Computer Networks 105 (2016) 150–165

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Performance modeling of next-generation WiFi networks☆

Antonios Michaloliakos*, Ryan Rogalin, Yonglong Zhang, Konstantinos Psounis, Giuseppe Caire

University of Southern California, Los Angeles, CA 90089 United States

ARTICLE INFO

Article history:
Received 8 January 2016
Revised 20 April 2016
Accepted 30 May 2016
Available online 31 May 2016

Keywords:
MIMO
Coordinated MU-MIMO
CSMA
Modeling
Analytic model

ABSTRACT

The industry is satisfying the increasing demand for wireless bandwidth by densely deploying a large number of access points which are centrally managed, e.g. enterprise WiFi networks deployed in university campuses, companies, airports etc. This "small cell" architecture is gaining traction in the cellular world as well, as witnessed by the direction in which 4G+ and 5G standardization is moving. Prior academic work in analyzing such large-scale wireless networks either uses oversimplified models for the physical layer, or ignores other important, real-world aspects of the problem, like MAC layer considerations, topology characteristics, and protocol overhead. On the other hand, for deployment purposes the industry is using on-site surveys and simulation tools which do not scale, cannot efficiently optimize the design of such a network, and do not explain why one design choice is better than another.

In this paper we introduce an analytical model which combines the realism and practicality of industrial simulation tools with the ability to scale, analyze the effect of various design parameters, and optimize the performance of real-world deployments. The model takes into account all central system parameters, including channelization, power allocation, user scheduling, load balancing, MAC, advanced PHY techniques (single and multi user MIMO as well as cooperative transmission from multiple access points), topological characteristics and protocol overhead. The accuracy of the model is verified via extensive simulations and the model is used to study a wide range of real world scenarios, providing design guidelines on the effect of various design parameters on performance.

1. Introduction

Modern wireless devices such as tablets and smartphones are pushing the demand for higher and higher wireless data rates while causing significant stress to existing wireless networks. While successive generations of wireless standards achieve continuous improvement, it is the general understanding of both academic research and the industry that a significant increase in wireless traffic demand can be met only by a dramatically denser spectrum reuse, i.e., by deploying more base stations/access points per square kilometer, coupled with advanced physical (PHY) layer techniques to reduce inter-cell interference.

Enterprise WiFi networks have been deployed following this paradigm for years. As a matter of fact, the density of access points (APs) has increased to a point where inter-cell interference is canceling any additional gains from even denser deployments. At the same time, advanced physical layer techniques have been incorporated into the standards, most notably single-user MIMO in 802.11n and multi-user MIMO in 802.11ac. Cellular networks, unable to satisfy the bandwidth demand of data plans, resort to WiFi offloading, i.e. they deploy WiFi networks to offload traffic from the cellular network. Future cellular network architectures will most likely follow a similar pattern, that is, they will consist of many small cells densely deployed and use advanced physical layer techniques, e.g. massive MIMO.

The industry has responded to the need to efficiently manage such networks with tools that are mostly based on on-site measurements, simulations, and over-simplistic analytical models. Based on the available public information about such tools in the enterprise WiFi market [1,2], these tools perform three main operations: (i) user load balancing among APs, (ii) interference management between APs by channel allocation and power control, and (iii) optimization of the Clear Channel Assessment (CCA) CSMA threshold to allow for concurrent transmissions which can tolerate interference from nearby APs. While such network management tools have increased the performance of enterprise WiFi networks, they do not scale well, cannot be used to efficiently optimize the

☆ This work is supported by the National Science Foundation, under the grant ECCS-1444060.

* Corresponding author.
E-mail address: michalol@usc.edu (A. Michaloliakos).

http://dx.doi.org/10.1016/j.comnet.2016.05.024
1389-1286/

system, and do not incorporate the effects of advanced physical layer techniques.

In this paper, we introduce an accurate and practical analytical model which takes into consideration all the important parameters affecting the performance of present and future wireless networks, and can be efficiently used in real-world setups. Specifically, we model and investigate the performance impact of physical layer features such as channelization, power allocation, topological characteristics (e.g. user density and AP distribution in various buildings/structures) and physical layer techniques like single-user (SU), multi-user (MU), and coordinated MU-MIMO [3,4] (where a number of remote APs coordinate and transmit concurrently and jointly to multiple users). Additionally, we model the performance impact of MAC and higher layer features such as user-AP association, MAC parametrization and adaptive coding/modulation.

Our main contributions are the following. We introduce and validate through simulations the first (to the best of our knowledge) analytical model which can be applied in real-world scenarios while taking into account all the important design parameters in the PHY, MAC and higher layers. Second, we include in the model next generation wireless networking technologies such as MU-MIMO and coordinated MU-MIMO for which there is currently no clear understanding of large scale network performance. Third, we apply the model to a variety of real world scenarios, including conference halls, office buildings, open spaces, large stadiums, etc., and study a number of important phenomena (see Section 5 for a detailed discussion). It is important to note that as of today there are no existing 3rd party simulators which support advanced PHY layer techniques like MU-MIMO and coordinated MU-MIMO, and no software defined radio testbeds supporting both MU-MIMO and WiFi, neither in the industry nor in academia.[1] Thus, our analytical model is, at the moment, the only way to study the performance of large scale WiFi networks utilizing advanced PHY layer techniques.

The outline of the paper is as follows. In the next section we discuss related work. Section 3 motivates and describes a unified analytic treatment of wireless network deployments consisting of an analytical model for various current and next generation PHY layer schemes as well as for the CSMA MAC. The validation and limitations of our model are studied via extensive simulations for tractable scenarios of interest in Section 4. Finally Section 5 applies the analytic model in various deployment scenarios of interest like conference halls, open and closed office floor plans and stadiums.

2. Related work

Traditional 802.11 analysis. When it comes to analyzing wireless network deployments, the wireless networking community has focused on the MAC layer and has generally ignored the PHY layer. Specifically, early work of Bianchi [5] on 802.11 MAC layer proposed an analytical model to analyze CSMA/CA overhead and performance. Meanwhile in [6] the authors investigated the performance of exponential backoff mechanism in terms of throughput and delay. In [7–10] similar Markov Chain models of CSMA/CA have been developed and employed to develop algorithms optimizing various performance metrics. Stochastic geometry approaches have been also used to model CSMA networks, using appropriate point processes (see [11,12] and references therein). Finally, in our own previous works [13,14], a full analytic model for computing the achievable rate region of CSMA in multi hop wireless networks has been presented. These works are mostly based on pure upper layer modeling and do not take the advances of physical layer into account.

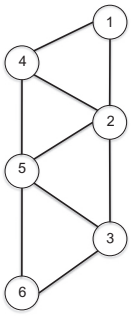
Random matrix theory approaches. There are two main trends the wireless communication literature follows in analyzing wireless network deployments. The first makes use of techniques based on random matrix theory to extract performance measures (such as achievable rates, SINRs etc.) that are used with combinatorial and convex optimization methods to solve problems that appear in multi-cell wireless networks. Such problems include but are not limited to: finding the optimal achievable rates under power control, massive MIMO system asymptotics, base station cooperation towards a coordinated MU-MIMO solution and more as can be found in prior work of ours [15–17] and others [18–23]. These works exhibit similarities to our work on the basis that we also adopt random matrix theory results and techniques to come up with analytically tractable PHY layer performance models. On the other hand our work differs from the aforementioned approaches because we use the analytically tractable models in combination with an accurate MAC layer approximation, to compare the various forthcoming WiFi technologies, and examine their applicability and efficiency under various use cases and MAC layer parameters.

Stochastic geometry approaches. The second approach is based on stochastic geometry and focuses on the random placement of APs and users according to some stochastic point process (see for example [24,25] and references therein). Most of these works do not consider advanced interference management schemes at the PHY layer since they introduce statistical dependence between the nodes, and this would break the independence on which most of these results are based. Recent progress has been made to model more advanced PHY schemes such as MU-MIMO [26,27], AP cooperation [28,29] and multi-cell coordination [30,31] using a stochastic geometry analysis, however these works are limited to non-coherent AP cooperation (also called single-user joint transmission) [29] or pairwise only coherent joint transmission [28]. In other words, the performance of full coherent coordination of a large number of APs serving concurrently users is not being captured in these works, although it is one of the most promising forthcoming PHY layer technologies. In contrast, our approach applies also to large coherent cooperative clusters of APs and that can serve multiple users simultaneously. Moreover, although stochastic geometry methodologies can give important insights for the average behavior of a network, they cannot encompass a specific AP/user placement topology since they analyze performance over network ensembles with certain distributions and densities.

Advanced PHY layer comparisons. Recent papers have taken steps towards comparing advanced PHY layer techniques, but lack the analytic simplicity of the model we proposed in this work. In [32] the authors compare a traditional Wi-Fi network with advanced cooperative cellular networks (coordinated MU-MIMO), they resort however to standard Monte Carlo simulation based approaches and fail to address the adoption of MU-MIMO in future Wi-Fi networks. In [33] the authors compare dense station deployments against coordinated approaches but only incorporate PHY layer characteristic and no MAC is included in the model. Moreover, they also rely on Monte Carlo Simulations over both the topology and the channels. In contrast, we propose a simple analytic approach based on random matrix theory that incorporates PHY layer advances in a single PHY/MAC layer model and accounts for a variety of network design choices. This approach allows for very accurate deterministic approximations of the users' peak link rates for a given geometry of the network.

Industry tools. Lastly, there are a number of tools that the industry currently uses for wireless network deployment guidance. For example, Fluke Networks has developed a product [34] which creates a model for the wireless environment so that an

[1] MU-MIMO enabled WiFi chipsets are expected to become available soon, but such chipsets can't be used in a coordinated MU-MIMO setup and come with the usual limitations of non-programmable hardware when it comes to experimentation.



(a) Interference graph

State	Limiting Probability
000000	$1/(1+6\rho+6\rho^2)$
100000	$\rho/(1+6\rho+6\rho^2)$
010000	$\rho/(1+6\rho+6\rho^2)$
⋮	⋮
000001	$\rho/(1+6\rho+6\rho^2)$
101000	$\rho^2/(1+6\rho+6\rho^2)$
100010	$\rho^2/(1+6\rho+6\rho^2)$
100001	$\rho^2/(1+6\rho+6\rho^2)$
010001	$\rho^2/(1+6\rho+6\rho^2)$
001100	$\rho^2/(1+6\rho+6\rho^2)$
000101	$\rho^2/(1+6\rho+6\rho^2)$

(b) Limiting probabilities of the CSMA CTMC

Fig. 1. Interference graph and limiting probabilities of the corresponding CSMA CTMC.

administrator may simulate and predict the performance. These tools mostly resort to real time on-site surveys which do not scale and do not incorporate state-of-the-art advances in the PHY layer.

3. Modeling of next generation wireless

In the following sections we construct an analytic model that captures both PHY and MAC layer characteristics of next-generation Wi-Fi networks. As is the consensus of both academia and the industry, we envision dense deployments for meeting the increasing wireless data needs. We model such deployments for both state-of-the art and forth-coming technologies. Specifically, in Subsection 3.1 we describe an analytic approximation for 802.11/CSMA-like scheduling which is the basis of all deployed and upcoming Wi-Fi architectures. We proceed in Subsection 3.2 where a short PHY primer is presented along with our assumptions for the channel modeling. Subsection 3.3 introduces the model for a single-user beamforming PHY which is the physical layer underlying most current wireless deployments (802.11n and 802.11ac first stage). In Subsection 3.4 a local MU-MIMO PHY is modeled analytically using random matrix theory lemmas. MU-MIMO is the technology adopted from the second wave of 802.11ac deployments, providing the ability to transmit to multiple users concurrently from a single AP. Finally, Subsection 3.5 describes a coordinated MU-MIMO architecture, the next-generation PHY technology that comes with the promise of mitigating all interference by tightly coordinating multiple APs and transmissions to users. The Section concludes with common system parameters used in the verification and results sections that follow (see Subsection 3.6).

3.1. CSMA modeling

Our goal is to deduce an analytically tractable yet reasonably accurate CSMA model that will be integrated with the PHY layer modeling presented in Sections 3.3 and 3.5. With this in mind, we assume that (i) the transmission and countdown times are exponentially distributed with means $1/\mu$ and $1/\lambda$ respectively, and (ii) the medium can be sensed instantaneously, as in [7–10].

The first assumption makes the system easy to analyze since it can be modeled as a Continuous Time Markov Chain (CTMC) and can be relaxed to milder and more natural system assumptions without affecting the results, as shown in [8]. The combination of the first and the second assumption implies that there will be no collisions between transmitting stations. [2] Clearly, this might lead to an overestimation of the throughput of CSMA when the system spends a sizable time in collisions. However, this won't be the case in practical, well-designed topologies as in optimized enterprise WiFi networks. Simulation results in NS-2 [35] validate the accuracy of the model despite the assumptions above (see Section 4).

Note that it is possible to incorporate in the model both collisions and the binary exponential backoff algorithm along with non-saturated transmission schedules as has been done in [36], but this escapes the purpose of this paper. We hasten to mention that our goal is not to develop a full model of the CSMA/CA, that would lead to non-tractable equations, but rather adopt good assumptions that lead to a provably good approximation while allowing a seamless integration with an advanced PHY model. Last, note that since our CSMA model operates on top of an SINR-based PHY model, we fully model phenomena like hidden terminals and channel capture. For example, if a hidden terminal situation arises, the receiving station will experience high interference and thus a low SINR and a very low rate.

The states of the aforementioned CTMC are all the different feasible transmission patterns for the N_s stations (N_a APs plus the users). That is, only non-conflicting stations, or, equivalently, independent sets of stations, can transmit at the same time. The transitions between the different states happen when a station that is not conflicting with the stations currently transmitting, gets out of countdown and starts transmitting, or when a station finishes its transmission. Let $\mathbf{m} = \{m_k | k = 1, \ldots, N_s\}$ be the binary state vector of the CTMC, where we let $m_i = 1$ if station i is transmitting and $m_i = 0$ otherwise. Let $\mathcal{M} \subseteq \{0,1\}^{N_s}$ be the state space of the CTMC. Following a similar approach with prior work, see, for example, [8,9], we compute the limiting stationary distribution of state $\mathbf{m} \in \mathcal{M}$ to be:

$$\pi_{\mathbf{m}} = \frac{\rho^{\|\mathbf{m}\|_1}}{\sum_{\mathbf{m}' \in \mathcal{M}} \rho^{\|\mathbf{m}'\|_1}}, \tag{1}$$

where $\rho = \lambda/\mu$ and $\|\cdot\|_1$ is the $\mathbf{L_1}$ norm, thus $\|\mathbf{m}\|_1$ is the number of stations transmitting during state $\mathbf{m}$.

As an example, in Fig. 1a we see 6 stations operating on the same channel and the corresponding contention graph, i.e. a graph showing the conflicts between stations in our network with an edge between two stations that cannot transmit concurrently. There are 13 feasible states for this graph and the probabilities for each state can be seen in Fig. 1b. Intuitively, one expects that

[2] Note that in this case all stations will pick a backoff value from the same range, and, as a result, all stations will have the same expected countdown time $1/\lambda$ value.

larger independent sets are more likely to be scheduled by CSMA, since once an station is selected, all sets not containing this station will not be scheduled. Indeed, the probabilities in Fig. 1B are much higher for larger independent sets because the transmission times are much larger than the countdown times and thus $\rho \gg 1$. [3] As a result, in large networks it is reasonable to take into account only states corresponding to *maximal independent sets* of the interference graph, rather than all feasible states. It should be noted here that identifying the maximal independent sets of a graph is known to be an NP-complete problem, nevertheless given the specific characteristics of the interference graphs of CSMA networks we can successfully run large examples within minutes by using the algorithm introduced in [37] and recently used in [36].

Having computed the fraction of time that the network spends in each transmission pattern $\mathbf{m} \in \mathcal{M}$, we next take into account the throughput resulting from CSMA/CA. To do so, we evaluate the spectral efficiency achieved for each transmission pattern, average these values with respect to the stationary distribution (see Eq. 1), and multiply the result by the channel bandwidth to obtain the throughput in bit/s.

3.2. PHY Primer for next-generation Wi-Fi networks

OFDM primer. Orthogonal Frequency Division Multiplexing (OFDM) [38] is the preferred PHY layer modulation technique of modern cellular and WLAN networks. OFDM consists of taking blocks of N coded modulation "frequency domain" symbols and transforming them into a block of N "time domain chips" via an Inverse Discrete Fourier Transform (IDFT). Each block of N time-domain chips is expanded by repeating the last L chips at the beginning of the block. This precoding technique, known as Cyclic Prefixing (CP), is able to turn the linear convolution of the transmitted signal with the channel impulse response into a block-by-block cyclic convolution, provided that the length of the multipath channel (in chips) is not larger than L. At the receiver, after block timing and carrier frequency synchronization, the CP is removed and a DFT is applied to the resulting blocks of N chips in order to recover the N frequency domain symbols. As a result, the time-domain multipath channel is transformed into a set of parallel frequency-flat channels (referred to as subcarriers) in which each frequency-domain symbol experiences only a complex multiplicative "fading" channel coefficient, corresponding approximately to the channel transfer function evaluated at each subcarrier center frequency (see [39] for a recent accurate and general model of OFDM including also non-ideal transmit/receive effects).

Small and large scale fading. It is well-known that typical wireless channels in WLAN environments are slowly-varying in time in the order of hundreds of milliseconds [38]. For a sequence of successive time slots, each one corresponding to a data packet, the channel coefficients are strongly correlated such that estimates obtained at a given point in time are accurate for a fairly large number of time slots.

Motivated by this, in the following we assume that the frequency-domain channel coefficients are random but constant over many data slots, and drop the time index when referring to a time slot for the sake of notation simplicity. We shall use index ν to indicate the OFDM subcarrier index. Hence, the received signal to a user terminal (UT) k equipped with a single antenna from an AP i equipped with M antennas at any given generic time slot and OFDM subcarrier ν is given by

$$y_k[\nu] = \sqrt{g_{ik}}\mathbf{x}_i[\nu]^{\mathrm{H}}\mathbf{h}_{ik}[\nu] + z_k[\nu], \tag{2}$$

[3] The increasing data rates of newer standards, e.g. 256-QAM mode of 802.11ac, have reduced the gap between transmission and countdown times. To maintain high efficiency, the new standards have introduced packet aggregation as a means to maintain high efficiency by keeping transmission times larger.

Table 1
WINNER II parameters (n_w is the number of walls separating the user from the AP).

Scenario	A	B	C	X
Conference Hall	13.9	64.4	20	0
Office floor with rooms (LOS)	18.7	46.8	20	0
Office floor with rooms (NLOS)	36.8	43.8	20	$5(n_w\text{-}1)$
Open floor office	13.9	64.4	20	0
Stadium	13.9	64.4	20	0

where $\mathbf{h}_{ik}[\nu]$ is an $M \times 1$ vector with Gaussian i.i.d. elements $\sim \mathcal{CN}(0,1)$,[4] representing the channel coefficients between the antenna array of AP i and the antenna of UT k, $\mathbf{x}_i[\nu]$ is a $M \times 1$ vector containing the (frequency-domain) coded modulation symbols sent by AP i in the downlink, on subcarrier ν, and $z_k[\nu]$ represents noise plus the interference caused by other APs transmitting on the same time slot and frequency channel at the receiver of UT k.

The channel gain, g_{ik}, is a function of the distance, d_{ik}, between AP i and UT k, the carrier frequency, f_c, and other "large-scale" effects, such as blocking objects, walls and trees. This "large-scale" channel gain is frequency-flat, i.e., it does not depend on the subcarrier index ν, and it is modeled using the widely accepted WINNER-II model [40]. According to this model the pathloss, g_{ik}, is given in dB from the formula below:

$$g_{ik}[dB] = A\log_{10}(d_{ik}[m]) + B + C\log_{10}(f_c[GHz]/5) + X,$$
$$3 \le d_{ik} \le 100,$$

where A, B, C and X are scenario-dependent parameters. For the scenarios we consider, the parameters are shown in Table 1 and can be found in [40] for the corresponding scenarios.

MIMO primer. Point-to-Point MIMO [41,42] consists of the techniques of extracting higher capacities from a given wireless bandwidth by adding antennas to the transmitter and receiver. MIMO is ideally suited to work together with OFDM. We describe the effective channel between a transmitter with M antennas and a receiver with N_r antennas via a set of $M \times N_r$ channel matrices $\mathbf{H}_{ik}[\nu]$ for all the subcarriers forming the given system channel. One of the advantages of having multiple transmit/receive antennas is the ability to increase the effective SNR, known as a *power gain*, through techniques such a conjugate beamforming and maximal ratio combining [38]. In addition, a variety of techniques known as *spatial precoding* [38] enable multiplexing in the spatial domain, such that the capacity at high SNR is increased by a multiplicative factor up to $\min(M, N_r)$, known as a *multiplexing gain*.

For the downlink where N_r may be small, a set of techniques known as MU-MIMO [43] can be used to attain higher multiplexing gain up to $\min\{M, KN_r\}$, for a system where an AP with M antennas serves K UTs with N_r antennas each. Since the UTs do not cooperate, most MU-MIMO schemes require channel state information at the transmitter (CSIT) in order to precode the data streams. This allows multiple data streams to be simultaneously transmitted on the same time-frequency slot, and yet interfere in a benign and controlled manner at each desired UT by using precoding schemes such as zero-forcing beamforming (ZFBF) [44].

3.3. Single-user beamforming

Without loss of generality, let AP i be active on channel c_i, and let $\nu = 1, \ldots, N$ denote the corresponding OFDM subcarrier. Assume that UT k is associated with AP i, i.e., $k \in \mathcal{S}_i$. The *instan-*

[4] $h \sim \mathcal{CN}(0,1)$ indicates a complex circularly symmetric Gaussian random variable with mean 0 and variance $\mathbb{E}[|h|^2] = 1$. This type of small-scale fading is usually referred to as Rayleigh fading [38].

taneous peak rate at which AP i can serve user k is given by

$$C_{ik}^{\mathbf{m}} = m_i \frac{1}{N} \sum_{\nu=1}^{N} \log\left(1 + \mathrm{SINR}_{ik}^{\mathbf{m}}[\nu]\right), \tag{3}$$

where $\mathbf{m}$ denotes the current state of the CMTC (such that this rate is zero if $m_i = 0$, i.e., AP i is not in the current active set according to the CSMA protocol). The rate $C_{ik}^{\mathbf{m}}$ is measured in bit/s/Hz, and it is referred to as "instantaneous" since it is a function of the realization of the fading channel coefficients. The term $\mathrm{SINR}_{ik}^{\mathbf{m}}[\nu]$ denotes the Signal to Interference plus Noise Ratio on subcarrier ν. With single-user beamforming, this is given by

$$\mathrm{SINR}_{ik}^{\mathbf{m}}[\nu] = \frac{g_{ik}\|\mathbf{h}_{ik}[\nu]\|^2 P_i}{1 + \sum_{j\in\mathcal{A}_{c_i}: j\neq i} m_j g_{jk} \left|\mathbf{v}_j^{\mathrm{H}}[\nu]\mathbf{h}_{jk}[\nu]\right|^2 P_j}, \tag{4}$$

where the set $\mathcal{A}_{c_i}$ includes the APs assigned to channel c_i, P_i denotes transmit power spectral density (energy per frequency domain symbol) of AP i and $\mathbf{v}_i^{\mathrm{H}}[\nu] = \mathbf{h}_{ik}[\nu]/\|\mathbf{h}_{ik}[\nu]\|$ is the unit-norm transmit beamforming vector of AP i serving user k, according to the conjugate beamforming scheme. The interfering APs $j \in \mathcal{A}_{c_i}$ use the same scheme, with a unit vector $\mathbf{v}_j[\nu]$ that depends on the channel vector to their own intended user. In the SINR denominator, the inner product $\mathbf{v}_j^{\mathrm{H}}[\nu]\mathbf{h}_{jk}[\nu]$ corresponds to the "spatial coupling" between the beamforming vector of AP j and the channel vector from AP j and user k. It is important to note that $\mathbf{v}_j[\nu]$ and $\mathbf{h}_{jk}[\nu]$ are statistically independent. Thus, $\mathbf{v}_j^{\mathrm{H}}[\nu]\mathbf{h}_{jk}[\nu]$ has the same distribution as $\mathbf{h}_{jk}[\nu]$ since $\mathbf{v}_j^{\mathrm{H}}[\nu]$ is a unitary vector. On the other hand, the coupling between the beamforming vector $\mathbf{v}_i[\nu]$ of AP i and its intended user channel $\mathbf{h}_{ik}[\nu]$ yields $\mathbf{v}_i^{\mathrm{H}}[\nu]\mathbf{h}_{ik}[\nu] = \|\mathbf{h}_{ik}[\nu]\|$ which provides the beamforming gain in the numerator of the SINR. It follows that $\left|\mathbf{v}_j^{\mathrm{H}}[\nu]\mathbf{h}_{jk}[\nu]\right|^2$ is a chi-squared random variable with 2 degrees of freedom, while $\|\mathbf{h}_{ik}[\nu]\|^2$ is a chi-squared random variable with $2M$ degrees of freedom.

In order to obtain simple deterministic analytical formulas for the rates, we consider that M is large enough such that the effect of small-scale fading disappears. As a result, the statistical fluctuations of the random quantities around their mean in Eq. (4) are quite small, see [45]. Motivated by this, we replace $\|\mathbf{h}_{ik}[\nu]\|^2$ and $|\mathbf{v}_j^{\mathrm{H}}[\nu]\mathbf{h}_{jk}[\nu]|^2$ with their expected values, M and 1 respectively. Thus, the resulting deterministic rate formula for the peak rate is:

$$C_{ik}^{\mathbf{m}} = m_i \log\left(1 + \frac{g_{ik} M P_i}{1 + \sum_{j\in\mathcal{A}_{c_i}: j\neq i} m_j g_{jk} P_j}\right). \tag{5}$$

With orthogonal downlink multiple access (TDMA) the rate region for the downlink of AP i for $m_i = 1$ (i.e., in the time slots where AP i is allowed to transmit because of CSMA/CA) is given by the set of non-negative rates $\{R_{ik}^{\mathbf{m}} : k \in \mathcal{S}_i\}$ such that

$$\sum_{k\in\mathcal{S}_i} \frac{R_{ik}^{\mathbf{m}}}{C_{ik}^{\mathbf{m}}} \leq 1. \tag{6}$$

With proportional fairness downlink resource allocation, the system operates at the rate point

$$R_{ik}^{\mathbf{m}} = \frac{m_i}{|\mathcal{S}_i|} \log\left(1 + \frac{g_{ik} M P_i}{1 + \sum_{j\in\mathcal{A}_{c_i}: j\neq i} m_j g_{jk} P_j}\right). \tag{7}$$

Eventually, the average spectral efficiency of user $k \in \mathcal{S}_i$ when also averaging with respect to the stationary distribution of the CTMC that describes the CSMA/CA MAC layer, is given by

$$R_{ik} = \sum_{\mathbf{m}\in\mathcal{M}} \pi_{\mathbf{m}} \cdot R_{ik}^{\mathbf{m}}. \tag{8}$$

In order to convert this number in the more usual average throughput in bit/s, it is sufficient to multiply R_{ik} by the channel bandwidth W_{c_i} (measured in Hz) of the channel c_i allocated to AP i.

Once the user spectral efficiencies R_{ik} are determined, we can present the results in terms of the throughput CDF for a given placement of the APs and of the UTs, pathloss realization, channel allocation and user-AP association. Letting $T_{ik} = W_{c_i} R_{ik}$ denote user k average throughput in bit/s, the throughput CDF over the user population is given by

$$F_T(r) = \frac{1}{\sum_i |\mathcal{S}_i|} \sum_{i,k} 1_{\{T_{ik}\leq r\}}. \tag{9}$$

It follows that $F_T(r)$ indicates the fraction of users with throughput less or equal to some number r.

3.4. Local MU-MIMO

We now consider the case where each AP i implements MU-MIMO in order to serve its associated users $k \in \mathcal{S}_i$. This scheme is inspired by the MU-MIMO mode of 802.11ac [46]. We denote this scheme as "local" MU-MIMO in order to stress the difference with respect to a coordinated MU-MIMO approach that shall be treated in the next section, and may be regarded as a future trend of WLANs and small cell networks. The main modeling difficulty here is represented by the fact that AP i can simultaneously serve a subset of users of size not larger than $\min\{M, |\mathcal{S}_i|\}$. Consistent with the 802.11ac standard [46], we consider MU-MIMO based on ZFBF.

Let $\widehat{\mathcal{S}}_i \subseteq \mathcal{S}_i$ denote the subset of users to be served on a given time slot in MU-MIMO mode, and let $S_i = |\widehat{\mathcal{S}}_i|$ indicate its size. The $M \times 1$ channels of users $k \in \widehat{\mathcal{S}}_i$ are assumed to be known at the AP i transmitter through some form of channel state feedback (e.g., as specified in the 802.11ac standard). Such channel vectors are collected as the columns of a $M \times S_i$ channel matrix $\mathbf{H}_i[\nu] = [\mathbf{h}_{i1}[\nu], \ldots, \mathbf{h}_{ik}[\nu]]$, for all the subcarriers ν forming channel c_i, as defined before. The ZFBF precoded signal vector on subcarrier ν is given by $\mathbf{x}_i[\nu] = \mathbf{V}_i[\nu]\mathbf{u}_i[\nu]$, where $\mathbf{u}_i[\nu]$ is a $S_i \times 1$ column vector of frequency-domain coded modulation symbols to be sent to users $k \in \widehat{\mathcal{S}}_i$ and $\mathbf{V}_i[\nu]$ is the ZFBF precoding matrix, of dimension $M \times S_i$, given by

$$\mathbf{V}_i[\nu] = \mathbf{H}_i[\nu]\left(\mathbf{H}_i^{\mathrm{H}}[\nu]\mathbf{H}_i[\nu]\right)^{-1} \boldsymbol{\Xi}_i^{1/2}[\nu], \tag{10}$$

where $\boldsymbol{\Xi}_i[\nu]$ is a column-normalizing diagonal matrix included in order to preserve the total AP transmit power. In particular, the k-th diagonal element of $\boldsymbol{\Xi}_i[\nu]$ is given by $\xi_{ik}[\nu] = 1/[(\mathbf{H}_i^{\mathrm{H}}[\nu]\mathbf{H}_i[\nu])^{-1}]_{kk}$ where $[\,\cdot\,]_{kk}$ denotes the k-th diagonal element of the matrix argument. Using the fact that $\mathbf{V}_i^{\mathrm{H}}[\nu]\mathbf{H}_i[\nu] = \boldsymbol{\Xi}_i^{1/2}[\nu]$ and letting $\lambda_{ik}[\nu] = g_{ik}\xi_{ik}[\nu]$ denote the effective channel coefficient including the large-scale gain from Eq. (2), the signal received at UT k takes on the form

$$\begin{aligned} y_k[\nu] &= \sqrt{g_{ik}}\mathbf{u}_i[\nu]^{\mathrm{H}}\mathbf{V}_i^{\mathrm{H}}[\nu]\mathbf{h}_{ik}[\nu] + z_k[\nu] \\ &= \sqrt{\lambda_{ik}[\nu]}u_{ik}[\nu] + z_k[\nu], \end{aligned} \tag{11}$$

which illustrates that the multi-access interference from signals generated by AP i and sent to the other users $k' \in \widehat{\mathcal{S}}_i$ with $k' \neq k$ is completely eliminated by ZFBF precoding.

Based on the effective channel (11), we have that the instantaneous user rate takes on the same form (3) with a different expression of $\mathrm{SINR}_{ik}^{\mathbf{m}}[\nu]$. In particular, in the case of local MU-MIMO with ZFBF precoding this is given by

$$\mathrm{SINR}_{ik}^{\mathbf{m}}[\nu] = \frac{\lambda_{ik}[\nu] P_i/S_i}{1 + \sum_{j\in\mathcal{A}_{c_i}: j\neq i} m_j g_{jk} \|\mathbf{V}_j^{\mathrm{H}}[\nu]\mathbf{h}_{jk}[\nu]\|^2 P_j/S_j}, \tag{12}$$

where we have assumed (for the sake of simplicity) uniform power allocation over the downlink data streams, i.e., each user gets $1/S_i$ of the total AP transmit power.

A major difficulty here is represented by the fact that the coefficient $\lambda_{ik}[\nu]$ depends on the channel matrix through the whole selected group of users $\widehat{\mathcal{S}}_i$, which means that the individual user rates do not "decouple": for finite dimension, we need to calculate the rates for all the $G_i = \binom{|\mathcal{S}_i|}{S_i}$ possible groups. In order to alleviate this problem and obtain a "decoupled" rate expression, we resort to asymptotic formulas. With this in mind, we consider the regime where both M and S_i become large while keeping the ratio $S_i/M \leq 1$ and fixed. Specifically, we scale all channel coefficients by $1/\sqrt{M}$ and multiply the transmit power by a factor M, which yields

$$\lambda_{ik}[\nu] \to [1 - (S_i - 1)/M] \cdot g_{ik},$$

and

$$\|\mathbf{V}_j^{\mathrm{H}}[\nu]\mathbf{h}_{jk}[\nu]\|^2 \to \frac{1}{M}\mathrm{tr}(\mathbf{V}_j^{\mathrm{H}}[\nu]\mathbf{V}_j[\nu]) = \frac{S_j}{M},$$

where the asymptotic convergence follows from the trace lemma of random matrices (see [47] Theorem 3.4, as well as [15,48,49] for similar asymptotic approaches).

With these limits, we obtain the SINR deterministic approximation (which becomes exact in the large-system regime)

$$\mathsf{SINR}_{ik}^{\mathbf{m}} \to \frac{(M - S_i + 1)g_{ik}P_i/S_i}{1 + \sum_{j\in\mathcal{A}_i : j\neq i} m_j g_{jk} P_j}. \tag{13}$$

Notice that, as in the case of single-user beamforming, this limit does not depend on the subcarrier index ν any longer. Also, for the sake of consistency, it is interesting to notice that for $S_i = 1$ we recover the expression for single user conjugate beamforming

Under this approximation, the vector of user rates for a given active group of users $\widehat{\mathcal{S}}_i \subseteq \mathcal{S}_i$ is given by

$$C_{ik}^{\mathbf{m}}(\widehat{\mathcal{S}}_i) = \begin{cases} 0 & \text{for} \quad k \in \mathcal{S}_i - \widehat{\mathcal{S}}_i \\ m_i \log\left(1 + \mathsf{SINR}_{ik}^{\mathbf{m}}\right) & \text{for} \quad k \in \widehat{\mathcal{S}}_i. \end{cases} \tag{14}$$

The achievable rate region in the case of local MU-MIMO is significantly more complicated to express than for the case of single-user beamforming with downlink orthogonal access. In this case, any group of users $\widehat{\mathcal{S}}_i \subseteq \mathcal{S}_i$ with size $S_i \leq M$ can be scheduled in the downlink. The individual average rate of user k is the convex combination of the rates achieved in each group, where the convex combination coefficients depend on the downlink scheduling scheme. In particular, we can order the groups $\widehat{\mathcal{S}}_i \subseteq \mathcal{S}_i$ of size $S_i \leq M$ in lexicographic order. Let $\mathbf{C}_i^{\mathbf{m}}(\widehat{\mathcal{S}}_i)$ denote the $1 \times |\mathcal{S}_i|$ row vector of user rates given in (14), and let $\mathbf{C}^{\mathbf{m}}(\mathcal{S}_i)$ denote the $G_i \times |\mathcal{S}_i|$ matrix obtained by stacking all the group rate rows on top of each other in the same group lexicographic order. Define the $1 \times |\mathcal{S}_i|$ rate vectors $\mathbf{R}_i^{\mathbf{m}}$ with components $R_{ik}^{\mathbf{m}}$ for all users $k \in \mathcal{S}_i$. The rate region obtained by applying TDMA downlink scheduling on top of MU-MIMO for each AP i is given by the union of all points $\mathbf{R}_i^{\mathbf{m}}$ satisfying

$$\mathbf{R}_i^{\mathbf{m}} \leq \boldsymbol{\rho}^{\top}\mathbf{C}^{\mathbf{m}}(\mathcal{S}_i) \tag{15}$$

for some non-negative time-sharing vector $\boldsymbol{\rho}$ of dimension $G_i \times 1$, with components satisfying $\sum_{k\in\mathcal{S}_i}\rho_k = 1$. Notice also that this region generalizes the single-user beamforming TDMA region, since in this case we have only $|\mathcal{S}_i|$ possible groups of size 1, and combining the inequalities $R_{ik}^{\mathbf{m}} \leq \rho_k C_{ik}^{\mathbf{m}}$ with $\sum_{k\in\mathcal{S}_i}\rho_k = 1$ yields again Eq. (6).

Focusing on proportional fairness downlink resource allocation, the AP gives to each user $k \in \mathcal{S}_i$ equal air time, such that the user rates are given by

$$R_{ik}^{\mathbf{m}} = m_i \frac{S_i}{|\mathcal{S}_i|}\log\left(1 + \mathsf{SINR}_{ik}^{\mathbf{m}}\right). \tag{16}$$

The number of downlink streams S_i may be fixed by some technology constraints (e.g., by the number of data streams that the AP chipset is able to precode), or it may be optimized for each AP, subject to the constraint $S_i \leq \min\{M, |\mathcal{S}_i|\}$. In this work, the latter was chosen as we maximize the sum-rate that every AP serves its users individually for every AP.

As before, the average user rate over all possible states of the CSMA/CA CTMC is given by using Eq. (16) in (8). From the resulting average rates $\{R_{ik}\}$, the throughput CDF can be obtained via Eq. (9).

3.5. Coordinated MU-MIMO

In this case we assume that all APs that operate in the same channel cooperate and act as a single virtual mega AP (no contention between the APs for the wireless medium). Therefore, the channel index is not necessary and is dropped for notation simplicity.

Assume that B APs with M antennas each form a cluster, pooling together all their BM antennas, and collectively serve the population of K UTs with a single antenna each. This type of coordinated MU-MIMO configuration has been widely studied as far as the PHY algorithms and the achievable rates from a communication theoretic viewpoint are concerned (see for example prior work of ours [15,16,39] and others [22,50]). Recently, experimental results with such systems have been published (see for example our own prior work [3] as well as work from others [4,51]). As before, we wish to characterize the rate per user in which a subset $\widehat{\mathcal{S}}$ of size $S \leq \min\{K, BM\}$ of users is served simultaneously by the coordinated MU-MIMO scheme. We shall then apply a TDMA scheduling over the user subsets in order to obtain a desired throughput K-tuple satisfying some required fairness criterion.

As far as the PHY schemes and corresponding performance are concerned, the analysis of coordinated MU-MIMO is significantly more involved than the previously treated cases because the channel vectors from all the AP antennas to any given user are not i.i.d. given the pathloss. In fact, letting g_{ik} denote again the channel gain coefficient from AP i to user k and letting $\mathbf{h}_{ik}[\nu]$ denote the channel small fading coefficient, we have that the composite channel vector is given by

$$\mathbf{h}_k[\nu] = \left[\sqrt{g_{1k}}\mathbf{h}_{1k}[\nu], \sqrt{g_{2k}}\mathbf{h}_{2k}[\nu], \ldots, \sqrt{g_{Bk}}\mathbf{h}_{Bk}[\nu]\right]^{\top}.$$

For a certain set of $\widehat{\mathcal{S}} = \{1, \ldots, S\}$ users to be served, the resulting $BM \times S$ channel matrix takes on the form

$$\mathbf{H}[\nu] = \begin{bmatrix} \sqrt{g_{11}}\mathbf{h}_{11}[\nu] & \sqrt{g_{12}}\mathbf{h}_{12}[\nu] & \cdots & \sqrt{g_{1S}}\mathbf{h}_{1S}[\nu] \\ \sqrt{g_{21}}\mathbf{h}_{21}[\nu] & \sqrt{g_{22}}\mathbf{h}_{22}[\nu] & \cdots & \sqrt{g_{2S}}\mathbf{h}_{2S}[\nu] \\ \vdots & & & \vdots \\ \sqrt{g_{B1}}\mathbf{h}_{B1}[\nu] & \sqrt{g_{B2}}\mathbf{h}_{B2}[\nu] & \cdots & \sqrt{g_{BS}}\mathbf{h}_{BS}[\nu] \end{bmatrix},$$

where the effective channel coefficient λ_k for user $k \in \widehat{\mathcal{S}}$ under joint ZFBF from all the B APs is given by

$$\lambda_k[\nu] = 1/[(\mathbf{H}^{\mathrm{H}}[\nu]\mathbf{H}[\nu])^{-1}]_{k,k}$$

and the SINR is given by

$$\mathsf{SINR}_k[\nu] = \lambda_k[\nu]\frac{P_{\mathrm{sum}}}{S}, \tag{17}$$

where we define the sum power of all APs in the cluster as $P_{\mathrm{sum}} = \sum_{i=1}^{B} P_i$ and assume equal power allocation on all downlink data streams. Note that since all APs operate as a single virtual mega AP, there is no interference between APs and thus there is no interference term in the SINR above.

The problem is that now the structure of the channel matrix depends on the actual set of users $\widehat{\mathcal{S}}$ being served. Hence, even using asymptotic formulas from random matrix theory, we should evaluate rates for each set of $\binom{K}{S}$ users, for $S = 1, \ldots, \min\{K, BM\}$. With coordinated MU-MIMO, relevant numbers are (for example) $K = 100$, $M = 4$ and $B = 16$. Clearly even enumerating all the possible active user subsets $\widehat{\mathcal{S}}$ is difficult.

Inspired by prior work of ours [15] we propose the following simplification: Under certain symmetry conditions in the pathloss coefficients, the effective channel coefficient $\lambda_k[\nu]$ for user k under joint ZFBF from all the B APs, in the limit for large M and large S with fixed ratio M/S, asymptotically takes on the form:

$$\lambda_k[\nu] \to [1-(S-1)/(BM)] \cdot \sum_{i=1}^{B} g_{ik}.$$

Once more, the limiting expression does not depend on the subcarrier index any longer. We use the above expression as an approximation of the ZFBF effective channel coefficients for general coordinated MU-MIMO. This has the following appealing interpretation: the coordinated MU-MIMO scheme behaves as a local MIMO system with pathloss coefficients equal to the sum of the pathloss coefficients and the total number of antennas (the number of antennas of all coordinated APs in the cluster). Thus, we obtain the (approximated) user downlink rate as:

$$C_k = \begin{cases} 0, & k \in \mathcal{S} - \widehat{\mathcal{S}} \\ \log\left(1 + \left(M - \frac{S-1}{B}\right)\left(\sum_{i=1}^{B} g_{ik}\right)\frac{P_{\text{sum}}}{S}\right), & k \in \widehat{\mathcal{S}}. \end{cases} \quad (18)$$

Notice that this is the effective performance of a virtual local MU-MIMO system with "resource pooling", with M effective antennas, effective load (number of active users per antenna) equal to S/B, effective channel gains $g_k = \sum_{i=1}^{B} g_{ik}$ and effective power P_{sum}. Notice also that for $B = 1$ the above formula coincides with the case developed for local MU-MIMO in the absence of inter-AP interference.

3.6. System level parameters

Frequency reuse/channel assignment. In our numerical examples, we have assumed that the system has 80 MHz of total bandwidth that can be partitioned into 20 MHz or 40 MHz non-overlapping channels or a single 80 MHz channel, with reference to the 802.11 2.4 GHz band and the 802.11ac revision [46]. Each AP is active on one of the system channels. The same channel can be allocated to multiple APs according to some suitable spatial frequency reuse scheme. The allocation of channels to APs is, in general, a hard problem. In this work, we have adopted a one-pass, greedy algorithm. APs are ordered according to a random sorting permutation π and choose their channel sequentially, such that AP $\pi(i)$ makes its choice at step i and chooses channel $c_{\pi(i)} \in \mathcal{C}$ where $\mathcal{C}$ is the set of available system channels (numbered as $1 \ldots |\mathcal{C}|$). The channel choice is made according to the rule: $c_{\pi(i)} = \arg\min_{c \in \mathcal{C}} \{\sum_{j<i:\pi(j) \in \mathcal{A}_c(i)} I^{c}_{\pi(i),\pi(j)}\}$ where $I^{c}_{\ell,r}$ denotes the interference power received by AP ℓ from AP r in channel c, and $\mathcal{A}_c(i)$ denotes the set of APs already assigned to channel c at step i of the assignment procedure.

User-AP association. The simplest, most widely used user-AP association algorithm connects each user to the AP with the highest Received Signal Strength Indication (RSSI). This solution does not take into account load balancing; some APs may be cluttered by a lot of users while others may be left almost idle. The typical industry solution to this problem is to force a new user to switch to another AP if the AP with the strongest RSSI has too many users [1].

The AP association and load balancing issue are relatively well-studied problems in academia, see, for example, [52–54]. In our examples (see Section 5), we have considered a heuristic user-AP assignment scheme aimed at improving the system fairness. Inspired by the work in [55], we allocate users based on "available capacity". A central controller orders the users according to a sorting random permutation π' and associates each user sequentially to the AP that offers the largest available capacity. In particular, the association algorithm starts by assigning to all APs $i = 1, \ldots, N_a$ an empty set $\mathcal{S}_i = \emptyset$ of associated users. Then, at step k, user $\pi'(k)$ is added to the set $\mathcal{S}_{i_k}$ if i_k is the AP index satisfying $i_k = \arg\max_{i=1,\ldots,N_a} \frac{C_{i,\pi'(k)}}{|\mathcal{S}_i \cup \{\pi'(k)\}|}$, where $C_{i,k}$ denotes the peak rate of the link from AP i to user k (as estimated from Eqs. (5), (14) and (18)). The association ends when all users have been assigned. Notice that we use the ratio $\frac{C_{i,\pi'(k)}}{|\mathcal{S}_i \cup \{\pi'(k)\}|}$ as a proxy of the *average* throughput that user $\pi'(k)$ can get from AP i. This is motivated by the fact that, under proportional fairness scheduling, each AP allocates its downlink channel resource in equal proportion to its users.

Note that while the channel allocation and user-AP association schemes adopted in this paper are sensible approaches, they do not correspond necessarily to optimal strategies. Nevertheless, the purpose of this paper is not that of proposing system optimization strategies, rather, our focus is to develop a general analytical model which works irrespectively of the specific channel allocation and user-AP association scheme and can be used to enable such a system optimization.

Coordinated MU-MIMO clusterization. For a network with N_a APs, letting $B = N_a$ with coordinated MU-MIMO yields that the whole network corresponds to a single coordination cluster. In practice, this may be too constraining since other system limitation aspects arise in coordinated MU-MIMO architectures. For example, since all user data need to be precoded jointly from all the coordinated APs in the cluster, eventually the wired backbone network connecting the APs will become the system bottleneck. Here we are not concerned with such practical implementation problems of coordinated MU-MIMO, but for the sake of generality we wish to take into account the case where the network is split into multiple coordination clusters. In particular, in some numerical results we consider the case where the network is split into 4 clusters, each of which with $B = N_a/4$ APs, and operating on a different 20 MHz channel, such that there is no inter-cluster interference. In this case, the same formula (18) applies, where we restrict the active user set $\widehat{\mathcal{S}}$ to be a subset of the users assigned to the given cluster. Similarly, we can cluster the APs in 2 clusters assigning the two 40 MHz non-overlapping channels to each one. User-cluster association can be done through a greedy scheme similar to what we have seen before for the user-AP association. Finally, a larger number of clusters can be chosen, and then inter-cluster interference has to be accounted for.

As with the single-user beamforming and local MU-MIMO systems, for coordinated MU-MIMO we are interested in the per-user downlink throughput under proportional fairness scheduling. Letting S denote the number of users simultaneously served by the cluster of APs and assuming that users are given equal air time, we have that each user k is served for a fraction of time equal to $\frac{S}{K}$. Hence, the proportional fairness scheduler with equal power per stream yields throughput

$$R_k = \frac{S}{K} \log\left(1 + \left(M - \frac{S-1}{B}\right)\left(\sum_{i=1}^{B} g_{ik}\right)\frac{P_{\text{sum}}}{S}\right). \quad (19)$$

These rates can be maximized with respect to the size of the active user set $S = 1, 2, \ldots, \min\{K, BM\}$. From the resulting average rates $\{R_{ik}\}$, the throughput CDF can be obtained again via Eq. (9).

4. Model validation

We will proceed with the validation of our model in two steps. First, we will evaluate the accuracy of the analytic PHY models using a custom Monte Carlo simulator implemented in Matlab, and then the accuracy of the MAC model (CSMA CTMC) using NS-2 [35]. Notice that no open-source 3rd party simulator supports the advanced PHY layer schemes that we consider, and thus this two

step validation process is required to validate our model in its entirety.

As an example, consider Eq. (8) which gives the rate of a user under single-user beamforming. This equation uses the rates computed in (7) and the stationary distribution of the CSMA CTMC computed in (1). In the first step of the validation process we will use our custom simulator to verify the accuracy of Eq. (7), and, in the second step we will use NS-2 to validate Eq. (1). A similar process is used for the cases of local MU-MIMO and coordinated MU-MIMO, where in the former case the user rates are given by (16) and in the later by (19).

We start with the first step of the validation process, namely the validation of Eqs. (7), (16), and (19). In all three cases, we compare the distribution of the average per-user rates as computed by the model and by our simulation for some representative tractable scenarios of interest, namely an open conference hall (see Section 5.2) and an office with rooms (see Section 5.4). We utilize the same user locations for both the analytic and simulation results, which are illustrated in Fig. 5a and c. For the scenarios which require APs to use orthogonal channels, the channel selection is performed as in Section 3.6, where 4 channels of 20MHz each are shared among the APs. Lastly, the user-AP association is performed as in Section 3.6 as well.

Our custom Monte Carlo simulator proceeds as follows. For each topology scenario it creates a large number of channel instantiations. Then, for a particular channel realization, it computes the received signal power and the received interference power for each user, calculates the SINR of each user under that channel realization, and computes the corresponding (Gaussian) user rates. Finally, it calculates the average achievable rate for each user across all the channel realizations and the empirical CDF of user rates. It is important to note that the instantaneous SINR values are computed using a different formula by the Monte Carlo simulator depending on the PHY model. For single-user beamforming we use (4), for local MU-MIMO (12), and for coordinated MU-MIMO (17) (and variations to take into consideration different cluster sizes).

For the SU-MISO simulations, each transmitter transmits to a random user located within its cell on each iteration. The transmitters are assumed to have channel state information, and thus beamform to their selected user. Physically adjacent APs operate on separate channels, such that users experience interference only from non-neighbor APs transmitting on the same frequency. Each AP transmits with power $P = 90$ dB above the noise floor, and at a bandwidth of 20 MHz. Including pathloss, the users receive their respective signals at typical Wi-Fi SNRs in the range of roughly 0–40 dB.

In the case of the 802.11ac-like approach, APs retain the same channel allocation as in the SU-MISO simulations, such that neighboring APs operate on different frequencies. Instead of transmitting to a single user, however, they beamform to a random subset of the users in each cell using ZFBF. Based on simulation trials, we determined the optimal number of users (on average) to serve simultaneously in each cell, which was typically 2 or 3. Again, each AP transmits with power $P = 90$ dB above the noise floor at a bandwidth of 20 MHz.

The coordinated MIMO simulations utilize a clustering approach in which groups of 5 APs achieve sufficient synchronization such that they act as a single 20-antenna AP. In this case, all interference is nulled within a given cluster, and each cluster is assigned to a channel as in the previous two simulations. We assume that the APs pool their power, such that total transmit power is ~ 97 dB within each cluster. This simplifies the treatment of the precoder calculation, but it may allow individual APs to transmit higher than 90 dB power above the noise floor. In practice, we would need to utilize results on ZFBF under per-antenna power constraints [56], though we avoid this issue since it is not treated in the model. As in the previous simulations, each cluster occupies 20 MHz of bandwidth.

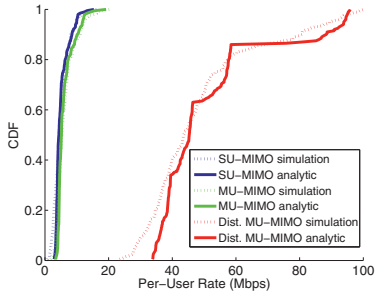
Figs. 2 and 3 display the CDFs of the average per-user rates for the simulations and the models. Although the analytic models make a number of assumptions, for example, in the single-user beamforming case that the number of antennas M per AP is large, in the local MU-MIMO case that the number of antennas and users per AP is large, and in the coordinated MU-MIMO case that, in addition, there is a specific geometric symmetry in the topology, we see that analytical results track closely the Monte Carlo simulations. We should note that specifically for the coordinated MU-MIMO case, the extra symmetric condition assumption that was necessary in order to break the dependence from the random channel coefficients, produces a larger gap between analytic and simulated results compared to the other results. The analytic results still, nevertheless, tracks with satisfying approximation the CDF of the simulation results. We have produced similar validation results for a variety of user and AP placements along with user-AP association schemes and varying AP numbers.

Lastly, we proceed to the second step of the validation process and evaluate the accuracy of the CSMA CTMC model used in this work for the open conference hall scenario depicted in Fig. 5a using NS-2 [35]. We want to compute the relative airtime of each AP, taking into account contention from other stations (APs and users). We assume a ratio of 90/10 for downlink and uplink traffic. This decision is motivated by the fact that the traffic patterns of enterprise WiFi deployments like the ones examined here are heavily asymmetric with the majority of the traffic being downlink, forcing network planners to give high priority to APs during contention to avoid congestion and packet losses (see [57] and references therein).

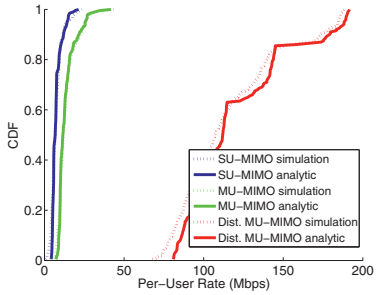
Stations form a contention graph by setting an SINR threshold below which the interference is treated as noise. (This is known as the Clear Channel Assessment (CCA) threshold.) Stations in the sensing range of each other that operate on the same channel cannot transmit concurrently, due to the CSMA/CA algorithm and thus they share a link in the contention graph.

In order to infer the APs relative airtime in NS-2, an 802.11 scenario is constructed where every AP and transmitting UL user is to transmit a UDP stream with constant bit rate (CBR) traffic of 7 Mbps. At the MAC layer, the maximum data rate was set to 11 Mbps [5], the ACKs and other control messages are transmitted as usual at the base rate of 1 Mbps, and the carrier sensing and receiving thresholds were arranged so that the same interference graph resulted for both the simulation and the CTMC model. Due to PHY and MAC overhead, e.g. airtime consumed by ACKs, NS-2 yields an effective rate of 5.3 Mbps when every station is transmitting in isolation. Thus, by simulating the achievable goodput under saturation conditions when all stations are transmitting, we can estimate the relative airtime of each AP. The results can been seen in Fig. 4a and b for 16 and 20 APs respectively all operating at the same channel. In the same figure, error bars indicating two standard deviation intervals are plotted from the measurements coming from different possible UL traffic scenarios (placement of users). As can be seen the analytic model closely tracks the NS-2 relative airtime allocations with their difference being mostly below 15%. (There are occasions where larger deviations are noticed, for example for a couple of APs in the 20 APs scenario, but even in

[5] Notice that the specific data rate does not impact the verification procedure since NS-2 is used only to verify the relative air-time between APs. Therefore, even in higher data rates, since the transmition simulated is a UDP flow of an infinite packet at a constant bit rate, the relative airtimes would remain the same. The choice of the 11 Mbps data rate comes from the fact that NS-2 models up to 802.11b transmit rates.

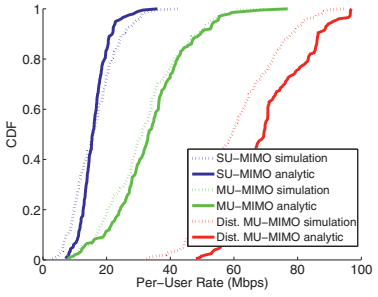


(a) 20 APs and 4 antennas per AP

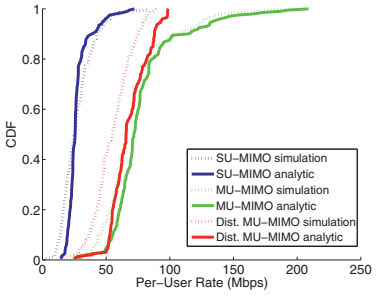


(b) 20 APs and 10 antennas per AP

Fig. 2. CDFs of the user rate for the analytic model vs. simulation for the conference hall scenario.

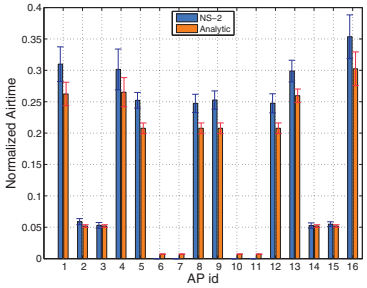


(a) 4 rooms, 20 APs, 4 antennas per AP

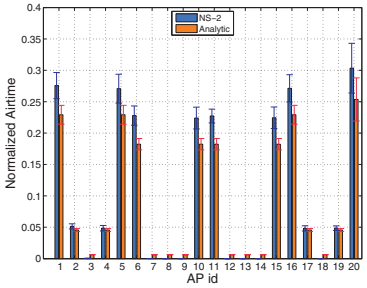


(b) 40 rooms, 20 APs, 4 antennas per AP

Fig. 3. CDFs of the user rate for the analytic model vs. simulation for the office floor scenario.



(a) 16 APs.



(b) 20 APs.

Fig. 4. NS-2 and analytic relative airtime allocations validation plots for the conference room scenario.

these cases the difference between the corresponding probabilities is rather small.) Note that a similar CTMC CSMA model has been validated under small scale scenarios in [8].

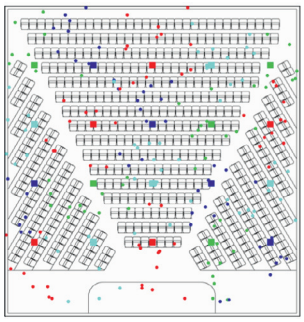
The formulas used in our Monte Carlo simulations (e.g. Eqs. (4), (12) and (17)) are stochastic formulas which depend on the channel realization and thus computing rates via them requires quite a bit of computational power. In contrast, the obtained deterministic formulas of our analytical model, e.g. Eq. (13) in place of (12), require far less computational power. For example, for the stadium scenario of Section 5.5 (500 APs and 20000 users), a Monte Carlo simulation for a single instantiation takes roughly an hour to run on a dual-core Intel i3 3.2GHz processor, compared to less than a second for the analytic computation. Moreover, NS-2 can't even handle such a large scale scenario. For a moderate scale scenario with, say, 100 APs, it takes NS-2 15 min to run a 2 min simulation interval whereas it takes a few seconds to compute the maximal independent sets and the associated stationary distribution of the CTMC. Taking into account that in an optimization setting multiple setups would have to be evaluated in terms of user/AP association, number of APs, transmit power of APs etc., it becomes apparent that the analytical model is the only practical way to proceed. What is more, the analytical model can also capture network dynamics, e.g. due to the mobility of nomadic users, in the sense that its runtime is orders of magnitude faster than the rate of change of such dynamics.

5. Network performance under practical scenarios

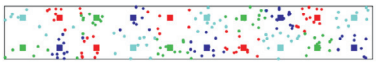
In the following section, results and insights from the application of the analytic model in various practical scenarios will be presented. Most of these results would be practically intractable to run using Monte Carlo and NS-2 simulation as noted earlier.

5.1. OFDM and MCS overhead

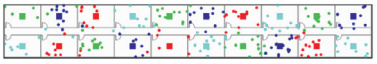
We discuss the rate degradation from having actual quantized rates based on the 802.11 standard's Modulation and Coding schemes (MCS) along with the overhead incurred by OFDM. There are of course other sources of overhead as well, e.g. the CSMA overhead included in our model already (see Section 3.1), and the overhead to collect channel state information and/or coordinate



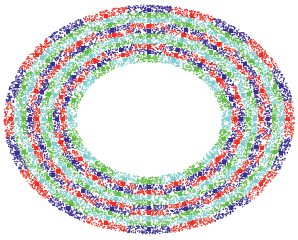
(a) Conference hall with 20 APs and 200 users.



(b) Open-floor office building with 20 APs and 200 users.



(c) Office building with rooms with 20 APs and 200 users.



(d) Stadium scenario with 200 APs and 20000 users.

Fig. 5. AP placement, channel allocation and user-to-AP assignment for the scenarios examined. In the figures users are dots and APs are squares. The 4 different colors account for the 4 non-overlapping channels. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Modulation/Coding pairs from IEEE 802.11ac and the corresponding SINRs at which they can be selected.

802.11ac MCS index	Modulation	Code rate	SNR
0	BPSK	1/2	$\geq$ 2 dB
1	QPSK	1/2	$\geq$ 5 dB
2	QPSK	3/4	$\geq$ 8 dB
3	16-QAM	1/2	$\geq$ 12 dB
4	16-QAM	3/4	$\geq$ 15 dB
5	64-QAM	2/3	$\geq$ 18 dB
6	64-QAM	3/4	$\geq$ 21 dB
7	64-QAM	5/6	$\geq$ 24 dB
8	256-QAM	3/4	$\geq$ 27 dB

remote APs in the context of coordinated MU-MIMO which will be discussed later (see Section 5.6).

Starting from the OFDM overhead we note that based on the 802.11ac standard [46] for channels of 20 MHz bandwidth, only 52 of the 64 subcarriers carry data (the rest are devoted to pilots or are nulled). For 40 MHz and 80 MHz channels the corresponding numbers are 108 out of 128 subcarriers, and 234 out of 256 subcarriers respectively (notice that these overheads slightly vary for 802.11n and 802.11ac, the numbers of the latter are adopted for reasons of simplicity). Finally, the OFDM Cyclic Prefix (CP) is included in the Guard Interval (GI) that is prepended to every OFDM symbol for a total overhead of an extra 20% (assuming a normal GI of 0.8 μs with total symbol duration of 4 μs).

As a more realistic approach to the rate calculation, we map the received SINRs into a discrete set of modulation and coding pairs. While we acknowledge that choosing the best among several discrete modulation and coding options (known as rate adaptation) is non-trivial, for the sake of simplicity we assume that we can choose the best scheme based on the received SINR. Table 2 provides one such mapping that corresponds to the 9 mandatory MCSs of 802.11ac [46], keeping in mind that mappings vary by vendor or may be dynamically chosen in practical scenarios.

The scenarios presented in the rest of the section cover typical yet interesting cases of WiFi deployments, along with more challenging situations. Specifically: i) a densely populated conference hall, ii) an open-floor plan office space, iii) an office building floor with separated rooms and iv) an event stadium with high capacity. For all scenarios 4 antennas per AP are assumed (the maximum number allowed by the 802.11n standard and supported by most 802.11ac implementations to date) and the UTs are equipped with 1 antenna, as is the typical case for handheld devices.

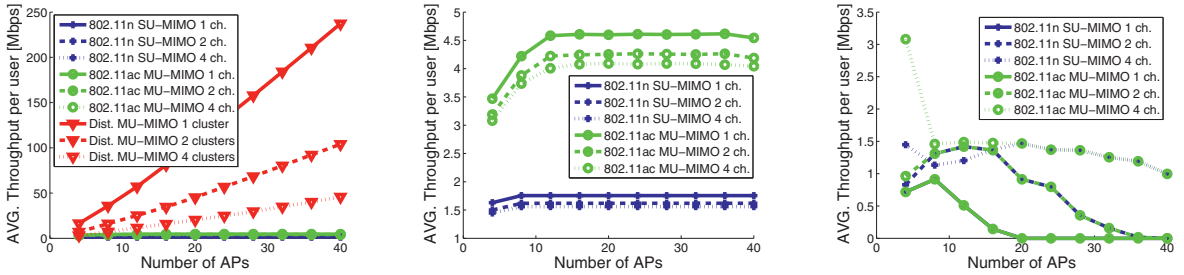
5.2. Conference hall

A conference hall of dimension 20 m$\times$20 m with 200 users and a varying number of APs placed in a canonical fashion (see Fig. 5a) is examined. Fig. 6a plots the average throughput per user against the number of APs. [6] It is obvious that in this scenario, a coordinated solution where interference is suppressed through the coordinated MU-MIMO system is highly favorable. As the number of APs increases, the average user throughput increases for the coordinated MU-MIMO technology. The greatest gains come if we utilize the whole bandwidth as a single channel and let both APs and user receivers transmit and receive in the whole 80 MHz band. Nevertheless, even in the case where we cluster neighboring APs to act as a single AP, assign a 20 MHz channel to each cluster, and assign each user to a single cluster (see Section 3.5) the gains are still enormous compared to the non-coordinated technologies. It is worth noting that in these scenario, due to the symmetry of the user and AP placement the standard deviation between the achievable rates of different users is very small, giving a variance coefficient (standard deviation over the mean) lower than 0.3 and a Jain's fairness index greater than 95% implying a rather fair rate allocation.

It is notable that even under the idealized CSMA scenario where collisions are assumed to be insignificant, the throughputs of 802.11n SU-MISO and 802.11ac MU-MIMO are quickly saturated. Thus, increasing the number of APs does not provide any extra gains as can be seen in Fig. 6b. Also, the small gains in the figure that different channelization options produce to non-coordinated systems under a strict CCA threshold are due to the reduced OFDM overheads when a smaller number of larger channels is used (see Section 5.1). Notice that this is only true under the assumption of no collisions.

CSMA impact as number of APs increases: In Fig. 6b and c we see the average per user throughput for different channelization options for the case of CSMA with the default CCA threshold of 10 dB above the noise floor and an unbounded CCA threshold (no CSMA present). It is worthwhile to notice that 802.11ac MU-MIMO, being more susceptible to interference, performs significantly worse under the regime of no CSMA. It performs similar to 802.11n's robust conjugate beamforming gains as the number of interferers grows (Fig. 6c). Indeed, in this interference limited regime, the MU-MIMO system selects only one user to be served

[6] To examine the network performance for a wide range of parameter values, e.g. number of APs, we will be reporting averages, e.g. average throughput, and, at times, variances and Jain's fairness index [58], rather than showing the complete CDF as we did in Figs. 2 and 3.

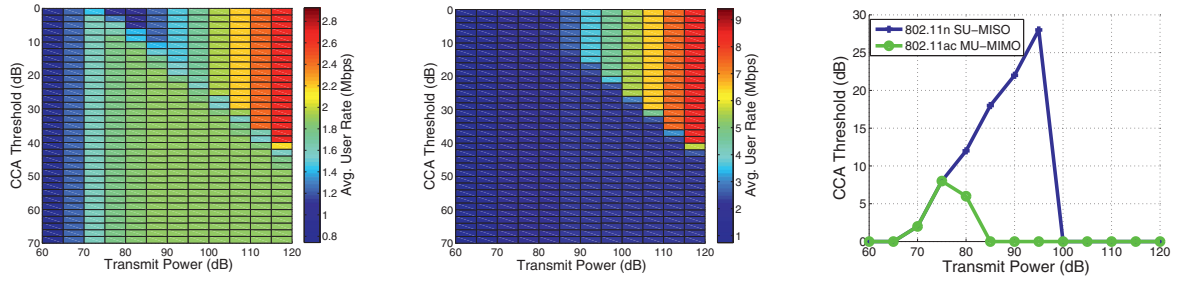


(a) All technologies with CCA = 10dB

(b) 802.11n and 802.11ac with CCA = 10dB

(c) 802.11n and 802.11ac with no CCA

Fig. 6. Average throughput per user in a conference hall with 200 users varying CCA thresholds and 90 dB transmit power.



(a) 802.11n SU-MISO

(b) 802.11ac MU-MIMO

(c) Minimum optimal CCA threshold for a range of transmit powers from 60 to 120 dB

Fig. 7. CCA threshold and transmit power optimization for 802.11n SU-MISO and 802.11ac MU-MIMO.

at a time, reverting to a SU- MISO scheme. On the other hand, as can be seen in Fig. 6b, using CSMA random access clearly benefits the MU-MIMO scheme as the multiplexing gains are larger. Finally, in terms of fairness between the users, when CSMA is turned off, as the number of APs grows, only a few users manage to have a measurable service rate and thus fairness degrades, and the variance coefficient of the user rates distribution grows to values larger than 1.

CCA threshold and power control: Fig. 7a and b show the change in the average user throughput as power increases and the CCA (Clear Channel Assessment) threshold is changed for 802.11n SU-MISO and 802.11ac MU-MIMO. For this optimization, Gaussian rates have been assumed instead of the discretized rate allocation we introduced. Also, a 4 channel system with 20 MHz per channel and 20 APs serving 200 users is assumed for these computations. We can see from these plots that SU-MISO is more robust to interference as the power of all APs increases, whereas for MU-MIMO to perform with a noticeable multiplexing gain a lower CCA threshold has to be chosen. In Fig. 7c we see the optimal CCA threshold for a range of transmit powers from 60 to 120 dB for the two non-coordinated technologies. Notice that thanks to the model we have introduced, such an optimization can be quickly performed and trade-offs can be efficiently explored.

User/AP association: In Fig. 8 we compare the throughput gains of different user/AP association schemes. For the purposes of this simulation a non-uniform user distribution (according to which users were concentrated towards the one half of the floor plan) was selected in order to exemplify the merits of different association schemes. We compare the simple high RSSI scheme with the "available capacity" scheme, introduced in Section 3.6. The gains of a more sophisticated association metric are pronounced in non-uniform user distributions. Specifically, for the SU-MISO case gains of 53% are realized whereas for the MU-MIMO and the coordinated MU-MIMO schemes the gains are 77% and 24% respectively.

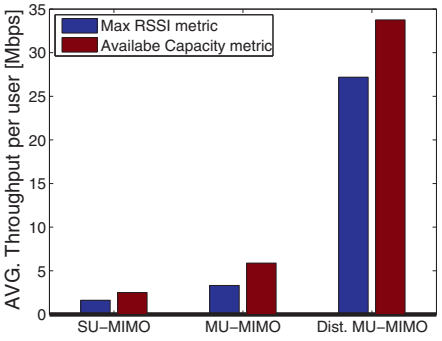


Fig. 8. User/AP association metrics comparison for a conference hall with 200 users and 20 APs.

Sectorization gains: In the previous results we saw that the coordinated MU-MIMO technology significantly outperforms the presented non-coordinated schemes that quickly become interference limited as the number of APs grows. To mitigate the interference between APs that operate on the same channel we can use a *sectorized* deployment. Such deployments can commonly be found in cellular networks, where the antennas used have a sub-circular sector shaped radiation pattern [59], and simplified variations are slowly making their appearance in enterprise WiFi deployments. We present here a simple scheme to illustrate the gains of mitigating the interference for non-coordinated technologies with such an approach. We shape the radiation patterns of all APs in the conference hall topology (Fig. 5a) such that they only radiate energy to a 90^o sector, oriented towards the same direction for all APs. This simple approach significantly improves the performance of SU-MISO 802.11n and MU-MIMO 802.11ac technologies as can be seen in Fig. 9, by a factor of 3x and 7x respectively.

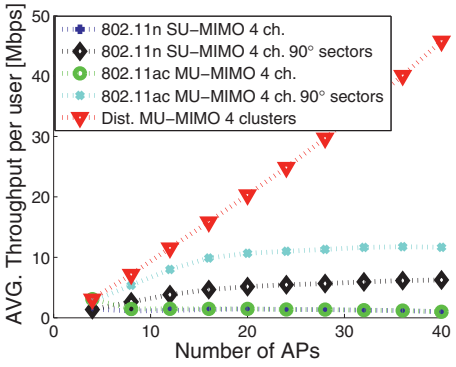


Fig. 9. Average throughput per user in a conference hall with and without sectorization with transmit power of 90 dB and no CCA.

Note that coordinated MU-MIMO, which uses omnidirectional antennas, still performs significantly better, especially in a 1 cluster setup. (Only the 4 cluster setup is shown in this plot to better distinguish the performance of the other schemes). As expected, this simple sectorized approach has an impact on the fairness of the network, lowering the fairness index more than 20% in some occasions.

5.3. Open-floor office building

The open-floor office scenario shares a few common characteristics with the conference hall but has some interesting differences. We define an office floor of dimensions 160 m × 23 m that is occupied by cubicles (essentially no walls obscure the line of sight from AP-to-AP or AP-to-user). APs are placed equally spaced in two rows and users are scattered uniformly and at random in the whole floor (see Fig. 5b). The coordinated MU-MIMO system clearly outperforms the non-coordinated solution in terms of average per user throughput as can be seen in Fig. 10a. It is interesting to notice that in contrast to the previous case, increasing the number of APs even up to 40 gives some gains for the non-coordinated systems under the CSMA scheduling scenario (see Fig. 10b). From Fig. 10c it is apparent that when the CCA threshold is removed all together, the regime becomes interference limited as before. This can be seen by the rate degradation of the single channel system when adding more than 20 APs in the topology. Finally, the fairness index is very high for the coordinated MU-MIMO case, and remains lower than 60% for the SU-MISO and MU-MIMO schemes in most settings with more than 8 APs.

At this point we must emphasize the impact of using quantized rates in comparison to the information theoretic Gaussian rates. Although the latter offer a starting point to solve complex optimization problems and attain intuition for physical layer effects, they might lead to misleading intuition if not combined with an actual quantized mapping. Specifically, the scenarios with no CCA threshold (see Figs. 10c and 6c) would look significantly different in their respective Gaussian versions. The rates would still saturate, but even at very low SINR, when the MCS mapping would give rates equal to zero, the Gaussian rate would assign some positive rate to every user. Further, increasing the number of APs increases the number of users served simultaneously (up to some limit), so the use of Gaussian rates leads to computing a higher average rate per user compared to the negligible rate a quantized scheme would achieve.

5.4. Office building floor with rooms

A typical office building floor with two rows of rooms and a corridor in between them (see Fig. 5c) is analyzed with our model. The length of the floor is 160 m, the width of each room is 10 m, and the corridor width is 3 m. For this series of experiments we fix the number of APs to 20 and users to 200 and investigate the behavior of the network as the number of rooms is varied. APs are placed in a canonical way to cover the whole area and users are scattered uniformly at random in the whole floor.

The increase of the number of rooms, while the floor dimensions and number of APs remains the same, essentially adds additional barriers (walls) between APs. This, in return, effectively increases the pathloss between APs that share the same channel and thus significantly lowers the interference. On the other hand, the users fall out of line-of-sight with more APs, and in fact their pathlosses to many of the APs substantially increase due to the walls in between them. Thus, the multiplexing gains of coordinated MU-MIMO disappear shortly after the number of rooms increases.

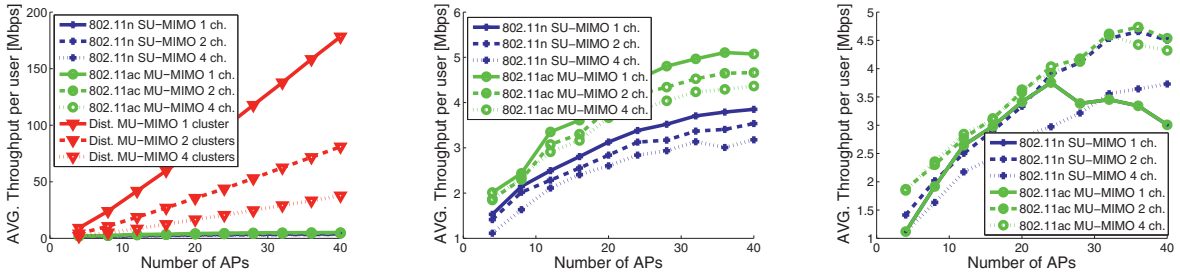
In Fig. 11 we can see that as the APs become better separated by adding walls between them, the gains of coordinated MU-MIMO are negligible and eventually trail behind the localized MU-MIMO scheme. This is expected since in this scenario, only the gain with the AP assigned to each user is significant and thus, trying to beam form to users from distant APs, that are not in their transmit range, actually hurts the system instead of providing multiplexing gains.

Finally, in terms of fairness, Jain's fairness index indicates that coordinated MU-MIMO achieves fairness larger than 90% for most of the topologies, with it dropping lower from 85% only in the highly spatially decoupled cases of a large number of rooms. On the other hand, both SU-MISO and local MU-MIMO suffer from lower fairness indexes, that go to as low as 50%. The non-coordinated technologies benefit in terms of fairness as the number of rooms increases moderately whereas the coordinated MU-MIMO system starts losing it's efficiency to serve all users equally.

5.5. Stadium

Having the analytic model makes it easy and fast to compute the expected average user throughput in even larger scenarios as would be in the case of a football stadium (with a radius of 100 m) during a big game (see Fig. 5d for topology and channel assignment). We can imagine that at moments of high interest, for instance during a questionable refereeing decision or a touchdown, a large percentage of the in stadium fans might want to see a replay or a different camera feed online. In Fig. 12a we can see the average throughput per user for a stadium of dimensions 200 m × 200 m, with 20000 users trying to access the internet and a range of APs. It is obvious in this case that only coordinated MU-MIMO has the potential to provide an acceptable user service under such a challenging yet typical scenario. Moreover, if the CCA threshold is completely removed we again end up in an interference limited scenario where non-coordinated technologies suffer from rate saturation/degradation as the number of APs increases (see Fig. 12b). Note that in this scenario the fairness index remains in most occasions lower than 80% for the non-coordinated MU-MIMO system and lower than 50% for the SU-MISO. It grows to values larger than 90% when activating CSMA.

In such a large deployment, coordinating hundreds of APs at the level of accuracy the coordinated MU-MIMO architecture requires may be impossible in practice. Thus, we consider clustering only a small number of closely located APs in a coordinated MU-MIMO cluster, and operating a large number of such clusters independently to cover the whole area. Clearly, clusters operating in the same frequency will interfere with each other. In Fig. 12c we plot the performance of coordinated MU-MIMO when APs are grouped in small clusters consisting of neighboring APs. As expected, the performance is lower than that of a system coordinating all the APs together (compare Fig. 12a and c). That said, it outperforms uncoordinated MU-MIMO. It is interesting to explain the trend in the



(a) All technologies with CCA = 10dB.

(b) 802.11n SU-MISO and 802.11ac MU-MIMO with CCA = 10dB.

(c) 802.11n SU-MISO and 802.11ac MU-MIMO with no CCA.

Fig. 10. Average throughput per user for an open-floor office plan with transmit power of 90 dB.

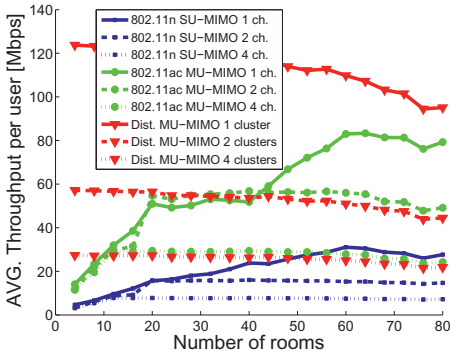


Fig. 11. Average throughput per user for an office building with rooms with transmit power of 90dB and CCA = 10 dB.

throughput curves in Fig. 12c. Taking as an example the clusters of 20 APs, when we have 100 APs (5 clusters in total) only 2 clusters interfere and since this is a stadium the two clusters that interfere are diametrically opposite to each other making the interference small. With 140 APs (7 clusters) 6 clusters interfere (in pairs of two) and so 3 channels experience interference which outbalances the increase in APs and hence the drop in the average throughput. With 180 APs (9 clusters) all channels experience interference and now interfering clusters can no longer be diametrically opposite in pairs so interference is increased significantly compared to the previous case. After 180 APs the increase in APs provides gains since we were already in the regime that all clusters experienced interference.

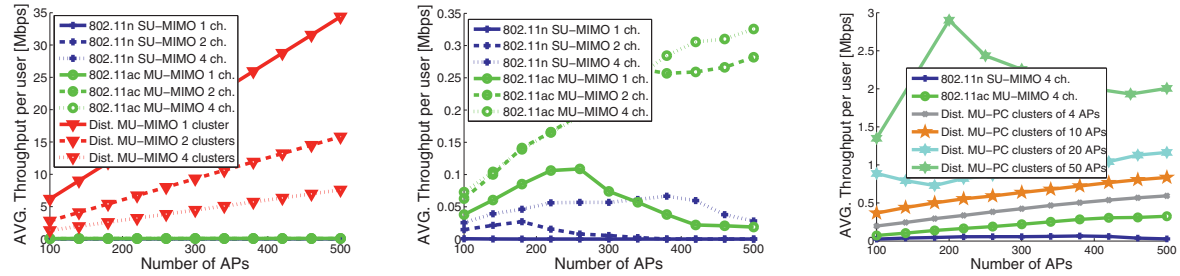
5.6. Signaling overhead

CSIT overhead: It is worth noting that all the PHY schemes considered require the knowledge of channel state information at the transmitter side (CSIT) [38]. The overhead of the collection of CSIT exists and is similar in all aforementioned systems, but differs per topology and thus can only be taken into account in a per-scenario basis, it is thus not included in the analytic model. It is however straightforward to compute and discount for such an overhead based on the 802.11n/ac standards [46] and on some recent work of ours [39].

Coordinated MU-MIMO overhead: Coordinated MU-MIMO requires additional overhead to coordinate the APs in the cluster and allow them to operate as a single, virtual AP. In particular, timing and frequency synchronization (necessary for coherent coordinated ZFBF) and uplink-downlink reciprocity (necessary to learn the downlink channel from the users uplink data packets) must be implemented through an efficient synchronization and calibration protocol. These problems have been recently addressed in depth in our work [39], as well as in a number of publications reporting software-defined radio prototype implementations of coordinated MU-MIMO by us [3] and others [4,51].

Practical synchronization between the APs is imperfect and residual carrier frequency offsets (CFOs) lead to a rate degradation that gets progressively worse over the course of transmission. Developing an analytical characterization of the tradeoff between the achievable rates and the synchronization overhead is difficult, since it is highly dependent upon the selection of “anchor” or leader nodes, and it involves non-convex optimization [39]. As such, this problem can be solved on a per-topology basis using simulations.

In addition to synchronization overhead, coordinated MU-MIMO utilizes calibration between the APs in order to achieve uplink-downlink reciprocity [39,60]. In general, for a given network con-



(a) All technologies with CCA = 2dB.

(b) 802.11n SU-MISO and 802.11ac MU-MIMO with no CCA.

(c) Avg. throughput per user for a stadium with small clusters of neighboring APs.

Fig. 12. Average throughput per user for a stadium with 20000 users and transmit power of 90 dB.

figuration and coordinated MU-MIMO clustering, it is possible to optimize the overhead incurred by synchronization and calibration off-line, and apply this overhead as a discount factor to calculate the goodput of the system.

6. Conclusions

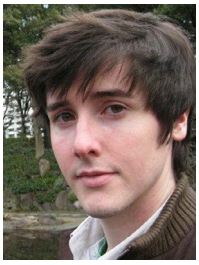
In this paper we have introduced an accurate and practical analytical model for next generation WiFi networks and applied it in a variety of real-world scenarios. An important result from this study is the significant performance gains that coordinated MU-MIMO has over non-coordinated approaches. That said, coordinated MU-MIMO incurs additional overhead, and its gains are smaller in the presence of walls and other barriers which reduce inter-cell interference, as well as when considering a cap on the total number of APs that can be efficiently coordinated. Other interesting results are the sizable gains from sectorization, though to achieve those gains one needs to use front ends of varying complexity and cost, as well as how fast non-coordinated approaches become interference-limited resulting in no additional gains as more APs are deployed. As a final point, it is evident from the analysis of the various practical scenarios in this work that our model can be effectively used to guide the deployment of future wireless networks and optimize their performance.

References

[1] Cisco, Cisco wireless LAN controller configuration guide, Technical Report, Cisco Systems, Inc., 2013.
[2] Aruba, ArubaOS 6.2, user guide, Technical Report, Aruba Networks, Inc., 2013.
[3] H.V. Balan, R. Rogalin, A. Michaloliakos, K. Psounis, G. Caire, AirSync: enabling distributed multiuser MIMO with full spatial multiplexing, IEEE/ACM Trans. Netw. PP (99) (2013), doi:10.1109/TNET.2012.2230449. 1–1
[4] H.S. Rahul, S. Kumar, D. Katabi, JMB: Scaling wireless capacity with user demands, ACM SIGCOMM (2012), doi:10.1145/2377677.2377722.
[5] G. Bianchi, Performance analysis of the IEEE 802.11 distributed coordination function, IEEE J. Sel. Areas Commun. 18 (3) (2000) 535–547, doi:10.1109/49.840210.
[6] E. Ziouva, T. Antonakopoulos, CSMA/CA performance under high traffic conditions: throughput and delay analysis, Comput. Commun. 25 (3) (2002) 313–321.
[7] R. Boorstyn, A. Kershenbaum, B. Maglaris, V. Sahin, Throughput analysis in multihop CSMA packet radio networks, IEEE Trans. Commun. 35 (3) (1987) 267–274, doi:10.1109/TCOM.1987.1096769.
[8] S.-C. Liew, C.H. Kai, H.C. Leung, P. Wong, Back-of-the-envelope computation of throughput distributions in CSMA wireless networks, IEEE Trans. Mobile Comput. 9 (9) (2010) 1319–1331, doi:10.1109/TMC.2010.89.
[9] L. Jiang, J. Walrand, A distributed CSMA algorithm for throughput and utility maximization in wireless networks, IEEE/ACM Trans. Netw. 18 (3) (2010) 960–972, doi:10.1109/TNET.2009.2035046.
[10] X. Wang, K. Kar, Throughput modelling and fairness issues in CSMA/CA based ad-hoc networks, IEEE INFOCOM, 2005, doi:10.1109/INFCOM.2005.1497875.
[11] M. Haenggi, Mean interference in hard-core wireless networks, IEEE Commun. Lett. 15 (8) (2011) 792–794, doi:10.1109/LCOMM.2011.061611.110960.
[12] H. ElSawy, E. Hossain, A modified hard core point process for analysis of random CSMA wireless networks in general fading environments, IEEE Trans. Commun. 61 (4) (2013) 1520–1534, doi:10.1109/TCOMM.2013.020813.120594.
[13] A. Jindal, K. Psounis, On the efficiency of CSMA-CA scheduling in wireless multihop networks, IEEE/ACM Trans. Netw. 21 (5) (2013) 1392–1406, doi:10.1109/TNET.2012.2225843.
[14] A. Jindal, K. Psounis, The achievable rate region of 802.11-scheduled multihop networks, IEEE/ACM Trans. Netw. 17 (4) (2009) 1118–1131, doi:10.1109/TNET.2008.2007844.
[15] H. Huh, A. Tulino, G. Caire, Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling, IEEE Trans. Inf. Theory 58 (5) (2012) 2911–2934, doi:10.1109/TIT.2011.2178230.
[16] H. Huh, G. Caire, H. Papadopoulos, S. Ramprashad, Achieving massive MIMO spectral efficiency with a not-so-large number of antennas, IEEE Trans. Wireless Commun. 11 (9) (2012) 3226–3239, doi:10.1109/TWC.2012.070912.111383.
[17] S. Ramprashad, G. Caire, Cellular vs. network MIMO: A comparison including the channel state information overhead, IEEE PIMRC, 2009, doi:10.1109/PIMRC.2009.5450085.
[18] S. Lakshminaryana, J. Hoydis, M. Debbah, M. Assaad, Asymptotic analysis of distributed multi-cell beamforming, IEEE PIMRC, 2010, doi:10.1109/PIMRC.2010.5671653.
[19] F. Rusek, D. Persson, B.K. Lau, E. Larsson, T. Marzetta, O. Edfors, F. Tufvesson, Scaling up MIMO: Opportunities and challenges with very large arrays, IEEE Signal Process. Mag. 30 (1) (2013) 40–60, doi:10.1109/MSP.2011.2178495.
[20] A. Castiglione, F. Palmieri, U. Fiore, A. Castiglione, A.D. Santis, Modeling energy-efficient secure communications in multi-mode wireless mobile devices, J. Comput. Syst. Sci. 81 (8) (2015) 1464–1478. http://dx.doi.org/10.1016/j.jcss.2014.12.022.
[21] J. Hoydis, S. Ten Brink, M. Debbah, Massive MIMO in the UL/DL of cellular networks: How many antennas do we need? IEEE J. Sel. Areas Commun. 31 (2) (2013) 160–171, doi:10.1109/JSAC.2013.130205.
[22] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, W. Yu, Multi-Cell MIMO cooperative networks: A new look at interference, IEEE J. Sel. Areas Commun. 28 (9) (2010) 1380–1408, doi:10.1109/JSAC.2010.101202.
[23] R. Zakhour, D. Gesbert, Optimized data sharing in multicell MIMO with finite backhaul capacity, IEEE Trans. Signal Process. 59 (12) (2011) 6102–6111, doi:10.1109/TSP.2011.2165949.
[24] H. ElSawy, E. Hossain, M. Haenggi, Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey, IEEE Commun. Surveys Tuts. 15 (3) (2013) 996–1019, doi:10.1109/SURV.2013.052213.00000.
[25] M. Haenggi, J. Andrews, F. Baccelli, O. Dousse, M. Franceschetti, Stochastic geometry and random graphs for the analysis and design of wireless networks, IEEE J. Sel. Areas Commun. 27 (7) (2009) 1029–1046, doi:10.1109/JSAC.2009.090902.
[26] A.K. Gupta, H.S. Dhillon, S. Vishwanath, J.G. Andrews, Downlink MIMO HetNets with load balancing, CoRR abs/1310.6795 (2013).
[27] H. Dhillon, M. Kountouris, J. Andrews, Downlink coverage probability in MIMO hetnets, in: IEEE ASILOMAR, 2012, pp. 683–687, doi:10.1109/ACSSC.2012.6489098.
[28] F. Baccelli, A. Giovanidis, A stochastic geometry framework for analyzing pairwise-cooperative cellular networks, IEEE Trans. Wireless Commun. 14 (2) (2015) 794–808, doi:10.1109/TWC.2014.2360196.
[29] R. Tanbourgi, S. Singh, J. Andrews, F. Jondral, A tractable model for noncoherent joint-transmission base station cooperation, IEEE Trans. Wireless Commun. 13 (9) (2014) 4959–4973, doi:10.1109/TWC.2014.2340860.
[30] S. Akoum, R. Heath, Multi-cell coordination: A stochastic geometry approach, in: IEEE SPAWC, 2012, pp. 16–20, doi:10.1109/SPAWC.2012.6292881.
[31] K. Huang, J. Andrews, An analytical framework for multicell cooperation via stochastic geometry and large deviations, IEEE Trans. Inf. Theory 59 (4) (2013) 2501–2516, doi:10.1109/TIT.2012.2232966.
[32] D.H. Kang, K.W. Sung, J. Zander, Cost efficient high capacity indoor wireless access: Denser Wi-Fi or coordinated pico-cellular? CoRR abs/1211.4392 (2012).
[33] Y. Liang, A. Goldsmith, G. Foschini, R. Valenzuela, D. Chizhik, Evolution of base stations in cellular networks: Denser deployment versus coordination, IEEE ICC, 2008, doi:10.1109/ICC.2008.775.
[34] Airmagnet, Best practices for wireless site design, Technical Report, AirMagnet, Inc., 2007.
[35] T. Issariyakul, E. Hossain, Introduction to Network Simulator NS2; 2nd ed., Springer, Dordrecht, 2012.
[36] M. Garetto, T. Salonidis, E.W. Knightly, Modeling per-flow throughput and capturing starvation in CSMA multi-hop wireless networks, IEEE/ACM Trans. Netw. 16 (4) (2008) 864–877, doi:10.1109/TNET.2007.902687.
[37] A. Kershenbaum, R. Boorstyn, M. Chen, An algorithm for evaluation of throughput in multihop packet radio networks with complex topologies, IEEE J. Sel. Areas Commun. 5 (6) (1987) 1003–1012, doi:10.1109/JSAC.1987.1146614.
[38] A. Molisch, Wireless communications, Wiley-IEEE, 2005.
[39] R. Rogalin, O. Bursalioglu, H. Papadopoulos, G. Caire, A. Molisch, A. Michaloliakos, V. Balan, K. Psounis, Scalable synchronization and reciprocity calibration for distributed multiuser MIMO, IEEE Trans. Wireless Commun. 13 (4) (2014) 1815–1831, doi:10.1109/TWC.2014.030314.130474.
[40] P. Kyösti, et. al, WINNER II Channel Models, Technical Report, Information Society Technologies, 2007.
[41] G. Foschini, M. Gans, On limits of wireless communications in a fading environment when using multiple antennas, Wireless Pers. Commun. 6 (1998) 311–335.
[42] E. Telatar, Capacity of multi-antenna Gaussian channels, Eur. T. Telecommun. 10 (6) (1999) 585–595, doi:10.1002/ett.4460100604.
[43] G. Caire, S. Shamai, On the achievable throughput of a multiantenna Gaussian broadcast channel, IEEE Trans. Inf. Theory 49 (7) (2003) 1691–1706, doi:10.1109/TIT.2003.813523.
[44] G. Caire, N. Jindal, M. Kobayashi, N. Ravindran, Multiuser MIMO Achievable Rates With Downlink Training and Channel State Feedback, IEEE Trans. Inf. Theory 56 (6) (2010) 2845–2866, doi:10.1109/TIT.2010.2046225.
[45] B. Hochwald, T. Marzetta, V. Tarokh, Multiple-antenna channel hardening and its implications for rate feedback and scheduling, IEEE Trans. Inf. Theory 50 (9) (2004) 1893–1909, doi:10.1109/TIT.2004.833345.
[46] IEEE Computer Society, IEEE Standard for IT - Telecommunications and Information Exchange Between Systems - LAN/MAN - Specific Requirements - Part 11: Wireless LAN MAC and PHY Layer Specifications - Amd 4: Enhancements for Very High Throughput for operation in bands below 6GHz, IEEE Std 802.11ac-2013.
[47] R. Couillet, M. Debbah, Random Matrix Methods for Wireless Communications, Cambridge University Press, 2011.
[48] K.-K. Wong, Z. Pan, Array gain and diversity order of multiuser MISO antenna systems, Int. J. of Wirel. Inf. Netw. 15 (2) (2008) 82–89, doi:10.1007/s10776-008-0078-5.
[49] N. Anand, J. Lee, S.-J. Lee, E. Knightly, Mode and user selection for multi-user MIMO WLANs without CSI, in: IEEE INFOCOM, 2015, pp. 451–459, doi:10.1109/INFOCOM.2015.7218411.

[50] H. Huang, M. Trivellato, A. Hottinen, M. Shafi, P. Smith, R. Valenzuela, Increasing downlink cellular throughput with limited network MIMO coordination, IEEE Trans. Wireless Commun. 8 (6) (2009) 2983–2989, doi:10.1109/TWC.2009.080179.
[51] X. Zhang, K. Sundaresan, M.A.A. Khojastepour, S. Rangarajan, K.G. Shin, NEMOx: Scalable Network MIMO for Wireless Networks, ACM MobiCom, 2013, doi:10.1145/2500423.2500445. New York, NY
[52] Y. Bejerano, S.-J. Han, L. Li, Fairness and load balancing in wireless LANs using association control, IEEE/ACM Trans. Netw. 15 (3) (2007) 560–573, doi:10.1109/TNET.2007.893680.
[53] J.G. Andrews, S. Singh, Q. Ye, X. Lin, H.S. Dhillon, An overview of load balancing in HetNets: Old myths and open problems, CoRR abs/1307.7779 (2013).
[54] H. Dhillon, R. Ganti, F. Baccelli, J. Andrews, Modeling and analysis of k-tier downlink heterogeneous cellular networks, IEEE J. Sel. Areas Commun. 30 (3) (2012) 550–560, doi:10.1109/JSAC.2012.120405.
[55] R. Murty, J. Padhye, R. Chandra, A. Wolman, B. Zill, Designing high performance enterprise Wi-Fi networks, USENIX NSDI, 2008. Berkeley, CA
[56] H. Huh, H. Papadopoulos, G. Caire, Multiuser MISO transmitter optimization for intercell interference mitigation, IEEE Trans. Signal Process. 58 (8) (2010) 4272–4285, doi:10.1109/TSP.2010.2048564.
[57] A. Gupta, J. Min, I. Rhee, Wifox: scaling WiFi performance for large audience environments, in: CoNEXT, ACM, New York, NY, USA, 2012, pp. 217–228, doi:10.1145/2413176.2413202.
[58] R.K. Jain, D.-M.W. Chiu, W.R. Hawe, A Quantitative Measure Of Fairness And Discrimination For Resource Allocation In Shared Computer Systems, Technical Report, Digital Equipment Corporation, 1984.
[59] K. Gilhousen, I. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver, C.E. Wheatley, On the capacity of a cellular CDMA system, IEEE Trans. Veh. Technol. 40 (2) (1991) 303–312, doi:10.1109/25.289411.
[60] C. Shepard, H. Yu, N. Anand, L. Li, T. Marzetta, R. Yang, L. Zhong, Argos: Practical many-antenna base stations, ACM SIGCOMM, 2012.

Antonios Michaloliakos graduated in 2009 from the department of Electrical and Computer Engineering of the National Technical University of Athens. He is currently pursuing a PhD in Electrical Engineering at the University of Southern California. His interests span the areas of probabilistic modeling and design of networks and cross-layer optimization in new generation wireless networks. He is currently working on developing a MAC protocol for a MU-MIMO system.



Ryan Rogalin received his BS degree in Electrical Engineering summa cum laude from the University of Oklahoma in 2008. In 2010 he received his MS degree, also in Electrical Engineering, from the University of Southern California. In 2015 he graduated with a PhD in Electrical Engineering from the University of Southern California as an Annenberg Fellow. His research interests lie at the intersection of Communication Theory, Information Theory, wireless networks and their implementation through software defined radio.

Yonglong Zhang graduated from the department of Electrical and Computer Engineering of the cole SupŽrieure des Sciences et Technologies de l'IngŽnieur de Nancy in 2012. He is a PhD student in Electrical Engineering at the University of Southern California advised by Konstantinos Psounis. His research interests are performance analyzing of computer networks and WLAN resource management. He is currently working on user-AP association protocol for enterprise Wi-Fi networks.

Konstantinos Psounis is an Associate Professor of Electrical Engineering and Computer Science at the University of Southern California. He received his first degree from the department of Electrical and Computer Engineering of National Technical University of Athens, Greece, in June 1997, and the M.S.and PhD degrees in Electrical Engineering from Stanford University, California, in January 1999 and December 2002 respectively. Konstantinos models and analyzes the performance of a variety of wired and wireless networks. He also designs methods, algorithms, and protocols to solve problems related to such systems. He is a senior member of both IEEE and ACM.

Giuseppe Caire was born in Torino, Italy, in 1965. He received the B.Sc. in Electrical Engineering from Politecnico di Torino (Italy), in 1990, the M.Sc. in Electrical Engineering from Princeton University in 1992 and the Ph.D. from Politecnico di Torino in 1994. He was a recipient of the AEI G.Someda Scholarship in 1991, has been with the European Space Agency (ESTEC, Noordwijk, The Netherlands) from May 1994 to February 1995, was a recipient of the COTRAO Scholarship in 1996 and of a CNR Scholarship in 1997. He has been visiting Princeton University in Summer 1997 and Sydney University in Summer 2000. He has been Assistant Professor in Telecommunications at the Politecnico di Torino, Associate Professor at the University of Parma, Italy, Professor with the Department of Mobile Communications at the Eurecom Institute, Sophia-Antipolis, France, and he is currently a professor of Electrical Engineering with the Viterbi School of Engineering, University of Southern California, Los Angeles, CA and an Alexander von Humboldt Professor with the Electrical Engineering and Computer Science Department of the Technical University of Berlin, Germany.