2015

A survey of single and multi-hop link schedulers for mmWave wireless systems

Jian Song
University of Wollongong, js198@uowmail.edu.au

Kwan-Wu Chin
University of Wollongong, kwanwu@uow.edu.au

Publication Details

J. Song & K. Chin, "A survey of single and multi-hop link schedulers for mmWave wireless systems," Ad Hoc Networks, vol. 33, pp. 269-283, 2015.

A survey of single and multi-hop link schedulers for mmWave wireless systems

Abstract
Wireless communication at 60. GHz, aka mmWave, provides extremely high data rates, i.e., several Gb/s. Moreover, devices have a much shorter transmission range as compared to those operating in the 2.4 and 5. GHz bands. Indeed, links can be treated as pseudo-wires with minimal interference leakage. As a result, future 60. GHz systems will have very high spatial reuse. This, however, is at the expense of high propagation loss, which can be overcome using directional or smart antennas. Another promising solution is to employ relays to boost the signal of weak links. In particular, if relays are properly selected, they are able to offer higher data rates than direct links, and also help circumvent obstacles. To this end, we review state-of-the-art schedulers that take advantage of the high spatial re-use afforded by 60. GHz wireless systems to activate multiple links within a channel time allocation. Moreover, we survey works that use passive and active relays to overcome obstacles and to facilitate novel applications. We also survey those that maximize both spatial reuse and throughput of both direct and indirect (relay) links simultaneously.

Keywords
multi, hop, link, single, schedulers, mmwave, survey, systems, wireless

Disciplines
Engineering | Science and Technology Studies

Publication Details
J. Song & K. Chin, "A survey of single and multi-hop link schedulers for mmWave wireless systems," Ad Hoc Networks, vol. 33, pp. 269-283, 2015.

A Survey of Single and Multi-Hop Schedulers for mmWave Wireless Systems

Jian Song and Kwan-Wu Chin
School of Electrical, Computers, and Telecommunication Engineering,
University of Wollongong, NSW, Australia
kwanwu@uow.edu.au

Abstract

Wireless communication at 60 GHz, aka mmWave, provides extremely high data rates, i.e., several Gb/s. Moreover, devices have a much shorter transmission range as compared to those operating in the 2.4 and 5 GHz bands. Indeed, links can be treated as *pseudo-wires* with minimal interference leakage. As a result, future 60 GHz systems will have very high spatial reuse. This, however, is at the expense of high propagation loss, which can be overcome using directional antennas. Another promising solution is to employ relays to forward data from senders to receivers. In particular, if relays are properly selected, they are able to offer higher data rates than direct links, and also help circumvent obstacles. To this end, we review state-of-the-art schedulers that take advantage of the high spatial re-use afforded by 60 GHz wireless systems to activate multiple links within a channel time allocation. Moreover, we survey works that use passive and active relays to overcome obstacles and facilitate novel applications. We also survey those that maximize both spatial reuse and throughput of both direct and indirect (relay) links simultaneously.

1 Introduction

Advances in 60 GHz, also known as Millimeter-Wave (mmWave), communications have attracted increasing attention from both academia and industry in the past decade. In particular, the Federal Communications Commission (FCC) in the year 2001 freed the unlicensed frequency spectrum between 57 GHz and 64 GHz for commercial use, offering up 7 GHz in bandwidth [7]. Consequently, devices are expected to achieve multi-Gbps data rates and will be capable of supporting bandwidth intensive, upcoming applications such as uncompressed High-Definition Television (HDTV). Moreover, mmWave links are ideal replacements for wires that interconnect a computer and its external peripherals, and also allow users to exchange audio/video files over an ad-hoc network [4]. Consequently, 60 GHz is a key enabler of future, gigabits wireless networks, which encompasses next generation cellular communications [28] and body area networks [5].

A prominent characteristic of the 60 GHz frequency band is its high free space path loss due to atmospheric gaseous absorption. It is also affected by precipitation. Measurements have shown that attenuation peaks at 24 and 60 GHz, which coincide with the resonant frequencies of gas molecules [7]. Consequently, 60 GHz wireless systems have short communication ranges; in indoor scenarios, the reported range is around 10 meters. In addition, mmWave lacks the ability to diffract around obstacles that are more than 5mm in length, meaning obstacles such as humans, furniture and walls can easily degrade the signal strength by 20-30 dB. This is also referred as the link blockage problem. Moreover, as documented in [32], Line-of-Sight (LOS) is necessary because channel delay is primarily affected by reflection rather than diffraction and penetration.

Henceforth, 60 GHz systems employ directional antennas to combat propagation loss [36]. Advantageously, the millimeter wavelength allows multiple elements to be placed on a device [4], which in turn enables Spatial Division Multiple Access (SDMA) and high gains. However, this requires expensive multiple Radio Frequency (RF) chains. To this end, the IEEE 802.15.3c task group focuses on using Multiple-Input Multiple Output (MIMO) to increase transmission range only. Nevertheless, a key advantage of directional antennas is the high spatial reuse, which enables multiple concurrent transmissions. This is particular true

for 60 GHz systems due to their high path loss, meaning power leakage from side lobes is very small, and hence, their effect can be neglected. Therefore, given a suitable interference management policy that pairs transceivers properly, a 60 GHz system is capable of achieving high network capacity; see Section 2. For example, in [13], the authors show how using *flyways*, relays that create an indirect route using 60 GHz links, can be employed dynamically to increase the capacity of switches in data centers. We remark that 60 GHz links are ideal for use in data centers due to the empty space between the ceiling and racks, meaning line-of-sight is readily available.

The authors of [36] show that directional antennas may be able to overcome human-body or non Line-of-Sight (NLOS) blockage. Nevertheless, as obstacles on LOS path reduces the transmitted signal by up to 20-30 dB, high gain directional antennas may not be sufficient to alleviate significant path loss. To this end, helper nodes are usually deployed to relay packets around obstacles. They provide alternative routes that help boost links with high path loss. The assumption is that a relay has better link quality to both the sender and receiver. Consequently, the resulting two-hop path may offer higher throughput and better coverage.

Next, we review 60 GHz works that exploit the inherent high spatial reuse afforded by directional antennas and high path loss. Section 2 reviews works that aim to maximize network capacity by intelligently scheduling non-interfering links in each time slot. After that, in Section 3, we discuss works that exploit active and passive relays to overcome blockages; e.g., humans. In Section 4, we review three works that consider both spatial reuse feature and multi-hop solution simultaneously. Some future works and our conclusions are presented in Section 5.

2 Single Hop

Currently, most single hop works are based on the IEEE 802.15.3c specification [17]. A IEEE 802.15.3c Wireless Personal Area Network (WPAN) consists of several devices (DEVs) and a PNC that manages all transmissions. The PNC and all DEVs rely on superframes to synchronize their transmissions. The structure of a superframe is shown in Figure 1. An IEEE 802.15.3c superframe consists of three parts: 1) a Beacon Period (BP), where DEVs are informed of access management information, and (ii) Contention Access Period (CAP), which is used by DEVs for asynchronous data communications. This is achieved using carrier sense multiple access/collision avoidance (CSMA/CA), and (iii) Channel Time Allocation Period (CTAP), whereby DEVs are provided with multiple CTA blocks and have exclusive use of their allocated CTAs. In other words, they access the channel in a Time Division Multiple Access (TDMA) manner.

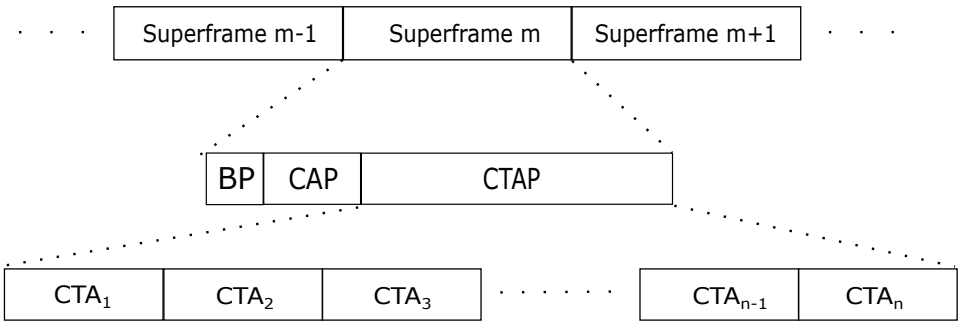
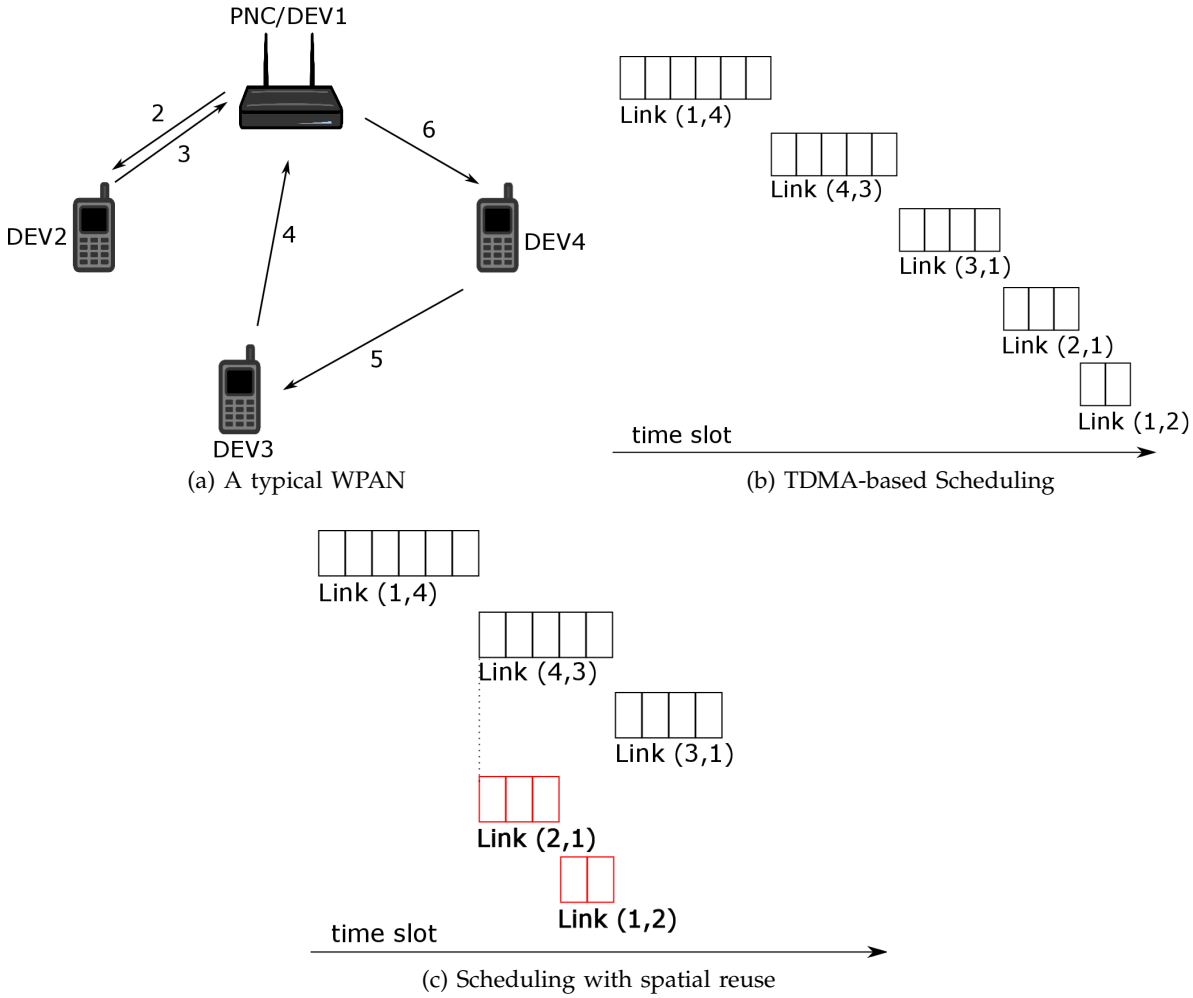


Fig. 1. IEEE 802.15.3c superframe structure.

A key problem in IEEE 802.15.3c is that a CTA block is dedicated to one DEV only. Ideally, we want multiple non-interfering links to co-exist in each CTA. Unfortunately, IEEE 802.15.3c does not allow the PNC to maximize spatial reuse. Figure 2 shows the significant improvement to be had if there is spatial reuse. In Figure 2a, we see four DEVs and DEV1 is selected as the PNC. All transmissions are direct since there are no relays. The arrows represent uni-directional links, and the number next to each link represents the transmission demand, in terms of slots, of each node pair. For example, DEV1 has requested six time slots for its transmission to DEV4. Figure 2b and 2c show possible scheduling results with and without exploiting spatial reuse respectively. Each block represents one time slot. As only one link is allowed to

transmit in each slot using TDMA, the schedule length is $6+5+4+3+2=20$ time slots. As we can see, when link $(4,3)$ is active, link $(2,1)$ or $(1,2)$ is able to transmit with link $(4,3)$ simultaneously. An improved schedule that exploits spatial reuse can be seen in Figure 2c. A total of $6+5+4=15$ time slots are required to finish all transmissions. The above results show that the current IEEE 802.15.3c standard lacks the ability to take advantage of spatial reuse. In fact, the link scheduling problem is left to vendors.



(a) A typical WPAN

(b) TDMA-based Scheduling

(c) Scheduling with spatial reuse

Fig. 2. Improving system throughput in a mmWave WPAN via spatial reuse.

We now review works that exploit spatial reuse. As we will see, the recurring theme is determining non-interfering links. The other is determining the combination of links that reduces total transmission time. In [1] and [3], the authors propose a directional transmission scheduling (DTS) algorithm for 60 GHz based WPANs. The key idea is to schedule multiple non-interfering links in the same CTA blocks to enable concurrent transmissions. To do this, a link co-existence test (LCT) is carried out before scheduling. LCT considers overlapping beam sectors when estimating the interference between two links. Consider Figure 3. There are three DEVs, namely $D0, D1, D2$, that are one-hop neighbors of each other. For D0, its azimuth plane is divided into $N_B = \frac{360}{\theta \div 2}$ sectors, where θ is the beamwidth. D1 and D2 are located in D0's sector 2 and 3 respectively. Denote $I_{D1}^{D0} = [1,2,3]$ as the directional signature of D1 to D0. Thus $I_{D2}^{D0} = [2,3,4]$ is the directional signature of D2 to D0. Here, $I_{D1}^{D0} \cap I_{D2}^{D0} \neq \emptyset$. Hence, LCT concludes that D1 and D2 interfere with each other if they both transmit to D0 simultaneously. In other words, two transmissions can co-exist only if they do not share the same receiver. The DTS procedure is divided into three parts. It first conducts a LCT if new link requests are received by the PNC. A new link is added if it does not interfere with scheduled links and if any of the following criteria are met: (i) the requested channel time of the new link is smaller than the current CTA, and (ii) when the requested channel time of the new link is larger than

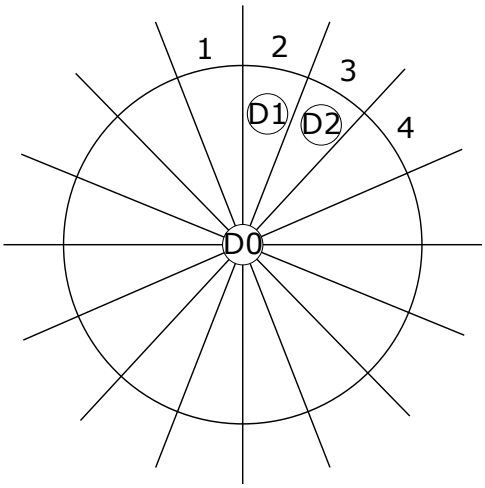


Fig. 3. The beam sector division of node D0.

the current CTA but smaller than the remaining channel time of CTAP. For criterion (ii), the length of the current CTA will be extended to be the same as the length of the new link's requested channel time. For example, a new link requesting two time slots and the current CTA has a length of only one time slot. In order to schedule the new link in the current CTA, the length of current CTA is extended to two time slots.

In [35] and [34], the authors propose a Virtual Time Slot Allocation (VTSA) scheme that allows multiple links to receive service in each CTA. In order to ensure the co-channel interference (CCI) between links is minimal, the PNC must be aware of the interference caused by each transmitting device to other devices. To this end, the first slot of the CTAP contains a Probing Signal Broadcasting Period (PSBP) whereby devices broadcast a short frame, consisting of a preamble and header, in their assigned slot in the PSBP. Other devices then measure the resulting received power and report it to the PNC. At the end of the PSBP, the PNC knows the level of interference caused by a transmitting link. The PNC then runs the VTSA algorithm. Recall that in IEEE 802.15.3c, each CTA is allocated to one link. Howevever, in the VTSA algorithm, once all CTAs are allocated, the PNC checks to see whether more links can be added into a CTA. If so, it is designated as a virtual CTA, and is assigned to one or more unscheduled links according to two methods: 1) minimum-CCI slotting method (MCSM), where an unscheduled link is allocated to the virtual CTA that causes the minimum CCI, and (2) random slotting method (RSM), where a CTA is allocated randomly. In both methods, after each allocation, the algorithm checks to determine if the CCI of the selected virtual CTA exceeds a threshold. If not, the link will be granted the current virtual CTA. The algorithm terminates when all links are scheduled. In a different work but with the same aim, Jin et al. [18] propose that devices monitor the transmission in each CTA. Consider the case where CTA-1 is used by device A for transmission to B. Other devices monitor the channel to determine where they can detect the transmission. If a device, say D, detects an idle channel, then it can potentially be a receiver in CTA-1; say from device C. This information is then relayed to the PNC, which then schedules both A to B and C to D simultaneously. The PNC also conveys to these devices the start and end data transmission and acknowledgment time to ensure there is no interference.

In [24], the authors propose a randomized exclusive region (ER) based scheme to schedule multiple senders simultaneously. An ER is defined as an interference region around a receiver. If the transmitter of a flow is located outside the ER of another flow's receiver, then these two flows are allowed to transmit concurrently. To find the size of an ER, the authors use the Shannon-Hartley theorem and free space path loss model. Specifically, for a given flow i, its average received power is modeled as

$$P_R(i) = k_1 G_T(i) G_R(i) d_i^{-\alpha} P_T(i) \tag{1}$$

Here, k_1 equals $(\frac{\lambda}{4\pi})^2$, where λ is the wavelength and α is the path loss exponent that ranges from 2 to 6. In

Equ. 1, $G_T(i)$ and $G_R(i)$ are the antenna gains of the transmitter and the receiver of flow i, respectively. The distance between the transmiter and receiver of flow i is denoted as d_{ij}. Lastly, $P_T(i)$ is the transmission power.

Assume there are N slots in a superframe and each slot has only one flow. Then the average data rate of a flow i over N slots is given by

$$R_i = \frac{k_2 W}{N} log_2 \left(1 + \frac{P_R(j)}{N_0 W}\right) \tag{2}$$

where k_2 is a coefficient that is governed by the efficiency of transceiver design and N_0 is the one-sided spectral density of white Gaussian noise. In the case where all flows transmit simultaneously in all slots, the achieveable data rate of flow i is

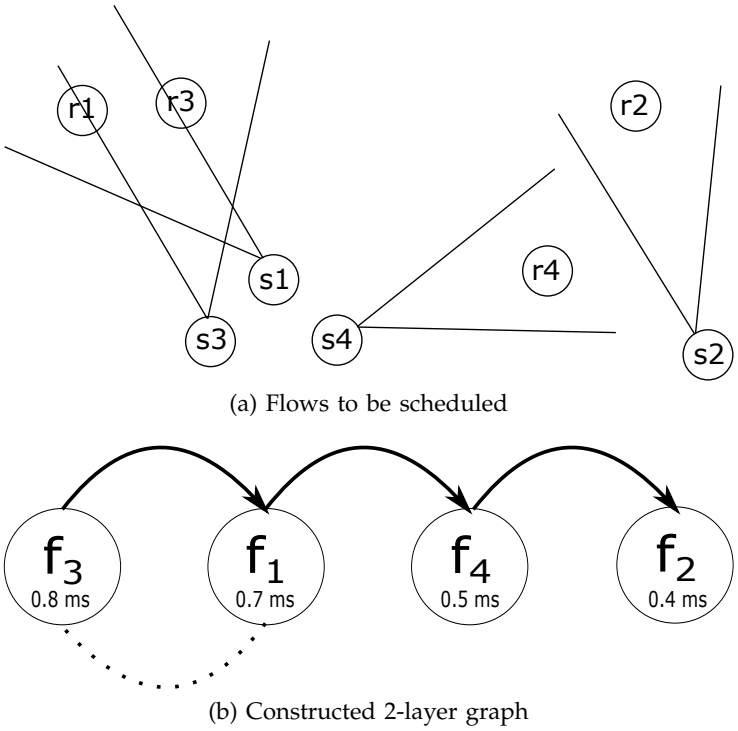
$$R_i^{'} = k_2 W log_2 \left(1 + \frac{P_R(i)}{N_0 W + \Sigma_{i \neq j} I_{j,i}}\right) \tag{3}$$

where $I_{j,i}$ denotes the interference between the transmitter of flow j and the receiver of flow i. To achieve $R_i^{'} \geq R_i$, meaning flow i has the same average rate as if it is the only flow in a given slot, the authors of [24] show that the term $I_{i,j}$ must be no larger than $N_0 W$. In particular, this condition is met when an interferer j is at separated from the receiver of flow i by the distance $r(i)$, which is formally defined as

$$r(i) = \left(\frac{k_1 G_0 G_T(i) G_R(j) P_T(i)}{N_0 W}\right)^{\frac{1}{\alpha}} \tag{4}$$

For a given flow i, its ER is defined as the region at distance $r(i)$ from the receiver of flow i. With Equ 4 in hand, the authors proceed to define the ER for scenarios when the transmitter/receiver uses omni or directional antennas. In other words, depending on the type of antenna used for transmission/reception, the ER of a flow will be different. After that, the authors propose a randomized exclusive region scheduling (REX) scheme to schedule concurrent peer-to-peer transmissions. Initially, REX allocates a counter $T_a = 0$ to each flow, where T_a is the number of slots allocated to a flow. REX randomly chooses a flow with the minimum T_a and schedules it to the current time slot. Then REX checks the remaining flows one by one with the minimum T_a and schedules the flows that are mutually outside the ER of other flows. REX terminates when all flows are scheduled. Similar to [24], the authors of [46] propose a power control based ER scheduling algorithm in order to increase system throughput. Instead of using a fixed transmission power, the authors consider different transmission powers, which yield variable ER ranges. The authors employ the Shannon-Hartley theorem and Saleh-Valenzula model (S-V model) [25] to derive the concurrent transmission condition; i.e., the interference at the receiver is smaller than the background noise. Then according to different transmission power levels, a set of ERs is derived, which is then fed to the scheduling algorithm in [24].

The scheduling algorithm in [6] aims to maximize spatial reuse in a CTAP. To describe the data flow information and the interference relationship between flows, the authors use a two-layer flow graph $G = (V, E_1, E_2)$ to capture flows and interfering flow pairs. In the two-layer flow graph, there is a set V of flows to be scheduled, E_1 represents flows sorted in descending order of requested transmission time slots and E_2 represents the interfering flow pairs. The algorithm consists of three phases: layer-1 and layer-2 edge construction followed by scheduling. In the layer-1 phase, upon knowing the transmission request of each flow, the PNC sorts all flows in descending order of requested transmission time and constructs a layer-1 flow graph. Then in the layer-2 edge construction phase, the PNC determines the interference between flows according to the following conditions. Consider two flows f_1 and f_2. If their respective sender is located outside the reception beams of the corresponding receiver, then both flows are non-interfering pairs. Otherwise, f_1 and f_2 interfere with each other. Lastly, in the scheduling phase, the algorithm picks the flow with the highest number of requested time slots from the layer-1 flow graph and assigns it one CTA. Then the algorithm moves to the next flow with the second highest number of requested time slots and checks the layer-2 interference graph to see if the current flow interferes with the scheduled flow. If interference exists, a new CTA is scheduled to the current flow. Otherwise, the current flow will be added to the old CTA to enable concurrent transmissions. The algorithm terminates when all flows are scheduled.



(a) Flows to be scheduled

(b) Constructed 2-layer graph

Fig. 4. An example of scheduling concurrent transmissions using 2-layer scheduling algorithm.

To explain the scheduling algorithm, consider Figure 4. There are four flows $f_1(s_1, r_1)$, $f_2(s_2, r_2)$, $f_3(s_3, r_3)$, $f_4(s_4, r_4)$, each requesting 0.7, 0.4, 0.8, 0.5 ms respectively from the PNC. As we can see, the transmission of flows $f_1(s_1, r_1)$ and $f_3(s_3, r_3)$ interferes with each other if they both transmit at the same time. This is because the receiver of flow f_3, i.e., r_3, is located within the transmission range of sender s_1 while receiver r_1 is located within the transmission range of sender s_3. The scheduling algorithm first constructs a two-layer flow graph. As shown in Figure 4b, the algorithm constructs layer-1 flow graph by sorting all flows in descending order as per their requested transmission time, i.e., f_3, f_1, f_4 and f_2. The arrows show the construction order of all flows. In the next phase, the algorithm constructs a layer-2 interference graph. This means the interference relationship of all flows is added to the layer-1 flow graph. As we can see from Figure 4b, a dotted curve connects flows f_3 and f_1 because they interfere with each other's transmission when scheduled to the same time slot. In the last phase, scheduling is performed. The algorithm first schedules flow f_3 with CTA_1. Then it moves to the next flow in the 2-layer flow graph, i.e., flow f_1. The algorithm checks if interference occurs when assigning f_1 to CTA_1. We can observe from Figure 4b that there is an interference edge between flow f_3 and f_1, which means f_3 and f_1 are not allowed to transmit at the same time. Therefore, flow f_1 is assigned CTA_2. The algorithm then moves to the next flow f_4 to check if it is possible to schedule f_4 to CTA_1. As there is no interference between f_4 and f_3, f_4 can transmit to f_3 concurrently. After f_4, the algorithm moves to the last flow f_2 and schedules f_2 to CTA_1 as f_2 does not interfere with any flow in CTA_1. The algorithm terminates at this point.

Son et al. [33] propose a Frame-based scheduling Directional Medium Access Control (FDMAC) protocol; a key departure from the IEEE 802.15.3c specification that uses the frame structure shown in Figure 5. FDMAC has two phases: 1) scheduling T_{sch}, and 2) transmission T_{tr}. In the scheduling phase, the PNC calculates a concurrent transmission schedule using collected transmission requests from all DEVs. Then in the transmission phase, all DEVs cooperate to transmit simultaneously according to the PNC's calculated schedule. As we can see from Figure 5, a T_{poll} period occurs before the scheduling period T_{sch} and a T_{push} period is after the scheduling period. During T_{poll}, the PNC polls each DEV for its transmission request. Then after the schedule is computed, the PNC sends the schedule to each DEV during T_{push}. The PNC constructs a demand matrix D based on the collected transmission requests during T_{poll}. The traffic

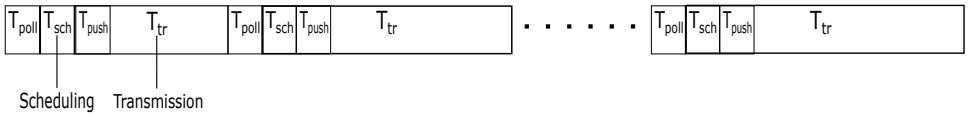


Fig. 5. A frame-based MAC structure.

demand matrix D consists of elements d_{ij} that correspond to the number of time slots required by the transmission from node i to j. According to the demand matrix D, the PNC then calculates a transmission schedule that maximizes concurrent transmissions. The authors assume all directional transmissions are pseudo-wires [26]. In other words, links are considered highly directional such that interference from neighboring transmissions is negligible. Figure 6 shows a directional WPAN consisting of four DEVs with DEV1 selected to act as the PNC. The arrows represent uni-directional transmissions between two nodes and the values on the arrows denote the traffic demands d_{ij} on each link. Before the scheduling period T_{sch}, the PNC polls each node i for its traffic demand vector d_i. Note, d_i is the traffic demand vector that describes all traffic demands from node i to other nodes. In this case, for DEV1, we have $d_1 = [0, 2, 0, 6]$, meaning it has two and six packets to DEV2 and DEV4, respectively. Upon collecting all traffic demand vectors from each node, for this example, the PNC has the following traffic demand matrix D,

$$D = \begin{bmatrix} 0 & 2 & 0 & 6 \\ 3 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 \end{bmatrix} \tag{5}$$

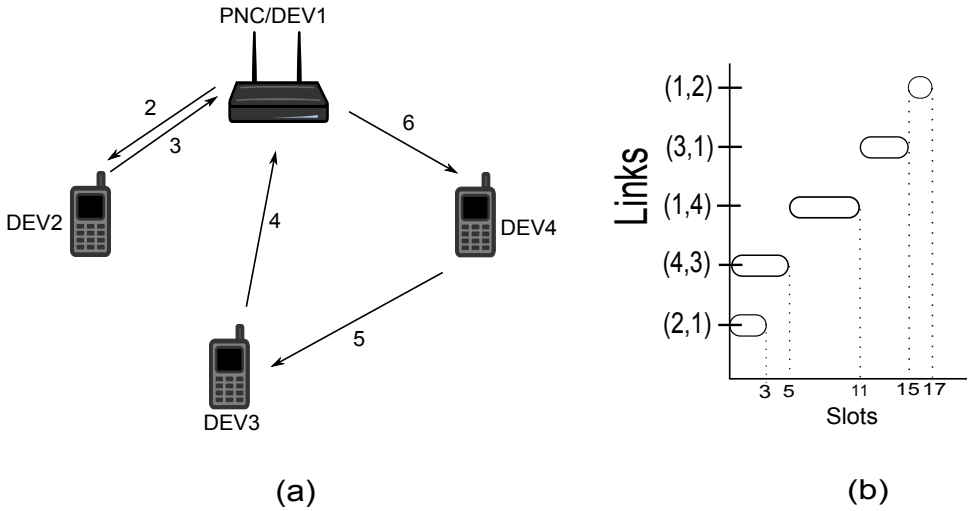


Fig. 6. An example 60 GHz WPAN; (a) number of packets to be transmitted from each device, and (b) a possible schedule.

A transmission schedule is then computed from the traffic demand matrix D; also see Figure 6(a). A possible concurrent transmission schedule is as follows; see Figure 6(b). The PNC schedules i) five time slots for link (2,1) and link (4,3), where link (2,1) is active for three time slots and idle for two time slots, ii) six time slots for link (1,4), iii) four time slots for link (3,1), and iv) two time slots for link (1,2). From the transmission schedule, a total of $5 + 6 + 4 + 2 = 17$ times slots are required to send all packets. In comparison, using a traditional TDMA-based scheduling policy, whereby one link is active in each time slot, we will need a total of $6 + 5 + 4 + 3 + 2 = 20$ time slots.

As we can see, a proper scheduling algorithm is necessary to schedule concurrent transmissions in order to minimize transmission delays. To this end, the authors of [33] formulate the problem as an NP-hard mixed integer non-linear programming (MINLP) problem. The Reformulation-Linearization Technique (RLT) is then applied to obtain a mixed integer linear programming (MILP) relaxation [21]. The authors

show that the optimal scheduling problem is similar to the K-edge coloring problem. The difference is that the objective is to minimize the total transmission time, whereas the goal of solving the conventional coloring problem is to find the edge chromatic number. To solve the said MILP problem in real-time, the authors propose a Greedy Coloring (GC) algorithm to compute a near-optimal solution. Initially, the algorithm obtains a directed and weighted multigraph $G(V, E)$ from the demand matrix D and sorts all links in descending order with respect to the number of requested time slots. Then the proposed algorithm carries out the following steps for each slot. Initially, the algorithm picks one link that has the highest weight and places it into a matching matrix, whereby a node belongs to only one link in a matching matrix; i.e., forms an independent set. The algorithm then moves to the next link with the highest weight, and adds it if the link is independent from nodes in the matching matrix.

In [30], the authors propose a spatial-time division multiple access (STDMA) algorithm that schedules both non-interfering and interfering links simultaneously. The authors introduce the concept of multi-user interference (MUI). Higher MUI is incurred when there are more concurrent transmission links in the same time slots. To achieve the minimum required data rate, a limited number of links is scheduled in the same time slots. The authors derive the achievable data rate of flow i, with source and receiver s_i and r_i respectively, over an Additive White Gaussian Noise (AWGN) channel as follows

$$R_i \leq \eta W log_2(1 + \frac{kP_t d_{I,i}^{-\gamma}}{WN_0 + b\Sigma_{l \neq i} f_{l,k} kP_t d_{l,i}^{-\gamma}}) \tag{6}$$

where η describes the transceiver design efficiency, ranging from zero to one, $k = 10^{PL(d_0)/10}$ is the constant scaling factor corresponding to the reference path loss, γ is the path loss exponent ranging from zero to six, b is the MUI factor, $f_{l,k} = 1$ if sender s_l and receiver r_k direct their beam towards each other. Given the above data rate characterization, the next problem is to maximize the number of concurrent links while satisfying each flow's minimum data rate requirement. The authors formulate the problem as a non-linear integer programming problem, which is similar to the NP-complete 0-1 Knapsack problem. However, the optimal solution always favors the flows with high data rates and starves flows with low data rates. To overcome this unfariness, the authors formulated the problem as a non-convex integer problem, and propose a polynomial time, slot based concurrent transmission approach. In each time slot k, there is a decision vector $U_k = [u_{k,1}, u_{k,2}, \cdots, u_{k,N}]$, where $u_{k,i} = 1$ indicates that flow i is scheduled in time slot k and $u_{k,i} = 0$ indicates flow i cannot be scheduled in time slot k. To determine which flow is active in each time slot, the authors propose a flip-based algorithm. A new link is only added to the active set if it does not reduce the data rate of other links; i.e., if adding the link does not cause the MUI experienced by other links to increase to a level that reduces their data rate. In addition, they aim to minimize fractional flows by ensuring flows that are active in the previous slot but have yet to receive their data rate requirement continue to remain in the active set. The algorithm terminates when all time slots are scheduled.

In [11], the authors consider a directional CSMA/CA protocol. Nodes first derive the beam weight that yields the maximum gain to each other. After that, they beamform toward the PNC, which then schedules an interference measurement phase to ascertain the set of links or pairs of nodes that can transmit simultaneously. It then groups non-interfering links together and assigns them an ID. Channel access is then carried out as follows. Consider a station labeled 'A'. It first informs the PNC its intention to transmit to say station 'B'. The PNC then broadcasts a Target Clear to Send (CTS) message to all stations, which contains a given transmission opportunity (TxOp) and group ID. Nodes belonging to the same group ID are then permitted to transmit simultaneously with station 'A' for TxOp time. Other nodes not in the same group set their Network Allocation Vector (NAV) accordingly. In a subsequent work, the authors consider beamforming to multiple users simultaneously [12]. To schedule downlink transmissions, the PNC transmits a Request to Send (RTS) message to all nodes, which is identified using a multicast address, that it can simultaneously beamform to. These nodes then reply with a mmWave CTS message to inform the PNC that they are ready to receive. The PNC then beamforms to these users. As for uplink transmissions, a node, say A, first sends a RTS to the PNC, which then replies with a uplink clear CTS message to all members in the same multicast group as node A. These nodes then beam form to the PNC to transmit their data.

As pointed out in [14], a key problem with the approach in [1] is that it assumes all devices are aligned on the same axes. In particular, upon booting up, a node will need to determine its coordinate, and orient its axes to that of the PNC, which is located at coordinate $(0, 0)$. This ensures the PNC is able to determine the coordinate of any transmitter and receiver pairs and decides whether they are interfering, and if not, schedules them to transmit concurrently. To this end, Hsu et al. [14] design a scheduler that considers both axis alignment and location of devices. A node aligns its axes as follows. A PNC transmits its beacon at an angle θ. The node then replies with an association request in the direction $(\theta + 180)$. If there is no response, it rotates its X-axis counterclockwise by α, and repeats the process until it receives a response. Once aligned, it and the PNC becomes an anchor used by other nodes to triangulate their position. Denote the angle of transmission from the PNC as θ_1, and from an anchor B as θ_2. Their respective coordinate is (0,0) and (x_B, y_B). A new node C, determines its coordinate as,

$$(x_C, y_C) = \left(\frac{y_B - (\tan\theta_2)X_B}{\tan\theta_1 - \tan\theta_2}, \frac{(\tan\theta_1)y_B - (\tan\theta_1)(\tan\theta_2)x_B}{\tan\theta_1 - \tan\theta_2} \right) \tag{7}$$

In addition, the authors also consider improving accuracy by selecting anchors whose beam yields the smallest overlap region. Nodes with packets to transmit then send a request, with the corresponding transmission time, to the PNC. The PNC then determines the set of simultaneously transmitting nodes. It also sets the transmission time to the maximum time requested by nodes in the set based on their coordinate.

2.1 Discussion

Table 1 summarizes and compares the key features of each method presented in Section 2. As we can see, all works aim to schedule concurrent transmissions in the same time slot to take advantage of the high spatial reuse afforded by the 60 GHz band. A fundamental problem is thus to determine the set of concurrent or non-interfering transmissions. In [35], the authors require devices to conduct measurement and pass the result to the PNC at the beginning of each CTAP. This ensures the channel condition is current but requires a dedicated sub-frame for channel probing. In [18], no such sub-frame is used. Instead, devices monitor the transmission of other devices in each CTA. In both works, the PNC schedules links in the same slot as long as their transmission does not cause the interference at receiving devices to exceed a given threshold. Other works, however, are more theoretical in nature. For example, LCT [1] divides an area based on the beamwidth of a node's directional antenna, and schedules links to transmit together as long their beams do not intersect. A key weakness, however, is that reference [1] does not consider the radio propagation characteristics of 60 GHz links. Morever, as noted in [14], it assumes all devices have the same coordinate axes. The authors of [24] introduce the concept of ER, where if no other devices transmit within a link's ER, then it experiences the same data rate as the case where there is no spatial reuse. In addition, the authors consider the interferences caused by omni and directional transmissions. A key limitation, however, is that they assume free space radio propagation. Moreover, they did not consider different data rates and assume devices have a fixed transmission power. The former weakness is addressed in [30], and the latter by [46]. In addition to maximizing spatial reuse, Chao et al. [6] also consider the traffic demands of flows. Specifically, their scheduler preferentially picks flows in order of requested transmission slots. Similarly, Son et al. [33] consider transmission times but aim to determine a schedule that ensures all links are scheduled in the quickest possible time; i.e., determine the set of transmitting links in each slot such that the total number of required slots is minimal. An interesting future research direction is to develop approximation algorithms that bound the schedule length given *any* set of flows and correponding demands. Apart from that, a key assumption of [33] is that links are effectively pseudo-wires. Thus, it will be interesting to see whether their approach is applicable over the communication model described in [24] or [46]. It is worth noting that the works discussed thus far do not aim to route around blocked links. Instead, if a link is blocked, meaning after measurement its signal strength is below an acceptable threshold, it is excluded from the scheduling process. As a result, the end nodes of the link will be disconnected from each other.

TABLE 1
Summary and comparison of spatial reuse methods.

Method	CTAP?	Interference Management	Aim
[1] [3]	Yes	Link Co-existence Test (LCT)	To maximize concurrent transmissions per CTA
[35]	Yes	Co-channel interference (CCI) monitoring table	To maximize concurrent transmissions subject to CCI_{all}
[18]	Yes	By monitoring the transmission of a CTA owner, and informing the PNC	To maximize spatial reuse
[24]	Yes	Exclusive Region (ER)	To maximize concurrent transmissions per time slot while ensuring fairness for each link
[46]	Yes	ER with power control	Same as ER-based REX [24]
[6]	Yes	Layer-2 interference graph	To maximize concurrent transmissions per CTA subject to layer-2 interference graph
[33]	No. Frame-based	None. Assume links are pseudo-wires	To minimize delays
[30]	Yes	Multi-user interference (MUI)	To maximize concurrent transmissions per time slot subject to MUI
[11]	No	Interference measurement and arbitration by an AP	Resolve deafness and maximize spatial reuse
[12]	No	AP transmits RTS and CTS messages to multicast group members	Exploit multi-user diversity
[14]	Yes	Use nodes' coordinate	Maximize spatial reuse

3 MULTI-HOP

As we discussed earlier, the link blockage problem is unique to 60 GHz systems, primarily due to its short, five millimeter wavelength. This means any moving objects (human) or fixed obstructions (furniture and walls) will block or severely degrade transmitted signals. This problem thus requires the use of alternative paths comprising of relays.

Generally, there are two types of relays: passive [39] and active [31][42][41][38]. Briefly, a passive relay simply reflects an incoming signal to an intended receiver. Passive relays are usually placed at fixed positions such as on a ceiling. As its name implies, passive relays are simply reflectors and thus are not capable of adjusting the direction of reflection to match any changes. Therefore, passive relays are only used in fixed networks; e.g., data centers. Apart from that, the deployment of passive relays introduces high signal attenuation. This is exacerbated by the high path loss of the 60 GHz signal coupled with elongated relayed paths. An active relay functions as a signal repeater and more importantly, it is able to forward a signal onto any direction using beamforming techniques or directional antennas if needed. Therefore, active relays are ideal in dynamic WPANs.

As mentioned in Section 2, the IEEE 802.15.3c specification does not support relays. This is not the case for ECMA-387 [9].

Next, we present works that use active relays in 60 GHz systems. After that, in Section 3.2, we review works that employ passive relays.

3.1 Active Relays

In [31], the authors propose a multi-hop MAC protocol using active relays to maintain network connectivity when a direct link is blocked by moving obstacles. As they assume 60 GHz WPAN devices are equipped with directional antennas, conventional carrier sensing methods are not applicable. The authors propose a multi-hop relay directional MAC framework that is composed of four procedures: 1) discovery process, 2) normal operation mode, 3) trailing control phase, and 4) lost node discovery. Next, we briefly explain these procedures. During the discovery process, the PNC broadcasts HELLO messages to discover DEVs. It then uses the responses from DEVs to create a network topology map. The PNC also iteratively designates each DEV to send HELLO messages, and requests the DEVs to create their own network topology map. Once the discovery process ends, every DEV sends back its network topology map to the PNC. After the discovery process, the PNC and DEVs operate in normal mode. Specifically, the PNC polls each DEV in a round robin fashion. A DEV must respond to the PNC even though it has no data to send. This helps the PNC detects connectivity to each DEV. Then in the trailing control phase, the PNC and DEVs sequentially send HELLO messages to track changes. Both the PNC and DEV update their network topology map

if they are any changes. If a DEV is disconnected, i.e., no reply after a poll, the PNC assigns it a relay based on the network topology map. The relay then sends a HELLO message to the disconnected DEV by steering its beam in the direction of the disconnected DEV. If the disconnected DEV receives the HELLO message and responds, the relay measures the received signal strength and reports the result to the PNC. Otherwise, if the discovery process fails, the PNC selects another DEV as a relay. Upon establishing a relayed path, the PNC reverts back to normal mode.

In [42], the authors propose a multi-hop routing protocol, which at its core is a hop discovery process. To detect link failure and discover relay nodes, the PNC applies multi-hop routing protocols such as dynamic source routing (DSR) or ad hoc on-demand distance vector routing (AODV) [27]. The route request (RREQ) packets are sent to all neighbors in an omni-directional manner; this is achieved by having DEVs send out RREQs in all directions, which introduces high overheads. To this end, the authors of [42] propose to let the PNC sweep its antenna elements one by one. For example in Figure 7, link (PNC, C) is blocked. The PNC first sends out RREQ packets via antenna element i to locate node C. As there is no node other than node C in element i, and node C is blocked, there is no response after a timeout period. The PNC will then send out RREQs via elements $i+1$ and $i-1$. In this case, node B replies to the PNC. In general, this discovery process does not stop until a relay is found or the PNC has attempted all elements.

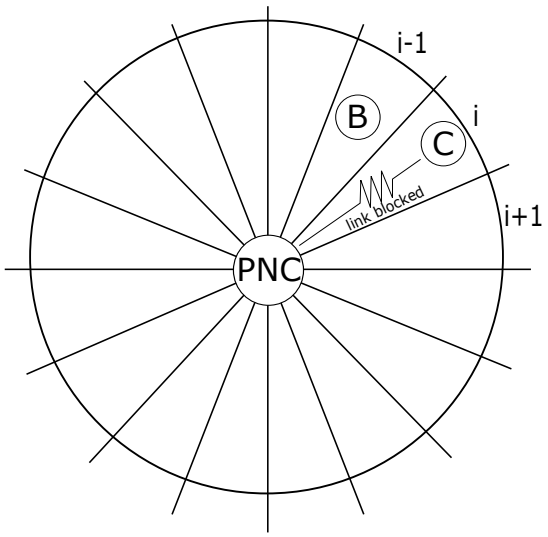
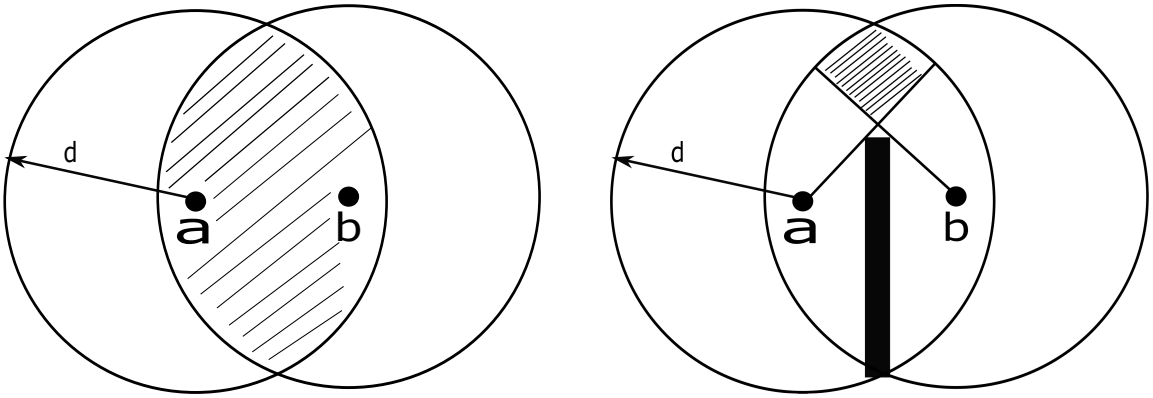


Fig. 7. An illustration of the relay discovery process.

Once the PNC discovers a relay node, in this case, node-B, the relay forwards the RREQ received from the PNC to the next hop. If the direction of node C is recorded in the look-up table of node B, then node B sends the RREQ directly to node C. Otherwise, node B transmits the RREQ omni-directionally. This process terminates when the PNC receives a reply message from node C. Upon receiving RREQs, node C has two options: 1) it chooses the first arriving RREQ and replies with a route reply, or 2) it waits for a certain amount of time to record all possible routes. The first option ensures that node C chooses the shortest path but may have less than ideal signal quality. For the second option, node C will have to delay its reply in order to collect RREQs before comparing the minimum SNR of each discovered route. It then sends a reply via the route with the maximum minimum SNR.

Instead of making use of existing devices as relays, the authors of [41] sought to place relays in positions that optimize a given metric. In particular, they address the (i) Robust Minimum Relay Placement (RMRP) problem; i.e., use fewest relays to satisfy minimum system requirements, and (ii) Robust Maximum Utility Relay Placement (RMURP); i.e., use no more than a fixed number of relays to maximize network utility. Candidate regions, in which relays are located, must have both LOS to both transmitter and receiver, and must have distance no further than a threshold d. Example candidate locations are shown in Figure 8 as shadowed areas. In Figure 8a, we can see that node a and b are visible to each other because they are within each other's transmission range d. If the LOS path between a and b becomes blocked, relays can

be deployed in the shadowed area to form an alternative link. In Figure 8b, the LOS path between node a and b is blocked. To recover the connection between node a and b, a relay is needed to form a multi-hop path to bypass the obstacle. Hence, a relay must be placed at a position where it can see both a and b; i.e., in the shadowed area.



(a) Candidate locations for relays with no obstacles

(b) Candidate locations for relays when an obstacle (solid rectangle) exists

Fig. 8. Candidate locations for placing a relay.

The authors formulate RMRP as a Mixed-Integer Linear Program (MILP), and solve it using CPLEX [16]. As for RMURP, the goal is to find the candidate regions for a finite number of relays such that they maximize a given network utility and satify robustness constraints. Here, the authors define network utility to be the ratio of the achievable rates over the base rate. The authors formulate RMURP as a Mixed-Integer Non-Linear Program (MINLP), and proposed two algorithms: bisection search and Generalized Benders' Decomposition (GBD) technique [15]. Their results show that using bisection search is faster than GBD but is not necessarily optimal. On the other hand, GBD yields the optimal solution at the expense of high computational complexity.

In [38], the authors propose to use k repeaters to form multi-hop paths. The authors argue that forming a new multi-hop path may degrade the throughput of other nearby links. Moreover, a repeater may be requested by two links at the same time. Therefore, an efficient scheduler is needed to carefully manage the interference of neighboring links and the assignment of repeaters. The problem is as follows: given a set of n communicating pairs and k repeaters, for each pair, determine whether they are to communicate directly or via one of the k repeaters. As the problem is NP-hard, the authors propose a distributed greedy algorithm that operates as follows. Initially, each pair connects directly and it measures the best achievable data rate. If a pair's data rate is below 800 Mbps, the algorithm randomly chooses a repeater, and a new connection for the pair is formed via the chosen repeater. This process repeats until all links with less than 800 Mbps rate are re-routed. It is possible that some links that had a higher than 800 Mbps rate experience a rate lower than the threshold. The algorithm considers these links and iterates until no further improvement is observed in two consecutive time slots. We remark that the 'randomness' step yields a simple rule, and incurs little or no signaling overheads between nodes when deciding which repeater to use. This is particular pertinent in a distributed setting where communication is costly; e.g., battery operated devices.

A key problem is switching a transmission quickly from the primary path to a secondary or back-up path quickly, especially in an environment with human movement. For example, when the stream from a video player to a display is momentarily blocked by a human, the player needs to quickly reroute the stream via a relay to ensure continuity. In [37], the authors consider devices equip with an antenna array. The array thus allows a device to beamform to a neighbor. In this respect, the transmit weight vector, which defines the weight for each antenna element, must ensure the resulting beam optimizes transmission and reception. This requires devices to monitor the channel continously and critically, record the best weight for their antenna elements for a given neighbor. Consequently, for each link with end nodes i and j, there is a beam pair: $<w_i, w_j>$, where w_i and w_j are the transmit and receive weight vector, respectively. A path,

with a single or multiple links, thus consists of one or more beam pairs. For each path, the authors associate a number of metrics. Specifically, (i) aggregate gain of links on the path, (ii) number of times in which communication was successful, and (iii) SNR at the receiver. Each device maintains these metrics for all discovered paths. Upon detecting a significant drop in SNR, a node switches to the "best" alternative path. Here, the "best" path is selected using a composite metric that is a weighted combination of metric (i), (ii) and (iii). The switching process is fast as the beam weights are pre-computed, and hence, devices can quickly reroute their transmission whenever a blockage or human obstructs an on-going communication. In addition, devices monitor their paths periodically and switch whenever there is a better path.

Thus far, the reviewed works have only considered improving link quality or avoiding blockages. We remark that researchers have also taken advantage of the high rate and spatial reuse of 60 GHz links to enable new applications. For example, Kim et al. [20] consider using a two-hop 60 GHz wireless system to deliver live videos from multiple video cameras located in a stadium to a receiver or broadcasting center. They study the performance in scenarios where sources (video cameras) and relays have varying number of antenna elements. Specifically, a source may form a single or multiple beams to one or more relays. Also, relays may form a single or multiple beams to the receiver or broadcasting center, and have the ability to aggregate multiple video sources. For each scenario, the authors consider route selection and coding rate control with the goal of maximizing video quality. They formulated the problem as an Integer Linear Program (ILP) with the goal of maximizing video quality, represented as a concave function on the amount of delivered data at the receiver subject to minimum data requirements and link capacities. For multiple beams scenarios, there is a further constraint that ensures the number of connections/beams from each relay or source is bounded by the number of antenna elements. The authors then employ the branch-and-refine based algorithms in [23] to solve the resulting formulation, which is a non-convex mixed-integer nonlinear program. Another example is to provide additional capacity in a data center. In [13], the authors show that 60 GHz links are ideal for use in data centers. In particular, they install "flyways", i.e., 60 GHz point-to-point links, that act as short-cuts between server racks. In particular, flyways are activated when there are bottleneck links and serve to reduce congestion at the network core. As an example, assume the downlink to the Top of Rack (ToR) switch p is congested due to demands from ToR switches A, B and C; the route taken by a switch $x \in \{A, B, C\}$ is denoted as x–⊞–p, where ⊞ denotes the network core. In addition, there are the following flyways: (A, p), (B, p) and (C, p). One possible solution involves A forwarding its demand directly to p via the flyway (A, p). This solution thus removes A's demand from the path A–⊞–p. An interesting observation is that B, C or both may use the flyway (A, p), meaning they use A as a transit point for the demand to p; for example C–⊞–A–p. Similarly, for the uplink case, a ToR switch p evaluates whether it is worthwhile to reroute some of its demands via a neighbor ToR switch z; i.e., by enabling the flyway (p, z). For each congestion downlink/uplink, the authors exploit the aforementioned observations to design an algorithm that picks the path with the highest capacity to send a demand; the path can be either via the core as per-normal, directly over a flyway, or via another ToR switch and its flyway to the target ToR switch.

3.2 Passive Relays

In [39], the authors propose a 3D beamforming approach for use in a data center equipped with 60 GHz wireless links. The key idea is to deploy ceiling reflectors to bounce signals from the transmitter to a receiver. In [19] and [13], antennas are placed on top of server racks in a data center. The deployment of ToR antennas using 60 GHz wireless multi-gigabits links helps avoid changing the data center architecture, and thus reduce the cost of maintenance. However, ToR antennas suffer from two problems: interference from nearby ToR antennas during simultaneously transmissions, and high delays caused by increased number of hops.

The interference problem is shown in Figure 9. Rack 3 is within rack 1's transmission range. In this case, rack 3's antenna is interfered by the transmissions between rack 1 and 2. Figure 10 shows the reason for high delays when using multi-hop routes. If a rack at the edge of the network topology requests to communicate with the center rack, more than one hop is needed to forward data. Moreover, ToR antennas need to first receive and then re-orient their direction toward the receiver before transmission, and thus

incurs non-negligible delays. From Figure 11, we see that rack 1 needs four hops to reach rack 5, which will incur a high delay due to antenna orientation adjustment.

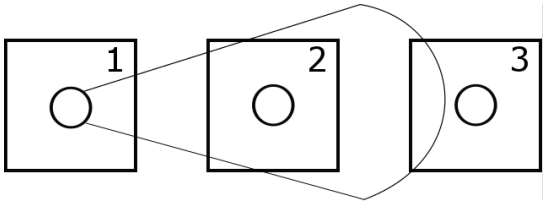


Fig. 9. Interference on nearby ToR antennas

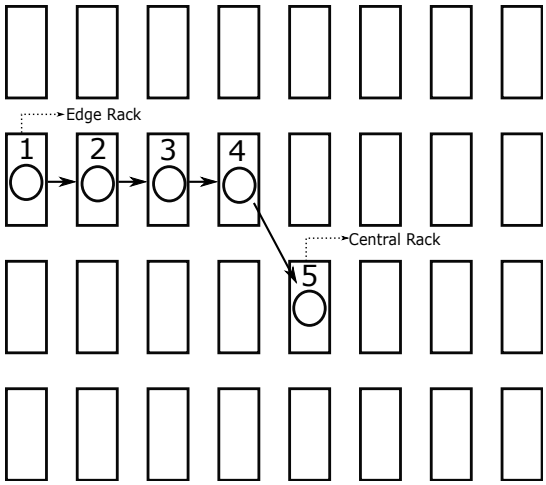


Fig. 10. Multi-hop with ToR antennas

To overcome the above problems, the authors propose to install reflectors on the ceiling to form 3D beamforming links. This approach can significantly reduce the interference range of the signal at the receiver. From Figure 11, we see that when rack 1 wants to communicate with rack 4, rack 1's antenna focuses its beam on the ceiling reflector so that the signal is redirected to rack 4's antenna. The formation of the link is simple: both the transmitter and the receiver point their beams at the ceiling reflector located halfway between them. In this case, only the area around the rack 4 is covered by the reflected beam. Therefore, the new path via the ceiling reflector both reduces the interference to nearby racks and decreases the delays caused by a multi-hop path.

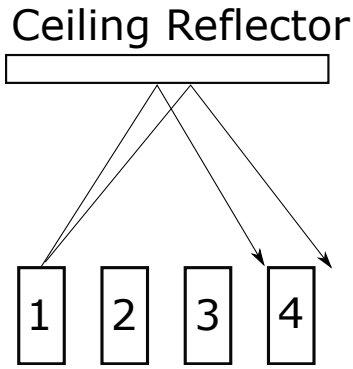


Fig. 11. Ceiling reflector.

As pointing antenna beams in both azimuth and elevation needs high accuracy, the authors suggest the use of rotators that achieve an accuracy of 0.09 degrees. For ceiling reflectors, they use microwave

reflectors such as flat metal or aluminum plates. To avoid local reflection on the receiver rack, they place electromagnetic absorbers on each rack to further reduce interference to nearby racks. To improve the performance of 3D beamforming technique, the authors propose a conflict-degree based greedy scheduling algorithm to maximize the number of concurrent transmissions while minimizing transmission delays in their later work [45]. Given a set of unscheduled links D^C, the conflict degree of link d_i is defined as the number of links from set D^C that cannot coexist with link d_i. To determine the coexistence of two links d_i and d_j on the same channel, the algorithm records the SINR at link d_i in the presence of link d_j. If the SINR at d_i falls below the threshold that degrades its data rate, then d_i and d_j cannot coexist. The scheduling problem is then formulated as a traditional graph coloring problem and is solved by techniques from [8] in a greedy fashion based on the conflict degree of each link.

In a similar work, the authors of [22] and [10] evaluate the use of passive relays placed strategically in different parts of a room. They consider amplify-and-forward relays, an AP on the ceiling and varying relays positions. They show relaying helps improve coverage and capacity in experiments with varying human densities. In [40], the authors consider a system comprising of a controller connected to multiple APs. The controller's job is to associate a station to an AP that results in the maximum reward or throughput subject to varying channel conditions caused by blockage. The key issue is to distinguish between interference and channel fading and blockage; an incorrect decision will cause stations to ping-pong between APs. To this end, they formulated the problem as a Partially Observable Markov Chain (POMC). In particular, the state space consists of the current state of APs; i.e., block or unblock. At each time slot, the controller must decide the AP to serve a given a station with the goal of maximizing its reward or throughput. In this respect, the authors show that the controller must balance between instantaneous reward and gaining more feedback in order to maximize future rewards. Consequently, they propose a simple threshold policy that is based on the number of failed transmissions.

The authors of [2] consider switching to a relay path when the direct link is blocked by a human, which manifests as fast attenuation, long fading duration and large fading amplitude. The authors investigate two beam switching mechanisms: instant decision and environment learning. In the former, the best alternative path is defined as one that has the highest SNR relative to the direct link and also has the widest angle to ensure that a human is not blocking the back-up path as well. In the second approach, the authors use an exponential weighted moving average to characterize relay paths according to their transmission sucesses and failures. The authors conclude that considering both beam direction and SNR yield superior performance.

3.3 Discussion

Table 2 summarizes and compares the similarities and differences between the multi-hop approaches presented in Section 3.1 and Section 3.2. In terms of active relays, a fundamental problem is to build a network or interference map, which is then used by the PNC to schedule concurrent transmissions. This map can be constructed via probing, e.g., [31], using HELLO messages. Unfortunately, due to the directionality of 60 GHz transmissions, multiple probe messages will be required to provide adequate coverage. Another problem is identifying the most suitable relays or routes quickly and also one that yields a high data rate. This can be carried out over existing relays or via manual placement; e.g., [41]. To date, only reference [20] has considered multi-path routing over relays. Their formulation, however, does not consider link outages. Moreover, it will be interesting to design a distributed solution for the same problem or system. As for passive relays, a number of researchers, e.g., [45], have applied 60 GHz links in data centers to increase the capacity between server racks. A fundamental problem is the configuration of reflectors, meaning a placement that maximizes SNR or a policy that exploits these reflectors for diversity gain in the presence of human movements. In this respect, all authors assume passive reflectors are pre-installed, and hence, incorporating dynamic relay placement strategies such as those in [41] will be an interesting future work.

In all works, there is an entity that determines when to use the direct or multi-hop path. Works such as [31] are centralized whereby the PNC actively sought out relays with the help of other devices. They tend to incur significant signaling overheads as the PNC and other devices must probe each other frequently

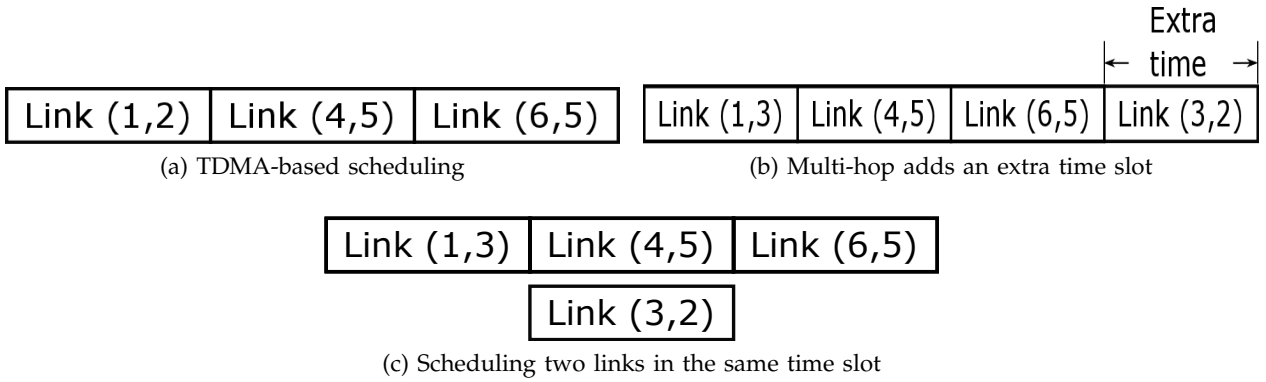
to ensure the topological information and channel condition are up to date. Also, works that use passive relays are centralized in nature. That is, all routes via relays are known and a source's objective is to adapt to changing channel condition by switching to a better path dynamically and quickly. On the other hand, works such as [13] and [38], use a distributed scheduling process where devices decide on their own the best route to use. The resulting schedulers are usually simple and require only local information; i.e., they do not require the complete topological information.

TABLE 2
Summary and comparison of multi-hop approaches.

Reference	Relay Type	Relay Device	Relay Discovery or Selection	Relay Location	Scheduling Control
[31]	Active	DEVs	Network topology map	Dynamic	Centralized
[42]	Active	DEVs	PNC sweeps its antenna in a clock-wise and anti-clockwise direction	Dynamic	Centralized
[41]	Active	DEVs	Solutions to RMRP and RMURP problems	Dynamic	Centralized
[38]	Active	Repeaters	Randomly chosen	Dynamic	Distributed
[37]	Active	DEVs	Learned dynamically during transmission and reception	Fixed	Distributed
[20]	Active	Relays	Routes determined by solving a non-convex mixed integer non-linear program	Fixed	Centralized
[13]	Active	ToR switches	Flyways that divert the most traffic away from a congested link	Fixed	Distributed
[39]	Passive	Reflectors	Beamform towards reflectors mounted on ceiling	Fixed	Centralized
[2]	Passive	DEVs	Best SNR and widest angle or transmission and reception successes	Fixed	Centralized
[22][10]	Passive	Reflectors	None. Relays amplify-and-forward signal	Fixed	Centralized
[40]	Passive	APs	Uses transmission errors for relay selection	Fixed	Centralized

4 HYBRID APPROACHES

To date, only three works have considered both spatial re-use and selecting the best relay paths concurrently. In [43], Zhou et al. propose two algorithms, random fit deflect routing (RFDR) and best fit deflect routing (BFDR) to schedule shared time slots to enable concurrent transmissions between direct links and multi-hop paths. The authors argue that although multi-hop relaying is effective in solving link degradation and blockage problems, the actual throughput of the network is decreased due to the extra airtime introduced by relaying. In Figure 12b, we see that $Link(1,2)$ is replaced with a multi-hop path $Link(1,3)$ and $Link(3,2)$ via relay node 3. This replacement adds an extra airtime if links are scheduled in a TDMA-based manner. However, if two links can be scheduled in the same time slot, the extra airtime in Figure 12b can be removed. In Figure 12c, $Link(3,2)$ is scheduled to share the same time slots with $Link(4,5)$. To schedule



(a) TDMA-based scheduling

(b) Multi-hop adds an extra time slot

(c) Scheduling two links in the same time slot

Fig. 12. Reducing the extra airtime incurred by multi-hop transmissions.

multiple links in the same time slot, the CCI on each link needs to be measured before scheduling. In particular, the authors derive the concurrent transmission condition as follows: a link is scheduled to a

time slot only if the CCI on every link in the same time slot is lower than a given threshold. The authors consider a network consisting of n pairs of DEVs. There are m pairs of DEVs requesting multi-hop operation from the AP, where $0 < m < n$. The problem is to find relays for these m pairs of DEVs and to schedule their transmission in a manner that maximizes system throughput. The authors propose two algorithms: RFDR and BFDR. For a $Link(d_s, d_d)$ that requests multi-hop operation, RFDR randomly selects a DEV d_j as a candidate relay. Assume the PNC has scheduled a timeslot t_u for pair (d_s, d_d). If the second hop path (d_j, d_d) can be scheduled in a timeslot t_l later than t_u and the CCI is acceptable, DEV d_j is chosen to be the relay for pair (d_s, d_d). RFDR repeats the procedure until all multi-hop requests are processed. On the other hand, BFDR checks every DEV to find the best relay. In particular, BFDR checks every time slot in order to locate one that causes the least CCI to links scheduled in a time slot. BFDR terminates when all m pairs are scheduled with a relay. In a subsequent work, Zhou et al. [44] present a Binary Integer Program (BIP) with an objective to maximize data rate for the same problem. As the problem is NP-hard, they reformulated the BIP as a maximum weight bipartite matching problem. The resulting problem is then solved using the well-known Kuhn-Munkres (Hungarian) algorithm.

The third work, see [29], presents a multi-hop concurrent transmission scheme that consists of a hop selection metric and a concurrent transmission scheme. The aim is to improve flow and network throughput. The problem is to find an optimal route such that the aggregated throughput is maximized and one that also load balances nodes. In addition, to explore spatial reuse efficiency, given a set of multi-hop routes, the problem is to schedule non-conflicting links on the routes to transmit simultaneously in a fair manner. To solve the multi-hop route selection problem, the authors use a weighted graph. The weight of an edge reflects both distance and traffic load; i.e.,

$$w(i,j) = \left[\frac{d(i,j)}{\bar{D}}\right]^2 + \frac{F(j)}{\bar{F}} \tag{8}$$

where $d(i,j)$ is the link distance between node i and node j, $F(j)$ is the traffic load on node j, $\bar{D}$ and $\bar{F}$ are the average link length and traffic load among all nodes respectively. The hop selection metric always selects the multi-hop route with the minimum aggregated link weights as computed by the Dijkstra algorithm. The hop selection method re-computes routes whenever the network topology changes or new flows are added. The authors then propose a concurrent transmission scheme that preferentially schedules links on a path in descending order of transmission loads. It also groups non-interfering links together, and assigns the group a time slot that is equal to the maximum scheduled time slot. It is possible that some links cannot be scheduled in the current superframe because there are insufficient time slots remaining in the transmission period. These links are removed from consideration and the corresponding flows are removed from the current schedule and are re-scheduled in the next superframe.

5 Conclusion

We have presented a didatic survey of approaches that aim to address the link blockage problem in 60 GHz systems via relays and also those that aim to exploit the high path loss to maximize network capacity. In general, the key problems covered include (i) maximizing the number of concurrent links within a time slot, whereby the links may belong to a direct or relayed paths, (ii) relays selection and placement, and (iii) identifying non-interfering links, which could be facilitated by a controller or carried out in a distributed manner, and involves channel measurement or deriving a suitable coordinate system. To date, as evidenced by our discussion in Section 4, not many works consider maximizing both spatial reuse and transmission data rate concurrently; in particular, no distributed solutions exist. Apart from that, there is limited work that sought to optimize the placement of passive relays in a manner that enhance decoding gains. Apart from that, no studies exist on video transmissions over direct and indirect paths in WPANs, especially on a scheduler that aims to minimize stalls or delays. In addition, no works have taken advantage of 60 GHz links to create high capacity multi-hop wireless networks whereby nodes with multiple antenna elements, as facilitated by millimeter wavelength, transmit *or* receive concurrently simultaneously.

REFERENCES

[1] X. An and R. Hekmat, "Directional MAC protocol for millimeter wave based wireless personal area networks," in *IEEE Vehicular Technology Conference*, Singapore, May 2008.

[2] X. An, C. Sum, R. Prasad, J. Wang, L. Zhou, J. Wang, R. Hekman, H. Harada, and I. Niemegeers, "Beam switching support to resolve link-blockage problem in 60 GHz WPANs," in *IEEE Intl. Symposium on Personal, Indoor and Mobile Communications (PIMRC)*, Tokyo, Sep. 2009.

[3] X. An, S. Zhang, and R. Hekmat, "Enhanced MAC layer protocol for millimeter wave based WPAN," in *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, Cannes, France, Sep. 2008.

[4] T. Baykas, C. Sum, Z. Lan, J. Wang, M. A. Rahman, H. Harada, and S. Kato, "IEEE 802.15.3c: The first IEEE wireless standard for data rates over 1 Gb/s," *IEEE Communications Magazine*, vol. 49, no. 7, pp. 114–212, 2011.

[5] N. Chahat, G. Valerio, M. Zhadobov, and R. Sauleau, "On-body propagation at 60 GHz," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 4, pp. 1876–1888, 2013.

[6] H. Chao, , and M. Hsu, "CTAP-minimized scheduling algorithm for millimeter-wave-based wireless personal area networks," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 8, pp. 3840–3852, 2011.

[7] R. C. Daniels, J. N. Murdock, T. S. Rappaport, and R. W. H. Jr, "60 GHz wireless: Up close and personal," *IEEE Microwave Magazine*, vol. 11, no. 7, pp. 44–50, 2010.

[8] R. Diestel, *Graph Theory*, 4th ed., ser. Graduate Texts in Mathematics (Book 173). Springer-Verlag, 2010.

[9] ECMA International, "Standard ECMA-387: High rate 60 GHz PHY, MAC and PALs," 2010.

[10] Z. Genc, G. M. Olcer, E. Onur, and I. Niemegeers, "Improving 60 GHz indoor connectivity with relaying," in *IEEE International Conference on Communications (ICC)*, Cape Town, South Africa, Jun. 2011.

[11] M. Gong, D. Akhmetov, R. Wang, and S. Mao, "Directional CSMA/CA protocol with spatial reuse for mmWave wireless networks," in *IEEE GLOBECOM*, Miami, FL, USA, Dec. 2010.

[12] M. Gong, D. Akhmetov, R. Want, and S. Mao, "Multi-user operation in mmwave wireless networks," in *IEEE ICC*, Kyoto, Japan, Jun. 2011.

[13] D. Halperin, S. Kandula, J. Padhye, P. Bahl, and D. Wetherall, "Augmenting data center networks with multi-gigabit wireless links," in *ACM SIGCOMM*, Toronto, Canada, Aug. 2011.

[14] M. Hsu and H. Chao, "Scheduling with reusability improvement for millimeter wave based wireless personal area networks," in *IEEE ICC*, Cape Town, South Africa, May 2010.

[15] C. Hua and R. Zheng, "Robust topology engineering in multi-radio multichannel wireless networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 3, pp. 492–503, 2012.

[16] IBM. ILOG CPLEX Optimization Studio. [Online]. Available: http://www-01.ibm.com/software/websphere/products/optimization/academic-initiative/

[17] IEEE 802.15.3 Working Group, "Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs) Amendment 2: Millimeter-Wave-based Alternative Physical Layer Extension," *IEEE Std 802.15.3c-2009 (Amendment to IEEE Std 802.15.3-2003)*, 2009.

[18] S. Jin, M. Choi, K. Kim, and S. Choi, "Opportunistic spatial reuse in IEEE 802.15.3c wireless personal area networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 2, pp. 824–834, Feb. 2013.

[19] Y. Katayama, K. Takano, Y. Kohda, N. Ohba, and D. Nakano, "Wireless data center networking with steered beam mmwave links," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Cancun, Mexico, Mar. 2011.

[20] J. Kim, Y. Tian, S. Mangold, and A. F. Molisch, "Joint scalable coding and routing for 60 GHz real-time live HD video streaming applications," *IEEE Transactions on Broadcasting*, vol. 59, no. 3, pp. 500–514, Sep. 2013.

[21] S. Kompella, S. Mao, T. Y. Hou, and H. Sherali, "On path selection and rate allocation for video in wireless mesh networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 17, no. 1, pp. 212–224, 2009.

[22] C. S. C. Leong, B. S. Lee, A. R. Nix, and P. Strauch, "A robust 60 GHz wireless network with parallel relaying," in *IEEE International Conference on Communications (ICC)*, Paris, France, Jun. 2004.

[23] S. Leyffer, A. Sartenaer, and E. Wanufelle, "Branch and refine for mixed integer nonconvex global optimization," Argonne National Laboratory, 9700 Soutch Cass Avenue, Argonne, Illinois 60439, USA, Tech. Rep. MCS-P1547-0908, Oct. 2008.

[24] X. Lin, C. Lin, X. Shen, and J. Mark, "REX: A randomized exclusive region based scheduling scheme for mmWave WPANs with directional antenna," *IEEE Transactions on Wireless Communications*, vol. 9, no. 1, pp. 113–121, 2010.

[25] C. Liu, E. Skafidas, and R. Evans, "Angle of arrival extended SV model for the 60 GHz wireless indoor channel," in *Australasian Telecommunication Networks and Applications Conference*, Christchurch, NZ, Dec. 2007.

[26] R. Mudumbai, S. Singh, and U. Madhow, "Medium access control for 60 GHz outdoor mesh networks with highly directional links," in *IEEE INFOCOM*, Rio de Janeiro, Brazil, Apr. 2009.

[27] C. Perkins, *Ad Hoc Networking*, 1st ed. Adison-Wesley, 2001.

[28] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, 2011.

[29] J. Qiao, L. X. Cai, and X. Shen, "Multi-hop concurrent transmission in millimeter wave WPANs with directional antenna," in *IEEE ICC*, Cape Town, South Africa, May 2010.

[30] J. Qiao, L. Cai, and J. Mark, "STDMA-based scheduling algorithm for concurrent transmissions in directional millimeter wave networks," in *IEEE International Conference on Communications (ICC)*, Ottawa, ON, Canada, June 2012.

[31] S. Singh, F. Ziliotto, U. Madhow, E. Belding, and M. W. Rodwell, "Millimeter wave WPAN: Cross-layer modeling and multi-hop architecture," in *IEEE INFOCOM*, Anchorage, AK, USA, May 2007.

[32] P. F. M. Smulders, "Broadband wireless LANs: a feasibility study," Ph.D. dissertation, Eindhoven University of Technology, The Netherlands, 1995.

[33] I. Son, , S. Mao, M. Gong, and Y. Li, "On frame-based scheduling for directional mmWave WPANs," in *IEEE INFOCOM*, Orlando, Florida, USA, Mar. 2012.

[34] C.-S. Sum and H. Harada, "Scalable heuristic STDMA scheduling scheme for practical multi-Gbps millimeter-wave WPAN and WLAN systems," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 2658–2669, May 2012.

[35] C. Sum, L. Zhou, R. Funada, J. Wang, T. Baykas, M. Rahman, and H. Harada, "Virtual time-slot allocation scheme for throughput enhancement in a millimeter-wave multi-Gbps WPAN system," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1379–1389, 2009.

[36] X. Tie, K. Ramachandran, and R. Mahindra, "On 60 GHz wireless link performance in indoor environments," in *Passive and Active Measurement Conference*, Vienna, Austria, March 2012.

[37] C. Wu, H. Zhang, X. Cui, and G. Zhang, "Codebook beam switching based relay scheme for 60 GHz anti-blockage communication," in *4th Intl. Conference on Emerging Intelligent Data and Web Technologies*, Xian, China, Sep. 2013.

[38] C. Yiu and S. Singh, "Link selection for point-to-point 60 GHz networks," in *IEEE International Conference on Communications (ICC)*, Cape Town, South Africa, May 2010.

[39] W. Zhang, X. Zhou, L. Yang, Z. Zhang, B. Zhao, and H. Zheng, "3D beamforming for wireless data centers," in *Proceedings of the 10th ACM Workshop on Hot Topics in Networks*, Cambridge, MA, USA, Nov. 2011.

[40] X. Zhang, S. Zhou, X. Wang, Z. Niu, X. Lin, D. Zhu, and M. Lei, "Improving network throughput in 60 GHz via multi-AP diversity," in *IEEE International Conference on Communications (ICC)*, Ontario, Canada, Jun. 2012.

[41] G. Zheng, C. Hua, R. Zheng, and Q. Wang, "A robust reflector placement framework for 60 GHz mmWave wireless personal area networks," *CoRR*, vol. abs/1207.6509, 2012.

[42] F. Zhong, "Wireless networking with directional antennas for 60 GHz systems," in *14th European Wireless Conference*, Prague, Czech Republic, Jun. 2008.

[43] L. Zhou, C. Sum, J. Wang, T. Baykas, J. Gao, H. Nakase, H. Harada, and S. Kato, "Deflect routing for throughput improvement in multi-hop millimeter-wave WPAN system," in *IEEE Wireless Communications and Networking Conference*, Budapest, Hungary, Apr. 2009.

[44] ——, "Directional relay with spatial time slot scheduling for mmWave WPAN systems," in *IEEE Vehicular Technology Conference*, Taipei, Taiwan, May 2010.

[45] X. Zhou, Z. Zhang, Y. Zhu, Y. Li, S. Kumar, A. Vahdat, B. Zhao, and H. Zheng, "Mirror mirror on the ceiling: Flexible wireless links for data centers," *ACM SIGCOMM Computer Communication Review*, vol. 42, no. 4, pp. 443–454, 2012.

[46] W. Zou, Y. Hu, B. Li, Z. Zhou, and Y. Ye, "An improved ER scheduling algorithm based spatial reuse scheme for mmWave WPANs," in *6th International ICST Conference on Communications and Networking (CHINACOM)*, Harbin, Hei Long Jiang, China, August 2011.