Momentum Survey Propagation: A Statistical Physics Approach to Resource Allocation in mMTC

Andrea Ortiz, Rostyslav Olshevskyi, Daniel Barragan-Yani

***Abstract*—The importance of Massive Machine-Type Communications (mMTC) in Beyond 5G and 6G networks is supported by the ever-increasing number of connected devices in what are known as massive Internet of Things (IoT) networks. These networks bring unprecedented challenges for the distribution of the available communication resources because the allocation problems often lead to combinatorial optimization formulations which are known to be NP-hard. A fact that limits the performance of state-of-the-art techniques when the network size increases. To address this challenge, we take a new direction and propose a method based on statistical physics to address resource allocation problems in large networks. To this aim, we first show that resource allocation problems have the same structure as the problem of finding specific configurations in spin glasses, a type of disordered physical systems. Based on this parallel, we propose Momentum Survey Propagation, a resource allocation method to minimize the interference in mMTC networks. Our proposed approach extends the Survey Propagation method of statistical physics. Specifically, it exploits the so-called momentum technique, widely used in the context of neural networks, to improve the convergence properties of Survey Propagation. Our implementation is the first application of Survey Propagation to a wireless communication network. Through numerical simulations we show that Momentum Survey Propagation is a promising tool for the efficient allocation of communication resources in mMTC.**

***Index Terms*—Massive Machine-Type Communications, Survey Propagation, Resource Allocation, Interference Minimization.**

I. Introduction

A. Motivation and Challenges

The emergence of massive Internet of Things (IoT) networks ratifies the persistent relevance of Massive Machine-Type Communications (mMTC) in Beyond 5G and 6G networks [1]–[3]. In mMTC, millions of devices per square kilometer occasionally transmitting small amounts of data are considered [4]. One of the key challenges to support such massive communications is the scalability of radio resource management methods that avoid congestion in the radio access network [5], [6]. This challenge comes from the fact that the allocation problems often lead to combinatorial optimization formulations which are known to be NP-hard. Moreover, the complexity of the allocation problem is not constant but depends on the number of available resources, the number of devices, and the network topology.

Recent research shows that data aggregation is a potential solution for the congestion problem in mMTC [7], [8] because the devices do not communicate directly with the base stations providing the connectivity. Instead, a group of devices, termed aggregators, are tasked with the collection of data from their associated devices and the subsequent transmission to the base stations. Such hierarchical structure reduces the number of connections at the base stations, the power consumption, and improves the resource allocation efficiency [9].

Research on data aggregation has mainly focused on analyzing its benefits, and on the distribution of resources within each aggregator [6]–[8], [10]–[15]. However, little attention has been paid to the allocation of resources to the aggregators or the reuse of resources among them. This is a challenging problem because the larger the number of connected devices, the higher the density of the required aggregators and the more difficult it is to formulate tractable optimization problems. Moreover, a higher density of aggregators means that the distance between them is reduced and the interference caused by the reuse of resources grows. In this paper, we aim at filling this void by proposing a resource allocation method to minimize the interference in a mMTC network using data aggregation. Our proposed approach is based on the Survey Propagation method from statistical physics. Survey Propagation is a heuristic solution developed to solve large discrete optimization problems in a type of disordered physical systems called spin glasses [16]. By drawing a parallel between mMTC networks and spin glasses, i.e., by considering that the aggregators can be seen as the particles in a spin glass, we show how Survey Propagation can be adapted to address resource allocation problems in mMTC networks.

B. Related Work

Research interest in data aggregation has focused on analyzing the benefits of the scheme [10], [11] and on the distribution of resources within each aggregator [6]–[8], [12]–[15]. Using stochastic geometry, the authors in [10] characterize the interference and coverage performance of a cellular network serving a massive number of Machine-Type Communication Devices (MTCDs). Specifically, they propose an analytical framework to evaluate the success probability of the MTCDs' transmissions, the average number of served MTCDs and the average channel utilization. This work is extended in [11], where the authors introduce a hybrid access protocol that exploits orthogonal and non-orthogonal multiple access for the communication between the MTCDs and the aggregators.

The allocation of available resources to the MTCDs served by a single aggregator is investigated in [6], [12], [13]. Specifically, in [6], the authors focus on how to divide the resources among the two communication phases, i.e., from MTCDs to

Andrea Ortiz is currently with the Institute of Telecommunications at Vienna University of Technology, andrea.ortiz@tuwien.ac.at, Rostyslav Olshevsky is with Rice University, ro22@rice.edu. At the time of this work, Daniel Barragan-Yani was with the Department of Physics and Materials Science, University of Luxembourg.

aggregator and from aggregators to base stations. A similar scenario in considered in [12] where a group-based random access and data transmission scheme is proposed. In [13], the trade-off between network utility and resource allocation fairness under outage probability constraints for the uplink communication is investigated. Moreover, the fulfillment of MTCDs' Quality of Service (QoS) requirements is considered in [7], [8], [14], [15]. All the aforementioned works assume that the aggregators have preallocated resources and focus on maximizing the efficiency in the communication with their associated MTCDs. However, they do not investigate how the resources are distributed to the aggregators or how interference is handled among neighboring ones.

Combinatorial optimization problems, such as the allocation of resources to aggregators in mMTC, are NP-hard. Therefore, finding the optimal solution at a large scale is an open research question. Nevertheless, research effort has been put in the development of efficient heuristics [17]–[22]. The available solutions can be classified depending on the approach followed, i.e., message-passing algorithms [17], random variable assignment [18], greedy heuristics [19], [20], and learning approaches [21], [22]. The Belief Propagation method described in [17] is a well-known message-passing algorithm to solve combinatorial problems. Its main characteristic is the calculation and exchange of status messages among the variables in the problem in order to assign their values. In [18], the authors propose Walksat, a heuristic method to solve binary combinatorial optimization problems. The main idea behind Walksat is to randomly flip the values of the variables until all constraints are satisfied. Greedy heuristics are exploited in [19], [20], where the optimization problem is formulated as a graph-coloring problem and the variables are fixed according to some predefined order. Recently, graph neural networks architectures have been used to find solutions to constraint satisfaction problems [21], [22]. While a promising approach, so far graph neural networks have been considered for smaller networks and when the ration between the number of values to be assigned and variables is large. In fact, the drawback of the aforementioned approaches is that their performance decreases when the problem size increases and when the number of available resources is very limited, a fact that hinders their applicability in mMTC networks. To overcome this drawback, we propose a novel resource allocation strategy based on Survey Propagation, a message-passing method from statistical physics developed by [16]. Although Survey Propagation has shown impressive results in solving large binary combinatorial problems [16], [23], [24], it has not been yet applied to problems outside the physics domain.

C. Contributions

We investigate the resource allocation problem in mMTC networks with data aggregation. To this aim, a network consisting of a single base station and a large number of MTCDs is considered. Our goal is to develop a scalable resource allocation method for the distribution of resources among the aggregators that minimizes the interference in the network by leveraging methods from statistical physics. The contributions of the paper are summarized as follows:

- *Innovative Network Model:* We propose a model of the mMTC network inspired by spin glasses. Moreover, we formulate the resource allocation problem as a Constraint Satisfiability Problem (CSP) to find the resource allocation solution that minimizes the interference. Our model highlights the parallels between resource allocation in mMTC networks and minimum energy configurations in spin glasses.
- *Momentum Survey Propagation Method:* We propose a resource allocation method to distribute radio resources to the aggregators in a mMTC network. Our approach, termed *Momentum Survey Propagation*, extends the Survey Propagation method proposed in [16] and minimizes the interference in the network. Specifically, it exploits the so-called momentum technique, widely used in neural networks, to improve the performance of Survey Propagation making it more effective for large-scale network applications. Ours is the first implementation of Survey Propagation in wireless communication networks.
- *Superior Performance:* Through numerical simulations, we show that our proposed Momentum Survey Propagation outperforms reference approaches, namely, Survey Propagation [16], Belief Propagation [17] and Walksat [18]. Moreover, it is able to find zero interference allocation solutions at least 40% of the times and when the reference schemes fail to do so.

The rest of the paper is organized as follows. In Sec. II, we discuss parallel between spin glasses and mMTC networks. The considered system model is described in Sec. III and the corresponding resource allocation problem is formulated In Sec. IV. Our proposed Momentum Survey Propagation is explained in Sec. V and its convergence properties are discussed in Sec. VI. Numerical performance results are presented in Sec. VII and Sec. VIII concludes the paper.

II. Spin Glasses and mMTC Networks

A. Spin Glasses

Spin glasses are a fundamental type of disordered system studied in statistical physics. They refer to magnetic systems in which conventional ferromagnetic or antiferromagnetic long-range orders cannot be established due to some structural disorder. As a result, the magnetic moments or spins composing the system prefer to be arranged in random directions [25]–[27]. One way to imagine, or model, spin glasses is to assume that the spins are located at fixed points in a regular lattice, see Fig. 1, with disorder introduced via random couplings, or interactions, that follow a suitable distribution [27]–[29]. The spin glass model can be very well applied to describe communication systems. In the spin glass model of the mMTC network, each spin represents one aggregator and its direction corresponds to the orthogonal resource allocated to it, e.g., a frequency band or a time slot. For example, in Fig. 1, we assume eight orthogonal resources, this means, eight possible directions (North, South, East, West, North-East, North-West, South-East and South-West). To highlight the different directions, we have used different colors and patterns for each of them. The random coupling between two

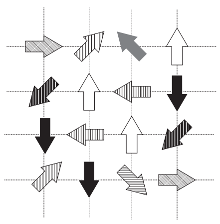


Fig. 1: Simplistic depiction of a spin glass in a grid.

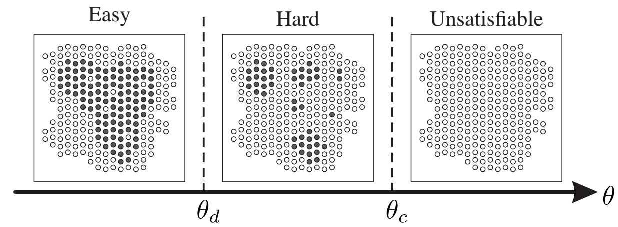


Fig. 2: Change in the solution space as θ increases. Figure based on [31], [32].

spins corresponds, in the mMTC case, to the interference caused by a pair of neighboring transmitting aggregators using the same orthogonal resource. Since nature has an invariable tendency to satisfy the principle of minimum energy [30], statistical physicists are often interested in finding minimal energy configurations in order to describe the equilibrium properties of relevant systems, like spin glasses [27]–[29]. From a communication perspective, we have a similar aim. Our goal is to find the resource allocation solution (orientation of spins) which minimizes the interference in the system (yields the minimum energy in the spin glass).

B. Constraint Satisfiability Problems (CSP)

CSPs are strongly related to the theory of spin glasses because the problem of finding the minimal energy configuration of the spin glass model can be written as a CSP. At the core of combinatorial optimization theory, CSPs deal with the question of whether a set Γ of constraints between a set $\mathcal{X}$ of discrete variables can be simultaneously satisfied. Each constraint is a clause formed by the logical disjunction (OR) of a subset of the variables or their negations [16]. The solution of the CSP is an assignment of the variables that guarantees that all the constraints in the problem are satisfied. The clauses in the CSP are associated to the interactions between neighboring spins and the orientation of each. The minimal energy configuration is an assignment that satisfies all clauses. Similarly, the resource allocation problem in mMTC networks can be formulated as a CSP. Details of such formulation are presented in Sec. IV.

Finding the minimal energy configuration and, similarly, the resource allocation that minimizes the interference are combinatorial problems. The design of algorithms to find configurations that fully satisfy the CSP and the determination of whether a given CSP can be satisfied, is a challenging task [16], [23], [32]. This is because finding a solution heavily depends on how constrained is the particular problem at hand. Let us define the ratio θ between the number $|\mathcal{X}|$ of variables and the number $|\Gamma|$ of constraints in the CSP as

$$\theta = \frac{|\Gamma|}{|\mathcal{X}|}. \tag{1}$$

There exists a critical threshold θ_c for which the CSP becomes unsolvable. When $\theta < \theta_c$, the CSP can be satisfied. Conversely, when $\theta > \theta_c$ the CSP is unsatisfiable [16], [23], [31]. Note that within the satisfiable region $\theta < \theta_c$ the complexity of finding a solution is not constant. Physicists have found that there exists an intermediate threshold θ_d that specifies a region $\theta_d < \theta < \theta_c$ where a CSP is still satisfiable but the solution is hard to find [16][1]. Such difficulty comes from the fact that when θ increases, the solution space becomes clusterized, as depicted in Fig. 2.

In Fig. 2, each circle represents a solution of the CSP. Filled circles mean the allocation satisfies the CSP, and empty circles denote that the allocation does not satisfy all the clauses. For $\theta < \theta_d$, the satisfying allocation (filled circles) are close to each other forming one single cluster. However, for larger values of θ, groups of satisfying allocations are apart forming many smaller clusters. The number of clusters increases exponentially until the CSP becomes unsatisfiable [16]. In the region $\theta_d < \theta < \theta_c$, the allocations in separate clusters are far apart and moving from one allocation in a cluster to some other allocation in another cluster requires simultaneously changing the value of many of the considered variables [17], [32].

For resource allocation in mMTC networks, the solution of the CSP is highly dependent on the network topology, the density and the number of available time-frequency resources. The threshold θ_c indicates the minimum number of time-frequency resources needed to find an allocation solution that minimizes the interference. Moreover, finding solutions in the range $\theta_d \leq \theta \leq \theta_c$ allows us to reduce the number of required resources compared to solutions in the range $\theta < \theta_d$. In this paper, we propose Momentum Survey Propagation to find a solution to the resource allocation problem, especially in the range $\theta_d \leq \theta \leq \theta_c$. The parallel between spin glasses and mMTC networks is summarized in Table I, where the main components of a spin glass model are mapped in a one-to-one manner to the main elements in the network.

III. System Model

We consider a mMTC network, as depicted in Fig. 3. Similar to [33]–[35], the network is formed by a single Base Station (BS) and K single-antenna MTCDs. We denote by $\mathcal{K}$ the set containing all MTCDs. To make an efficient use of the communication resources, a subset $\mathcal{M} \subset \mathcal{K}$ of MTCDs act as aggregators[2]. According to their locations, the remaining MTCDs $k \in \mathcal{M}^{\mathrm{C}}$ are associated to one aggregator $m \in \mathcal{M}$, with $\mathcal{K} = \mathcal{M} \cup \mathcal{M}^{\mathrm{C}}$. In our model, each MTCD communicates with a single aggregator. The aggregator transmits the aggregated data from different MTCDs to the BS through orthogonal dedicated channels depicted as dotted lines in Fig.

[1]Other works have identified more intermediate thresholds to make a finer characterization of the behavior of the solution space [31].

[2]Note that the aggregators in our model can be replaced by femto or pico base stations.

	Spin Glass	mMTC Network
Scenario	Spin Orientation Interaction	MTCD Time-frequency resource Interference
Goal	Find the minimal energy configuration	Find the allocation that minimizes the interference
Problem Formulation	Constraint Satisfiability Problem	
Solution Method	Momentum Survey Propagation	

TABLE I: Parallel between spin glasses and mMTC networks.

3. We assume that the BS has N time-frequency resources available for the communication between the aggregators and the MTCDs. These N resources are divided into $Q \in \mathbb{N}$, $Q \ll M$, resource pools. Every aggregator is assigned a resource pool $q \in \mathcal{Q}$, where $\mathcal{Q}$ is the set of available resource pools.

To minimize the interference in the network, we aim at orthogonal resource allocation, i.e., neighboring aggregators cannot share the same resource pool q. Two aggregators m and n are said to be neighbors if the received power $p^{\mathrm{Rx}}_{m,n} \in \mathbb{R}$ of a test interference signal $s_{m,n} \in \mathbb{C}$ sent from m and received at n is above a given threshold $\mu \in \mathbb{R}$. The wireless channel between two aggregators m and n is characterized by the channel coefficient $h_{m,n}$. Moreover, we assume each aggregator uses a fixed transmit power $p^{\mathrm{Tx}}_{m} \in \mathbb{R}$ for the communication with its associated MTCDs[3]. The power $p^{\mathrm{Rx}}_{m,n}$ of the received interference signal is calculated

$$p^{\mathrm{Rx}}_{m,n} = |h_{m,n}|^2 p^{\mathrm{Tx}}_{m} + \sigma^2_m, \tag{2}$$

where σ^2_m is the noise power of aggregator m. Additionally, channel reciprocity between aggregators is assumed, i.e., $h_{m,n} = h_{n,m}$. Consequently, $p^{\mathrm{Rx}}_{m,n} = p^{\mathrm{Rx}}_{n,m}$.

The BS controls the resource allocation, i.e., it decides which resource pool q is assigned to each aggregator. The binary variable $x_{q,m} \in \{0, 1\}$, indicates if resource pool q is assigned to aggregator m. To minimize the interference in the network and make an efficient use of the available resources, we impose two constraints on the resource allocation. The first constraint ensures that each aggregator m is assigned one resource pool q, i.e.,

$$\sum_{q \in \mathcal{Q}} x_{q,m} = 1, \quad \forall m \in \mathcal{M}. \tag{3}$$

The second constraint hinders the allocation of the same resource pool q to two neighboring aggregator m and n, i.e.,

$$x_{q,m} x_{q,n} = 0, \quad \forall m, n \in \mathcal{M}, q \in \mathcal{Q}. \tag{4}$$

We model the considered network as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertices and $\mathcal{E}$ is the set of edges. In our case, the set $\mathcal{V}$ of vertices contains all the aggregators $m \in \mathcal{M}$ such that $\mathcal{V} = \mathcal{M}$. The set $\mathcal{E}$ of edges $e = (m, n)$ contains the links between neighboring aggregator m, n, i.e., $\mathcal{E} = \{e = (m, n) : m, n \in \mathcal{V}, p^{\mathrm{Rx}}_{m,n} \geq \mu\}$. If the received power $p^{\mathrm{Rx}}_{m,n} \in \mathbb{R}$ of the test interference signal sent from m and received at n is above the threshold μ, an edge between aggregators m and n is established in the network graph $\mathcal{G}$ to indicate that they are neighbors. Note that although a single BS is considered, the model can be easily extended to multiple BSs by considering each of them separately. The potential interference between neighboring aggregators that connect to different BSs, can be included by adding these connections in the network graph $\mathcal{G}$ of each of BS.

[3]Power allocation can be done at each aggregator to maximize the performance of the communication with its respective MTCDs.

IV. Problem Formulation

In this section we formulate the time-frequency resource allocation problem. Initially, we present the corresponding optimization problem and describe its properties. Next, and to facilitate the application of our proposed Momentum Survey Propagation method, we reformulate the problem as a CSP and explain its corresponding factor graph representation.

A. Optimization Problem

Our goal is to find a resource allocation solution that minimizes the interference in the network. As the Q available resource pools are orthogonal, interference occurs when neighboring aggregators communicate using the same resource pool q. We term such event as a *conflict*. Finding the allocation solution that minimizes the interference is equivalent to finding the solution that minimizes the number of conflicts. Considering the constraints in (3) and (4), the optimization problem is written as

$$\underset{\{x_{q,m}\}_{q \in \mathcal{Q},\, m \in \mathcal{M}}}{\text{minimize}} \quad \sum_{q=1}^{Q} \sum_{n=1}^{M} \sum_{m=1}^{M} x_{q,m} x_{q,n} \tag{5a}$$

$$\text{subject to} \quad \sum_{q \in \mathcal{Q}} x_{q,m} = 1, \qquad \forall m \in \mathcal{M}, \tag{5b}$$

$$x_{q,m} \in \{0, 1\}, \quad \forall q \in \mathcal{Q}, m \in \mathcal{M} \tag{5c}$$

The problem in (5) is a non-linear integer programming problem which is known to be NP-hard. As a consequence, up to now, there is no polynomial-time algorithm to find its optimal solution.

B. Formulation as Constraint Satisfiability Problem

We propose Momentum Survey Propagation to solve the problem in (5). Our approach is a message-passing algorithm based on Survey Propagation. To facilitate its description in Sec. V, we reformulate the problem as a CSP.

Consider the graph $\mathcal{G}$ introduced in Sec. III. The CSP, denoted by γ, is formed by the set $\mathcal{X} = \{x_{q,m} : m \in \mathcal{V}, q \in \mathcal{Q}\}$ of the binary variables $x_{q,m}$ in (5c) and the set Γ of logical constraints in (5a) and (5b). The goal is to find a resource allocation solution $\mathbf{X} = \big((x_{1,1}, ..., x_{Q,})^{\mathrm{T}}, ..., (x_{q,M}, ..., x_{Q,M})^{\mathrm{T}}\big)$, where $(\cdot)^{\mathrm{T}}$ is the transpose operation, that ensures all the constraints in Γ are satisfied, in other words, $\gamma = \texttt{true}$. In the following, we rewrite the objective function in (5a) and the constraint (5b) as logical clauses.

The objective function in (5a) aims at the minimization of the number of conflicts. To this aim, we introduce constraint α_e for all edges $e \in \mathcal{E}$. Formally, α_e is defined as

$$\alpha_e = (\bar{x}_{1,m} \vee \bar{x}_{1,n}) \wedge (\bar{x}_{2,m} \vee \bar{x}_{2,n}) \wedge ... \wedge (\bar{x}_{Q,m} \vee \bar{x}_{Q,n}), \tag{6}$$

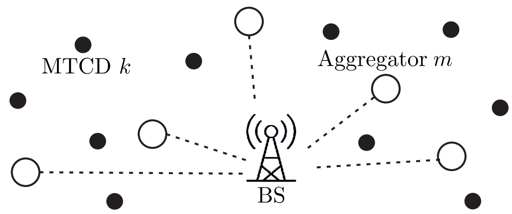


Fig. 3: Example of the considered network with one BS, $M = 5$ aggregators and a total of $K = 15$ MTCDs.

where $\wedge$ and $\vee$ are the logical conjunction (AND) and disjunction (OR) operators, respectively, and $\bar{x}_{q,n}$ is the negated variable $x_{q,n}$. The set of all α_e constraints in the CSP is denoted by $\mathcal{A}$.

Constraint (5b) states that each aggregator should be assigned one resource pool. This constraint is rewritten for every aggregator m as the logical clause β_m defined as

$$\beta_m = (x_{1,m} \vee x_{2,m} \vee ... \vee x_{Q,m}). \tag{7}$$

The set of all β_m constraints is denoted by $\mathcal{B}$. The CSP γ is then written as the logical conjunction of all the α_e and β_m constraints as

$$\gamma = \left(\bigwedge_{e \in \mathcal{E}} \alpha_e \right) \wedge \left(\bigwedge_{m \in \mathcal{M}} \beta_m \right). \tag{8}$$

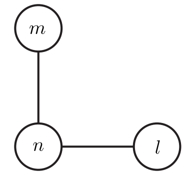
Let us now define the cost function C as the number of unsatisfied clauses in (8). A solution $\mathbf{X}$ is optimal when $\gamma(\mathbf{X}) = \texttt{true}$ and consequently, $C = 0$. This means, there is no interference between neighboring aggregators because no resource pools are shared among them. Given a network graph $\mathcal{G}$, the corresponding CSP γ is formed by $|\mathcal{X}| = QM$ variables and $|\Gamma| = |\mathcal{A}| + |\mathcal{B}| = Q|\mathcal{E}| + M$ constraints.

C. Factor Graph Representation

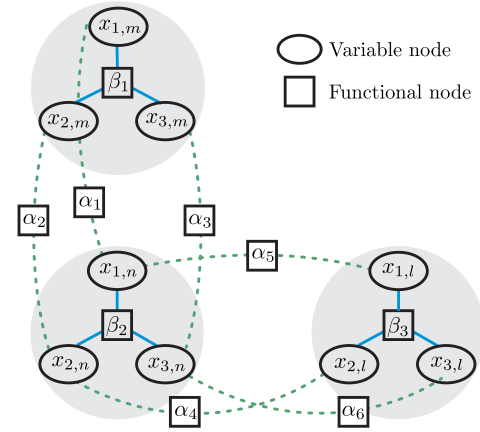
A better understanding of γ can be achieved by using the factor graph representation of the CSP [24]. A factor graph is a bipartite graph formed by two types of vertices, namely, variable nodes and functional nodes[4]. This means, the factor graph does not represent the mMTC network but the constraints that need to be fulfilled in order to achieve minimum interference. The factor graph provides a graphical description of the relation between the aggregators in the network based on the constraints in the CSP, and it is helpful to explain the resource allocation method proposed in Sec. V.

Consider the example in Fig. 4. On the left side, in Fig. 4a, a small mMTC network with three aggregators is depicted. The set of vertices is given by $\mathcal{V} = \{m, n, l\}$ and the set of edges by $\mathcal{E} = \{(m, n), (n, l)\}$. To build the corresponding factor graph, the constraints α_e and β_m, defined in (6) and (7) respectively, are considered. Specifically, the variable nodes, depicted with circles in Fig. 4b, are the variables $x_{q,m} \in \mathcal{X}$ that determine whether resource pool q is allocated to aggregator m or not. The set of all variables nodes is then the set $\mathcal{X}$. The functional nodes, shown with squares in Fig. 4b, correspond to the constraints $\alpha_e \in \mathcal{A}$ and $\beta_m \in \mathcal{B}$ defined in (6) and (7), respectively. The set of all functional nodes is $\Gamma = \mathcal{A} \cup \mathcal{B}$, and we term by $\zeta \in \Gamma$ any functional node. An edge between a variable and a functional node means that the variable node is included in the clause represented by the functional node. We denote by $\mathcal{X}(\zeta)$ the set containing all variable nodes $x_{q,m}$ considered in clause ζ. Similarly, we denote by $\Gamma(x_{q,m})$ the set of functional nodes ζ to which the variable node $x_{q,m}$ is connected. The type of line in the factor graph representation indicates whether the variable $x_{q,m}$ or its negation $\bar{x}_{q,m}$ is used. In Fig. 4b, the edges connecting the variable nodes with the functional nodes $\zeta = \alpha_e$ are depicted with dashed lines because the negated variables $\bar{x}_{q,m}$ are considered in this clause. On the contrary, the edges connecting the variable nodes with the $\zeta = \beta_m$ constraints are depicted with solid lines because the non-negated variables $x_{q,m}$ are considered. Edges between two variable nodes are not possible since variable nodes can only be related through clauses, i.e., through functional nodes.

[4]Note that the variable nodes and functional nodes do not correspond to the aggregators in the mMTC network, but to the variables and constraints associated to the CSP.



(a) Conventional graph.



(b) Factor graph.

Fig. 4: Conventional graph and factor graph representations of a network with $M = 3$ aggregators and $Q = 3$ available resources.

To highlight the relationship between functional and variable nodes, let us introduce some additional definitions. We define $\Gamma^+(x_{q,m})$ as the set of functional nodes in which $x_{q,m}$ appears non-negated (solid lines in Fig. 4b), and $\Gamma^-(x_{q,m})$ as the set of functional nodes in which $x_{q,m}$ appears negated (dashed lines in Fig. 4b), i.e., $\Gamma(x_{q,m}) = \Gamma^+(x_{q,m}) \cup \Gamma^-(x_{q,m})$. Additionally we define $\Gamma^{\mathrm{s}}_{\zeta}(x_{q,m})$ and $\Gamma^{\mathrm{u}}_{\zeta}(x_{q,m})$ as the sets containing all the functional nodes connected to $x_{q,m}$ which tend to make variable node $x_{q,m}$ satisfy ($\Gamma^{\mathrm{s}}_{\zeta}(x_{q,m})$) or not satisfy ($\Gamma^{\mathrm{u}}_{\zeta}(x_{q,m})$) clause ζ, respectively. From the definitions of α_e and β_m in (6) and (7), it is clear that a variable node $x_{q,m}$ can be connected to multiple α_e nodes but only to one

β_m node. Consequently, for functional nodes $\zeta = \alpha_e$ the sets $\Gamma^{\mathrm{s}}_{\alpha_e}(x_{q,m})$ and $\Gamma^{\mathrm{u}}_{\alpha_e}(x_{q,m})$ are defined as

$$\Gamma^{\mathrm{s}}_{\alpha_e}(x_{q,m}) = \mathcal{A}(x_{q,m}), \tag{9}$$

$$\Gamma^{\mathrm{u}}_{\alpha_e}(x_{q,m}) = \beta_m, \tag{10}$$

where $\mathcal{A}(x_{q,m})$ is the set of functional nodes α_e connected to $x_{q,m}$. In contrast, for functional nodes $\zeta = \beta_m$ the sets $\Gamma^{\mathrm{s}}_{\beta_m}(x_{q,m})$ and $\Gamma^{\mathrm{u}}_{\beta_m}(x_{q,m})$ are defined as

$$\Gamma^{\mathrm{s}}_{\beta_m}(x_{q,m}) = \beta_m, \tag{11}$$

$$\Gamma^{\mathrm{u}}_{\beta_m}(x_{q,m}) = \mathcal{A}(x_{q,m}). \tag{12}$$

V. Momentum Survey Propagation

In this section, we present our proposed Momentum Survey Propagation for resource allocation in mMTC networks. First, we describe Survey Propagation as presented in [16] in Sec. V-A. In Sec. V-B we describe the update procedure for the surveys, and introduce the proposed resource allocation algorithm in Sec. V-C.

A. Survey Propagation

The Survey Propagation method proposed in [16] is a physics-based approach to solve CSPs. In a nutshell, Survey Propagation is an iterative message-passing algorithm. The main idea behind it is the propagation of statistical information, or surveys, between the vertices of the factor graph. Note that Survey Propagation does not run over the mMTC network represented by the graph $\mathcal{G}$, but over the factor graph representation of the resource allocation problem associated to it. Moreover, the surveys are not physically transmitted messages between the BS and aggregators, but are an abstraction of messages "interchanged" between the nodes in the factor graph, i.e., between variable and functional nodes.

The surveys are used to identify the state of the variables $x_{q,m}$ in clusters that satisfy the CSP, i.e., whether the variables $x_{q,m}$ are fixed to a certain logical value "0" or "1", or if they are in an indifferent "*" state. In other words, if resource pool q cannot be allocated to aggregator m ($x_{q,m} =$ "0"), if resource pool q must be allocated to aggregator m ($x_{q,m} =$ "1"), or if aggregator m is indifferent to the allocation of q ($x_{q,m} =$ "$*$"). For example, such cases occur when a neighboring aggregator has been already allocated resource pool q, when the only available resource pool for aggregator m is q, and when aggregator m has many available resource pools and none of its neighbors are already using them, respectively. When the state of a variable $x_{q,m}$ is fixed to "0" or "1", it can be removed from the problem. This is because, for the satisfaction of the CSP, its value has been set. This means, only the logical value of the variables in the indifferent state remains to be identified. By repeating this procedure, the original problem is reduced. In every iteration, a smaller problem is considered until the space of satisfying solutions is formed by a single cluster.

The survey transmitted from any ζ to a connected variable node $x_{q,m}$ is a real number $\eta_{\zeta\to x_{q,m}} \in [0,1]$. The survey transmitted from a variable node $x_{q,m}$ to a connected functional node ζ is a triplet $\pi_{x_{q,m}\to\zeta} = (\pi^{\mathrm{u}}_{x_{q,m}\to\zeta}, \pi^{\mathrm{s}}_{x_{q,m}\to\zeta}, \pi^{*}_{x_{q,m}\to\zeta})$ formed by three real numbers $\pi^{\mathrm{u}}_{x_{q,m}\to\zeta}$, $\pi^{\mathrm{s}}_{x_{q,m}\to\zeta}$ and $\pi^{*}_{x_{q,m}\to\zeta}$. These surveys can be interpreted as probabilities of warning [17]. Specifically, $\eta_{\zeta\to x_{q,m}}$ can be seen as the probability that functional node ζ warns the variable node $x_{q,m}$ to take the correct value in order to satisfy clause ζ. Similarly, $\pi^{\mathrm{u}}_{x_{q,m}\to\zeta}$, $\pi^{\mathrm{s}}_{x_{q,m}\to\zeta}$ and $\pi^{*}_{x_{q,m}\to\zeta}$ can be interpreted as the probabilities that variable node $x_{q,m}$ sends a warning to functional node ζ informing it that $x_{q,m}$ cannot satisfy the clause ($\pi^{\mathrm{u}}_{x_{q,m}\to\zeta}$), that it can satisfy it ($\pi^{\mathrm{s}}_{x_{q,m}\to\zeta}$) or that it is indifferent ($\pi^{*}_{x_{q,m}\to\zeta}$). All these surveys are recursions calculated from the messages received from neighboring nodes in the factor graph of the CSP. Let us introduce the following definitions

$$P^{\mathrm{s}}_{\zeta,x_{q,m}} = \prod_{\xi\in\Gamma^{\mathrm{s}}_{\zeta}(x_{q,m})} \left(1-\eta_{\xi\to x_{q,m}}\right), \tag{13}$$

$$P^{\mathrm{u}}_{\zeta,x_{q,m}} = \prod_{\xi\in\Gamma^{\mathrm{u}}_{\zeta}(x_{q,m})} \left(1-\eta_{\xi\to x_{q,m}}\right), \tag{14}$$

where the sets $\Gamma^{\mathrm{s}}_{\zeta}(x_{q,m})$ and $\Gamma^{\mathrm{u}}_{zeta}(x_{q,m})$ are defined in (9) and (10) for $\alpha_e \in \mathcal{A}$, and in (11) and (12) for the $\beta_m \in \mathcal{B}$. The survey $\pi_{x_{q,m}\to\zeta} = (\pi^{\mathrm{u}}_{x_{q,m}\to\zeta}, \pi^{\mathrm{s}}_{x_{q,m}\to\zeta}, \pi^{*}_{x_{q,m}\to\zeta})$ transmitted from $x_{q,m}$ to ζ is calculated as [17]

$$\pi^{\mathrm{u}}_{x_{q,m}\to\zeta} = \left[1-P^{\mathrm{u}}_{\zeta,x_{q,m}}\right] P^{\mathrm{s}}_{\zeta,x_{q,m}}, \tag{15}$$

$$\pi^{\mathrm{s}}_{x_{q,m}\to\zeta} = \left[1-P^{\mathrm{s}}_{\zeta,x_{q,m}}\right] P^{\mathrm{u}}_{\zeta,x_{q,m}}, \tag{16}$$

$$\pi^{*}_{x_{q,m}\to\zeta} = P^{\mathrm{s}}_{\zeta,x_{q,m}} P^{\mathrm{u}}_{\zeta,x_{q,m}}. \tag{17}$$

The survey from ζ to $x_{q,m}$ is calculated as [17].

$$\eta_{\zeta\to x_{q,m}} = \prod_{x_{q,n}\in\mathcal{X}(\zeta)\setminus x_{q,m}} \left[\frac{\pi^{\mathrm{u}}_{x_{q,n}\to\zeta}}{\pi^{\mathrm{u}}_{x_{q,n}\to\zeta} + \pi^{\mathrm{s}}_{x_{q,n}\to\zeta} + \pi^{*}_{x_{q,n}\to\zeta}}\right]. \tag{18}$$

B. Momentum-based Survey Update

The surveys in (18) are calculated for each node in the factor graph according to the surveys received from neighboring nodes. That means, their calculation is an iterative process that concludes when the surveys reach a steady value, i.e., the difference between the surveys calculated in two consecutive iterations is less than a given threshold ϵ for all nodes, or when the maximum number T_{SP} of iterations is reached.

The convergence of the surveys is guaranteed on graphs with a tree structure [17]. However, the graphs associated to mMTC networks contain loops. Therefore, to improve the convergence properties of Survey Propagation when applied to mMTC networks, we investigate the concept of momentum. The idea of momentum was originally introduced in [36] as a tool to stabilize the optimization process of Belief Propagation. It is also widely used in the context of Neural Networks. Here, we adapted it to Survey Propagation.

Algebraically, momentum is defined as the weighted sum of the surveys at two consecutive times stamps t_{SP} and $t_{\mathrm{SP}}-1$

$$\eta_{\zeta\to x_{q,m}}(t_{\mathrm{SP}}) = (1-\mu)\eta_{\zeta\to x_{q,m}}(t_{\mathrm{SP}}) + \mu\eta_{\zeta\to x_{q,m}}(t_{\mathrm{SP}}-1) \tag{19}$$

where $\eta_{\zeta \rightarrow x_{q,m}}(t_{\mathrm{SP}})$ is the value of survey $\eta_{\zeta \rightarrow x_{q,m}}$at time stamp t_{SP} and μ is the momentum factor. The value of μ controls the rate at which the surveys are updated[5].

Algorithm 1 summarizes the procedure used to calculate the surveys. First, the surveys are randomly initialized for all the nodes in the factor graph, and the maximum number of iterations T_{SP}, the convergence threshold ϵ, the momentum factor μ and the indicator variable convergenceSP are set (lines 1-2). While the surveys have not yet converged, a random permutation is generated for the order in which an individual survey will be updated (lines 3-4). Using this order, the surveys are updated using (18)-(19) (line 5). Then, for every survey, we evaluate whether the value has converged according to the predefined threshold ϵ (lines 6-12). If the condition is not fulfilled for all the nodes in the factor graph, the procedure is repeated (line 10). On the contrary, if the values of $\eta_{\zeta \rightarrow x_{q,m}}$ have reached convergence, then these values are used as surveys (line 15). If the values of the surveys do not converge and the maximum number of iterations T_{SP} is reached, the surveys are initialized once again and Algorithm 1 runs once more.

Algorithm 1: Surveys update

1 Randomly initialize the surveys $\eta_{\zeta \rightarrow x_{q,m}}$ for all $\zeta \in \Gamma$ and $x_{q,m} \notin \mathcal{X}^*$
2 Initialize $T_{\mathrm{SP}},\ \epsilon,\ \mu$
3 Set convergenceSP = 0 and counter $t_{\mathrm{SP}} = 1$
4 **while** $(t_{\mathrm{SP}} \leq T_{\mathrm{SP}})$ **and** (*convergenceSP = 0*) **do**
5 Generate a random permutation for the order in which the surveys $\eta_{\zeta \rightarrow x_{q,m}}$ will be updated
6 Update $\eta_{\zeta \rightarrow x_{q,m}}$ using the defined order // Eq. (18)–(19)
7 Set convergenceSP = 1
8 **for** *every survey* $\eta_{\zeta \rightarrow x_{q,m}}$ **do**
9 **if** $|\eta_{\zeta \rightarrow x_{q,m}}(t_{\mathrm{SP}}) - \eta_{\zeta \rightarrow x_{q,m}}(t_{\mathrm{SP}} - 1)| > \epsilon$ **then**
10 Set convergenceSP = 0
11 Go to line 14
12 **end**
13 **end**
14 Set $t_{\mathrm{SP}} = t_{\mathrm{SP}} + 1$
15 **end**
16 **return** $\eta_{\zeta \rightarrow x_{q,m}}$

Algorithm 2: Proposed resource allocation algorithm

1 Build CSP and factor graph from mMTC network // Sec. IV
2 Initialize parameters T, T' and set $t = 1$, $t' = 1$
3 **while** $\mathcal{X}^* \neq \mathcal{X}$ **do**
4 **while** $(t \leq T) \wedge (\mathcal{X}^* \neq \mathcal{X}) \wedge (t' \leq T')$ **do**
5 Calculate the surveys $\eta_{\alpha_e \rightarrow x_{q,m}}$ // Alg. 1
6 **if** *the surveys converge to a steady value* **then**
7 **if** *all the surveys have values larger than zero* **then**
8 Set $t' = 1$ and calculate $W^+_{x_{q,m}}$, $W^-_{x_{q,m}}$ for each $x_{q,m}$ // Eq. (25)–(22)
9 Find variable node $x^*_{q,m}$ with largest bias difference $|W^+_{x_{q,m}} - W^-_{x_{q,m}}|$
10 **if** $W^+_{x^*_{q,m}} > W^-_{x^*_{q,m}}$ **then**
11 Set $x^*_{q,m} = 1$ and $x_{r,m} = 0,\ \forall r = 1, ..., Q, r \neq q$
12 **else**
13 Set $x^*_{q,m} = 0$
14 **end**
15 Remove $x^*_{q,m}$ and its edges from the factor graph
16 Update $\mathcal{X}^*$ and set $t = t + 1$
17 **else**
18 Use a heuristic for the remaining nodes, e.g., [18]
19 **end**
20 **else**
21 Set $t' = t' + 1$
22 **end**
23 **end**
24 **end**
25 **return** $x^*_{q,m} \forall q \in \mathcal{Q},\ m \in \mathcal{M}$

C. Resource Allocation Algorithm

In this section, we describe the proposed Momentum Survey Propagation algorithm to find the resource allocation solution that minimizes the interference in mMTC networks. It is an iterative approach that aims at fixing the value of the variables $x_{q,m}$ based on the exchanged survey messages. In other words, in each iteration $t = 1, ..., T$, where T is the maximum number of iterations, the algorithm fixes the allocation of a resource pool q to one or more aggregators m according to the values of the surveys.

The proposed algorithm is presented in Algorithm 2. The first step is to formulate the resource allocation problem as a CSP and build the corresponding factor graph, as explained in Sec. IV (line 1). Second, the algorithm parameters and variables are initialized (line 2). Let $\mathcal{X}^* \subseteq \mathcal{X}$ be the set of variable nodes $x_{q,m}$ whose value has been fixed. The execution continues while there are still aggregators without an allocated resource pool q, i.e., $\mathcal{X}^* \neq \mathcal{X}$, while the number t of iterations is less than the maximum value T, and while the number t' of times Algorithm 1 has been run in a single iteration does not exceed a maximum value T' (line 3-4). The last condition comes from the fact that, as mentioned in Sec. V and VI, the convergence of the Survey Propagation method is not guaranteed. Therefore, we include the variable T' as the maximum number of times the Survey Propagation method can be run in a single iteration. Once the surveys $\eta_{\zeta \rightarrow x_{q,m}}$ converge to a steady value (line 5-6), we evaluate if their value is equal to zero (line 7). If this is the case, we face the trivial solution and heuristics, such as [18] or [19], are used to allocate the resources pools to the remaining aggregators (line 18). If the surveys $\eta_{\zeta \rightarrow x_{q,m}} > 0$ then the bias of each variable $x_{q,m}$ towards the possible logical states "0" and "1" is calculated (line 8). The biases represent, according to the surveys, how sure is aggregator m about being allocated (or not) resource pool q. The bias towards the logical state "0" is denoted as $W^-_{x_{q,m}}$ and the bias towards "1" is denoted by $W^+_{x_{q,m}}$. Let us introduce $\pi^+_{x_{q,m}}$, $\pi^-_{x_{q,m}}$ and $\pi^0_{x_{q,m}}$ as

$$\pi^+_{x_{q,m}} = \left[1 - P^+_{x_{q,m}}\right] P^-_{x_{q,m}}, \tag{20}$$

$$\pi^-_{x_{q,m}} = \left[1 - P^-_{x_{q,m}}\right] P^+_{x_{q,m}}, \tag{21}$$

$$\pi^0_{x_{q,m}} = \prod_{\zeta \in \Gamma(x_{q,m})} (1 - \eta_{\zeta \rightarrow x_{q,m}}), \tag{22}$$

[5]Intuitively, the momentum factor is analogous to the learning rate in machine learning algorithms

where,

$$P^{+}_{x_{q,m}} = \prod_{\zeta \in \Gamma^{+}(x_{q,m})} (1 - \eta_{\zeta \to x_{q,m}}), \quad (23)$$

$$P^{-}_{x_{q,m}} = \prod_{\zeta \in \Gamma^{-}(x_{q,m})} (1 - \eta_{\zeta \to x_{q,m}}). \quad (24)$$

The bias are then calculated as [17]

$$W^{-}_{x_{q,m}} = \frac{\pi^{-}_{x_{q,m}}}{\pi^{+}_{x_{q,m}} + \pi^{-}_{x_{q,m}} + \pi^{0}_{x_{q,m}}}, \quad (25)$$

$$W^{+}_{x_{q,m}} = \frac{\pi^{+}_{x_{q,m}}}{\pi^{+}_{x_{q,m}} + \pi^{-}_{x_{q,m}} + \pi^{0}_{x_{q,m}}}. \quad (26)$$

The next step is then to find the variable node $x^{*}_{q,m}$ with the largest absolute difference $|W^{+}_{x_{q,m}} - W^{-}_{x_{q,m}}|$ between biases (line 9). If $W^{+}_{x^{*}_{q,m}} \geq W^{-}_{x^{*}_{q,m}}$, resource pool q is allocated to aggregator m (line 11). If $W^{+}_{x^{*}_{q,m}} < W^{-}_{x^{*}_{q,m}}$ variable $x^{*}_{q,m} = 0$ (line 13). After fixing the value of $x^{*}_{q,m}$, the variable node and its edges are removed from the factor graph (line 15). Next, the set $\mathcal{X}^{*}$ is updated (line 16). If all the variables nodes in the factor graph have been set, then the algorithm returns this assignment as the resource allocation solution. Otherwise, another iteration starts.

VI. Convergence of Survey Propagation

As reported by the proposers of Survey Propagation, there is no general proof of convergence of the algorithm for arbitrary CSPs [16]. This is because the survey update in (18) is based on the cavity method from Statistical Physics. The cavity method is a state-of-the-art non-rigorous approach to calculate minimum energy states in spin glasses. Turning the cavity method into a rigorous theory is an open research question [17], [24]. As in any message-passing algorithm, the assumption that the messages transmitted among neighboring nodes in the factor graph representation are statistically independent, can only be guaranteed when the considered graphs have a tree structure. This means, the solution of the cavity method, and in turn, the Survey Propagation approach is exact only for factor graphs with a tree structure. Nevertheless, numerical results have shown that the application of Survey Propagation is not limited to only trees, but is able to find solutions, and outperform other state-of-the-art techniques, in cases when the considered graph has a local tree-like structure [16], [17]. The intuition behind this is that when the considered network has a tree-like structure, the length of the existing loops in the graph grows with the number of considered nodes. Thus for large mMTC networks the correlation between the exchanged surveys among neighboring nodes is reduced. The use of momentum improves the convergence of Survey Propagation because including the estimates of the previous surveys in the survey update rule can help stabilize it values.

VII. Numerical Evaluation

In this section, we evaluate the performance of the proposed Momentum Survey Propagation algorithm via numerical simulations. For the evaluation, 1000 independent realizations are considered. Each realization corresponds to the generation of a random mMTC network. To guarantee the tree-likeness of the generated networks, we consider an Erdös-Renyi model for each graph $\mathcal{G}$ representing the mMTC network. To ensure that the CSPs lie in the solvable but difficult interval $\theta_d \leq \theta \leq \theta_c$ and based on the results in [31], we assume that the probability P that two aggregators are neighbors in the Erdös-Renyi model is $P = \omega/M$, where ω is the critical threshold introduced in [31] and M is the number of considered aggregators.

For our simulations, we consider, unless otherwise specified, a network with $M = 10^3$ aggregators and $Q = 4$ available resource pools. Furthermore, $\omega = 7.5$. For our Momentum Survey Propagation we consider a momentum factor $\mu = 0.3$, a threshold $\epsilon = 10^{-3}$ and a maximum of $T_{\mathrm{SP}} = 10$ iterations for the convergence of the surveys. Additionally, the maximum number T of iterations is set to $T = IQ$ iterations. To evaluate the performance of our proposed method, we compare it to the following approaches:

- Survey Propagation [16]: This approach corresponds to the original Survey Propagation algorithm as proposed in [16] for the solution of CSPs.
- Walksat [18]: Heuristic method to solve binary combinatorial optimization problems in which the values of the variables in the CSP are randomly flipped until all constraints are satisfied or a maximum number $T_{\mathrm{flip}} = 10^5$ of flips is reached.
- Belief Propagation [17]: In this case, the resource allocation problem is solved using the Belief Propagation algorithm described in [17]. Belief Propagation is a message passing algorithm in which only the probability of each variable node $x_{i,q}$ taking the logical "0" or the logical "1" value is calculated

Figure 5 shows the number of times a zero interference allocation is found, in percentage, as a function of the critical threshold ω. Zero interference allocation means that the available resource pools are allocated without incurring in any conflict, i.e., the interference is minimized. In the figure it can be seen that the performance of all the approaches decreases as ω increases. As the probability of two nodes being neighbors is $P = \omega/M$, the larger the ω, the larger the probability that two aggregators are neighbors. For a fixed Q, the larger the probability P means that the allocation problem becomes harder as more constraints need to be simultaneously fulfilled in the CSP. At $\omega = 7.5$, our proposed approach is able to find zero interference allocations 98% of the time, which corresponds to a gain of 6%, 18% and 25% with respect to Survey Propagation, Walksat and Belief Propagation, respectively. Note, however, that the gains of our proposed increase with ω. For $\omega = 8$, the performance of Momentum Survey Propagation is at least three times higher than Survey Propagation, which is the best reference scheme. For $\omega = 8.2$, the performance of Momentum Survey Propagation is approximately six times higher than that of Survey Propagation. This means, the harder the allocation problem, the larger the gains obtained by using our proposed scheme. The good performance of Momentum Survey Propagation is based on the fact the allocation of resources is not done by considering only the number of neighbors each node has, like Walksat does. The exchange

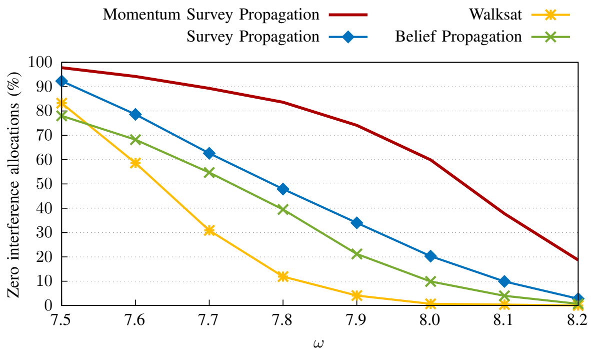


Fig. 5: Percentage of the number of times a zero interference allocation is found versus the critical threshold ω when $Q = 4$.

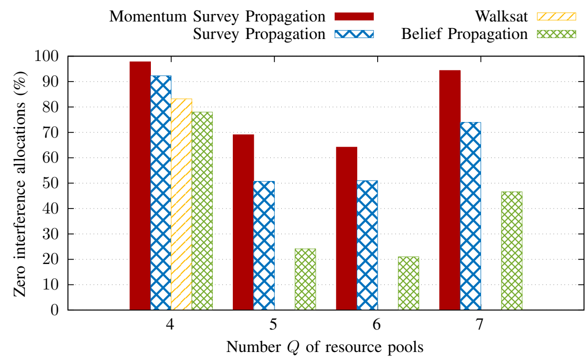


Fig. 6: Percentage of the number of times a zero interference allocation is found versus the number Q of available resource pools.

Resource Pools	Critical Threshold
$Q = 4$	$\omega = 7.5$
$Q = 5$	$\omega = 12.5$
$Q = 6$	$\omega = 17.5$
$Q = 7$	$\omega = 22.5$

TABLE II: Critical threshold ω according to the number Q of available resource pools.

of surveys in the proposed approach allows us to consider the impact an allocation will have on the satisfiability of the whole problem, and the use of a momentum factor helps to improve their convergence compared to Survey Propagation. Although Belief Propagation is also a message-passing algorithm, it performance is greatly reduced as it assumes a tree structure for the network, which is clearly not valid.

We evaluate the impact of the number Q of resource pools on the performance in Figure 6. For the simulations, we vary the number of available resource pools from $Q = 4$ to $Q = 7$. Note that for a fixed P, a larger Q results in simpler allocation problems because there are more resources to distribute. Therefore, to ensure that the complexity of the allocation problems lies in the range $\theta_d \leq \theta \leq \theta_c$, for each considered Q we increase ω as shown in Table II. The non-monotonic behaviour observed in all the considered approaches is due to the fact that the complexity of the allocation problems for each ω and Q combination is not equal. Nevertheless, it can be seen that our proposed approach outperforms the reference schemes across all the range of considered Q. For $Q = 5$ Momentum Survey Propagation achieves a performance two times higher than Belief Propagation and 36% higher than Survey Propagation. For $Q > 4$, Walksat is not able to find zero interference allocation solutions. This means that message-passing algorithms are more suitable to solve resource allocation problems when the number Q of resource pools increases.

In Figure 7, we vary the number M of aggregators in the network from $M = 10^3$ to $M = 10^4$, while keeping the number of resource pools constant with $Q = 4$ and $\omega = 8$. It can be seen that our proposed Momentum Survey Propagation outperforms the reference schemes in all the network size range. Moreover, Walksat and Belief Propagation are not able to find zero interference allocations for $M > 2 \times 10^3$ and $M > 3 \times 10^3$, respectively. Additionally, as observed in Figure 5, the gains of our proposed approach, with respect to the original Survey Propagation strategy, increase when the allocation problem becomes more difficult, i.e., for larger M. Considering a larger M makes the resource allocation problem more difficult because more constraints have to be simultaneously considered in the CSP as the number of constraints depend on the number of aggregators in the network. Furthermore, note that for our proposed approach, increasing the network size ten times only causes a decrease of 10% in the performance.

The performance of our proposed Momentum Survey Propagation as a function of the momentum factor μ is shown in Figure 8. In this case, we consider two values for Q, namely, $Q = 4$ and $Q = 7$. It can be seen that both of the considered values of Q exhibit a monotonic behaviour, where $\mu = 0$ corresponds to the original Survey Propagation algorithm. The figure shows that the inclusion of momentum in the update of the surveys improves the performance as long as $\mu \leq 0.7$. The maximum performance is achieved for $\mu = 0.3$. However, this maximum is only empirically found over the considered range. The larger μ, the larger the impact the value of the survey at time $t-1$ has on the survey update at time t. For $t \geq 0.7$, the use of momentum hinders the performance compared to the original Survey Propagation because more weight is put into past updates affecting the survey's convergence.

VIII. Conclusions

Motivated by the scalability challenges posed by the next generation of wireless networks, in this paper the resource allocation problem for large mMTC networks is studied. To address this problem, we propose a novel approach termed Momentum Survey Propagation, which is based on statistical physics. Specifically, we first introduce a model of the wireless network inspired by spin glasses, a type of disordered physical systems. Based on this model, we show how the resource allocation problem can be written as a constraint satisfiability problem. This formulation allows us to exploit the Survey Propagation method from statistical physics to find resource allocation solutions that minimize the interference. Our proposed approach extends Survey Propagation and improves its convergence properties by introducing the concept of momentum. The main advantage of our proposed Momentum

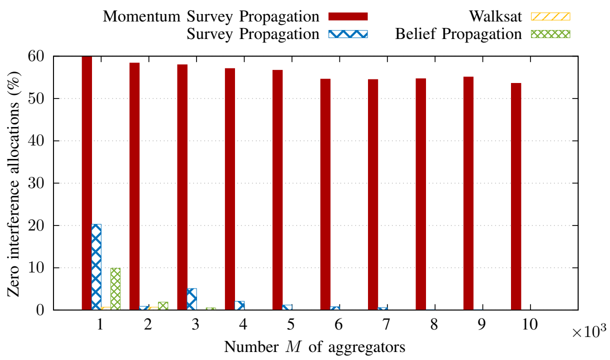


Fig. 7: Percentage of the number of times a zero interference allocation is found versus the number M of aggregators.

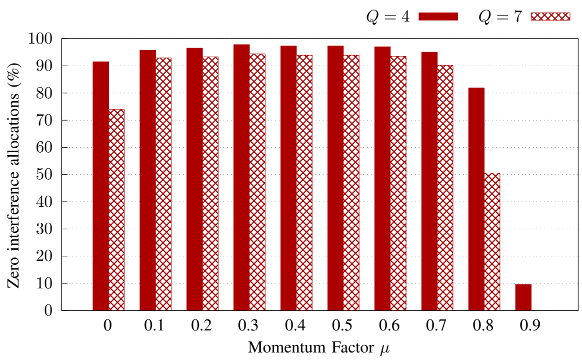


Fig. 8: Percentage of the number of times a zero interference allocation is found versus the momentum factor μ for $Q = 4$ and $Q = 7$.

Survey Propagation is that it can be applied to networks with a large number of nodes. For a fixed number of resources, it is able to find zero interference allocation solutions at least 40% of the time and when the reference schemes fail to do so In our opinion, these results should be the starting point and, at the same time, motivate a joint effort of both communities. From this joint work, an improvement of the available methods from statistical physics can be achieved and a new set of tools to meet the challenges posed by next generation wireless communication networks can be developed. Furthermore, our proposed Momentum Survey Propagation can be used to generate training data to develop learning approaches considering large networks. Future work aims at the consideration of joint resource and power allocation problems a well as dynamic environments.

Acknowledgement

This work has been funded by the Vienna Science and Technology Fund (WWTF) [Grant ID: 10.47379/VRG23002], by the BMBF project Open6GHub, grant number 16KISK014, and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Projektnummer 210487104 - SFB 1053 MAKI.

References

[1] U. M. Malik, M. A. Javed, S. Zeadally, and S. u. Islam, "Energy-efficient fog computing for 6G-enabled massive IoT: Recent trends and future opportunities," *IEEE Internet of Things J.*, vol. 9, no. 16, pp. 14 572–14 594, 2022.

[2] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. Dobre, and H. V. Poor, "6G internet of things: A comprehensive survey," *IEEE Internet of Things J.*, vol. 9, no. 1, pp. 359–383, 2022.

[3] X. Zhou, W. Liang, K. Yan, W. Li, K. I.-K. Wang, J. Ma, and Q. Jin, "Edge-enabled two-stage scheduling based on deep reinforcement learning for internet of everything," *IEEE Internet of Things J.*, vol. 10, no. 4, pp. 3295–3304, 2023.

[4] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55 765–55 779, 2018.

[5] N. Xia, H.-H. Chen, and C.-S. Yang, "Radio resource management in machine-to-machine communications—a survey," *IEEE Commun. Surveys & Tutorials*, vol. 20, no. 1, pp. 791–828, 2018.

[6] G. Bartoli, R. Fantacci, and D. Marabissi, "Resource allocation approaches for two-tiers machine-to-machine communications in an interference limited environment," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 9112–9122, 2019.

[7] T. Salam, W. U. Rehman, and X. Tao, "Cooperative data aggregation and dynamic resource allocation for massive machine type communication," *IEEE Access*, vol. 6, pp. 4145–4158, 2018.

[8] W. U. Rehman, T. Salam, A. Almogren, K. Haseeb, I. Ud Din, and S. H. Bouk, "Improved resource allocation in 5G MTC networks," *IEEE Access*, vol. 8, pp. 49 187–49 197, 2020.

[9] T. Salam, W. U. Rehman, and X. Tao, "Data aggregation in massive machine type communication: Challenges and solutions," *IEEE Access*, vol. 7, pp. 41 921–41 946, 2019.

[10] J. Guo, S. Durrani, X. Zhou, and H. Yanikomeroglu, "Massive machine type communication with data aggregation and resource scheduling," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 4012–4026, 2017.

[11] O. L. A. López, H. Alves, P. H. J. Nardelli, and M. Latva-aho, "Aggregation and resource scheduling in machine-type communication networks: A stochastic geometry approach," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4750–4765, July 2018.

[12] T. Wang, Y. Wang, C. Wang, Z. Yang, and J. Cheng, "Group-based random access and data transmission scheme for massive MTC networks," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8287–8303, 2021.

[13] D. Han, H. Minn, U. Tefek, and T. J. Lim, "Network dimensioning, QoE maximization, and power control for multi-tier machine-type communications," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 859–872, 2019.

[14] N. M. Ahmed and N.-E. Rikli, "QoS-based data aggregation and resource allocation algorithm for machine type communication devices in next-generation networks," *IEEE Access*, vol. 9, pp. 119 735–119 754, 2021.

[15] R. Chai, C. Liu, and Q. Chen, "Energy efficiency optimization-based joint resource allocation and clustering algorithm for M2M communication systems," *IEEE Access*, vol. 7, pp. 168 507–168 519, 2019.

[16] M. Mézard, G. Parisi, and R. Zecchina, "Analytic and algorithmic solution of random satisfiability problems," *Science Mag.*, vol. 297, pp. 812–815, 2002.

[17] A. Braunstein, M. Mézard, and R. Zecchina, "Survey propagation: An algorithm for satisfiability," *Random Structures & Algorithms*, vol. 27, no. 2, pp. 201–226, 2005.

[18] B. Selman, H. A. Kautz, and B. Cohen, "Local search strategies for satisfiability testing," in *Cliques, Coloring, and Satisfiability*, 1993, p. 521–532.

[19] S. Ramanathan, "A unified framework and algorithm for channel assignment in wireless networks," *Wireless Networks*, vol. 5, no. 2, pp. 81–94, 1999.

[20] H. Liu, J. Zhang, S. Jin, and B. Ai, "Graph coloring based pilot assignment for cell-free massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9180–9184, 2020.

[21] J. Tönshoff, M. Ritzert, H. Wolf, and M. Grohe, "Graph neural networks for maximum constraint satisfaction," *Frontiers in Artificial Intelligence*, vol. 3, 2021.

[22] J. Tönshoff, B. Kisin, J. Lindner, and M. Grohe, "One model, any csp: Graph neural networks as fast global search heuristics for constraint satisfaction," in *Int. Joint Conf. Artificial Intelligence, IJCAI-23*, 2023, pp. 4280–4288.

[23] M. Mézard and R. Zecchina, "Random k-satisfiability problem: From an analytic solution to an efficient algorithm," *Physical Review E*, vol. 66, p. 056126, Nov 2002.

[24] A. Braunstein, M. Mézard, M. Weigt, and R. Zecchina, "Constraint satisfaction by survey propagation," in *Computational Complexity and Statistical Physics*, 2006.

[25] K. Binder and A. P. Young, "Spin glasses: Experimental facts, theoretical concepts, and open questions," *Reviews of Modern Physics*, vol. 58, pp. 801–976, Oct 1986.

[26] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond: An Introduction to the Replica Method and its Applications*. World Scientific, 1987.

[27] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*. Oxford University Press, 2001.

[28] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *J. Physics F: Metal Physics*, vol. 5, no. 5, pp. 965–974, may 1975.

[29] D. Sherrington and S. Kirkpatrick, "Solvable model of a spin-glass," *Physical Review Lett.*, vol. 35, pp. 1792–1796, Dec 1975.

[30] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*. Wiley, 1985.

[31] L. Zdeborová and F. Krząkała, "Phase transitions in the coloring of random graphs," *Physical Review E*, vol. 76, p. 031131, Sep 2007.

[32] E. Maneva, E. Mossel, and M. J. Wainwright, "A new look at survey propagation and its generalizations," *J. ACM*, vol. 54, no. 4, pp. 17–58, 2007.

[33] X. Bian, Y. Mao, and J. Zhang, "Grant-free massive random access with retransmission: Receiver optimization and performance analysis," *IEEE Trans. Commun.*, vol. 72, no. 2, pp. 786–800, 2024.

[34] Z. Ma, W. Wu, F. Gao, and X. Shen, "Model-driven deep learning for non-coherent massive machine-type communications," *IEEE Trans. Wireless Commun.*, vol. 23, no. 3, pp. 2197–2211, 2024.

[35] Y. Liu, B. Clerckx, and P. Popovski, "Network slicing for eMBB, URLLC, and mMTC: An uplink rate-splitting multiple access approach," *IEEE TRans. Wireless Commun.*, vol. 23, no. 3, pp. 2140–2152, 2024.

[36] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," p. 467–475, 1999.

Andrea Ortiz is a Vienna Research Group Leader holding a Tenure-Track Position funded by WWTF at TU Wien. Previously, she was a Post-Doctoral Researcher at TU Darmstadt. Her research interests include the use of reinforcement learning and signal processing for resource allocation in wireless communications. She is the recipient of several awards, including the "Dr. Wilhelmy-VDE-Preis" given by German Association for Electrical, Electronic and Information Technologies (VDE) and the WWTF Vienna Research Group for Young Investigators Grant.

Rostyslav Olshevskyi holds a Master's Degree in Information and Communication Engineering from TU Darmstadt. Currently, he is a PhD student at Rice University in Texas. His research interests lie in distributed optimization and the application of graph neural networks in wireless communication systems.

Daniel Barragan-Yani is a Post-Doctoral researcher at the German Aerospace Center (DLR). Previously, he was a Post-Doctoral Researcher at the University of Luxembourg. His research focuses on advancing computational materials science to understand technologically relevant materials and their defects, and on finding new paths to use techniques of condensed matter physics for interdisciplinary problems. He was a recipient of the Dean's prize for excellence in Research (TU Darmstadt) and was awarded a Marie Skłodowska-Curie individual fellowship in 2020.