bayes Risk detector for a general linear system and present exemplary results for a specific Machine-to-Machine communication scenario.

I. Introduction

Machine-to-Machine (M2M) communication is expected to grow tremendously in the next years, raising new challenges for nowadays communication systems which are mainly designed to meet the demands for human based communication. In particular, M2M devices are often devices of reduced complexity and require only low data-rates for communication. A broad field of application for future M2M networks comes from industrial approaches such as smart metering, surveillance purposes or logistics [1]. For these applications, a huge number of M2M devices are connected to a central aggregation node and transmit only on occasion or event driven, making communication sporadic. One major problem in state of the art communication systems such as LTE is that control signaling or sophisticated resource management is not well suited for low-rate communication of sensor nodes. As an example, in the current frame structure of LTE-A, at most a few tens of M2M devices can be supported with control channel elements (CCEs) allocated within a subframe [2]. This number is far below of what future M2M applications will require.
One approach to reduce control signaling and management overhead is to shift activity detection to the physical layer by performing a joint activity and data detection. Thus, we propose to extend common state of the art detectors to perform activity and data detection jointly.
In this paper we consider a scenario where sensor nodes sporadically transmit their data to a central aggregation node. The aggregation node detects if a node is active or not. In this setup, inactive nodes can simply be modeled as transmitting zeros instead of modulation symbols. This means, the multi-user signal that is composed of all sensor nodes no matter if active or inactive is a sparse signal. In [3] the authors derived the sparsity aware Maximum a posteriori detector for this setup. One major drawback of this detector is that the activity detection tends to decide in favor of inactivity. As a consequence, it is very likely that active nodes are wrongly estimated to be inactive, especially at low Signal-to-Noise Ratio (SNR).
As the activity detection severly impacts the overall system performance, we would like to stress that an active node that is wrongly estimated to be inactive is generally worse than the opposite. In the first case data is simply lost, whereas in the latter case wrong transmissions can be identified by higher layer error handling mechanisms such as check codes. Consequently, from an intuitive point, the so called False Inactive errors have a higher impact on the system performance than the False Active errors.
In this paper we address optimal joint activity and data detection by applying a Bayes Risk Sphere Detector that allows to balance the activity detection by a tuning parameter. This tuning parameter can be controlled by higher layer protocols, resulting in a cross-layer design.

II. Bayes Risk Maximum a Posteriori Detection

A. Sporadic Communication Model

In the following we consider an uplink system, where K sensor nodes transmit data to a central aggregation node. We assume that the nodes have a very low activity factor and assume that each node is active with probability p_a. We further assume that the system is linear and can be described by the canonical input-output relation

$$\mathbf{y} = \mathbf{T}\mathbf{x} + \mathbf{w}. \tag{1}$$

Here $\mathbf{y} \in \mathbb{R}^M$ is the received signal at the aggregation node. The matrix $\mathbf{T} \in \mathbb{R}^{M \times K}$ summarizes the channels between the K nodes and the aggregation node. Without giving any further specifications on the elements of $\mathbf{T}$, we restrict to the case where $M \geq K$ holds. Scenarios where $M < K$ are strongly related to the field of Compressed Sensing and are (in another context) addressed in [4]. The noise vector $\mathbf{w} \sim \mathcal{N}\left(\sigma_n^2, 0\right)$ obeys an uncorrelated zero mean Gaussian distribution with variance σ_n^2. The source vector $\mathbf{x}$ contains the modulated

symbols from the sensor nodes and the kth symbol is denoted by x_k. Inactive nodes are modeled as transmitting zeros and we have $x_k = 0$. Active nodes modulate their data using the modulation alphabet $\mathcal{A}$ which can be any real alphabet such as Amplitude Shift Keying (ASK) or Binary Phase Shift Keying (BPSK) and we have $x_k \in \mathcal{A}$. To ease notation, we consider the communication system to be a real valued system which is not a general restriction. The scheme we propose can also be applied to complex systems using a equivalent real valued system description.

B. Generalized Likelihood Ratio Test

With the previously introduced sporadic communication model, the goal of the detector is to detect symbols from the augmented modulation alphabet $\mathcal{A}_0 = \mathcal{A} \cup 0$. In [3] the authors derived the MAP detector for the system in (1) given the knowledge about the a priori distribution of the source vector $\mathbf{x} \in \mathcal{A}_0^K$. However, this approach simply minimizes the total symbol errors on the augmented alphabet $\mathcal{A}_0$ and activity errors and symbol errors are considered to have the same impact on the system.

In the following we consider the activity detection in detail. For each estimate $\hat{x}_k$ the detector has to estimate whether the node was active, i.e., $\hat{x}_k \in \mathcal{A}$ or inactive, i.e., $\hat{x}_k = 0$. The possible outcomes for the activity detection task are summarized as follows. Table I shows the so called Confusion

	$x_k \in \mathcal{A}$	$x_k = 0$
$\hat{x}_k \in \mathcal{A}$	True Active	False Active
$\hat{x}_k = 0$	False Inactive	True Inactive

TABLE I
CONFUSION MATRIX FOR ACTIVITY DETECTION

Matrix of the possible mappings between the true classes (node is active or inactive) and the hypothesized classes (node is active or inactive) [5]. The two error events (False Active and False Inactive) are of special interest. If a node is wrongly estimated to be inactive, the data of this node is simply lost and cannot be recovered. However, in the the opposite case, when a node is wrongly estimated to be active, higher layer error detection methods can sort out wrong packets. Thus, we observe that these error events have different impact on the overall system performance. This necessitates a detector that can be controlled by an additional parameter to balance the activity errors.

To include this balancing in the detector, we consider a binary hypothesis testing problem to determine whether a node is active or inactive.

$$H1 : x_k \in \mathcal{A} \rightarrow \text{Node is active}$$
$$H2 : x_k = 0 \rightarrow \text{Node is inactive}$$

Following Table I we assign the following costs for wrong activity detection.

- C_{Fi} : Cost for estimating false inactive
- C_{Fa} : Cost for estimating false active

Costs for correct decisions are implicitly set to zero. We cast the problem as the minimization of the Bayes Risk defined as

$$\mathcal{R} = C_{\text{Fa}} \Pr(x_k = 0) \int_{Z_{\mathcal{A}}} p(y_m | x_k = 0) \, dy_m +$$
$$C_{\text{Fi}} \Pr(x_k \in \mathcal{A}) \int_{Z_0} p(y_m | x_k \in \mathcal{A}) \, dy_m. \quad (2)$$

The Bayes Risk given in (2) expresses the overall risk of wrong activity detection. $p(y_m | x_k)$ denotes the density for an observation y_m under the hypothesis x_k, $\Pr(\cdot)$ is an event probability and Z_0 and $Z_{\mathcal{A}}$ are the regions in the observation space where the detector would estimate the node to be inactive $\hat{x}_k = 0$ or active $\hat{x}_k \in \mathcal{A}$. For the sake of clarity we omit lengthy derivations and refer the reader to [6]. With the results from [7] we cast (2) as a Generalized Likelihood Ratio Test (GLRT) which has the following form

$$\frac{p(y_m | x_k = 0) \Pr(x_k = 0) C_{\text{Fa}}}{\max_{x \in \mathcal{A}} p(y_m | x_k) \Pr(x_k) C_{\text{Fi}}} \underset{H1}{\overset{H2}{\gtrless}} 1 \quad (3)$$

The GLRT in (3) evaluates the weighted a posteriori probabilities for both hypothesis and assigns the observation to the hypothesis with higher probability. Moreover, hypothesis $H1$ covers the set $\mathcal{A}$ and is therefore a so called composite hypothesis. As (3) shows we compare the best model in $H1$ with H2. Instead of taking the best model in H2 one could take the average model in H2 by replacing the max by a mean operator. This would lead to the so called Bayes Factor which is not investigated further in this paper. It can be seen that (3) can be rewritten as

$$\hat{x}_k = \underset{x_k \in \mathcal{A}_0}{\text{argmax}} \; p(y_m | x_k) \Pr(x_k) C(x_k), \quad (4)$$

with

$$C(x_k) = C_{\text{Fi}}^{\mathbf{1}_{\mathcal{A}}(x_k)} C_{\text{Fa}}^{1 - \mathbf{1}_{\mathcal{A}}(x_k)}. \quad (5)$$

Here, $\mathbf{1}_{\mathcal{A}}(\cdot)$ is the indicator function which takes the value 1 if the argument is contained in the set $\mathcal{A}$. The solution to the optimization problem given in (4) is the optimal scalar estimate for x_k given a scalar observation y_m. Assuming i.i.d. variables x_k and i.i.d. observations y_m (4) can be extended for vector observations and vector hypothesis and we have

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathcal{A}_0}{\text{argmax}} \; \prod_{m=1}^{M} p(y_m | \mathbf{x}) \underbrace{\prod_{k=1}^{K} \Pr(x_k) C(x_k)}_{a(\mathbf{x})}. \quad (6)$$

In the following we consider the system given in (1). The noise vector $\mathbf{w}$ is a vector of uncorrelated zero mean Gaussian distributed random variables having variance σ_n^2. The a posteriori probability for the element y_m with vector hypothesis $\mathbf{x}$ can now be expressed as

$$p(y_m | \mathbf{x}) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{1}{2\sigma_n^2} |y_m - \mathbf{T}_m \mathbf{x}|^2\right], \quad (7)$$

and

$$\prod_{m=1}^{M} p(y_m | \mathbf{x}) = \frac{1}{(2\pi)^{M/2} \sigma_n^M} \exp\left[-\frac{1}{2\sigma_n^2} \|\mathbf{y} - \mathbf{T}\mathbf{x}\|_2^2\right]. \quad (8)$$

Here $\mathbf{T}_m$ is the mth row vector of the matrix $\mathbf{T}$. The elements of the source vector x_k obey a Bernoulli / Uniform distribution. In particular, we have

$$\Pr\left(x_k \in \mathcal{A}\right) = 1 - \Pr\left(x_k = 0\right) = p_a \tag{9}$$

$$p\left(x_k | x_k \in \mathcal{A}\right) = \frac{1}{|\mathcal{A}|} \tag{10}$$

and we write the a priori probability of the source vector $a(\mathbf{x})$ weighted with the costs from (5) as

$$a(\mathbf{x}) = \prod_{k=1}^{K} \left[C_{\mathrm{Fa}}\left(1 - p_a\right)\right]^{1 - \mathbf{1}_{\mathcal{A}}(x_k)} \left(C_{\mathrm{Fi}} \frac{p_a}{|\mathcal{A}|}\right)^{\mathbf{1}_{\mathcal{A}}(x_k)}$$

$$= \left[C_{\mathrm{Fa}}\left(1 - p_a\right)\right]^{K - \sum_k \mathbf{1}_{\mathcal{A}}(x_k)} \left(C_{\mathrm{Fi}} \frac{p_a}{|\mathcal{A}|}\right)^{\sum_k \mathbf{1}_{\mathcal{A}}(x_k)}, \tag{11}$$

with (8) and (11) we can finally rewrite the optimization problem given in (6) by taking logarithms and rearranging we obtain

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x} \in \mathcal{A}_0} \|\mathbf{y} - \mathbf{T}\mathbf{x}\|_2^2 + 2\sigma_n^2 \sum_{k=1}^{K} \mathbf{1}_{\mathcal{A}}\left(x_k\right) \log\left(\frac{C_{\mathrm{Fa}}}{C_{\mathrm{Fi}}} \frac{1 - p_a}{p_a/|\mathcal{A}|}\right). \tag{12}$$

Moreover, the indicator function $\mathbf{1}_{\mathcal{A}}\left(x_k\right)$ can be replaced by the zero pseudo-norm defined as $\|\mathbf{x}\|_0 = \#\{x_k : x_k \neq 0\}$. Thus we obtain the non-convex integer optimization problem

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x} \in \mathcal{A}_0} \|\mathbf{y} - \mathbf{T}\mathbf{x}\|_2^2 + 2\sigma_n^2 \|\mathbf{x}\|_0 \log\left(\frac{C_{\mathrm{Fa}}}{C_{\mathrm{Fi}}} \frac{1 - p_a}{p_a/|\mathcal{A}|}\right). \tag{13}$$

To ease notation, we express the ratio of costs as $\Omega = \frac{C_{\mathrm{Fa}}}{C_{\mathrm{Fi}}}$. Choosing $\Omega = 1$ (13) leads to the MAP detector from [3]. This detector generalizes MAP detection by the additional weighting parameter Ω. This parameter allows to adjust the cost for the activity detection adaptively and system dependent. The penalty term $\lambda(\Omega) = 2\sigma_n^2 \log\left(\Omega \frac{1 - p_a}{p_a/|\mathcal{A}|}\right)$ reflects the a priori assumption about the source vector $\mathbf{x}$, and scales with the noise power. Additionally, Ω scales the a priori assumption and can be interpreted as an additional parameter that determines whether the detector is conservative $\Omega > 1$ or liberal $\Omega < 1$. A conservative detector will decide in favor of inactivity and will produce less False Active errors. In contrast a liberal detector will decide in favor of activity and produce less False Inactive errors than a conservative detector.

C. Bayes-Risk Sphere-Detector (BR-SD)

In this section we briefly show the possible realization of the Bayes Risk detector as a Sphere Detector which we later call Bayes Risk Sphere Detector (BR-SD). Thus, we rewrite the non-convex optimization problem (13) as the application of the QR-decomposition of $\mathbf{T} = \mathbf{Q}\mathbf{R}$ with $\mathbf{Q}$ being an unitary matrix, i.e., $\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}_M$, and $\mathbf{R}$. We can rewrite (13)

as

$$\operatorname*{argmin}_{\mathbf{x} \in \mathcal{A}_0} \|\mathbf{y} - \mathbf{Q}\mathbf{R}\mathbf{x}\|_2^2 + \|\mathbf{x}\|_0 \lambda(\Omega)$$

$$\operatorname*{argmin}_{\mathbf{x} \in \mathcal{A}_0} \|\mathbf{Q}^T\mathbf{y} - \mathbf{Q}^T\mathbf{Q}\mathbf{R}\mathbf{x}\|_2^2 + \|\mathbf{x}\|_0 \lambda(\Omega)$$

$$\operatorname*{argmin}_{\mathbf{x} \in \mathcal{A}_0} \|\tilde{\mathbf{y}} - \mathbf{R}\mathbf{x}\|_2^2 + \|\mathbf{x}\|_0 \lambda(\Omega). \tag{14}$$

Note that implementation via Sphere Decoding [8] is only valid if $\lambda(\Omega) \geq 0$ holds. If $\lambda(\Omega) < 0$, the fundamental assumption of an accumulating metric is violated [8]. Therefore, if and only if $\lambda(\Omega) < 0$ holds, we rewrite (14) as

$$\operatorname*{argmin}_{\mathbf{x} \in \mathcal{A}_0} \|\tilde{\mathbf{y}} - \mathbf{R}\mathbf{x}\|_2^2 - \|\mathbf{x}\|_0 |\lambda(\Omega)| + K|\lambda(\Omega)|$$

$$\operatorname*{argmin}_{\mathbf{x} \in \mathcal{A}_0} \|\tilde{\mathbf{y}} - \mathbf{R}\mathbf{x}\|_2^2 + |\lambda(\Omega)| \left[K - \|\mathbf{x}\|_0\right]$$

$$\operatorname*{argmin}_{\mathbf{x} \in \mathcal{A}_0} \|\tilde{\mathbf{y}} - \mathbf{R}\mathbf{x}\|_2^2 + |\lambda(\Omega)| \sum_{k=1}^{K} \mathbf{1}_0\left(x_k\right). \tag{15}$$

Note that with (14) and (15) implementation via Sphere Decoding is possible for $\lambda(\Omega) \in \mathbb{R}$ by selecting (14) if $\lambda(\Omega) \geq 0$ and (15) if $\lambda(\Omega) < 0$.

III. CDMA SYSTEM SETUP

In the following we apply the Bayes Risk Sphere Detector to a specific M2M system setup. We assume an uplink scenario where K sensor nodes are connected to a central aggregation node. More specifically, we assume a star topology. The sensor nodes access the medium via a Code Division Multiple Access (CDMA) scheme with independent Pseudo Noise (PN) sequences applied at the nodes. The application of CDMA is not a general restriction, but allows for adaptive and flexible support of different data rates, and flexible and scalable number of supported devices which is very attractive for M2M scenarios [9].

Furthermore, we assume the same sporadic communication scenario where nodes are only active on occasion to transmit their data to the central aggregation node. As CDMA is employed, the data symbols are spread by a node specific PN sequence of length N. After spreading, the symbols are transmitted over a frequency selective Rayleigh fading channel with impulse response length $L_{h,k}$ chips. The channel coefficients $h_{k,n}$, $1 \leq n \leq L_{h,k}$ are assumed to obey a zero mean uncorrelated Rayleigh Fading statistic with variance $1/L_{h,k}$. Without loss of generality, we assume $L_{h,k} = L_h \; \forall k$. CDMA allows for a certain degree of asynchronicity on the chip level between the nodes. However, to simplify the system description, we assume the nodes to transmit synchronously in time. These assumptions allow to formulate the CDMA specific spreading and convolution with the underlying frequency selective channel by the matrix $\mathbf{A} \in \mathbb{R}^{(N+L_h-1) \times K}$. This eases the formulation of a matched filter by simply filtering the received signal with $\mathbf{A}^T$. In this formulation we assume perfect Channel State Information (CSI) at the receiver. This could be obtained via training phases which are repeated periodically.

To ensure white uncorrelated Gaussian noise at the output of the matched filter, we apply the pre-whitening filter matrix [10] $\mathbf{P} \in \mathbb{R}^{K \times K}$ and obtain the canonical input-output relation at symbol-rate as

$$\mathbf{y} = \mathbf{PA}^T\mathbf{Ax} + \mathbf{PA}^T\mathbf{w}$$
$$= \mathbf{Tx} + \tilde{\mathbf{w}}. \qquad (16)$$

Here $\mathbf{y} \in \mathbb{R}^K$ is the received signal at symbol-rate. We assume that $L_h < N$ which allows to neglect inter-symbol interference in the symbol-rate model [11]. The system model with pre-whitening (16) and the underlying model for the Bayes Risk MAP from (1) are now equivalent.

IV. NUMERICAL EVALUATION

A. Classification of Error Events

Performing a joint data and activity detection necessitates the distinction between activity errors and data-symbol errors. To measure the performance of the joint activity and data detection, we define the following classes of symbol errors:

- Net Symbol error (NSE) rate
- False Active (FA) rate
- False Inactive (FI) rate
- Gross Symbol error (GSE) rate

We term the case where a node transmits a data symbol from $\mathcal{A}$ that is detected wrongly as a Net Symbol Error (NSE). Additionally, activity error events are classified by the False Active rate and the False Inactive rate. The wrong detection of a symbol from the augmented alphabet $\mathcal{A}_0$ is counted as a Gross Symbol Error (GSE). This class summarizes symbol and activity errors [3].

B. Performance of the Bayes-Risk Sphere-Detector

To illustrate the behavior of the Bayes-Risk Sphere-Detector for joint activity and data detection, we investigate the system setup summarized in Table II, exemplary.

Simulation Parameters	
Number of Nodes	$K = 10$
Spreading Gain	$N = 16$
Length of Channel Impulse Resp.	$L_h = 4$ chips
Channel Type	real valued Rayleigh Fading
Channel State Information	full
Activity Probability	$p_a = 0.2$
Modulation Type	BPSK
Pre-Whitening	Cholesky Decomposition
Detector	Bayes-Risk Sphere-Detector

TABLE II
SIMULATION PARAMETER

First, we investigate the performance and, especially, the impact of the weighting parameter Ω on the activity detection. Figure 1 shows the FA and the FI rates for three different values for Ω. Analyzing the curves for setting $\Omega = 1$ shows that the FA rate is much lower than the FI rate resulting in a conservative performance since the detector tends to estimate a node to be inactive even if it is active. This behavior becomes immediately clear by observing that in the low SNR region, the FI rate goes to one while producing a low FA rate. Even
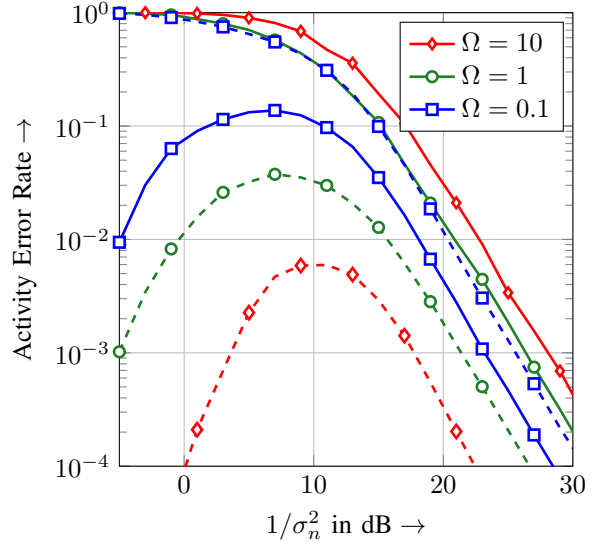

Fig. 1. Performance comparison of the activity detection parametrized by Ω. False Inactive rate (**Solid**), False Active rate (**Dashed**)

at higher SNR, the detector still tends to decide in favor of inactivity. This setting correspond to the MAP detector from [3]

Fig. 1 also shows the FA and FI rates exemplary for a more conservative ($\Omega = 0.1$) detector and for a liberal ($\Omega = 10$) detector. For $\Omega = 10$ the imbalance between FA and FI rate becomes even more severe compared to the preceding example. Consequently, the FA rate is much lower than the FI rate. In this setup, FA events occur rarely and the maximum is below 10^{-2}, contrary to FI events which occur often and converge to one in the low SNR range.

Setting $\Omega = 0.1$ leads to a more liberal detector and nodes are more likely detected to be active. In this setup, the FI rate is always lower than the FA rate. At low SNR, all nodes are estimated to be active and the FA rate tends to one whereas the FI rate is much lower.

As this example shows, choosing Ω too conservative will trim the detector such that many active nodes are estimated to be inactive. This setting will consequently lead to a loss of data. In the opposite, choosing Ω too liberal makes it complicated to decide if nodes are detected erroneously or if the detector is simply too sensitive to noise. A too liberal detector is therefore not very reliable and often detects signals that are only noise. Besides the performance of the activity detection, we are also investigating the performance of data detection. We start by considering the NSE which gives us an indication for the wrong estimation of data symbols that were transmitted by the active nodes. Fig. 3 shows the NSE for the BR-SD for different values of Ω. Starting again with the performance of the detector parametrized by $\Omega = 1$. We already noted that this setup results in a conservative performance. Consequently, especially at low SNR the detector will tend to estimate all nodes to be inactive and estimate the zero symbol. Accordingly, the NSE goes up to 1. Increasing Ω makes the detector even more conservative and the NSE is increased. In contrast, setting
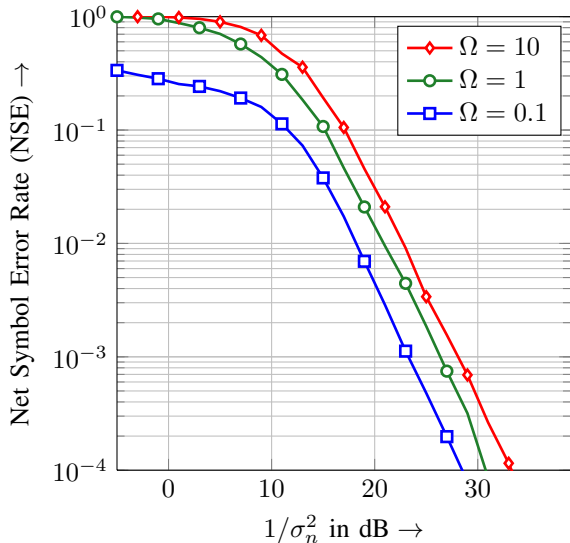
Fig. 2. Performance comparison of the Net Symbol Error rate parametrized by Ω.

$\Omega = 0.1$ nodes are more likely estimated to be active. At first glance, this strategy seems to be favorable since the NSE is much lower. However, one should note that with $\Omega = 0.1$ the FA rate is increasing as well as Fig. 1 shows.

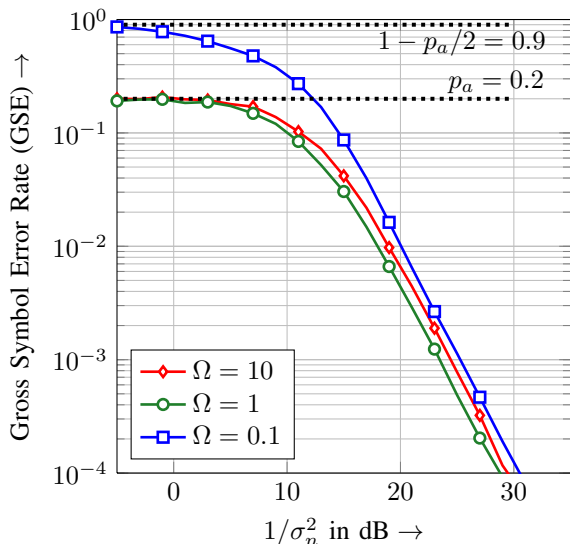The increase of the FA rate can be observed by investigating the Gross Symbol Errors (GSE) as shown in Fig. 3. The GSE

Fig. 3. Performance comparison of the Gross Symbol Error rate parametrized by Ω.

shows that choosing $\Omega = 1$ results in the lowest overall error rate. This is obvious as $\Omega = 1$ results in the MAP detector on the symbol alphabet $\mathcal{A}_0$. For low SNR, the detector estimates all nodes to be inactive and, makes p_a wrong decisions on average. For $\Omega = 10$, the detector is even more conservative and the number of nodes that are correctly estimated to be active is decreased. Additionally, the GSE is higher than for $\Omega = 1$. The liberal detector, finally, tends to detect nodes erroneously to be active. At very low SNR, all nodes are estimated to be active and $50\ \%$ of the symbols from the

active nodes are detected wrongly on average. This leads to the asymptotic GSE of $1 - p_a/2$ here and generally to a GSE of $1 - p_a/|\mathcal{A}|$.

V. CONCLUSION

In this work, we introduced a novel Bayes Risk Sphere-Detector that minimizes the risk of an erroneous decision with respect to the activity detection. As a degree of freedom the activity detection is controlled by an additional parameter. This provides a system depended flexible management of the False Active and False Inactive rate and can be adjusted by higher layers such that certain system specific error rates are met. We showed that this detector can easily be implemented as Bayes Risk Sphere-Detector. Additionally, we have proved the general concept exemplary on a CDMA uplink scenario. However, our results can be applied to any system where activity detection plays a fundamental role.

REFERENCES

[1] W. Lang, R. Jedermann, D. Mrugala, A. Jabbari, B. Krieg-Brueckner, and K. Schill, "The intelligent container a cognitive sensor network for transport management," *IEEE Sensors Journal*, vol. 11, no. 3, pp. 688 – 698, Mar. 2011.

[2] S. Lien, K. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," *Communications Magazine, IEEE*, vol. 49, no. 4, pp. 66 –74, Apr. 2011.

[3] H. Zhu and G. Giannakis, "Exploiting sparse user activity in multiuser detection," *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 454 – 465, february 2011.

[4] H. F. Schepker and A. Dekorsy, "Sparse Multi-User Detection for CDMA transmission using greedy algorithms," in *8th International Symposium on Wireless Communication Systems (ISWCS)*, Aachen, Germany, Nov., pp. 291 –295.

[5] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*, 2nd ed. Springer, jul. 2011.

[6] H. L. van Trees, *Detection, estimation and modulation theory*. Wiley, 1967.

[7] A. S. Willsky and H. L. Jones, "A generalized likelihood ratio approach to state estimation in linear systems subjects to abrupt changes," in *IEEE Conference on Decision and Control including the 13th Symposium on Adaptive Processes*, vol. 13, nov 1974, pp. 846 –853.

[8] M. O. Damen, H. E. Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2389 – 2402, oct. 2003.

[9] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, November 1998.

[10] F. Monsees, C. Bockelmann, D. Wübben, and A. Dekorsy, "Sparsity aware multiuser detection for machine to machine communication," in *Second International Workshop on Machine-to-Machine Communication at IEEE Globecom 2012*, Anaheim, USA, Dec 2012.

[11] V. Kühn, *Wireless Communications over Mimo Channels: Applications to CDMA And Multiple Antenna Systems*. John Wiley & Sons, 2006.