A-Duplex: Medium Access Control for Efficient Coexistence Between Full-Duplex and Half-Duplex Communications

Aimin Tang and Xudong Wang, *Senior Member, IEEE*

***Abstract*—As full-duplex wireless communication evolves into a practical technique, it will be built into communication nodes in many application scenarios. However, it is difficult to do so for legacy communication nodes. Thus, full-duplex communication nodes will coexist with half-duplex communication nodes in the same application environment. In this paper, a wireless local area network with a full-duplex access point (AP) and half-duplex clients is studied, and a media access control (MAC) protocol called asymmetrical duplex (A-Duplex) is developed to support efficient coexistence between half-duplex clients and the full-duplex AP. A-Duplex explores packet-alignment-based capture effect to establish dual links between the AP and two different clients. In this way, the capability of a full-duplex AP can be utilized by half-duplex clients, which leads to much improved network throughput. Moreover, to ensure fairness of the MAC protocol, a virtual deficit round-robin algorithm is proposed for the AP to select appropriate half-duplex clients for dual-link setup. A-Duplex does not require any change in the physical layer of half-duplex clients; only an update of MAC driver is necessary. Thus, it is well suited for coexistence between half-duplex clients and a full-duplex AP. Both analysis and simulations are conducted to evaluate performance of A-Duplex. Results show that it improves the throughput by 48% and 188% and reduces the average packet delay by 26% and 22%, as compared to the IEEE 802.11 Distributed Coordination Function with and without RTS/CTS, respectively. Moreover, the throughput remains steady as the number of clients grows. A-Duplex also maintains a high level of fairness.**

***Index Terms*—Medium access control, full-duplex communication, wireless LAN, coexistence between full-duplex and half-duplex communications.**

I. Introduction

FULL duplex wireless communication is evolving into a more practical technique [1]–[5]. In [1], three antennas are used to achieve about 30 dB antenna cancellation. Combined with path loss attenuation, analog cancellation and digital cancellation, this design can achieve about 100 dB self-interference cancellation within 5 MHz bandwidth for IEEE 802.15.4. In [2], Balun cancellation is leveraged for an improved design of self-interference cancellation, in which only two antennas are used. With the help of path loss attenuation and digital cancellation, about 110 dB cancellation can be achieved within 20 MHz. In [3], an additional transmitting chain is added to generate an invert analog signal to cancel the self-interference. So far the best design is given in [4], where only one antenna is used to achieve almost perfect self-interference cancellation for full duplex WiFi radio. In this design, a passive circuit is first used to achieve 60 dB self-interference cancellation and then both linear and nonlinear digital cancellation are used to cancel the remaining self-interference. In [5], a balanced cancellation circuit is designed for RF front-end. This design can achieve 60 dB cancellation at the RF front-end. On the other hand, it can be integrated with the design in [4] to further improve the self-interference cancellation. Full duplex can nearly double the capacity of a point-to-point communication link and thus significantly improves the spectrum efficiency. However, to fully leverage the capability of full duplex communications in a network, an efficient media access control (MAC) protocol plays a critical role.

So far a few MAC protocols have been proposed to support full duplex communications [6]–[9], but they assume that all nodes in the network are equipped with full duplex radios. However, in many application scenarios full duplex and half duplex radios have to coexist in the same network for two main reasons. First, despite fast progress in the development of full duplex radios, challenges still remain when applying full duplex radios to nodes such as smart phones or laptops. For example, in [1]–[3], [5], all radios need more than one antennas to achieve full duplex communications, and the distance between antennas are more than 20 cm. In [4], only one antenna is needed, but implementation of the circuit is rather complicated. These factors make it difficult for a smart phone or personal digital assistant (PDA) to adopt a full duplex radio. Second, many legacy devices only support half duplex communications. It is fine to replace a legacy AP by a full duplex AP, but replacing all legacy devices by full duplex ones is neither economical nor acceptable. Thus, coexistence between full duplex and half duplex communications becomes indispensable. For example, supporting a wireless LAN with a full duplex AP and half duplex clients is extremely meaningful for the next generation WiFi networks.

In this paper, we study a full duplex wireless LAN where the AP supports full duplex communications but all clients are half duplex nodes. The full duplex AP is assumed to have perfect self-interference cancellation [4]. Since all clients have

Manuscript received October 21, 2014; revised February 25, 2015 and June 6, 2015; accepted June 7, 2015. Date of publication June 12, 2015; date of current version October 8, 2015. The associate editor coordinating the review of this paper and approving it for publication was C.-F. Chiasserini. *(Corresponding author: Xudong Wang.)*

The authors are with the UM–SJTU Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: wxudong@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2015.2443792

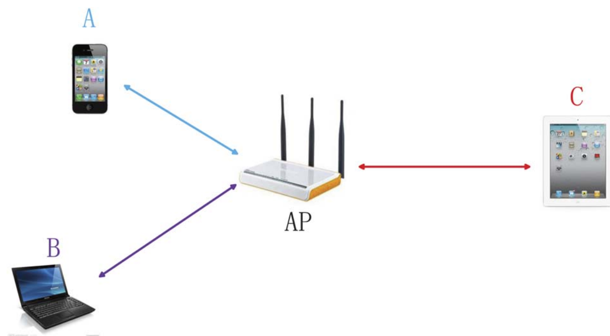


Fig. 1. A wireless LAN with a full duplex AP.

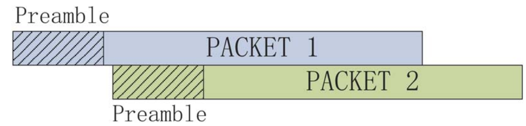


Fig. 2. Capture effect without collision in preamble time.

no capability of full duplex communications, only asymmetric dual links can be utilized to boost network throughput. For example, as shown in Fig. 1, while Client A sends a packet to the AP, the AP sends a packet to another client (e.g., Client C). In [6], asymmetric dual links (i.e., A → AP → C) do not allow collisions at both receivers, i.e., the signal from Client A does not interfere Client C. However, such a condition cannot be easily satisfied in a wireless LAN, so the capability of the full duplex AP cannot fully utilized. In order to fully leverage the capability of the full duplex AP, we develop a novel MAC protocol to explore *capture effect* at the client side to improve opportunity of full duplex communications in a wireless LAN. Capture effect [10]–[12] means, when a receiver receives two colliding packets, the stronger one can still be decoded correctly. For example, Client A and Client B may not be hidden to each other, but a packet from Client A to the AP may not corrupt the packet from the AP to Client B, due to capture effect at Client B. In this situation, capture effect is utilized to form full duplex communications at the AP, even though Clients A and B are not hidden from each other.

The performance of capture effect can be significantly improved through alignment of two colliding packets. To leverage *packet-alignment* for better capture effect,[1] there are two options: 1) the desired packet is sent first; 2) the interfering packet is sent first. In either type, if the preamble of the first packet does not collide with the second packet as shown in Fig. 2, it is relatively easy for the first packet to be decoded correctly. However, it is rather difficult to decode the second packet (This case is also called message in message (MIM) in [13]). For example, under the basic rate in IEEE 802.11a, the first packet can be decoded when its signal strength is 0 dB stronger than that of the second packet [12]. If we want to decode the second packet correctly, the signal strength of the second packet need to be 11 dB stronger than that of the first packet [12].

However, the asymmetric dual links supported via capture effect may not always improve the network performance. When asymmetric dual links are established (e.g., A → AP → C), the transmission rate from the AP to Client C is lower than that in the half duplex case since Client C is interfered by Client A. Thus, the transmission time of the packet from the AP to Client C will be longer than that in the half duplex case. The concurrent transmission in dual links can save some time; however, if the transmission time from the AP to Client C is much longer than that of the packet from Client A to the AP, the dual links may reduce the throughput. In other words, the capture effect can help establish dual links but it cannot ensure throughput improvement. As a result, to better utilize the capture effect in a wireless LAN with a full duplex AP and half duplex clients, the MAC protocol must take into account three requirements: 1) the packet from the AP needs to be transmitted first; 2) the received signal strength of the packet from the AP to a client is stronger than the interference from the packet sent from another client to the AP; 3) the AP needs to choose a client with a proper transmission rate to establish dual links so that throughput can be improved. To this end, the AP needs to collect relative signal strength between nodes to build an information map from which clients can be properly selected to establish dual links effectively. In this paper, a dynamic information update scheme is designed to establish such a map.

Thanks to the asymmetric dual links, our MAC protocol supports efficient co-existence between a full duplex AP and half duplex clients, so it is called asymmetrical-Duplex (A-Duplex) in this paper. A-Duplex not only improves the throughput but also reduces packet delay. Simulation results show that the throughput is improved by 48% and 188% as compared to IEEE 802.11 DCF [14] with RTS/CTS and without RTS/CTS, respectively. The packet delay is reduced by 26% and 22%, respectively.

A-Duplex is distinct with several features. First, it leverages packet-alignment based capture effect to establish asymmetric dual links so as to fully utilize the capability of the full duplex AP. As a result, throughput is improved effectively as compared to the case of only using the hidden nodes to establish dual links. Second, a map of signal-to-interference ratio (SIR) is built at the AP; based on this map asymmetric dual links are established to take advantage of capture effect. The SIR map can be updated dynamically to capture the variable network environment. Third, a virtual deficit round robin algorithm is developed for the AP to select a downlink in dual link setup, which improves fairness of the MAC protocol. Finally, no physical layer change is needed at legacy nodes; only MAC driver update is needed. To the best of our knowledge, this is the first MAC that holds these features.

The rest of the paper is organized as follows. Related work is introduced in Section II. The protocol design of A-Duplex is described in Section III. Detailed procedures of establishing dual links in A-Duplex are elaborated in Section IV. The fairness issue is studied in Section V. Simulation results are presented in Section VII. Compatibility and practicality issues are discussed in Section VIII. The paper is concluded in Section IX.

[1] If successive interference cancellation (SIC) is taken into account, then packet alignment is not necessary. However, the client side consists of legacy nodes, so any change to the physical layer is infeasible. Thus, SIC is not considered in this paper.

II. Related Work

A. Capture effect

In random access, if more than one frames are transmitted concurrently, collision occurs. However, it does not always destroy all the frames. If the receiving power of one of the colliding frames is larger enough than that of the other, then the receiver can decode this frame correctly. Such a process of receiving a frame from colliding frames is called capture effect. The physical layer model for capture effect has been explored by various experiments in [10]–[12], [15]. Usually if the stronger frame comes earlier than other frames, it is easier to decode this frame since the receiver has already been synchronized with the first frame. However, if a weaker frame comes first, MIM is needed to decode the stronger frame. In this case, since the receiver has been synchronized with the weaker frame, the signal strength of the stronger frame must be much larger than the weak one so that the receiver can re-synchronize with the stronger one to utilize capture effect [12], [13], [16]. Since the receiver takes a certain amount of time to achieve synchronization, even if the stronger frame arrives earlier, it is necessary to make sure the lead time is sufficient. In [12], the experimental results show that even two frames come at the same time, capture effect may still succeed in 802.11a networks. However, if the first frame is earlier by a preamble time, the success probability of capture effect can be improved, because a clean preamble is helpful to conduct synchronization and channel estimation.

So far the work in [12] provides the best model for capture effect in 802.11a networks. It provides the measurement-based SIR capture thresholds for all the 802.11a bit rates.

B. Full Duplex MAC

So far there exist a few MAC protocols that support full duplex communications [6]–[8]. In [6], both symmetric and asymmetric dual links are considered. A set of symmetric dual links are formed by two nodes transmitting signal to each other simultaneously. An example is the link (A → AP → A) shown in Fig. 1. A set of asymmetric dual links are established by three nodes in a two-hop setup, such as the links (B → AP → C) shown in Fig. 1. In [6], the MAC protocol mainly explores symmetric dual links to improve network throughput, and in the asymmetric case (e.g., links (B → AP → C) in Fig. 1), dual links can be established only under the following conditions: a) the uplink transmission (B → AP) does not collide with the downlink transmission (AP → C) at the Client C; b) the AP transmits to Client C later.

In [7], a reservation based MAC protocol was developed for a network where all nodes have full duplex radios. All nodes are allocated with some time slots by the AP. In this MAC protocol, asymmetric dual links are allowed even for non-hidden nodes. For example, dual links A-AP-C in Fig. 1 are allowed even if they are not hidden nodes, since the interference by Client A is small at Client C. In this scheme, AP first collects the interference information of clients and then schedules the transmission order according to the interference relationship. However, packet-alignment based capture effect are not considered

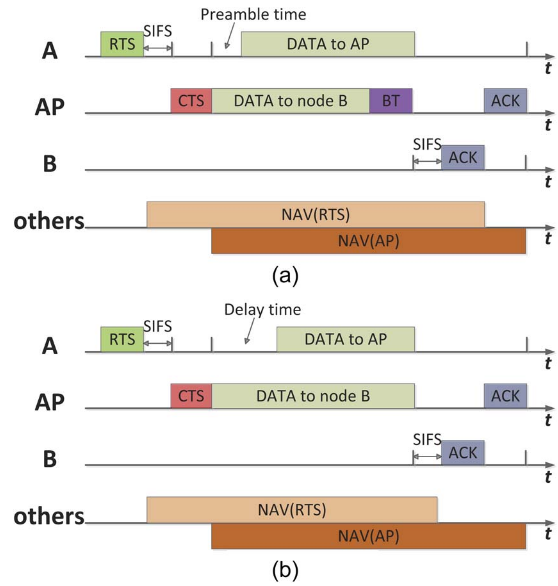


Fig. 3. Dual link establishment. (a) The "AP-shorter" scenario: the transmission time of the packet from the AP to B is shorter than the packet transmission time of the other link plus a preamble time. (b) The "AP-longer" scenario: the transmission time of the packet from the AP to B is longer than the packet transmission time of the other link plus a preamble time.

in this scheme. Thus, the performance of capture effect is limited. For example, in the basic operation mode (i.e., lowest rate with robust modulation/coding), the SIR needs to be 10 dB, which is consistent with [12]. If packet alignment-based capture effect is leveraged, then 1 dB SIR is sufficient [12]. For other modes (i.e., higher rate), the situation is similar.

In [8], the MAC protocol addresses the exposed terminal problem in a full duplex network. It adopts the pseudo-random noise (PN) sequence-based signatures to identify the concurrent links on-the-fly and explore the chance of exposed transmissions. As a result, the throughput of a full duplex network can be improved. The protocol also assumes all nodes have full duplex capability.

III. MAC Protocol Design

In this section, a MAC protocol called A-Duplex is designed for a wireless LAN where the AP has full duplex capability, but all clients can only work in the half duplex mode. To utilize full duplex capability in this wireless LAN, asymmetrical dual links need to be established. As explained before, to explore capture effect in an efficient way, the packet from the AP to a client needs to start first. However, since a client is in the half duplex mode, once it detects such a transmission, it considers the channel busy and cannot start a second link with the AP. To resolve this issue, an RTS/CTS mechanism is added into the MAC protocol to start dual links where the AP-to-client link can be started first. As shown in Fig. 3, the procedure to set up a full duplex link always starts with an RTS frame by a client.

Three cases are considered below. When a client gets the channel via the RTS frame, two cases need to be considered: 1) the AP can send a packet to another client to establish dual links; 2) the AP does not have a packet for another client. The third case is that the AP acquires the channel first. Note that

RTS is only initiated by a client, so the AP does not generate any RTS frame but responds to a client's RTS with a CTS frame.

The *first case* is shown in Fig. 3. Client A first transmits an RTS frame to the AP. When the AP finds that it has a suitable packet for Client B, it replies a CTS frame to Client A and then transmits the packet to client B immediately. Note that in each frame there is a field in the MAC header called duration, which is used for other nodes to update the network allocation vector (NAV) value. When Client A receives the CTS frame, it finds out that the duration in the CTS frame is longer than its duration. Thus, Client A knows that the AP intends to establish dual links. It then transmits the packet to the AP after a certain delay, which is at least longer than a preamble time. The delay time can be computed by Eq. (2). There are two scenarios in this case. If the transmission time of the packet from the AP to Client B is shorter than the packet transmission time of the other link plus a preamble time, as denoted by the "AP-shorter" scenario in Fig. 3(a), then the AP transmits a busytone signal to ensure two transmissions finish at the same time. Otherwise, as denoted by the "AP-longer" scenario in Fig. 3(b), Client A delays its transmission for enough time such that two transmissions finish simultaneously. When Client A and the AP finish their transmissions, Client B first returns an ACK frame to the AP and then the AP returns an ACK frame to Client A upon receiving the ACK frame from Client B. To ensure correct operation of A-Duplex, the duration in RTS and CTS needs to be computed properly. Since in the beginning Client A does not know whether or not the AP will establish dual links, it computes the duration as if the AP does not set up dual links. In case that the AP does not set up dual links, the duration field in the RTS frame works exactly as that in the half duplex case. If the AP does set up dual links, it just uses the duration in the RTS frame to compute the duration of the CTS frame as explained below. Since the AP's transmission covers all clients, the duration field in the CTS frame can gracefully protect the transmissions between Client A and the AP. As a result, the duration in RTS frame is computed as

$$Duration_{RTS} = 3T_{SIFS} + T_{CTS} + T_1 + T_{ACK}, \quad (1)$$

where T_{SIFS} is the Short Inter-Frame Spacing (SIFS) time, T_{CTS} is the time of CTS frame, T_{ACK} is the time of ACK frame, and T_1 is the time of data packet to the AP. With the duration from RTS, the AP computes the duration in CTS as follows. In the case of Fig. 3(a), the duration in CTS is computed as

$$Duration_{CTS} = Duration_{RTS} - T_{CTS} - 2T_{SIFS} + T_p + T_{ACK},$$

where T_p is the preamble time. In the case of Fig. 3(b), the duration in CTS is computed as

$$Duration_{CTS} = T_2 + T_{SIFS} + 2T_{ACK},$$

where T_2 is the transmission time of the packet from the AP to Client B. When A receive the CTS frame, it decides the delay time according to

$$Delay = Duration_{CTS} - T_1 - T_{SIFS} - 2T_{ACK}. \quad (2)$$

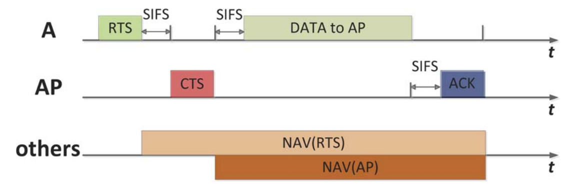


Fig. 4. An example showing no dual links.

The delay for Client A serves two purposes. First, for a legacy node, when it finishes a data packet transmission, it waits a fix time for ACK. If no ACK is received after the preset timeout, it starts the retransmission procedure. Second, the delay ensures the packet to Client B is sent first and its preamble can be protected without interference.

In the above procedure, Client A in Fig. 3 waits a time period of $T_{SIFS} + T_{ACK}$ before it receives its ACK. This time period is longer than Distributed Coordination Function Inter Frame Spacing (DIFS) time. Thus, the ACK may collide with the packet from the other client. In our design, the CTS frame can solve this problem. The duration in CTS covers the entire period of the above procedure, so all other clients receiving the CTS frame update the NAV value according to the duration in the CTS frame. As a result, the ACK from the AP can also be protected.

In the *second case* when the AP has no packet for a client, the operation procedure is shown in Fig. 4. When AP finds no suitable packet for all other clients, it only replies a CTS frame. From the CTS frame, Client A finds out that the duration in CTS frame is the same with its duration. Thus, Client A knows that the AP will not establish dual links, so it only delays an SIFS time and then starts a packet transmission. When the packet is received correctly, the AP returns an ACK to Client A.

In the *third case*, the AP may get the channel first by sending a data packet without RTS/CTS exchange. No dual link can be set up in this case for two reasons. First, a client does not support full duplex communications, so symmetrical dual links cannot be formed. Second, clients cannot decide among themselves about which one can start asymmetrical links to explore capture effect. If the decision is done by the AP, then the queue status as well as the length of the first packet in queue of each client must be reported to the AP. To avoid such complexity, dual link setup is ignored in the third case.

In A-Duplex, the RTS/CTS frame exchange procedures fulfill the following functions. Firstly, from the RTS frame, the AP knows exactly which client is about to send a packet and then decides how to establish dual links with another client. In [6], the AP transmits packet to another client to establish dual links when it decodes the header of the received packet. However, if the packet is transmitted by OFDM symbols such as in 802.11a network, the AP cannot decode the packet header to get the transmitting client address until it receives the whole packet. Secondly, the RTS/CTS mechanism guarantees that the packet from the AP to Client B is transmitted first and the preamble of the packet is not interfered by the packet from Client A to the AP. Thus, Client B can decode the packet more easily through capture effect. Thirdly, the RTS frame contains signal strength information that can be used by the AP to select a proper client to set up a dual link. Details of this function are explained in the next section. Fourthly, ACK frames in dual links are protected via the duration in RTS/CTS frames (i.e., the NAV mechanism).

A-Duplex may experience two types of collisions. In the first type of collisions, more than one clients send an RTS to the AP. Without correct reception of an RTS, AP does not reply a CTS frame. As a result, the competing clients know that collision happens and do not start data packet transmission. In the second type of collisions, the AP starts a transmission simultaneously with an RTS frame from a client. Since the AP can simultaneously transmit and receive, it can find out collisions while it is transmitting. In this case, the AP stops transmitting immediately and let the client's RTS frame proceed, so the client can still contend the channel successfully.

IV. Establishment of Asymmetrical Dual Links

The critical step of setting up dual links is to ensure capture effect to be properly utilized. For example, if dual links (from A to AP and from AP to C) are to be set up, the AP needs to make sure that the signal strength at Client C is stronger enough than interference from Client A according to the transmission rate. In other words, the AP needs to find out the signal-to-interference ratio (SIR) for Client C and decides the transmission rate based on the SIR value. To this end, it is necessary for the AP to build an information map of SIR for all clients. In what follows, such an information map is called the SIR MAP.

A. Establishment of SIR MAP

The SIR MAP contains the SIR information of all clients. Considering Clients A and B in a wireless LAN, SIR at Client A given the interference from B can be represented by SIR of B to A. Thus, if a wireless LAN has 3 clients: clients A, B, and C, then SIR map table must have 6 items such as SIR of B to A, C to A, A to B, C to B, A to C, and B to C, because there are 6 different scenarios of interference between two clients.

To measure the SIR value experienced by a client, two values are needed: the signal strength from the AP and that from the interfering client. However, an AP cannot get such information by itself, so it has to rely on clients to collect such information. When the AP transmits a packet to a client (e.g., Client B), the client can get the signal strength of this on-going transmission. When another client (e.g., Client A) transmits a packet to the AP, the current client (i.e., Client B) can also get the interference strength by overhearing the packet. Since we make no change to the physical layer of legacy nodes, the signal strength that can be collected in the MAC layer is the Received Signal Strength Indicator (RSSI) when a client succeeds in receiving a packet in the MAC layer.

MAC layer messages are used to support SIR measurements at clients and help build the SIR MAP at the AP. As explained in the previous section, when a client (e.g., Client A) accesses the channel, it first transmits an RTS frame to the AP. All other clients who receive this RTS frame record the interference strength. After receiving the RTS frame, the AP replies a CTS frame and records this client address. All other clients update the signal strength from AP according to the CTS frame, and then they compute the SIR value with the two recorded values: the signal strength from the AP and interference strength. In the next round of RTS/CTS/Data/ACK, if one client (e.g., Client B) contends the channel successfully, it transmits its recorded SIR value via the RTS frame to the AP. Since the AP records which client (i.e., Client A) started the last round of RTS/CTS/Data/ACK, it knows that the SIR value carried in the RTS frame is the SIR of a new client (i.e., Client B) given the interference from Client A. Thus, it updates such an SIR value (of A to B) in its SIR MAP. If the same client contends the channel successfully, the SIR value is ignored. As explained in the protocol design, the AP may contend the channel successfully. In this scenario, all clients just update the signal strength from the AP and recompute the SIR value. If the AP continuously gets the channel, all clients update the SIR value until a client contends the channel successfully. Since then, the above procedure is followed to record or update the SIR value in the AP's SIR MAP. By building up SIR MAP, the AP can have a dynamic view of SIR for each client.

In the MAC layer, RSSI information is only available when a MAC packet is received. Given a client, in order to detect signal strength from an interfering client, it needs to get the RTS frame from the interfering client. To achieve this goal, we use the basic rate to transmit the RTS frame so that other clients can rely on this frame to detect strength of an interfering signal. This approach is effective for two reasons. First, in a wireless LAN the transmit power of an RTS frame is usually set the same as that of a regular data packet. Second, an RTS frame uses the basic rate for transmission, so it can mostly cover the large area of a wireless LAN. In case a client can sense a busy channel but cannot receive an RTS frame, it relies on a CTS frame to get RSSI. More specifically, when the client receives a CTS frame, it can conclude that there was an RTS frame before. Thus, the client can use the minimum required signal strength for decoding an RTS frame to approximate the interference signal strength.

As for the collision case, all clients can find out whether a collision happens via RTS/CTS exchange. More specifically, if a client does not receive a CTS frame from the AP after receiving an RTS frame from a client, it knows collision has happened and thus skips updating its interference signal strength.

Due to channel fading, an instantaneous SIR value cannot always accurately reflect the channel quality very well. Thus, when the AP updates the MAP table, it cannot simply substitute the value. Instead, a moving average strategy is necessary. One easy approach is to use the average of the most recent values. However, this demands the AP to store too many SIR values. For example, if we want to get the average of 5 SIR values, AP needs to stores 5 SIR values for each client. In order to reduce the storage of AP, a simple moving average scheme is given as follows:

$$SIR_{New} = SIR_{Old} \times (1 - \theta) + SIR_{update} \times \theta, \tag{3}$$

where θ is a weight factor from 0 to 1, the value SIR_{Old} is from the current SIR MAP, SIR_{Update} is newly learned from frame exchange, and SIR_{New} is the new SIR in the MAP. With this scheme, the AP only needs to store one SIR value for each

Octets	2	2	6	6	1	4
	Frame Control	Duration	RA	TA	SIR	FCS

Fig. 5. RTS frame structure.

client. However, if θ is too small, the SIR cannot reflect the change of the channel in time. Thus, we need to find a suitable θ for a specific environment, as discussed in performance evaluation.

From the establishment procedure of SIR MAP, we know that the SIR stored in the MAP is not the real time SIR. Thus, when the AP uses this SIR information to establish dual links, the capture link may fail. From the simulation results in Section VII-B2, we know that the failure rate is less than 20%, which reveals that the SIR is not perfect but is effective to capture dual-link opportunities. When capture effect fails in dual links, the receiving client cannot successfully receive a packet from the AP. Thus, the AP will retransmit the packet later, which is acceptable. However, if we try to achieve 100% success rate for the capture link, the AP needs to collect the real time SIR of all clients for all dual links, which will lead to very high overhead. In contrast, the overhead from setting up the SIR MAP in A-Duplex is small.

B. RTS Frame Structure and SIR Values

Since the SIR value is sent in an RTS frame, we make a slight change to the RTS frame in the IEEE 802.11 standard. As shown in Fig. 5, we add one byte in the RTS frame. The first bit is used to indicate the sign of the SIR, i.e., 0 represents positive and 1 represents negative. The last 7 bits give an absolute SIR value. Since 64 dB SIR is enough for a practical system and considering a granularity of 0.5 dB, all SIR levels can be represented by 7 bits (i.e., from level 0 till level 127).

C. Setup of Asymmetric Dual Links

When the AP receives an RTS frame from a client, it needs to select a client for downlink transmission so that dual links are set up. The basic condition to set up such a downlink is that the capture effect at the selected client exists, i.e., the SIR at the selected client must exceeds a capture threshold. However, this basic condition is insufficient to ensure throughput improvement. For example, when Client A uses the highest rate to send packet 1 to the AP, the AP may establish a downlink and send packet 2 to Client B with the lowest rate. If the payloads of the two packets are the same, then the time needed by packet 2 is much longer than packet 1. As a result, the dual links will decrease the throughput. Thus, when AP sets up dual links, it needs to make sure dual links can improve the throughput. One solution is that the AP chooses the client such that the overall transmission rate of dual links is higher than the case of a single link in half duplex communications [17]. This problem can be viewed in another way. The packet sent from the AP to Client B under dual-link setup leads to the cost of additional transmission time. If the additional time is less than the time needed for sending the packet in the half duplex link from the AP to Client B, then the throughput can be improved. When the AP transmits a packet to a client, the transmission time T_{AP} is given by

$$T_{AP} = T_{data} + T_{SIFS} + T_{ACK} + T_{DIFS}, \tag{4}$$

where T_{DIFS} is the DIFS time and T_{data} is the data frame time. When a client gets the channel and AP does not establish dual links as shown in Fig. 4, the transmission time T_A is given by

$$T_A = T_{RTS} + 3T_{SIFS} + T_{CTS} + T_1 + T_{ACK} + T_{DIFS}, \tag{5}$$

where T_1 is the time of data packet to the AP. However, when the AP establishes dual links in the "AP-shorter" scenario in Fig. 3(a), the transmission time T_{A1} is given by

$$T_{A1} = T_{RTS} + 2T_{SIFS} + T_{CTS} + T_p + T_1 + 2T_{ACK} + T_{DIFS}, \tag{6}$$

where T_p is the preamble time. If the AP establishes dual links in the "AP-longer" scenario in Fig. 3(b), the transmission time T_{A2} is given by

$$T_{A2} = T_{RTS} + 2T_{SIFS} + T_{CTS} + T_2 + 2T_{ACK} + T_{DIFS}, \tag{7}$$

where T_2 is the transmission time of the packet from the AP to Client B. Note that, in both "AP-shoter" and "AP-longer", the transmission time of the packet from AP to Client B (i.e., T_2) is determined by the corresponding capture rate, which is further determined by the SINR. Thus, when the AP establishes the downlink for dual-link setup, it is necessary that the additional time T_{add} (i.e., $T_{add} = T_{A1} - T_A$ or $T_{A2} - T_A$) is less than T_{AP}.

Considering the case of "AP-shoter" in Fig. 3(a), from Eqs. (5) and (6) we get

$$T_{add} = T_{A1} - T_A = T_p - T_{SIFS} + T_{ACK}.$$

Thus, we know that $T_{add} < T_{AP}$ can always be satisfied in this case. In the case of "AP-longer" in Fig. 3(b), from Eqs. (5) and (7) we get

$$T_{add} = T_{A2} - T_A = T_2 - T_1 - T_{SIFS} + T_{ACK}.$$

Thus, $T_{add} < T_{AP}$ is not always guaranteed. To ensure throughput improvement, the AP needs to make sure $T_{add} < T_{AP}$ is satisfied when selecting a downlink for dual-link setup. The AP can set a condition $T_{add} \leq T_{AP}/\beta$ to improve the throughput, where $\beta \geq 1$. Note that β needs to be determined properly; a larger β potentially increases throughput, but the condition of dual-link setup cannot be easily satisfied. To satisfy both the basic condition and the condition of $T_{add} \leq T_{AP}/\beta$, our protocol takes the following procedure to select a client for downlink transmissions: 1) determine a set of clients that satisfy the capture threshold; 2) from this set of clients, select the client that meets the requirement of $T_{add} \leq T_{AP}/\beta$.

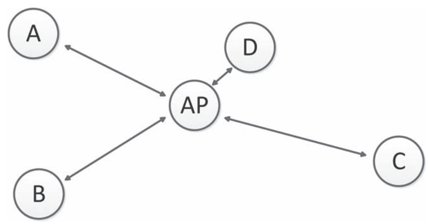


Fig. 6. An example of the fairness issue.

V. Fairness: The Virtual Deficit Round Robin Algorithm

When selecting a client for the downlink of dual-link setup, the AP prefers clients that have more hidden nodes or are closer to the AP, which favors throughput improvement. However, a fairness issue arises. An example is shown in Fig. 6, where there are 4 clients in the network. When Clients A or B gets the channel (for uplink transmission), the AP can establish dual links with Client C or D, which is called a dual-link chance. When Client C gets the channel, the AP can establish the downlink with any of three other nodes. However, when Client D gets the channel, only half duplex link is allowed, i.e., the AP cannot select any other client to set up the downlink for dual-link setup. As a result, Client D has more chances of receiving packets from the AP, which leads to unfairness in downlink.

On the other hand, for each chance the throughput improvement is also different. For example, when Clients A or B gets the channel, the AP can select either Client C or D for dual links. However, since Client D is very close to the AP and has higher SIR than Client C, more throughput improvement can be achieved if the AP choose Client D for dual links. The same situation can occur when Client C gets the channel. As a result, if the AP wants to achieve higher throughput improvement, Client D is the best choice to get the chance each time, which makes the fairness worse. Thus, there exists a tradeoff between fairness and throughput.

Fairness in channel access can be defined in different ways. In this paper, it is evaluated on the channel access time [7], i.e., the channel access time of a downlink is expected to be equal for all clients.

In order to improve fairness, two mechanisms are considered. First, when the AP accesses a channel for data transmission, its internal queueing system gives a higher priority to clients that have a lower chance of being selected for dual-link setup, e.g., Clients A and B in Fig. 6. Second, when the AP establishes a downlink during dual-link setup, it also gives a higher priority to clients that have a lower chance of being selected for dual-link setup. To fulfill the above two mechanisms, the AP needs to look into the dual-link setup chances for each client and then balance the utilization of these chances. To this end, a virtual deficit round robin algorithm is developed for the AP to select an appropriate client for downlink transmission.

For a network with variable packet transmission time, the deficit round robin algorithm [18] can achieve fairness with low complexity. More specifically, during round robin, whether or not a client can get a chance is determined by a time parameter called *deficit*. The value of deficit for Client k at round n, denoted by $T_k(n)$, is updated as

$$T_k(n) = \text{quantum} + T_k(n-1)\, I(n-1), \tag{8}$$

where $I(n-1)$ is 1 if the deficit is not enough to transmit the packet in the previous round or 0 if the queue becomes empty, and quantum is a constant number. If the deficit $T_k(n)$ is insufficient to send a packet from the AP to Client k, then the AP turns to Client $k+1$. If the deficit $T_k(n)$ is enough to transmit a packet, then Client k can be selected in round n. After a packet is sent, the transmission time is deducted from $T_k(n)$. This process is repeated until the deficit $T_k(n)$ is insufficient for sending another packet.

However, due to capture effect, the above algorithm is not applicable to client selection for downlink transmission. For example, when the AP finishes its transmission to Client A, it gives the downlink chance to Client B according to round robin. However, if Client A gets the channel successfully, then the AP can only give the downlink chance to Client C or D for dual-link setup, as selecting Client B does not satisfy the condition of capture effect. Thus, the deficit round robin scheme is not working in this case. To this end, a virtual deficit round robin algorithm is developed. It works with the following procedures:

a) When the AP contends the channel successfully, it follows the deficit round robin scheme to send a packet. After each transmission, it needs to find and record the *next client* according to the deficit. Suppose the Client k is the recorded *next client* in deficit round robin. After this packet transmission, if the remaining deficit is enough to send another packet, Client k is still recorded as the *next client*. Otherwise, the deficit round robin goes to Client $k+1$. If the deficit of Client $k+1$ is enough for a packet, Client $k+1$ is recorded as the *next client*. If the deficit is insufficient, Client $k+2$ is considered. This process is repeated until a client with enough deficit is found and is recorded as the *next client*.

b) If a client wins the channel access, the AP needs to select a client for downlink transmission by considering capture effect. In this case, it does not follow the order of deficit round robin, but it still relies on the deficit to select a client. More specifically, the AP selects the client that has the maximum deficit among all clients that satisfy the capture condition. After the downlink transmission, the deficit is deducted by the time that is needed when half-duplex communication is conducted. The reason for deducting such a time is to make sure the deficit is updated consistently with case a). As a result, winning a downlink transmission in full duplex communications by a client will give more opportunities to other clients including those that cannot participate in asymmetrical dual links. It is possible that the client selected by the AP is the one recorded as the *next client* in case a). In this situation, the AP needs to find a new *next client* in the similar way as done in case a).

It should be noted that the AP in case b) does not check if the deficit is enough to transmit the packet, so the deficit can become negative after it is deducted by the downlink transmission time. Even after it is updated according to Eq. (8), it may still be negative. As a result, even if the procedure in case a) is similar to the deficit round robin algorithm, the deficit can be negative. When the AP in case a) finds a client has negative deficit, it just skips this client to the next one. Actually we can set a limit to the negative value of the deficit. In other words, if the deficit of a client is not enough for a packet transmission (by considering the negative limit), the client cannot be selected in case b). This limit improves fairness, but results in loss of dual link chances. Thus, fine-tuning such a limit provides a tradeoff between fairness and throughput.

We call our scheme a virtual deficit round robin scheme for several reasons: 1) Although the deficit is updated in the same way as that in the deficit round robin algorithm, the deficit in case b) is not used in a round robin way; 2) The deficit in cases a) and b) can be negative; 3) Since the round robin order is not strictly followed (due to exploring dual links via capture effect), perfect fairness like deficit round robin cannot always be ensured.

VI. Performance Analysis

In this section we analyze saturation throughput of A-Duplex to study its throughput improvement over 802.11 DCF. "Saturation" means the AP and clients always have packets for transmission. In [19], the performance of 802.11 DCF is analyzed via a Markov model. The same analysis can be adopted here, but two major differences in A-Duplex need to be considered. First, when a client and the AP collide, the client can still get the channel successfully due to full duplex capability at the AP. Second, capture effect exists in A-Duplex.

In a practical system, the Markov chains of the AP and clients can be different since the contention windows for them are different. We set the minimum contention window W_0 and maximum $2^{m_0}W_0$ for the AP and the minimum contention window W and maximum $2^m W$ for clients. Following similar Markov chain analysis in [19], we can get the transmission probability P_t for a client in a slot as

$$P_t = \frac{2}{1 + W + pW\sum_{i=0}^{m-1}(2p)^i}, \tag{9}$$

where p is the conditional collision probability for a client. In A-Duplex, a successful transmission from a client occurs when one of the clients gets the channel no matter whether the AP gets the channel or not. Thus, p is given by

$$p = 1 - (1 - P_t)^{N-1}, \tag{10}$$

where N is the number of clients.

In the same way, we can get the transmission probability P_{t0} for the AP in a slot as

$$P_{t0} = \frac{2}{1 + W_0 + p_0 W_0\sum_{i=0}^{m_0-1}(2p_0)^i}, \tag{11}$$

where p_0 is the conditional collision probability for the AP and is given by

$$p_0 = 1 - (1 - P_t)^N, \tag{12}$$

as the AP transmits successfully when all clients keep silent. Following [19], the throughput S is defined as the fraction of time for successfully sending payload bits. It can be calculated as follows

$$S = \frac{(P_A + P_c + P_c P_{ca})E\{L_p\}}{(1 - P_{tr})T_{slot} + P_A T_{s1} + P_c T_{s2} + P_c P_{ca} T_{add} + P_{col} T_c}, \tag{13}$$

where the parameters are defined as follows. P_{tr} is the probability of transmission from at least one node. P_A and P_c are defined as the probability of successful transmission at the AP and a client, respectively. P_{ca} is the average conditional capture probability, and P_{col} is the collision probability. T_{s1} and T_{s2} are the average time of successful transmission for the AP and for a client, respectively. T_{add} is the additional time for capture effect, as defined in Section IV-C. Moreover, T_c is the collision time and $E\{L_p\}$ is the average payload size in a data frame.

The above equation reflects that the throughput is contributed by three components, i.e., data transmission from the AP to clients without utilizing capture effect, data transmission from clients to the AP, and data transmission from the AP to clients with capture effect, but it must exclude overhead from RTS, CTS, ACK, etc. Such overhead is captured by T_{s1} and T_{s2} as follows

$$\begin{cases} T_{s1} = E\{T_{AP}\} = E\{T_{data}\} + T_{SIFS} + T_{ACK} + T_{DIFS} \\ T_{s2} = E\{T_A\} = T_{RTS} + 3T_{SIFS} + T_{CTS} + E\{T_{data}\} + T_{ACK} + T_{DIFS} \\ T_c = T_{RTS} + T_{DIFS}. \end{cases} \tag{14}$$

Since there are N clients and one AP, P_{tr} is given by:

$$P_{tr} = 1 - (1 - P_{t0})(1 - P_t)^N. \tag{15}$$

When all clients keep silent and only the AP sends packet, the AP sends packet successfully. Thus, P_A is given by:

$$P_A = P_{t0}(1 - P_t)^N. \tag{16}$$

The client transmits packet successfully when only one of the clients transmits no matter whether the AP keeps silent or active. Thus, P_c is given by:

$$P_c = NP_t(1 - P_t)^{(N-1)}. \tag{17}$$

The collision probability is given by:

$$P_{col} = P_{tr} - P_A - P_c. \tag{18}$$

The average conditional capture probability P_{ca} is associated with the capture threshold, i.e., P_{ca} decreases as the capture threshold increases. Considering a propagation model with deterministic power attenuation and Rayleigh fading, the instantaneous power $\mathcal{P}$ is exponentially distributed as:

$$f_p(\mathcal{P}) = \frac{1}{\mathcal{P}_0} e^{-\frac{\mathcal{P}}{\mathcal{P}_0}}, \mathcal{P} > 0, \tag{19}$$

where $\mathcal{P}_0$ is the received local-mean power and is determined by $\mathcal{P}_0 = Ar^{-n}\mathcal{P}_{tx}$, where Ar^{-n} is the deterministic path-loss law [20], n is the path loss index (usually with a range $2-4$), $\mathcal{P}_{tx}$ is the transmit power at the transmitter, and r is the distance between the transmitter and the receiver. We assume A, n, and $\mathcal{P}_{tx}$ are the same for all nodes. To guarantee success of capture effect, the SIR needs to exceed the capture threshold z. In our design, the SIR ($\gamma = \mathcal{P}_u/\mathcal{P}_i$) at a client u is the ratio of signal power from the AP over that from an interfering Client i. The conditional capture probability is given by [21]

$$P_{ca}(\gamma > z|i) = \frac{1}{1+z(r_i/r_u)^{-n}}, \tag{20}$$

where r_u is the distance between the AP and the receiving client u, and r_i is the distance between the interfering client i and the receiving client u. Considering the clients are uniformly distributed around the AP, the probability density function (PDF) of r_u is given by

$$h(r_u) = 2r_u, 0 < r_u \le 1. \tag{21}$$

The PDF of r_i is calculated by [22]

$$h(r_i) = \frac{1}{2}\frac{1}{B(2,2.5)}\frac{r_i}{2}\left(1-\frac{r_i}{2}\right)^{\frac{3}{2}}, 0 < r_i \le 2, \tag{22}$$

where $B(\cdot,\cdot)$ is the Beta function. The approximate average conditional capture probability can be computed as follows [22]:

$$P_{ca} = \int_0^1\int_0^2 \frac{h(r_i)h(r_u)}{1+z\left(\frac{r_i}{r_u}\right)^{-n}}\mathrm{d}r_i\mathrm{d}r_u. \tag{23}$$

From the above equation we know that P_{ca} is determined by the capture threshold z, and z is determined by the capture link rate. As explained in Section IV-C, setting up a dual link needs to check an additional condition, i.e., $T_{add} \le T_{AP}/\beta$. If the β is set larger, then the threshold z is higher, which means it demands a higher capture link rate. When the link rate and data frame length are variable, it is difficult to derive a closed-form relationship between β and z. However, given a fixed link rate and data frame length, the value of z can be determined from β as follows. First, the allowed additional time is determined, i.e., $T_{add} = T_{AP}/\beta$, where $T_{AP} = T_{S1}$ since link rate and data frame length are fixed. Following that, the transmission time of the capture link T_2 is determined as $T_2 = T_{add} + T_1 + T_{SIFS} - T_{ACK}$, where T_1 is the transmission time for data frame in half duplex link. Since the link rate and the data frame length are fixed, T_1 is also a fixed value. With T_2 and the data frame length, we can get the transmission rate of downlink. Finally, the capture threshold z is determined by the transmission rate.

Substituting Eqs. (14)–(18) and (23) into Eq. (13) and let $T_{add} = T_{s1}/\beta$, we can derive the throughput. Note the above derivation assumes the maximum additional time that satisfies $T_{add} \le T_{AP}/\beta$, so the derived throughput is the minimum throughput that can be achieved by our protocol.

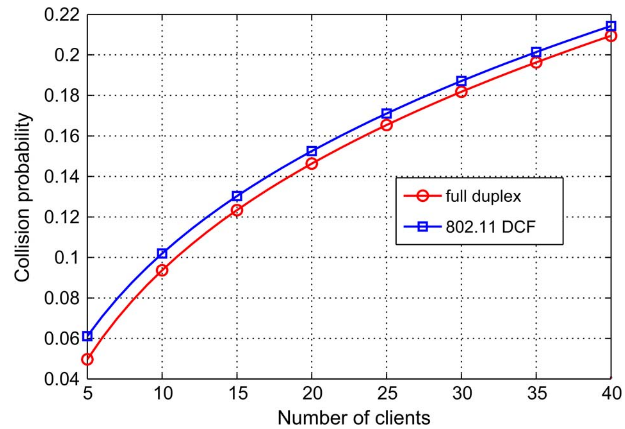


Fig. 7. The probability of collision.

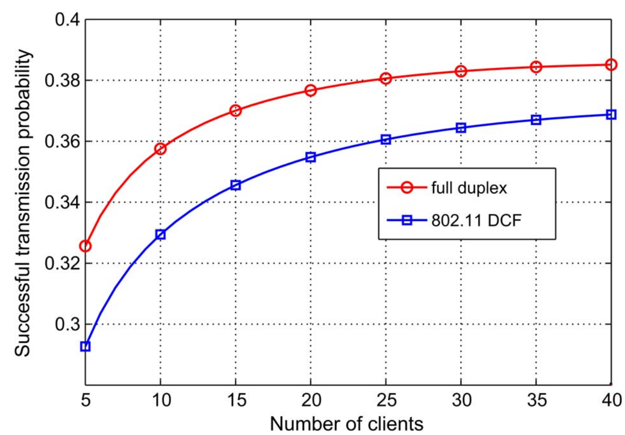


Fig. 8. The probability of successful transmission.

VII. Performance Evaluation

In this section the performance of A-Duplex is first evaluated based on a protocol model, and then a physical model is used to evaluate A-Duplex in a more practical setup.

A. *Performance Results Based on the Protocol Model*

We set the minimum contention window to 32 for both the AP and clients and the maximum contention window to 128 for the AP and 1024 for clients. The collision probabilities (Eq. (18)) of our protocol and 802.11 DCF are shown in Fig. 7, where the result of 802.11 DCF follows the analysis in [19]. The probability of successful transmission ($P_A + P_c$ given by Eqs. (16) and (17)) are shown in Fig. 8. Thus, the collision probability of A-Duplex is lower than 802.11 DCF and the successful transmission probability is higher than 802.11 DCF, which leads to higher throughput in A-Duplex even if the capture effect is not considered as shown in Fig. 9. However, without capture effect, the improvement over 802.11 DCF with RTS/CTS is small.

To evaluate the saturation throughput, the system parameters are selected according to Table I. We set the link rate to 18 Mbps. We set $\beta = 2.2$. Considering the requirement of $T_{add} = T_{AP}/2.2$, the link rate considering capture effect needs

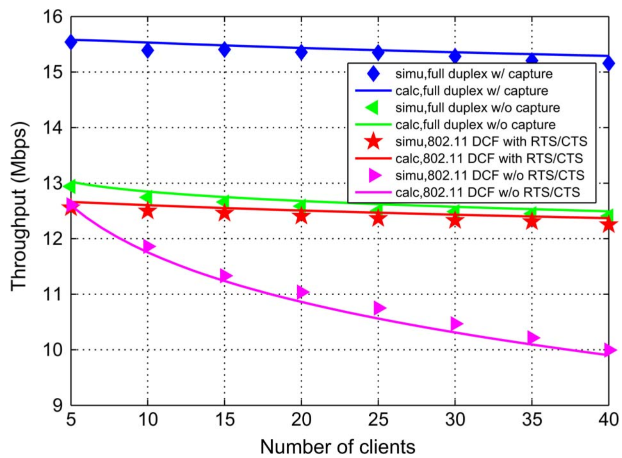


Fig. 9. The throughput results under the protocol model.

TABLE I
PARAMETERS USED IN SIMULATIONS

Parameter	Value	Parameter	Value
PHY header (PH)	20 μs	RTS(FD)	21 Byte + PH
PHY preamble	16 μs	RTS(802.11)	20 Byte + PH
PHY signal	4 μs	CTS	14 Byte + PH
Payload (default)	1500 Byte	ACK	14 Byte + PH
Link rate	18 Mbps	W	16
Capture rate	12 Mbps	W_0	16
Slot time	9 μs	m	6
SIFS	16 μs	m_0	3
DIFS	34 μs	capture threshold	5 dB

to be 12 Mbps. To support such a rate, the capture threshold is 5 dB [12]. From Eq. (23) we can get $P_{ca} = 0.4371$. Both simulation (marked by symbols) and analytical (represented by line) results of throughput are shown in Fig. 9. From these results, we know that: 1) The analytical model is valid, as simulation results are consistent with analytical results; 2) A-Duplex improves throughput by 23% and 24% over 802.11 DCF with RTS/CTS for 5 clients and 40 clients, respectively, and 24% and 54% over 802.11 DCF without RTS/CTS for 5 clients and 40 clients, respectively.

B. Performance Results Based on the Physical Model

1) Physical Model Setup: The physical model simulation is conducted in Matlab. In addition to throughput, other performance metrics such as fairness, packet loss, and packet delay are also considered in performance evaluation. The system parameters are given in Table I except that link rate, capture rate (i.e., the link rate under capture effect), and capture threshold are determined dynamically. The clients are set in a wide area uniformly distributed around the AP. We consider a propagation model with deterministic power attenuation and Rayleigh fading. Thus, the channel is a time varying channel. However, we assume that the channel only changes between packets, i.e., the channel parameters remain the same during the transmission of each packet.

TABLE II
THE MINIMUM REQUIRED SNR FOR A CLEAN CHANNEL (γ_1 (dB)) AND THE MINIMUM SIR FOR CAPTURE EFFECT (γ_2 (dB))

Mode	Modulation	Coding rate	γ_1 (dB)	γ_2 (dB)
1	BPSK	1/2	5-6	1-4
2	BPSK	3/4	6-7	4-5
3	QPSK	1/2	7-9	5-7
4	QPSK	3/4	9-13	7-12
5	16-QAM	1/2	13-17	12-16
6	16-QAM	3/4	17-20	16-21
7	64-QAM	2/3	20-22	21-23
8	64-QAM	3/4	≥ 22	≥ 23

The path loss in the channel model is given by

$$PL(d) = PL(d_0) + 10 \times n \times log(d/d_0) \text{ [dB]},$$

where we set $d_0 = 1$ m and $n = 3$, and let $PL(1) = 48$ dB to reflect the indoor environment. Moreover, the minimum required SNR to decode different modulation and coding schemes for a clean channel (γ_1 (dB)) and the minimum SIR for capture effect (γ_2 (dB)) are given in Table II, where γ_1 is adopted from [23] and γ_2 is from experimental results in [12].

The rate adaptation scheme adopted in our simulation is ARF [24]. More specifically, when a client misses an ACK three times consecutively, its transmission rate moves to the next lower level. When a client succeeds in 5 consecutive transmissions or the rate is not changed over 10 times, the transmission rate rises to the next higher level. The capture rate is determined according to the SIR MAP table. Note that the AP does not establish dual links when the SIR MAP is empty in the beginning.

2) Saturation Throughput and Fairness: To conduct experiments, we first need to determine an appropriate value for θ in Eq. (3). Since it is hard to find an optimal value for θ in a practical system, we consider discrete values of $\theta = 1/n$, ($n = 1, 2, 3 \cdots$) and then determine a proper value of n. When we set $n = 1$, which means the new SIR value is the latest updated value $SIR_{new} = SIR_{update}$, the successful transmission ratio of dual links is about 55%. When we set $n = 3$, the successful ratio increases to 80%. However, when we increase n further, little improvement can be achieved. As a result, we set $n = 3$ in the following simulations.

In the first set of experiments, 10 clients shown in Fig. 10 are considered, and 40% clients (s1, s2, s3, s4) have hidden nodes. The frame length is set to 1500 Bytes. In this scenario, 8 cases are simulated and compared as shown in Table III. The first case is that all clients and the AP transmit packets based on the 802.11 DCF without using RTS/CTS. The second case is the same as the first case except that RTS/CTS is enabled. The third case is our MAC but setting no limit to the additional time T_{add}. The fourth case is that the AP only establishes dual links with hidden nodes. The next three cases study the performance of A-Duplex when the parameter β is set to 1, 2, and 3, respectively. The last case is that the AP always chooses the client with the best capture rate to establish dual links. As shown in Fig. 11, if we set no limit to T_{add} (i.e., case three), the throughput is

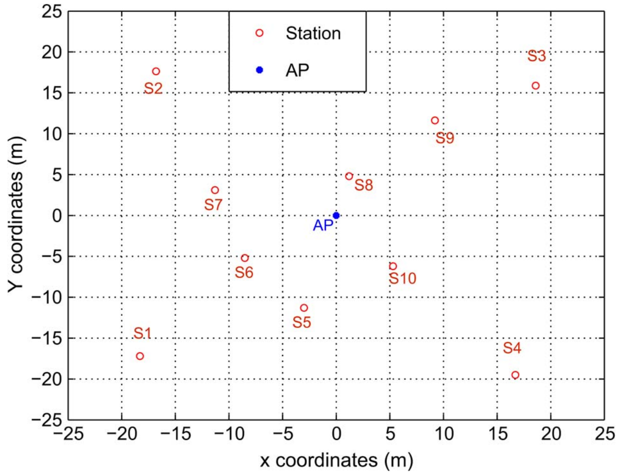


Fig. 10. A typical topology with 40% hidden nodes.

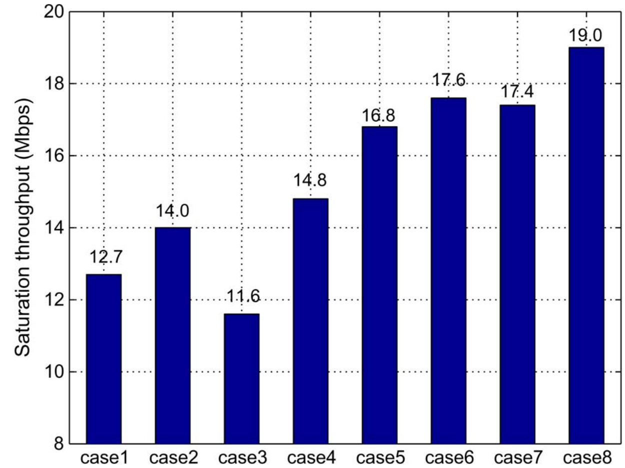


Fig. 11. The throughput in a wireless LAN with 40% hidden nodes.

TABLE III
THE EXPLANATION OF THE 8 CASES IN THE FIRST SET OF EXPERIMENTS

case 1	802.11 DCF without using RTS/CTS
case 2	802.11 DCF with using RTS/CTS
case 3	FD with no limit to the additional time T_{add}
case 4	FD with dual links only between hidden nodes
case 5	FD with $\beta = 1$
case 6	FD with $\beta = 2$
case 7	FD with $\beta = 3$
case 8	FD with the best dual links choice

TABLE IV
THE SHARE TIME OF DOWNLINKS FOR EACH CLIENT FOR THE CASES IN FIG. 11

	case1	case2	case3	case4	case5	case6	case7	case8
I_{Jain}	1	1	1	0.65	0.91	0.99	1	0.38

even lower than that of case one and case two, because some dual links lead to lower transmission rate. When the AP only establishes dual links with hidden nodes (i.e., case four), the throughput is improved by 16% and 6% as compared to case one and case two, respectively. However, with capture effect, the throughput of case six can be improved by up to 39% and 26% as compared to case one and case two. The throughput of case six ($\beta = 2$) is a little higher than that of case five or case seven. The reason is as follows. β of case five is lower than that of case six, so the average transmission rate of dual links in case five is lower than case six. However, for case seven, the average transmission rate of dual links is higher than case six but the capture probability is lower, which leads to a little lower throughput than case six. The throughput of the last case is the highest, since the AP establishes dual links with a client with the best capture rate.

We consider fairness of downlink transmissions among all clients with respect to transmission time. Jain index [25] is used to measure the fairness. The results are shown in Table IV. In case one and case two, the deficit round robin algorithm is applied, so perfect fairness is achieved. From case three to case eight, the virtual deficit round robin algorithm is applied. In case three, the throughput is the lowest, but perfect fairness can be achieved. The reason is that, without setting a limit on T_{add}, most clients can be selected by the AP for dual-link setup, so the virtual deficit round robin algorithm works like the deficit round robin algorithm. In case eight, the highest throughput is achieved, but the fairness is lowest, because no fair scheduling is considered in downlink transmissions. In case four, neither throughput nor fairness is satisfactory. In case four, only hidden nodes are selected for dual-link setup. When the virtual deficit round robin algorithm is applied, it is possible that the default round robin order is changed but no proper client can be selected for dual-link setup, which leads to poor performance in both throughput and fairness. This result indicate that the virtual deficit round robin algorithm is not fit for case four. However, as explained in the end of Section V, we can set a limit to the negative deficit to improve the performance of case four. In cases five to seven, the virtual deficit round robin algorithm works effective, as both hidden nodes and capture effect are considered in dual-link setup, so high performance is achieved in both throughput and fairness.

In the second set of experiments, we evaluate throughput with respect to the number of clients. For each number of clients, we create 50 random topologies and calculate the average throughput among all these topologies. β is set to 2, which is shown to be a proper value in the first set of experiments.

In the first case, we set the data frame length to 1500 Bytes. As shown in Fig. 12, the throughput of 802.11 DCF without RTS/CTS decreases as the number of clients grows, since collisions occur more frequently. The throughput of 802.11 DCF with RTS/CTS is stable even if the number of clients increases. In comparison, A-Duplex achieves stable and much higher throughput. For example, considering 25 clients, our MAC protocol improves throughput by about 48% and 188% as compared to the 802.11 DCF with RTS/CTS and without RTS/CTS, respectively. We also simulate the scheme proposed in [6] where only the historical information is used to search asymmetric dual links and the capture effect is not considered. This scheme is indicated by “full duplex without SIR MAP”. Although the MAC protocol in [7] considers non-hidden nodes for capture effect, it is not compared with our MAC protocol, because it is a reservation based MAC protocol instead of random access. As shown in Fig. 12, its throughput decreases

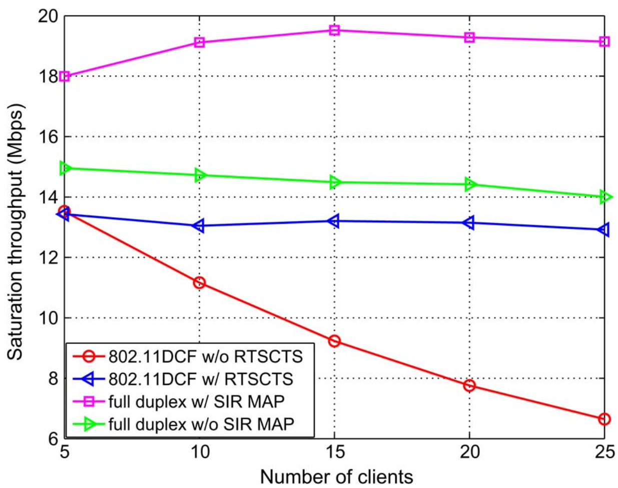


Fig. 12. The throughput under the same frame length.

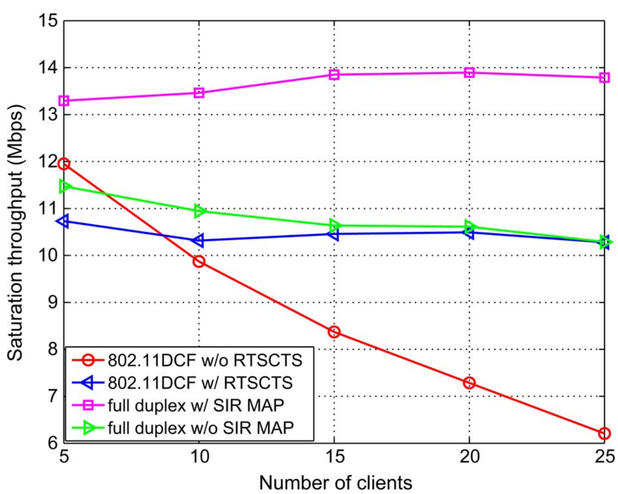


Fig. 13. The throughput under different frame lengths.

slightly with an increasing number of clients, because the overhead of collecting the historical information of dual links increases. Considering 25 clients, the throughput of A-Duplex is 37% higher than this scheme.

In the second case, we consider different frame lengths, i.e., 240, 576, 1000, and 1500 Bytes. The frame length of a data packet is randomly selected from these values. The results of this case are shown in Fig. 13. The advantage of A-Duplex is clearly illustrated. For example, with 25 clients, A-Duplex improves throughput by about 34% and 122% as compared to that of the 802.11 DCF with RTS/CTS and without RTS/CTS, respectively. Furthermore, the saturation throughput of each scheme in the second case is lower than that in the first case. The reason is that in the second case the frame length is shorter than that in the first case, which makes the overhead in the second case larger and results in a lower throughput. As for the scheme without SIR MAP [6], the AP collects the historical information of dual links via blind trial. More specifically, the AP searches the historical information of dual links to select clients to set up asymmetric dual links. Such historical information is accumulated via trial-and-error. Failure in trails leads to overhead. When there are a larger number of clients, more trials are needed to collect historical information. Thus, as the number of clients increases, the overhead increases and the throughput drops, as show in both Figs. 12 and 13. However, in the scenario with different frame lengths (i.e., in Fig. 13), there exists additional overhead. For example, when Client A sends a packet to the AP with the frame length of 240 Bytes, the AP tries to select Client B to form dual links and then send a packet with a frame length of 1500 Bytes. Whether this trial succeeds or not, Client A experiences additional overhead, because its transmission finishes much earlier than the AP's transmission. Due to blind trial, the chance of having such additional overhead grows if the number of clients increases. As a result, for the scheme without SIR MAP [6], the throughput in Fig. 13 drops more quickly. Considering 25 clients, the scheme without SIR MAP does not have any gain as compared to 802.11 DCF with RTS/CTS. However, throughput performance of A-Duplex remains steady as the number of clients grows.

3) Packet Loss and Delay: We evaluate packet loss and delay under the influence of client movement. In A-Duplex the success of dual-link setup depends on timely update of the SIR MAP. If a client moves away from the previous place but its SIR MAP is not updated in time, the dual-link setup may fail. To evaluate the impact of mobility to SIR MAP, we only consider unsaturation mode. The reason is simple: when nodes constantly send packets (i.e., in saturation mode), the SIR MAP can always be updated in time, because the SIR information can be brought back to the AP in a timely fashion.

The random waypoint mobility model [26] is used to simulate client mobility, i.e., each client can randomly choose a direction and a distance for each step. Each client moves a step every 0.25 s, but the step distance varies in different experiments to simulate different moving speed. In our simulations, we set the maximum moving speed to 4 meters per second to evaluate the performance of our proposed MAC protocol since we consider the situation of the people's movement in a wireless LAN under a practical environment. The network topology is shown in Fig. 10 where there are 10 clients. The total number of packets generated in the network is controlled to be in an unsaturation mode. More specifically, every 0.25 s we allocate 25 packets to the 10 clients in a random way and assign 25 packets to the AP with destinations selected randomly from the 10 clients. The packet length is set to 1500 Bytes.

In our simulation, the number of packet retransmission is set to 7. The packet loss rate and the packet delay are shown in Figs. 14 and 15, respectively. In the case of full duplex with SIR MAP(1), a retransmitted packet in a dual-link is not counted as a retransmission. However, in the case of full duplex with SIR MAP(2), whenever a packet is retransmitted, it is regarded as a retransmission. Thus, in full duplex with SIR MAP(1), the number of retransmissions for some packets from the AP to a client may exceed 7 times. As a result, the full duplex with SIR MAP(1) achieves the lowest packet loss. IEEE 802.11 DCF without RTS/CTS results in the highest packet loss due to highest collisions. The packet loss of full duplex with SIR MAP(2) is nearly equal to IEEE 802.11 DCF with RTS/CTS, but both achieve lower packet loss than that of full duplex without SIR MAP.

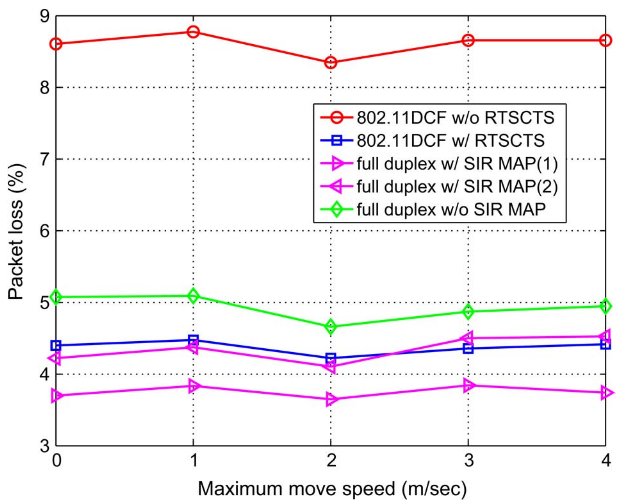


Fig. 14. Packet loss.

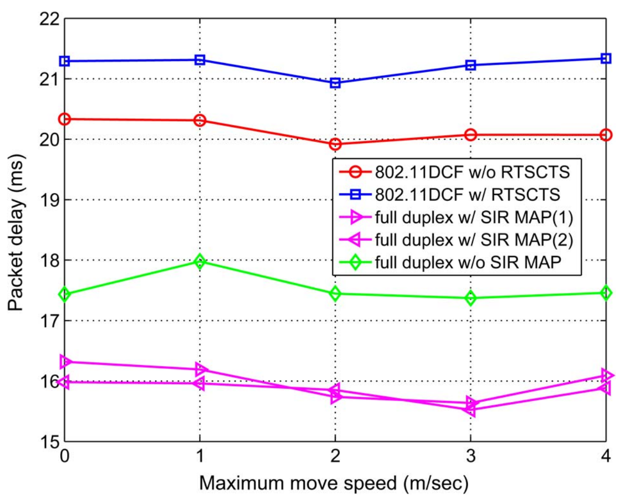


Fig. 15. Packet delay.

The packet delays of full duplex with SIR MAP(1) and full duplex with SIR MAP(2) are almost the same and are the lowest among all scenarios shown in Fig. 15. 802.11 DCF with RTS/CTS has the highest delay. As for full duplex without SIR MAP, its delay is between 802.11 DCF and MAC protocol. Compared to 802.11 DCF without RTS/CTS and with RTS/CTS, the delay of A-Duplex is reduced by about 22% and 26%, respectively.

From the above results we can conclude that the full duplex with SIR MAP(1) is a better option than full duplex with SIR MAP(2), since it leads to lower packet loss and achieves similar delay. Moreover, the results show that the movement of the clients almost does not obviously impact the network performance; The packet loss remains stable as the moving speed varies, which means that the MAP can be dynamically updated in time.

4) The Effect of Path Loss Exponent: In A-Duplex the throughput improvement is mainly attributed to capture effect. However, the SIR depends on the path loss exponent of the environment. Thus, the performance of A-Duplex is evaluated with respect to different path loss exponents. We choose 802.11 DCF with RTS/CTS as benchmark. The ratios of throughput improvement are shown in Fig. 16. As the path loss exponent increases, the throughput improvement becomes more significant.

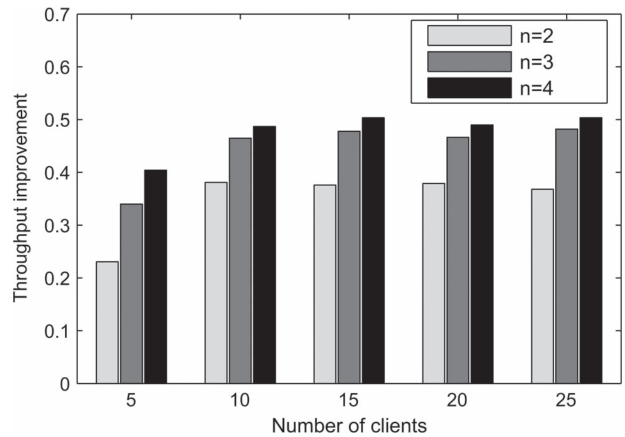


Fig. 16. The ratio of throughput improvement of A-Duplex under different pass loss exponents; 802.11 DCF with RTS/CTS is chosen as the benchmark.

VIII. Discussion

A. Accommodating Full-Duplex Clients

A-Duplex supports coexistence between a full duplex AP and legacy half duplex nodes. However, it can be easily extended to support a network including both full duplex and half duplex clients. First, the SIR MAP is established in the same way as that for the network with only half duplex clients. Second, when the AP selects a client for downlink transmissions during dual-link setup, it treats full duplex clients and half duplex clients similarly. If the AP establishes dual links with a full duplex client, the transmission rate in the downlink is determined according to a rate adaptation algorithm. In other words, the downlink rate does not rely on the SIR MAP, since no capture effect needs to be utilized in symmetrical full duplex dual links. However, if the AP establishes asymmetrical dual links with two different clients, the SIR MAP is still needed to determine the rate for downlink transmission.

Whether the AP establishes symmetrical or asymmetrical dual links depends on the following conditions: 1) If a half duplex client wins the channel access, then the AP can only establish asymmetrical dual links; 2) If a full duplex client wins the channel, then the AP can establish either symmetrical or asymmetrical dual links; 3) If the AP wins the channel, it can establish symmetrical dual links with a full duplex client or a single link with a half duplex client. In all three cases, two strategies can be followed by the AP to select a client for downlink transmission. In the first strategy, the AP always selects a full duplex client to set up symmetrical dual links in cases 2) and 3), and selects a half duplex client that supports the highest downlink transmission rate in case 1). Thus, throughput of the network is much improved. In the second strategy, the AP selects a client according to the virtual deficit round robin algorithm in all three cases. In this strategy, even if a full duplex client wins the channel, the AP may still select another client

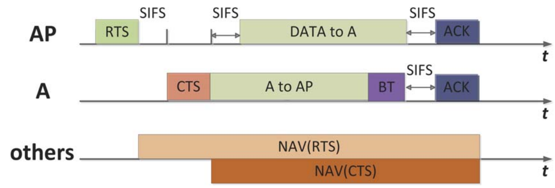


Fig. 17. Symmetric dual links for full duplex clients (Client A finishes data packet transmission earlier).

(either half duplex or full duplex). Thus, the second strategy is more focused on fairness performance instead of throughput.

Similar to the asymmetrical case, symmetrical dual links can be established in two ways. If a full duplex client wins the channel after sending an RTS frame, the AP knows that this is a full duplex client.[2] It then establishes symmetric dual links if it has a packet to the full duplex client. If the AP wins the channel, it sets up symmetrical dual links as depicted in Fig. 17. First, the AP sends an RTS frame to the full duplex client A. When Client A receives the RTS frame and finds that the destination address (DA) for itself, it replies a CTS frame. Moreover, if it has a packet to the AP, the packet is sent after the CTS frame. After the AP receives the CTS frame, it waits an SIFS and then sends the data packet to Client A. When both transmissions finish, the AP and Client A reply ACK simultaneously. In this case, if either the AP or Client A finishes its data transmission earlier, it needs to send a busytone (BT) signal to make sure the two data packets finish at the same time. Suppose Client A has no packet to the AP, it only replies a CTS frame, and then the AP sends a data packet to Client A in a half duplex mode.

B. The Effect of Multiple APs

In this paper, A-Duplex is designed and evaluated by considering only one AP. However, it works well under an environment with multiple APs in the nearby area. We assume all APs that can potentially impact each other can hear each other. In this network, multiple APs and their associated clients contend the same channel. We know that an RTS frame from a client contains the associated AP's address. Thus, when other clients overhear such an RTS frame, they can find out if the transmitting client is associated with the same AP or not. The clients with the same AP follows the procedure of A-Duplex to compute the SIR, but the clients from other APs just ignore the received RTS frame.

It is possible that two clients associated with different APs cannot hear each other. A-Duplex still works properly in this case. For example, when AP 1 has established dual links and then the client of the AP 2 sends an RTS frame to AP 2. In this case, AP 2 will not return a CTS frame, because it has updated its NAV when it receives the CTS frame from AP 1.

The above analysis shows that the clients of different APs get the channel alternately following the random access protocol. Thus, considering one particular wireless LAN with an AP and its associated clients, it works as if there exists no other wireless LANs except that a certain amount of time is left idle for other LANs. As a result, A-Duplex can properly adapt to the situation of multiple APs in the same area. We also conduct simulations to evaluate A-Duplex with multiple APs in the range of each other. Since multiple APs within the range of each other use the same channel, there exist a much larger number of contending nodes. Thus, as shown in Table V, the overall throughput drops but slightly, thanks to the mechanisms of RTS/CTS. However, the throughput for each cell covered by one AP is simply diluted. These simulation results confirm that our protocol can still work well when multiple APs running the A-Duplex protocol are in the range of each other.

[2]We assume the AP has classified the client type (either full duplex or half duplex) during the association process.

TABLE V
THE OVERALL SATURATION THROUGHPUT FOR MULTIPLE APs

Number of AP	1	2	3	4
Saturation throughput ($Mbps$)	19.0	18.6	18.3	18.2

C. A-Duplex in the Case of Imperfect Self-Interference Cancellation

In this paper, we assume that the AP can achieve perfect self-interference cancellation. A-Duplex can also be modified to adapt the case of imperfect cancellation. If the AP cannot achieve perfect self-interference cancellation, uplink reception in dual links is impacted. Thus, the client needs to reduce the uplink rate properly so that the AP can decode the packet correctly. To achieve this goal, the following procedure is taken. During self-calibration of the full duplex radio on the AP, the remaining self-interference can be quantified as a system parameter. This parameter is carried by the CTS frame from the AP to a client, so an additional byte is added in the CTS frame. In case the AP sets up dual links, the value of this byte represents the remaining self-interference; otherwise, the value is just zero. Finally, when a client receives the value of remaining self-interference, it can adjust the transmission rate of the next frame by checking the modulation and coding table. For example, given the SNR of the current transmission rate and the remaining self-interference, the adjusted rate can be determined by checking Table II.

IX. CONCLUSION

A MAC protocol called A-Duplex was developed in this paper to support a wireless LAN with a full duplex AP and half duplex clients. Packet-alignment based capture effect was considered in A-Duplex to establish dual links to fully leverage full duplex capability of the AP. To this end, a map of network topology and relative signal strength of different links was set up via a dynamic SIR update scheme. Such a map helps the AP form dual links while exploring the advantage of capture effect. To improve fairness of A-Duplex, a virtual deficit round robin algorithm was designed for the AP to select an appropriate client for downlink transmissions. Both analytical and simulation results showed that A-Duplex effectively improved the performance of throughput, packet loss, and delay, as compared to existing MAC protocols. Besides, it maintained a high level of fairness. A-Duplex can be easily applied to legacy half duplex nodes, because it makes no change to the physical layer.

REFERENCES

[1] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proc. ACM MobiCom*, 2010, pp. 1–12.

[2] M. Jain *et al.*, "Practical, real-time, full duplex wireless," in *Proc. ACM MobiCom*, 2011, pp. 301–312.

[3] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[4] E. M. Dinesh Bharadia and S. Katti, "Full duplex radios," in *Proc. ACM SIGCOMM*, 2013, pp. 375–386.

[5] A. Tang and X. Wang, "Balanced RF-circuit based self-interference cancellation for full duplex communications," *Ad Hoc Netw.*, vol. 24, pp. 214–227, Jan. 2015.

[6] N. Singh *et al.*, "Efficient and fair MAC for wireless networks with self-interference cancellation," in *Proc. IEEE WiOpt*, 2011, pp. 94–101.

[7] J. Y. Kim, O. Mashayekhi, H. Qu, M. Kazadiieva, and P. Levis, "Janus: A novel MAC protocol for full duplex radio," Stanford Univ. Comput. Sci., Stanford, CA, USA, Tech Rep., 2013.

[8] W. Zhou, K. Srinivasan, and P. Sinha, "RCTC: Rapid concurrent transmission coordination in full duplex wireless networks," in *Proc. IEEE ICNP*, 2013, pp. 1–10.

[9] X. Wang, A. Tang, and P. Huang, "Full duplex random access for multi-user OFDMA communication systems," *Ad Hoc Netw.*, vol. 24, pp. 200–213, Jan. 2015.

[10] A. Kochut, A. Vasan, A. U. Shankar, and A. Agrawala, "Sniffing out the correct physical layer capture model in 802.11b," in *Proc. IEEE ICNP*, 2004, pp. 252–261.

[11] K. Whitehouse, A. Woo, F. Jiang, J. Polastre, and D. Culler, "Exploiting the capture effect for collision detection and recovery," in *Proc. IEEE 2nd Workshop Embedded Netw. Sens.*, 2005, pp. 45–52.

[12] J. Lee *et al.*, "An experimental study on the capture effect in 802.11a networks," in *Proc. 2nd ACM Int. Workshop Wireless Netw. Testbeds, Exp. Eval. Character.*, 2007, pp. 19–26.

[13] N. Santhapuri *et al.*, "Message in message MIM: A case for reordering transmissions in wireless networks," in *Proc. 7th ACM Workshop Hot Topics Netw.*, 2008, pp. 25–30.

[14] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Std. 802.11-1999, 1999.

[15] C. Gezer, C. Buratti, and R. Verdone, "Capture effect in IEEE 802.15.4 networks: Modelling and experimentation," in *Proc. IEEE Int. Symp. Wireless Pervasive Comput.*, 2010, pp. 204–209.

[16] C. Ware, J. Chicharo, and T. Wysocki, "Simulation of capture behaviour in IEEE 802.11 radio modems," in *Proc. IEEE VTC*, 2001, vol. 3, pp. 1393–1397.

[17] A. Tang and X. Wang, "Medium access control for a wireless LAN with a full duplex AP and half duplex stations," in *Proc. IEEE GLOBECOM*, 2014, pp. 4732–4737.

[18] M. Shreedhar and G. Varghese, "Efficient fair queuing using deficit round-robin," *IEEE/ACM Trans. Netw.*, vol. 4, no. 3, pp. 375–385, Jun. 1996.

[19] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[20] W. C. Jakes and D. C. Cox, *Microwave Mobile Communications*. Piscataway, NJ, USA: IEEE Press, 1994.

[21] M. Zorzi and R. R. Rao, "Capture and retransmission control in mobile radio," *IEEE J. Sel. Areas Commun.*, vol. 12, no. 8, pp. 1289–1298, Oct. 1994.

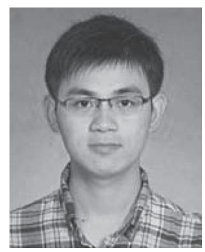
[22] Z. Hadzi-Velkov and B. Spasenovski, "Capture effect in IEEE 802.11 basic service area under influence of Rayleigh fading and near/far effect," in *Proc. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, 2002, vol. 1, pp. 172–176.

[23] "Cisco wireless mesh access points, design and deployment guide, Release 7.3," Cisco, San Jose, CA, USA, Cisco Rep., p. 55, 2012.

[24] A. Kamerman and L. Monteban, "WaveLAN-II: A high-performance wireless LAN for the unlicensed band," *Bell Lab. Tech. J.*, vol. 2, no. 3, pp. 118–133, Summer 1997.

[25] R. Jain, D.-M. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer system," Eastern Res. Lab., Digital Equip. Corp., Hudson, MA, USA, 1984.

[26] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Commun. Mobile Comput.*, vol. 2, no. 5, pp. 483–502, Aug. 2002.



Aimin Tang received the B.S. degree in electrical and computer engineering in 2013 from Shanghai Jiao Tong University (SJTU), Shanghai, China, where he is currently working toward the Ph.D. degree with the Wireless and Networking Laboratory. His current research interests include full-duplex communications, rateless coding, D2D communications, and software-defined wireless networks.

Xudong Wang (S'00–M'03–SM'08) received the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2003. He is currently with the UM–SJTU Joint Institute, Shanghai Jiao Tong University (SJTU), Shanghai, China. He is a Distinguished Professor (Shanghai Oriental Scholar) and is the Director of the Wireless and NetworkinG (WANG) Laboratory, SJTU. He is also an affiliate faculty member with the Department of Electrical Engineering, University of Washington, Seattle, WA, USA. Since August 2003, he has been working as a Senior Research Engineer, Senior Network Architect, and R&D Manager in several companies. He has been actively involved in R&D, technology transfer, and commercialization of various wireless networking technologies. He holds several patents on wireless networking technologies, and most of his inventions have been successfully transferred to products. His research interests include wireless communication networks, smart grid, and cyber–physical systems. Dr. Wang is an Editor of the IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and Elsevier *Ad Hoc Networks*. He was the Demo Co-Chair for the ACM International Symposium on Mobile Ad Hoc Networking and Computing (ACM MOBIHOC 2006), a Technical Program Co-Chair for the Wireless Internet Conference (WICON) 2007, and a General Co-Chair for WICON 2008. He has been a Technical Committee Member for many international conferences.