Low On Air: Inherent Wireless Channel Capacity Limitations

Paul Schmitt
University of California, Santa Barbara
pschmitt@cs.ucsb.edu

Elizabeth Belding
University of California, Santa Barbara
ebelding@cs.ucsb.edu

ABSTRACT

Wireless connectivity has fundamentally changed the way we connect and interact with the world. Over the past fifteen years there has been an exponential increase in wireless data usage, a trend that is predicted to continue. The overall capacity for wireless connectivity is limited in that it operates over electromagnetic spectrum, and the usable range of spectrum is both finite and already scarce. We argue that the growth in demand that we currently see is unsustainable in the long-term, as spectrum resources will become fully exhausted. While current lines of research seek to increase spectrum efficiency, increases in the future will achieve diminishing returns. In this work we present current technologies as well as cutting-edge research related to maximizing the efficiency of wireless systems, and offer research questions that will become critical as we near the limits of wireless connectivity.

1. INTRODUCTION

It is difficult to overstate the profound impact that wireless data communication has had on the way we connect and interact with the world around us. Users now expect always-available, high-quality connectivity in virtually any location, something that would have been seemingly impossible just a few decades ago. The shift in connectivity availability and the applications that now operate on mobile devices has manifested in dramatic, exponential increases in data consumption over wireless networks, a trend that appears likely to continue for the foreseeable future. Any system that faces exponential growth in consumption of a resource requires a corresponding exponential increase in the availability of the resource itself. Unfortunately, the medium that wireless communication operates on, electromagnetic spectrum, is finite and includes fundamental capacity limitations related to the channel bandwidth and quality. In this work we explore the variables that impact wireless channel capacity, advances that have been achieved to increase usage efficiency, and discuss the long-term challenges facing wireless connectivity.

For brevity, we focus on growth related to cellular data usage and corresponding growth in access link speeds that have been achieved in the past few decades. We examine the technology advances that have thus far enabled access link speeds to maintain pace with exponential growth in usage. We also study the current lines of research in the field that are needed in order to deliver the next generation of access link speeds.

Unfortunately, there appears to be scant room for substantial spectral efficiency increases beyond modern, efficient systems such as LTE and MIMO as these technologies operate near the underlying fundamental capacity limit. We believe that, as with other physical limitation scenarios (e.g. non-renewable resources), wireless link speed increases will slow and begin to cost more than is justifiable as we near the fundamental limits of channel capacity. Resultingly, assuming continued exponential growth in usage, we will fully exhaust all of the available wireless spectrum at a particular time and place in the future.

In this work we offer our vision for wireless connectivity in the near and long-term, and we argue that indefinite exponential increase in link capacities are unsustainable. In the medium-term, foundational changes in the ways that spectrum is allocated and shared will become critical in order to meet demand. In the long-term, we ultimately do not know what the reality of spectrum exhaustion will be. This paper is an attempt to open the discussion for wireless networking systems researchers to take a long horizon view of the field, and begin to consider the limited nature of wireless connectivity.

LIMITS '17, June 22-24, 2017, Santa Barbara, CA, USA
© 2017 ACM. ISBN 978-1-4503-4950-5/17/06...$15.00
DOI: http://dx.doi.org/10.1145/3080556.3080558

2. BACKGROUND

Users in traditionally well-connected regions now anticipate high-speed wireless connectivity in almost any location, at any time. The evolution of wireless connectivity, as well as devices (e.g. smartphones, tablets, etc.) that are designed to take advantage of the available capacity, drives our expectations. However, wireless communication channels have fundamental capacity limits based on the channel bandwidth and quality. In this section we provide background concerning the drivers of wireless growth as well as the looming capacity challenges facing the field due to spectrum scarcity and limited overhead for large increases in system efficiency.

2.1 Mobile data growth

The unprecedented growth in mobile data network usage has been well-documented. Over the past fifteen years, there was a 400 million-fold increase in cellular data traffic [4]. Ericsson forecasts a compound annual growth rate (CAGR) of 45%, a rate that would result in doubling every 1.87 years, between 2016 and 2022, with smartphone traffic increasing by 10 times and total mobile traffic for all devices by 8 times [6].

What is driving such demand? It is at least partly attributable to simply more users connecting to the Internet. Networks continue to

	2015	2016	2017	2018	2019	2020	Compound Annual Growth Rate (CAGR)
Global							
Global speed: All handsets	2.0	2.4	3.1	3.9	5.1	6.5	26%
By Region							
Asia Pacific	2.4	3.6	4.6	5.7	7.0	8.6	29%
Latin America	1.5	1.9	2.5	3.1	3.9	4.9	27%
North America	5.9	7.9	9.9	12.1	13.7	15.3	21%
Western Europe	4.1	6.1	8.3	10.5	12.2	14.1	28%
Central and Eastern Europe	2.3	3.4	5.6	7.8	9.1	10.6	36%
Middle East and Africa	0.8	1.3	1.9	2.6	3.6	4.8	45%

Table 1: Average Projected Mobile Network Connection Speeds (Mbps) [3].

add users, with particularly high growth in developing regions, who in-turn consume more data resources. Of course, we anticipate the trend of adding users will begin to slow as eventually everyone on the planet will be within coverage areas of wireless connectivity, at which point the increase in the number of users will likely follow global population growth trends. If data usage was in lockstep with the number of users, we may not reach spectrum exhaustion, or exhaustion may take hundreds of years. However, the applications that run on mobile devices have drastically increased their reliance and expectation of high-throughput connectivity as link capacities have grown. The applications and devices that represent the largest consumers of mobile bandwidth are diverse [12]. The overall trend toward high-quality multimedia such as streaming video represents perhaps the largest challenge for networks, as multimedia typically requires high-throughput connectivity with quality of service (e.g. latency) guarantees. Likewise, smartphones are increasingly used to deliver virtual or augmented reality environments, technologies that often require enormous data throughput to deliver real-time video streams.

It can be argued that exponential data growth will not necessarily continue unfettered, as the human brain itself has throughput limitations [17]. If humans are the only users of the system and screen sizes and densities remain relatively stable, there would be little sense in providing more information (i.e. higher resolution video streams) than is actually perceivable. However, humans are not the sole users of wireless networks. Machine-to-machine (M2M) communication has quickly grown to become a major user of networks and is expected to increase to 45% of all Internet traffic by 2022 [8]. While M2M typically has lower throughput and quality of service (QoS) needs compared with user-originated traffic, the sheer volume of data associated with M2M will require wireless networks to provision appropriately moving forward. Ultimately, wireless data growth is expected to continue increasing at exponential rates, driving industry and researchers to design wireless access link technologies that are able to deliver ever-higher throughput to users.

2.2 Mobile connectivity growth

In order to meet the demand placed on wireless networks, mobile access link speeds must increase accordingly. Up to this point, industry and researchers have found ways to increase access capacity. As shown in Table 1, Cisco expects the average access mobile connectivity speeds to increase globally by a CAGR of roughly 26% in the near future. Given a CAGR of 26%, connectivity speeds will double every 3 years. When discussing 5G, the forthcoming generation of cellular technologies, researchers and industry often discuss increasing speeds by $1000\times$. While such a jump in capacity would appear on its face to provide "enough" capacity for a very long time, a CAGR of 26% means that we would fully consume a thousand-fold increase in roughly 30 years. Exponential growth is not unique to mobile data; it has been observed in traditional broadband connectivity for many years. Nielson's law [18] states that traditional wired broadband speeds have a 50% compound annual growth rate, and has proven to be accurate for more than 30 years.

Mobile data growth and access link speeds are components in a positive feedback loop. Link capacities are increased and new, more demanding applications are developed that take advantage of the increased link speeds. In turn, link capacities become consumed, and so on. This positive feedback loop makes it difficult to assign responsibility for growth. Is usage growth a response to capacity growth, or does capacity grow in response to usage? Perhaps the two drive each other symbiotically. Unfortunately for wireless technologies, the capacity of the wireless medium itself is inherently limited, whereas it does not appear that usage growth will be for the foreseeable future. The wireless medium is itself unique and provides different challenges than are found with wired networking. We explore the reasons behind this in the following sections.

2.3 Wireless channel capacity

Wireless demand forces us to design systems that offer ever-higher capacity. However, wireless capacity is not infinite. Shannon's law states that the error-free capacity of any communications channel is a function of the signal bandwidth, received signal power, and noise [24], as shown in Equation 1, where C is the theoretical maximum capacity of a channel in bits per second, B is the signal bandwidth in hertz, and $\frac{S}{N}$ is the signal-to-noise ratio.

$$C = B \log_2 \left(1 + \frac{S}{N}\right) \tag{1}$$

What Shannon's law tells us is that we have relatively few knobs available to turn in order to increase the capacity of a given channel. Wireless spectrum that is usable for communications is finite and shared by all users in a given location, therefore we are limited in terms of the amount of bandwidth we can assign for a given channel. The other variable that we can attempt to control is the signal-to-noise ratio (SNR), as it is a major limiting factor in channel capacity. A naïve solution follows that we should simply increase the signal power in order to increase the SNR. However, such a solution proves impossible in reality due to a host of associated problems (e.g. power concerns, inter-cell interference, etc.).

Another major challenge currently facing wireless researchers is that most modern access technologies already approach the limit defined by Shannon. Even legacy technologies such as 1xEV-DO, HSDPA, and WiMAX are within roughly 2 or 3 decibels of the Shannon limit [7]. Likewise, LTE uses a highly efficient physical layer implementation that operates near this limit. This leaves little room for improvement for future generations of technology. We revisit this problem and contemporary solutions (e.g. MIMO, MU-MIMO, coordinated multipoint transmissions) for increasing

				Max. link spectral efficiency ((bits/s)/Hz)			
Service	**Standard**	**Max. net bitrate per carrier per one spatial stream(Mbit/s)**	**Bandwidth per carrier(MHz)**	**SISO**	**MIMO**	**Typical reuse factor1/K**	**System spectral efficiency (R/B)/K ((bit/s)/Hz per site)**
2G	GSM	0.013×8 timeslots = 0.104	0.2	0.52	N/A	1/9	0.17
2.75G	GSM + EDGE	0.384	0.2	1.92	N/A	1/3	0.33
3G	WCDMA FDD	0.384	5	0.077	N/A	1	0.51
3G	CDMA2000 1xEVDO Rev.A	3.072	1.2288	2.5	N/A	1	1.3
3.5G	HSDPA	21.1	5	4.22	N/A	1	4.22
4G	LTE	81.6	20	4.08	16.32 (4x4)	1	16.32
4G	LTE-Advanced	75	20	3.75	30.00 (8x8)	1	30

Table 2: Cellular technology spectral efficiencies [26].

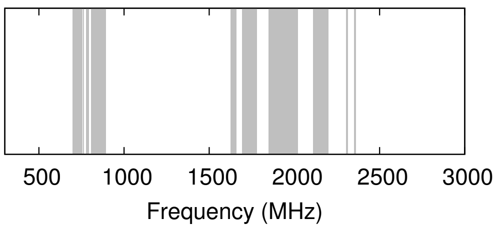

Figure 1: UHF cellular frequencies in the United States. Shaded regions indicate allocated spectrum.

wireless channel capacity in § 3.2, but the foundational problem remains that any wireless channel has a fundamental capacity limit for the amount of information that can be transmitted, and current generation technologies are near enough to the limit that we will obtain diminishing returns as we employ more sophisticated and expensive techniques to close the gap.

2.4 Spectrum allocation

Given what we know about channel capacity, we seek to identify performance that we can expect given the limited wireless spectrum. Electromagnetic spectrum is the range of all known frequencies and their related wavelengths, ranging from wavelengths near the Planck length on the short-end and the size of the universe on the long-end. For simplicity, we focus on cellular spectrum allocation in the United States[1]. Until recently, cellular spectrum has largely been confined in the ultra high frequency (UHF) band of the radio range of the spectrum. UHF is defined as frequencies between 300 MHz and 3,000 MHz. Figure 1 displays the UHF band and the frequency ranges that have been allocated for cellular usage by the FCC. As shown, only a fraction of the UHF frequency space has been allocated for cellular communications, with large amounts of spectrum set aside for other technologies such as broadcast television, radio navigation, and military use. The current cellular spectrum allocation in the United States totals slightly more than 560 MHz. However, that number is misleading, as the allocated spectrum is not contiguous. There are 14 different contiguous regions, ranging from 5 MHz to 145 MHz of contiguous spectrum.

To illustrate the current challenge, we can perform back-of-the-envelope calculations. In this idealized scenario, let us imagine we have a single carrier that has exclusive rights to all of the currently allocated U.S. cellular spectrum. We assume the use of 5 MHz HSDPA channels, as HSDPA exhibits high spectral efficiency and operates using 5 MHz channels, which will fit within the smallest contiguous cellular region. In this situation we would be able to have 110 unique 5 MHz HSDPA channels available. A 5MHz HSDPA channel is able to offer roughly 21 Mbps to share amongst connected users. Therefore, using all of the 110 channels available the HSDPA capacity would be roughly 2.3 Gbps. Clearly, this capacity would not do even today, let alone as data demand rises in a sufficiently dense user environment. One obvious solution for the problem is increasing the amount of spectrum that we can use for mobile connectivity. We discuss this in § 3.1. Additionally, spectrum is reusable in the spatial domain and with modern interference mitigation techniques, which we explore in § 3.2.

[1]Note that the U.S. does not use all of the bands specified by 3GPP [1].

2.5 Spectral efficiency

Because the portion of RF spectrum we use for mobile communications is a scarce, finite resource, it is imperative to utilize spectrum as efficiently as possible. In the cellular domain, "spectral efficiency" is often used to characterize different systems and technologies. It can be understood as the information rate that can be transmitted over a channel with a given bandwidth using a particular physical layer protocol, and is normally expressed in $(bits/sec)/hertz$ [11]. Table 2 shows spectral efficiency values for different generations of cellular technology. For example, 3G HSDPA has a maximum spectral efficiency of 4.22 bps / Hz, while LTE with 4×4 MIMO (multiple-input, multiple-output) can reach 16.32 bps / Hz in an ideal scenario. Researchers view LTE as highly efficient, which means achieving drastic increases in spectral efficiency will prove to be difficult in practice [13]. Yet, given the exponential rise in demand (assuming a 26% CAGR), a 10-fold increase will be *required* in roughly ten years.

2.6 Wireless limit consequences

We know that wireless throughput demands are increasing at an exponential rate, and that wireless link access capacities are governed by fundamental limits. Given this, how do we foresee wireless communication as we eventually reach the limits of the medium?

Fortunately, wireless spectrum is very different compared with material resources. Whereas exponential consumption of non-renewable resources, such as mineral ore found in the Earth's crust, may lead to overshoot or catastrophic collapse [16], spectrum is unique in that it is instantaneously renewable and impossible to overshoot in terms of consumption. Our use of wireless technologies does not reduce the amount of usable spectrum for future users. This provides us the opportunity to reconsider, and drastically alter, how spectrum is used for communication at any time with benefits carried forward from that point on. Spectrum is also spatially-reusable; therefore, complete consumption is only likely to occur in densely connected areas (e.g. cities). Accordingly, we do not anticipate wireless capacity exhaustion in rural areas in the near or medium-term, as it has previously been found that spectrum in rural areas is widely available compared with urban locations [21]. Complete capacity exhaustion will first occur in user-dense, urban areas at peak usage times.

It is our belief that volatile collapse is unlikely. However, as capacities near the limits defined by Shannon's law, gains will become more difficult and costly to realize, resulting in a sigmoidal approach to the capacity limit rather than overshoot. This will result in an analogue with the discussion of non-renewable resource consumption. Different areas of spectrum are "more rich" in resources than others (i.e. they offer more capacity and desirable propagation characteristics). Just as with copper ore mining, the "better" quality portions of spectrum are the first we use for wireless communication, and as we move forward the ranges of available spectrum will be less desirable and potentially more costly to utilize. We believe that, in a spectrum exhaustion scenario, it will be necessary to match application needs and subsequent spectrum usage in order to use "less rich" spectrum where we can, while conserving desirable spectrum for applications that rely upon it. We discuss this in § 4.1.

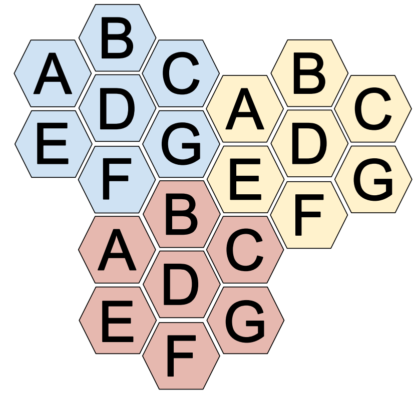

Figure 2: Cellular network comprises multiple basestations grouped into clusters. Cell colors denote clusters. Letters denote frequency. Adjacent cells avoid interference by utilizing separate frequencies in legacy cellular systems.

3. MAXIMIZING WIRELESS CAPACITY

What can we do to increase wireless access link capacities as demand skyrockets? It is commonly accepted that in the next few years we, researchers and industry, must develop a fifth generation (5G) of mobile network to meet imminent latency and throughput demands [19]. Going back to Shannon's law, there are two variables that we can attack: bandwidth and SNR. In fact, there is a third variable we have not yet discussed: antenna count. In this section we explore modern advances in access link technologies and difficulties researchers and industry faces as we attempt to meet next-generation spectral efficiency goals.

It is important to note that we do not believe that wireless capacity limits will be reached in the immediate future. It may take decades, as new breakthroughs are constantly increasing the spectral efficiency of wireless systems. However, we do foresee a future in which we near or reach the limit.

3.1 Bandwidth

To increase capacity of a channel, we can increase bandwidth. From a system-level perspective, if we are able to use more spectrum overall (i.e. spectrum allocation), or more effectively use the spectrum we already have available (i.e. spatial reuse), we can increase network capacity.

3.1.1 Spectrum allocation

As discussed in § 2.4, cellular systems have restricted their usage to the UHF band, which offers relatively small amounts of frequency range. In the search for more spectrum, researchers have recently tabbed higher frequencies as an area for exploration [22]. In the past, higher frequencies have been viewed as poor choices for wireless communications as they typically have poor propagation characteristics related to path loss, rain fade, and strict line-of-sight requirements [23]. However, further exploration has revealed that in small-cell, dense, urban environments, millimeter-wavelength (e.g. 60 GHz) wireless channels can offer significant channel capacities. The challenges facing this line of research are related to the propagation characteristics. Link loss with such systems is much more likely due to blocking caused by physical objects and oxygen absorption. Accordingly wireless networks must be redesigned in order to manage a higher probability of ephemeral connectivity loss. We envision millimeter-wave, and higher frequency use in general, to be very promising in terms of capacity gains, particularly for indoor environments. However, it is not a panacea; exponential growth will eventually demand even more than this technology can offer.

Regulators have also recognized the need for additional spectrum in order to meet capacity goals. Advances in software-defined radio technology, which allows for agile use of frequency spectrum, have led to new shared spectrum licensing and occupancy models in recent years. Essentially, the new models allow for "secondary" users, those that may not have exclusive license to operate over a specified frequency in a location, to utilize idle spectrum when and where incumbents (i.e. license holders) are not operating. Much work in this space has focused on "TV white space" frequencies made available by the digitization of broadcast television as those UHF frequencies have favorable propagation characteristics. Researchers have conducted trials of the spectrum sharing models and shown that they can be successful [20]. In addition to new models, the FCC in the United States released nearly 11 GHz of high-frequency (>24 GHz) spectrum intended to aid in reaching 5G capacity goals in 2016 [5]. While these developments are certainly welcome and will have a large impact on wireless capacities in the near-term, spectrum remains finite and exponential growth in usage will eventually exhaust the additional resources.

3.1.2 Spatial reuse

When we discuss spectral efficiency in cellular networks, an additional dimension is added for the "system" spectral efficiency, which allows us to measure the capacity of a system to serve end-users (typically per sector antenna, or cell). This value is governed by the frequency reuse factor.

Cellular networks are based around the concept of cells [15], each of which transmit signals on some portion of wireless spectrum. A common depiction of cellular networks is shown in Figure 2, where clusters of 7 cells, each operating at a unique frequency are grouped. In order to avoid interference, adjacent cells must use different frequencies. Reuse is limited by cell range (i.e. signal power). The frequency reuse factor is denoted as $1/K$, where K is the number of cells that cannot use the same frequencies. In Figure 2, K is 7, resulting in a frequency reuse factor of this system of 1/7. Common values for legacy cellular technologies are 1/3, 1/4, 1/7, 1/9, 1/12. The system spectral efficiency value for an area is the calculated spectral efficiency multiplied by the reuse factor. Accordingly, high frequency reuse greatly impacts the system spectral efficiency for an area. Table 2 displays the system spectral efficiency of various cellular technologies. As shown, modern cellular technologies such as LTE and 3G have achieved frequency reuse factors of 1, meaning adjacent cells are able to use the same frequencies. Such performance is enabled through the use of coordinated interference mitigation and code division multiplexing. However, there remain limitations to such techniques.

We can also increase spatial reuse by employing cells with smaller coverage areas than legacy cellular designs traditionally offer [13]. With small cells (e.g. femtocells, picocells), frequency reuse can be increased because inter-cell interference is reduced, therefore system capacity will rise. Likewise, long-distance wireless links tend to have poorer line-of-sight and path loss problems, leading to lower connectivity speeds for faraway users. With small cells, line-of-sight between users and basestations is more likely, leading to higher speeds. Small cells will be heavily relied upon to increase efficiency moving forward; however, we cannot simply decrease cell size indefinitely. As we densify the wireless network, overhead in the form of control traffic is increased, as the network must manage user mobility and small cells can share spectrum with macro cells, which requires coordinating time-frequency use. Likewise, inter-cell interference will be a limiter as we introduce more and more adjacent basestations.

3.2 SNR and CoMP

Naïvely, the desire to increase SNR can be focused on two areas: mitigating noise and interference, or increasing the signal strength. Unfortunately, increasing signal strength is an unlikely avenue for improvement, as mobile devices are often energy-constrained (i.e. battery-powered), and inter-cell interference can severely limit SNR in dense networks, where transmitting with higher signal strength would only serve to reduce SNR for adjacent cells. Therefore, we focus on mitigating noise and interference.

A fundamental nature of wireless signal propagation is that signals within a cell are not perfectly confined to the cell's intended coverage area (i.e. signals leak into adjacent cells). The resulting inter-cell interference lowers SNR for user devices, thus limiting channel capacity. Recently, the cellular industry has pushed for an updated LTE, LTE-Advanced, which allows for coordinated multipoint (CoMP). CoMP *leverages* inter-cell signal leakage, rather than attempting to avoid it altogether. Essentially, a user device can exchange data with multiple nearby basestations simultaneously. In a sense, CoMP can be viewed as multi-cell MIMO, where the additional antennas are spread across multiple physical basestations from the network point of view. CoMP is made possible by the fact that basestation infrastructure is immobile, leading to relatively stable channel state. We discuss MIMO in § 3.3.

CoMP requires significant coordination between neighboring cells in a cluster. Control traffic must ensure synchronization between basestations and up-to-date channel information. As a result, high-capacity backhauls are required to support a CoMP cluster. Depending on the number of cells in the cluster and the network design, backhaul requirements can quickly reach *tens of Gbps* for centralized networks or *thousands of Gbps* for distributed network designs [9]. Accordingly, the backhaul requirements prevent CoMP from realizing MIMO-like linear capacity increases.

3.3 Additional antennas

Shannon's law can include one additional variable that we have not yet considered: additional antennas. This is attributable to MIMO technologies, where multiple antennas at the sender and receiver transmit different signals across the same wireless channel, exploiting multipath differences in signal reception. In an ideal situation, MIMO can essentially increase the channel capacity linearly as the number of antennas used increases. This is shown in Equation 2.

$$C = B \cdot a \cdot \log_2 \left(1 + \frac{S}{N} \right) \tag{2}$$

MIMO has been widely explored in recent years, and has led to significant increases in spectral efficiency for wireless link technologies. For example, Table 2 shows that LTE-Advanced using 8x8 MIMO can theoretically realize an 8-fold increase in spectral efficiency, up to 30 bps / Hz. Unfortunately, while MIMO has greatly improved spectral efficiency, it does not often approach the touted theoretical capacities for a few key reasons. First, maximum theoretical efficiency gains are typically calculated for situations where there is a single user, which is not a common occurrence for cellular systems. Further, in order for MIMO to reach the maximum spectral efficiency, the sender and receiver both must have *perfect instantaneous knowledge* of the channel state information. In reality, such a scenario simply does not exist [10], and capacity gains can vary greatly depending on SNR. In some cases, the cost and complexity of additional antennas introduced by MIMO are not justified by marginal gains.

Multi-user MIMO (MU-MIMO) is an extension of the MIMO concept, where multiple antennas at the basestation can simultaneously send signals to multiple receivers, rather than multiple antennas at a single receiver. MU-MIMO does not fundamentally change the capacity gains, it only spreads them across multiple users. For MU-MIMO to increase the overall capacity across multiple users, the basestation must increase the number of antennas proportionally to the number of users in the cell (i.e. eight basestation antennas to send signals to two clients with four antennas each). As cellular basestations often serve dozens of users, the number of antennas needed at a basestation could quickly become unruly.

MIMO gains are impressive; but are, even in the best case, linear with the number of antennas at the sender and receiver. If MIMO is to be leveraged as the answer to exponential growth in demand, it stands to reason that the number of antennas per device must exponentially grow. Great increase in MIMO antennas has been termed "masive MIMO" [14]. As the number of antennas grows, power consumption, synchronization with users, and combining low-precision components in an effective way is a challenge. Additionally, physical size may become untenable as neighboring antennas must be sufficiently diverse from each other so as to differentiate signals intended for each antenna. We believe that MIMO will continue to provide many gains in capacity, but it is unlikely to be an infinitely scalable solution.

4. RESEARCH AGENDA

Eventually, we will reach the capacity limits for wireless communications, particularly in dense, urban network environments. Of course, it would be myopic to assume that current generation technologies have essentially reached the limits defined by Shannon's Law, and there is nowhere to improve. We will continue to achieve impressive access link speed growth in the near-term. However, we believe that large capacity increases will become more and more difficult to attain. As networking researchers, it is our responsibility to continue to pursue technologies that maximize connectivity with the limited spectrum. In this section we provide an overview of related technical research that we believe will have the highest impact in the coming years. We also include a brief discussion of social questions related to capacity exhaustion. In each area, we include what we see as open questions.

4.1 Spectrum usage

Most current spectrum allocation and regulation occurs at scales that are broader than necessary (e.g. often nation-wide). Further, spectrum bands were largely allocated before the widespread use of wireless technologies, which has lead to large ranges of 'valuable' spectrum that are unavailable due to regulations. A complete re-allocation of all usable spectrum could provide a drastic increase in efficiency, as many frequency ranges have been set aside for technologies that either do not use the spectrum in many locations, or use it inefficiently. Unfortunately, a complete re-allocation is highly unlikely, as it would require the full cooperation of national bodies that manage spectrum (e.g. FCC), incumbent users that have invested vast financial resources into spectrum licenses (e.g. cellular providers), and incumbents that enjoy large areas of spectrum without competition (e.g. military).

On a smaller scale, (i.e. within already-allocated bands), we believe that more agile, intelligent use of spectrum can be achieved. For instance, spectrum range allocation based on application throughput need or disconnection tolerance could increase overall spectral efficiency. Agile, physical layer implementations, enabled by software-defined radio technologies, would also prove beneficial (e.g. different bandwidths depending on needs, CDMA versus OFDM in low-SNR environments). For example, low-throughput devices, such as many Internet of Things devices, could potentially be assigned low frequencies to achieve long-distance, low-throughput channels, or high frequencies for high-throughput, short-range, disruptable channels.

There are also opportunities to expand wireless connectivity outside of traditional RF spectrum. Free-space optical communications [2] allow us to operate wireless channels in light portions of the spectrum and can achieve extremely high throughputs over short ranges. Free space optics include some drawbacks, principally due to sensitivity to signal blockage by physical objects and receiver mobility. However, early work in this line of research has shown great potential for wireless connectivity.

Questions:

- *How can we achieve dynamic, distributed, heterogeneous spectrum usage that is fair to all parties?*
- *What cross-layer mechanisms must be built to intelligently select appropriate physical-layer implementations?*
- *What are the implications of mixed-spectrum systems capable of operating in both RF and optical spectrum?*

4.2 Signal coexistence

Current wireless technologies are often SISO (single-input single-output) and the chosen medium is often shared; users must take turns to communicate. We believe the next generation must allow for both coordinated and uncoordinated heterogeneous signal coexistence. A current example of coordinated coexistence would be multiuser MIMO, where a single basestation can send signals to multiple users simultaneously over a single channel. Beyond the benefits available through additional antennas, recent breakthroughs have been achieved by leveraging orbital angular momentum (OAM) multiplexing, a different physical layer modulation implementation where multiple signals are 'twisted' and bundled together. Using OAM, researchers have achieved terabit throughputs in optical spectrum [25]. Thus far, OAM has proven to be less-usable in RF spectrum; however, there is active work in this space.

Uncoordinated signal coexistence includes heterogeneous systems, even differing physical layer implementations, simultaneously sharing the same spectrum. A recent example is a system that allows for data communication using *occupied* UHF broadcast television bands [27]. Uncoordinated coexistence is also a goal of LTE-Advanced, where cellular channels are able to occupy unlicensed frequencies (e.g. WiFi frequency bands) and coexist with incumbents. Such coexistence introduces many research challenges, such as quality of service guarantees and interference mitigation, as many current media access control (MAC) layer implementations do not account for sharing the channel medium with different, non-cooperative peers. However, we believe this line of research has much potential for increasing the overall capacity for wireless communications as we near the limits of the medium.

Questions:

- *Can multiple physical layer implementations simultaneously coexist peacefully in the same time and frequency space to form multiple channels?*
- *Are there novel modulation schemes that would allow for much higher levels of signal coexistence in RF bands?*
- *Can programmatic interference mitigation techniques such as CoMP be accomplished in uncoordinated coexistence situations?*
- *What MAC layer mechanisms are necessary to enable distributed, heterogeneous coexistence?*

4.3 Non-technical considerations

This paper has focused on the technical limitations facing wireless networking. However, as we reach fundamental capacity limits, societal behaviors and expectations will almost certainly inform systems solutions. For instance, traditional medium access layer protocols and congestion control algorithms in network transport layers strive to ensure fairness, often defined as each of N network hosts receiving roughly $1/N$ of any shared resource such as bandwidth or time. In a capacity-limited scenario, such fairness assumptions will likely be questioned and new sharing algorithms, perhaps based on societal utility of communications may be necessary (i.e. machine-to-machine communication may receive lower-priority and less bandwidth than user-initiated traffic). As wireless networking researchers, we believe cross-disciplinary research and discussion is necessary in order to find meaningful solutions to the broader questions that arise as wireless resources are exhausted.

Questions:

- *What types of traffic are "more valuable," and therefore more worthy of spectrum consumption, during congestion events?*
- *How much is "enough" with regards to connectivity? Can it be defined?*
- *Should some types of traffic be entirely prevented from access during peak usage?*
- *In a distributed access environment, will behavior-based policies be self-enforced or must a centralized enforcement mechanism be introduced?*
- *What mechanisms to incentivize lower resource consumption can we explore?*

5. CONCLUSION

Demand for wireless capacity is unrelenting. As we move forward it will become increasingly difficult, and eventually impossible, to increase access link speeds to maintain pace. We must take drastic steps to overhaul wireless technologies and policies to maximize the use of finite spectrum in the near-term. We also must consider a future where spectrum exhaustion is probable, and begin designing workable, alternative wireless options for connectivity before we inevitably reach capacity limits.

6. ACKNOWLEDGMENTS

This work was funded through NSF Network Science and Engineering (NetSE) Award CNS-1064821.

7. REFERENCES

[1] 3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception. TS 36.101, Jan. 2017.

[2] V. W. S. Chan. Free-space optical communications. *Journal of Lightwave Technology*, 24(12):4750–4762, Dec 2006.

[3] Cisco. The Zettabyte Era - Trends and Analysis. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html, June 2016. Accessed: 2017-01-30.

[4] Cisco. White paper: Cisco VNI Forecast and Methodology, 2015-2020. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html, June 2016. Accessed: 2017-01-30.

[5] F. C. Commission. FCC adopts rules to facilitate next generation wireless technologies, 2016. Accessed: 2017-02-24.

[6] Ericsson. Ericsson mobility report: On the pulse of the networked society. Technical report, Nov. 2016.

[7] A. M. B. Explosion. The 3GPP wireless evolution. *Rysavy Research/4G Americas*, 2013.

[8] J. Gantz and D. Reinsel. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future*, 2007(2012):1–16, 2012.

[9] X. Ge, H. Cheng, M. Guizani, and T. Han. 5G wireless backhaul networks: challenges and research advances. *IEEE Network*, 28(6):6–11, Nov 2014.

[10] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of mimo channels. *IEEE Journal on selected areas in Communications*, 21(5):684–702, 2003.

[11] D. N. Hatfield. Measures of spectral efficiency in land mobile radio. *IEEE Transactions on Electromagnetic Compatibility*, EMC-19(3):266–268, Aug 1977.

[12] M. Hazas, J. Morley, O. Bates, and A. Friday. Are there limits to growth in data traffic?: on time use, data generation and speed. In *Limits 2016*, Irvine, CA, June 2016.

[13] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfrojd, and T. Svensson. The role of small cells, coordinated multipoint, and massive MIMO in 5G. *IEEE Communications Magazine*, 52(5):44–51, May 2014.

[14] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta. Massive MIMO for next generation wireless systems. *IEEE Communications Magazine*, 52(2):186–195, February 2014.

[15] V. H. Mac Donald. Advanced mobile phone service: The cellular concept. *Bell System Technical Journal*, 58(1):15–41, 1979.

[16] D. Meadows, J. Randers, and D. Meadows. *The limits to growth: the 30-year update*. Chelsea Green, 2004.

[17] G. A. Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.

[18] J. Nielsen. Nielsen's Law of Internet Bandwidth. https://www.nngroup.com/articles/law-of-bandwidth/. Accessed: 2017-01-30.

[19] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren. Scenarios for 5G mobile and wireless communications: the vision of the METIS project. *IEEE Communications Magazine*, 52(5):26–35, May 2014.

[20] M. Palola, M. Matinmikko, J. Prokkola, M. Mustonen, M. Heikkilä, T. Kippola, S. Yrjölä, V. Hartikainen, L. Tudose, A. Kivinen, J. Paavola, and K. Heiska. Live field trial of licensed shared access (LSA) concept using LTE network in 2.3 GHz band. In *Dynamic Spectrum Access Networks (DYSPAN)*, Mclean, VA, USA, Apr. 2014.

[21] V. Pejovic, D. L. Johnson, M. Zheleva, E. M. Belding, and A. Lysko. *VillageLink: A Channel Allocation Technique for Wide-Area White Space Networks*, pages 249–280. Springer International Publishing, Cham, 2015.

[22] Z. Pi and F. Khan. An introduction to millimeter-wave mobile broadband systems. *IEEE Communications Magazine*, 49(6):101–107, June 2011.

[23] S. Rangan, T. S. Rappaport, and E. Erkip. Millimeter-wave cellular wireless networks: Potentials and challenges. *Proceedings of the IEEE*, 102(3):366–385, March 2014.

[24] C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[25] J. Wang, J.-Y. Yang, I. M. Fazal, N. Ahmed, Y. Yan, H. Huang, Y. Ren, Y. Yue, S. Dolinar, M. Tur, et al. Terabit free-space data transmission employing orbital angular momentum multiplexing. *Nature Photonics*, 6(7):488–496, 2012.

[26] Wikipedia. Spectral efficiency — wikipedia, the free encyclopedia, 2016. Accessed: 2017-02-10.

[27] X. Zhang and E. W. Knightly. WATCH: WiFi in active TV channels. In *MobiHoc '15*, Hangzhou, China, 2015.