Early Results from Next-Generation Broadband Wireless Systems

Subir Varma and Tassos Michail
Aperto Networks
1637 South Main Street, Milpitas CA 95035

ABSTRACT

The first generation of Broadband Wireless systems were derived from cable modem architectures, as a result they were not able to handle wireless specific channel impairments. This restricted first generation systems to line-of-sight (LOS) and single cell type of deployments. The next generation of systems are specifically designed for wireless deployment, and hence incorporate a number of advanced features that enable them to function under highly impaired channel conditions.

Field testing and early deployments of next generation Fixed Broadband Wireless Access (FBWA) systems developed over the past two years is beginning to occur under a wide variety of circumstances and in multiple sub-10GHz frequency bands in global markets. These trials include variable equipment configuration conditions LOS, obstructed line-of-sight (OLOS) and non-line-of-sight (NLOS) locations, heavily loaded network traffic, symmetrical voice, data and video transmissions, and variable cell radii.

As these systems are beginning to get deployed, experimental results that show the benefits of the advanced features are being collected. The objective of this paper is to share some of these results with a wider audience.

The authors propose to present results of these field tests, the changes being made in architectures, configurations and protocols based on these results, and anecdotal data on benchmarks for IP network functions over FBWA architectures.

1.0 Introduction

FBWA is an emerging technology, which seeks to solve the last-mile access problem through bypassing the local loop infrastructure, owned by incumbent telecom carriers. Even though FBWA products have been available for last few years, they have yet to reach the ubiquity of cable modem or DSL technology. One of the main reasons for this state of affairs is that first generation FBWA products have been derived using cable modem technology, which puts the following restrictions on their deployment:

- First generation products work, provided very little wireless related link impairments, such as fading and multipath
- First generation products are not able to handle interference very well
- First generation products have limited Quality of Service (QoS) capabilities

Due to the first two restrictions, first generation products can only be deployed in single cell configurations, with a very high base station antenna placement and LOS links with every subscriber. This severely restricts the coverage, and subsequently the viability of first generation FBWA technology. The third restriction limits the services that can be delivered over FBWA, and hence the profitability of FBWA service providers.

Next generation FBWA products, which are architected from ground up to operate in wireless environments, are now beginning to appear in the market. They include a number of features that help to increase the robustness of the wireless link, and thus enable the system to keep operating under highly impaired conditions. Next generation systems incorporate innovations in both the Medium Access Control (MAC) and Physical Layer (PHY) technologies. Next generation FBWA MACs are designed to overcome wireless related impairments and make the link more robust. As compared to traditional point to multipoint MACs, they have a greater ability to control the PHY and change its parameters in response to link conditions. Next generation FBWA PHYs are designed to withstand deep fades and multipath introduced due to OLOS/NLOS situations.

The capabilities present in Next generation FBWA products allow the service provider to deploy them in dense cellular configurations with a high degree of frequency reuse. It allows the Customer Premise Equipment (CPE) to be installed indoors or under the eaves, thus reducing major source of cost to the service provider. It also allows the service provider to realize additional revenues by using the QoS capabilities in the next generation equipment. With these capabilities, the FBWA market should grow rapidly and become a serious competitor to cable and DSL.

The objective of this paper is to provide experimental validation of several aspects of next generation FBWA product behavior, using equipment currently being shipped by Aperto Networks. These experiments have been carried out using a combination of laboratory and outdoor environments. The laboratory set up allows us to regulate the amount of impairments, such as noise, introduced into the system, and thus plot the system behavior over a range of conditions. Outdoor results on the other hand, provide a sample data point, which is representative of the conditions that exist at the time the experiment is carried out. In several cases we validate the curves obtained from the laboratory experiments, by showing that the results from the outdoor tests also fall on these curves, for similar values of the control variable.

The next generation FBWA systems whose performance has been described in this paper can be categorized as Adaptive Time Division Multiple Access (TDMA). As shown through measurements conducted on functioning equipment, Adaptive TDMA is capable of efficiently solving the problems associated with FBWA deployment, while delivering QoS and bandwidth management support. Major standards organizations such as IEEE 802.16 working group on broadband wireless access standards [1] and ETSI HIPERMAN [2] have chosen to base their specifications for FBWA on Adaptive TDMA technology.

In the rest of the paper we present a collection of experimental results that emphasize the importance of advanced features of next generation FBWA systems. In the next section we discuss the effect of Automatic Repeat Request (ARQ) mechanisms on system performance. In section 3, we illustrate the advantages of an adaptive modulation/coding scheme and present a set of laboratory and field results. Section 4 includes a brief discussion on fading conditions often encountered in OLOS/NLOS sites and a set of data that reveal how diversity techniques can potentially help to provide acceptable performance. Finally, in section 5, we discuss the QoS capabilities of next generation systems and the advantages they provide both to the service provider and the end user.

2.0 Automatic Repeat Request (ARQ)

FBWA systems are expected to operate in environments that are often characterized by severe fading and interference conditions, which render physical layer error-correction coding schemes (such as Forward Error Correction (FEC)) insufficient. The use of automatic repeat-request (ARQ) mechanisms is a well-known technique to handle transmission errors, which occur when data is transmitted over noisy channels [3]. ARQ schemes rely on a feedback control channel to acknowledge packet reception by the receiver; packets received in error are retransmitted at the data-link layer, which makes the link appear error-free to higher layer protocols.

The existence of an ARQ scheme at the MAC layer is particularly important for the performance of flow control protocols, such as TCP. When TCP operates over links with relatively high error rates, it may experience severe performance degradation. TCP employs a sliding window algorithm to control the amount of outstanding (and unacknowledged) data in order to detect and react to network congestion. A traffic source increases its TCP window size until a packet loss is detected. At this point the source reduces the window size and the cycle repeats. In wireless networks however, packet loss may frequently be due to link-level errors caused by channel impairments. It has been shown [4] that the effective throughput of a TCP connection decreases as the probability of packet loss increases, because of the retransmissions at the TCP layer. ARQ addresses this problem by triggering retransmissions at the MAC layer, which make the link appear error free. The Aperto system implements ARQ in hardware, which allows it to do several re-transmissions within duration of tens of milliseconds. This facility is very important, because if the TCP packet is not recovered within half a second, the TCP source times out and re-transmits the packet.

In what follows, we present both laboratory and field results that illustrate the importance of ARQ mechanism on fixed wireless systems. Given that most of the popular Internet applications use TCP as the preferred flow control algorithm, our performance measure is TCP throughput; we compare the performance of flows running over ARQ against flows running without ARQ support.

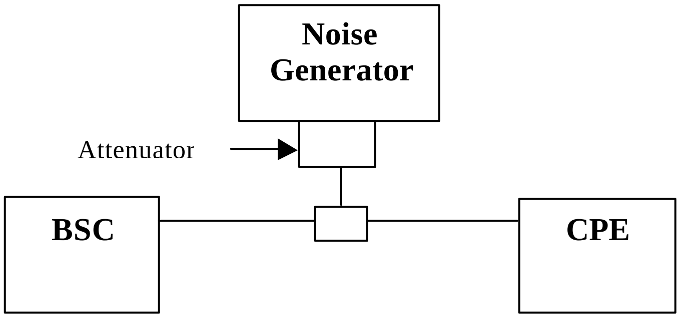

Figure 1: Experimental laboratory setup to control system SNR

The first set of results was collected using the experimental test-bed shown in Figure 1. A base station controller (BSC) and a CPE are connected with cables, via a Noise source that generates additive white Gaussian noise. By varying the attenuation at the output of the noise source, we are able to control the signal to noise ratio (SNR) of the system with granularity of 0.1dB.

We have measured the average TCP throughput rate and the MAC packet error rate for a set of SNR values, with the modulation level fixed at 16QAM. We have assumed two different FEC coding levels, referred to as Low and High FEC. High FEC offers more protection at the cost of increased overhead. Figure 2 plots the normalized TCP throughput rate versus the packet error rate, with and without ARQ. The improvement we observe when using ARQ is impressive. With ARQ, the system is capable of sustaining packet error rates up to 10 times higher, for an equivalent throughput.

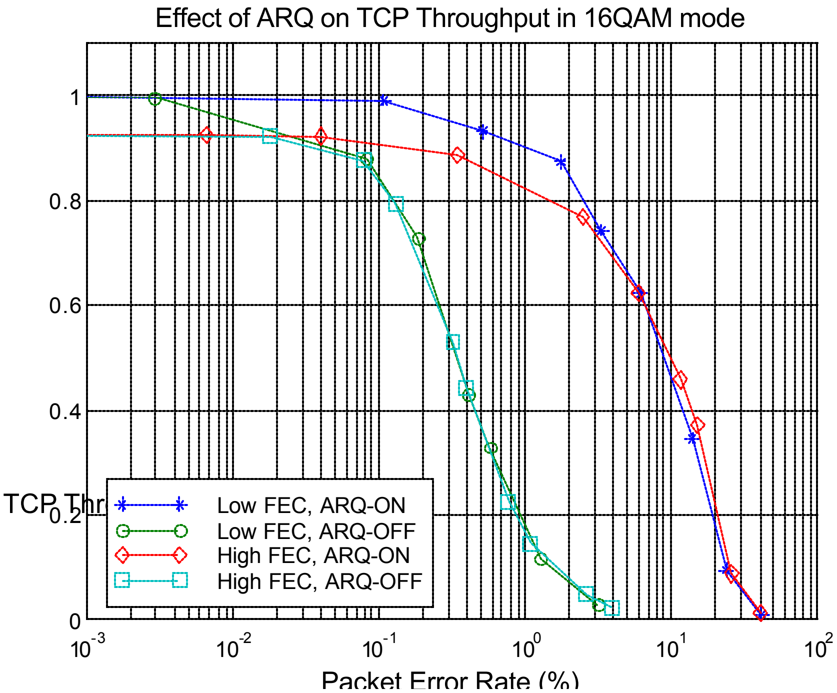

Figure 2: Effect of ARQ on TCP throughput in 16QAM mode

Next we evaluate the effect of the number of ARQ retries on throughput performance. Figure 3 shows the throughput efficiency versus the SNR level for different numbers of ARQ retransmissions (between 0 and 6). Clearly the higher the number of retransmissions gets, the higher the probability of a packet going through which explains the increase in throughput. However, the improvement becomes smaller as we increase the number of retries, since more bandwidth is taken away by the increased amount of ARQ retransmissions. Hence, the number of retries is a critical parameter for system performance and such experiments provide valuable insights on setting it to an optimal value. Note also, that there is a substantial improvement in performance even with a single retry. This implies that applications such as packetized voice, that are very sensitive to latency, can still gain the benefits of ARQ by setting the number of re-transmissions to one.

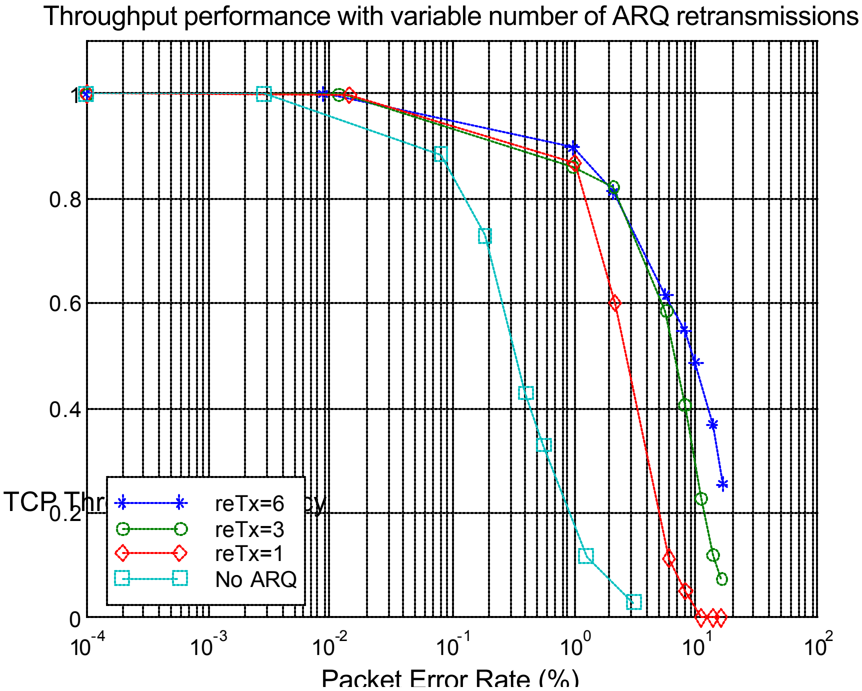

Figure 3: TCP Throughput performance with variable number of ARQ retransmissions

The effect of ARQ is further illustrated through the data presented in Table 1. This data has been collected as part of field tests at two subscriber sites situated approximately 2 to 3 miles away from the BSC, which are operating under OLOS conditions. A comparison between the ARQ and the non-ARQ throughput rates indicates that we get a significant amount of improvement by enabling ARQ. These sample points can also be matched against the curves of Figure 2. A small deviation from the curves could be explained since the curves were generated by calculating average throughput rates.

Table 1: Throughput comparison with and without ARQ at 2 OLOS field-test sites

	CPE-1	Throughput		CPE-2	Throughput	
	Packet Error Rate	ARQ ON	ARQ OFF	Packet Error Rate	ARQ ON	ARQ OFF
High FEC	0.52%	0.90	0.39	0.61%	0.89	0.33
Low FEC	9.68%	0.43	0.0	4.55%	0.72	0.02

2.1 Effect of Wireless Packet Size

The effectiveness of ARQ protocols when used over relatively high bit error rate channels is tied to the packet size used during the transmissions. When too large a packet size is employed, the probability of packet error increases rapidly with the bit-error rate, whereas small packet sizes may be inefficient because of the fixed overhead required per packet. Moreover, the spectrum available to wireless service providers varies in different markets and from provider to provider and the possibility to operate over multiple channelization schemes illustrates the need for adjustable MAC packet sizes.

Aperto Networks' MAC protocol addresses these issues by optimizing the Wireless Protocol Data Unit (WPDU) size to maximize performance. Higher layer packets may be fragmented to span across multiple WPDUs or may be concatenated (packed) into a single WPDU, depending on their size. The WPDU size selected by the MAC is therefore independent of higher layer packet sizes so that it is tailored to the current channel conditions.

As an example consider the case that multiple small IP packets are being transmitted over a "clean" channel. The possibility to pack as many of these packets into a single WPDU would result in a reduced overhead. On the other hand, a large IP packet that is to be transmitted across high error rate link, could be fragmented into multiple small WPDUs which would be more robust during transmission and would likely result in fewer ARQ retransmissions.

Using the same setup of Figure 1, we have conducted a set of measurements of the TCP throughput for various WPDU sizes and for five different SNR values. The attenuation at the output of the noise generator controls the system SNR. The objective of this experiment is to illustrate the variation in throughput that is caused by different WPDU sizes and to get some insights on the optimal WPDU size selection. In Figure 4, we plot TCP throughput versus WPDU size. The five curves correspond to five different SNR levels. At the highest SNR, the channel is clean of errors and the larger WPDU results in higher throughput since the effect of overhead is lower. However, in the presence of errors (SNR lower than 19 dB) it is clear that a WPDU size of approximately 600 bytes results in the highest throughput.

As a result of concatenation and fragmentation at the MAC layer, the system is able to operate efficiently over a range of IP packet sizes. The system independently chooses the WPDU size that leads to optimal performance under the current channel conditions.

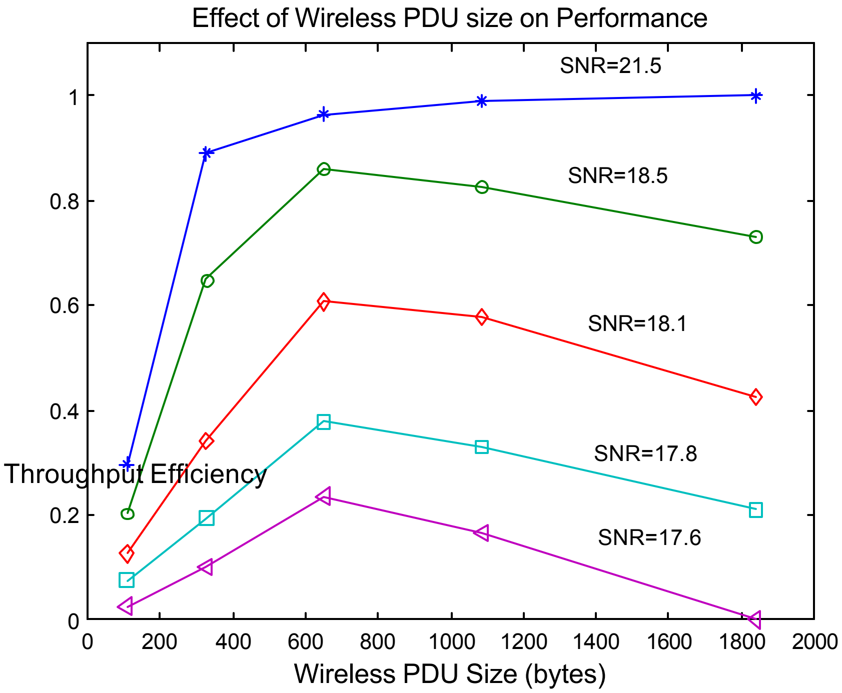

Figure 4: Effect of Wireless PDU size selection on TCP throughput

3.0 Adaptive Coding and Modulation

Next generation FBWA systems have the ability to dynamically adapt modulation and coding levels as a function of the link between the BSC and each CPE. Hence at any point in time, each subscriber may have a different set of values for these parameters, which may independently evolve over time, as link conditions change. The Aperto system allows flexibility in the following set of link parameters:

- *Dynamic Modulation*: The system adaptively switches between QPSK and 16QAM modulation. Dynamic modulation allows the link margin to increase by 7-8 dB, thus leading to the following benefits:
  - Dynamic modulation extends the radius of coverage of the cell, so that subscribers that are closer to the base station receive 16QAM reception, while those further away receive QPSK.
  - Dynamic modulation increases robustness against co-channel interference, thus enabling more aggressive frequency re-use plans and smaller cell sizes.
  - Dynamic modulation helps to enable reception in OLOS/NLOS cases, by enabling the link to function with lower received power.
- *Dynamic Coding:* The system adaptively switches between two levels of FEC coding capabilities. Dynamic coding allows the link margin to increase by about 2-3 dB, and leads to the same set of benefits as listed for dynamic modulation. Since dynamic coding leads to a smaller decrease in link speed, it is used as the first line of defense against link impairments, before dynamic modulation kicks in.

The different modulation and coding levels can be arranged in the following order of increasing robustness:

1. 16QAM, Low FEC
2. 16QAM, High FEC
3. QPSK, Low FEC
4. QPSK, High FEC

Figure 5 plots the normalized TCP throughput as a function of SNR, for each of the four combinations enumerated above. The reader will observe that Level 1 has the highest throughput, but also is the first to fall in performance, as SNR decreases. Conversely, Level 4 has the lowest throughput, but is able to survive the longest as the SNR is decreased. A system that does not have adaptation will be forced to operate in one of these two extreme states:

- If set to Level 1, it will have limited coverage, but will serve those that it can reach, with the best throughput,
- If set to Level 4, it will have the best coverage, but half the bandwidth will be unavailable

Adaptive coding and modulation helps to capture the best of both worlds, since the system is able to supply Level 1 to all subscriber units that have a good link to the base station, and Levels 2 and higher to subscriber units whose link is not as good. Hence the service provider is able to maximize his coverage, without sacrificing any capacity. This is illustrated in Figure 6, in which the thick line shows the response of the system as the SNR is changed. The system is able to optimally jump from one level to another as the SNR changes, and thus closely track the upper envelope of the individual throughput curves.

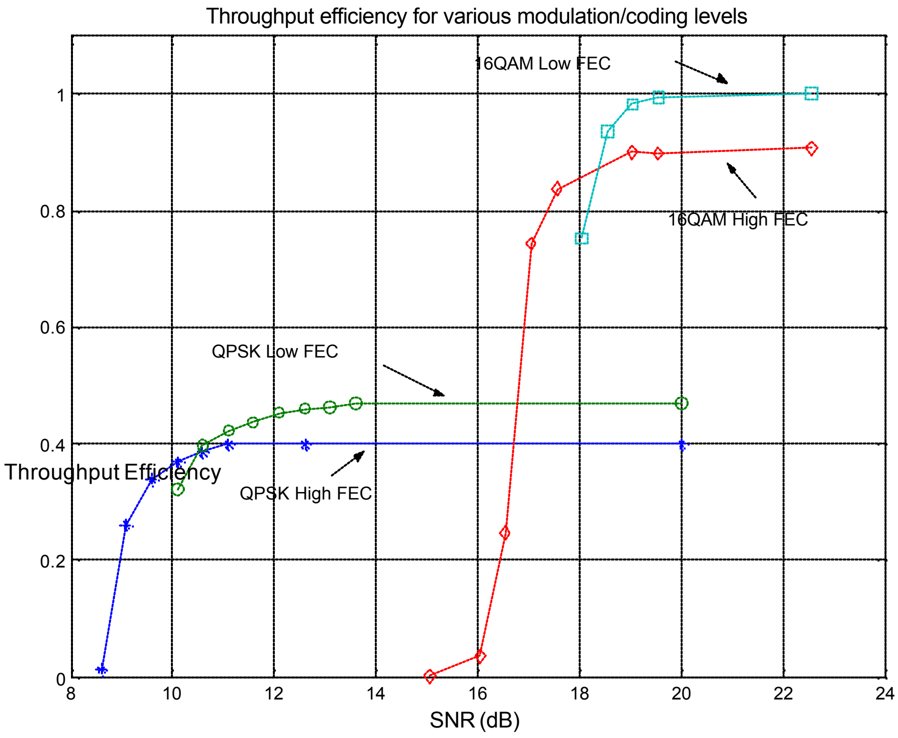

Figure 5: TCP Throughput versus SNR for Static Modulation and Coding

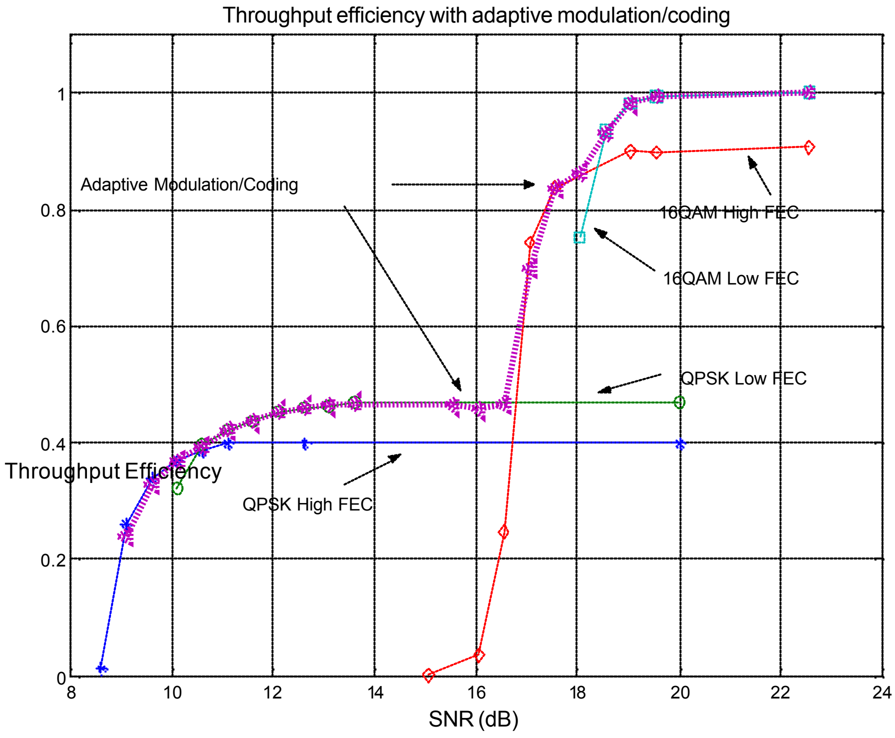

Figure 6: TCP Throughput versus SNR for Adaptive Modulation and Coding

The operation of Adaptive Coding and Modulation is further illustrated through the data presented in Tables 2 and 3. This data has been collected from two subscriber sites using the Aperto Networks product. One of the sites is situated two miles away and has a clean LOS to the base station, while the other site is three miles away and has its LOS obstructed by a tree.

Table 2 shows that the LOS subscriber spends more than 98 % of its time in the 16QAM and Low FEC state, however it does transition to the lower states on rare occasions in response to transient channel impairments. Table 3 shows that the packet error rate for the LOS subscriber is negligible in all four states. The OLOS subscriber experiences very high error rate in the 16QAM state, and hence spends most of its time in the QPSK and Low FEC state. As Table 2 shows, it occasionally transitions to the 16QAM state to probe it for usability, but quickly transitions back to QPSK.

Table 2: Percentage of time spent at each Modulation/Coding level

Location	OLOS Site ~3miles		LOS Site ~2miles	
Modulation/Coding Level	Direction		Direction	
	Up	Down	Up	Down
QPSK, High FEC	0	14.48	0.14	0.20
QPSK, Low FEC	98.59	85.38	0.46	0.55
16QAM, High FEC	1.41	0.14	0.62	0.73
16QAM, Low FEC	0	0	98.78	98.53

Table 3: Packet error rate (percentage) measured at each Modulation/Coding level

Location	OLOS Site~3miles		LOS Site ~2miles	
Modulation/Coding Level	Direction		Direction	
	Up	Down	Up	Down
QPSK, High FEC		0.92	0	0
QPSK, Low FEC	0.31	0.37	0	0
16QAM, High FEC	53.52	98.77	0	0
16QAM, Low FEC			0.004	0.01

4.0 Antenna Diversity and Fading Channels

OLOS and NLOS locations present a special challenge for FBWA installations, especially if the signal propagates through foliage. Figure 7 shows the received signal in a LOS situation (as observed on a spectrum analyzer), and the reader can observe that there is less than 1 dB variation in the signal power. Figure 8 shows the received signal after it has propagated though two lines of trees with heavy foliage (over a one minute interval), and it can be seen that the received signal strength varies by as much as 20 dB. The type of fading that takes place in FBWA systems occurs due to scattering through leaves, and takes place over relatively large time intervals, for example a fade may easily last for hundreds of milliseconds. A photograph of the site where this measurement was taken is included in Figure 13. Deep fades of this sort are extremely detrimental to the operation of FBWA systems, unless preventive counter-measures are taken.

In order to enable OLOS/NLOS operation, the wireless link should be able to continue operating in the presence of deep fades. Next generation FBWA systems incorporate receive diversity techniques to combat deep fades. This involves using dual antenna at the subscriber site, which are spaced one wavelength apart. The signal received at Antenna 1 is very highly un-correlated from that received at Antenna 2, as shown in Figure 8. When Antenna 1 is in deep fade, then the receiver detects this condition, and switches over to Antenna 2, and thus is able to continue operation.

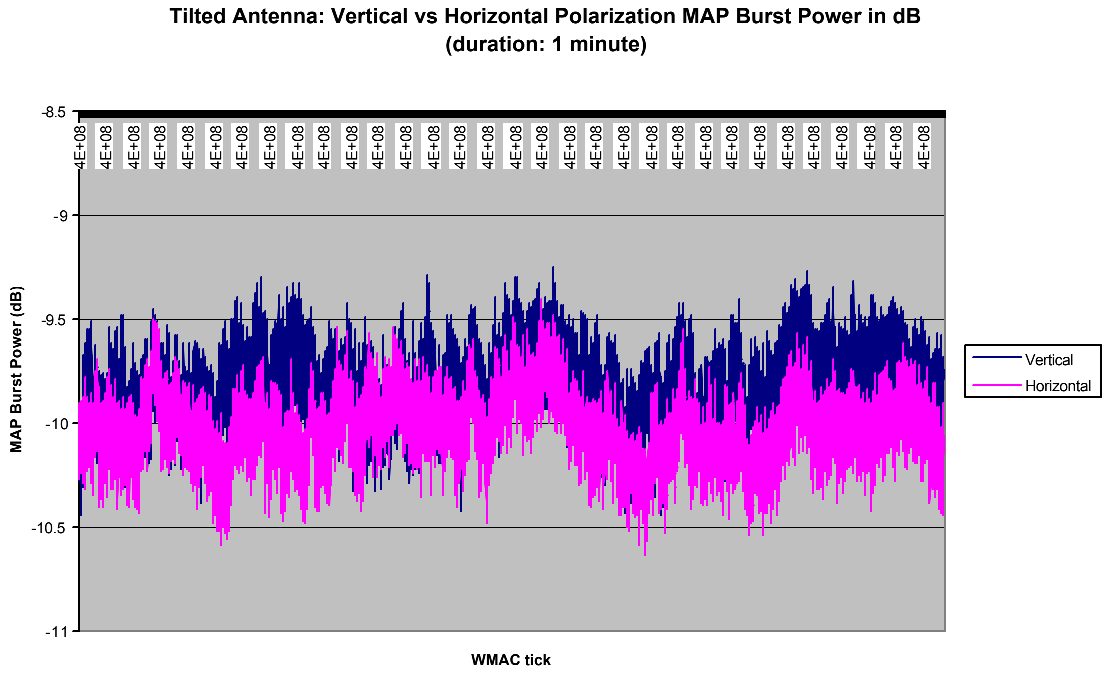

Figure 7: Received signal power at LOS subscriber site

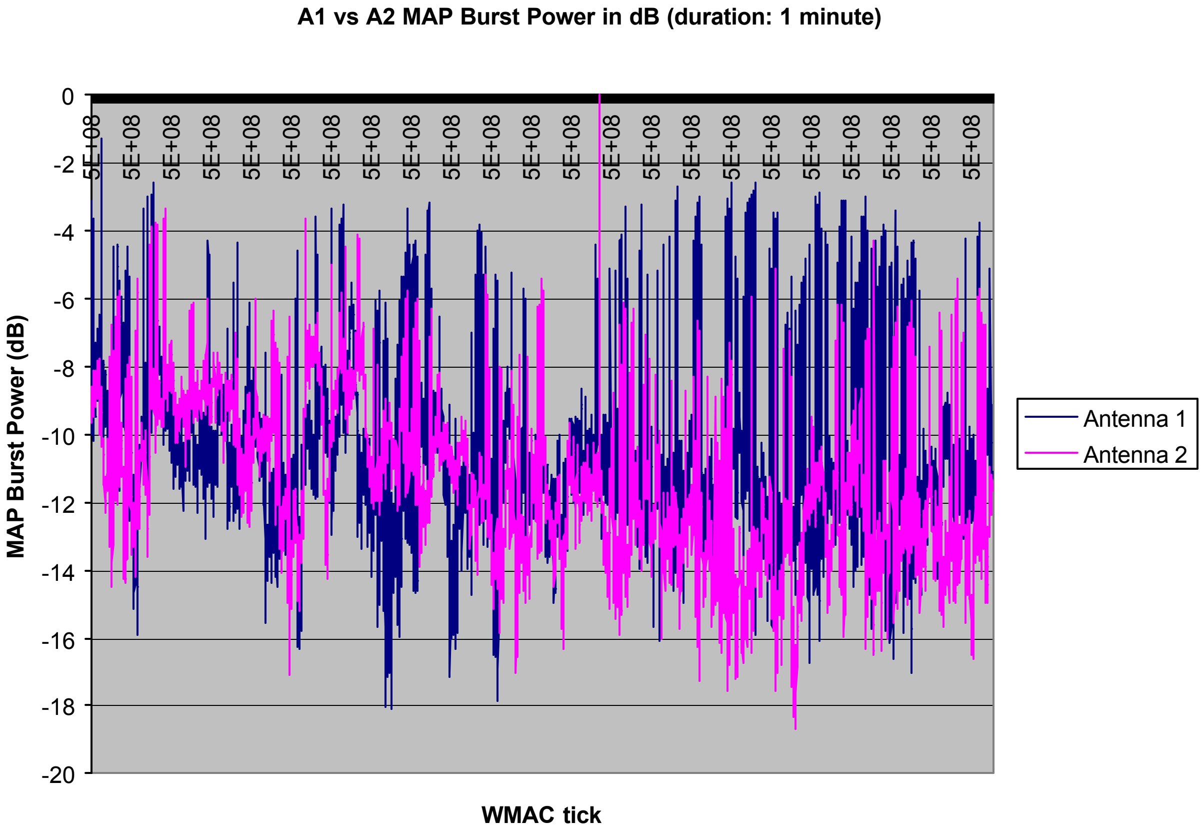

Figure 8: Received signal power at OLOS/NLOS subscriber site

Field data that we have collected show that in cases of intense fading, ARQ and adaptive coding and modulation are not sufficient by themselves to keep the system operational. Antenna diversity support is an essential element to ensure acceptable performance under these conditions. Our results show that simple switched diversity using a two-element antenna is sufficient, in contrast to more sophisticated (and expensive) techniques that have been proposed in recent years.

5.0 QoS Provisioning

Next generation FBWA systems are being designed to provide sophisticated QoS capabilities to the end users and the service providers. All traffic is classified into service flows and a scheduling mechanism operating at the MAC layer is responsible for assigning bandwidth to these flows according to their class of service and their QoS requirements. Aperto's system supports three classes of service:

- Constant Bit Rate (CBR) suitable for voice and video applications
- Committed Information Rate (CIR) most appropriate for businesses and high-end residential users that require a minimum service guarantee
- Best Effort (BE), for users that do not require any minimum bandwidth guarantee (typical applications include file transfer, browsing, e-mail etc).

CBR and CIR flows are subject to admission control and once a flow is admitted it is guaranteed a QoS for its duration. The available link bandwidth is partitioned between CBR, CIR and BE classes. If the bandwidth allocated for a particular service class is exhausted, then newly arriving flows of this class will be temporarily denied service of the particular class and will be treated as Best Effort flows until bandwidth is released. All bandwidth reservations are done strictly on demand, i.e., when traffic from the flow is present, and the reserved BW is released as soon as the traffic ends. This allows the service provider to achieve a very high degree of statistical multiplexing and thus realize more revenue. The QoS support in the Aperto product works seamlessly with the adaptive coding and modulation scheme, such that the bandwidth guarantees continue to be met even when the modulation or coding for the link changes.

In this section we describe a set of experiments that illustrate the provisioning of CIR service and its performance in the presence of Best Effort traffic. Our experimental setup consists of four subscriber units and the BSC. We have configured three of the CPEs as CIR subscribers and the fourth CPE is used to run background BE traffic. We assume that the total bandwidth is allocated between CIR and BE (60% CIR and 40% BE).

For the first experiment we reserve 20% of the total bandwidth for each CIR CPE; for simplicity we refer to the three CIR CPEs as "CIR1", "CIR2" and "CIR3". We have considered FTP as our source traffic and focus on the following three scenarios:

- Scenario 1: CIR1, CIR2, CIR3 active; BE idle.
- Scenario 2: CIR1, CIR2, CIR3, BE active
- Scenario 3: CIR1, CIR2, BE active; CIR3 idle.

Figure 9 plots the measured throughput (normalized to reflect the portion of the bandwidth used) for each CPE for all three scenarios described above. During scenario 1, CIR users get about 10% more than their committed rate, since no BE traffic is running. This shows that bandwidth allocation is dynamic and no bandwidth is wasted if certain flows are idle. As soon as BE traffic becomes active, CIR users are restricted to the 60%

of the CIR allocated bandwidth and they all get approximately a 20% proportion. Note that the actual number may be slightly less than 20% due to TCP window limitations and IP overhead. BE traffic at the same time occupies roughly 40% of the BW. Finally, during scenario 3, CIR3 remains idle and the excess 20% of CIR bandwidth is allocated to the CIR flows in proportion of their committed rates (which happen to be equal here).

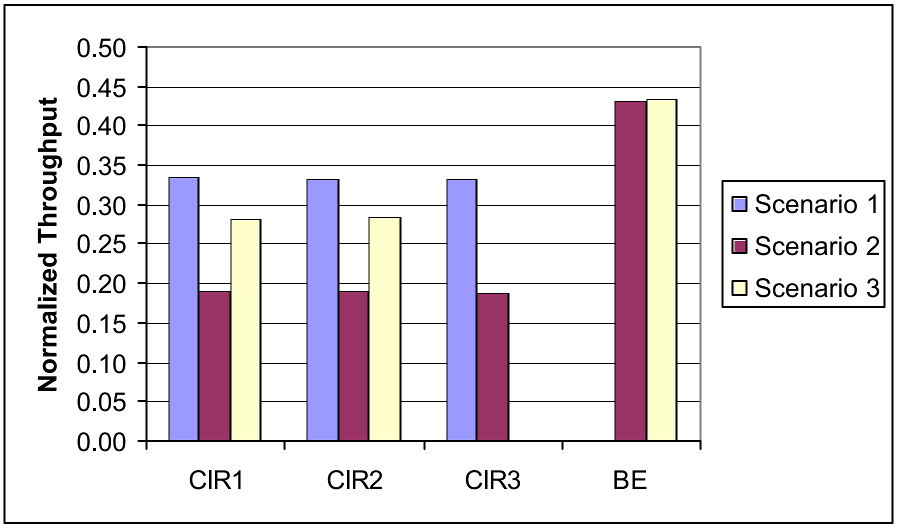

Figure 9: Bandwidth allocations between CIR and BE for experiment 1

In the next experiment we have changed the CIR reservations to 10, 20 and 30% of the total bandwidth for CIR1, CIR2 and CIR3 respectively. The fourth CPE is still running BE traffic which if active occupies up to 40% of the available bandwidth. In a similar fashion to experiment 1, we use FTP as our source of traffic and consider the following three scenarios:

- Scenario 1: CIR1, CIR2, CIR3 active; BE idle.
- Scenario 2: CIR1, CIR2, CIR3, BE active.
- Scenario 3: CIR1, CIR2 active; CIR3, BE idle.

Figure 10 plots the normalized throughput for all three scenarios. For scenario 1, since BE is idle the CIR flows benefit of the available bandwidth and experience throughput higher than their committed rates. The excess bandwidth is allocated in proportion to their CIR rates. During scenario 2, BE becomes active and occupies 40% of the bandwidth; the CIR users get their committed rate, which is 10, 20 and 30% respectively. Finally, during scenario 3, CIR 3 and BE become idle and the excess bandwidth is allocated to the remaining active CIR users.

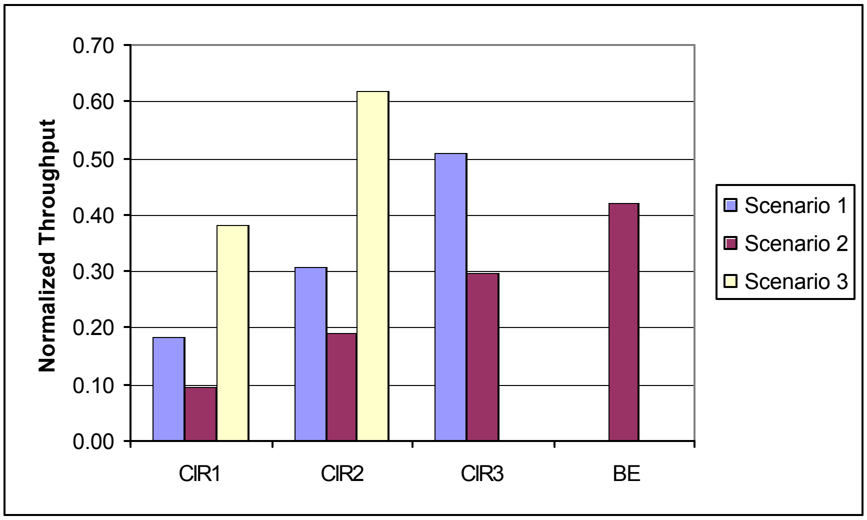

Figure 10: Bandwidth allocations between CIR and BE for experiment 2

6.0 Conclusions

In this paper we have presented measurements taken from a functioning next generation FBWA system, which is currently being shipped by Aperto Networks. These results illustrate some powerful new features that are now available in FBWA products. These features are enabling new capabilities such as dense cellular deployments, OLOS/NLOS operation, BW management support etc., which are crucial in making FBWA truly ubiquitous. The results presented in this paper also show that Adaptive TDMA technology is very well capable of solving the problems associated with FBWA, and hence specifications that are being written with an Adaptive TDMA base, such as IEEE 802.16ab and ETSI HIPERMAN, are on the right track.

A typical installation of the Aperto CPE, using the 3.5 GHz band product, is shown in Figure 11, with a view facing the base station tower in Figure 12 (the CPE antenna is the white panel in the balcony of the first floor home). This site is at a distance of about 3.5 miles from the base station tower (which is about 80 feet high), and the signal propagates over typical Silicon Valley suburban neighborhoods before reaching the CPE. We believe that installations such as these are key to increasing the penetration of FBWA, since the subscriber can carry them out without requiring a service provider truck-roll. There is an alternative vision of indoor CPEs that is being promoted by a section of the industry. We do not believe that this is the right approach to the problem, for the following reasons:

- The technology uses CDMA, which has well known limitations in supporting bursty packet traffic. The per-subscriber bit rates have to be reduced in order to gain the benefits of CDMA, which is a fundamental limitation.
- Signal propagation indoors is blocked if aluminum siding or chicken wire based stucco is used to build the home.
- The cell radius is much smaller, which leads to a much larger number of base stations to cover the same area. The resulting increase in infrastructure costs is enough to negate any cost savings realized due to indoor installation.

There is another section of the industry that has pinned its hopes on modulation schemes such as Orthogonal Frequency Division Multiplexing (OFDM) to get around LOS problems. Field results from the first set of OFDM deployments, that are beginning to become available, show that OFDM by itself is not sufficient. In spite of the larger delay spread tolerance, obstructions such as trees are able to block reception. A large measure for this failure is accounted for by the fact that several of these products used the DOCSIS MAC, which was not designed for wireless environments.

The Aperto approach has been to adopt a systems based design methodology, in which equal attention is paid to the MAC, PHY and Radio technologies, without over-burdening any particular layer. This has succeeded in keeping the system costs down, and has led to a product that solves the important problems of wireless deployment and is available today.

REFERENCES

[1] IEEE 802.16ab Working group on Broadband Wireless Access Standards http://ieee802.org/16/
[2] European Telecommunications Standard Institute (ETSI) HIPERMAN Project http://www.etsi.org/
[3] M. Zorzi, R.Rao and L.B. Milstein. ARQ Error Control for Fading Mobile Radio Channels. *IEEE Transactions on Vehicular Technology, vol. 46, no.2, pp. 445-455, May 1997.*
[4] T.V. Lakshman and U. Madhow. Performance Analysis of Window-based Flow Control Using TCP/IP: the Effect of High Bandwidth-Delay Products and Random Loss. *IEEE Transactions C-26, High Performance Networking V*, North Holland, 1994, pp.135-150.

APPENDIX A: Pictures from CPE installations

Figure 11

Figure 12

Figure 13