Fair and Efficient Coexistence of Heterogeneous Channel Widths in Next-Generation Wireless LANs

Sihui Han, Xinyu Zhang, and Kang G. Shin

Abstract—To meet the diverse traffic demands of different applications, emerging WLAN standards have been incorporating a variety of channel widths ranging from 5 to 160 MHz. The coexistence of variable-width channels imposes a new challenge to the 802.11 protocols, since the 802.11 MAC is agnostic of, and thus incapable of, adapting to the PHY-layer spectrum heterogeneity. To address this challenge, we uncover the cause and effect of variable-width channel coexistence, and develop a MAC-layer scheme, called *Fine-grained Spectrum Sharing* (FSS), that solves the general problem of fair and efficient spectrum sharing among users with heterogeneous channel-widths. Instead of deeming its spectrum band as an atomic block, an FSS user divides the spectrum into chunks, and adapts its chunk usage on a per-packet basis. FSS's spectrum adaptation is driven by a decentralized optimization framework. It preserves the 802.11 CSMA/CA primitives while allowing users to contend for each spectrum chunk, and can opportunistically split a wide-band channel or bond multiple (discontiguous) chunks to ensure fair and efficient access to available spectrum. In making such adaptation decisions, it balances the benefit from discontiguous chunks and the cost of guardband – a unique tradeoff in WLANs with heterogeneous channel-widths. Our in-depth evaluation demonstrates that FSS can improve throughput by multiple folds, while maintaining fairness of spectrum sharing for heterogeneous WLANs.

Index Terms—Fine-grained channel access, dynamic spectrum access, heterogeneity, proportional fairness, next-generation WLANs

1 Introduction

The explosive growth of mobile devices and application traffic has been a driving force for high-throughput wireless Internet access. The 802.11ac, a new Gbps WiFi standard, is expected to play a crucial role in meeting such demands in next-generation wireless networks. It can support up to 6.93 Gbps data-rate by integrating advanced communications technologies, such as multi-user MIMO, frame aggregation, and wideband transmission [9]. Meanwhile, the legacy WiFi devices (802.11 a/g/n) will continue their dominant role over the next few years, and are unlikely to fade away because of their unique advantages, such as energy-efficiency and low cost. Therefore, the Gbps wireless devices must maintain compatibility and support coexistence with legacy devices from the early stage of their development.

The distinct PHY-layer spectrum usage of Gbps and legacy WiFi poses a major challenge for their coexistence. The high throughput of 802.11ac mainly comes from its 80 and 160 MHz wideband transmissions, whereas legacy devices commonly adopt 20 MHz channel by default [9]. As wireless spectrum remains a limited/scarce resource, wideband Gbps networks must *partially share* spectrum with narrowband legacy networks as depicted in Fig. 1a. Basic MAC operations in 802.11ac, such as carrier sensing, treat each 160 MHz-channel as an atomic block. Hence, the entire band needs to suspend its transmission even when only part of it is occupied by a legacy narrowband channel. As a result, a wideband 802.11ac network may severely under-utilize its spectrum, and may even be starved when it shares spectrum with multiple narrowband legacy networks. This is a major impediment to the evolution of WiFi toward a Gbps WLAN.

One way to deal with this problem is to allow the wideband transmitters to use the remaining idle part of the spectrum unoccupied by narrowband users. The current 802.11ac standard considers this solution in case when a 20 MHz channel coexists with 40/80 MHz ones. However, the resulting spectrum may be fragmented and discontiguous, leading to new challenges for MAC-layer spectrum sharing.

In particular, how to ensure fair and efficient access to the non-contiguous spectrum remains an open problem. Following the 802.11 CSMA primitives, each transmitter must make its spectrum access decision locally, considering the idle/busy status of each spectrum chunk. Such status, in turn, depends on the decision of neighboring transmitters who may share part or all of the spectrum. Eventually, the local decisions will be interwoven across the entire network, raising a central question for the network stack: "Will the local spectrum access decision of each transmitter lead to network-wide efficiency and fairness?" Besides, to prevent the adjacent channel interference, a guardband is needed between neighboring spectrum used by different transmitters (Fig. 2), which leads to spectrum wastage while providing users more chance to access the spectrum. This tradeoff between guardband cost and spectrum access opportunity must be taken into account in next-generation WLANs with heterogeneous spectrum widths.

- *S. Han and K.G. Shin are with the Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109. E-mail: {sihuihan, kgshin}@umich.edu.*
- *X. Zhang is with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706. E-mail: xyzhang@ece.wisc.edu.*

Manuscript received 5 Apr. 2015; revised 1 Dec. 2015; accepted 7 Jan. 2016. Date of publication 13 Jan. 2016; date of current version 28 Sept. 2016.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TMC.2016.2517654

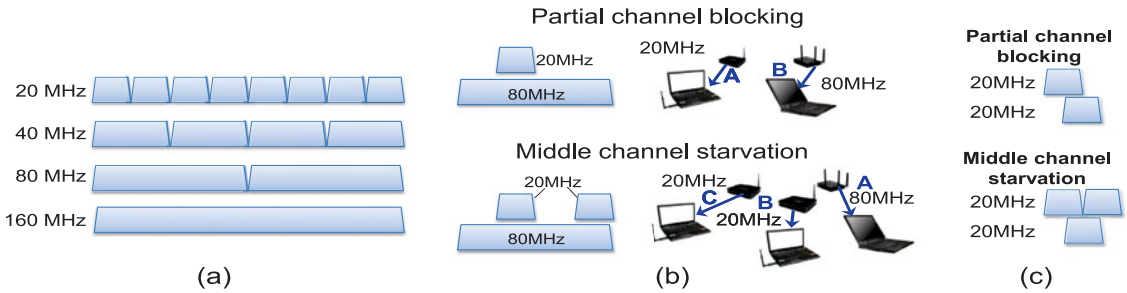

Fig. 1. 802.11 spectrum sharing: (a) Channel overlapping patterns in the 5 GHz band (for 802.11 a/n/ac channels). (b) Problems occurring when narrowband channels coexist with a wideband channel. (c) Similar problems occur in the 2.4 GHz band with partially-overlapping channels.

In this paper, we address the aforementioned problems and make the following main contributions:

- Transforming the problem of heterogeneous WLAN coexistence into a fine-grained spectrum sharing (FSS) problem. Prior work [15], [16], [21] focused on the *PHY-layer* realization of packet transmission through part of the (possibly non-contiguous) available spectrum. In contrast, we design *spectrum management* primitives that allow transmitters to contend and access heterogeneous spectrum segments in a fair and efficient manner.
- Formulation of the spectrum access problem as a utility maximization framework, derivation of a distributed solution, and conversion to a practical executable MAC-layer protocol to manage the spectrum sharing and asynchronous channel access for transmitters. FSS is compatible with 802.11 CSMA and enables WLANs to receive a fair rate of access to spectrum chunks, realizing Gbps networks even in a crowd of low-rate legacy networks, thus solving the coexistence problem.
- Capturing and addressing a fundamental tradeoff in heterogeneous WLANs, i.e., the opportunity from accessing non-contiguous spectrum chunks and the channel wastage due to guardband.

The rest of the paper is organized as follows. Section 2 reviews the current WLAN with variable-width channels. Section 3 presents an overview of FSS, and Section 4 details its design. Section 5 describes our experimental validation and evaluation. Related work is discussed in Section 6. Finally, Section 7 concludes the paper.

2 Background and Motivation

Fundamental limitations of CSMA in heterogeneous spectrum. The current 802.11 relies on CSMA/CA to arbitrate the contention among transmitters on the same channel, but it is not designed to coordinate users with heterogeneous channel widths – a central problem when Gbps and legacy WiFi networks coexist. This severely degrades spectrum efficiency, causing the following MAC-level pathologies.

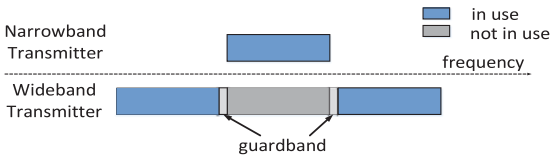

Fig. 2. If coexisting with a narrowband transmitter, a wideband transmitter can use the remaining idle spectrum and place guardband near the spectrum boundaries.

(i) Partial channel blocking: With 802.11, a transmitter is blocked even when only part of its spectrum is occupied by a neighboring transmitter, wasting the remaining idle spectrum and thus resulting in severe spectrum underutilization. In the top example in Fig. 1b, WLANs A and B have 20 and 80 MHz bandwidths, respectively. At any time, either A or B can transmit, but not both. Suppose they transmit packets of the same size, then the transmission takes only one time slot for the 80 MHz WLAN, and four for the 20MHz WLAN. Using 802.11, both WLANs have an equal chance to access the medium, resulting in mean spectrum utilization of $\frac{1}{5}(20 \times 4 + 80 \times 1) = 32\,\text{MHz}$, which is only 40 percent of the nominal utilization (80 MHz). In contrast, the optimal utilization could approach 1 if B could use the remaining idle spectrum during A's transmission.

(ii) Middle-channel starvation: The 802.11 CSMA mechanism may starve a wideband WLAN that resides in the middle of multiple narrowband WLANs. The bottom example of Fig. 1b illustrates the case when an 80 MHz WLAN A partially overlaps with two orthogonal 20 MHz channels B and C. With 802.11, A can transmit only if both B and C are idle, which occurs only when they complete transmissions approximately at the same time, and subsequently A wins the contention over both. This chance occurs rarely if B and C have backlogged traffic, and A will remain starved most of the time, although it nominally has a wider bandwidth and should have a higher throughput. Other simulation studies [10] have also made a similar observation.

Legacy solutions. Prior to the 802.11ac proposal, the IEEE 802.11n had already recognized the importance of managing coexistence of heterogeneous spectrum, especially 40 and 20 MHz channels. A MAC-level patch called Dynamic Channel Allocation (DCA) [18] was proposed, designating the two halves of a 40 MHz channel as *primary* and *secondary* bands, respectively. When the secondary band is occupied, the transmitter switches to the primary band and notifies the receiver. This alleviates the partial channel blocking problem, and has been migrated to 802.11ac to manage the coexistence of 20 and 80/160 MHz channels [9].

In practice, however, most deployments of 802.11n use the 20 MHz-mode, since the effectiveness of the legacy solution is quite limited [11]. First, the wideband channel loses its advantage if either primary or secondary band suffers from intensive contention with a narrow band channel. Second, an 802.11n transmitter can switch from the primary band to the entire band only on a coarse-grained time-scale,

based on the idleness of the secondary band. Third, the solution does not apply to the general case involving 5 MHz/10 MHz spectrum widths. In summary, when narrow- and wide-band networks share spectrum, both bands should ideally have a fair rate of access to the shared spectrum, but current solutions simply reduce the wideband to the narrowband to "avoid" the coexistence of heterogeneous spectrum but not solve the problem. Recent studies such as [12] considered new channel bonding schemes to improve legacy solutions, but none of them considered fair spectrum sharing among heterogeneous WLANs.

Broader implications. The above problem has wider implications than Gbps & legacy WiFi coexistence. In general, whenever two or more channels partially overlap, partial channel blocking and middle-channel starvation can occur. For example, the 2.4 GHz ISM band allows partial overlapping between neighboring 20 MHz channels, which can cause partial channel blocking and middle-channel starvation as well (see Fig. 1c).

Even if future WiFi devices are all capable of wide-band transmission, the use of heterogeneous channel widths will persist. A wider channel increases throughput, but reduces the communication range due to less concentrated energy [20] and also incurs higher energy consumption due to a higher sample rate, whereas a narrowband channel improves energy-efficiency at the cost of reduced capacity and more guardband cost. Instead of fixing at a narrowband or wideband, different applications may prefer different spectrum widths. For instance, an AP might communicate with an energy-limited device (such as smartphones) on a narrowband 5 MHz channel to conserve energy, and communicate with HDTV device on a wider 160 MHz channel to achieve high capacity, even if both devices are compliant with 802.11ac and capable of wide-band transmission. Such heterogeneous user demands will again lead to heterogeneous channel widths.

As heterogenous spectrum sharing becomes a common feature in contemporary and future WLANs, the coexistence problem will become more prevalent. It is therefore imperative to establish a rigorous framework that performs spectrum aggregation to opportunistically create links of high throughput while ensuring fair access to small spectrum chunks.

3 System Model and Solution Overview

3.1 System Model

We consider a network that contains n transmitter-receiver (TX_i, RX_i) pairs, where $i \in \mathcal{N} = \{1, 2, \ldots, n\}$. Each pair may have different channel bandwidths, which could be chosen from the allowable options defined in 802.11 (5, 10, 20, 40, 80 or 160 MHz) [9]. To enable the coexistence between heterogeneous spectra, we divide a transmitter's channel into spectrum chunks, each with bandwidth B. The transmitters can finely share spectrum as shown in the example in Fig. 3. In 802.11 protocols, the minimum channel width is 5 MHz, and the channel center frequencies are specified such that the overlapping between any two channels is a multiple of 5 MHz. Therefore, by default, we set the atomic chunk size B to 5 MHz.

Our PHY-layer model is commensurate with the state-of-the-art OFDM spectrum access technologies [10], [15], [16], [21], which enable per-frame spectrum shaping by simply suppressing certain subcarriers without changing a channel's center frequency, and enable the receivers to sense the spectrum occupied by the frame and decode it accordingly. FSS builds on such PHY technologies that allow simultaneous use of discontiguous spectrum chunks. Our main focus is placed on the MAC-layer problems on top of such PHY-layer solutions.

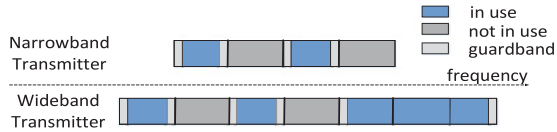

Fig. 3. The wideband and narrowband transmitters share spectrum chunks in a fine-grained (possibly discontiguous) manner.

3.2 Overview

Our MAC design for fine-grained spectrum sharing is driven by a utility maximization framework which guides each transmitter to adjust its spectrum access probability, in order to achieve efficient and fair channel utilization. Let $\mathcal{M} = \{1, 2, \ldots, m\}$ denote the set of spectrum chunks in the network, where m is the number of chunks available. Each TX_i tries to access a spectrum chunk j with probability $p_{i,j}$. If TX_i cannot sense the chunk or transmit through the chunk, $p_{i,j}$ is set to 0. This models transmitters with heterogeneous bandwidths, such as coexisting 802.11a and 802.11ac transmitters with 20 and 160 MHz channel width, respectively.

We further use matrix $\mathbf{p} = (p_{i,j})_{n \times m}$ to concatenate access probability variables and vector $\boldsymbol{p}_i$ to denote the access probability of TX_i, where $\boldsymbol{p}_i = \{p_{i,1}, p_{i,2}, \ldots, p_{i,m}\}$. FSS is a CSMA-based probabilistic contention protocol, but as will be clarified later, it can be translated into a backoff-based 802.11-compatible CSMA/CA protocol. FSS aims to achieve proportional fairness among transmitters (links). Intuitively, the access rate of a transmitter should be proportional to the following two constraints. The first is the maximum number of spectrum chunks it can sense and aggregate. Wideband transmitters should be given a chance to access more spectrum chunks, as they can "see" more chunks and anticipate heavier channel usage than narrowband transmitters. The second is the congestion status of each spectrum chunk. If a transmitter can only aggregate spectrum chunks that are shared by many others, then its throughput should be low accordingly. Following the definition in [5], we can define proportional fairness with respect to access rate in FSS as follows.

Definition (Proportional Fairness). *The access probability matrix* $\mathbf{p} = (p_{i,j})_{n \times m}$ *is* proportionally fair *if it is feasible and if, for any other feasible vector* $\mathbf{p}'$*, the aggregate of proportional changes is 0 or negative:*

$$\sum_{i \in \mathcal{N}} \frac{\sum_{j \in \mathcal{M}} p'_{i,j} - \sum_{j \in \mathcal{M}} p_{i,j}}{\sum_{j \in \mathcal{M}} p_{i,j}} \leq 0. \tag{1}$$

The proportional fairness metric can overcome the limitations of conventional 802.11 WLAN technologies, which suffer from the partial channel blocking and middle-channel starvation problems (Section 2). This metric has been widely used in practical networks (e.g., 3 GPP LTE), and in many

different aspects of network optimization including scheduling [3], [4], rate control [1], [2], and resource allocation [6], [7]. Extending FSS to other fairness metrics such as max-min fairness [8] entails a routine reformulation of the constraints on access rate, and is beyond the scope of this paper.

4 Fair and Efficient Access to Heterogeneous Spectra

The overarching goal of FSS is to ensure optimal and fair channel access for all transmitters. This optimization problem can be cast as:

$$\max_{\mathbf{p}} \quad \sum_{i\in\mathcal{N}} \alpha_i U(\mathbf{p}_i),$$

where $U(\cdot)$ is a utility function. When $U(\mathbf{p}_i) = \log(f_i(\mathbf{p}_i))$, maximizing the sum utility is proven to achieve global optimality *w.r.t.* $\mathbf{p}_i$ and proportional fairness [1]. Thus, we adopt the log utility function in FSS. $f_i(\mathbf{p}_i)$ is a function mapping transmitter i's access probability vector $\mathbf{p}_i$ to its throughput and α_i is the weight of transmitters.

We address three key challenges to achieve the optimization objective: (1) modeling the guardband-related tradeoff in non-contiguous spectrum access, (2) designing a decentralized contention algorithm, and (3) mapping the algorithm to a practical 802.11-compatible MAC protocol.

4.1 Contiguous versus Non-Contiguous Access

The 802.11 standards require transmitters to access contiguous spectrum, and guardband is needed on the leftmost and rightmost sides of a spectrum block to prevent interference to/from adjacent channels. For contiguous spectrum access, the optimization problem becomes:

$$\max_{\mathbf{p}} \quad \sum_{i\in\mathcal{N}} \alpha_i \log(Eb_i), \tag{2}$$

$$\text{s.t.} \quad \omega \sum_{j=1}^{m-1} |p_{i,j+1} - p_{i,j}| = 0, \tag{3}$$

where Eb_i is the expected throughput of TX_i, which can be modeled as $Eb_i = B\sum_{j=1}^{m} p_{i,j} - \omega_l p_{i,1} - \omega_r p_{i,m}$. ω_l and ω_r denote the left and right guardband size, respectively, and $\omega = \omega_l + \omega_r$. Constraint (3) models the fact that 802.11 deems an entire channel as an atomic block: all constituent spectrum chunks within that channel have the same access probability (i.e., $p_{i,j} = p_{i,j+1}$ for all $j \in [0, m-1]$), and guardband is needed on the left and right edges, but not in between.

In the widely used OFDM-based 802.11 standards (e.g., IEEE 802.11 a/g), a channel comprises 64 small spectrum units called *subcarriers*. The default guardband size is 6 subcarriers on the left and 5 on the right edge of the channel spectrum band [9]. FSS adopts this standard guardband size, such that $\omega_l = \frac{6}{64}B$ and $\omega_r = \frac{5}{64}B$ for each spectrum chunk.

The term $\omega\sum_{j=1}^{m-1}|p_{i,j+1} - p_{i,j}|$ in constraint (3) can model the guardband cost in the general case of *non-contiguous channel access*. To understand this better, let us consider a simple case, where the access probability can be either 0 or 1. Suppose TX_i accesses four spectrum chunks with probability $[0, 1, 0, 1]$, respectively. The guardband cost is $\omega\sum_{j=1}^{m-1}|p_{i,j+1} - p_{i,j}| = \omega$. Assuming the access probability is $[0, 1, 1, 0]$, the guardband cost is $\omega\sum_{j=1}^{m-1}|p_{i,j+1} - p_{i,j}| = 2\omega$, which increases since the spectrum becomes more fragmented.

More generally, when the access probability falls in between 0 and 1, we can consider the term as the expectation of the guardband cost. If the adjacent chunks j and $j+1$ have similar characteristics (i.e., the access probabilities are close to each other), then they are very likely to be accessed together with no guardband in between. Accordingly, the probability of needing extra guardband is small, yielding a small guardband cost $\omega|p_{i,j+1} - p_{i,j}|$. A zero guardband cost corresponds to contiguous access when all chunks have the same access probability.

Contiguous spectrum access limits the chunks transmitters could choose, and decreases the access probability. Even worse, the transmitters may have to contend for chunks under severe contention as they have no other choice. It imposes an extra constraint on the optimization problem, reduces the feasible region of the problem, and decreases the throughput. Allowing transmitters to access non-contiguous chunks could alleviate this problem, but this comes at the cost of guardband.

To incorporate the guardband cost into the utility maximization framework, we remodel the optimization objective as:

$$\max_{\mathbf{p}} \quad \sum_{i\in\mathcal{N}} \alpha_i(\log(Eb_i) - \mu G_i), \tag{4}$$

where G_i denotes the extra guardband cost of transmitter TX_i and $G_i = \omega\sum_{j=1}^{m-1}|p_{i,j+1} - p_{i,j}|$ and μ is a parameter representing the guardband cost of discontiguous spectrum access.

When $\mu = 0$, the transmitters can freely access each chunk, without considering the savings of guardband owing to spectrum aggregation. As μ increases, the guardband cost weighs more, and hence more penalty is imposed when accessing non-contiguous chunks. Hence the transmitters have less freedom to use non-contiguous chunks. An extremely large μ corresponds to the case where the transmitters only access contiguous spectrum chunks. *Hence, the value of μ can capture the tradeoff between contiguous and non-contiguous accesses.* We will detail the impact of μ in Section 5.

4.2 Decentralized Spectrum Access

4.2.1 Spectrum Interference Constraint

Besides the objective function that incorporates the guardband cost, the utility-optimization framework must account for interference between links sharing the same spectrum chunks. Due to the CSMA-based contention protocol, at most one transmitter can be active in each contention domain. A contention domain can be characterized by a maximal clique – a complete graph in which there exists an edge between two links if their transmission can interfere with each other. For each spectrum chunk, the sum contention probability of all transmitters belonging to the same maximal clique should not exceed 1. Let Q_j be the collection of all maximal cliques in the interference graph comprised of links using channel j, and let $|Q_j| = S_j$. Let $q_j^{(k)} \in Q_j$ be

the kth maximal clique, $k = 1, \ldots, S_j$. Then, the complete utility-optimization framework becomes:

$$\max_{\mathbf{p}} \quad \sum_{i \in \mathcal{M}} \alpha_i U_i(\mathbf{p}_i), \tag{5}$$

$$\text{s.t.} \quad \sum_{i \in q_j^{(k)}} p_{i,j} \leq 1, \forall q_j^{(k)} \in Q_j, \forall j \in \mathcal{M}, \tag{6}$$

where $U_i(\mathbf{p}_i) = \log(Eb_i) - \mu G_i$, and it is a concave function. Eq. (6) represents the aforementioned interference constraint for each spectrum chunk. In practice, the interference constraint between each transmitter-receiver pair could be obtained by allowing nodes to overhear each others' packet preambles and accordingly build a conflict map as in [27].

4.2.2 Decentralized Solution for Transmitters

We proceed to derive a decentralized fine-grained spectrum access (FSS) algorithm by solving the above optimization problem and decomposing it with respective to individual transmitters.

First, we introduce a Lagrangian multiplier $\boldsymbol{\lambda}$ to relax the interference constraint (6):

$$L(\boldsymbol{p}, \boldsymbol{\lambda}) = \sum_{i \in \mathcal{N}} \alpha_i U_i(\boldsymbol{p}_i) + \sum_{j \in \mathcal{M}} \sum_{k=1}^{S_j} \lambda_{jk} \left(1 - \sum_{l \in q_j^{(k)}} p_{l,j} \right). \tag{7}$$

By reordering the summation, we can obtain

$$\sum_{j \in \mathcal{M}} \sum_{k=1}^{S_j} \lambda_{jk} \sum_{l \in q_j^{(k)}} p_{l,j} = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}} \sum_{k: i \in q_j^{(k)}} \lambda_{jk} p_{i,j}.$$

Let $\boldsymbol{\lambda}^{(i)} = (\lambda_1^{(i)}, \ldots, \lambda_M^{(i)})$ and $\lambda_j^{(i)} = \sum_{k: i \in q_j^{(k)}} \lambda_{jk}$. We decompose (7) with respect to each transmitter TX_i:

$$L(\boldsymbol{p}, \boldsymbol{\lambda}) = \sum_{i \in \mathcal{N}} L_i(\boldsymbol{p}_i, \boldsymbol{\lambda}^{(i)}), \tag{8}$$

where

$$L_i(\boldsymbol{p}_i, \boldsymbol{\lambda}^{(i)}) = \alpha_i U_i(\boldsymbol{p}_i) - \sum_{j \in \mathcal{M}} \lambda_j^{(i)} p_{i,j} + \sum_{j \in \mathcal{M}} \sum_{k=1}^{S_j} \lambda_{jk}, \tag{9}$$

λ_{jk} is the Lagrange multiplier for the kth maximal clique of spectrum chunk j. We can write λ_{jk} as $\lambda_{jk} = \lambda F_{jk}^{(i)}$, where $F_{jk}^{(i)}$ is the collision probability of transmitter TX_i in clique k of chunk j.

The overall collision probability of transmitter TX_i in spectrum chunk j is $F_{ij} = 1 - \prod_{k: i \in q_j^{(k)}} (1 - F_{jk}^{(i)})$ since different transmitters in general operate independently. Assume $F_{jk}^{(i)}$ is kept sufficiently small, then we have:

$$F_{ij} \approx \sum_{k: i \in q_j^{(k)}} F_{ik}^{(i)}. \tag{10}$$

The assumption of a small $F_{jk}^{(i)}$ requires that the collision probability of any contention domain be kept low. This can be satisfied by ensuring no transmitter becomes overly aggressive during the contention. CSMA networks have built-in mechanisms (e.g., exponential backoff in 802.11) to enforce such etiquette. We will discuss in Section 4.3 how this can be achieved in a practical way within our spectrum access protocol.

Following the approximation in (10), the value of F_{ij} can be easily estimated by allowing transmitter TX_i to locally keep track of its overall contention/transmission failure probability, without knowledge of loss probabilities in each of its maximal cliques. Now, substituting for $\lambda_j^{(i)}$ and following (9), we have

$$L_i(\boldsymbol{p}_i) = \alpha_i U_i(\boldsymbol{p}_i) - \lambda \sum_{j \in \mathcal{M}} F_{ij} p_{i,j} + \sum_{j \in \mathcal{M}} \sum_{k=1}^{S_i} \lambda_{jk}. \tag{11}$$

While α_i can be used to prioritize different transmitters, we set $\alpha_i = \alpha$ for an equal bias. Note that $L_i(\boldsymbol{p}_i)$ is a concave function and it is maximized when

$$\frac{\partial L_i(\boldsymbol{p}_i)}{\partial \boldsymbol{p}_i} = 0 \iff \alpha \frac{\partial U_i}{\partial \boldsymbol{p}_i} - \lambda \sum_{j \in \mathcal{M}} F_{ij} = 0 \tag{12}$$

$$\iff \alpha \frac{\partial U_i}{\partial p_{i,j}} - \lambda F_{ij} = 0, \quad \forall j \in \mathcal{M}. \tag{13}$$

The third term $\sum_{j \in \mathcal{M}} \sum_{k=1}^{S_i} \lambda_{jk}$ on the right of (11) is removed when taking the partial derivative $\frac{\partial L_i(\boldsymbol{p}_i)}{\partial \boldsymbol{p}_i}$, while the constraint (6) can be satisfied with the backoff method which will be discussed in Section 4.3. We now devise a subgradient-based method to solve (13). We begin with deriving $\frac{\partial U_i}{\partial p_{i,j}}$ when $j \neq 1, m$.

Define $H := \frac{1}{\sum_{j=1}^{m} p_{i,j} - \omega_l p_{i,1} - \omega_r p_{i,m}}$. When $p_{i,j} \neq p_{i,j-1}$ and $p_{i,j} \neq p_{i,j+1}$, U_i is differentiable w.r.t. $p_{i,j}$, and when $p_{i,j} = p_{i,j-1}$ or $p_{i,j} = p_{i,j+1}$, U_i is not differentiable w.r.t. $p_{i,j}$. We derive its subgradient instead:

$$\partial U_i = \begin{cases} [\frac{1}{H}, \frac{1}{H} + 2\mu\omega], & p_{i,j} = p_{i,j-1},\ p_{i,j} < p_{i,j+1} \\ [\frac{1}{H} - 2\mu\omega, \frac{1}{H}], & p_{i,j} = p_{i,j-1},\ p_{i,j} > p_{i,j+1} \\ [\frac{1}{H} - 2\mu\omega, \frac{1}{H} + 2\mu\omega], & p_{i,j} = p_{i,j-1}.\ p_{i,j} = p_{i,j+1} \\ [\frac{1}{H} - 2\mu\omega, \frac{1}{H}], & p_{i,j} > p_{i,j-1},\ p_{i,j} = p_{i,j+1} \\ [\frac{1}{H}, \frac{1}{H} + 2\mu\omega], & p_{i,j} < p_{i,j-1},\ p_{i,j} = p_{i,j+1}. \end{cases}$$

At each break point, we can pick an arbitrary value within the interval corresponding to ∂U_i. For simplicity, we can pick the subgradient of U_i w.r.t. $p_{i,j}$ at all break points (i.e., $p_{i,j} = p_{i,j-1}$.) to be $\frac{1}{H}$.

In the case of $j = 1$ or $j = m > 1$, we can apply a similar approach and pick the subgradient of U_i as $\frac{1-\omega_l}{H} + \mu\omega$ and $\frac{1-\omega_r}{H} + \mu\omega$, respectively. For simplicity and without loss of generality, we could actually treat ω_r and ω_l equally as ω in the analysis for $j = 1$ or m.

Given the partial derivative or subgradient $\frac{\partial U_i}{\partial p_{i,j}}$, each transmitter TX_i can approach the utility-optimal point (13) following a gradient $\frac{\partial p_{i,j}}{\partial t} = \alpha \frac{\partial U_i}{\partial p_{i,j}} - \lambda F_{ij}$, which can be approached by adjusting its access probability to spectrum chunk j as:

$$p_{ij}(t+1) = p_{ij}(t) + \alpha \frac{\partial U_i}{\partial p_{i,j}} - \lambda F_{ij}. \tag{14}$$

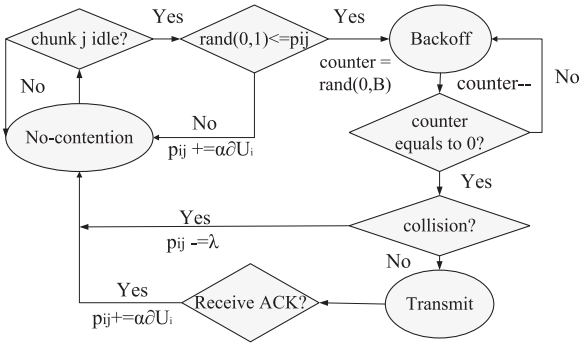

Fig. 4. State-transition diagram for transmitters.

Here α and λ can be considered as the "utility weight" and "penalty weight", respectively, and can be tuned to make a trade-off between maximizing utility and minimizing collision rate. A large α yields large utility. On the other hand, a large λ incurs a large penalty to $p_{i,j}$ to minimize the collision rate. But setting either α or λ too large causes oscillation around the optimum. In addition, each transmitter TX_i can estimate F_{ij} using its contention failure rate on spectrum j, and hence no explicit message exchange is needed.

4.3 Mapping Optimization Solution to MAC Protocol

We now show how to translate the update rule in Eq. (14) into a practical CSMA-compatible MAC protocol.

Fig. 4 shows the state-transition diagram of each transmitter. Similar to 802.11, a transmitter updates its states on a per-time-slot basis. But it needs to maintain states for each spectrum chunk separately. Initially, all chunks have an identical contention probability, set to 0.1 by default. There are three states: *no-contention*, *contention*, *transmission*. In the *no-contention* state, the transmitter senses the idle/busy status on that chunk. If it is idle, the transmitter transits to *contention* state with a contention probability $p_{i,j}$. With probability $(1 - p_{i,j})$, it remains in the *non-contention* state, and increases its contention probability by:

$$p_{i,j} \leftarrow p_{i,j} + \alpha \frac{\partial U_i}{\partial p_{i,j}}, \tag{15}$$

where $\frac{\partial U_i}{\partial p_{i,j}}$ simply depends on $p_{i,j}$, $p_{i,j+1}$ and $p_{i,j-1}$ as discussed above, and can be obtained by each transmitter independently since it only requires local information.

After a spectrum chunk enters the *contention* state, the transmitter starts the backoff procedure for it by selecting a number R uniformly distributed between 0 and the backoff window BW. It suspends itself and counts R slots before transmission. If the backoff counter expires on time and the chunk is idle across the R slots, the transmitter begins transmission, and enters the *transmission* state. Otherwise, it declares a contention failure and switches back to *non-contention* state, and reduces the contention probability as:

$$p_{ij} \leftarrow p_{ij} - \lambda. \tag{16}$$

Upon completing transmission and receiving an ACK, the transmitter returns to its initial non-contention state and advances its contention probability on chunk j according to equation (15). Otherwise, it suffers a transmission failure and reduces its contention probability according to (16). During the adaptation, transmitters should truncate p_{ij} to make it fall in the range $[0, 1]$.

Note that multiple spectrum chunks may be idle and transit to the *contention* state simultaneously. In such a case, the transmitter should choose the same backoff counter R for all such chunks. The maximum backoff window size BW remains intact during the adaptation, since the transmitters have already adapted their aggressiveness through the contention probability p_{ij}.

4.4 Convergence Analysis

We show that the above adaptation mechanism of individual transmitters converges to the utility-optimal point. Note that the update rule for p_{ij} is essentially an iterative scheme:

$$\partial \boldsymbol{p}_i = \alpha \partial U_i(\boldsymbol{p}_i) - \lambda \sum_{j \in \mathcal{M}} F_{ij}, \tag{17}$$

where $\partial U_i(\boldsymbol{p}_i)$ is the subgradient of U_i *w.r.t.* $\boldsymbol{p}_i$, and $\partial \boldsymbol{p}_i := (\frac{\partial p_{i,1}}{\partial t}, \ldots, \frac{\partial p_{i,M}}{\partial t})$.

The equilibrium of this iterative system is equal to the optimal value $\boldsymbol{p}_i^{\star}$ that satisfies $\partial L_i(\boldsymbol{p}_i^{\star}) = 0$, which is unique due to concavity. Note also that

$$\begin{aligned}\frac{\partial L_i(\boldsymbol{p}_i)}{\partial t} &= \partial L_i(\boldsymbol{p}_i)' \partial \boldsymbol{p}_i \\ &= \left(\alpha \partial U_i(\boldsymbol{p}_i) - \lambda \sum_{j \in \mathcal{M}} F_{ij}\right)' \left(\alpha \partial U_i(\boldsymbol{p}_i) - \lambda \sum_{j \in \mathcal{M}} F_{ij}\right) \geq 0.\end{aligned}$$

The last inequality holds because the identity matrix $\mathbf{I}$ is positive definite, i.e., $\mathbf{x}'\mathbf{x} = \mathbf{x}'\mathbf{I}\mathbf{x} \geq 0$ for any vector $\mathbf{x}$. Hence, $L_i(\boldsymbol{p}_i)$ is a Lyapunov function under (17). Following iteration (17), the objective function L_i converges to the unique stable point. Due to the effect of subgradient, it converges to within some range of the optimal value and finds an ϵ-suboptimal point. ϵ decreases with the step sizes α and λ.

At differentiable points, the above iteration is similar to the steepest ascent gradient method which is known to have linear convergence. Further, the non-differentiable parts of our objective function are concave and piecewise linear. Under subgradient methods, the optimal points of such a function is approached at a linear convergence rate [14]. So, our solution also has linear convergence. Note that the adaptation mechanism only needs to be executed on a coarse time-scale. It stops after convergence, and only needs to be rerun when a new node joins the network, or when certain nodes decide to change their maximum channel width (e.g., a node degrading itself from 160 MHz 802.11ac mode to 20 MHz 802.11a mode), such that p_{ij} needs to be updated.

4.5 Discussion

Our previous analysis focused on the coexistence of heterogeneous channel widths where all devices are capable of per-frame spectrum shaping and discontiguous spectrum access. Legacy 802.11 devices, though not integrated with such advanced PHY-layer techniques, can be viewed as a special case in our framework, where all spectrum chunks share the same access probability and transmit simultaneously. Therefore, by further constraining the access

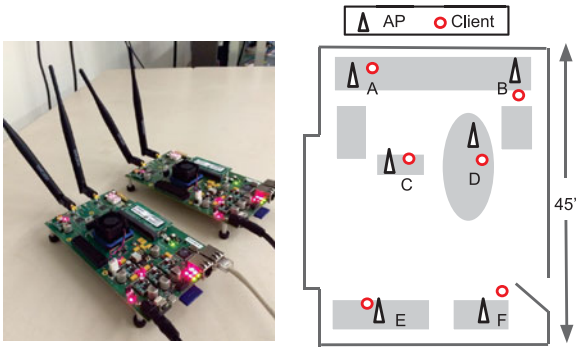

Fig. 5. Testing environments: (a) WARP testbeds and (b) Testing locations. Experiments are conducted in a typical lab environment with abundant mutli-path reflections. The grey blocks in the figure represents the pillars, tables, racks, and other obstacles.

probability of transmitters in the optimization formulation (e.g., by setting a large μ), FSS can deal with the coexistence of legacy WiFi devices of different bandwidths. In what follows, we will stay focused on the general case where all devices are capable of flexible spectrum shaping/access.

5 Evaluation

To validate FSS, we built a detailed packet-level simulator with trace-driven simulation capability. The simulator adopts an SINR-based interference module that accumulates the power level of all interfering packets and declares a collision only if the SINR is below the decoding threshold (6 dB for BPSK [13]). The collision model takes into account possible partial overlap between packets from different channels. The PHY-layer spectrum access capabilities follow existing physical-layer solutions [10], [15], [16], [21] (Section 3). FSS's MAC protocol is implemented following the description in Section 4.3.

To augment trace-driven simulation, we built a 12-node software-radio testbed with the WARP boards [25] (shown in Fig. 5a) that can collect packet-level traces and feed them into the simulator. We deploy the nodes following the floormap in Fig. 5b. Then, we follow the approach in [26] to collect pairwise RSS (or interference power) and receiver noise floor, by sending/receiving a narrowband signal between each pair of nodes. Based on the collected statistics, we can compute the SINR of each link when one or more links are transmitting concurrently. The transmission power is within the limit specified by the 802.11 standard. In the resulting interference map, we deem link i as interfering link j if the latter's SINR is below the decoding threshold when both are used simultaneously for packet transmission. We fix a link to its basic modulation level as long as its SINR is above the threshold. This isolates the rate-adaptation problem and allows us to focus on the impact of heterogeneous spectrum sharing. However, FSS can coexist with an existing rate-adaption scheme as it makes no assumption on transmission rate. To avoid irrelevant disturbances from external networks, the trace collection is conducted on a 5 GHz band unused by ambient WiFi devices. According to our measurements with WARP boards, there is no interference between location A and F, B and E, so that APs at these locations could exploit spatial reuse and transmit packets concurrently. Other than this, APs will interfere with each other and only one could transmit packets using the same spectrum.

We compare the proposed scheme with the following spectrum sharing schemes: (i) 802.11 legacy MAC protocol without Dynamic Channel Allocation [18], (ii) 802.11 legacy MAC protocol with DCA [18], (iii) WiFi-NC MAC protocol [21], (iv) FLUID-like spectrum access.

Among these benchmarks, WiFi-NC [21] shares a similar PHY layer as FSS. It allows a transmitter to split its spectrum into 5 MHz chunks. Each chunk performs carrier sensing and packet transmission/reception. A transmitter can simultaneously transmit packets on one chunk and receive on the other, and isolation between such chunks is ensured using filters with a steep transition band. Unlike WiFi-NC, however, FSS can opportunistically combine contiguous narrow chunks into a wide channel to reduce the guardband cost.

FLUID [24] also supports flexible channel widths like FSS, but an AP cannot split a frame across non-contiguous spectrum and requires a central controller for spectrum assignment. We did not implement the exact centralized heuristic of FLUID, but used the distributed adaptive spectrum access algorithm similar to FSS, so that the performance difference between FSS and FLUID comes from fine-grained spectrum access, not because of centralization or decentralization.

Also, we used 802.11's MAC-level timing parameters and simulate the detailed packet-level CSMA/CA operations. We also account for the DCA feature in legacy 802.11ac [18]. Without DCA, wide-band devices perform carrier sensing on the entire wide band. With DCA [18], the wide-band transmission can select one 20 MHz channel as the primary channel and uses it for channel sensing and packet transmission even though the other parts of the channel might be busy. Without loss of generality, we assume the leftmost 20 MHz channel as the primary channel of wide-band transmission.

For each spectrum-sharing scheme we consider the following three heterogeneous spectrum band coexistence cases: (i) two coexisting WLANs, each consisting of an AP-client pair, (ii) multiple coexisting WLANs, each consisting of an AP-client pair, (iii) two or multiple coexisting WLANs, each consisting of multiple AP-client pairs.

Throughput and access rate are used as performance metrics. The access rate is defined as the number of transmission attempts (upon winning contention) per second on all spectrum chunks of a transmitter. We assume transmitters have backlogged traffic and packet size is fixed at 1 KB. All our experiments run for 100 seconds in simulation time and the results are averaged over 30 repetitions with different random seeds. Following 802.11, the basic data rate is 1.5 Mbps for 5 MHz spectrum chunk and increases linearly with the spectrum width [9]. The guardband is set as described in Section 4.1.

- *Two WLANs partially sharing spectrum:* We start with the case where two co-located WLANs share part of the spectrum. Without loss of generality we choose two WLANs with corresponding AP location A, B. Figs. 6a and 7 show the spectrum deployment and simulation results, respectively. In Fig. 7, "802.11" represents the result of 802.11 protocol with/without

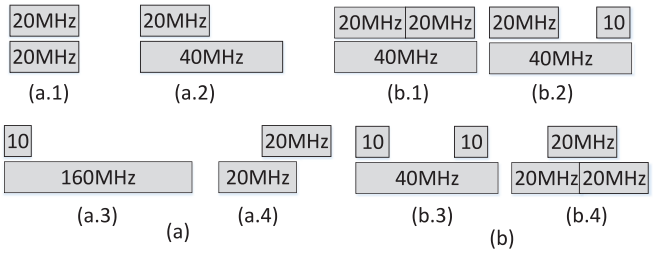

Fig. 6. Experimental configuration with heterogeneous channel widths or partially-shared channels: (a) two WLANs and (b) multiple WLANs.

DCA. DCA degenerates to a legacy protocol in the case of two WLAN spectrum sharing due to its inefficient primary channel selection. Current 802.11 standard [9] does not provide any efficient primary channel selection scheme, which is a main limitation of current DCA.

When both WLANs use 20 MHz channels that are fully overlapping (Fig. 6a.1), both achieve about same throughput and 802.11 achieves similar performance as FSS, WiFi-NC and FLUID. The aggregate throughput increases when a 20 MHz WLAN shares its channel with a 40MHz channel (Fig. 6a.2), since the 40 MHz WLAN takes less time to send a packet. However, running 802.11, the 40 MHz WLAN has almost the same throughput as the 20 MHz WLAN, although it has a wider bandwidth. This reflects the partial channel blocking problem (Section 2), where the 40 MHz WLAN treats its entire channel as a single entity, and accesses the non-overlapping 20 MHz band at the same rate as the shared one, causing severe under-utilization of spectrum. By contrast, with FLUID and FSS, both halves of the 40 MHz channel can be opportunistically exploited at any time. Compared to the legacy 802.11, FLUID increases the throughput of the 40 MHz WLAN by 95 percent. FSS enhances the throughput of the 40 MHz WLAN by 153 percent, and maintains similar throughput for the 20 MHz WLAN, confirming the intuition that wider channels should gain more throughput. WiFi-NC enables each 5 MHz spectrum chunk to transmit packets independently. The throughput of 20 MHz WLAN decreases slightly due to extensive contention between 20/40 MHz WLANs, and its total throughput is lower than FSS since it incurs a large guardband cost.

When the width of the narrowband WLAN reduces from 20 to 10 MHz (Fig. 6a.3), and using 160 MHz wideband, although more non-overlapping spectrum is available, the total network throughput remains almost the same when running 802.11. FSS improves the throughput of the 160 MHz WLAN by almost 8× compared to the 802.11 legacy and around 13 percent (10 percent) compared to FLUID (WiFi-NC).

In summary, *the spectrum underutilization of 802.11 gets more severe as the ratio of the shared spectrum to the channel bandwidth decreases. FSS can improve throughput by multiple folds in such cases. It outperforms variable channel width protocols, such as FLUID, that do not allow non-contiguous spectrum access. It also outperforms protocols, like WiFi-NC that rely on narrowband transmission, by intelligently reducing guardband costs.*

We evaluate fairness by using the ratio between the wide-band and narrow-band clients' access rate. We study the topology in Fig. 6a.2 as an example, and plot the results in Fig. 8a. Although the sub-optimal subgradient-based algorithm leads to small occasional jitters, FSS still approximates the optimal solution well and has a higher level of fairness than 802.11, FLUID and WiFi-NC. The optimal access ratio is 2, whereas the access ratio of 802.11 is close to 1 due to the inherent fairness performance of CSMA mechanism. FLUID builds on a heuristic spectrum access protocol which does not provide any fairness guarantee. WiFi-NC does not address fairness issues, either.

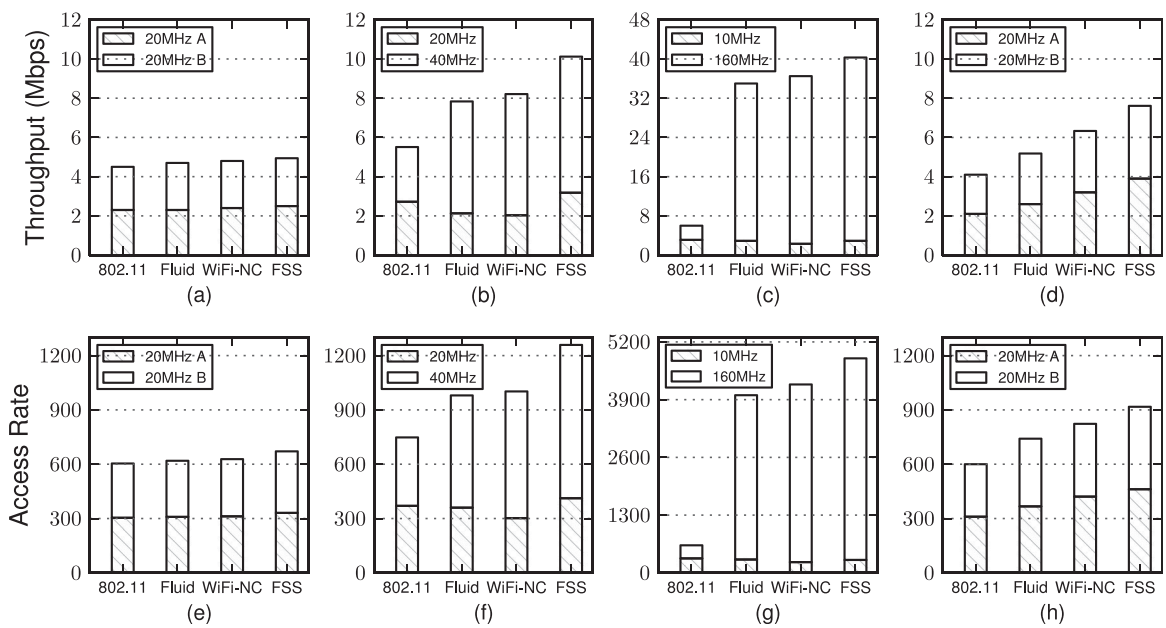

Fig. 7. Throughput and access rate for two WLANs: single AP-client pair in each WLAN.

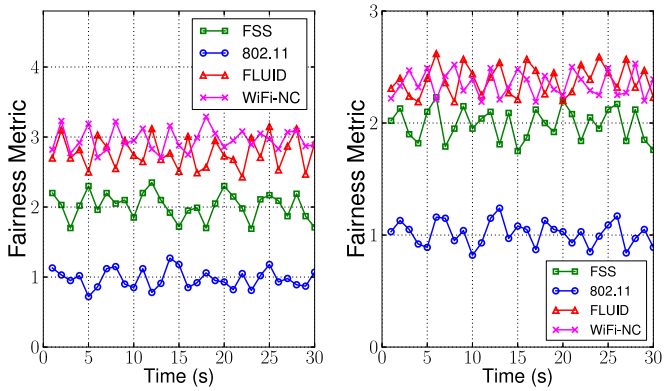

Fig. 8. Variation of fairness of FSS versus other MAC protocols: (a) homogeneous traffic and (b) heterogenous traffic.

The other topologies also reveal similar results.

To demonstrate the fairness performance in an extreme case where 20/40 MHz WLANs carry heterogeneous traffic (i.e., different number of transmission pairs), we consider the following topology without loss of generality: one 20 MHz transmission pair at location A and three 40 MHz transmission pairs at location B, C, D, respectively. The four transmission pairs interfere with each other and contend for the use of the spectrum (Fig. 6a.2). We calculate the wide/narrow-band access ratio (of a single user) and Fig. 8b plots the fairness results. FSS still provides a fairness level close to the optimal regardless of the heterogeneous traffic for different WLANs. The access ratio of 802.11 is still close to 1. The access ratios of WiFi-NC and FLUID are closer to the optimal than Fig. 8a since the highly intensive contention among wideband transmissions themselves intrinsically reduces the per-user access rate, and the wide/narrow-band access ratio decreases correspondingly. Therefore, FSS ensures proportional fairness even when different WLANs carry heterogeneous traffic.

- *Multiple WLANs sharing spectrum:* We explore the example multi-WLAN coexistence cases shown in Fig. 6b and use interfering AP locations A, B and C. Fig. 9 indicates the throughput and access rate in different cases. When a 40 MHz WLAN coexists with two orthogonal 20 MHz WLANs running 802.11, its throughput approaches 0, whereas the 20 MHz WLANs have similar throughput to the case without contenders. This essentially verifies the middle-channel starvation, where the 40 MHz WLAN can hardly find any slot when both 20 MHz contenders are idle. In the case of multiple (> 2) DCA-enabled WLANs sharing spectrum, 40 MHz WLAN can opportunistically transmit on the primary 20 MHz channel, but the throughput of such a wideband WLAN is limited since the secondary channel rarely has a transmission chance. Besides, the right 20 MHz WLAN experiences less contention than the left one and gets higher throughput than others. Therefore, DCA is neither efficient nor fair. With FLUID, the 40 MHz WLAN can opportunistically transmit over contiguous spectrum chunks, and achieve 1.4× the 20 MHz WLAN throughput. The total network throughput decreases slightly in this case since assigning only contiguous spectrum leads to inefficient spectrum usage. WiFi-NC also greatly improves the 40 MHz WLAN throughput over 802.11, but its total throughput is hampered by the large guardband cost. Using FSS, the 40 MHz WLAN can opportunistically transmit over any of the spectrum chunks, thus achieving 2× the throughput of 20 MHz WLAN. We make a tradeoff between guardband spectrum wastage and non-contiguous spectrum access with FSS, and hence the total network throughput is larger than WiFi-NC and FLUID.

  When the two narrowband WLANs reduce their channel width (e.g., 20/10/40 and 10/10/40 MHz coexistence cases), the 40 MHz channel remains starved when running 802.11. DCA still suffers from the inefficiency and unfairness problems. In contrast, FSS improves the throughput by an order-of-magnitude. Its efficient usage of non-overlapping spectrum also enhances the throughput.
- *Two or multiple WLANs with multiple APs sharing spectrum:* We study the example spectrum coexistence cases shown in Figs. 6a.2 and 6b.2, respectively. Correspondingly and without of loss of generality, for

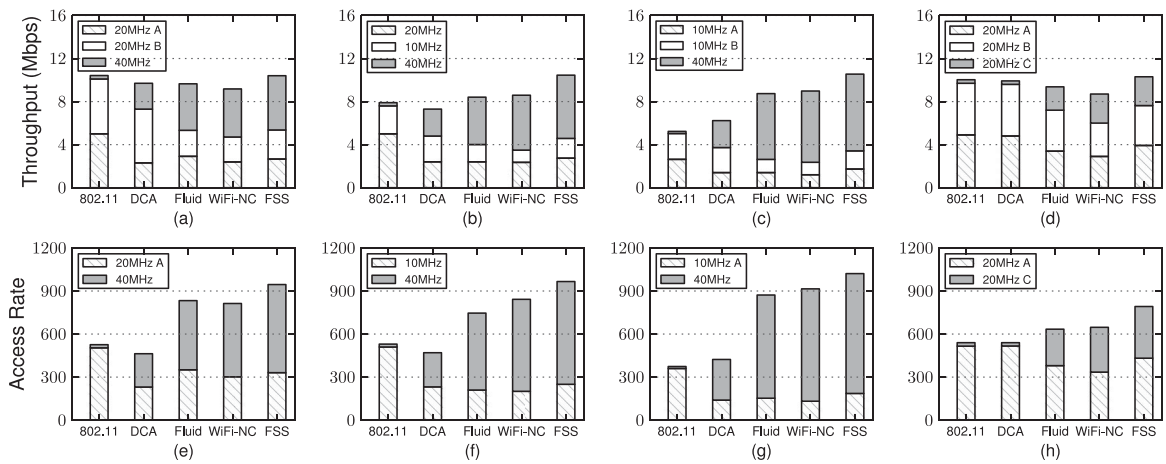

Fig. 9. Throughput and access rate for three WLANs: single AP-client pair in each WLAN.

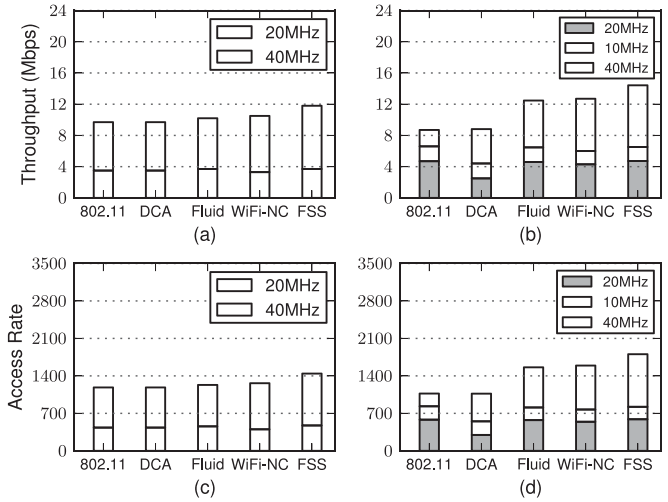

Fig. 10. Throughput and access rate in multiple WLANs: multiple AP-client pairs in each WLAN.

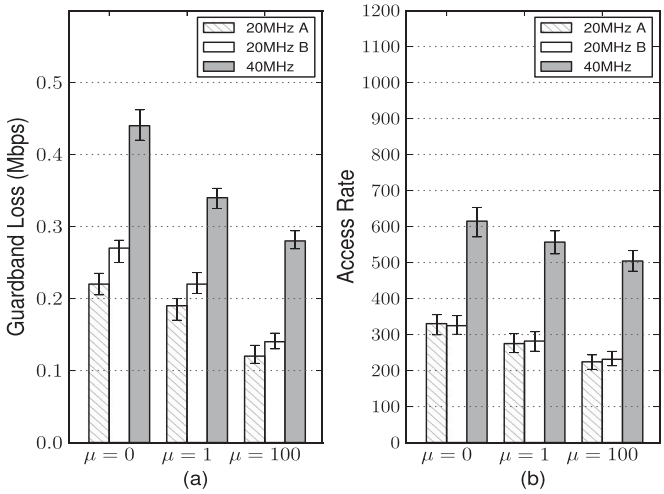

Fig. 11. Tradeoff between guardband cost and spectrum access.

the two-WLAN topology, we deploy the 20 MHz WLAN consisting of three 20 MHz APs at locations A, B, C and the 40 MHz WLAN with APs at E, D, F. We sum up the throughput and access rate of all APs in each WLAN and the simulation results are plotted in Figs. 10a and 10c. For the multi-WLAN topology, we deploy the 20 MHz WLAN with APs at locations A and B, 10 MHz WLAN with APs at C and D and 40 MHz WLAN with APs at E and F. Figs. 10b and 10d plot the throughput and access rate, respectively. In both cases, FSS further enhances the throughput of wideband WLAN while maintaining that of narrowband WLAN. In particular, as the spectrum becomes more heterogeneous (Figs. 10b and 10d), FSS gains more from fine-grained spectrum access. Note that the above evaluation has focused on small-scale benchmark topologies that represent typical heterogeneous spectrum sharing scenarios. Increasing the number of APs on each channel and forming a large-scale network will not reveal more or better insights.

- *Tradeoff between guardband cost and spectrum access opportunity:* We evaluate the tradeoff between guardband cost and spectrum access opportunity, focusing on the topology in Fig. 6b.1. As shown in Fig. 11, when μ increases, more weight and penalty are imposed on use of guardbands, encouraging the transmitters to access contiguous chunks. With a smaller μ, the guardband cost decreases, but the transmitters will have less chance to access the available spectrum. Since throughput decreases with higher guardband cost and increases with higher access rate, we can make a tradeoff between these two factors by manipulating μ. Fig. 12 shows how μ impacts the network throughput in the setting as above. Setting μ either too large or too small may underutilize the channel, and an intermediate value of around 18 maximizes network throughput. Note that this value is optimal only under the 802.11guardband setting and needs to be redesigned for different cases. Following similar measurements, we can obtain easily optimal values of μ for different cases and apply it to balance the tradeoff between guardband cost and spectrum access opportunity.
- *Realistic traffic trace-based analysis:* We further evaluate FSS's performance under intermittent and bursty traffic loads, and focus on intermittent TCP Web downloads as in [22]. To augment trace-driven simulation, we periodically download an entire webpage `www.apple.com` (following the same procedure in [22]) and use Wireshark to capture the realistic narrowband traffic traces. We calculate the data packets transmitted within every 100 ms, treat it as a narrowband packet-arrival pattern and feed the traces into our simulator. The wideband channel is assumed to have backblogged traffic. We study the topology in Fig. 6a.2 as an example, and consider the 802.11 protocol as the baseline. Fig 13 shows the result.

  Fig. 13a shows the narrowband throughput (averaged over 100 ms) without wideband transmission. The narrowband transmission starts at around 12 s and is intermittent. The peak throughput of narrowband is around 2 Mbps since the TCP downloads are too short. Fig. 13b shows the narrowband throughput when it coexists with FSS wideband transmission. Narrowband TCP packets are short,

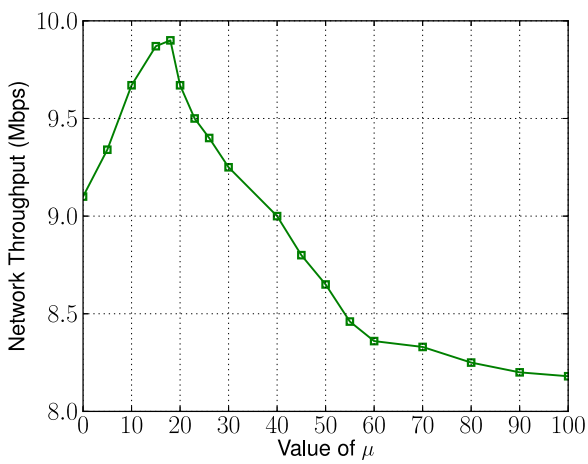

Fig. 12. Network throughput when μ varies.

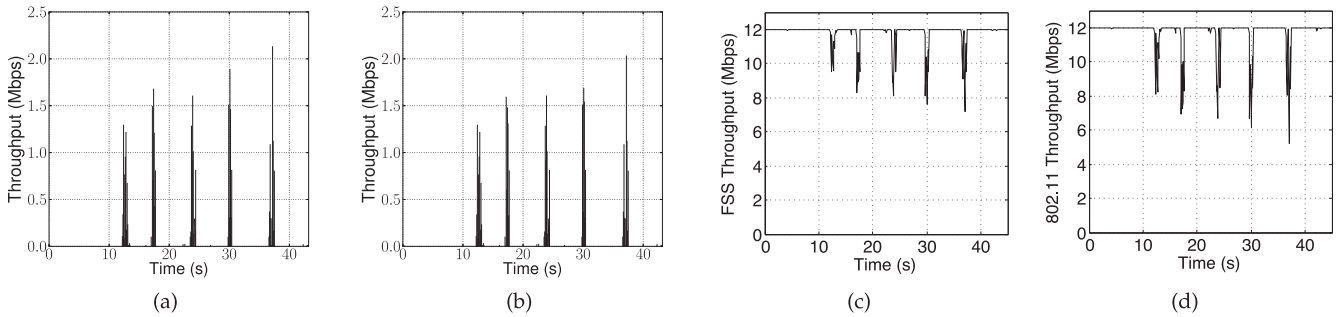

Fig. 13. FSS's reactions to TCP web downloads: (a) Narrowband throughput without wideband transmission. (b) Narrowband throughput with FSS wideband transmission. (c) FSS wideband throughput. (d) 802.11 wideband throughput.

and therefore the narrowband throughput does not change much even in the presence of wideband contention (it shows similar performance when coexisting with 802.11 wideband transmission). Figs. 13c and 13d show the throughput of wideband transmission with FSS and 802.11, respectively. During the arrival interval of a narrowband packet, the throughput of wideband transmission reaches its maximum value (12 Mbps) for both FSS and 802.11. At the burst point of narrowband packets (at which the narrowband throughput peaks), wideband throughput decreases due to the spectrum contention with narrowband packets. As shown in Figs. 13c and 13d, 802.11 wideband throughput decreases more than FSS when narrowband reaches its peak throughput.

To demonstrate the throughput differences, Fig. 14 plots the throughput of FSS and 802.11 wideband at the five sample timestamps where the bursty narrowband gets its peak throughput (time: 12.4-12.5s; 17.2-17.3s; 23.9-24.0s; 30.1-30.2s; 37.2-37.3s). FSS improves wideband throughput by an average of 24 percent among the five points. Therefore, the narrow/wide-band coexistence problem is present even when under intermittent and bursty traffic patterns, and FSS deals with the problem efficiently, providing larger throughput than 802.11. The SWIFT protocol in [22] allows wideband radios to identify busy channels and then nulls them to prevent interference to narrowband. However, using SWIFT under intermittent and bursty traffic, the wideband throughput decreases during the arrival interval of narrowband packets ([22], Section 7.5), so it is not as efficient as FSS.

Downlink/uplink symmetry is another key characteristic of realistic WLAN traffic. The authors of [29] proposed a spectrum-sharing mechanism to eliminate contention asymmetry due to traffic asymmetry. Traffic asymmetry is not a focus of FSS, but it can be dealt with by complementing FSS with the results in [29].

- *Legacy and wide-band devices coexistence:* We consider the case when FSS devices coexist with legacy 802.11 devices that can only deem its spectrum band as an atomic block. We consider two example topologies: (1) one 20 MHz legacy WLAN and one 40 MHz FSS WLAN (as in Fig. 6a.2); (2) two 20 MHz legacy WLANs with one 40 MHz FSS WLAN (as in Fig. 6b.1). These two examples correspond to the coexistence of two/multiple WLANs. The spectrum chunk size is set to 20 MHz in FSS due to the coarse-granularity of legacy devices. We compare the throughput performance with the legacy 802.11 and plot the results in Fig. 15. In Fig. 15a, FSS improves the throughput of 40 MHz WLAN by 140 percent, solving the partial blocking problem when one 40 MHz transmission coexists with one 20 MHz transmission. Fig. 15b further validates FSS's capability to deal with the middle-channel starvation problem when 40 MHz coexists with two 20 MHz transmissions. In summary, FSS solves the heterogeneous bandwidths coexistence problem even when it runs together with legacy devices.

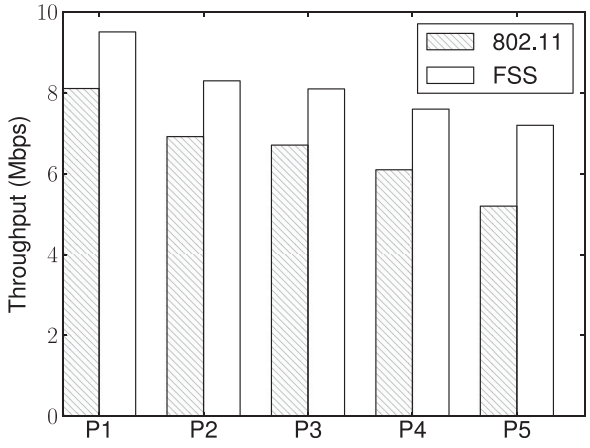

Fig. 14. Wideband throughput under realistic traffic: FSS versus 802.11.

6 Related Work

FSS is closely related to two broad areas: fine-grained channel access and dynamic spectrum access (DSA).

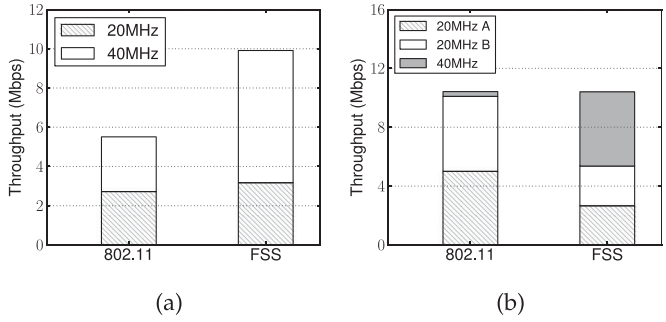

Fig. 15. Throughput of legacy and wide-band devices under two topologies: (a) two WLANs and (b) multiple WLANs.

- *Fine-grained channel access.* FICA [19] reduces the MAC-layer overhead of high-rate WLANs by splitting a channel into multiple subchannels and allowing contention for use of subchannels. However, it uses a frequency-domain backoff algorithm which is very different from the traditional CSMA and cannot coexist directly with current 802.11 WLANs. The authors of [15] proposed to divide a frame into multiple mini-frames and divide the entire spectrum into multiple mini-channels. They use a greedy spectrum-assignment algorithm to determine where to place the mini-frames, without considering the guardband cost of non-contiguous mini-channels.
- *Dynamic spectrum access.* RODIN [16] is a per-frame spectrum-shaping mechanism that runs in the PHY-layer and supports DSA in COTS wireless devices. The authors of [20] show that dynamically changing channel width achieves better throughput than using a static wide channel. However, the transmitters can only access contiguous spectrum in a coarse-grained manner. FLUID [24] is a more comprehensive heterogeneous channel-assignment scheme but is limited by the same constraints as RODIN. In contrast, FSS allows the transmitters to access non-contiguous spectrum and makes spectrum adaptation decisions on a per-packet basis. WiFi-NC [21] proposes use of narrow channels to improve the spectrum efficiency. FSS builds on similar PHY mechanisms as WiFi-NC, but solves the MAC-layer utility-optimal channel access problem. It allows narrow channels to opportunistically bond and form wide channels. FSS optimizes the guardband cost and further increases network throughput while ensuring fairness. SWIFT [22] allows wideband radios to identify busy channels and then null them to prevent interference to WiFi. However, it identifies busy spectrum by poking the WiFi devices with a jamming tone and observing their backoff reaction. It enables coexistence, but it is conservative, thus limiting the performance. Its MAC protocol suffers from the scalability problem.

JELLO [23] dynamically allocates spectrum according to traffic demand, which works only in a coarse-grained way and requires coordination of transmitters. In contrast, FSS transmitters adjust their spectrum usage and access probability on a per-frame basis, and without explicit signaling. B-smart [28] introduces a time-frequency allocation protocol for cognitive radio networks, which requires each node to know all its neighbors' spectrum-usage via an RTS/CTS handshake executed on a dedicated control channel. In contrast, FSS aims to resolve the coexistence between the different generations of 802.11 protocols with heterogeneous spectrum widths. FSS' spectrum adaptation is in-band and purely based on distributed carrier sensing. An approach of optimal CSMA was proposed in [30] to achieve proportional fairness under heterogeneous traffic, collisions, etc. FSS assumes all users have backlogged traffic in order to focus on the proportional fairness. It accounts for the differences of channel bandwidth between users, which had not been considered before. To achieve proportional fairness under both heterogeneous channel bandwidths and traffic is an interesting problem in its own right, while it is beyond the scope of this paper and left as our future work.

7 Conclusion

In this paper, we investigated the inefficiency and unfairness of 802.11 when variable-width WLANs coexist. We proposed FSS, an optimization-driven framework that enables fine-grained spectrum sharing among heterogeneous WLANs. FSS builds on top of the 802.11 CSMA protocol, but allows transmitters to contend for use of small spectrum chunks that can be discontiguous across an entire channel. The contention algorithm is fully decentralized, but can achieve optimal efficiency and proportional fairness, and converges at a linear rate. It solves the partial spectrum sharing and middle-channel starvation problems, and achieves multi-fold throughput gain over current 802.11. In future, we would like to implement FSS as a real-time MAC-layer module on top of an existing software-radio implementation of discontiguous channel access [10], [23].

Acknowledgments

The work reported in this paper was supported in part by the NSF under grants CNS-1317411, CNS-1318292, and CNS-1343363; DGIST Global Research Laboratory Program through NRF funded by MSIP of Korea (2013K1A1A2A02078326).

References

[1] F. P. Kelly, A .K. Maulloo, and D. K. H. Tan,"Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, no. 3, pp. 237–252, 1998.
[2] X. Wang and K. Kar, "Cross-layer rate control for end-to-end proportional fairness in wireless networks with random access", in *Proc. 6th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2005, pp. 157–168.
[3] L. Li, M. Pal, and Y. R. Yang, "Proportional fairness in multi-rate wireless LANs," in *Proc. IEEE Conf. Comput. Commun.*, 2008.
[4] S. B. Lee, I. Pefkianakis, A. Meyerson, and S. Xu, "Proportional fair frequency-domain packet scheduling for 3GPP LTE Uplink" in *Proc. 29th Conf. Inform. Commun.*, 2009, pp. 2597–2605.
[5] F. P. Kelly, "Charging and rate control for elastic traffic," *Eur. Trans. Telecommun.*, vol. 8, pp. 33–37, Jan./Feb. 1997.
[6] A. Eryilmaz and R. Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control", in *Proc. 24th Annu. Joint Conf. IEEE Comput. Commun. Soc.* 2005, pp. 1794–1803.
[7] A. Abdel-Hadi and C. Clancy, "A utility proportional fairness approach for resource allocation in 4G-LTE", in *Proc. Int. Conf. Comput., Netw. Commun.*, 2014, pp. 1034–1040.
[8] T. Lan, D. Kao, M. Chiang, and A. Sabharwal, "An axiomatic theory of fairness in network resource allocation", in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.
[9] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Amendment 4*, IEEE Standard 802.11TM, 2013.
[10] X. Zhang and K. G. Shin, "Adaptive subcarrier nulling: Enabling partial spectrum sharing in wireless LANs," in *Proc. IEEE Int. Conf. Netw. Protocols*, 2011, pp. 311–320.
[11] Texas Instruments, "WLAN Channel bonding: Causing greater problems than it solves," White Paper, 2003.
[12] L. Deek, E. Villegas, E. Belding, S. Lee, and K. Almeroth, "The Impact of channel bonding on 802.11n network management", in *Proc. ACM CoNEXT*, 2011, pp. 11:1–11:12.
[13] M. Y. Arslan, K. Pelechrinis, I. Broustis, S. V. Krishnamurthy, S. Addepalli, and K. Papagiannaki, "Auto-Configuration of 802.11n WLANs," in *Proc. ACM CoNEXT*, 2010, pp. 27:1–27:12.

[14] Dimitri P. Bertesekas, *Nonlinear Programming*, 3rd ed. Belmont, MA, US: Athena Scientific, 1995.
[15] S. Yun, D. Kim, and L. Qiu, "Fine-grained spectrum adaption in WiFi networks," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, 2013, pp. 327–338.
[16] E. Chai, K. G. Shin, J. Lee, S. Lee, and R. Etkin, "Building efficient spectrum-agile devices for dummies," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 149–160.
[17] L. Yang, B. Y. Zhao, and H. Zheng, "The spaces between us: Setting and maintaining boundaries in wireless spectrum access," in *Proc. ACM MobiCom*, 2010, pp. 37–48.
[18] M. Park, "IEEE 802.11 ac: Dynamic bandwidth channel access." in *Proc. Int. Conf. commun.*, 2011, pp. 1–5.
[19] K. Tan, J. Fang, Y. Zhang, S. Chen, L. Shi, J. Zhang, and Y. Zhang, "Fine-grained channel access in wireless LAN," in *Proc. ACM SIGCOMM*, 2010, pp. 147–158.
[20] R. Chandra, R. Mahajan, T. Moscibroda, R. Raghavendra, and P. Bahl, "A case for adapting channel width in wireless networks," in *Proc. ACM SIGCOMM Conf. Data Commun.*, 2008, pp. 135–146.
[21] K. Chintalapudi, B. Radunovic, V. Balan, and M. Buettener," WiFi-NC: WiFi over narrow channels," in *Proc. 9th USENIX Conf. Netw. Syst. Des. Implementation*, 2012, p. 4.
[22] H. Rahul, N. Kushman, D. Katabi, C. Sodini, and F. Edalat, "Learning to share: Narrowband-friendly wideband wireless networks," in *Proc. ACM SIGCOMM Conf. Data Commun.*, 2008, pp. 147–158.
[23] L. Yang, W. Hou, L. Cao, B. Zhao, and H. Zheng, "Supporting demanding wireless applications with frequency-agile radios," in *Proc. 7th USENIX Conf. Netw. Syst. Des. Implementation*, 2010, p. 5.
[24] S. Rayanchu, V. Shrivastava, S. Baneree, R. Chandra, "FLUID: Improving throughputs in entreprise wireless LANs through flexible channelization," in *Proc. 7th Annu. Int. Conf. Mobile Comput. Netw.*, 2011, pp. 1–12.
[25] A. Khattab, J. Camp, C. Hunter, P. Murphy, A. Sabharwal, and E. W. Knightly. Warp, "A flexible platform for clean-slate wireless medium access protocol design," *SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 12, pp. 56–58, 2008.
[26] L. Cheng, B.E. Henty, D.D. Stancil, and F. Bai, "Mobile vehicle-to-vehicle narrow-band channel measurement and characterization of the 5.9 GHz dedicated short range communication (DSRC) frequency band", *IEEE J. Select. Areas Commun.*, vol. 25, no. 8, pp. 1501–1516, 2007.
[27] Mythili Vutukuru, Kyle Jamieson, and Hari Balakrishnan "Harnessing exposed terminals in wireless networks", in *Proc. 5th USENIX Symp. Netw. Syst. Des. Implementation*, 2008, pp. 59–72.
[28] Yuan Yuan, P. Bahl, R. Chandra, T. Moscibroda, and Y. Wu, "Allocating dynamic time-spectrum blocks in cognitive radio networks", in *Proc. 8th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2007, pp. 130–139.
[29] A. Flores and E. Knightly, "Virtual duplex: Scaling dense WLANS and eliminating contention asymmetry", in *Proc. IEEE 22nd Int. Conf. Netw. Protocols*, 2014, pp. 603–611.
[30] B. Nardelli, E. Knightly, "Robust CSMA: Adapting to channel and traffic asymmetry," in *Proc. 12th Annu. IEEE Int. Conf. Sens., Commun., Netw.*, 2015, pp. 127–135.

Sihui Han received the bachelor's degree in information engineering from the Shanghai Jiaotong University, China, in 2013. She is currently working toward the PhD degree in computer engineering from the Department of Electrical Engineering and Computer Science, University of Michigan. Her research focuses on enabling high-efficiency wireless networks, including NetMIMO networks, MU-MIMO networks, and wireless LANs.

Xinyu Zhang received the BE degree in 2005 from the Harbin Institute of Technology, the MS degree in 2007 from the University of Toronto, and the PhD degree in 2012 from the University of Michigan. He is an assistant professor in the Department of Electrical and Computer Engineering, University of Wisconsin-Madison. He was a research intern at the Microsoft Research Asia from May to August 2010, and at the NEC Labs American from May to December 2011. His research interest lies in designing cross-layer protocols that improve wireless network performance, as well as mobile applications that enable fine-grained context sensing. His work spans the areas of wireless networking, communications engineering, and mobile computing, involving both mathematical analysis and system implementation. He received the ACM MobiCom Best Paper Award in 2011, and the US NSF CAREER award in 2014.

Kang G. Shin is the Kevin & Nancy O'Connor professor of computer science in the Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor. His current research focuses on QoS-sensitive computing and networking as well as on embedded real-time and cyber-physical systems. He has supervised the completion of 75 PhDs, and authored/coauthored more than 830 technical articles, a textbook and more than 30 patents or invention disclosures, and received numerous best paper awards, including the Best Paper Awards from the 2011 ACM International Conference on Mobile Computing and Networking (MobiCom11), the 2011 IEEE International Conference on Autonomic Computing, the 2010, and 2000 USENIX Annual Technical Conferences, as well as the 2003 IEEE Communications Society William R. Bennett Prize Paper Award, and the 1987 Outstanding IEEE Transactions of Automatic Control Paper Award. He has also received several institutional awards, including the Research Excellence Award in 1989, Distinguished Faculty Achievement Award in 2001, and Stephen Attwood Award in 2004 from The University of Michigan (the highest honor bestowed to Michigan Engineering faculty); a Distinguished Alumni Award of the College of Engineering, Seoul National University in 2002; 2003 IEEE RTC Technical Achievement Award; and 2006 Ho-Am Prize in Engineering (the highest honor bestowed to Korean-origin engineers).