ON Semiconductor

Is Now

To learn more about onsemi™, please visit our website at
www.onsemi.com

Optimizing Efficiency and Throughput in 11ax

Originally published in Mar. 2018 under the company name Quantenna® Communications, Inc.

Optimizing Efficiency and Throughput in 11ax

Introduction

802.11ax defines two transmission modes that enable simultaneous transmissions to multiple client devices (MU). These modes are Orthogonal Frequency Division Multiple Access (OFDMA) and Multi−User MIMO (MU−MIMO). In this paper, we will review the basics of each transmission mode and discuss how and under which conditions these modes may improve upon the performance of single−user (SU) transmission. In general, no single mode of operation is optimal for all situations. SU, OFDMA and MU−MIMO all show benefits in different scenarios.

OFDMA Basics

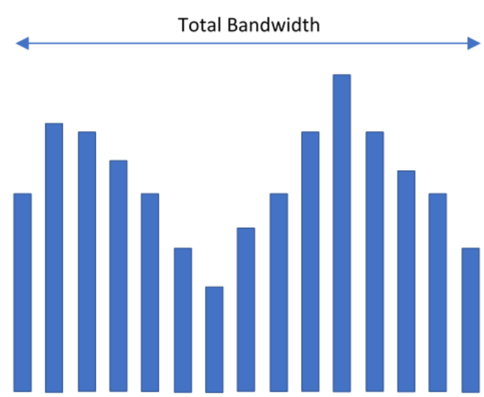

Figure 1. Illustration of OFDM: the Total Bandwidth is Subdivided in a Large Number of Narrow−Band Channels

Even though the full bandwidth conceptually consists of many parallel channels, all subcarriers are modulated with data intended for a single receiver. The data throughput in basic OFDM is determined by the total number of subcarriers and the number of bits modulated on each subcarrier.

OFDM offers many benefits in a wireless environment. It allows flexible use of the spectrum, easy one−tap equalization of the individual subcarriers and a convenient mechanism to deal with multi−path channels.

OFDMA (Orthogonal Frequency Division Multiple Access) is a mechanism that allows a single transmitter to send data to several different receivers at the same time. It uses OFDM as an underlying modulation scheme, but in this case, not all subcarriers carry data for the same receiver. Instead, groups of subcarriers are allocated to different receivers, usually in contiguous blocks, as illustrated in Figure 2. Within 802.11ax, such a group of subcarriers allocated to a user is referred to as a resource unit (RU). In OFDMA, the total throughput does not change relative to OFDM as the total number of tones is the same in each case.

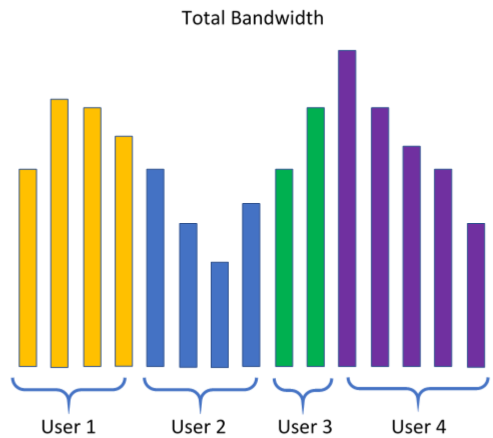

Figure 2. Illustration of OFDMA: Different Subcarriers are Allocated to Different Users

MU−MIMO Basics

Multiple−input Multiple−Output (MIMO) refers to communication systems where both transmitter and receiver have more than one antenna. The use of multiple antennas at both sides of the link creates a multi−propagation environment between transmitter and receiver that can be used to significantly improve the system throughput.

The origin of the improvement can be intuitively understood by thinking of the multi−propagation environment as the equivalent of several separate paths between transmitter and receiver, each of which can be independently used for data transmission. This is referred to as spatial multiplexing. Each such path uses the full bandwidth available to the system. Data associated with a single path is called a “stream”. The number of streams that can be supported depends on the characteristics of the channel and on the number of transmit and receive antennas. If we denote the number of transmit antennas as N_{TX} and the number of receive antennas as N_{RX}, no more than $min\ (N_{TX}, N_{RX})$ streams can be sent.

In SU−MIMO, all streams enabled by spatial multiplexing are sent from a single transmitter to a single receiver. If both sides have N antennas, the link throughput could theoretically be N times higher than the throughput over the same link using single−antenna devices.

A downside of SU–MIMO is that the spatial multiplexing gain is limited by the configuration of both transmitter and receiver, as evidenced by the formula above. In Wi–Fi® networks, there typically exists a significant asymmetry between Access Points (APs) and client devices (STAs). APs may support 4 or 8 antennas, while portable devices may be limited to one or at most two antennas. In such a case, spatial diversity is limited by the number of antennas at the client side. As a result, even in a rich scattering environment with an AP supporting many antennas, spatial multiplexing gains are limited by the client although channel information could still be used to improve performance through beamforming.

MU–MIMO exploits spatial diversity in a different way that is not limited by the number of antennas at individual client devices. An MU–MIMO transmitter uses the spatial diversity to send *N* independent streams to *N* independent receivers. Conceptually, one could think of this as a transmission to a receiver with *N* antennas, with the antennas in different physical locations. MIMO theory tells us that the total downlink throughput in such a case can increase by a factor of *N* compared to the throughput to a single one–antenna device. Because the client devices are in separate physical locations, the transmitter has to apply proper precoding to ensure that each stream does not create interference at the receivers for which it is not intended.

Precoding requires knowledge of the channels between the AP and the various STAs. In practice, this information is obtained by sending dedicated training sequences and responses, referred to as “sounding frames” and “sounding feedback”.

MIMO OFDM Resources

When combining OFDM and MIMO, there are essentially three types of resources that determine the system performance:

1. Bandwidth: the total spectrum available for transmission
2. Time: the amount of time available for data transmission, excluding various types of overhead
3. Spatial multiplexing: the total number of spatial streams that can be sent

The differences between SU, OFDMA and MU–MIMO can essentially be understood in terms of how these three elements contribute differently in each of the transmission modes.

Figure 3 illustrates the usage of the various resources in SU, MU and OFDMA transmission mode. It assumes that clients have a single antenna, while the AP has multiple antennas (like 4 or 8), as is typically the case in practice.

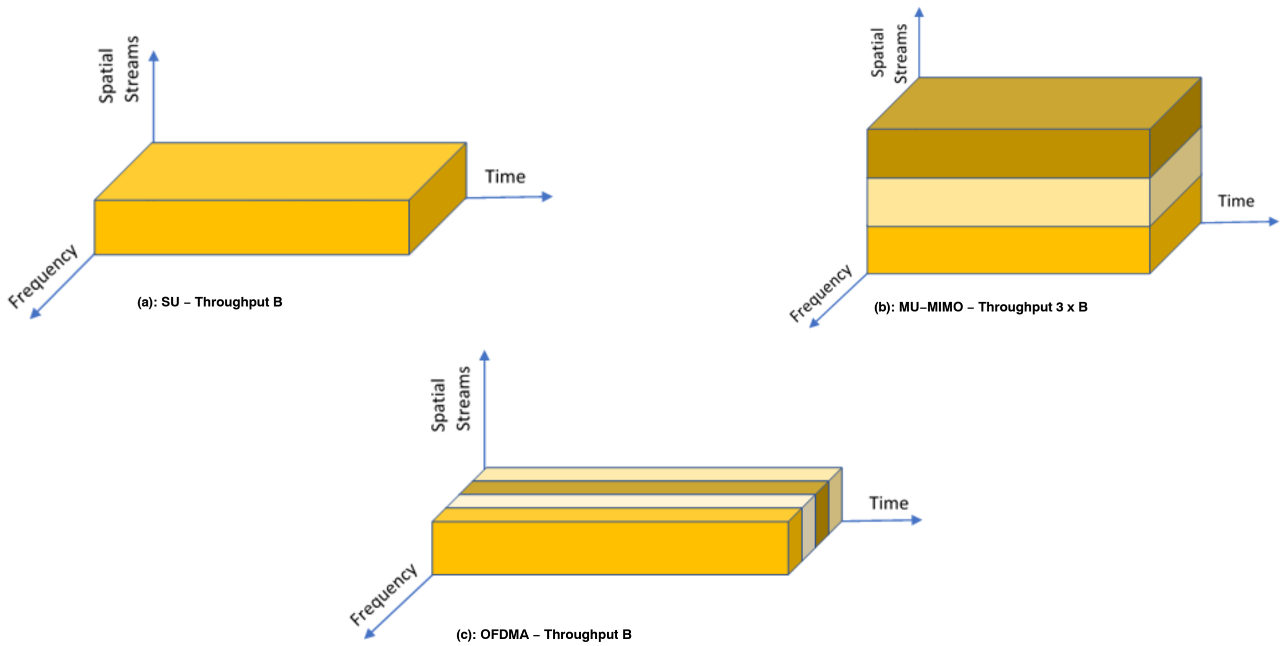

Figure 3. Illustration of Resource Usage for SU, MU and OFDMA Mode of Transmission

In SU (Figure 3(a)), the total spectrum is used for transmission of a single stream (limited by the receive capability of the client) to a single receiver. In OFDMA (Figure 3(c)), the same total spectrum is divided up (as illustrated by the various colors) such that four users each receive a single stream over part of the spectrum. In MU (Figure 3(b)), spatial multiplexing allows transmission of multiple streams. Each stream occupies the total spectrum and is intended for a separate receiver (as illustrated by the various colors). In all three cases shown in Figure 3, the combined volume of the blocks is measure of the total downlink throughput.

Understanding Transmission Overhead

Both OFDMA and MU−MIMO improve upon SU by increasing the usage efficiency of the available resources. In the case of OFDMA, the improvement comes mainly from more efficient usage of time by reducing certain types of overhead that are present in Wi−Fi transmissions. MU−MIMO enjoys the same improvement, but also gets significant benefit from the use of spatial multiplexing.

The time occupied by a Wi−Fi transmission on the wireless medium is only partially used for actual data transmission. Each transmission is burdened with a certain amount of overhead that reduces the effective throughput over the channel. A typical transmission timeline for an SU packet is illustrated in Figure 4.

Figure 4. Wi−Fi Packet Transmission

The timeline can be broken down in a number of stages. Before signal transmission starts, every Wi−Fi packet must wait a random period of time before being allowed access to the medium. This is shown in Figure 4 as "CSMA back−off". CSMA stands for "Carrier Sense Multiple Access", which is the protocol that is used to regulate the access of multiple devices to a shared medium. Typical back−off times are of the order of 100−200 μs. The Wi−Fi packet itself starts with a preamble that allows the receiver(s) to estimate the channel and that also conveys basic information about the format of the frame that follows. Typical preamble times are of the order of 50−100 μs. While the preamble is essential for the reception of the frame, no data is conveyed during this time. The data frame itself follows the preamble and is the only part of the full sequence that contains useful payload for the receiver. Following (successful) reception of the packet, the receiver acknowledges reception of the frame by sending an ACK−frame 16 μs (Short Interframe Spacing or SIFS) after the end of the received frame.

In summary, every time data is transmitted, a portion of the transmission time is spent sending overhead that reduces the effective throughput. Especially for small payloads, the combined overhead of CSMA back−off, preamble and acknowledgement can be significant. For instance, if the payload is only a few 100 bytes long and can be carried in one or two data symbols, the entire transmission might look like the one illustrated in Figure 5.

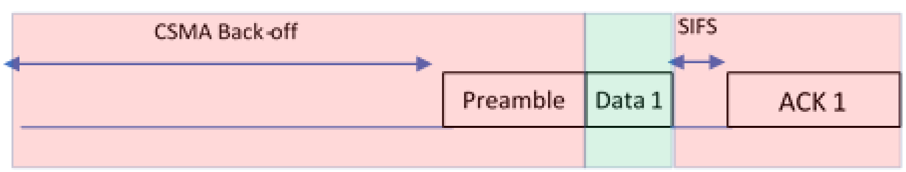

Figure 5. Overhead Versus "Useful" Transmission Time for the Case of a Short Data Frame

In Figure 5, the area highlighted in red represents overhead, while the area highlighted in green is the actual data transmission. It is obvious that in this case, the overhead significantly exceeds the useful part of the packet. Luckily, this only happens for short to very short data payloads. When the payload is larger and the data portion of the packet is much longer, the transmission overhead becomes virtually irrelevant. This case is illustrated in Figure 6.

Figure 6. Overhead Versus "Useful" Transmission Time for the Case of a Long Data Frame

Both OFDMA or MU−MIMO (henceforth collectively denoted as “MU”) allow multiple simultaneous transmissions to share some of the overhead shown in Figure 4 to Figure 6. Specifically, an MU packet containing payloads for multiple users only needs to contend for the medium once and a single common preamble is used for the entire frame. This means that CSMA back−off overhead and preamble overhead are now shared between N users, instead of being replicated N times as for SU.

For MU, the acknowledgment of the data uses a mechanism that is different from the one shown in Figure 4. Instead, 802.11ax defines two procedures that can be used to acknowledge reception of MU packets. The overhead of these acknowledgments also needs to be factored in to see whether MU provides a performance benefit over SU. The two acknowledgement protocols are illustrated in Figure 7.

Figure 7. Acknowledgment Procedures Illustrated for Two Users

The first type of MU acknowledgment has the first receiver respond immediately after the end of the received packet, similar to the way SU acknowledges reception of a data frame (see Figure 4). Acknowledgements for subsequent receivers are sent in response to a Block Ack Request (BAR) sent by the original transmitter. One BAR is needed for each receiver. This is illustrated in the top part of Figure 7.

Alternatively, as illustrated in the bottom part of Figure 7, the transmitter can send a Trigger frame of type MU−BAR that elicits an Uplink OFDMA frame where all outstanding individual acknowledgments are multiplexed into a single reception.

Which of these two mechanisms is more efficient depends on the number of users and the specific formats used for the ACK, BAR, MU−BAR and UL OFDMA response frames.

Comparison of MU Benefits

The reduction of overhead associated with Wi−Fi transmission is a key benefit of MU transmission. In addition, MU−MIMO brings the extra gain of spatial multiplexing. In this section, we illustrate how the various benefits play out in a number of different scenarios.

The gain from improved time efficiency will depend on the amount of overhead that can be saved. Obviously, for short data payloads where the amount of overhead is significant (see for instance Figure 5), more improvement can be expected than for cases where the transmission overhead wasn’t that large in the first place. Figure 8 shows an illustration that compares SU, OFDMA and MU transmission for a short data payload.

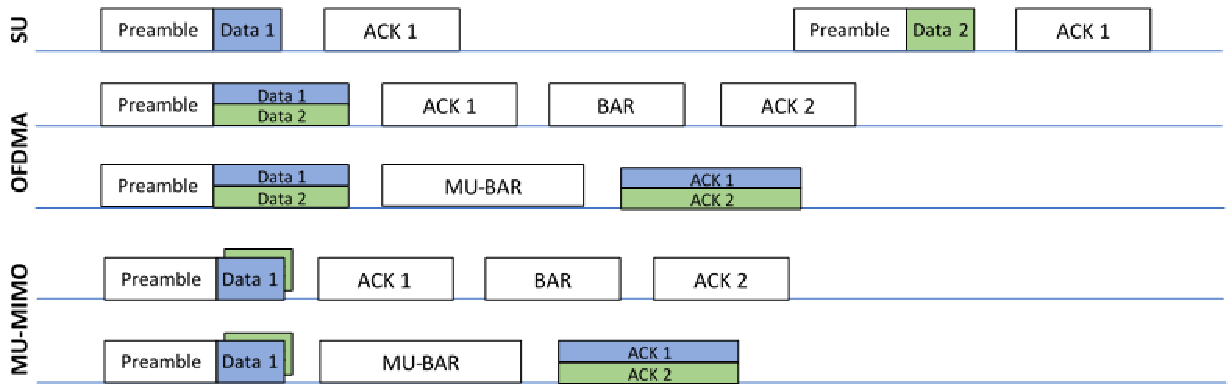

Figure 8. SU vs. MU Transmission for Short Data Payload. Overhead Can Be Reduced Significantly by Using MU Transmission, Leading to Lower Transmission Times

The top part of Figure 8 shows the transmission of two short data payloads using SU. Each individual transmission incurs the various types of overhead shown in Figure 4. The remainder of Figure 8 shows the equivalent transmissions using MU (either OFDMA or MU−MIMO) with different ACK schemes. In all cases, the same amount of data is transmitted to the two users. For the case illustrated here, the transmission sequence is notable shorter when MU is used vs. when SU is used. There are minor differences between MU and OFDMA and total transmission time also depends on the acknowledgment protocol that is used.

MU−MIMO and OFDMA both reduce transmission overhead by the same amount, especially CSMA back−off time and preamble. In addition to this, MU−MIMO is able to send the data payloads in a shorter amount of time since it can use the full spectrum for both users. OFDMA on the other hand allocates half of the spectrum to each user, making it less efficient than MU−MIMO.

Note that for MU−MIMO to operate, channel information is needed. This results in some additional overhead that is not shown in Figure 8 since it happens on a much longer timescale than the time illustrated here. However, the sounding overhead must be factored in to establish the actual throughput in MU−MIMO mode.

Considering all factors contributing to transmission overhead, it would appear that for the case of short payloads shown in Figure 8, either OFDMA or MU−MIMO could give the best performance. A short payload could be made up of short, infrequent packets that cannot be aggregated efficiently. These payloads could be a couple of hundreds of bytes long,

A similar illustration can be made for larger data payloads. These payloads could be 100,000 bytes or more per frame. In this case, the relative transmission overhead will become less significant. In fact, as we approach the situation illustrated in Figure 6, relative overhead will be all but negligible.

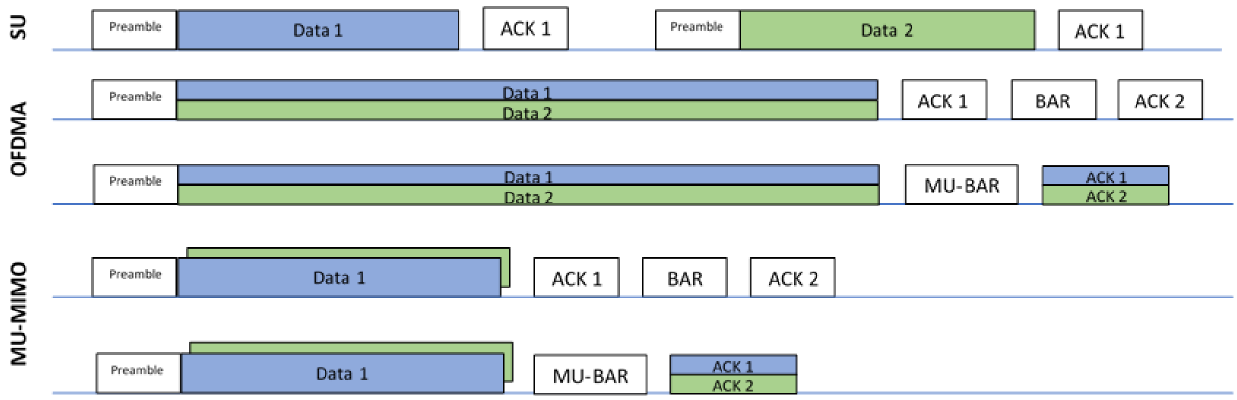

Figure 9. SU vs. MU Transmission for Longer Data Payload. Since Overhead is Relatively Less Important, the Gain from OFDMA Diminishes Relative to SU. MU−MIMO Benefits from Spatial Multiplexing

In Figure 9, the advantage of OFDMA diminishes since the relative contribution of the various transmission overheads is less in the case of larger data payloads. This also affects MU−MIMO, but with data transmission taking up more time overall, the impact of spatial multiplexing provides significant gain over other modes of transmission, even after sounding overhead has been accounted for.

In contrast to Figure 8, we see that OFDMA is not necessarily better than even SU and may in fact yield lower performance, depending on the ACK mechanism that is used. MU−MIMO on the other hand continues to enjoy a significant performance benefit over both SU and OFDMA thanks to the power of spatial multiplexing. Sounding overhead will slightly reduce the raw performance gain, but MU−MIMO looks like the preferred mode of operation for the case illustrated in Figure 9.

Conclusions

802.11ax defines several transmission modes: SU, OFDMA and MU−MIMO. To optimize overall performance, it is important to select the right mode for every situation as there is no single mode that will provide the best performance in all cases. While this is a multi−faceted problem, we can employ some rules of thumb to decide when to use which mode.

First of all, for MU (either OFDMA or MU−MIMO) to work efficiently, concurrent traffic to multiple users needs to be available at the same time. If traffic for different users is bursty in nature and arrives at different times, it can obviously not be combined into a single MU transmission. In such cases, SU is the transmission mode of choice.

When traffic for multiple users is available simultaneously, MU becomes a viable option. Which mode to use will then depend mostly on the type and volume of the traffic. The payload carried by a Wi−Fi frame will consists of a number of application packets. The transmitter will attempt to group these application packets into larger aggregates as long as the latency requirements of the application are not exceeded. If application packets are small or they arrive infrequently, the resulting aggregates (and thus the Wi−Fi payload) will be small. Based on this payload size, we can consider the following cases:

- For very small packet payloads, MU provides performance improvement through reduction of transmission overheads, as discussed above. This applies to both OFDMA and MU−MIMO. However, the added sounding overhead needed for MU−MIMO may reduce the effective throughput for this mode. In these cases, either MU−MIMO or OFDMA could be the preferred transmission mode.
- For small to medium−size payloads, the contribution of transmission overhead diminishes and so does the gain of OFDMA. MU−MIMO starts to show larger gains thanks to the increased airtime taken up by the data payload and spatial multiplexing. The gain of spatial multiplexing will outweigh the overhead of channel sounding. In these cases, MU−MIMO is the preferred transmission mode.
- For large payload sizes, there is very little benefit from using OFDMA. In fact, due to restrictions in packet duration and available RU allocations, OFDMA may in some cases perform worse than SU. The gain for MU−MIMO continues to grow since the gain of spatial multiplexing far exceeds the sounding overhead. In these cases, MU−MIMO is the preferred transmission mode.

Quantenna, now a division of ON Semiconductor, has pioneered the use of MU−MIMO with its 4x4 QSR1000 chipset back in 2014, and also was the first to introduce OFDMA in its 802.11ax chipset QSR10G−AX in 2016. Based on our experience, we have seen that both MU−MIMO and OFDMA provide key benefits for different use cases.

PUBLICATION ORDERING INFORMATION

LITERATURE FULFILLMENT:
Email Requests to: orderlit@onsemi.com

ON Semiconductor Website: www.onsemi.com

TECHNICAL SUPPORT
North American Technical Support:
Voice Mail: 1 800−282−9855 Toll Free USA/Canada
Phone: 011 421 33 790 2910

Europe, Middle East and Africa Technical Support:
Phone: 00421 33 790 2910
For additional information, please contact your local Sales Representative