2021 19th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt)

BEAMWAVE: Cross-layer Beamforming and Scheduling for Superimposed Transmissions in Industrial IoT mmWave Networks

Luis F. Abanto-Leon, and Matthias Hollick, and Gek Hong (Allyson) Sim

Secure Mobile Networking Lab, Technische Universität Darmstadt, Germany

{labanto, mhollick, asim}@seemoo.tu-darmstadt.de

***Abstract*—The omnipresence of IoT devices in Industry 4.0 is expected to foster higher reliability, safety, and efficiency. However, interconnecting a large number of wireless devices without jeopardizing the system performance proves challenging. To address the requirements of future industries, we investigate the cross-layer design of beamforming and scheduling for layered-division multiplexing (LDM) systems in millimeter-wave bands. Scheduling is crucial as the devices in industrial settings are expected to proliferate rapidly. Also, highly performant beamforming is necessary to ensure scalability. By adopting LDM, multiple transmissions can be non-orthogonally superimposed. Specifically, we consider a *superior-importance control multicast message* required to be ubiquitous to all devices and *inferior-importance private unicast messages* targeting a subset of scheduled devices. Due to NP-hardness, we propose `BEAMWAVE`, which decomposes the problem into *beamforming* and *scheduling*. Through simulations, we show that `BEAMWAVE` attains near-optimality and outperforms other competing schemes.**

***Index Terms*—cross-layer, beamforming, scheduling, unicast, multicast, layered-division multiplexing, industrial IoT, mmWave.**

I. Introduction

Industry 4.0 envisions automated factories with a massive number of interconnected industrial internet-of-things (IoT) devices [1], such as sensors, actuators, programmable logic devices, and access points. Such degree of interconnectivity is expected to facilitate ultra-precise control and seamless coordination, thus enabling extremely efficient and dependable manufacturing processes [2]. In the existing industrial settings, the majority of stationary devices are interconnected through redundant wired connections to guarantee communications with high reliability. However, with the upsurge of devices in smart industries, wired solutions will encounter the following problems: *(i)* intricate implementation complexity to interconnect a massive number of devices, *(ii)* increased operational costs due to hard-wiring, *(iii)* limited maneuverability of articulated robots, and *(iv)* communication infeasibility with autonomous mobile freight transport. In contrast, wireless solutions can substantially simplify the deployment complexity and reduce maintenance costs while promoting the adoption of more flexible mechanics and mobile apparatus. *Thus, the transformation from wired to wireless infrastructure is an appealing strategy towards the evolution of industries.*

By harnessing *millimeter-wave* (mmWave) and *massive multiple-input multiple-output* (mMIMO), high spectral efficiency has been demonstrated (e.g., [3], [4]). Specifically, mmWave is an attractive substitute for the saturated sub-6 GHz spectrum due to broad bandwidth availability. Also, because of the shorter wavelength, mmWave requires miniature antennas that can be easily embedded onto small industrial devices. Further, mmWave exhibits high spatial reuse due to severe path-loss and sparse propagation, making it ideal for short-range communications in extremely dense scenarios such as the industrial settings. Besides, owing to increased degrees of freedom, mMIMO renders extraordinary interference mitigation [5], [6] that enables augmented spectral efficiency and exceptional multiplexing capability, which are desirable features to support the future industrial landscape.

In factories of the future, industrial devices will require two types of information: *shared safety/control messages* (multicast signal) and *private messages* (unicast signals). Such a requirement could be addressed by orthogonal multiple access (OMA) schemes, wherein multicast and unicast signals would be transmitted in disjoint time or frequency resources. Nevertheless, with the anticipated escalation, OMA schemes will struggle to accommodate a large number of devices in orthogonal resources. Thus, *non-orthogonal multiple access* (NOMA) schemes are envisaged as a remedy to cope with the scarcity of radio resources. In particular, NOMA can boost the spectral efficiency by admitting superposed transmissions in the power or code domain. Among the plethora of NOMA variants [7], layered-division multiplexing (LDM) has been recognized as a promising candidate to meet the growing spectrum demands. LDM is a power-domain NOMA scheme capable of conveying multiple layers of information simultaneously while using the same time-frequency resources. By harnessing LDM in industrial settings, multicast and unicast information can be disseminated concurrently without resorting to OMA schemes such as time/frequency-division multiplexing (T/FDM).

Several NOMA schemes have recently been intertwined with mmWave and mMIMO, showing remarkable synergy in many use cases (e.g., [8]–[10]). Also, preliminary studies on the usage of NOMA [11] and mmWave [24] for smart industries have shown favorable results. Based on this evidence, it is expected that by jointly leveraging mmWave, mMIMO and LDM, the stringent requirements of future industrial ecosystems can be fulfilled. However, the synthesis of these technologies poses challenges that require further study when considered in the context of Industry 4.0.

Challenges: The following summarizes relevant aspects that need to be considered in the envisaged industrial landscape.

ISBN 978-3-903176-37-9 © 2021 IFIP

- The maximum number of devices that can be simultaneously served with individual signals is limited by the number of radio frequency (RF) chains at the transmitter (e.g., base station). Hence, with the forecasted rapid escalation of devices in industrial sectors [12], the problem aggravates. Most existing works on beamforming consider sufficient RF chains to serve all devices, thus rendering scheduling unnecessary. *However, as networks densify, scheduling will be pivotal in exerting substantial improvement in the system performance. Thus, considering the cross-layer optimization of beamforming and scheduling is of utmost importance.*
- Multicast and unicast transmissions give rise to conflicting objectives. From the multicast perspective, the transmitter consumes lesser power while the spectral efficiency improves when the devices have correlated channels. From the unicast perspective, we observe the opposite effect, i.e., correlated channels yield low spectral efficiency while demanding higher power. *As a result, selecting a suitable set of devices (i.e., scheduling) in superimposed multicast-unicast LDM systems requires special consideration.*
- Problems dealing with cross-layer optimization of beamforming and scheduling are challenging to solve due its inherent nature of involving integer and continuous variables.

Research problem: Due to safety reasons, the superior-importance multicast signal (e.g., control messages) is not subject to scheduling but is required to be ubiquitous to all IoT devices. Contrastingly, the inferior-importance unicast signals (e.g., software updates) are conveyed to only a specific subset of devices (i.e., scheduling) subject to RF chains availability. As a result, two superimposed beamformers are designed. One beamformer transmits the control signal to all devices. The second beamformer caters a selected subset of devices with private unicast signals, where the selection of devices is inspired by the max-min criterion.

Related work: Beamforming in LDM systems has been studied for *(i)* transmit power minimization [8], [13], [14], *(ii)* energy efficiency improvement [15], *(iii)* joint beamforming and base station clustering [16], [17], *(iv)* sum-rate maximization [18], *(v)* simultaneous wireless information and power transfer (SWIPT) [19], [20], and *(vi)* fairness improvement [21]. *To the best of our knowledge, the cross-layer optimization problem for joint design of beamforming and scheduling in LDM systems has not been studied before. Further, the combination of mmWave, mMIMO and LDM has neither been studied in industrial settings.*

Contributions: Our contributions are the following.

- We formulate a NP-hard problem ($\mathcal{P}$) that jointly optimizes *beamforming* and *scheduling* for multicast-unicast LDM transmissions, where we impose a signal-to-interference-plus-noise ratio (SINR) constraint on the multicast signal to ensure that every IoT device correctly decodes the ubiquitous safety message.
- To solve problem $\mathcal{P}$ we propose BEAMWAVE, which decomposes $\mathcal{P}$ into two problems $\mathcal{S}$ and $\mathcal{D}$. We propose a novel scheduling scheme $\mathcal{S}$ based on new pair-wise metrics, PAWN, ROOK, KING, that we devise to guide the decision. Essentially, these metrics represent the discordance of co-scheduling two devices together. To solve $\mathcal{D}$, we devise an approach based on the convex-concave procedure (CCP). Through simulations, we show that the proposed BEAMWAVE can attain near-optimality when compared to an exhaustive search approach.
- We motivate the need for scheduling in LDM systems, specially when the number of RF chains is insufficient to serve a significantly larger number of devices (which is expected in future industrial settings). In addition, we apply our proposed scheduler $\mathcal{S}$ to T/FDM systems to find the set of devices co-scheduled in the same time or frequency resource. Through simulations, we show the importance of scheduling when compared to more trivial schemes such as random selection.

II. System Model

We assume a system, where a next-generation Node B (gNodeB) serves K devices indexed by $\mathcal{K} = \{1, \cdots, K\}$. The gNodeB transmits a signal composed of two non-orthogonal layers. The *primary layer* is a multicast signal that conveys a shared control message intended for every device $k \in \mathcal{K}$. The *secondary layer* is a composite signal consisting of multiple unicast messages intended for a subset of devices $\mathcal{K}' \subseteq \mathcal{K}$, where $K' = |\mathcal{K}'|$. Thus, K' *dual-layer* devices are catered with simultaneous unicast and multicast transmissions, whereas $K - K'$ *single-layer* devices are served with multicast information only. The gNodeB possesses a precoder (i.e., *transmit beamformer*) consisting of N_{tx} antennas and $N_{\text{tx}}^{\text{RF}} << N_{\text{tx}}$ RF chains. Without loss of generality, we assume that $N_{\text{tx}}^{\text{RF}} = K'$. Besides, each IoT device in the system is equipped with a single RF chain (i.e., $N_{\text{rx}}^{\text{RF}} = 1$) and N_{rx} antennas.

The downlink signal from the gNodeB is denoted by $\mathbf{x} = [\mathbf{B}|\mathbf{m}]\left[\mathbf{s}^T|z\right]^T$. The unicast and multicast precoders are represented by $\mathbf{B} \in \mathbb{C}^{N_{\text{tx}} \times K'}$ and $\mathbf{m} \in \mathbb{C}^{N_{\text{tx}} \times 1}$, respectively. In addition, $\mathbf{s} \in \mathbb{C}^{K' \times 1}$ denotes the unicast symbols for the *dual-layer* devices while $z \in \mathbb{C}$ is the shared multicast symbol intended for all K devices, with $\mathbb{E}\left\{\left[\mathbf{s}^T, z\right]^H \left[\mathbf{s}^T, z\right]\right\} = \mathbf{I}$. More specifically, $\mathbf{B} = \widetilde{\mathbf{B}}\mathbf{U}$ where $\widetilde{\mathbf{B}} = [\mathbf{b}_1, \ldots, \mathbf{b}_K] \in \mathbb{C}^{N_{\text{tx}} \times K}$ and $\mathbf{U} \in \mathbb{B}^{K \times K'}$ is a binary matrix. Also, $\mathbf{s} = \mathbf{U}^T\widetilde{\mathbf{s}}$ where $\widetilde{\mathbf{s}} = [s_1, \ldots, s_K]^T \in \mathbb{C}^{K \times 1}$. Concretely, the matrix $\mathbf{U}$ selects the *dual-layer* devices that will be served with both unicast and multicast signals. Thus, it must hold that $\mathbf{1}^T\mathbf{U}\mathbf{1} = K'$, $\mathbf{U}\mathbf{1} \preccurlyeq \mathbf{1}$ and $\mathbf{U}^T\mathbf{1} \preccurlyeq \mathbf{1}$. As a result, $\mathbf{U}\mathbf{U}^T = \text{diag}\left([\mu_1, \cdots, \mu_K]\right)$ is a square matrix whose k-th diagonal element is 1 when k is a *dual-layer* device (i.e., $\mu_k = \left[\mathbf{U}\mathbf{U}^T\right]_{k,k} = 1$, if $k \in \mathcal{K}'$). Otherwise, $\mu_k = \left[\mathbf{U}\mathbf{U}^T\right]_{k,k} = 0$ when k is a *single-layer* device. Assuming flat fading, the signal received by device $k \in \mathcal{K}$ is given by

$$y_k = \underbrace{\mathbf{w}_k^H\mathbf{H}_k\mathbf{m}z}_{y_k^{\text{M}}\text{: multicast signal}} + \underbrace{\mathbf{w}_k^H\mathbf{H}_k\sum_{j\in\mathcal{K}'}\mathbf{b}_j s_j}_{y_k^{\text{U}}\text{: aggregate unicast signal}} + \underbrace{\mathbf{w}_k^H\mathbf{n}_k}_{\eta_k\text{:noise}}, \tag{1}$$

where $\mathbf{w}_k^H\mathbf{H}_k\sum_{j\in\mathcal{K}\setminus\mathcal{K}'}\mu_j\mathbf{b}_j s_j = 0$ since $\mu_j = 0$, $\forall j \in \mathcal{K}\setminus\mathcal{K}'$. Besides, $\mathbf{w}_k \in \mathbb{C}^{N_{\text{rx}} \times 1}$ represents the combiner (i.e., *receive beamformer*) of the k-th device, $\mathbf{n}_k \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$ symbolizes circularly symmetric Gaussian noise whereas $\mathbf{H}_k \in \mathbb{C}^{N_{\text{rx}} \times N_{\text{tx}}}$

ISBN 978-3-903176-37-9 © 2021 IFIP

denotes the channel between the gNodeB and the k-th device, defined as

$$\mathbf{H}_k = \sqrt{\frac{N_{\rm rx}N_{\rm tx}}{L_k}} \sum_{l=1}^{L_k} \rho_k^{(l)} \mathbf{a}_{\rm rx}\left(\psi_k^{(l)}\right) \mathbf{a}_{\rm tx}\left(\phi_k^{(l)}\right)^H. \quad (2)$$

Here, L_k is the number of paths in $\mathbf{H}_k$, whereas $\psi_k^{(l)}$ and $\phi_k^{(l)}$ represent the angle of arrival (AoA) and angle of departure (AoD) of the l-th path in $\mathbf{H}_k$, respectively. The array vector responses at the k-th device and gNodeB, in the directions of $\psi_k^{(l)}$ and $\phi_k^{(l)}$, are respectively defined as $\mathbf{a}_{\rm rx}\left(\psi_k^{(l)}\right) = \frac{1}{\sqrt{N_{\rm rx}}}\left[1, \cdots, e^{-j(N_{\rm rx}-1)\frac{2\pi}{\lambda}d\cos(\psi_k^{(l)})}\right]^T$ and $\mathbf{a}_{\rm tx}\left(\phi_k^{(l)}\right) = \frac{1}{\sqrt{N_{\rm tx}}}\left[1, \cdots, e^{-j(N_{\rm tx}-1)\frac{2\pi}{\lambda}d\cos(\phi_k^{(l)})}\right]^T$. Also, $\frac{d}{\lambda} = 0.5$ and $\rho_k^{(l)}$ is the complex gain of the l-th path in $\mathbf{H}_k$, which is represented as a random variable following a complex Gaussian distribution $\mathcal{CN}(0,1)$.

Due to the superposed structure of the transmitted signal, successive interference cancellation (SIC) is performed by the *dual-layer* devices in order to extract multicast and unicast information. Every device $k \in \mathcal{K}$ decodes the multicast symbol first by treating the aggregate unicast signal as noise. In addition, if k is a *dual-layer* device (i.e., $k \in \mathcal{K}'$), then the device applies SIC decoding. Essentially, the k-th device reconstructs the multicast signal $y_k^{\rm M}$ using the decoded symbol z, and then subtracts $y_k^{\rm M}$ from y_k. Thereupon, the remaining byproduct consists solely of unicast components ($y_k^{\rm U}$) and noise (η_k), from where the *dual-layer* device can decode its intended symbol s_k. The SINR of the multicast and unicast signals at the k-th device are respectively defined as

$$\text{SINR}_k^{\rm M} = \frac{\left|\mathbf{w}_k^H \mathbf{H}_k \mathbf{m}\right|^2}{\sum_{j\in\mathcal{K}'}\left|\mathbf{w}_k^H \mathbf{H}_k \mathbf{b}_j\right|^2 + \sigma^2\left\|\mathbf{w}_k\right\|_2^2}, \forall k \in \mathcal{K}, \quad (3)$$

$$\text{SINR}_k^{\rm U} = \frac{\left|\mathbf{w}_k^H \mathbf{H}_k \mathbf{b}_k\right|^2}{\sum_{j\neq k, j\in\mathcal{K}'}\left|\mathbf{w}_k^H \mathbf{H}_k \mathbf{b}_j\right|^2 + \sigma^2\left\|\mathbf{w}_k\right\|_2^2}, \forall k \in \mathcal{K}'. \quad (4)$$

III. Problem Formulation

We present a joint formulation that encompasses the optimization of *(i)* scheduling, *(ii)* precoders and *(iii)* combiners,

$$\begin{aligned}
\mathcal{P}: \max_{\substack{\mathbf{W},\mathbf{m},\\ \widetilde{\mathbf{B}},\boldsymbol{\mu}}} \quad & \min_{k\in\mathcal{K}} \frac{\left|\mathbf{w}_k^H \mathbf{H}_k \mathbf{b}_k\right|^2 g(\mu_k)}{\sum_{j\neq k, j\in\mathcal{K}}\left|\mathbf{w}_k^H \mathbf{H}_k \mathbf{b}_j\right|^2 \mu_j + \sigma^2\left\|\mathbf{w}_k\right\|_2^2} \\
\text{s.t.} \quad & \text{C}_1: \frac{\left|\mathbf{w}_k^H \mathbf{H}_k \mathbf{m}\right|^2}{\sum_{j\in\mathcal{K}}\left|\mathbf{w}_k^H \mathbf{H}_k \mathbf{b}_j\right|^2 \mu_j + \sigma^2\left\|\mathbf{w}_k\right\|_2^2} \geq \gamma_{\min}, \forall k \in \mathcal{K}, \\
& \text{C}_2: \sum_{k\in\mathcal{K}} \left\|\mathbf{b}_k\right\|_2^2 \mu_k + \left\|\mathbf{m}\right\|_2^2 \leq P_{\rm tx}, \\
& \text{C}_3: \sum_{k\in\mathcal{K}} \mu_k = K', \\
& \text{C}_4: [\mathbf{w}_k]_l \in \mathcal{W}, l \in \mathcal{L}, \forall k \in \mathcal{K}, \\
& \text{C}_5: \mu_k \in \{0,1\},
\end{aligned}$$

where $g(\chi)$ is defined as

$$g(\chi) = \begin{cases} 1, & \text{if } \chi = 1, \\ \infty, & \text{if } \chi = 0. \end{cases}$$

and $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_K]$, $\widetilde{\mathbf{B}} = [\mathbf{b}_1, \cdots, \mathbf{b}_K]$, $\boldsymbol{\mu} = [\mu_1, \cdots, \mu_K]$.

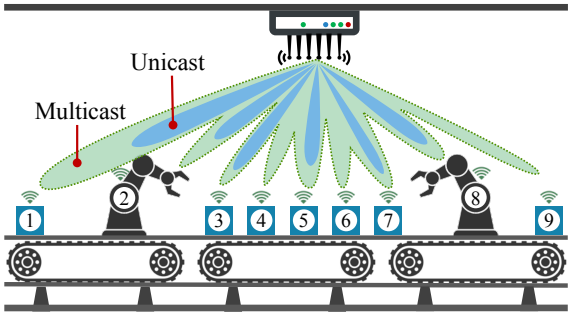


Figure 1: Multicast-unicast LDM system with $K = 9$ devices. *The multicast signal is intended for all devices whereas only a subset of $K' = 6$ devices is served with private unicast signals.*

The objective function of $\mathcal{P}$ aims to find the subset $\mathcal{K}' \subseteq \mathcal{K}$ that maximizes the minimum $\text{SINR}_k^{\rm U}$, $k \in \mathcal{K}'$. The constraint C_1 requires $\text{SINR}_k^{\rm M}$ to be above a threshold $\gamma_{\min}$ for all devices, whereas C_2 limits the transmit power to $P_{\rm tx}$. The constraint C_3 selects K' devices for *dual-layer* transmissions while C_4 enforces beamforming restrictions on the combiners. Specifically, only a small number of $L_{\rm rx}$ constant-modulus phase shifts are admitted for designing the combiners. Every phase shift $[\mathbf{w}_k]_l$ is confined to $\mathcal{W} = \left\{\delta_{\rm rx}, \ldots, \delta_{\rm rx} e^{j\frac{2\pi(L_{\rm rx}-1)}{L_{\rm rx}}}\right\}$, $l \in \mathcal{L} = \{1, \ldots, N_{\rm rx}\}$[1]. Finally, C_5 enforces the Boolean nature of μ_k. We consider limited receive power $P_{\rm rx}$ at each device. Thus, $P_{\rm rx} = \|\mathbf{w}_k\|_2^2 = \sum_{l=1}^{N_{\rm rx}} \left|[\mathbf{w}_k]_l\right|^2 = N_{\rm rx}\delta_{\rm rx}^2$, where $\delta_{\rm rx} = \sqrt{P_{\rm rx}/N_{\rm rx}}$.

To solve $\mathcal{P}$, one possibility is to adopt an exhaustive search approach (XHAUS). This procedure consists in generating every subset of devices of size K' from a total of K, thus yielding $J = \binom{K}{K'}$ possibilities for $\boldsymbol{\mu}$, i.e., $\{\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_J\}$. Then, $\mathcal{P}$ is solved for each of the combinations, i.e., $\left\{\mathcal{P}\left(\mathbf{W}, \mathbf{m}, \widetilde{\mathbf{B}}, \boldsymbol{\mu}_1\right), \cdots, \mathcal{P}\left(\mathbf{W}, \mathbf{m}, \widetilde{\mathbf{B}}, \boldsymbol{\mu}_J\right)\right\}$ and the choice that attains the *max-min* unicast SINR is selected as optimal. While XHAUS yields the best scheduling, it is computationally expensive. Therefore, in Section IV, we propose a scheme, wherein $\boldsymbol{\mu}$ is determined in advance by a novel scheduler. Then, $\mathbf{W}, \mathbf{m}, \widetilde{\mathbf{B}}$ are designed for the resulting selection of devices[2]. Problem $\mathcal{P}$ is illustrated in Fig. 1.

IV. BEAMWAVE: Proposed Scheme

We divide $\mathcal{P}$ into two problems: $\mathcal{S}$ (Section IV-A) and $\mathcal{D}$ (Section IV-B). First, $\mathcal{S}$ finds a subset $\mathcal{K}'$ of *dual-layer* devices, thus rendering the binary scheduling variables available. Subsequently, $\mathcal{D}$ designs the precoder and the combiners.

A. Scheduling

Selecting an optimal subset of *dual-layer* devices $\mathcal{K}'$ that leads to the maximization of the minimum unicast SINR is intrinsically of combinatorial nature. In order to circumvent the exhaustive search, we propose a novel scheduling scheme

[1]Realize that $\mathcal{W}$ consists of equally-distributed phase rotations with magnitude $\delta_{\rm rx}$, where $L_{\rm rx}$ defines the phase resolution.

[2]Notice that even for a given $\boldsymbol{\mu}'$, the problem $\mathcal{P}\left(\mathbf{W}, \mathbf{m}, \widetilde{\mathbf{B}}, \boldsymbol{\mu}'\right)$ is nonconvex and challenging to solve.

ISBN 978-3-903176-37-9 © 2021 IFIP

$\mathcal{S}$, which is based on the minimization of an aggregate pairwise device-specific channel metric. The objective is to find the variables $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ such that $f_{\mathcal{S}}(\boldsymbol{\nu})$ is minimized.

$$\begin{aligned}
\mathcal{S}: \min_{\boldsymbol{\mu},\boldsymbol{\nu}} \quad & f_{\mathcal{S}}(\boldsymbol{\nu}) \triangleq \sum_{j=1}^{K-1}\sum_{l=j+1}^{K} \theta_{j,l}\cdot\nu_{j,l} \\
\text{s.t.} \quad & \mathrm{Q}_1: \mu_j \geq \nu_{j,l}, \forall j<l, \\
& \mathrm{Q}_2: \mu_j+\mu_l \leq 1+\nu_{j,l}, \forall j<l, \\
& \mathrm{Q}_3: \sum_{j=1}^{K}\mu_j = K', \\
& \mathrm{Q}_4: \mu_j \in \{0,1\}, \forall j, \\
& \mathrm{Q}_5: \nu_{j,l} \in \{0,1\}, \forall j<l.
\end{aligned}$$

In particular, $\theta_{j,l}$ denotes a positive metric between two devices $j \in \mathcal{K}$ and $l \in \mathcal{K}$, representing the discordance of co-scheduling the two devices. The auxiliary variable $\nu_{j,l}$, assumes the value of 1, if devices j and l are co-scheduled for *dual-layer* transmissions. Otherwise, $\nu_{j,l}=0$. As defined in $\mathcal{P}$, the variable μ_j denotes with 1 that $j \in \mathcal{K}$ is a *dual-layer* device. The constraints Q_1 and Q_2 have been included in order to bind the two sets of variables, i.e., $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. Specifically, Q_1 states that $\nu_{j,l}$ is upper-bounded by μ_j since $\nu_{j,l}$ can only be 1 when the devices j and l are co-scheduled. Similarly, Q_2 is a lower bound for $\nu_{j,l}$ in terms of μ_j and μ_l. Besides, Q_3 restricts the maximum number of *dual-layer* devices to K'. Constraints $\mathrm{Q}_4-\mathrm{Q}_5$ denote the Boolean nature of the variables.

We denote the solution of $\mathcal{S}$ by $(\boldsymbol{\mu}^\star,\boldsymbol{\nu}^\star)$. In the following, we propose three metrics $\theta_{j,l}$ (i.e., PAWN, ROOK, KING), based on channel correlation and channel energy, which will support the scheduling decision.

CORR: Channel correlation has been extensively used for multiuser unicast scheduling in prior literature (e.g., [22]). Given any two devices j and l, CORR is computed as $\theta_{j,l}=\frac{\left|\mathbf{h}_j^H\mathbf{h}_l\right|}{\|\mathbf{h}_j\|_2\|\mathbf{h}_l\|_2}$, where $\mathbf{h}_j=\mathrm{vec}(\mathbf{H}_j)$. Intuitively, a large value of $0\leq\theta_{j,l}\leq 1$ implies that the two devices have correlated channels and therefore they are prone to generate more interference to each other. CORR has conventionally been used in a greedy manner, where users/devices are sequentially chosen based on the cumulative correlation with respect to the already selected devices. In contrast, herein we use CORR in combination with our proposed scheduler $\mathcal{S}$, thus allowing to find the best set $\mathcal{K}'$ of *dual-layer* devices that renders the least aggregate pair-wise channel correlation in the sense of $f_{\mathcal{S}}(\boldsymbol{\nu})$.

PAWN: We propose this metric as a generalization of CORR, where we compute the channel correlation between all the rows of $\mathbf{H}_j$ and $\mathbf{H}_l$. For two devices j and l, the metric is expressed as $\theta_{j,l}=\sum_{n_1=1}^{N_{\mathrm{rx}}}\sum_{n_2=1}^{N_{\mathrm{rx}}}\frac{1}{N_{\mathrm{rx}}^2}\frac{|\mathbf{H}_j(n_1)\mathbf{H}_l^H(n_2)|}{\|\mathbf{H}_j(n_1)\|_2\|\mathbf{H}_l(n_2)\|_2}$, with $\mathbf{H}_j(n)$ denoting the n-th row of $\mathbf{H}_j$. Note that for the special case of $N_{\mathrm{rx}}=1$, CORR and PAWN are equivalent.

ROOK: We devise this metric as a combination of two components. One of the constituents leverages the channel energy difference between two devices. The second component is the metric PAWN. Thus, ROOK is defined as $\theta_{j,l}=\omega\frac{\left|\|\mathbf{H}_j\|_{\mathrm{F}}^2-\|\mathbf{H}_l\|_{\mathrm{F}}^2\right|}{\|\mathbf{H}_j\|_{\mathrm{F}}^2+\|\mathbf{H}_l\|_{\mathrm{F}}^2}+(1-\omega)\sum_{n_1=1}^{N_{\mathrm{rx}}}\sum_{n_2=1}^{N_{\mathrm{rx}}}\frac{1}{N_{\mathrm{rx}}^2}\frac{|\mathbf{H}_j(n_1)\mathbf{H}_l^H(n_2)|}{\|\mathbf{H}_j(n_1)\|_2\|\mathbf{H}_l(n_2)\|_2}$ with $0\leq\omega\leq 1$. The rationale for this metric is that devices with uncorrelated channel vectors and comparable channel energy are desirable for scheduling.

KING: We also devise this metric as a combination of two components. Specifically, we combine PAWN with the ratio between the channel energy of a device and the largest channel energy among all the devices. Thus, $\theta_{j,l}=\omega\left(\frac{\|\mathbf{H}_{\max}\|_{\mathrm{F}}^2-\|\mathbf{H}_j\|_{\mathrm{F}}^2}{\|\mathbf{H}_{\max}\|_{\mathrm{F}}^2}+\frac{\|\mathbf{H}_{\max}\|_{\mathrm{F}}^2-\|\mathbf{H}_l\|_{\mathrm{F}}^2}{\|\mathbf{H}_{\max}\|_{\mathrm{F}}^2}\right)+(1-\omega)\sum_{n_1=1}^{N_{\mathrm{rx}}}\sum_{n_2=1}^{N_{\mathrm{rx}}}\frac{1}{N_{\mathrm{rx}}^2}\frac{|\mathbf{H}_j(n_1)\mathbf{H}_l^H(n_2)|}{\|\mathbf{H}_j(n_1)\|_2\|\mathbf{H}_l(n_2)\|_2}$, where $\|\mathbf{H}_{\max}\|_{\mathrm{F}}^2=\max_{j\in\mathcal{K}}\|\mathbf{H}_j\|_{\mathrm{F}}^2$ and $0\leq\omega\leq 1$. In contrast to ROOK, this metric measures the relative difference with respect to the largest energy, which compensates for the cases when the devices have uncorrelated channels but commensurable low energy.

Rationale: Intuitively, the aim of $\mathcal{S}$ is to place in $\mathcal{K}\backslash\mathcal{K}'$ (i.e. set of multicast-only devices) those devices that hinder more significantly the maximization of the minimum unicast SINR. This is achieved by $f_{\mathcal{S}}(\boldsymbol{\nu})$, which aims to minimize the total discordance of the co-scheduled devices. Whether such devices *(i)* have highly-correlated channels among themselves or *(ii)* have strongly attenuated channels and thus require high power, by not including them in $\mathcal{K}'$, the devices in $\mathcal{K}'$ can gain the highest profit (i.e., the minimum $\mathrm{SINR}_k^{\mathrm{U}}$, $k\in\mathcal{K}'$ is maximized).

B. Optimization of precoder and combiners

Once the scheduling variables $\boldsymbol{\mu}^\star$ are known, we replace them in $\mathcal{P}\left(\mathbf{W},\mathbf{m},\widetilde{\mathbf{B}},\boldsymbol{\mu}^\star\right)$. Thus, the remaining problem optimizes the unicast and multicast precoders (at the gNodeB) and combiners (at the devices) as shown in

$$\begin{aligned}
\mathcal{D}: \max_{\mathbf{W},\mathbf{m},\mathbf{B},} \quad & f_{\mathcal{D}}(\mathbf{W},\mathbf{B}) \triangleq \min_{k\in\mathcal{K}'}\frac{\left|\mathbf{w}_k^H\mathbf{H}_k\mathbf{b}_k\right|^2}{\sum_{j\neq k,j\in\mathcal{K}'}\left|\mathbf{w}_k^H\mathbf{H}_k\mathbf{b}_j\right|^2+\sigma^2\|\mathbf{w}_k\|_2^2} \\
\text{s.t.} \quad & \frac{\left|\mathbf{w}_k^H\mathbf{H}_k\mathbf{m}\right|^2}{\sum_{j\in\mathcal{K}'}\left|\mathbf{w}_k^H\mathbf{H}_k\mathbf{b}_j\right|^2+\sigma^2\|\mathbf{w}_k\|_2^2}\geq\gamma_{\min},\forall k\in\mathcal{K}, \\
& \sum_{k\in\mathcal{K}'}\|\mathbf{b}_k\|_2^2+\|\mathbf{m}\|_2^2\leq P_{\mathrm{tx}}, \\
& [\mathbf{w}_k]_l\in\mathcal{W}, l\in\mathcal{L},\forall k\in\mathcal{K},.
\end{aligned}$$

where $\mathbf{B}=\widetilde{\mathbf{B}}\mathbf{U}$ and $\mathbf{U}\mathbf{U}^T=\mathrm{diag}(\boldsymbol{\mu})$ as defined in Section II. Due to coupling between $\{\mathbf{b}_k\}_{k\in\mathcal{K}'}$ and $\{\mathbf{w}_k\}_{k=1}^K$, the optimization of $\mathcal{D}$ is challenging. To cope with it, we first design the combiners $\{\mathbf{w}_k\}_{k=1}^K$ based on the channels $\{\mathbf{H}_k\}_{k=1}^K$, which are assumed to be invariant for a few channel uses. Then, we jointly optimize the unicast precoders $\{\mathbf{b}_k\}_{k\in\mathcal{K}'}$ and the multicast precoder $\mathbf{m}$.

B.1 Optimization of combiners $\{\mathbf{w}_k\}_{k=1}^K$

We define $\mathcal{D}_1\triangleq\cup_{k\in\mathcal{K}}\mathcal{D}_{1,k}$, where

$$\mathcal{D}_{1,k}: \max_{\mathbf{w}_k}\left\|\mathbf{w}_k^H\mathbf{H}_k\right\|_2^2 \quad \text{s.t.} \quad \left|[\mathbf{w}_k]_l\right|=\delta_{\mathrm{rx}}, l\in\mathcal{L}. \tag{8}$$

Problem $\mathcal{D}_1$ designs the combiners $\{\mathbf{w}_k\}_{k=1}^K$ for all IoT devices in an independent manner. Therefore, each device can self-optimize its own combiner without need of the gNodeB. This problem admits a close-form solution that can be obtained using the Lagrange multipliers method. Specifically, the solution collapses to the principal eigenvector $\mathbf{r}_{\max}$ of $\mathbf{H}_k\mathbf{H}_k^H$. Then, to enforce the constant-modulus finite-resolution phase

ISBN 978-3-903176-37-9 © 2021 IFIP

shifts, $\mathbf{r}_{\max}$ is projected onto $\mathcal{W}$. Therefore, for the k-th device, $\mathbf{w}_k$ is obtained via $[\mathbf{w}_k]_l = \operatorname{argmax}_{\phi \in \mathcal{W}} \Re\mathfrak{e}\left\{\phi^* [\mathbf{r}_{\max}]_l\right\}$, $\forall l \in \mathcal{L}$. The solution of $\mathcal{D}_1$ is denoted by $\mathbf{W}^\star = [\mathbf{w}_1^\star, \cdots, \mathbf{w}_K^\star]$.

B.2 Optimization of $\{\mathbf{b}_k\}_{k\in\mathcal{K}'}$ *and* $\mathbf{m}$

Assuming that $\mathbf{g}_k = \mathbf{H}_k^H \mathbf{w}_k^\star$, the objective function of $\mathcal{D}$ depends only on $\mathbf{B}$. Note that $f_{\mathcal{D}}(\mathbf{W}^\star, \mathbf{B})$ is the minimum of several SINRs, which can be translated as a constraint as

$$\begin{aligned}
\mathcal{D}_2: \max_{\mathbf{B},\mathbf{m},\alpha} \quad & \alpha \\
\text{s.t.} \quad & \mathrm{R}_1: \frac{\left|\mathbf{g}_k^H \mathbf{b}_k\right|^2}{\sum_{j\neq k, j\in\mathcal{K}'} \left|\mathbf{g}_k^H \mathbf{b}_j\right|^2 + \sigma^2 \left\|\mathbf{w}_k^\star\right\|_2^2} \geq \alpha, \forall k \in \mathcal{K}', \\
& \mathrm{R}_2: \frac{\left|\mathbf{g}_k^H \mathbf{m}\right|^2}{\sum_{j\in\mathcal{K}'} \left|\mathbf{g}_k^H \mathbf{b}_j\right|^2 + \sigma^2 \left\|\mathbf{w}_k^\star\right\|_2^2} \geq \gamma_{\min}, \forall k \in \mathcal{K}, \\
& \mathrm{R}_3: \textstyle\sum_{k\in\mathcal{K}'} \|\mathbf{b}_k\|_2^2 + \|\mathbf{m}\|_2^2 \leq P_{\mathrm{tx}}.
\end{aligned}$$

where $\mathrm{R}_1 - \mathrm{R}_2$ are nonconvex whereas R_3 is convex.

Note that $\mathcal{D}_2$ poses a difficulty in finding a solution as it cannot be addressed by known frameworks in its current form. In the following, we propose a reformulation of the problem that allows tailoring an algorithm to solve it. In particular, we transform $\mathcal{D}_2$ into a difference-of-convex (DC) programming problem, where the objective and/or constraints are convex or DC functions. Then, by harnessing the convex-concave procedure (CCP), a local optimal solution of the resulting DC programming problem can be obtained.

Reformulation: With respect to R_1, if we bound from above the denominator with $\sum_{j\neq k, j\in\mathcal{K}'} \left|\mathbf{g}_k^H \mathbf{b}_j\right|^2 + \sigma^2 \left\|\mathbf{w}_k^\star\right\|_2^2 \leq t_k$ and the numerator from below with $\left|\mathbf{g}_k^H \mathbf{b}_k\right|^2 \geq r_k$, then R_1 can be equivalently rewritten as the intersection of the following constraints

$$\mathrm{R}_1 = \begin{cases}
\mathrm{R}_{1-1}: \underbrace{r_k}_{\text{convex}} - \underbrace{\left|\mathbf{g}_k^H \mathbf{b}_k\right|^2}_{\text{convex}} \leq 0, \forall k \in \mathcal{K}', \\
\mathrm{R}_{1-2}: \underbrace{\sum_{j\neq k, j\in\mathcal{K}'} \left|\mathbf{g}_k^H \mathbf{b}_j\right|^2 + \sigma^2 \left\|\mathbf{w}_k^\star\right\|_2^2 - t_k}_{\text{convex}} \leq 0, \forall k \in \mathcal{K}', \\
\mathrm{R}_{1-3}: \underbrace{\alpha t_k - r_k}_{\text{nonconvex}} \leq 0, \forall k \in \mathcal{K}'.
\end{cases}$$

In addition, we observe that the nonconvex constraint R_{1-3} can be recast as

$$\mathrm{R}_{1-3}: \underbrace{(\alpha + t_k)^2 - 4r_k}_{\text{convex}} - \underbrace{(\alpha - t_k)^2}_{\text{convex}} \leq 0, \forall k \in \mathcal{K}',$$

which stems from the difference of squares: $\frac{(x+y)^2 - (x-y)^2}{4} = xy$. Adopting a similar procedures as for R_1 reformulation, then R_2 can be expressed as,

$$\mathrm{R}_2 = \begin{cases}
\mathrm{R}_{2-1}: \underbrace{p_k}_{\text{convex}} - \underbrace{\left|\mathbf{g}_k^H \mathbf{m}\right|^2}_{\text{convex}} \leq 0, \forall k \in \mathcal{K}, \\
\mathrm{R}_{2-2}: \underbrace{\sum_{j\in\mathcal{K}'} \left|\mathbf{g}_k^H \mathbf{b}_j\right|^2 + \sigma^2 \left\|\mathbf{w}_k^\star\right\|_2^2 - q_k}_{\text{convex}} \leq 0, \forall k \in \mathcal{K}, \\
\mathrm{R}_{2-3}: \underbrace{\gamma_{\min} q_k - p_k}_{\text{convex}} \leq 0, \forall k \in \mathcal{K}.
\end{cases}$$

Observe that R_{1-2}, R_{2-2}, R_{2-3}, R_3 are convex whereas R_{1-1}, R_{1-3}, R_{2-1} are DC functions. Thus, with the transformations above, $\mathcal{D}_2$ is now a DC programming problem.

Solution: Optimization problems that have convex or DC objective/constraints can be efficiently tackled by means of the CCP procedure, which guarantees a stationary solution of the original problem.

> The CCP procedure [23] guarantees a stationary point of a DC programming problem. The main idea of CCP is to iteratively solve a sequence of convex subproblems, each of which is constructed by replacing the concave terms with first-order Taylor approximations. Consider the DC programming problem
>
> $$\begin{aligned} \mathcal{Z}: \max_{\mathbf{z}_1,\mathbf{z}_2} \quad & f(\mathbf{z}_1, \mathbf{z}_2) \\ \text{s.t.} \quad & h_i(\mathbf{z}_1) - g_i(\mathbf{z}_2) \leq, i = 1, \cdots, I, \end{aligned}$$
>
> where $f(\mathbf{z}_1, \mathbf{z}_2)$ is concave in $\mathbf{z}_1$, $\mathbf{z}_2$ whereas $h_i(\mathbf{z}_1)$ and $g_i(\mathbf{z}_2)$ are convex in $\mathbf{z}_1$ and $\mathbf{z}_2$, respectively. To convexify $\mathcal{Z}$, the concave terms, i.e. $-g_i(\mathbf{z}_2)$, are linearized. The resulting convexified DC programming problem is therefore expressed as
>
> $$\begin{aligned} \mathcal{Z}^{(\ell)}: \max_{\mathbf{z}_1,\mathbf{z}_2} \quad & f(\mathbf{z}_1, \mathbf{z}_2) \\ \text{s.t.} \quad & h_i(\mathbf{z}_1) - \tilde{g}_i(\mathbf{z}_2) \leq, i = 1, \cdots, I, \end{aligned}$$
>
> where $\tilde{g}_i(\mathbf{z}_2) = g_i\left(\mathbf{z}_2^{(\ell-1)}\right) + \nabla_{\mathbf{z}_2}^T g_i\left(\mathbf{z}_2^{(\ell-1)}\right)\left(\mathbf{z}_2 - \mathbf{z}_2^{(\ell-1)}\right)$ denotes a linearized version of $g_i(\mathbf{z}_2)$ around a given point $\mathbf{z}_2^{(\ell-1)}$. Since every instance of the resulting problem $\mathcal{Z}^{(\ell)}$ is convex, it can be solved using general-purpose solvers via interior-point methods. The process is repeated iteratively, each time refining the initial point $\mathbf{z}_2^{(\ell-1)} \leftarrow \mathbf{z}_2$ until a stop criterion is satisfied. Let N_{conv} be the maximum number of iterations that $\mathcal{Z}^{(\ell)}$ can be solved, and let $\epsilon \geq 0$ be a small number (e.g., $\epsilon = 0.001$). Thus, the iterative process stops when $\ell = N_{\mathrm{conv}}$ or $\left|f(\mathbf{z}_1, \mathbf{z}_2) - f\left(\mathbf{z}_1^{(\ell-1)}, \mathbf{z}_2^{(\ell-1)}\right)\right| \leq \epsilon$. Further, to guarantee convergence, an initial feasible point (i.e., when $\ell = 0$) is required, which we discuss in Appendix A.

According to the CCP procedure described above, to solve $\mathcal{D}_2$, we need solve the convex problem $\mathcal{D}_2^{(\ell)}$ iteratively until a stop criterion is met. Thus, for a given iteration ℓ, the convex problem $\mathcal{D}_2^{(\ell)}$ is defined as

$$\mathcal{D}_2^{(\ell)}: \max_{\substack{\mathbf{B},\mathbf{m},\alpha \\ \mathbf{r},\mathbf{t},\mathbf{p},\mathbf{q}}} \ \alpha \quad \text{s.t.} \quad \mathrm{R}_{1-1}{}^{(\ell)}, \mathrm{R}_{1-2}, \mathrm{R}_{1-3}{}^{(\ell)}, \mathrm{R}_{2-1}{}^{(\ell)}, \mathrm{R}_{2-2}, \mathrm{R}_{2-3}, \mathrm{R}_3.$$

$$\mathrm{R}_{1-1}{}^{(\ell)}: r_k + \left|\mathbf{g}_k^H \mathbf{b}_k^{(l-1)}\right|^2 - 2\Re\mathfrak{e}\left\{\mathbf{b}_k^{(l-1)\,H} \mathbf{g}_k \mathbf{g}_k^H \mathbf{b}_k\right\} \leq 0, \forall k \in \mathcal{K}',$$

$$\mathrm{R}_{1-3}{}^{(\ell)}: (\alpha + t_k)^2 - 4r_k + \left(\alpha^{(l-1)} - t_k^{(l-1)}\right)^2 - 2\left(\alpha^{(l-1)} - t_k^{(l-1)}\right)(\alpha - t_k) \leq 0,$$

$$\mathrm{R}_{2-1}{}^{(\ell)}: p_k + \left|\mathbf{g}_k^H \mathbf{m}^{(l-1)}\right|^2 - 2\Re\mathfrak{e}\left\{\mathbf{m}^{(l-1)\,H} \mathbf{g}_k \mathbf{g}_k^H \mathbf{m}\right\} \leq 0, \forall k \in \mathcal{K}.$$

At the completion of each iteration ℓ, the obtained solutions $\mathbf{B}$, $\mathbf{m}$, α, $\mathbf{t}$ are passed to $\mathbf{B}^{(\ell)}$, $\mathbf{m}^{(\ell)}$, $\alpha^{(\ell)}$, $\mathbf{t}^{(\ell)}$, which are used as the new initializations for the subsequent iteration $\ell + 1$. The solution of this stage is $\mathbf{B}^\star$ and $\mathbf{m}^\star$. For completeness, we summarize in *Algorithm* the complete optimization procedure of $\mathcal{S}$ and $\mathcal{D}$.

ISBN 978-3-903176-37-9 © 2021 IFIP

Algorithm: BEAMWAVE optimization

Input: $\{\mathbf{H}_k\}_{k=1}^{K}$, $\gamma_{\min}$, N_{conv}, ϵ

Execute:

1: Find $\boldsymbol{\mu}^\star$ by solving the scheduling problem $\mathcal{S}$.
2: Design the combiners $\{\mathbf{w}_k^\star\}_{k=1}^{K}$ for all devices by solving problem $\mathcal{D}_{1,k}$, $\forall k \in \mathcal{K}$.
3: Design the multicast precoder $\mathbf{m}^\star$ and the unicast precoders $\{\mathbf{b}_k^\star\}_{k\in\mathcal{K}'}$ by solving $\mathcal{D}_2^{(\ell)}$.

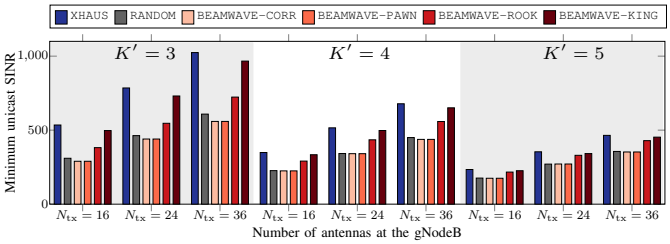


Figure 2: Achievable minimum unicast SINR for varying N_{tx} at the gNodeB.

V. Simulation Results

Throughout the simulations, we consider the geometric channel model defined in (2), with $L = L_1 = \cdots = L_K = 3$ propagation paths. This assumption is compliant with the results of a measurement campaign in an industrial environment [24], where the number of propagation paths is usually between 1 to 3. The angles of arrival are uniformly distributed as $\psi_k^{(l)} \in [-\pi;\pi]$ whereas the angles of departure are distributed as $\phi_k^{(l)} \in [-\pi/3;\pi/3]$. The power assigned to the combiners is $P_{\text{rx}} = 0$ dBm, the noise power is $\sigma^2 = 10$ dBm, and the multicast QoS requirement is $\gamma_{\min} = 4$ (~ 6 dB). Also, $\omega = 0.5$, $N_{\text{conv}} = 20$ and $\epsilon = 0.001$. The results in this section show the average performance over 100 simulations. For the selected settings, in all the channel realizations, we have obtained feasible solutions. To solve the optimization problems, we have used CVX. Specifically, CVX and GUROBI were used to solve the integer linear program $\mathcal{S}$. The convex problem $\mathcal{D}_2^{(\ell)}$ was solved by means of CVX and SDPT3. In the following, we examine scenarios, in which we evaluate the performance of BEAMWAVE.

A. Minimum unicast SINR for various N_{tx}

Fig. 2 depicts the impact of different N_{tx} configurations on the minimum unicast SINR when the total number of devices in the system is $K = 6$ and the number of *dual-layer* devices $K' = \{3, 4, 5\}$ varies. In this case, we have assumed that the IoT devices are equipped with a single antenna, i.e., $N_{\text{rx}} = 1$ and the gNodeB can transmit with a maximum power $P_{\text{tx}} = 35$ dBm.

As a general trend, we observe that increasing the number of *dual-layer* devices K' decreases the minimum unicast SINR. This occurs because the limited power P_{tx} at the gNodeB is divided into a greater number of scheduled devices, thus reducing the individual allocation of power for each *dual-layer* device. Also, serving more *dual-layer* devices translates to producing more interference, thus impacting the SINR. On the contrary, increasing N_{tx} improves the minimum unicast SINR. Essentially, a larger N_{tx} reduces the beamwidth that can be produced by the antenna array at the gNodeB, thus allowing to form more directional transmissions with reduced interference.

Another general trend in Fig. 2 is that XHAUS (exhaustive search) exhibits the highest performance in all configurations as it schedules the optimum subset of K' *dual-layer* devices. By leveraging the channel correlation, BEAMWAVE-CORR[3] only performs slightly better than RANDOM. Thus, scheduling decisions based solely on the channel correlation are insufficient to devise an optimal scheduler for LDM systems. On the contrary, BEAMWAVE-ROOK and BEAMWAVE-KING, which additionally include channel energy information, clearly outperform RANDOM. These two schemes achieve up to 60.38% and 77.68% higher SINR, respectively, compared to RANDOM. Noteworthily, throughout all the results in Fig. 2, BEAMWAVE-ROOK and BEAMWAVE-KING perform at worst 30.4% and 14.13% below XHAUS, respectively.

B. Minimum unicast SINR for various N_{rx}

Fig. 3 shows the impact of varying N_{rx} and L_{rx} on the minimum unicast SINR when $K = 6$, $K' = 5$, $N_{\text{tx}} = 16$, and $P_{\text{tx}} = 35$ dBm. In this setting, the IoT devices have a single RF chain that is connected to N_{rx} antennas. As a result, the devices are not capable of implementing any type of linear processing for interference mitigation but can perform constrained beamsteering due to constraint C_4 in $\mathcal{P}$.

In all subfigures in Fig. 3, we observe that the minimum unicast SINR improves as the number of receive antennas increases. With larger N_{rx}, the devices can shape more directional reception patterns to mitigate undesired signals. In particular, up to 60% gain can be achieved with $L_{\text{rx}} = 4$ when varying N_{tx} from 1 to 2. Also, since augmenting L_{rx} results in higher-resolution phase shifts, we observe performance improvement through Fig. 3a to Fig. 3d. In particular, gains up to 16.00%, 30.70% and 49.47% are achieved when increasing L_{rx} from 2 to 4, 4 to 8 and 8 to 16, respectively.

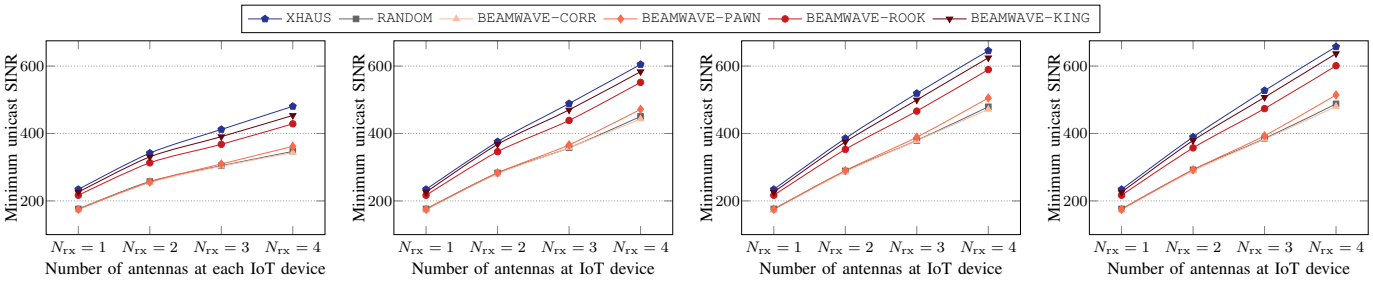
By comparing the performance of the proposed scheduling schemes under all assessed settings, the scheme that attains superior performance is BEAMWAVE-KING. In particular, BEAMWAVE-KING is outperformed by at most 5.60% when compared to the optimal highly complex XHAUS.

C. Spectral efficiency

In this scenario we consider $N_{\text{rx}} = 1$, $N_{\text{tx}} = 32$, , $P_{\text{tx}} = 45$ dBm and a varying number of devices $K = \{8, 12, 16, 20, 24, 28, 32, 36\}$. In particular, the number of scheduled *dual-layer* devices changes according to $K' = K/4$. In Fig. 4, we show the unicast spectral efficiency (SE) attained by BEAMWAVE. Due to the exponential growth in the number of scheduling combinations, the results with XHAUS are not

[3]As mentioned in Section IV-A, when $N_{\text{rx}} = 1$, PAWN and CORR result in the same value. For this reason, we observe that BEAMWAVE-CORR and BEAMWAVE-PAWN attain the same performance.

ISBN 978-3-903176-37-9 © 2021 IFIP



(a) $L_{\mathrm{rx}} = 2$ (b) $L_{\mathrm{rx}} = 4$ (c) $L_{\mathrm{rx}} = 8$ (d) $L_{\mathrm{rx}} = 16$

Figure 3: Achievable minimum SINR for varying N_{rx} and L_{rx} at each IoT device.

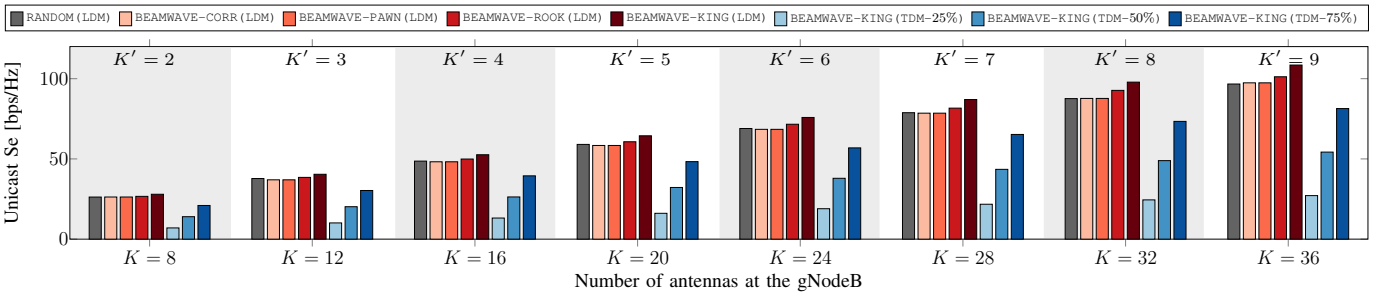


Figure 4: Spectral efficiency performance for varying K and $K' = K/4$.

presented in this scenario. However, BEAMWAVE-KING is taken as reference as it was shown in previous scenarios that its performance is at most 14.13% below the optimality of XHAUS. Further, we also use our proposed scheduler with T/FDM systems.

Note that while RANDOM scheduling performs as equally well as BEAMWAVE for small K' (since the generated interference is low), we observe that when K' is large (e.g., $K' = 16$) there is a significant performance gap. This shows that scheduling exerts a critical task, specially in LDM systems which generate additional inter-layer interference between unicast and multicast signals. Besides, we observe that LDM outperforms TDM, where the time allotted for unicast transmissions is 25%, 50% and 75% of the total available[4]. The remaining time is used for transmitting the multicast signal. Specifically, in the TDM case, we have also used BEAMWAVE to make the selection of unicast devices that yields the *max-min* SINR.

D. Computational complexity

In Table I, we show the complexity of the benchmarked schemes. In particular, $\mathcal{C}_{\mathcal{S}}$ is the complexity of the proposed scheduler, where $M = \frac{K(K+1)}{2}$ is the number of $0-1$ variables and $C = K^2 - K + 1$ is the number of constraints. As a reference, we have used the runtime of Vaidya's algorithm for the linear program, which GUROBI solves via the branch and bound (BnB) procedure. The complexity $\mathcal{C}_{\mathcal{D}_{1,k}}$ stems from the singular value decomposition (SVD) used to obtain the principal eigenvector, as described in Section IV-B1. Also, $\mathcal{C}_{\mathcal{D}_2}$ is derived based on the complexity required by interior point methods. Finally, $\mathcal{C}_{\texttt{XHAUS}}$, $\mathcal{C}_{\texttt{BEAMWAVE}}$ and $\mathcal{C}_{\texttt{RANDOM}}$ denote the overall complexities of the schemes XHAUS, BEAMWAVE and RANDOM respectively.

Table I: Computational complexity

Notation	Complexity
$\mathcal{C}_{\mathcal{S}}$	$\mathcal{O}\left(2^M (M+C)^{1.5} M\right)$
$\mathcal{C}_{\mathcal{D}_{1,k}}$	$\mathcal{O}\left(N_{\mathrm{rx}}^3\right)$
$\mathcal{C}_{\mathcal{D}_2}$	$N_{\mathrm{conv}} \cdot \mathcal{O}\left(\left(N_{\mathrm{tx}} K' (K+K')\right)^{3.5}\right)$
$\mathcal{C}_{\texttt{RANDOM}}$	$\mathcal{C}_{\mathcal{D}_2} + K \cdot \mathcal{C}_{\mathcal{D}_{1,k}}$
$\mathcal{C}_{\texttt{BEAMWAVE}}$	$\mathcal{C}_{\mathcal{D}_2} + K \cdot \mathcal{C}_{\mathcal{D}_{1,k}} + \mathcal{C}_{\mathcal{S}}$
$\mathcal{C}_{\texttt{XHAUS}}$	$\binom{K}{K'} \cdot \mathcal{C}_{\mathcal{D}_2} + K \cdot \mathcal{C}_{\mathcal{D}_{1,k}}$

VI. Conclusions

In this paper we investigated the cross-layer optimization of beamforming and scheduling for mmWave LDM systems, aiming to support future Industry 4.0 scenarios. In particular, through the adoption of LDM, multiple signal layers can be transmitted simultaneously using the same radio resources. For smart factory settings, we assumed that a superior-importance safety/control multicast message is required to be ubiquitous to all the devices in the system. In addition, due to insufficient RF chains, inferior-importance private unicast information is simultaneously transmitted to a selected group of scheduled devices with the aim of maximizing the minimum SINR. Due to NP-hardness of the problem, we proposed

[4]In the TDM case, the IoT devices are served in two time windows. In the first window, with duration T_m, all IoT devices in $\mathcal{K}$ are served with the multicast control signal. In the second window, with duration T_u, a subset of devices $\mathcal{K}'$ are served with unicast signals (e.g., software updates), such that $T_m + T_u = 1$. In our simulations, we have varied $T_u = \{0.25, 0.50, 0.75\}$.

ISBN 978-3-903176-37-9 © 2021 IFIP

BEAMWAVE which partitions the problem into *(i)* beamforming and *(ii)* scheduling. For device scheduling, we proposed a novel formulation, where we devised three metrics based on channel features, namely PAWN, ROOK, and KING to guide the selection decision. Further, we designed a precoder (i.e., transmit beamformer) with remarkable performance adopting the convex-concave procedure. We showed that our proposed scheme attains high spectral efficiency and outperforms orthogonal multiplexing schemes such as T/FDM.

ACKNOWLEDGMENT

The research is in part funded by the Deutsche Forschungsgemeinschaft (DFG) within the B5G-Cell project in SFB 1053 MAKI and by the LOEWE initiative (Hesse, Germany) within the emergenCITY center.

REFERENCES

[1] X. Chen, *Massive Access for Cellular Internet of Things Theory and Technique*. Berlin, Germany: Springer, 2019.

[2] L. F. Abanto-Leon, M. Hollick, and G. H. Sim, "HydraWave: Multi-group Multicast Hybrid Precoding and Low-Latency Scheduling for Ubiquitous Industry 4.0 mmWave Communications," in *IEEE WoWMoM*, 2020, pp. 98–107.

[3] E. Björnson, L. V. der Perre, S. Buzzi, and E. G. Larsson, "Massive MIMO in Sub-6 GHz and mmWave: Physical, Practical, and Use-Case Differences," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 100–108, April 2019.

[4] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. F. de Almeida, "Energy Efficiency of mmWave Massive MIMO Precoding With Low-Resolution DACs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 298–312, April 2018.

[5] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for Next Generation Wireless Systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.

[6] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO Has Unlimited Capacity," *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 574–590, January 2018.

[7] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal Multiple Access for 5G and Beyond," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2347–2381, December 2017.

[8] J. Zhao, O. Simeone, D. Gunduz, and D. Gomez-Barquero, "Non-Orthogonal Unicast and Broadcast Transmission via Joint Beamforming and LDM in Cellular Networks," in *IEEE GLOBECOM*, December 2016, pp. 1–6.

[9] X. Chen, Z. Zhang, C. Zhong, R. Jia, and D. W. K. Ng, "Fully Non-Orthogonal Communication for Massive Access," *IEEE Transactions on Communications*, vol. 66, no. 4, pp. 1717–1731, April 2018.

[10] Z. Ding, P. Fan, and H. V. Poor, "Random Beamforming in Millimeter-Wave NOMA Networks," *IEEE Access*, vol. 5, pp. 7667–7681, February 2017.

[11] 3GPP, "Technical Specification Group Services and System Aspects," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 21.916, 03 2020, version 0.4.0. [Online]. Available: www.3gpp.org/ftp/Specs/archive/21_series/21.916

[12] M. R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, "Internet of Things in the 5G Era: Enablers, Architecture, and Business Models," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 510–527, March 2016.

[13] J. Zhao, D. Gündüz, O. Simeone, and D. Gómez-Barquero, "Non-Orthogonal Unicast and Broadcast Transmission via Joint Beamforming and LDM in Cellular Networks," *IEEE Transactions on Broadcasting*, vol. 66, no. 2, pp. 216–228, June 2020.

[14] Y. Liu, C. Lu, M. Tao, and J. Wu, "Joint Multicast and Unicast Beamforming for the MISO Downlink Interference Channel," in *IEEE SPAWC*, July 2017, pp. 1–5.

[15] Y. Li, M. Xia, and Y. Wu, "Energy-Efficient Precoding for Non-Orthogonal Multicast and Unicast Transmission via First-Order Algorithm," *IEEE Transactions on Wireless Communications*, vol. 18, no. 9, pp. 4590–4604, September 2019.

[16] E. Chen and M. Tao, "Backhaul-Constrained Joint Beamforming for Non-Orthogonal Multicast and Unicast Transmission," in *IEEE GLOBECOM*, December 2017, pp. 1–6.

[17] E. Chen, M. Tao, and Y. Liu, "Joint Base Station Clustering and Beamforming for Non-Orthogonal Multicast and Unicast Transmission With Backhaul Constraints," *IEEE Transactions on Wireless Communications*, vol. 17, no. 9, pp. 6265–6279, September 2018.

[18] J. Wang, H. Xu, B. Zhu, L. Fan, and A. Zhou, "Hybrid Beamforming Design for mmWave Joint Unicast and Multicast Transmission," *IEEE Communications Letters*, vol. 22, no. 10, pp. 2012–2015, October 2018.

[19] W. Hao, G. Sun, Z. Chu, P. Xiao, Z. Zhu, S. Yang, and R. Tafazolli, "Beamforming Design in SWIPT-Based Joint Multicast-Unicast mmWave Massive MIMO With Lens-Antenna Array," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1124–1128, August 2019.

[20] W. Hao, G. Sun, F. Zhou, D. Mi, J. Shi, P. Xiao, and V. C. M. Leung, "Energy-Efficient Hybrid Precoding Design for Integrated Multicast-Unicast Millimeter Wave Communications With SWIPT," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10 956–10 968, November 2019.

[21] L. F. Abanto-Leon and G. H. Sim, "Fairness-Aware Hybrid Precoding for mmWave NOMA Unicast/Multicast Transmissions in Industrial IoT," in *IEEE ICC*, June 2020, pp. 1–7.

[22] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-Antenna Downlink Channels with Limited Feedback and User Selection," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1478–1491, September 2007.

[23] T. Lipp and S. Boyd, "Variations and Extension of the Convex–Concave Procedure," *Optimization and Engineering*, vol. 17, no. 2, pp. 263–287, June 2016.

[24] A. Loch, C. Cano, G. H. Sim, A. Asadi, and X. Vilajosana, "A Channel Measurement Campaign for mmWave Communication in Industrial Settings," *IEEE Transactions on Wireless Communications*, September 2020.

[25] W. Hao, M. Zeng, Z. Chu, and S. Yang, "Energy-Efficient Power Allocation in Millimeter Wave Massive MIMO With Non-Orthogonal Multiple Access," *IEEE Wireless Communications Letters*, vol. 6, no. 6, pp. 782–785, December 2017.

APPENDIX A
INITIAL FEASIBLE POINT FOR $\mathcal{D}_2$

In order to find $\mathbf{B}^{(0)}$, $\mathbf{m}^{(0)}$, $\alpha^{(0)}$, $\mathbf{t}^{(0)}$ we proceed as follows. First, let us define a as the power of the multicast precoder $\mathbf{m}$, such that $\mathbf{m} = \sqrt{a}\widehat{\mathbf{m}}$, $\|\widehat{\mathbf{m}}\|_2^2 = 1$. Similarly, we define a_k as the power of the unicast precoder $\mathbf{b}_k$, $k \in \mathcal{K}'$ such that $\mathbf{b}_k = \sqrt{a}_k\widehat{\mathbf{b}}_k$, $\left\|\widehat{\mathbf{b}}_k\right\|_2^2 = 1$. Now, we let $\left\{\widehat{\mathbf{b}}_k\right\}_{k\in\mathcal{K}'}$ be the zero-forcing precoders [25]. On the other hand, we let $\widehat{\mathbf{m}}$ be the principal eigenvector of the aggregate channels of all users. Thus, we define

$$\begin{aligned}\mathcal{D}_2^{\text{ini}}: \min_{\{a_k\}_{k\in\mathcal{K}'},a} \quad & \sum_{k\in\mathcal{K}'}\sum_{j\neq k,j\in\mathcal{K}'} a_j\left|h_{k,j}\right|^2 \\ \text{s.t.} \quad & \frac{a\left|h_k\right|^2}{\sum_{j\in\mathcal{K}'} a_j\left|h_{k,j}\right|^2+\sigma^2\left\|\mathbf{w}_k^\star\right\|_2^2} \geq \gamma_{\min}, \forall k\in\mathcal{K}, \\ & \textstyle\sum_{k\in\mathcal{K}'} a_k\left\|\widehat{\mathbf{b}}_k\right\|_2^2 + a\left\|\widehat{\mathbf{m}}\right\|_2^2 \leq P_{\text{tx}},\end{aligned}$$

where $h_{k,j} = \mathbf{g}_k^H\widehat{\mathbf{b}}_j$ and $h_k = \mathbf{g}_k^H\mathbf{m}$. Note that $\mathcal{D}_2^{\text{ini}}$ is a linear programming problem. Also, observe that any feasible solution for $\mathcal{D}_2^{\text{ini}}$ will be feasible for $\mathcal{D}_2$. In particular, the objective function of $\mathcal{D}_2^{\text{ini}}$ minimizes the total unicast interference perceived by all IoT devices (i.e., sum of all terms in the denominator of R_1 in $\mathcal{D}_2$). Once $\mathcal{D}_2^{\text{ini}}$ is solved, we obtain a solution $\left(\{a_k^\star\}_{k\in\mathcal{K}'}, a^\star\right)$. Harnessing this outcome, we obtain the initial feasible points for $\mathcal{D}_2^{(0)}$ by defining $\mathbf{b}_k^{(0)} = a_k^\star\widehat{\mathbf{b}}_k$, $\mathbf{m}^{(0)} = a^\star\widehat{\mathbf{m}}$, $t_k^{(0)} = \sum_{j\neq k,j\in\mathcal{K}'} a_j^\star\left|h_{k,j}\right|^2+\sigma^2\left\|\mathbf{w}_k^\star\right\|_2^2$, $\alpha^{(0)} = \min_{k\in\mathcal{K}'} \frac{a_k^\star|h_{k,k}|^2}{\sum_{j\neq k,j\in\mathcal{K}'} a_j^\star|h_{k,j}|^2+\sigma^2\|\mathbf{w}_k^\star\|_2^2}$.

ISBN 978-3-903176-37-9 © 2021 IFIP