MESSAGE PASSING WITH ACTIVITY-AWARE DYNAMIC SCHEDULING FOR JOINT ACTIVITY DETECTION AND CHANNEL ESTIMATION FOR MACHINE-TYPE COMMUNICATIONS

Roberto B. Di Renna and Rodrigo C. de Lamare

CETUC, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil
{robertobrauer, delamare}@cetuc.puc-rio.br

ABSTRACT

In this work, we study the problem of joint active device detection and channel estimation in an uplink grant-free massive MIMO system scenario. Based on the hybrid generalized approximate message passing (HyGAMP) algorithm, we propose the message-scheduling GAMP (MSGAMP), where we apply different sequential scheduling techniques in which messages are generated using the latest available information. We present three scheduling techniques that update the messages based on the activity detection performed at each new iteration. Simulations show that MSGAMP-type schemes exhibits good performance in terms of activity error rate and normalized mean squared error (NMSE), while outperforming HyGAMP and requiring a small number of iterations for convergence.

***Index Terms*—** Compressed sensing, message-scheduling, joint activity detection and channel estimation, mMTC.

1. INTRODUCTION

Massive machine-type communications (mMTC) is one of the key application scenarios of fifth generation (5G) and beyond cellular networks. Since the total number of machine-type devices (MTDs) is much larger than the receive processing resources, conventional scheduling-based orthogonal multiple access schemes is not suitable. Due to the high probability of frame collisions, the scheme in which the base station (BS) allocates orthogonal time/frequency resources to each device is impractical for the mMTC scenario. Moreover, specific features of the mMTC scenario as low data rates and sporadic transmission of short packets [1], are compromised due to signalling overhead and excessive latency. The uncoordinated access proposed in recent years can solve those issues. Based on a grant-free *non-orthogonal multiple access* (NOMA) [2], active devices transmit frames without previous scheduling, in order to eliminate the need for round-trip signaling. Thus, it is up to the BS to estimate the channels, detect active devices and transmitted signals.

Due to the intermittent activity pattern of the mMTC devices, the transmitted vector and the channel matrix can be modeled as a sparse vector. In this way, the multi-user detection (MUD) problem can be seen as a sparse signal recovery problem. Several approaches have been proposed to formulate joint user activity and data detection as the works in [3, 4, 5, 6]. In most of these studies, the uplink channel state information (CSI) from the MTD to the BS is assumed to be perfectly known to the BS, which allows interference mitigation techniques [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]. However, in practice, the uplink CSI should be estimated before data detection. In order to address this issue, various joint active user detection (AUD) and channel estimation (CE) schemes have been proposed.

Exploiting the *a priori* distribution of the sparse vector to be recovered, the works in [20, 21, 22] presents denoising-based AMP algorithms verifying the activity error rate performance, while the work in [23] uses the expectation propagation (EP) and computes iteratively the moment matching. As an extension of the GAMP [24] algorithm, the inference algorithm HyGAMP [25] is then developed. Since the components of the channel are particularly independent with conditional distribution, the combination of a loopy belief propagation (LBP) part for user activity detection and a GAMP-type strategy for channel estimation makes HyGAMP outperform other existing algorithms in terms of mean square error (MSE). However, HyGAMP considers a completely parallel update of the messages, where each iteration performs exactly one update of all edges.

Commonly applied in LDPC decoders, the sequential scheduling, in which messages are generated using the latest available information, significantly improve the efficiency in terms of convergence and error rates [26]. The main idea is to find the best sequence of message updates, focusing on the part of the graph that has not converged. Thus, in order to find a more efficient implementation or/with better convergence solution, in this work, we propose three message scheduling strategies of a HyGAMP based algorithm (MSGAMP). The proposed MSGAMP algorithm and strategies exploit the *a priori* distribution of the sparse channel matrix and use the number of antennas in the base station to enhance the activity detection. Simulations show that MSGAMP results in an improved performance over HyGAMP in terms of NMSE while requiring a small number of iterations for convergence and a lower computational cost than HyGAMP.

2. SYSTEM MODEL

In this section, we detail the uplink system model. We assume there are N single-antenna devices communicating with a BS equipped with M antennas. The problem of interest here is to estimate the channel matrix $\mathbf{H} \in \mathbb{C}^{N\times M}$ from a received signal $\mathbf{Y} \in \mathbb{C}^{L\times M}$ obtained through the following model

$$\mathbf{Y} = \sum_{n=1}^{N} \frac{\mathbf{a}_n}{\|\mathbf{a}_n\|}\mathbf{h}_n^T + \mathbf{W} = \mathbf{\Phi}\,\mathbf{H} + \mathbf{W}, \tag{1}$$

where $\mathbf{W} \in \mathbb{C}^{L\times M}$ is the independent complex-Gaussian noise matrix with $\mathcal{CN}\left(0, \sigma_w^2\right)$. The pilot matrix $\mathbf{\Phi} \in \mathbb{C}^{L\times N}$ is composed by independent pilot sequences $\mathbf{a}_n = \exp\left(j\pi\boldsymbol{\kappa}\right)$ of each device, where each element of vector $\boldsymbol{\kappa} \in \mathbb{R}^L$ is drawn uniformly at random in

This study has been funded in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

$[-1,1]$ and L is the length of the pilot sequence. Each active device transmits L pilot symbols, which we denote here as a frame. Since the frame size of each device is typically very small, we assume that all devices are synchronized in time.

Each element of $\mathbf{H}$ represents the channel gain between the n-th device and the m-th BS antenna. Since mMTC is a sparse scenario, we denote the Boolean variable $\xi_n \in \{0,1\}$ that indicates if the device is active when $\xi_n = 1$ and inactive otherwise. Thus, considering as ρ_n the probability of being active of the n-th device,

$$P(\xi_n = 1) = 1 - P(\xi_n = 0) = \rho_n, \tag{2}$$

where all ξ_n are considered i.i.d. and each device has its own probability of being active. Thus, given the vector $\boldsymbol{\xi}$, the components of $\mathbf{H}$ are independent with the conditional densities

$$h_{nm}|\boldsymbol{\xi} \sim \begin{cases} \delta(h_{nm}), & \xi_n = 0, \\ \mathcal{CN}(h_{nm}|0,\beta_n), & \xi_n = 1, \end{cases} \tag{3}$$

where $\delta(\cdot)$ is Dirac delta function. As in the mMTC scenario N is larger than M, the system is overloaded. However, due to the low activity probability of devices, $\mathbf{H}$ is sparse, which makes its recovery possible through the theory of compressed sensing (CS) [27]. Then, we propose MSGAMP for joint activity detection and channel estimation.

3. JOINT CHANNEL AND USER ACTIVITY ESTIMATION

In order to present the message updating rules of the MSGAMP algorithm, first we introduce some statistics of the system model. Setting $\mathbf{Z} = \boldsymbol{\Phi}\,\mathbf{H}$ and given the independence of the devices, the prior distribution of $\mathbf{h}_m$ and the likelihood are given by $p(\mathbf{h}_m|\boldsymbol{\xi}) = \prod_{n=1}^{N} p(h_{nm}|\xi_n)$ and $p(\mathbf{Y}|\mathbf{Z}) = \prod_{l=1}^{L}\prod_{m=1}^{M} p(y_{lm}|z_{lm})$, respectively. Since the main objective of HyGAMP is to approximate the marginal posterior density by a product of the prior and a Gaussian distribution, the minimum mean squared error (MMSE) estimate of h_{nm}, $\hat{h}_{nm} = \mathbb{E}_{h_{nm}|\mathbf{y}}[h_{nm}]\,\forall\, n,m$, is given by

$$p(h_{nm}|\mathbf{Y}) = \int p(\mathbf{H},\boldsymbol{\xi}|\mathbf{Y})\,\mathrm{d}\boldsymbol{\xi}\,\mathrm{d}\mathbf{H}_{\backslash nm} \tag{4}$$

where $\mathbf{H}_{\backslash nm}$ denotes all elements except h_{nm} and $p(\mathbf{H},\boldsymbol{\xi}|\mathbf{Y})$ denotes the posterior distribution, given by the Bayes' rules

$$\begin{aligned} p(\mathbf{H},\boldsymbol{\xi}|\mathbf{Y}) =& \frac{1}{p(\mathbf{Y})} p(\mathbf{Y}|\mathbf{H},\boldsymbol{\xi})\, p(\mathbf{H}|\boldsymbol{\xi})\, p(\boldsymbol{\xi}) \\ =& \frac{1}{p(\mathbf{Y})}\left[\prod_{l=1}^{L}\prod_{m=1}^{M} p\left(y_{lm}|\sum_{n=1}^{N}\phi_{ln}\,h_{nm}\right)\right] \\ &\times \left[\prod_{n=1}^{N}\prod_{m=1}^{M} P(h_{nm}|\xi_n)\right]\left[\prod_{n=1}^{N} P(\xi_n)\right], \end{aligned} \tag{5}$$

where $P(h_{nm}|\xi_n)$ is the conditional density for the random variable in (3).

3.1. Factor graph approach

As HyGAMP can be seen as an approximate employment of the sum-product loopy belief propagation (BP), we first write the messages to marginalize the problem. Fig. 1 shows a factor graph (FG) that represents the factorization of (5), where variable nodes ξ_n and h_{nm} are depicted as spheres and the factor nodes are represented as cubes. Factor nodes are the density functions, prior and likelihood. Factor nodes are the density functions, prior and likelihood. The approach that will be presented in what follows is inspired by work on low-density parity-check (LDPC) codes [28, 29, 30] and message passing algorithms [26, 31, 15, 32, 33].

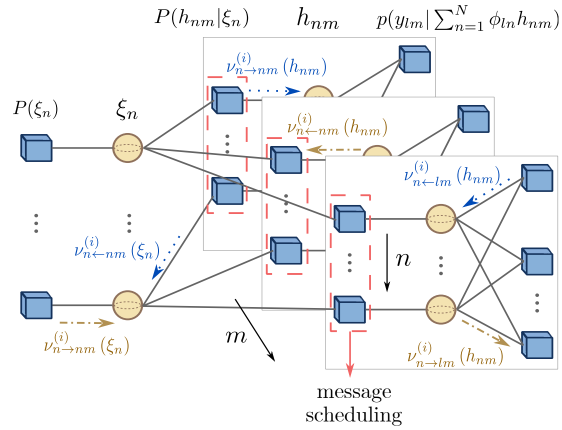


Fig. 1. The factor graph of joint distribution $p(\mathbf{H},\mathbf{Y},\boldsymbol{\xi})$ where cubes denote factor nodes and spheres variable nodes.

As the sensing matrix in our problem ($\boldsymbol{\Phi}$) is a dense matrix, the FG in Fig. 1 is fully connected. Computing the messages in fully connected graphs is tricky as the messages themselves are functions. The messages are approximated by prototype functions that resemble Gaussian density functions which can be described by two parameters only. Thus, message passing reduces to the exchange of the parameters of a function instead of the function itself. Therefore, it is possible to iteratively approximate, for a FG with cycles as in Fig. 1, the marginal posteriors passing messages between different nodes. So, we can define the messages from $p(y_{lm}|\cdot)$ to h_{nm} and to the opposite direction as

$$\begin{aligned} \nu^{(i)}_{n\leftarrow lm}(h_{nm}) \propto& \int p\left(y_{lm}|\sum_{k=1}^{N}\phi_{lk}\,h_{km}\right) \\ &\times \prod_{j\neq n}^{N} \nu^{(i)}_{j\to lm}(h_{jm})\,\mathrm{d}h_{jm} \end{aligned} \tag{6}$$

$$\nu^{(i+1)}_{n\to lm}(h_{nm}) \propto \nu^{(i)}_{n\to nm}(h_{nm}) \prod_{k\neq m}^{M} \nu^{(i)}_{n\leftarrow lk}(x_{nk}) \tag{7}$$

and, considering $\propto$ as proportional, the messages from $P(h_{nm}|\xi_n)$ to h_{nm} and to the opposite direction are given by

$$\nu^{(i)}_{n\leftarrow nm}(h_{nm}) \propto \prod_{k=1}^{M} \nu^{(i)}_{n\to lk}(h_{nm}), \tag{8}$$

$$\nu^{(i+1)}_{n\to nm}(h_{nm}) \propto \int p(h_{nm}|\xi_n)\,\nu^{(i)}_{n\to nm}(\xi_n)\,\mathrm{d}\xi_n. \tag{9}$$

Specifically, each iteration of the Algorithm 1 has three stages. The first stage, labelled as "GAMP approximation" contains the updates of the GAMP based on expectation propagation algorithm, which treats the components h_n as independent with the probability of being active $\hat{\rho}_n$. Drawing inspiration in [23, 34, 35], we incorporated the EP in the process of LBP to relaxed belief propagation and then to GAMP. At iteration i, the MSGAMP algorithm produces estimates $\hat{\mathbf{h}}^{(i)}$ and $\hat{\mathbf{z}}^{(i)}$ of the vectors $\mathbf{h}$ and $\mathbf{z}$. Several other intermediate vectors, $\hat{\mathbf{p}}^{(i)}$, $\hat{\mathbf{r}}^{(i)}$ and $\hat{\mathbf{s}}^{(i)}$, are also produced. Associated of each of these vectors are matrices like $\mathbf{Q}^{h(i)}$ and $\mathbf{Q}^{z(i)}$ that represent covariances. In a nutshell, in order to reduce the complexity of $O(LNM)$ to $O(NM)$, the message in (6) is firstly mapped to Gaussian distribution based on the central limit theorem and Taylor expansions. So, $\nu_{n\leftarrow lm}^{(i)}(h_{nm})$ is updated by the Gaussian reproduction property. Following the same procedure in the messages of (7), (8) and (9), relaxed BP is obtained by the combination of the approximated messages. Since many of these messages only differ slightly from each other, in order to fill up those differences, new variables are produced and ignoring the infinitesimals, the GAMP based on EP is obtained.

The second stage of Algorithm 1, labelled as "sparsity-rate update", refers to the LBP part of the FG in Fig. 1 and updates the estimates of each probability of being active $\hat{\rho}_{nm}$. Computed using Gaussian approximations of likelihood functions, these estimates are then used to define the message scheduling proposed in this work. The messages in the "sparsity-rate update" stage are given by

$$\nu_{n\leftarrow nm}^{(i)}(\xi_n) \propto \int p(h|\xi_n)\, \nu_{n\leftarrow nm}^{(i)}(h)\, \mathrm{d}h, \tag{10}$$

$$\nu_{n\rightarrow nm}^{(i)}(\xi_n) \propto P(\xi_n) \prod_{k\neq m}^{M} \nu_{n\leftarrow nk}^{(i)}(\xi_n), \tag{11}$$

where (10) refers to the message from $P(h_{nm}|\xi_n)$ to ξ_n while (11) denotes the message in opposite direction and each belief at ξ_n is given by $\nu_n^{(i)}(\xi_n) \propto P(\xi_n)\prod_{m=1}^{M}\nu_{n\leftarrow nm}^{i}(\xi_n)$.

The message in (10) can be approximated as a likelihood function given by $\nu_{n\leftarrow nm}^{(i)}(\xi_n) = \mathcal{CN}\left(h_{nm}|\hat{r}_{nm}^{(i)}, Q_{nm}^{r(i)}\right)$. Applying the Gaussian reproduction property [35] enables us to define

$$\begin{aligned}\mathrm{LLR}_{n\leftarrow nm}^{(i)} &= \log \frac{\nu_{n\leftarrow nm}^{(i)}(\xi_n = 1)}{\nu_{n\leftarrow nm}^{(i)}(\xi_n = 0)} \\ &= \log \frac{\mathcal{CN}\left(0\big|\hat{r}_{nm}^{(i)}, Q_{nm}^{r(i)} + \beta_n\right)}{\mathcal{CN}\left(0\big|\hat{r}_{nm}^{(i)}, Q_{nm}^{r(i)}\right)}.\end{aligned} \tag{12}$$

Similarly to (12), we have $\mathrm{LLR}_n \triangleq \log \frac{\nu_n^{(i)}(\xi_n=1)}{\nu_n^{(i)}(\xi_n=0)}$ and $\mathrm{LLR}_{n\rightarrow nm}^{(i)} \triangleq \log \frac{\nu_{n\rightarrow nm}^{(i)}(\xi_n=1)}{\nu_{n\rightarrow nm}^{(i)}(\xi_n=0)}$. Substituting (12) in (11) and in each belief, $\mathrm{LLR}_{n\rightarrow nm}^{(i)}$ is given by

$$\mathrm{LLR}_{n\rightarrow nm}^{(i)} = \log\left(\frac{\rho_n}{1-\rho_n}\right) + \sum_{d\neq m}^{M} \mathrm{LLR}_{n\leftarrow nd}^{(i)}. \tag{13}$$

Thereby, the message in (9) is described by

$$\begin{aligned}\nu_{n\rightarrow lm}^{(i+1)}(h_{nm}) &= P(h_{nm}|\xi_n = 1)\, \nu_{n\rightarrow nm}^{(i)}(\xi_n = 1) \\ &\quad + P(h_{nm}|\xi_n = 0)\, \nu_{n\rightarrow nm}^{(i)}(\xi_n = 0) \\ &= \hat{\rho}_{nm}^{(i)}\, \mathcal{CN}(h_{nm}|0, \beta_n) + \left(1 - \hat{\rho}_{nm}^{(i)}\right)\delta(h_{nm}),\end{aligned} \tag{14}$$

where

$$\hat{\rho}_{nm}^{(i)} \triangleq \nu_{n\rightarrow lm}^{(i+1)}(\xi_n = 1) = 1 - \frac{1}{1 + \exp\left(\mathrm{LLR}_{n\leftarrow nm}^{(i)}\right)}. \tag{15}$$

With HyGAMP established for mMTC, the next step is to use the estimates obtained in (15) in the message-scheduling techniques.

3.2. Message-scheduling schemes

We propose three ordering schemes for designing the message scheduling of MSGAMP. Given the device activity detection in (15), the idea of the schemes is to only update the messages of the active devices. Since mMTC is in a massive MIMO scenario, it is possible to refine the activity detection using the estimates of each BS antenna, as $\hat{\rho}_n^{(i)} = \sum_{m=1}^{M}\hat{\rho}_{nm}^{(i)}/M$. If $\hat{\rho}_n^{(i)}$ is higher than a threshold, the device is considered active and included in the set $\mathbf{S}^{(i)}$. Thus, MSGAMP proceeds until i reaches the maximum number of iterations I or $\left(\mathrm{tol}/M < 10^{-4}\right)$, where tol is given by

$$\mathrm{tol} = \sum_{m=1}^{M} \frac{\|\hat{\underline{\mathbf{h}}}_m^{(i)} - \hat{\underline{\mathbf{h}}}_m^{(i-1)}\|}{\|\hat{\underline{\mathbf{h}}}_m^{(i)}\|}, \tag{16}$$

where $\hat{\underline{\mathbf{h}}}_m^{(i)}$ is a $|\mathbf{S}^{(i)}| \times 1$ vector that corresponds to the estimated channel gains between the $|\mathbf{S}^{(i)}|$ devices and the m-th BS antenna.

The goal of this stopping criterion and message updating schedule is to focus on the messages that belongs to the activity devices, thus reducing computational complexity and possibly reducing the number of iterations to reach convergence. A great advantage of using these message scheduling schemes is that, unlike the well-known parallel message update of HyGAMP that, in each iteration, $O(MN)$ messages must be computed, in our schemes we need only $O(M|\mathbf{S}^{(i)}|)$. Taken into account that it is expected up to $300,000$ devices per cell [36] in future mobile communication systems and the sporadic transmission pattern of each device, the gain in scheduling schemes is evident, since $|\mathbf{S}^{(i)}| << N$. With the activity detec-

Algorithm 1 MSGAMP-AUD for mMTC

initialize
$i = 1$, $\hat{s}_{lm}^{(0)} = 0$, $\hat{r}_{nm}^{(0)} = 0$, $Q_{nm}^{r(0)} = 1$, $\hat{\rho}_{nm}^{(0)} = \rho_n$, $\mathbf{S}^{(0)} = [1, \ldots, N]$
repeat
% GAMP approximation
for $\left(n = 1, \ldots, |\mathbf{S}^{(i-1)}|\right) \forall n \in \mathbf{S}^{(i-1)}$
  for $(m = 1, \ldots, M)$
    $\hat{h}_{nm}^{(i)} = \mathbb{E}\left[\mathcal{X}_{nm}|\hat{r}_{nm}^{(i-1)}, Q_{nm}^{r(i-1)}; \hat{\rho}_n^{(i-1)}\right]$
    $Q_{nm}^{h(i)} = \mathrm{Var}\left[\mathcal{X}_{nm}|\hat{r}_{nm}^{(i-1)}, Q_{nm}^{r(i-1)}; \hat{\rho}_n^{(i-1)}\right]$
    for $(l = 1, \ldots L)$
      $Q_{tm}^{p(i)} = \sum_{n=1}^{N}|\Phi_{tn}|^2 Q_{nm}^{h(i)}$
      $p_{tm}^{(i)} = \sum_{n=1}^{N}\Phi_{tn}\,\hat{h}_{nm}^{(i)} - Q_{tm}^{p(i)}\,\hat{s}_{tm}^{(i-1)}$
      $\tilde{z}_{tm}^{(i)} = \left(y_{lm}\,Q_{tm}^{p(i)} + \sigma_w^2 p_{tm}^{(i)}\right) / \left(Q_{tm}^{p(i)} + \sigma_w^2\right)$
      $Q_{tm}^{z(i)} = \left(\sigma_w^2 Q_{tm}^{p(i)}\right) / \left(Q_{tm}^{p(i)} + \sigma_w^2\right)$
      $\hat{s}_{tm}^{(i)} = \left(\tilde{z}_{tm}^{(i)} - p_{tm}^{(i)}\right) / Q_{tm}^{p(i)}$
      $Q_{tm}^{s(i)} = Q_{tm}^{-p(i)}\left(1 - Q_{tm}^{z(i)}/Q_{tm}^{p(i)}\right)$
    end for
    $Q_{nm}^{-r(i)} = \sum_{l=1}^{L}|\Phi_{tn}|^2 Q_{tm}^{s(i)}$
    $r_{nm}^{(i)} = \hat{h}_{nm}^{(i)} + Q_{nm}^{r(i)}\sum_{l=1}^{L}\Phi_{tn}^{*}\hat{s}_{tm}^{(i)}$
% Sparsity-rate update
    $\mathrm{LLR}_{n\leftarrow nm}^{(i)} = \log \frac{\mathcal{CN}\left(0|\hat{r}_{nm}^{(i)}, Q_{nm}^{r(i)} + \beta_n\right)}{\mathcal{CN}\left(0|\hat{r}_{nm}^{(i)}, Q_{nm}^{r(i)}\right)}$
    $\mathrm{LLR}_{n\rightarrow nm}^{(i)} = \log\left(\rho_n/1-\rho_n\right) + \sum_{d\neq m}^{M}\mathrm{LLR}_{n\leftarrow nd}^{(i)}$
    $\hat{\rho}_{nm}^{(i)} = 1 - 1/1+\exp\left(\mathrm{LLR}_{n\rightarrow nm}^{(i)}\right)$
  end for
end for
% Message scheduling update
$\hat{\rho}_n^{(i)} = \sum_{m=1}^{M}\hat{\rho}_{nm}^{(i)}/M$
$\mathbf{S}^{(i)} = \mathrm{update}\left[\mathbf{S}^{(i-1)}\right]$ *% chosen AUD message scheduling type*
$\mathrm{tol} = (1/M)\sum_{m=1}^{M}\|\hat{\underline{\mathbf{h}}}_m^{(i)} - \hat{\underline{\mathbf{h}}}_m^{(i-1)}\| / \|\hat{\underline{\mathbf{h}}}_m^{(i)}\|$
$i = i + 1$
until $\left(i > I \text{ or } \mathrm{tol} < 10^{-4}\right)$

Algorithm 2: Message scheduling based on AUD

initialize: $p = 1$

if $\left(\hat{\rho}_n^{(i)} >= \text{threshold}\right)$, $\forall n \in N$:

  $s_p^{(i)} = n;\ p = p + 1;$

else $\forall m \in M$:

  $\hat{h}_{nm}^{(i+1)} = \hat{h}_{nm}^{(i)};\ Q_{nm}^{h(i+1)} = Q_{nm}^{h(i)};$

  $\hat{r}_{nm}^{(i+1)} = \hat{r}_{nm}^{(i)};\ Q_{nm}^{r(i+1)} = Q_{nm}^{r(i)};$

end

tion procedure and stopping criterion defined, we present the first message scheduling scheme.

3.2.1. MSGAMP-AUD

The Message scheduling based on activity user detection (MSGAMP-AUD) is a dynamic scheduling strategy that updates all messages of the devices detected as active in parallel, repeating the values of the messages of other devices. As described in Algorithm 2, at each new iteration, the set $\mathbf{S}^{(i)}$ is updated based on the new values of $\hat{\rho}_n$, thus modifying the messages to be updated in parallel.

3.2.2. MSGAMP-GAUD

In this strategy, we have a message scheduling inside the group of messages that belongs to devices detected as active. If in MSGAMP-AUD we update all the messages of the group in parallel, in the Message Scheduling GAMP based on group activity user detection (MSGAMP-GAUD) we do not update the messages that belong to the first device of the active group.

When $i = 2$, we have the first values of $\hat{\boldsymbol{\rho}}$, thus enabling update of the set $\mathbf{S}^{(i)}$. In the next iteration, all messages that belong to $\mathbf{S}^{(i)}$, except for $s_1^{(i)}$ will be updated. Then, the index that refers to the group of messages that had been updated is removed of $\mathbf{S}^{(i)}$ as in

$$\mathbf{S}^{(i)} = \left[s_2^{(i-1)}, \ldots, s_{|\mathbf{S}^{(i-1)}|}^{(i-1)}\right]. \tag{17}$$

Therefore, we exclude a group of messages that belong to a specific device to be updated, one at a time. When $\mathbf{S}^{(i)}$ is empty, a new update of the set $\mathbf{S}^{(i)}$ is performed. In a nutshell, we reduce the set $\mathbf{S}^{(i)}$ that is updated in parallel, until there is no message to update.

3.2.3. MSGAMP-GAUDp

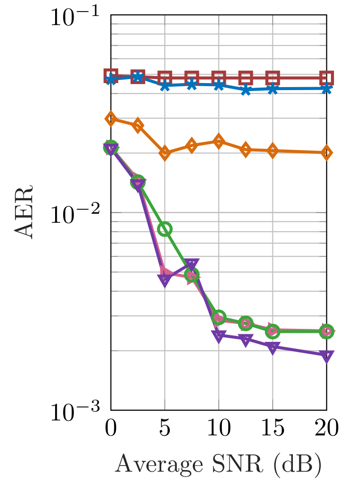
Since the last message scheduling techniques do not update the group of messages outside the set $\mathbf{S}^{(i)}$, they could suffer if the activity detection is not properly done. In order to address such possible errors, when a new set $\mathbf{S}^{(i)}$ is computed, the Message Scheduling GAMP based on group activity user detection with parallel update (MSGAMP-GAUDp) modifies the messages that do not belong to the set, as given by

$$\begin{cases} \mathbf{S}^{(i)} = [1, \ldots, N], & |\mathbf{S}^{(i-1)}| = 0, \\ \textit{proceed as in MSGAMP-GAUD}, & \text{otherwise.} \end{cases} \tag{18}$$

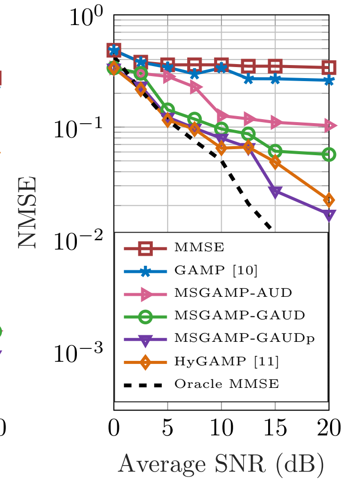
After the first update of the set $\mathbf{S}^{(i)}$ is made, MSGAMP-GAUDp proceeds as MSGAMP-GAUD. When $\mathbf{S}^{(i)}$ is empty, all messages are updated and a new activity detection is performed, as a new $\mathbf{S}^{(i)}$.

4. SIMULATION RESULTS

We simulate an mMTC system with $N = 128, M = 32$ and $L = 64$, designed as described in Section 2. The average SNR is set to $1/\sigma_w^2$, while the activity probabilities p_n are drawn uniformly at random in $[0.01, 0.05]$. The threshold chosen to detect the activity of devices is 0.9. The variations of MSGAMP are compared to the well-known minimum mean squared error (MMSE), generalized approximate message passing (GAMP) [24] and the state-of-the-art algorithm, HyGAMP [25]. An MMSE detector with perfect activity knowledge (oracle MMSE) is used as a lower bound. Figs. 2a and 2b show results of AER and NMSE. One can see that the use of the BS antennas in order to refine the activity detection improved the AER performance since the proposed MSGAMP-type techniques outperform HyGAMP. Regarding NMSE, Fig. 2b shows that the message scheduling schemes have a competitive performance with MSGAMP-GAUDp outperforming HyGAMP. Fig. 3 depicts the convergence rate of MSGAMP-type techniques and HyGAMP. One can notice that for different values of SNR, MSGAMP-type techniques converge faster and to lower values of NMSE than HyGAMP.

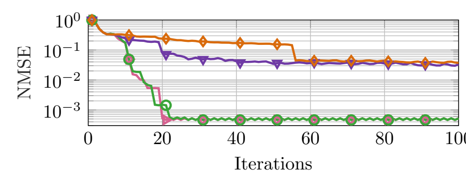


(a)

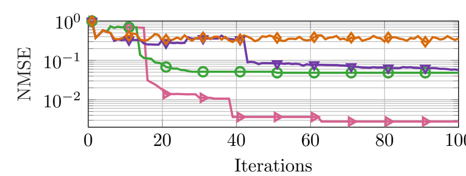


(b)

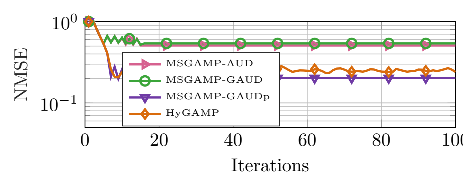
Fig. 2. Performance in terms of AER and NMSE versus average SNR. The NMSE considered only the active devices.



(a) Average SNR = 5 dB



(b) Average SNR = 7.5 dB.



(c) Average SNR = 10 dB.

Fig. 3. Convergence rate of MSGAMP-type techniques and HyGAMP. The NMSE considered all channel gains, not only the ones of active devices.

5. CONCLUSION

In this paper we have presented MSGAMP-type techniques that perform joint activity detection and channel estimation for mMTC. Exploiting the BS antennas, we developed three scheduling techniques for MSGAMP that update the messages based on the activity detection. The results indicate that MSGAMP-type techniques outperform other solutions in terms of NMSE and AER, while requiring a small number of iterations for convergence.

6. REFERENCES

[1] R. B. Di Renna, C. Bockelmann, R. C. De Lamare, and A. Dekorsy, "Detection techniques for massive machine-type communications: Challenges and solutions," *IEEE Access*, pp. 1–1, 2020.

[2] M. Shirvanimoghaddam *et al.*, "Massive non-orthogonal multiple access for cellular iot: Potentials and limitations," *IEEE Commun. Magazine*, vol. 55, no. 9, pp. 55–61, 2017.

[3] H. Zhu and G. B. Giannakis, "Exploiting Sparse User Activity in Multiuser Detection," *IEEE Transactions on Communications*, vol. 59, no. 2, pp. 454–465, 2011.

[4] B. Knoop and F. Monsees and C. Bockelmann and D. Wuebben and S. Paul and A. Dekorsy, "Sparsity-Aware Successive Interference Cancellation with Practical Constraints," in *WSA 2013; 17th International ITG Workshop on Smart Antennas*, Stuttgart, Germany, 2013, pp. 1–8.

[5] R. B. Di Renna and R. C. de Lamare, "Adaptive Activity-Aware Iterative Detection for Massive Machine-Type Communications," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1631–1634, 2019.

[6] R. B. Di Renna and R. C. de Lamare, "Iterative List Detection and Decoding for Massive Machine-Type Communications," *IEEE Transactions on Communications*, vol. -, no. -, pp. –, 2020.

[7] Rodrigo C. De Lamare, Raimundo Sampaio-Neto, and Are Hjorungnes, "Joint iterative interference cancellation and parameter estimation for cdma systems," *IEEE Communications Letters*, vol. 11, no. 12, pp. 916–918, 2007.

[8] R. C. De Lamare and R. Sampaio-Neto, "Minimum mean-squared error iterative successive parallel arbitrated decision feedback detectors for ds-cdma systems," *IEEE Transactions on Communications*, vol. 56, no. 5, pp. 778–789, 2008.

[9] Rodrigo C. de Lamare and Raimundo Sampaio-Neto, "Adaptive reduced-rank processing based on joint and iterative interpolation, decimation, and filtering," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2503–2514, 2009.

[10] Yunlong Cai, Rodrigo C. de Lamare, Benoit Champagne, Boya Qin, and Minjian Zhao, "Adaptive reduced-rank receive processing based on minimum symbol-error-rate criterion for large-scale multiple-antenna systems," *IEEE Transactions on Communications*, vol. 63, no. 11, pp. 4185–4201, 2015.

[11] Peng Li, Rodrigo C. de Lamare, and Rui Fa, "Multiple feedback successive interference cancellation detection for multiuser mimo systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2434–2439, 2011.

[12] Rodrigo C. de Lamare, "Adaptive and iterative multi-branch mmse decision feedback detection algorithms for multi-antenna systems," *IEEE Transactions on Wireless Communications*, vol. 12, no. 10, pp. 5294–5308, 2013.

[13] Patrick Clarke and Rodrigo C. de Lamare, "Transmit diversity and relay selection algorithms for multirelay cooperative mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 3, pp. 1084–1098, 2012.

[14] Andre G. D. Uchoa, Cornelius T. Healy, and Rodrigo C. de Lamare, "Iterative detection and decoding algorithms for mimo systems in block-fading channels using ldpc codes," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2735–2741, 2016.

[15] Roberto B. Di Renna and Rodrigo C. de Lamare, "Adaptive activity-aware iterative detection for massive machine-type communications," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1631–1634, 2019.

[16] Roberto B. Di Renna and Rodrigo C. de Lamare, "Iterative list detection and decoding for massive machine-type communications," *IEEE Transactions on Communications*, vol. 68, no. 10, pp. 6276–6288, 2020.

[17] Zhichao Shao, Rodrigo C. de Lamare, and Lukas T. N. Landau, "Iterative detection and decoding for large-scale multiple-antenna systems with 1-bit adcs," *IEEE Wireless Communications Letters*, vol. 7, no. 3, pp. 476–479, 2018.

[18] Roberto B. Di Renna, Carsten Bockelmann, Rodrigo C. de Lamare, and Armin Dekorsy, "Detection techniques for massive machine-type communications: Challenges and solutions," *IEEE Access*, vol. 8, pp. 180928–180954, 2020.

[19] Zhichao Shao, Lukas T. N. Landau, and Rodrigo C. de Lamare, "Dynamic oversampling for 1-bit adcs in large-scale multiple-antenna systems," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3423–3435, 2021.

[20] L. Liu and W. Yu, "Massive Connectivity With Massive MIMO—Part I: Device Activity Detection and Channel Estimation," *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 2933–2946, 2018.

[21] Z. Chen and F. Sohrabi and W. Yu, "Multi-Cell Sparse Activity Detection for Massive Random Access: Massive MIMO Versus Cooperative MIMO," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4060–4074, 2019.

[22] K. Senel and E. G. Larsson, "Grant-Free Massive MTC-Enabled Massive MIMO: A Compressive Sensing Approach," *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 6164–6175, 2018.

[23] J. Ahn and B. Shim and K. B. Lee, "EP-Based Joint Active User Detection and Channel Estimation for Massive Machine-Type Communications," *IEEE Transactions on Communications*, vol. 67, no. 7, pp. 5178–5189, 2019.

[24] Sundeep Rangan, "Generalized Approximate Message Passing for Estimation with Random Linear Mixing," *Retrieved August 13, 2012, from the arXiv database.*, 2012.

[25] S. Rangan, A. K. Fletcher, V. K. Goyal, E. Byrne, and P. Schniter, "Hybrid approximate message passing," *IEEE Transactions on Signal Processing*, vol. 65, no. 17, pp. 4577–4592, 2017.

[26] A. I. V. Casado, M. Griot and R. D. Wesel, "LDPC Decoders with Informed Dynamic Scheduling," *IEEE Transactions on Communications*, vol. 58, no. 12, pp. 3470–3479, 2010.

[27] J. W. Choi *et al.*, "Compressed sensing for wireless communications: Useful tips and tricks," *IEEE Commun. Surveys & Tut.*, vol. 19, no. 3, pp. 1527–1550, 2017.

[28] Andre G. D. Uchoa, Cornelius Healy, Rodrigo C. de Lamare, and Richard D. Souza, "Design of ldpc codes based on progressive edge growth techniques for block fading channels," *IEEE Communications Letters*, vol. 15, no. 11, pp. 1221–1223, 2011.

[29] C. T. Healy and R. C. de Lamare, "Decoder-optimised progressive edge growth algorithms for the design of ldpc codes with low error floors," *IEEE Communications Letters*, vol. 16, no. 6, pp. 889–892, 2012.

[30] Cornelius T. Healy and Rodrigo C. de Lamare, "Design of ldpc codes based on multipath emd strategies for progressive edge growth," *IEEE Transactions on Communications*, vol. 64, no. 8, pp. 3208–3219, 2016.

[31] Jingjing Liu and Rodrigo C. de Lamare, "Low-latency reweighted belief propagation decoding for ldpc codes," *IEEE Communications Letters*, vol. 16, no. 10, pp. 1660–1663, 2012.

[32] Cornelius Healy, "Knowledge-aided informed dynamic scheduling for ldpc decoding of short blocks," *IET Communications*, vol. 12, pp. 1094–1101(7), June 2018.

[33] Roberto B. Di Renna and Rodrigo C. de Lamare, "Dynamic message scheduling based on activity-aware residual belief propagation for asynchronous mmtc," *IEEE Wireless Communications Letters*, vol. 10, no. 6, pp. 1290–1294, 2021.

[34] P. Schniter, "A message-passing receiver for bicm-ofdm over unknown clustered-sparse channels," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 8, pp. 1462–1474, 2011.

[35] Q. Zou, H. Zhang, D. Cai, and H. Yang, "Message passing based joint channel and user activity estimation for uplink grant-free massive mimo systems with low-precision adcs," *IEEE Signal Processing Letters*, vol. 27, pp. 506–510, 2020.

[36] 3GPP, "Study on provision of low-cost machine-type communications (MTC) ues based on LTE," TS 36.888 V1.0.0, 2013.