

Clarke, Richard N.

Conference Paper

Expanding mobile wireless capacity: The challenges presented by technology and economics

19th Biennial Conference of the International Telecommunications Society (ITS): "Moving Forward with Future Technologies: Opening a Platform for All", Bangkok, Thailand, 18th-21th November 2012

Provided in Cooperation with:

International Telecommunications Society (ITS)

Suggested Citation: Clarke, Richard N. (2012) : Expanding mobile wireless capacity: The challenges presented by technology and economics, 19th Biennial Conference of the International Telecommunications Society (ITS): "Moving Forward with Future Technologies: Opening a Platform for All", Bangkok, Thailand, 18th-21th November 2012, International Telecommunications Society (ITS), Calgary

This Version is available at:

<https://hdl.handle.net/10419/72529>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The 19th ITS Biennial Conference 2012

“Moving Forward with Future Technologies: Opening a Platform for All”

18 - 21 November 2012, Thailand

Expanding Mobile Wireless Capacity: The Challenges Presented by Technology and Economics

Richard N. Clarke

AT&T

rnclarke@att.com

**Expanding Mobile Wireless Capacity:
The Challenges Presented by Technology and Economics**

Richard N. Clarke*
AT&T – Global Public Policy

Abstract

As demand for mobile broadband services continues to explode, mobile wireless networks must expand greatly their capacities. This paper describes and quantifies the economic and technical challenges associated with “deepening” wireless networks to meet this exploding demand. Methods of capacity expansion divide into three general categories: the deployment of more radio spectrum; more intensive geographic reuse of spectrum; and increasing the throughput capacity of each MHz of spectrum within a given geographic area. The paper first provides a brief technical background on mobile wireless networks and these basic methods to deepen their capacity. It goes on to measure the contribution of each of these methods to historical capacity growth within U.S. networks. The paper then describes the capabilities of 4G LTE wireless technology, and further innovations off of it, to increase network capacity. The capacity expansion capability of LTE-Advanced along with traditional spectrum reuse are quantified, and are compared to forecasts of future demand to evaluate the ability of U.S. networks to match this future demand. We find that without significantly increased allocations of spectrum, wireless capacity expansion will be wholly inadequate to accommodate expected demand growth. This conclusion is in contrast to those that claim that the U.S. faces no spectrum shortage.

1. Overview

Demand for mobile wireless services continues to explode. Cisco’s latest *Visual Networking Index (VNI)* reports that “global mobile data traffic grew 2.3-fold in 2011, more than doubling for the fourth year in a row.”¹ Further, Cisco estimates that by 2016, global mobile data traffic will exceed its 2011 level by a factor of 18. While in many parts of the world, significant portions of expansion in mobile wireless network capacity will continue to

* The analyses and data presented in this paper are intended to portray the U.S. mobile wireless industry on a national average basis. They may not be representative of any particular U.S. geographic region or mobile operator, including AT&T. No proprietary AT&T data were used in performing these analyses. The conclusions developed in this paper are those of the author, alone, and should not be construed as representing any official position of AT&T. I am indebted to my colleagues at AT&T and Peter Rysavy for valuable assistance in preparing this paper. All errors are my own.

¹ See Cisco (2012a), p. 1.

be due to expansions in the geographic coverage of wireless data networks, in developed countries such as the U.S., advanced mobile broadband networks already cover 98.5% of potential subscribers.² Thus, the network expansions necessary to accommodate demand growth in developed countries will be focused most greatly on “deepening” network capacities. Technically and economically, this presents a different set of challenges from simply expanding coverage scope – a topic that has been addressed extensively in universal service research.³

The purpose of this paper is to describe and quantify the challenges particularly associated with wireless network “deepening.” This includes an analysis of the technical issues concerning what techniques for capacity deepening are feasible, and also consideration of the costs of these techniques to determine the economic capability of these techniques to keep up with growing demand. In particular, methods of capacity expansion divide into three general categories: those that involve the deployment of more radio spectrum; those that permit more intensive geographic reuse of spectrum; and those that involve increasing the throughput capacity of each megahertz (MHz) of spectrum within a given geographic area. While certain parties have suggested that technology to increase throughput capacity per MHz of spectrum and geographic reuse of spectrum will be adequate to address wireless demand growth in the U.S. over the next five to ten years, this analysis finds otherwise.⁴ Although methods to increase throughput capacity per MHz or spectrum reuse may be technologically feasible, and are expected to continue to be used intensively by wireless providers, by themselves they are likely to be inadequate or to become uneconomic absent significant increases in mobile wireless pricing. Thus, even under the conservative assumptions of this paper, substantial quantities of new spectrum will need to be deployed for mobile wireless use if currently forecasted demand is to be satisfied over the next decade without significant service quality degradation or “rationing” from price increases.

This finding is consistent with the conclusions developed by several other studies in the literature that have examined the abilities of current or expected spectrum assignments and technologies to accommodate forecasted demand.⁵ The analysis presented in this study will differ from these prior efforts both by improving on the accuracy of their analyses and by

² See FCC (2011), ¶ 46.

³ See, for example, Williams et al. (2011).

⁴ See Bazinet and Rollins (2011), Bode (2012), Burstein (2011), Chen (2012a and 2012b) and Reardon (2010) for arguments suggesting that the U.S. faces no serious shortage of mobile wireless spectrum.

⁵ See FCC (2010b), Feldman et al. (2011), Lawson (2012), Rysavy (2011) and Rysavy Research (2010, 2011a and 2011c) for analyses suggesting that mobile wireless spectrum will become critically scarce.

projecting certain enhancements in the ability of evolving wireless technologies to carry more mobile traffic.

This paper will begin by providing a brief technical background on how mobile wireless capacity can be measured, and the basic categories of techniques that may be used to expand mobile wireless capacity. These include increasing raw amounts of available radio spectrum, increasing the absolute carrying capacity of each MHz of spectrum, reducing the bandwidth required to carry popular applications and increasing the utilization of each MHz of spectrum or unit of infrastructure through cell-splitting, sharing or multiple use. An Appendix describes how mobile wireless technology has evolved over the past thirty years to take advantage of these techniques. In section 3, this history of technological evolution is contrasted with both the growth in available mobile wireless spectrum and the growth in mobile wireless demand. The paper goes on to catalog the possible forward-going capabilities and economics of several of the most well-known potential Fourth Generation (4G) Long Term Evolution (LTE) wireless technology innovations, including innovations whose effects remain highly speculative. By comparing the joint capacity expansion capabilities of these new and old techniques with demand growth estimates, it is possible to evaluate their ability to accommodate demand growth and to reduce upwards pressure on current wireless pricing. In the end, we conclude that by themselves, these methods will be inadequate to accommodate fully expected demand growth at today's prices. Thus, increased assignment of radio spectrum to mobile wireless will be essential. This is in contrast to suggestions from certain parties that spectrum scarcity should not be a terribly significant concern for government policymakers.

2. Mobile wireless network basics

Modern mobile wireless networks are hybrids of wireless and wireline links.⁶ Within a “last-mile” local area (often called a “cell”), the network uses radio waves to convey signals between a cell site or tower and the mobile customer's wireless device. In the U.S., these cell sites are then most commonly linked back (or “backhauled”) to a base station controller⁷ and then to a mobile switch or packet data router by high capacity wireline facilities.⁸ High

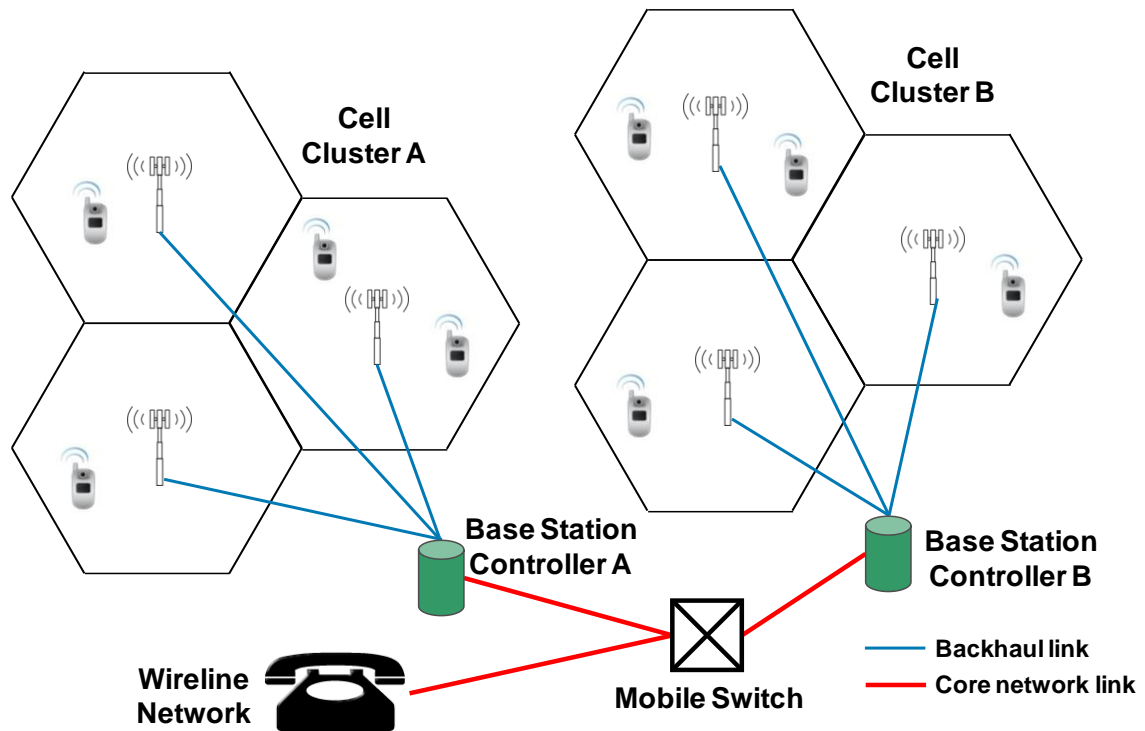
⁶ Additional technical information is presented in an Appendix to this paper. For a more complete discussion of wireless mobile network technology, see textbooks such as Forouzan (2003) or Korowajczuk (2011). Certain of the material developed in this section and in the Appendix is from these sources, in addition to Rysavy Research (2011b and 2012).

⁷ Base station controllers are sometimes also called radio network controllers.

⁸ It is also possible for these backhaul facilities to be fixed microwave radio technology. This is more common in less developed countries than in developed countries with pre-existing ubiquitous wireline

capacity wireline facilities are also used to link these regional mobile network facilities with fixed telecommunications networks or with mobile network facilities in other regions. This mobile network schematic is displayed in Exhibit 1.

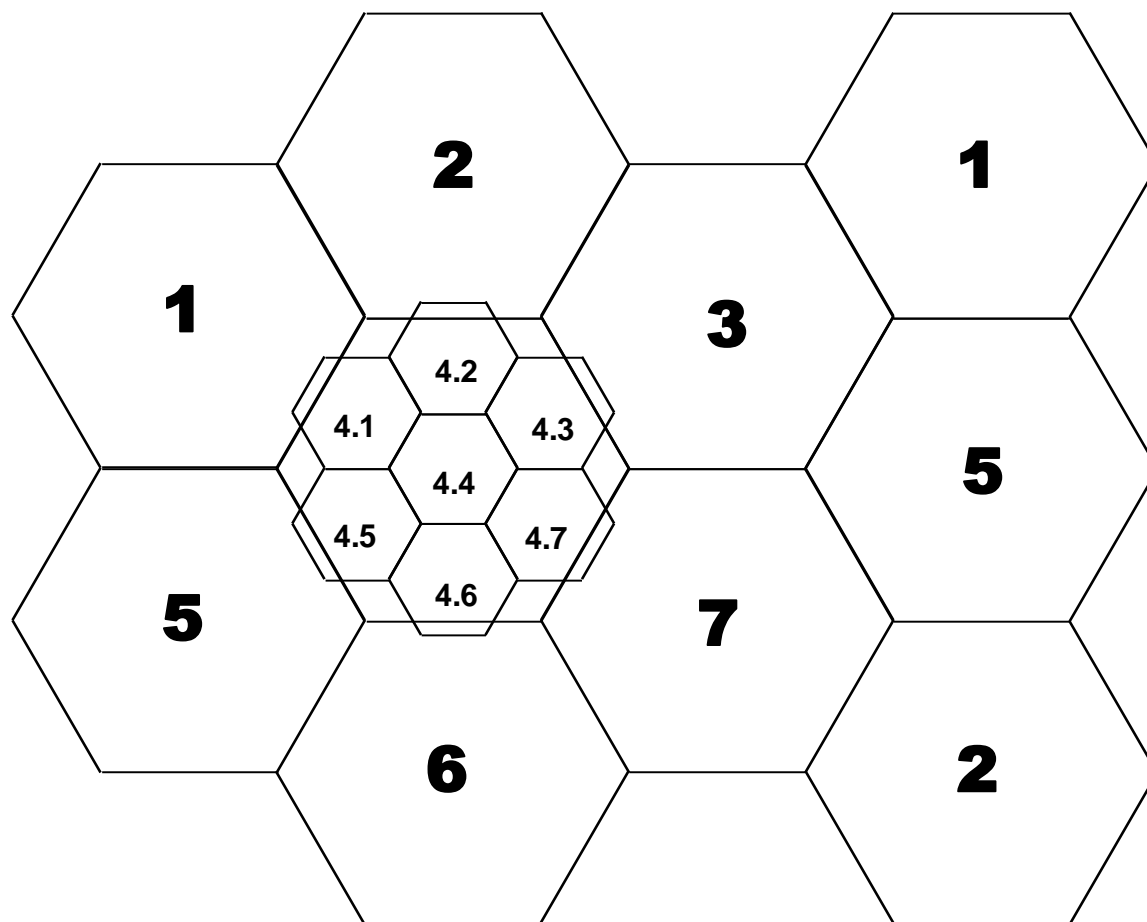
EXHIBIT 1: Mobile Network Schematic



Wide mobile network coverage using limited radio spectrum is obtained by placing multiple cell towers in an array of last-mile radio service areas. Commonly, these arrays are depicted diagrammatically as packed hexagons. This is displayed in Exhibit 2.

networks. In developed countries, wireline backhaul is transitioning from lower capacity copper T1 or E1 links to more modern and higher capacity fiber Ethernet links. This upgrade in backhaul has become necessary due to the large increases in mobile data traffic handled by each cell site.

EXHIBIT 2: Hexagonal Cellular Array



The carrying capacity of a mobile wireless system is the total amount of data or voice traffic that the system is able to transfer to and from customers.⁹ For data traffic, this is commonly measured in bytes. These bytes divide into two categories: user data and radio network overheads. The former are data that actually are transferred from the originating user to the receiving user. The latter are data that are “consumed” by the radio network for the purpose of enabling and managing the user data flows.¹⁰ Among the most numerous

⁹ In general, the capacity of the total network will be limited by the capacity of its last-mile radio access network (RAN). Although congestion in backhaul links may possibly occur, increased availability of fiber backhaul facilities should allow backhaul bottlenecks to be engineered away.

¹⁰ Such “overhead” data may include synchronization bytes, error-correction codes, framing bytes, control channel messages, etc. In mobile wireless networks such overhead bytes, depending on signal conditions, may account for over half of the total number of bytes transmitted by the radio network (see Forouzan, 2003 or Korowajczuk, 2011). These amounts of overhead bytes are much greater than are generally required in wireline networks. There are two reasons. The first is that the transmission medium used for wireless (i.e., the air) is much less protected against outside interferers or other signal impairments than the copper or fiber cables that constitute wireline transmission media. The second challenge comes from mobility. When a signal source or receiver is moving, this not only creates additional sources of signal degradation, but degradations that are constantly changing. Each of these

“overhead” bytes carried by wireless networks are those associated with forward error correction (FEC). These bytes (which may include checksum digits or other redundant data) enable radio networks to more efficiently increase their transmission of actual unerrored user data (“goodput”) than may be obtained by reducing gross data transmission speed to increase signal-to-noise ratios. While the carriage of these and other overhead bytes is essential to the effective operation of wireless networks, any improvement in transmission protocols or practices that allows more unerrored user data bytes to be carried per overhead byte consumed will improve the ability of the network to serve its customers.

Wireless data are carried by radio waves. Such waves undulate with a periodicity (i.e., frequency) measured in Hertz (Hz) or cycles per second. These radio waves are made to carry data by modulating or distorting them from otherwise uniform patterns.¹¹ Thus, one pattern of distortion may be employed as code for a digital “zero,” while another distortion pattern may be code for a digital “one.” The more waves a system can modulate in a second, the more coded zeros or ones it can send. The quantity of waves (or amount of spectrum) a wireless system is allowed to modulate each second is called its bandwidth, and is measured in Hz. Everything else equal, a signal with a higher bandwidth (i.e., more Hz) can carry more data per second than a signal of lower bandwidth (i.e., less Hz).

The total amount of data that a network may transfer over a given period of time relates closely to the rate at which it transfers data bytes. All things equal, a “faster” network will transfer more bytes than a “slower” network. Rates of data transfer are measured in terms of bits per second (bps).¹² Note, however, that in addition to raw transmission speed, the total amount of data transfer will be higher on a network that operates as a higher usage/fill factor (i.e., transfers data during more seconds of the measurement period). This can be achieved if a network has traffic offered more uniformly to it over the measurement period – either because the network serves multiple users whose patterns for offering traffic to the network are less than perfectly correlated, or because the network employs packet

challenges makes it less assured that a bit sent across the airwaves will be received as accurately as a bit sent through a cable. As a result, to achieve a given level of accurate throughput, more overhead bytes are required by wireless networks than by wireline networks.

¹¹ While beyond the scope of this paper, popular ways of modulating radio waves to carry data are by altering their amplitude (i.e., signal strength), their frequency (i.e., the time between wave peaks), or their phase (i.e., their peaks versus troughs).

¹² By convention, data transfer rates are typically measured in terms of bits (b), and data quantities in terms of bytes (B). Because there are eight bits in a byte, a transfer rate of 8 bps corresponds to transferring one byte per second (1 Bps).

scheduling protocols that efficiently divide traffic into different queues based on the immediacy of their need for transmission.¹³

Perhaps the most well-known way for cellular networks to increase the amount of data they carry is by dividing or splitting cells to reduce cell size, and thus increase the number of cells serving a given area. This is done by deploying more radio towers/antennas and shrinking the reach of each tower by reducing the radiated power of its radio transmissions. This is shown in Exhibit 2 by noticing that cell 4 is divided into seven smaller cells: 4.1, 4.2, ..., 4.7. By doing so, radio spectrum may be “reused” for multiple simultaneous transmissions within a larger geographic area, rather than just one. Thus by subdividing cells, the amount of traffic that a Hz of spectrum can carry within an overall geographic area (measured by bps/km²) is increased. But while very effective at deepening wireless network capacity, this method is also expensive – requiring the construction of extra towers/antennas, deploying more radios and base station equipment; as well as extending additional backhaul links to link new towers back into the mobile operator’s core network.¹⁴

Over the past thirty years, mobile wireless networks have evolved from First Generation (1G) to Fourth Generation (4G) technologies that make increasingly efficient use of available radio spectrum. This is described in greater detail in the Appendix. In the following sections, the contributions of these various technical enhancements are developed – and presented in comparison with contributions to capacity growth from raw spectrum growth and increased reuse of available radio frequencies.

3. Analysis of historical demand and spectrum growth

3.1. Growth in mobile usage

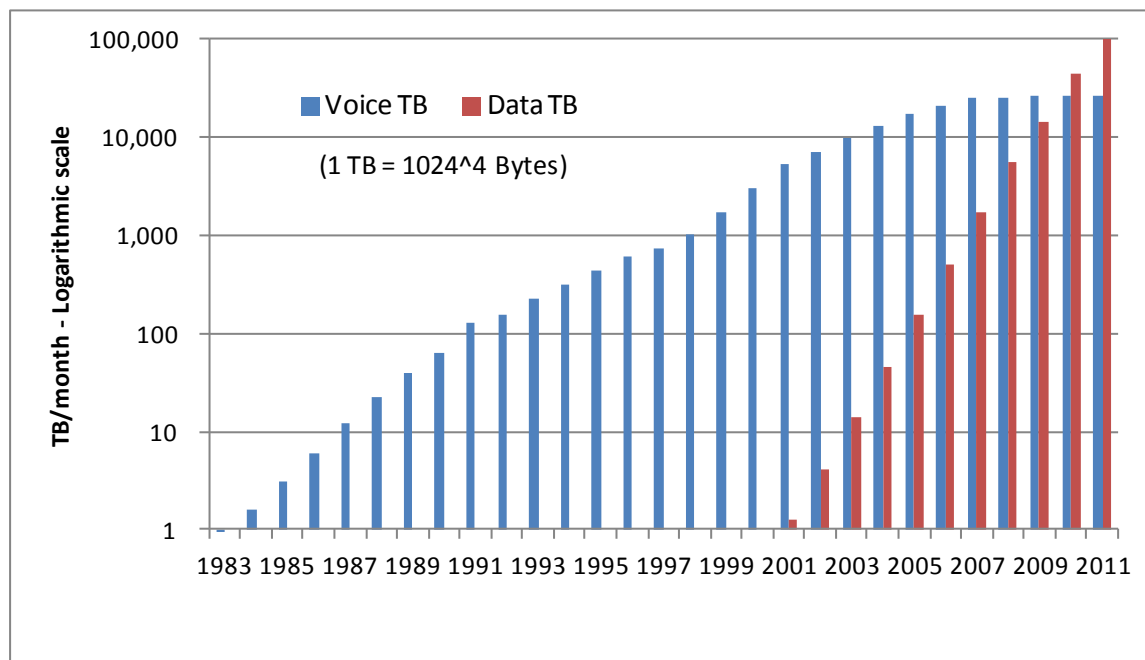
The history of U.S. mobile services demand has been one of continuous and increasingly rapid growth. As noted in the Appendix, the payload carried by 1G networks was almost exclusively voice. This continued to be the case with Second Generation (2G) networks – fed both by large increases in subscribership and by increasing numbers of

¹³ Not all traffic requires immediate handling. If data bytes supporting a voice conversation are not transferred immediately, the conversation will become broken up and unintelligible. But data bytes supporting certain file transfers may tolerate delays quite well. Indeed, the speed at which a file is transferred is perhaps best indexed by how soon its last byte is transferred, and not by how quickly any of its intermediate bytes arrive. Thus, if file transfer bytes are held briefly and interspersed into gaps in a voice transmission, it is quite possible for a joint-use wireless channel to operate far more efficiently than if separate channels were used to satisfy each demand. See Yuksel et al. (2010).

¹⁴ Reuse techniques may also be slow to implement – as regulatory approvals may need to be obtained before new cell sites are deployed or additional equipment is placed at pre-existing sites.

monthly voice minutes used by each subscriber. But with the advent of advanced mobile devices and Third Generation (3G) networks in the middle 2000s, data usage began to explode. Indeed, if voice minutes are converted into data traffic equivalents, it is possible to chart historical rates of mobile network demand growth.¹⁵ This is shown in Exhibit 3. As can be seen, U.S. data usage outstripped voice usage by the end of 2009, and by 2011 data usage exceeded voice usage by a factor of 4.¹⁶

EXHIBIT 3: History of Voice and Data Demand Growth



Notes: Voice TBs calculated from CTIA (2012b) data on mobile minutes, adjusted to data equivalents assuming 20 kbps encoding. Data TB from Cisco (2009, 2010, 2011 and 2012a) for 2008-2011, and extrapolated backwards for 2001-2007.

This explosive growth in data demand has been the product of two forces: increasing numbers of subscribers adopting advanced mobile data devices and increasing monthly use of mobile data services by the subscribers that are using these devices. Because adoption of smartphones is now nearing 70% in the U.S., it is possible that the first force driving mobile usage growth may begin to stabilize, but the latter shows no hint of flagging.¹⁷ Cisco

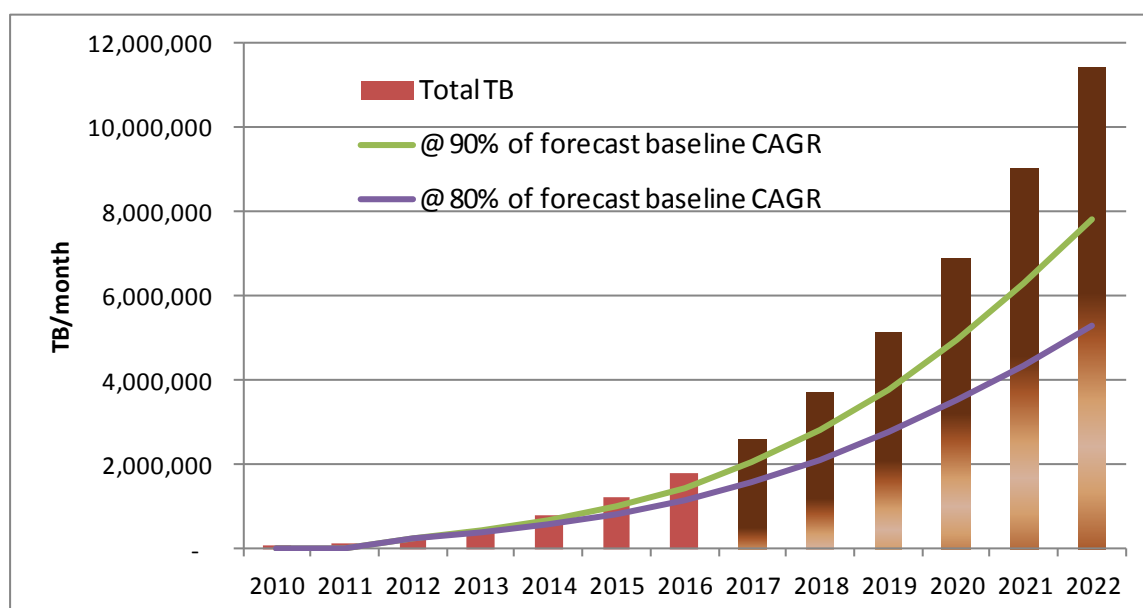
¹⁵ This analysis assumes that each voice signal is the rough equivalent of a 20 kbps packet digital signal.

¹⁶ Similar global figures are shown in Ericsson (2012a and 2012b).

¹⁷ See Credit Suisse (2012), p. 65 and Ericsson (2012b). But although smartphone penetration may be close to saturation, customers are adding still additional mobile devices like tablets or netbooks to their portfolio. Further, we are just beginning to see the widescale adoption of machine-to-machine devices

estimates that while current mobile data growth rates (that still exceed 100% annually) may be the maximum that we shall see in the advanced countries, these growth rates are projected only to decline slowly, and will still exceed 50% in 2016.¹⁸ Total mobile traffic growth is shown in Exhibit 4.¹⁹ For sensitivity purposes, this chart also shows trends assuming data growth at annual rates that are 10% and 20% less than the growth rates implied by the adjusted baseline Cisco forecast. This second alternative cuts by over half the projected amount of traffic that U.S. mobile networks will carry at the end of our analysis period in 2022.²⁰

EXHIBIT 4: Forecasted Growth in Mobile Demand



Notes: Forecasted total demand growth assumes Voice TB remain flat at 2011 levels. Cisco (2012a) figures for Data TB used for 2010-2016. These Data TB figures are extrapolated forward for 2017-2022 assuming that Cisco’s forecasted demand growth rate for 2015-16 decays by 4 percentage points per year over the 2017-2022 period.

(OECD, 2012). As we evolve to the “Internet of Things,” there may be no limits to how many new data devices will seek to be connected to mobile networks.

¹⁸ See Cisco (2012a). Cisco currently projects figures out to 2016. To continue the projection out to 2022, Cisco’s projected 2015-16 growth rate of 51% is reduced by 4 percentage points per year to reach a growth rate of 27% in 2021-2022. This latter growth rate matches reasonably current growth rates in wireline broadband use per subscriber (Cisco, 2012b).

¹⁹ Based on recent experience, we assume that voice traffic remains flat at its 2011 level throughout the 2012-2022 period.

²⁰ Note that Qualcomm believes that mobile data growth from 2010-2020 will be a factor of 1000x. This is much higher than our Cisco-based assumption of a 100x growth factor over this same period. See <http://www.qualcomm.com/media/blog/2012/08/09/heard-1000x-challenge-hint-it-s-about-mobile-data-growth>. This suggests that even the baseline traffic growth assumption used in this analysis may be an understatement.

3.2. Growth in spectrum resources

In contrast to the exponential growth in mobile bandwidth demand, U.S. mobile spectrum has been doled out at a far more moderate pace. The following Exhibit illustrates the significant assignments of wireless spectrum by the Federal Communications Commission (FCC) for U.S. mobile use.

EXHIBIT 5: Timeline of U.S. Mobile Spectrum Allocations

Year	Allocation	Band	Name
1983	40 MHz	850 MHz	Cellular
1989	10 MHz	850 MHz	Cellular
1993	14 MHz	800 MHz	SMR
1995	130 MHz	1900 MHz	PCS
2005	194 MHz ²¹	2500 MHz	EBS/BRS
2006	90 MHz	1700/2100 MHz	AWS-1
2008	70 MHz	700 MHz	700 MHz
Totals	548 MHz (assuming all EBS/BRS spectrum to be usable) 409.5 MHz (assuming only 55.5 MHz of EBS/BRS spectrum)		

Notes: Data in this Exhibit are assembled from Bazinet and Rollins (2011), FCC (2011), ¶¶ 270-275 and Feldman et al. (2011). Use of different data sources may result in slightly different implied figures for quantities of U.S. mobile wireless spectrum and their allocation dates. These small differences in quantity or in timing are inconsequential for the purposes of the following analysis.

While Exhibit 5 displays current mobile wireless spectrum allocations by the approximate year the assignments were made, it does not give an accurate indication as to the amount of spectrum actually built-out by mobile wireless operators. Build-outs typically lag

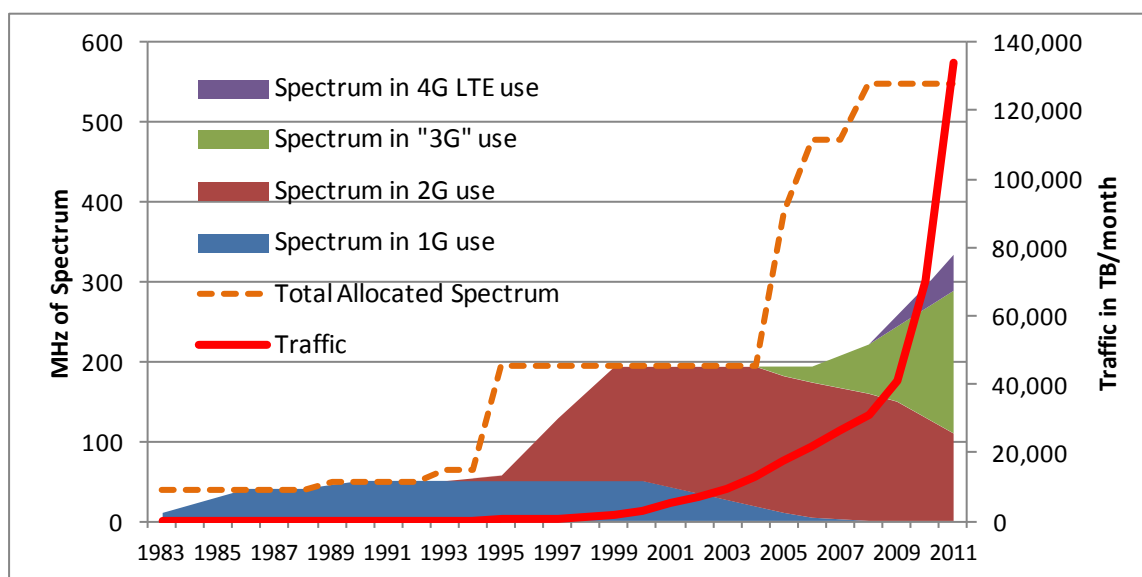
²¹ The EBS/BRS spectrum noted in this Exhibit is leased by Clearwire from its primary licensees, who are educational or religious institutions and providers of wireless CATV services. Because current FCC rules require certain portions of this spectrum to remain in educational use and may involve other encumbrances, the FCC considers only 55.5 MHz of this spectrum to be immediately deployable for mobile wireless use. See http://wireless.fcc.gov/services/index.htm?job=service_home&id=ebs_brs; http://wireless.fcc.gov/services/index.htm?job=licensing_1&id=ebs_brs and <http://wirelesspectrumreview.com/wireless-spectrum-bands/brsebs/>.

spectrum assignment dates by a number of years.²² There are several reasons. First, some of the spectrum bands were not cleared of their previous occupants as of the date of their reassignment to mobile wireless. This was particularly true of the EBS/BRS and AWS-1 bands. Second, build-outs, especially of new technologies, take planning and time. Thus, for the purpose of evaluating the actual amount of spectrum available to meet customer demand, we shall employ a rough rule that build-outs of spectrum in new frequency bands do not start until a year or more following its assignment, and that its full deployment into the market then takes place over an interval of four years.²³ Based on these assumptions it is possible to better compare growth in mobile traffic versus spectrum resources. This is presented in Exhibit 6. It is important to emphasize the difference in the scales used on this chart to measure spectrum growth versus traffic growth. The former is on the left, the latter on the right. Thus, if one looks at the mid-1990s as a base, since that time allocated spectrum has increased from 64 MHz to 548 MHz (assuming all of the EBS/BRS spectrum will be usable) – a factor of 8.5. But over the same period, total traffic has risen from a bit over 300 TB/month to 133,807 TB/month – a factor of about 425.

²² See Peter Rysavy (2012).

²³ The exceptions we make to this general rule are as follows. Buildouts of the initial Cellular allocation are assumed to have begun in the year of assignment. For the extra 10 MHz added to the Cellular band in 1989, deployment is assumed to have begun in the same year as assignment and completed the following year. For the AWS-1 and 700 MHz spectrum bands, it is assumed that buildouts did not commence until two years after the year of assignment. Note that these assumptions are quite rough, but even moderate changes to them will not alter general patterns in any significant sense.

EXHIBIT 6: Spectrum Growth versus Traffic Growth



Notes: Gross spectrum quantities based on Exhibit 5, with author’s estimates to divide by technology deployed and date of build-out. Traffic data are from Exhibit 3.

To be sure, while Exhibit 6 provides information about the amounts of spectrum licensed and built out to provide mobile services, it does not inform directly as to the amount of this “lit” spectrum that was actually being used to carry the mobile wireless demands that were offered to these networks in the given years. Such an estimate is essential to determine the amount of “headroom” that may exist in built-out U.S. mobile networks – and may remain available to accommodate increases in customer demand. For the purposes of evaluating the ability of U.S. mobile networks to accommodate increasing future demand, it is most useful to have a relatively recent estimate for this headroom. For several reasons, we choose 2010. Not only is 2010 recent, but it is a year that just precedes significant rollouts of services provided over AWS-1 or 700 MHz spectrum. Further, it is relatively soon after Clearwire began deploying small quantities of service over its EBS/BRS spectrum. Thus, in 2010 we shall assume that the amount of spectrum that was actually being “consumed” to offer mobile services was 194 MHz (= 50 MHz Cellular + 14 MHz SMR + 130 MHz PCS).²⁴ But to be especially conservative in estimating needs for additional spectrum, we shall make an additional assumption that the 194 MHz of spectrum consumed in 2010 was not used as

²⁴ While it was likely that in 2010 some service was being provided over EBS/BRS and AWS-1 spectrum, these quantities were quite small. Further, it is also likely that there was still some modest amount of headroom remaining within Cellular, SMR and PCS spectrum as of that date. We assume that the effects of these two over-simplifications roughly cancel each other. This assumption of 194 MHz occupancy in 2010 is also somewhat consistent with the estimate of 170 MHz occupancy in 2009 assumed in FCC (2010b) and the assumption of 192 MHz use in 2011 by Bazinet and Rollins (2011).

intensively (i.e., with as high a fill factor) as may be the case in the future. Based on analyst reports that “lit” U.S. mobile networks are operating at only 80% of their capacity on average, we shall adjust the figure this analysis employs for 2010 baseline spectrum usage down to 155.2 MHz (= 0.80 * 194 MHz).²⁵

Unfortunately, since 2008 there have been no significant additional allocations of mobile wireless spectrum by U.S. regulatory authorities.²⁶ Depending on further development of mobile wireless standards and FCC action, it is possible that some additional spectrum may be deployed over the next several years. This includes 12 MHz of unpaired lower 700 MHz spectrum that was previously used for broadcast fixed or mobile TV, 20 MHz of 2300 MHz WCS spectrum and 40 MHz of 2200 MHz S-Band MSS spectrum. In addition, there are several other bands currently under study by the FCC and other government authorities for conversion to possible mobile wireless use.²⁷ While all of these efforts are promising, the exact contributions they may offer remain speculative as successful repurposing of these spectrum bands for two-way terrestrial mobile wireless is by no means assured. Technical or regulatory roadblocks may occur. For future analysis, we shall assume, alternatively, that: (a) the FCC is unsuccessful at reallocating additional spectrum to mobile wireless so that it may be built out in time to serve customer demand over the 2012-2022 period; and (b) the FCC is successful in reallocating an additional 300 MHz of spectrum to mobile wireless and it is built out over the 2014-2022 period.²⁸ The former may be considered a pessimistic

²⁵ See Credit Suisse (2011) for survey results that North American mobile networks are operating at 80% of capacity. By reducing our baseline figure for effectively utilized spectrum, it implies that a greater amount of past FCC spectrum allocations remain available to serve future demand – thus reducing implied future needs for additional spectrum.

²⁶ Certain attempts have been made, however. In January 2011, the FCC issued an Order permitting LightSquared to repurpose 66 MHz of L-Band Mobile Satellite Service (MSS) spectrum to terrestrial mobile use. See http://hraunfoss.fcc.gov/edocs_public/attachmatch/DA-11-133A1.pdf. This spectrum, however, was adjacent to spectrum used by Global Positioning System (GPS) satellites and use by LightSquared of this spectrum was ultimately adjudged to interfere with neighboring GPS devices. As a result, in February 2012 the FCC withdrew LightSquared’s conditional authorization to use this spectrum for terrestrial mobile wireless. See http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-312479A1.pdf.

²⁷ These include 120 MHz in the UHF band currently used by broadcast television service and a portion of the 95 MHz in the 1750-1850 MHz band currently used by government agencies. More immediately, it is possible that 20 MHz of government spectrum in the 1755 MHz band could be paired with 20 MHz in the 2155 MHz AWS-3 band. For a fuller discussion of these future possibilities, see <http://transition.fcc.gov/statelocal/presentations/Incentive-Auctions-MB.ppt>, NTIA (2012), and more generally FCC (2010a), pp. 85-89.

²⁸ The FCC’s *National Broadband Plan* (issued in March 2010) calls for an additional 300 MHz of spectrum to be allocated to mobile wireless by 2015 and for a total of 500 MHz to be added by 2020. See FCC (2010a), p. 84.

assumption and the latter an optimistic one.²⁹ It is likely that actual new allocations will range somewhere between these two figures.³⁰

3.3. Intensity of spectrum reuse

Exhibit 6, showing that mobile usage has grown orders of magnitude faster than available spectrum, provides telling evidence that so far the largest contributor to increased mobile capacity in the U.S. has been investment by mobile operators to evolve networks to more spectrally-efficient technologies and to erect more cell towers that enable more intensive reuse of the radio frequencies that have been allocated to mobile wireless. When one examines the data, it is clear that most of this capacity uplift has been due to more intensive reuse.

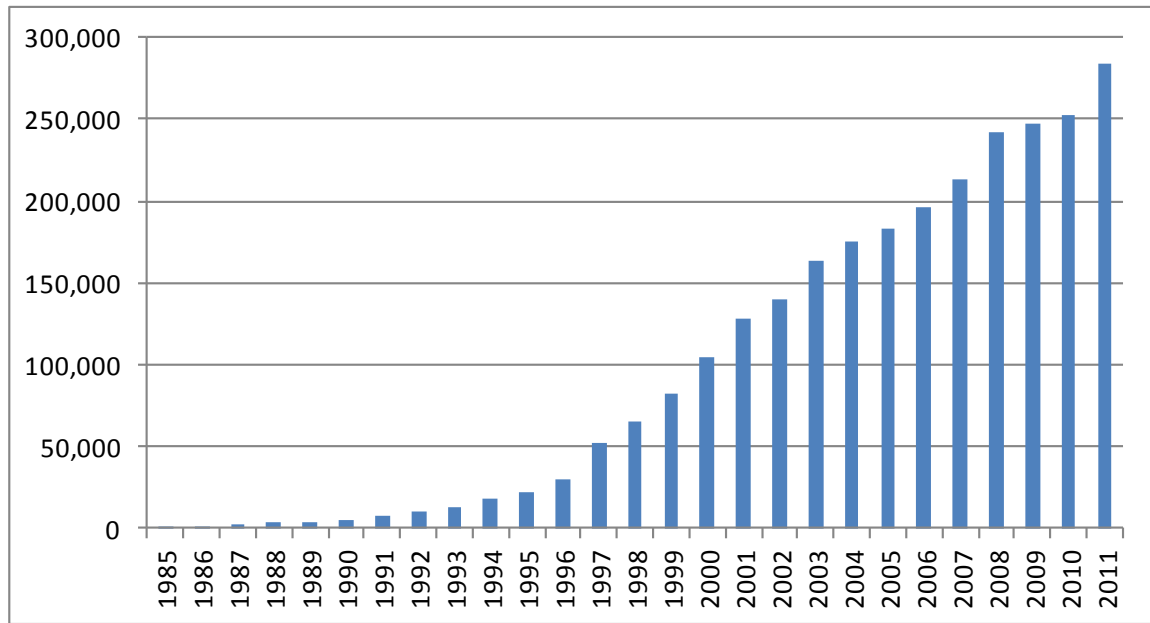
As noted in the Appendix, moving from 1G to 2G and 3G technologies has improved the efficiency with which a given amount of spectrum carries voice calls by a factor of less than 10. Improvements in data-carrying efficiency may be roughly double this. In contrast, much greater advance has been due to reuse of spectrum by splitting cells.³¹ From 1985 to 2011, cell site counts have grown by a factor 330. This history of cell site expansion is displayed in Exhibit 7.

²⁹ Realistically, given that no new allocations have been effected since the date of the *National Broadband Plan*, without significant changes in U.S. spectrum regulatory policy there is very little probability that more than 300 MHz of spectrum could be reallocated and built out within the 2014-2022 timeframe. Lead times for identifying, auctioning, clearing and building out spectrum are generally on the order of 8-10 years. See FCC (2010a), p. 79.

³⁰ A recent U.S. government report (PCAST, 2012) recommends that 1000 MHz of government-controlled spectrum within the sub-4000 MHz bands be made available for shared, secondary use by commercial mobile wireless. While ambitious, it is extremely unclear as to whether this proposal is practicable from a technical or economic point of view, and even if it is, whether it can have any appreciable effect within the 2014-2022 time period. As noted earlier, North American mobile networks already operate at 80% utilization within their exclusively licensed bands. Thus, the opportunity to share, on an idiosyncratic secondary basis, some additional spectrum, poses substantial economic challenges. New network equipment and cell sites would need to be deployed, only to be able to make partial unassured use of these new bands. This suggests that the additional costs of using these shared bands may be disproportionate to their effective capacity uplift. Further, it is unlikely that all frequencies would be available throughout the country – or in the particular locations most in need of additional capacity.

³¹ See “Cooper’s Law” at <http://www.arraycomm.com/technology/coopers-law> .

EXHIBIT 7: Number of U.S. Cell Sites



Notes: CTIA (2012a) count of cell sites. Note that CTIA’s count is by network technology by physical location.

While this may suggest that capacity reuse has also risen by a factor of 330, the capacity effects of this proliferation of cell sites are hard to quantify precisely. For example, in the early years of a mobile technology’s deployment, it is likely that most new sites were deployed to extend coverage of that technology, and not simply to increase the capacity of that technology within already covered areas. In the later years of a technology’s deployment, capacity increases are more likely to be the purpose of new sites. But, later deployments may be more likely to be of micro or pico cells, which are designed to serve smaller areas and traffic quantities, thus suggesting decreasing incremental capacity expansion per-site as cell site counts rise.³² But if we generously assume that all recent site deployments have been strictly to improve capacity, and that these deployments provide an effective uplift in network capacity in direct proportion to their number, this suggests that if site deployment rates observed over the last five to ten years continue, traditional reuse will provide about an 8% per year increase in capacity to serve demand growth.

³² See Chapin and Lehr (2011), p. 32, FCC (2010a), p. 77, Goldstein (2012) and Paolini (2011).

4. Spectral efficiency and unfolding 4G LTE technological capabilities

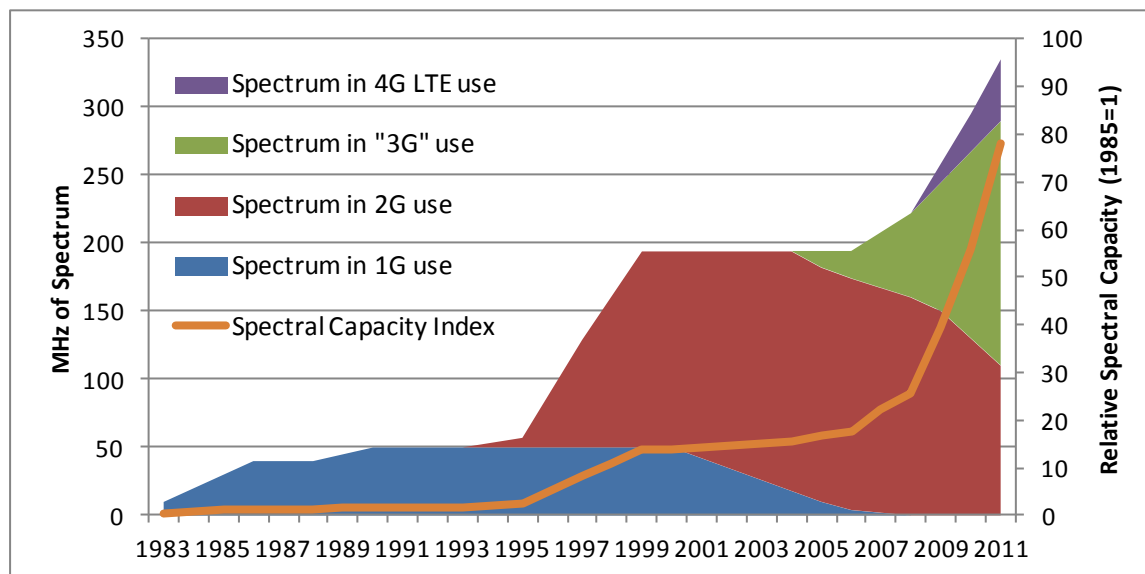
4.1. Evolution of U.S. spectral efficiency and reuse

Over the past thirty years there have been two sources of improved spectral efficiency of U.S. mobile networks: developments of wireless technologies that are more spectrally efficient; and increasing the quantity of spectrum used by these more efficient technologies relative to the quantity of spectrum used by older, less efficient technologies. For the purposes of the analysis presented here, we shall assume that 1G technologies have a spectral efficiency of less than 0.1 bps/Hz on a sector basis; 2G technologies have an efficiency of 0.25 bps/Hz; “3G” technologies an initial efficiency of 0.5 bps/Hz – rising to between 0.90 bps/Hz and 1.0 bps/Hz for advanced implementations such as EV-DO, Rev. A or 4G HSPA+;³³ and current 4G LTE technologies 1.4 bps/Hz.³⁴ Based on these spectral efficiency figures and estimates of the quantities of U.S. mobile spectrum deployed to different technologies, it is possible to map the evolution of the potential capacity of U.S. mobile networks to handle voice and data traffic. This is displayed by the Spectral Capacity Index in Exhibit 8 (scale on right side of the chart). As can be seen, the deployment of more spectrum and the implementation of more efficient 3G and 4G technologies have provided a substantial uplift in mobile network carrying capacity.

³³ Note 66 in the Appendix explains why, for the purposes of the spectral efficiency analysis used in this paper, we include 4G HSPA+ in the “3G” technology category.

³⁴ See Rysavy Research (2012), pp. 54-55. While these figures may be imprecise indicators of the actual rates of throughput achieved by different in-place network technologies, all that is absolutely necessary is that these figures provide a reasonably accurate picture of the *relative* spectral efficiency of the different technologies.

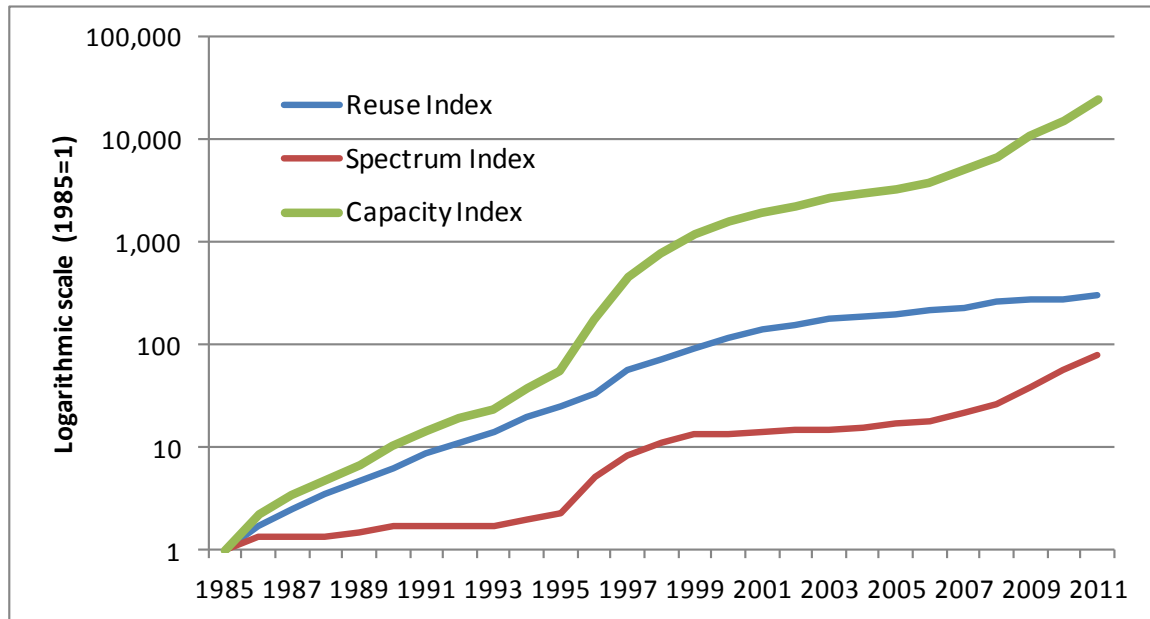
EXHIBIT 8: Capacity Expansion from Spectrum Expansion and Migration of Use to More Spectrally-Efficient Technologies



Notes: Division of spectrum by technology and dates of build-out are from Exhibit 6. Spectrum efficiency factors used to compute spectral capacity are from Rysavy Research (2012), p. 55.

But in addition to capacity growth due to spectrum expansion and the migration from older technologies to newer ones, there has also been increased reuse of frequencies due to the deployment of additional cell sites (Reuse Index). These two capacity expansion factors can be combined to estimate the total growth in U.S. mobile network capacity (Capacity Index). This is displayed in Exhibit 9.

EXHIBIT 9: Contributions of Increasing Spectral Efficiency and Frequency Reuse to Overall Mobile Network Capacity



Notes: Relative reuse figures from Exhibit 7, relative spectral capacity figures from Exhibit 8. Overall capacity index is the product of these two component indices.

While this progression of past expansion in overall mobile network capacity is certainly impressive, it faces a daunting task in keeping up with the projections of demand growth displayed in Exhibit 4. Whether or not it will be successful will depend importantly on future contributions to capacity growth from 4G LTE technology. These will be discussed in the following section.

4.2. Efficiency improvements from future 4G LTE innovations

While current LTE networks are significantly more spectrum-efficient in carrying mobile traffic than earlier network technologies, LTE is capable of even further improvements as its technology progresses to LTE-Advanced (LTE-A or 4G+).³⁵ These improvements divide into three categories: increasing the raw transmission throughputs over

³⁵ The progression of LTE technology is indexed by different releases of 3GPP specifications – with initial LTE-A specifications in Release 10. See Rysavy Research (2012), pp. 24-26. It is possible that certain of the enhancements that we term “4G+” are part of Releases earlier than 10, but since they are only now in the process of first implementation in networks, we shall, for convenience, call them “4G+.”

LTE radio links; further increasing the possibilities for spectrum reuse; and packing offered data more efficiently into available transmission capacity.³⁶

4.2.1.Improved transmission throughputs

Perhaps the most dramatic potential improvements in LTE's capacity to handle traffic derive from higher-order Multiple Input Multiple Output (MIMO) implementations.³⁷ Current LTE deployments use 2x2 MIMO. This places two antennas at the base station and two antennas in the user device.³⁸ Because of the slight physical displacement of each transmitting antenna from the other transmitting antenna(s) and of each receiving antenna from the other receiving antenna(s), each sent and received signal will be subject to different multipath characteristics.³⁹ By examining the four signals together, more of the originally encoded information may be extracted. In addition, MIMO technology may be used to send multiple concurrent transmission streams between the base station and user device. This too may increase throughput over a given amount of spectrum⁴⁰ Implementations of 4x2, 4x4, and even 8x8 MIMO are already part of LTE-A specifications. While theoretically, 4x4 MIMO should provide twice the throughput of 2x2 (although at higher power usage), and 8x8 four times the throughput of 2x2, field tests suggest that actual uplifts are substantially less.⁴¹ However, these important advances come with their own challenges. First, mobile user devices generally are quite small. Although it has been possible to place two diverse antennas within their form factor, placing four or eight diverse antennas is much more problematic.

³⁶ Lawson (2012) and Real Wireless (2010) also provide catalogs of capacity expansion techniques. See Bhat et al. (2012) for more technical discussions as to how LTE-A achieves these improvements.

³⁷ MIMO technology employs multiple antennas on cell towers and within user devices to enable two or more parallel radio links connecting a tower to a user device. See Korowajczuk (2011), chapter 14 and Rysavy Research (2012), pp. 74, 102-103.

³⁸ The first figure in Ax B MIMO represents the number of antennas at the network base station. The second figure represents the antenna count at the user device. 2x2 MIMO results in four effective signals: one from transmitting antenna 1 to receiving antenna 1; a second from transmitting antenna 1 to receiving antenna 2; a third from transmitting antenna 2 to receiving antenna 1; and a fourth from transmitting antenna 2 to receiving antenna 2.

³⁹ Multipath interference is perhaps the greatest reason why mobile technologies fail to attain throughput rates that match their theoretical maxima. This interference results from environmental obstacles like trees, buildings or mountains reflecting radio beams so that all of a signal's energy does not take the same path from its source antenna to its destination antenna. Because the whole signal does not take the same path (multipath), portions of the signal arrive at the destination both faded and slightly delayed, creating a less intelligible received signal. This becomes most acute when the signal is of high bandwidth because the portion of a data symbol whose receipt is delayed by multipath may interfere with the portion of a subsequent data symbol that arrives "on time."

⁴⁰ There are even more reasons why MIMO technology may improve throughputs between two points. Their discussion is beyond the scope of this paper. For further information, see Korowajczuk (2011), pp. 262-281.

⁴¹ See Rysavy Research (2012), p. 56.

Indeed, it may be that throughput advances from higher-order MIMO will only be won at the cost of physically larger and more power-hungry mobile devices. Second, placing additional radio transceivers at cell sites also bears increased costs for equipment, tower structure and electric power.

Another technology being developed is called Coordinated Multi-Point transmission/reception (CoMP). This technology leverages what would otherwise be undesirable interfering transmissions from adjacent cells or sector antennas by having user devices purposefully communicate with several cell sites (or sectors at a single cell site) simultaneously. Because CoMP coordinates these transmissions, what was previously interference from adjacent cell sites or sector antennas is actually converted into useful information-carrying bandwidth, and a more robust total signal received.⁴² While potentially valuable, CoMP is expected only to enter LTE-A specifications via 3GPP Release 11 in late 2012, and its potential implementation in commercial networks is at least several years in the future.

However, these two significant throughput-improving advances associated with LTE-Advanced are not achieved just through software tweaks. Higher-order MIMO requires new antennas at base stations, and completely new user devices that incorporate both the increased number of antennas and the chipsets necessary for the more complex processing of received and sent signals. New chipsets will also be needed for advanced versions of CoMP. As a result, we are unlikely see commercial networks fully incorporating these advances until 2016 or beyond.

Taking all of these potential enhancements together, LTE-A technology is expected to achieve a spectral efficiency of 2.25 bps/Hz – a lift of nearly 61% over the 1.4 bps/Hz efficiency of current 2x2 MIMO LTE deployments (Rysavy Research, 2012, p. 55).

4.2.2. Increasing spectrum reuse

As noted earlier, the prevalent form of spectrum reuse in mobile networks has been from “cellularizing” their served geographies into smaller and smaller units, and reusing radio frequencies in adjacent units. While this technique is certainly expected to continue, there are special ways in which LTE can foster even more intensive cellularization.

When mobile cellular networks were first deployed, cells were placed primarily for coverage purposes. Such cells are often called “macro cells.” As wide coverage was achieved and service demands in dense areas grew, macro cells were split into smaller cells,

⁴² See Rysavy Research (2012), pp. 103-105.

sometimes called “micro cells.” New micro cells were also simply overlaid on dense areas within the coverage of a previously-place macro cell. But macro and micro cells are expensive. They require a tower or rooftop to place their antennas, their own managed backhaul facilities, a substantial equipment hut, DC power supplies and possibly backup batteries or electric generating equipment. As individual cell tower service areas grow smaller, it may become less sustainable economically to continue spectrum reuse by the placement of more micro cells. Instead, the possible solutions are even smaller cells, known, variously, as pico cells/Distributed Antenna Systems (DAS) or femto cells.⁴³ The former are cells designed to serve only a single office building in a complex, or even just a floor or two of a particular building. They may be located in closets or alcoves rather than on rooftops or towers. Their demands for electric power and electronics rack space are low. But these cells still require their own engineered backhaul to the core network and are managed from base station controllers so that they can take and give soft handoffs to other cell sites as a subscriber device moves in or out of their service area.

Femto cells are more basic. Typically they are designed to serve only devices located within about a dozen meters. But because they operate at very low power, many of them can be packed into a tight geography – all reusing the same frequencies. These devices generally are not managed by a base station controller and do not accept handoffs from other cell sites.⁴⁴ Despite these technical deficiencies, femto cells may be attractive to mobile network operators because, typically, they do not bear the full cost of these cells’ deployment. Rather, the femto cell device is commonly purchased by a subscriber, perhaps at a discount from the operator. The subscriber is then responsible for furnishing local power to the device, and most importantly, the subscriber must also provide a broadband backhaul line (commonly DSL or cable modem) to the Internet. Thus, signals from femto cells enter and leave the mobile operator’s core network via the Internet and not via dedicated managed backhaul. But note, too, DAS or femto cells have typically been placed in mobile networks to improve signal coverage within homes or commercial buildings, not necessarily to deepen traffic capacity. Further, because of issues related to use of the broadband backhaul line furnished by the subscriber such as capacity, security and liability, femto cell access has commonly been restricted to just a predetermined list of subscriber devices – and is not “open” for general use.

⁴³ See Chapin and Lehr (2011) and Goldstein (2010 and 2012).

⁴⁴ See Chen et al. (2010) for a discussion of femto cell architectures and issues.

Another method of spectrum reuse by which mobile networks may be able to accommodate demand growth is to offload last-mile access traffic onto Wi-Fi networks – which are connected by fixed broadband access lines to the Internet, and then into mobile operators’ core networks. While apparently quite similar to femto cell offload, Wi-Fi offload differs in several important respects. The first is spectrum. Femto cells use spectrum allocations that are licensed to mobile wireless networks – thus their use “consumes” assigned mobile wireless spectrum. In contrast, Wi-Fi offload uses unlicensed spectrum. Further, Wi-Fi has access to a very substantial amount of spectrum – with allocations of several hundred MHz in the 2400 MHz and 5000 MHz bands. Because of the low permitted power of Wi-Fi signals, it is possible to make very substantial reuse of these frequencies. To take advantage of this, many mobile wireless devices are capable of connecting to local Wi-Fi networks in addition to signals from cell towers. But while use of Wi-Fi access is highly salutary, it may not affect the analysis presented here. This is because Cisco’s *VNI* forecasts for Global Mobile Traffic which we use to measure mobile demand are intended to capture only traffic flowing between a mobile operator’s RAN and its routed core network – and not entering/exiting by Wi-Fi or femto cell offload.⁴⁵

Future implementations of LTE-A are expected to support soft handoffs of connections between femto cell or Wi-Fi access and licensed cellular spectrum access – which may cause the amount of mobile wireless traffic offloaded to Wi-Fi or femto cells to accelerate. Further, these LTE-A specifications are expected to support more dynamic load-balancing across the pico or femto cells that may exist within macro or micro cell coverage areas.⁴⁶ If these are implemented successfully, Het-nets are likely to proliferate and the effective reuse of available radio frequencies will be intensified. Because of the very important capacity-expansion capabilities of this LTE-A innovation, we shall assume that it results in an annual growth rate for “effective” LTE-A cell sites of 16% – double the assumed growth increment of 8% for non-LTE-A sites.⁴⁷

⁴⁵ See Cisco (2012a).

⁴⁶ Networks that more gracefully manage and distribute loads across different access configurations (i.e., macro, micro, pico and femto cells) are called “heterogeneous networks” (Het-nets). See Bhat et al. (2012), Qualcomm (2012) and Rysavy Research (2012), pp. 106-111.

⁴⁷ By “effective,” we mean sites that are equivalent to traditional macro or micro cells in their capacity uplift. But note that because each small cell is expected to handle less traffic than a large cell, this implies that actual small cell counts will be substantially larger than the effective cell counts that we measure here.

4.2.3. More efficient data packing

Opportunities to pack data more efficiently into total available radio access spectrum are limited in mobile networks. Perhaps the greatest source of inefficiency in data packing is due to geography. In less populated areas of the country, portions of assigned mobile wireless spectrum may be fallow because local demand is insufficient to exhaust its carrying capacity. In contrast, in dense urban areas there are concentrated sources of demand that exhaust available mobile network capacity. While it is possible (and economic) for large users of resources like electricity (e.g., aluminum smelters) to locate close to remote hydroelectric dams, subscribers using mobile devices want to use mobile services where they are currently located (which generally is in denser population areas) and not in areas where they are not located.⁴⁸

One way to gain packing efficiencies is to optimize data content sought by mobile subscribers into lower bandwidth format (e.g., stripped down web pages or video files downconverted to resolutions more appropriate for smaller format mobile devices).⁴⁹ Efficiency can also be gained from providing different mobile data applications with Quality of Service (QoS) that is tailored more precisely to their individual needs. As noted earlier, VoIP or videoconferencing needs a consistent low-latency stream of data bytes. Errored packets, while they may create some degradation, can typically be accommodated by these applications without retransmission. In contrast, web browsing may function best if intermittent bursts of data packets arrive at high speed. Finally, file transfer bytes may be unforgiving of packet errors, but accommodate latency quite well. If a mobile network is offered a variety of these different application types by its subscribers, then by using QoS-based packet scheduling it can accommodate simultaneously these diverse demands using less total bandwidth than if each application is provided with the same QoS. By incorporating QoS-based packet scheduling into its specifications, newer releases of LTE will be able to allot bandwidth to different applications in a more efficient manner than earlier mobile technologies. This will permit operators to offer high quality Voice over LTE (VoLTE) service – and eventually to decommission their separate 2G and 3G voice networks. By enabling increases in the total amount of service customers can be offered through a limited

⁴⁸ Of course the classic exception to this rule is when a mobile customer is in an area of poor reception and moves a short distance (say, to a higher floor of his house or out of a sheltered building alcove) to get a better signal.

⁴⁹ Transcoding video traffic and limiting streaming buffer sizes is an optimization technique already employed by certain mobile operators. See Verizon Wireless: “Data Plans & Features - Terms and Conditions,” available at: http://support.verizonwireless.com/terms/products/vz_email.html and “Optimization Deployment - Terms & Conditions,” available at: http://support.verizonwireless.com/terms/network_optimization.html.

amount of spectrum, it is possible that LTE QoS and packet scheduling could improve effective throughput by up to 20%.⁵⁰

In addition to more efficient data packing enabled by advanced technology, it is important to note that more ordinary changes in subscriber use patterns may result in more efficient network fills. For example, as mobile data services evolve from ones with a business focus to consumer focus, it is possible that today's data are being offered to the network more evenly throughout the day than in prior years. Thus it is possible that a significant amount of the new video traffic that is flooding U.S. mobile networks may be off-peak load – and as such be carried without the requirement for increased network capacity. While it is certainly plausible that increased fill-in of underused capacity may continue, this is limited by available peak capacity and is very difficult to quantify.

5. Overall ability of technology and spectrum to meet forecasted demand

To measure the ability of these various expected enhancements to mobile network capacity to meet forecasted growth in demand, we employ a model similar to that used in FCC (2010b). The basis for our model is to compare indices of forecast demand growth with indices of possible overall capacity expansion. The latter index is a product of separate indices for growth in raw spectrum resources, customer migration to more spectrally-efficient higher-G mobile technologies, growth in spectrum reuse, and additional capacity growth due to more effective LTE-A network packing.

Two alternative indices of raw spectrum growth are developed. One assumes that no more spectrum allocations are forthcoming to the U.S. mobile wireless industry beyond the 548 MHz already allocated – and that all of this spectrum is usable. A second assumes that an additional 300 MHz of usable spectrum will become available to the industry over the 2014-2022 time period. The reality will likely be a figure somewhere in the middle. The year 2010 is assumed to be the model's base year. Although we estimate that 294.5 MHz of allocated spectrum was built-out as of that year, we assume that only 155.2 MHz of this spectrum was "filled." This results in the initial Spectrum Index for 2010 being 1.90 (= 294.5 MHz / 155.2 MHz) rather than 1.00. This is shown in column (c) of Exhibit 10. Column (d) shows the assumed timeline for building out the additional 300 MHz that may be offered to the

⁵⁰ See Yuksel et al. (2010) for a more in-depth explanation for why multi-use networks incorporating QoS require less bandwidth for equivalent service quality than multiple single-purpose networks. LTE-A carrier aggregation may be an additional source of packing efficiency. By permitting non-contiguous spectrum blocks to be aggregated into wider channels, and by permitting asymmetric uplink and downlink block sizes, it will be possible for mobile operators to increase the throughput capacity of their spectrum holdings. This uplift is also considered in the proposed 20%.

industry. An augmented Spectrum Index incorporating these additional allocations is given in column (f).

EXHIBIT 10: Contributions of Raw Spectrum to Mobile Capacity Growth

(a)	(b)	(c)	(d)	(e)	(f)
Year	Raw Spectrum				
	Built-Out Allocated Spectrum (MHz)	Base Spectrum Index	Additional Allocations (MHz)	Augmented Spectrum Quantity (MHz)	Augmented Spectrum Index
2010	294.5	1.90			
2011	334.5	2.16			
2012	374.5	2.41			
2013	426.6	2.75			
2014	478.8	3.08	30	508.8	3.28
2015	513.4	3.31	30	573.4	3.69
2016	548.0	3.53	30	638.0	4.11
2017	548.0	3.53	30	668.0	4.30
2018	548.0	3.53	40	708.0	4.56
2019	548.0	3.53	40	748.0	4.82
2020	548.0	3.53	40	788.0	5.08
2021	548.0	3.53	30	818.0	5.27
2022	548.0	3.53	30	848.0	5.46

Notes: Time pattern for deployment of already allocated spectrum and 300 MHz of additional spectrum is author’s estimate.

Indices of effective spectral efficiency presented in Exhibit 11 are developed by weighting the relative spectral efficiencies of each technology by estimates of the fractions of total built-out spectrum that will be in use by each technology in each year. Note that this analysis assumes that if the additional 300 MHz of spectrum is released to the industry over the 2014-2022 period, almost all of this will be built out as LTE-A – thus raising the industry’s average spectral efficiency. Column (g) of Exhibit 11 shows the index for capacity growth related to migration of use to higher G technologies assuming current total spectrum allocations.⁵¹ Column (m) shows the equivalent index assuming spectrum allocations are augmented by 300 MHz over the 2014-2022 period.

⁵¹ This migration of customers to higher G technologies (principally from 2G to “3G”, 4G or 4G+), assumes no government regulations restricting U.S. mobile operators from migrating their customers. In fact, it was not until 2002 that the FCC issued an Order permitting U.S. networks to retire their 1G networks in 2008. See http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-225216A1.pdf and Lawson (2008).

EXHIBIT 11: Contributions of Migration to More Spectrally-Efficient Higher G Technologies to Mobile Capacity Growth

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)	(m)
	Migration to Higher G											
	w / Currently Allocated Spectrum						w / Augmented Spectrum					
Year	Pct 2G	Pct "3G"	Pct 4G LTE	Pct 4G+	Weighted Average Efficiency (bps/Hz)	Spectral Efficiency Index	Pct 2G	Pct "3G"	Pct 4G LTE	Pct 4G+	Weighted Average Efficiency (bps/Hz)	Spectral Efficiency Index
2010	0.44	0.46	0.09	0.00	0.57	1.00						
2011	0.33	0.54	0.14	0.00	0.70	1.23						
2012	0.21	0.59	0.20	0.00	0.86	1.52						
2013	0.12	0.60	0.28	0.00	0.99	1.75						
2014	0.07	0.54	0.39	0.00	1.07	1.89	0.07	0.53	0.40	0.00	1.08	1.90
2015	0.04	0.46	0.39	0.11	1.24	2.19	0.03	0.47	0.40	0.10	1.24	2.18
2016	0.02	0.40	0.37	0.21	1.37	2.42	0.02	0.39	0.36	0.23	1.40	2.46
2017	0.01	0.38	0.37	0.25	1.43	2.52	0.01	0.35	0.35	0.29	1.48	2.61
2018	0.00	0.33	0.35	0.32	1.52	2.68	0.00	0.30	0.31	0.39	1.60	2.81
2019	0.00	0.28	0.32	0.39	1.61	2.83	0.00	0.25	0.28	0.47	1.69	2.97
2020	0.00	0.23	0.28	0.49	1.72	3.02	0.00	0.18	0.25	0.56	1.80	3.17
2021	0.00	0.16	0.23	0.61	1.85	3.26	0.00	0.13	0.22	0.65	1.90	3.34
2022	0.00	0.08	0.19	0.73	1.98	3.49	0.00	0.07	0.17	0.77	2.02	3.56
	0.25	0.95	1.40	2.25			0.25	0.95	1.40	2.25		
	Efficiencies (bps/Hz)						Efficiencies (bps/Hz)					

Notes: Fractions of total spectrum used by different technologies and deployment dates are estimates of the author. Spectral efficiencies of different technologies are estimates from Rysavy Research (2012), p. 55. Because of migration to 4G HSPA+, blended efficiencies for “3G” technologies are assumed to be 0.7 bps/Hz in 2010, 0.8 bps/Hz in 2011, 0.9 bps/Hz in 2012 and 0.95 bps/Hz afterward.

Exhibit 12 displays forecast growth in spectrum reuse. Column (b) shows forecasted cell site counts assuming continued 8% annual growth and current raw spectrum allocations. The index associated with this is in column (c). But as noted, LTE-A enables Het-net development, thus facilitating accelerated growth of small LTE-A cells. This acceleration is assumed to double the annual growth of effective LTE-A cells to 16%. This is shown in columns (d), (e) and (f). If an additional 300 MHz of spectrum is supplied, this changes upwards the fraction of cell sites that will ultimately be LTE-A. Adjusted growth figures to represent this are provided in columns (g), (h) and (i).

EXHIBIT 12: Contributions of Additional Frequency Reuse to Mobile Capacity

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
Reuse w / Het-nets								
Year	Cell Sites	Cell Site Reuse Index	w / Currently Allocated Spectrum			w / Augmented Spectrum		
			4G+ Het-net Growth Augment	Augmented Effective Sites	Reuse w / Het-net Augment Index	4G+ Het-net Growth Augment	Augmented Effective Sites	Reuse w / Het-net Augment Index
2010	253,086	1.00						
2011	283,385	1.12						
2012	306,056	1.21						
2013	330,540	1.31						
2014	356,983	1.41						
2015	385,542	1.52	1.0891	388,794	1.54	1.0882	388,453	1.53
2016	416,386	1.65	1.0967	426,382	1.68	1.0981	426,556	1.69
2017	449,696	1.78	1.0997	468,913	1.85	1.1034	470,656	1.86
2018	485,672	1.92	1.1055	518,398	2.05	1.1111	522,928	2.07
2019	524,526	2.07	1.1114	576,134	2.28	1.1174	584,332	2.31
2020	566,488	2.24	1.1194	644,925	2.55	1.1252	657,471	2.60
2021	611,807	2.42	1.1289	728,049	2.88	1.1323	744,463	2.94
2022	660,752	2.61	1.1383	828,754	3.27	1.1413	849,636	3.36
			Baseline site CAGR: 8.0%					
			LTE-A Het-net site CAGR: 16.0%					

Notes: Assumption for effective reuse increase from LTE-A Het-net support is author’s estimate.

The final source of increased capacity is more effective network packing. As noted, we assume that QoS-based packet scheduling, carrier aggregation and VoLTE implemented in LTE permits a phased-in 20% improvement in LTE and LTE-A throughputs. This is shown in Exhibit 13 – assuming LTE and LTE-A prevalence both with and without additional spectrum.

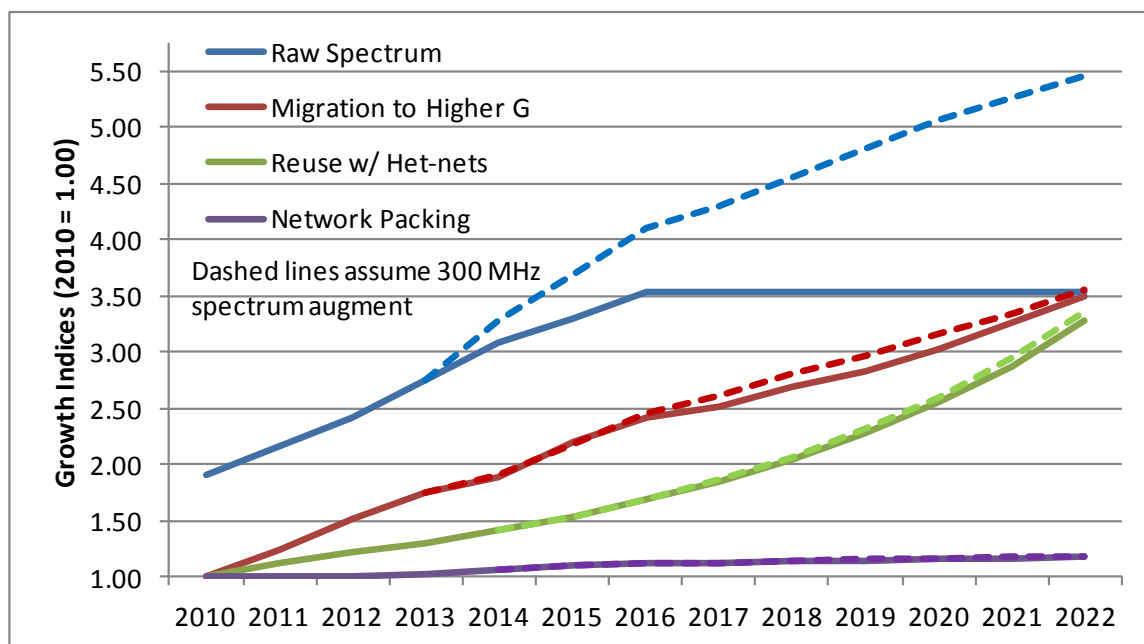
EXHIBIT 13: Contributions of Improved Network Packing to Mobile Capacity

(a)	(b)	(c)	(d)	(e)
	Network Packing			
	w/ Current Spectrum		w/ Augmented Spectrum	
Year	Adjusted to reflect % of 4G & 4G+ traffic	Network Packing Index	Adjusted to reflect % of 4G & 4G+ traffic	Network Packing Index
2010		1.00		
2011		1.00		
2012	1.0%	1.01		
2013	2.8%	1.03		
2014	5.9%	1.06	6.0%	
2015	10.0%	1.10	10.0%	1.10
2016	11.5%	1.12	11.8%	1.12
2017	12.3%	1.12	12.8%	1.13
2018	13.4%	1.13	14.0%	1.14
2019	14.3%	1.14	15.0%	1.15
2020	15.4%	1.15	16.3%	1.16
2021	16.9%	1.17	17.4%	1.17
2022	18.3%	1.18	18.7%	1.19
	Ultimate 4G & 4G+ packing efficiency lift: 20%			

Notes: Improvement in LTE and LTE-A network packing is assumed to be 5% in 2012, 10% in 2013, 15% in 2014 and 20% thereafter. Assumption for ultimate effect of QoS-based packet scheduling is author’s estimate based on VoLTE implementation, carrier aggregation and Yuksel et al. (2010).

Exhibit 14 shows the various contributions to capacity growth with solid lines assuming no additional spectrum and dashed lines assuming an additional 300 MHz of spectrum. As can be seen, migration to higher G technologies and site growth including Het-nets each have an individual effect of increasing total capacity by a factor of about 3.5 over the study period. These factors are roughly equal to the growth factor offered by completing build-out and filling of already allocated spectrum (assuming all EBS/BRS spectrum is usable). Adding an extra 300 MHz of spectrum will raise raw spectrum’s contribution to a factor of nearly 5.5.

EXHIBIT 14: Contributions to Mobile Capacity Growth



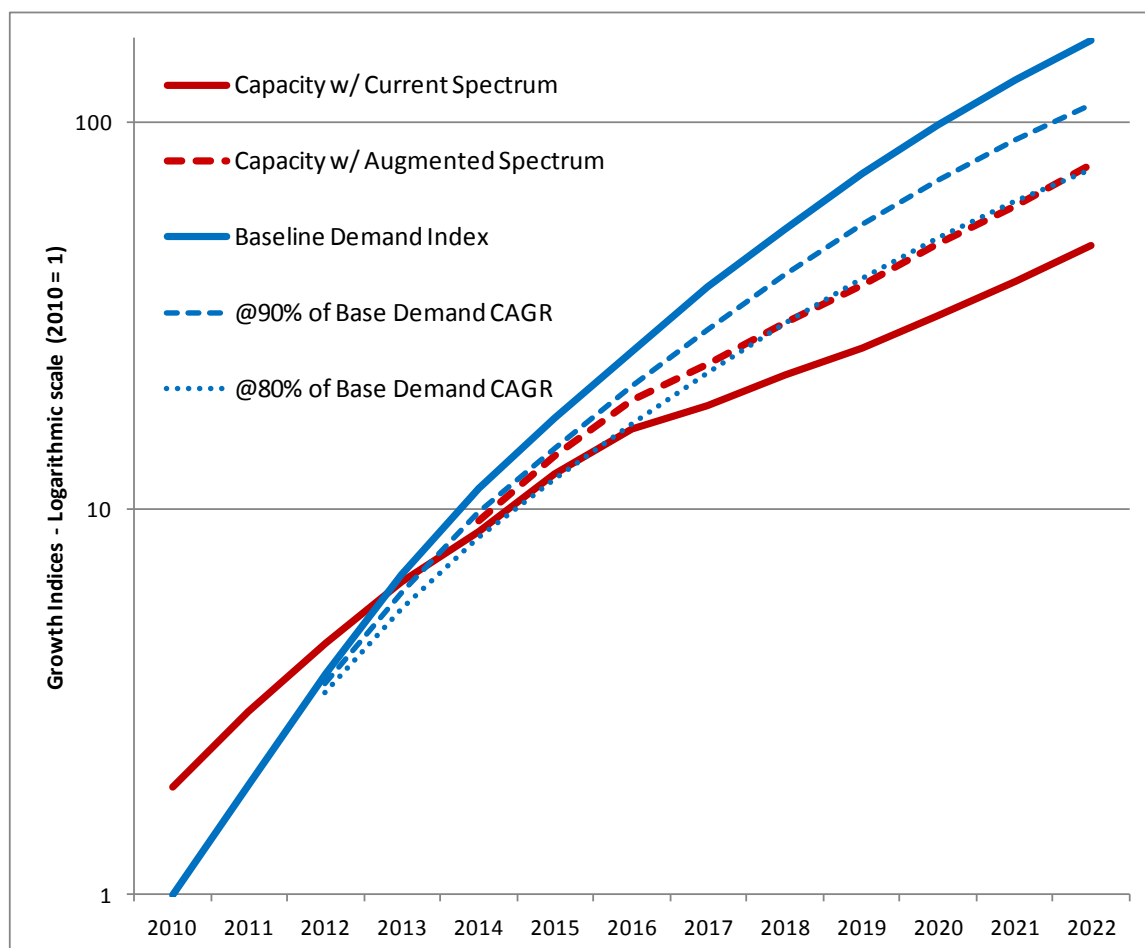
Notes: Solid lines assume no additional spectrum allocations. Dotted lines show contributions if 300 MHz of additional spectrum is allocated.

If we assume that the U.S. mobile industry will be fully successful at: (a) allocating all of the raw spectrum assumed; (b) intensively reusing this spectrum; (c) achieving all of the stated LTE and LTE-A efficiency lifts; and (d) quickly migrating customers to these higher G technologies; these multiple contributions may be combined and their joint lift on mobile capacity compared with forecasted traffic demand. This is done in Exhibit 15a. The solid red line shows capacity growth assuming only current spectrum allocations. The dashed red line shows this growth assuming an extra 300 MHz of spectrum. This capacity growth is in comparison to our baseline demand forecast for U.S. mobile traffic from Cisco (2012a), extrapolated for 2017-2022 and shown in solid blue. Two additional blue lines display more conservative variations on baseline demand: the dashed one assumes that growth rates are only 90% of the baseline growth forecast, and the dotted one assumes that these growth rates are only 80% of the baseline growth forecast.

As can be seen, under the baseline demand forecast, the U.S. goes into mobile capacity deficit by the end of 2013; and by 2017 baseline demand will be more than double available capacity. By 2022 it will be over triple available capacity. Under either the 90% or 80% of base growth demand forecasts, capacity deficits do not arrive until 2014 or 2015 and grow more slowly, but they remain severe. While beginning to add 300 MHz of additional spectrum in 2014 narrows slightly the capacity gap, even this is inadequate to keep the U.S.

out of deficit through the rest of the study period – no matter whether the baseline or 90% of base demand forecasts are used. Only if we assume that actual demand growth will be just 80% of base growth will rapidly adding 300 MHz of spectrum beginning in 2014 keep capacity in fairly close balance with demand over the study period.⁵² The index figures underlying Exhibit 15a are provided in Exhibit 15b.

EXHIBIT 15a: Comparing Mobile Capacity Growth to Demand Growth



Notes: Logarithmic scale.

⁵²Note that for this 80% demand forecast to be accurate, it requires Cisco’s baseline growth forecast to be overstated by a quarter for each year beginning in 2012 and running through 2022.

EXHIBIT 15b: Contributions to Mobile Capacity Growth

(a)	(b)	(c)	(d)	(e)	(f)
	Total Indices				
	Comparing Capacity Growth with Demand Growth				
Year	Capacity w/ Current Spectrum	Capacity w/ Augmented Spectrum	Baseline Demand Index	@90% of Base Demand CAGR	@80% of Base Demand CAGR
2010	1.90		1.00		
2011	2.98		1.91		
2012	4.47		3.71	3.53	3.35
2013	6.47		6.71	6.08	5.49
2014	8.72	9.32	11.21	9.74	8.40
2015	12.26	13.60	17.12	14.33	11.89
2016	16.07	19.07	25.58	20.67	16.55
2017	18.52	23.57	37.31	29.17	22.57
2018	22.02	30.24	53.03	40.20	30.13
2019	26.00	38.04	73.32	54.01	39.30
2020	31.40	48.58	98.53	70.69	50.06
2021	38.69	60.89	128.51	90.02	62.20
2022	47.80	77.57	162.53	111.44	75.33

While this analysis suggests that evolving demand for U.S. mobile wireless services is likely to be stymied by inadequate capacity growth, it is also useful to consider possible reasons why this may fail to occur – or if it does occur, how the market will equilibrate.

One possibility is that the forecasted demand figures are wrong and mobile service usage will not expand at the rates forecasted by Cisco. While this is certainly a possibility, it should be noted that each of Cisco’s last three updates of its Global Mobile Data Traffic Forecast have found actual Total Global Mobile Data Traffic to be larger than what its forecast from the previous year predicted.⁵³ Further, the sensitivities that we employed (assuming demand growth rates that are only 90% or 80% of baseline forecast growth rates) result in substantial reductions in forecasted traffic over the study period – but any reduction that falls short of the 53.65% traffic reduction implied by the 80% assumption will not be adequate to bring demand down to match even generously forecasted capacity growth.

Another possibility is that far greater load-shifting, and thus improved network packing, will take place than we suspect. To the extent that disproportionate amounts of new customer demand are for mobile services at times-of-day and in geographical locations where network capacities are not at their limit, it is possible that networks could absorb increased traffic without requiring proportional capacity reinforcements. Absent very granular network traffic data, it is impossible to know the extent of this possible mitigating effect.

⁵³ See Cisco (2009, 2010, 2011 and 2012a). Also remember that Qualcomm projects 1000x growth from 2010-2020 – far greater than the growth factors of 98.53x, 70.69x and 50.06x implicit in our baseline, 90% and 80% of Cisco VNI-based figures for 2010-2020. See note 20.

On the other hand, it is more likely that reductions in service quality or price increases will end up being the principal forces for equilibrating the market. When wireless capacities are tight, data connections will slow down. Either customers will accept slower performance of their mobile applications, or they will discontinue their use, or transfer their use to off-peak periods or locations. It is also possible that they will eschew particularly data-hungry applications in favor of less-desirable substitute applications that have the virtue of reduced data use. Either way, the effective service quality that customers receive will be reduced. Further, it is quite certain that prices will also be a major equilibrator of the market. This is because Cisco's demand forecasts are predicated on customer adoption and use trends that assume a continuation of today's price trajectory – which has been sharply declining prices per MB of traffic. As such, Cisco's *VNI* should be considered to provide forecasts of “notional” demand. To the extent that capacities go into deficit, this will attenuate, and possibly reverse, current downward price trajectories. If this occurs, future demands will be repressed to match more closely to available supply.

6. Conclusions

Expanding mobile wireless capacity in an economic and effective manner requires multifaceted efforts. The analysis presented in this paper has demonstrated the techniques that have been used to “deepen” mobile wireless networks. These include the allocation of additional spectrum; the development of more spectrally-efficient wireless technologies and the migration of customers to these technologies; increased reuse of available radio frequencies enabled both by cell site splitting and LTE-A support for enhanced small cell and Wi-Fi integration; and by tighter packing of offered data into available transmission capacity. But even though we have tried to be generous in our projections of capacity growth and cautious in our projections of demand growth, it appears that U.S. mobile wireless markets will face significant deficits in capacity from 2014 onward. In order to keep these deficits to manageable proportions, it will be essential for U.S. regulatory authorities to allocate quickly all of the 300 MHz of increased spectrum that was proposed in the *National Broadband Plan* for exclusive mobile use. If less than this sum is allocated, or if allocations are delayed until towards the end of our decade-long prospective study period, shortfalls will be especially severe. Indeed, in order to keep U.S. wireless markets on their current trajectory of virtuous growth, substantially more than 300 MHz needs to be added.

References

- Bazinet, Jason B. and Michael Rollins (2011), “Wireless Supply and Demand: Spectrum Control, Not Availability, Is the Real Constraint,” Citigroup Global Markets: Citi Investment Research and Analysis, September 22, 2011.
- Bhat, Prakash et al. (2012), “LTE-Advanced: An Operator Perspective,” *IEEE Communications Magazine*, Vol. 50, Issue 2, February 2012, pp. 104-114.
- Bode, Karl (2012), “It's Time to Stop Buying the Capacity Crisis Myth Like the 'Exaflood,' Looming Wireless Apocalypse a Lie,” DSLReports.com, January 27, 2012, available at: <http://www.dslreports.com/shownews/Its-Time-to-Stop-Buying-the-Capacity-Crisis-Myth-118099?nocomment=1>.
- Burstein, Dave (2011), “70-90% of AT&T Spectrum Capacity is Unused,” DSL Prime, March 22, 2011, available at: <http://www.dslprime.com/a-wireless-cloud/61-w/4193-70-90-of-atat-spectrum-capacity-unused>.
- Chapin, John M. and William H. Lehr (2011), “Mobile Broadband Growth, Spectrum Scarcity, and Sustainable Competition,” paper presented at *Telecommunications Policy Research Conference*, Arlington, VA, September 25, 2011, available at: <http://ssrn.com/abstract=1992423>.
- Chen, Jen et al. (2010), “Femtocells – Architecture & Network Aspects,” Qualcomm, January 28, 2010, available at: <http://www.qualcomm.com/media/documents/files/femtocells-architecture-network-aspects.pdf>.
- Chen, Brian X. (2012a), “Carriers Warn of Crisis in Mobile Spectrum,” *The New York Times*, April 17, 2012, available at: <http://www.nytimes.com/2012/04/18/technology/mobile-carriers-warn-of-spectrum-crisis-others-see-hyperbole.html?tnemail0=y&r=2emc=tnt&pagewanted=all>.
- Chen, Brian X. (2012b), “Q.&A.: Martin Cooper, Father of the Cellphone, on Spectrum Sharing,” *The New York Times*, May 31, 2012, available at: <http://bits.blogs.nytimes.com/2012/05/31/qa-marty-cooper-spectrum-sharing/>.
- Cisco (2009), “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update,” Appendix A, January 29, 2009, available at: <http://mobiletvworld.com/documents/Global%20Mobile%20Data%20Traffic%202009.pdf>.
- Cisco (2010), “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2009-2014,” Appendix A, February 9, 2010, available at: <http://theruckusroom.typepad.com/files/cisco-rmobile-trends-report.pdf>.
- Cisco (2011), “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010-2015,” Appendix A, February 1, 2011, available at: http://newsroom.cisco.com/ekits/Cisco_VNI_Global_Mobile_Data_Traffic_Forecast_2010_2015.pdf.
- Cisco (2012a), “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011-2016, Appendix A” and “General Questions About the Cisco Visual Networking Index,” February 14, 2012, available at: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf and http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI-Forecast_QA.html.

Cisco (2012b), “Cisco Visual Networking Index: The Zettabyte Era,” Appendix A, May 30, 2012, available at:

http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hype_reconnectivity_WP.pdf.

Credit Suisse (2011), “Global Wireless Capex Survey – A Multi-Year Spending Cycle,” July 2011.

Credit Suisse (2012), “A Closer Look at Spectrum Needs and the Impact on M&A,” Credit Suisse Securities Research & Analytics, June, 8 2012.

CTIA (2012a), “Semi-Annual Wireless Industry Survey, Year End 2011,” April 13, 2012, available at: http://files.ctia.org/pdf/CTIA_Survey_Year_End_2011_Graphics.pdf.

CTIA (2012b), “Wireless Industry Indices: Year-End 2011,” May 2012, available at: <http://www.ctia.org/advocacy/research/index.cfm/AID/10317>.

Ericsson (2012a), “Interim Update: Traffic and Market Data Report,” February 2012, available at: http://www.ericsson.com/res/docs/2012/tmd_report_feb_web.pdf.

Ericsson (2012b), “Traffic and Market Data Report,” June 2012, available at: http://www.ericsson.com/res/docs/2012/traffic_and_market_report_june_2012.pdf.

FCC (2010a), “Connecting America: The National Broadband Plan,” Federal Communications Commission, March 2010, available at: <http://www.broadband.gov/download-plan/>.

FCC (2010b), “Mobile Broadband: The Benefits of Additional Spectrum,” FCC Staff Technical Paper, October 2010, available at http://transition.fcc.gov/Daily_Releases/Daily_Business/2010/db1021/DOC-302324A1.pdf.

FCC (2011), “Fifteenth Annual Report and Analysis of Competitive Market Conditions With Respect to Mobile Wireless, Including Commercial Mobile Services,” Federal Communications Commission (WT Docket No. 10-133), June 27, 2011, available at: http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-11-103A1.pdf.

Feldman, Brett et al. (2011), “Coping With the Spectrum Crunch: Part 1,” Deutsche Bank Securities, U.S. Wireless Services, September 29, 2011.

Forouzan, Behrouz A. (2003), *Data Communications and Networking*, 3rd ed., McGraw–Hill, 2003.

Goldstein, Phil (2010), “Distributed Antenna Systems: From Niche to Necessity,” *Fierce Wireless*, March 4, 2010, available at: <http://www.fiercewireless.com/node/59880/print>.

Goldstein, Phil (2012), “Sprint CTO Embraces Small Cells, But Tempers Enthusiasm,” *Fierce Wireless*, September 21, 2012, available at: <http://www.fiercewireless.com/story/sprint-cto-embraces-small-cells-tempers-enthusiasm/2012-09-21>.

ITU (2010), “World Radiocommunication Seminar Highlights Future Communication Technologies,” International Telecommunications Union, December 6, 2010, available at: http://www.itu.int/net/pressoffice/press_releases/2010/48.aspx.

Korowajczuk, Leonhard (2011), *LTE, WiMax and WLAN: Network Design, Optimization and Performance Analysis*, John Wiley & Sons, 2011.

Lawson, Stephen (2008), “Most Analog Cellular To Fade Away on Monday,” *Infoworld*, February 14, 2008, available at: <http://www.infoworld.com/print/32423>.

Lawson, Stephen (2012), “11 Ways Around Using More Spectrum for Mobile Data,” *Computerworld*, August 16, 2012, available at: http://www.computerworld.com/s/article/9230345/11_ways_around_using_more_spectrum_f_or_mobile_data.

NTIA (2012), “An Assessment of the Viability of Accommodating Wireless Broadband in the 1755 – 1850 MHz Band,” National Telecommunications and Information Administration, March 2012, available at: http://www.ntia.doc.gov/files/ntia/publications/ntia_1755_1850_mhz_report_march2012.pdf.

OECD (2012), “Machine-to-Machine Communications: Connecting Billions of Devices,” OECD Digital Economy Papers, No. 192, OECD Publishing, January 30, 2012, available at: <http://dx.doi.org/10.1787/5k9gsh2gp043-en>.

Paolini, Monica (2011), “The Big Challenge for Small Cells: Backhaul,” *FierceBroadbandWireless*, November 9, 2011, available at: <http://www.fiercebroadbandwireless.com/story/paolini-big-challenge-small-cells-backhaul/2011-11-09>.

PCAST (2012), “Report to the President: Realizing the Full Potential of Government-Held Spectrum to Spur Economic Growth,” President’s Council of Advisors on Science and Technology, July 2012, available at: http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast_spectrum_report_final_july_20_2012.pdf.

Qualcomm (2012), “LTE Advanced,” February 2012, available at: <http://www.qualcomm.com/media/documents/files/lte-advanced-the-global-4g-solution.pdf>.

Real Wireless (2010), “Strategies for Mobile Network Capacity Expansion,” November 7, 2010, available at: <http://realwireless.files.wordpress.com/2010/11/strategies-for-mobile-network-capacity-expansion-v1-0.pdf>.

Reardon, Marguerite (2010), “Rethinking the Wireless Spectrum Crisis,” *CNET News*, May 25, 2010, available at: http://news.cnet.com/8301-30686_3-20005831-266.html.

Rysavy, Peter (2011), “Unleashing the Wireless Power of Long-Term Evolution: Spectrum, and Lots of It,” *BNA Daily Report for Executives*, The Bureau of National Affairs, December 19, 2011, available at: http://www.rysavy.com/Articles/2011_12_19_LTE_and_Spectrum.pdf.

Rysavy, Peter (2012), “No Silver Bullets for FCC, NTIA Spectrum Challenge,” *BNA Daily Report for Executives*, The Bureau of National Affairs, September 2012, available at: http://www.rysavy.com/Articles/2012_09_No_Spectrum_Silver_Bullets.pdf.

Rysavy Research (2002), “Voice Capacity Enhancements for GSM Evolution to UMTS,” July 18, 2002, available at: http://www.rysavy.com/Articles/GSM_voice_capacity_71802.pdf.

Rysavy Research (2010), “Spectrum Shortfall Consequences,” April 21, 2010, available at: http://www.rysavy.com/Articles/2010_04_Rysavy_Spectrum_Shortfall_Filing.pdf.

Rysavy Research (2011a), “The Spectrum Imperative: Mobile Broadband Spectrum and Its Impacts for U.S. Consumers and the Economy -- An Engineering Analysis,” March 16, 2011, available at: http://www.rysavy.com/Articles/2011_03_Spectrum_Effects.pdf.

Rysavy Research (2011b), “Efficient Use of Spectrum,” May 4, 2011, available at: http://www.rysavy.com/Articles/2011_05_Rysavy_Efficient_Use_Spectrum.pdf.

Rysavy Research (2011c), “Comments on Citi ‘Wireless Supply and Demand,’” October 13, 2011, available at: http://www.rysavy.com/Articles/2011_10_13c_Rysavy_Citi_Comments.pdf.

Rysavy Research (2012), “Mobile Broadband Explosion: The 3GPP Wireless Evolution,” August 2012, available at:
http://www.rysavy.com/Articles/2012_09_Mobile_Broadband_Explosion.pdf.

Seybold, Andrew M. (2011), “Broadband 101 and Real-World Tests Results,” 2011, available at: <http://www.foxcomm-wi.com/system/files/Alameda9-15Presentation%2520revised%2520final.pdf>.

Williams, Mark D. J. et al. (2011), “Africa’s ICT Infrastructure: Building on the Mobile Revolution,” The World Bank, 2011, available at:
http://siteresources.worldbank.org/INFORMATIONANDCOMMUNICATIONANDTECHNOLOGIES/Resources/AfricasICTInfrastructure_Building_on_MobileRevolution_2011.pdf.

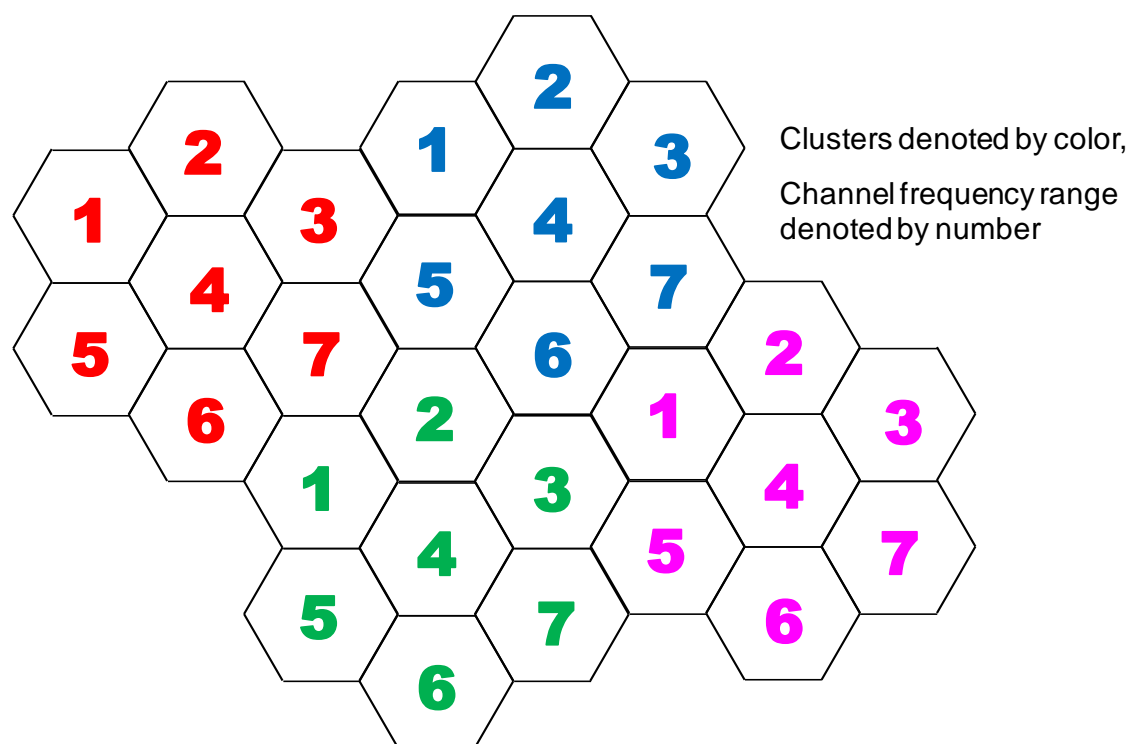
Yuksel, Murat et al. (2010), “Quantifying Overprovisioning vs. Class-of-Service: Informing the Net Neutrality Debate,” *Proceedings of IEEE International Conference on Computer Communication Networks*, pp. 1-8, Zurich, Switzerland, August 2010, available at:
<http://www.cse.unr.edu/~yuksemy-papers/iccn10.pdf>.

Appendix on Technological Evolution

First generation technology

Widespread introduction of mobile wireless services first occurred in advanced countries in the 1980s. The First Generation (1G) technology employed in the U.S. in 1983 was called Advanced Mobile Phone System (AMPS) and was deployed in radio spectrum located in the 850 MHz band. This analog technology was designed specifically to carry voice radiotelephone calls. A network's total bandwidth was split, using Frequency Division Multiplexing (FDM), into separate 30 kHz channels – with these channels divided among radio cells so that no cell used the same channel as adjacent cell. This commonly resulted in an array of cells where each cell used only 1/7th of the total number of channels. This is displayed in Exhibit A1.

EXHIBIT A1: AMPS 1:7 Frequency Reuse



AMPS technology used frequency modulation (FM) signal encoding to carry a single direction (uplink or downlink) of a voice call within each 30 kHz channel, thus two unidirectional channels were required to create a duplex channel (60 kHz total) necessary for

a two-way voice communications link.⁵⁴ Using a technique called Frequency Division Multiple Access (FDMA), these duplex channels could be separately assigned to individual subscribers within the range of the radio tower who were seeking to place or receive a call – so long as all channels were not occupied in handling the calls of other subscribers.⁵⁵ While it was possible for these channels to be used to carry data signals, it was cumbersome. In general, the subscriber needed to obtain a separate analog wireless data modem that would modulate his data device’s digital signals to analog, which could then be sent out over the voice channel at a data rate that could not exceed 9.6 kbps, but due to network error correction was usually significantly lower than this figure.

Within the basic AMPS framework, there were basically only two ways to deepen network capacity: increase spectrum reuse by building more radio towers or antennas; or by increasing the total amount of spectrum usable for mobile communications. Both of these were pursued. In 1989, the U.S. Federal Communications Commission (FCC) increased the size of the frequency band available for AMPS from 40 MHz to 50 MHz. In addition, more radio towers were built to subdivide cells in dense-use areas and increase spectrum reuse.⁵⁶ But in the end, exploding customer demand for mobile wireless services required 1G analog systems to be succeeded by more efficient Second Generation (2G) systems that used digital technologies to enable increased traffic capacity per unit of spectrum.

Second generation technology

By the early 1990s, it was clear both that demand for mobile wireless services was not going to flag anytime soon and spectrum adequate for analog networks to accommodate this growing demand was not likely to be available. Further, digital wireless technology had developed such that much more efficient 2G systems were now feasible. Instead of using FDM-created 30 kHz channels to carry a single voice signal in analog format, it was possible

⁵⁴ The total of 50 MHz of 850 MHz spectrum eventually allocated to A-block and B-block cellular operators in the U.S. was divided into 832 duplex channels (30 kHz uplink and 30 kHz downlink). Roughly 42 of these channels were used for network control purposes, with the remaining 790 available to carry voice signals. But due to interference issues, spectrum used to provide service in one cell could not be reused in adjacent cells. This was handled by distributing cells in a roughly hexagonal array (see Exhibit A1), and generally only allowing 1/7th of the total spectrum to be used within each cell. This restricted U.S. cellular networks to carrying (on average) only 56 simultaneous voice calls per radio cell.

⁵⁵ “Multiple Access” describes how the segmented spectrum resource is shared among users. FDMA shares these 30 kHz channels among users by using radio tuners in user devices to “tune” individually to the assigned channel – similar to how an FM radio or television tuner operates.

⁵⁶ In certain geographies it was also possible to improve the reuse factor by reducing separation between cells using the same frequencies from two buffer cells to one buffer cell. This increased the amount of spectrum capable of being deployed in each cell from 1/7th of the total available spectrum to 1/4th.

to combine FDM multiplexing technique with either Time Division Multiple Access (TDMA) or Code Division Multiple Access (CDMA) and use digital encoding of the voice signal to carry several voice signals within an equivalent amount of spectrum.

In general, there were three ways in which this was accomplished. One method, which had a relatively short lifetime, was to use TDMA to divide AMPS' 30 kHz frequency channels into three time slots, and to use digital compression of voice waveforms to place a separate voice signal within each time slot. This method was called, variously, Digital AMPS (D-AMPS), IS-54 or IS-136.⁵⁷ Because three voice signals could be carried within each 30 kHz of spectrum, this 2G technology was three times more spectrally efficient than AMPS (using only 10 kHz per voice signal) and permitted roughly 169 simultaneous two-way voice to be carried within each radio cell.

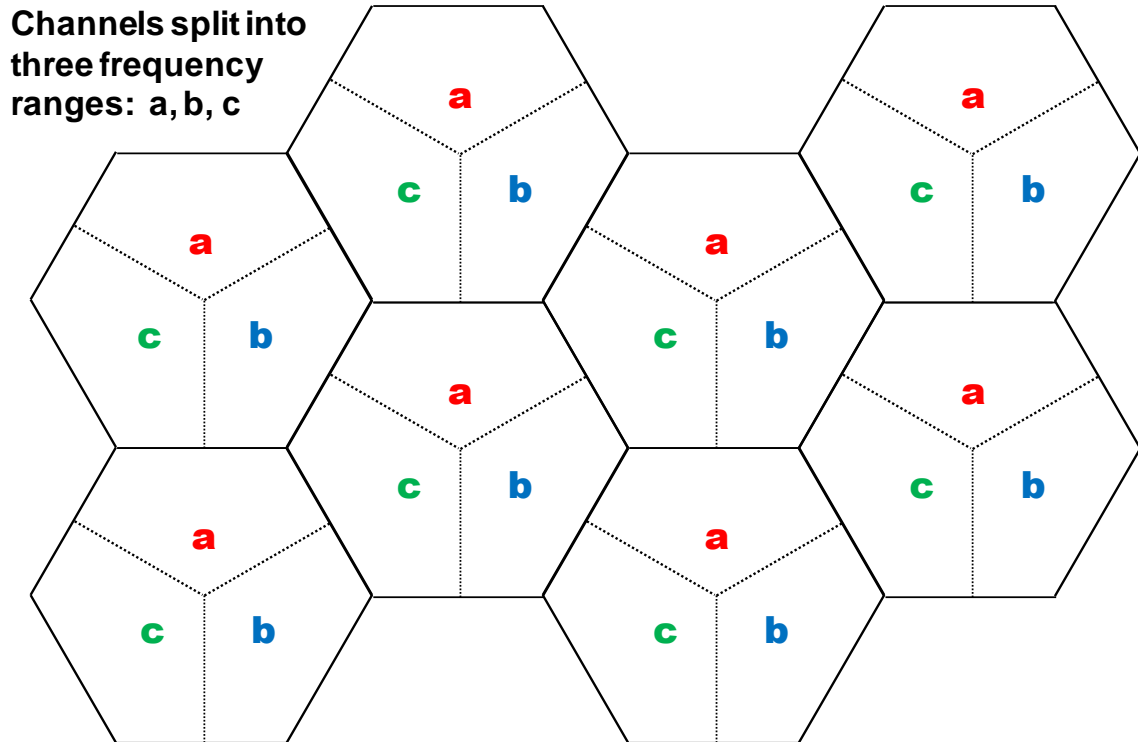
A second method, initially developed in Europe and called GSM,⁵⁸ used FDM to create larger frequency channels of 200 kHz, and used TDMA to divide these into 8 time slots, each holding a separate digitally-compressed voice signal. While this may have seemed to create only a small uplift in carrying capacity as carrying 8 voice signals in 200 kHz only reduced a signal's spectrum occupancy from 30 kHz under AMPS to 25 kHz ($= 200 \text{ kHz} / 8$) under GSM, there was an additional significant benefit. Because GSM's digital signals were less prone to interfering with each other, it was possible to increase the reuse of frequency bands from 1:7 employed by AMPS. This was done, generally, by eliminating the requirement that adjacent radio towers had to use completely different frequencies. Rather, radio towers could be sectored by adding additional antennas – and dividing the total frequencies available to the system separately among the sectors. Thus, rather than using a single omnidirectional antenna that broadcast all of the frequencies available at the cell site out to its complete extent, three directional antennas could be mounted on a tower – with each broadcasting one-third of the total number of available frequencies. This is shown in Exhibit A2. But because each frequency potentially could be reused on an adjacent tower (so long as it was used in a sector that was not beamed directly at the first tower), effective

⁵⁷ This paper will use the term “D-AMPS” as a shorthand to refer to all three of these technologies. In the U.S., this technology was adopted by AT&T Wireless, Southwestern Bell Mobile Systems and BellSouth Mobility in the mid-1990s.

⁵⁸ GSM initially stood for *Groupe Spécial Mobile*. It has since been renamed *Global System for Mobile Communications*. It has become the most widely adopted mobile wireless standard around the world. In the U.S., this technology was adopted initially in the mid-1990s by a new entrant company called Onmipoint (an operator that eventually became T-Mobile USA). It was later adopted by the major operators that had previously chosen D-AMPS – AT&T Wireless, Southwestern Bell Mobile Systems and BellSouth Mobility – a group of companies that later became known as Cingular Wireless and are now known simply as AT&T.

frequency reuse was now 1:3. This enabled GSM cells potentially to carry 330 simultaneous two-way voice calls.⁵⁹

EXHIBIT A2: Sectorization Example with 1:3 Frequency Reuse



The third major 2G technology combined FDM-created broad 1.25 MHz frequency channels with CDMA within these frequency channels to carry multiple voice signals. This technology was called IS-95, or more commonly just CDMA.⁶⁰ Rather than splitting the frequency channel based on different time slots to carry multiple voice signals, CDMA assigns each signal a separate identification code which is used by the terminals to distinguish the signals associated with one voice call versus another. Each 1.25 MHz frequency channel may have up to 64 different codes assigned to it, 9 for control purposes and 55 for voice signals. This results in each direction of a voice call requiring as little as 22 kHz of frequency spectrum ($22 \text{ kHz} = 1.25 \text{ MHz} / 55$). But because codes may also be used to differentiate between signals emanating from different radio towers, it is possible to have 1:1 frequency reuse. This means that 50 MHz of spectrum can theoretically support up to 1100

⁵⁹ This can be calculated by dividing 50 MHz of total spectrum into 248 separate one-way 200 kHz channels. But with 1:3 reuse, this reduces to 83 channels per tower. Each of these channels could carry 8 call signals, but since two (uplink and downlink) are necessary for a single call, this yields about 330 simultaneous calls within a cell.

⁶⁰ The name trademarked for this technology is *cdmaOne*. In the U.S., the major wireless carriers employing this technology are Verizon and Sprint.

simultaneous two-way voice calls within a single cell.⁶¹ While this might suggest that 2G CDMA technology is more spectrally efficient than either of the TDMA-based 2G technologies, this is not always assured and comes at some additional costs. Among the most well-known of these issues is that engineering CDMA networks is more complex because as the number of simultaneous callers within a cell rises, transmit power levels need to be reduced to limit interference between call signals. This causes the radius of the area that the cell tower can serve to decrease with increased use and increase with decreased use.⁶² Further, because handsets need to “listen” to and parse broad 1.25 MHz signals in order to extract and decode the particular voice signal frames destined for them, the handset’s chipset processing is complex, and its electrical demands higher and battery life more limited than with other 2G technologies.⁶³

To review, by harnessing digital technology 2G wireless systems were able to carry more voice conversations within a given amount of radio spectrum than 1G analog technologies. They did this by overlaying an additional multiplexing/multiple access technique (either TDMA or CDMA) on top of basic FDM and by digitally compressing voice to fit into a smaller spectrum “space.” Additionally, because digital signals are more immune to interference than analog signals, these 2G technologies could increase reuse of frequencies within a given geographic area. These enhancements “deepened” greatly the overall capacity of 2G networks relative to 1G networks. But a further capability of 2G networks was their ability to handle data transmissions more gracefully than 1G networks. In 1G networks, the only way to handle data was to use an auxiliary modem to convert a digital data signal into an analog voice waveform. In contrast, because 2G networks encoded voice waveforms into digital signals, it was simpler for these networks to carry data directly.

It is also useful to consider the potential quantities of data that 2G wireless services could carry. Whether it was possible to use the full frequency slices created by each 2G technology to carry data depended on that technology’s evolution. Early technical

⁶¹ This can be calculated by dividing 50 MHz of total spectrum into 40 separate one-way 1.25 MHz channels. Each of these channels can carry up to 55 call signals, but since two of these (uplink and downlink) are necessary for a single call, this yields up to 1100 simultaneous calls within a cell. Note, though, that CDMA voice quality degrades as the number of simultaneous calls rises – so that in the real world a 1.25 MHz CDMA channel would never actually approach handling its theoretic maximum of 55 calls. See Rysavy Research (2002), p. 4 and Seybold (2011), p. 9.

⁶² The term used to describe this phenomenon is “cell breathing.”

⁶³ Other factors also served to attenuate the popularity of CDMA. These included the fact that its initial specifications did not support text messaging (a popular revenue-generating service), and the patents associated with CDMA were held by a single manufacturer, Qualcomm, which concerned many wireless operators who worried about supply costs and lack of secondary sources. See <http://www.fundinguniverse.com/company-histories/qualcomm-incorporated-history/>.

specifications allowed only single 2G voice channels to be used for data, limiting potential throughputs to between 9.6 and 14.4 kbps. As these 2G technologies improved to what became known as 2.5G, it became possible to use complete FDM-created frequency slices for data – yielding a potential gross D-AMPS throughput of 48.6 kbps over 30 kHz of spectrum (1.62 bps/Hz); a potential gross GSM throughput of 270.8 kbps over 200 kHz of spectrum (1.35 bps/Hz); and a potential gross CDMA throughput of 1228.8 kbps over 1228.8 kHz of spectrum (1.00 bps/Hz). But due to overheads from various sources (e.g., framing, synchronization and FEC, etc.) and other technical operating concerns and conditions, typical levels of data throughput (goodput) that 2.5G technologies could offer customers within a sector were in the range of 50 kbps to 100 kbps for both GSM and CDMA – depending on the technical status of the network.⁶⁴

Third generation technology

While the data speeds and throughput capacities of 2G and 2.5G systems were adequate for small ASCII monochrome screen applications (such as Blackberry email), or communications with websites designed specially to interact with limited capability mobile devices, user experiences with non-specialized websites or employing devices with high resolution displays were likely to be unsatisfying. Over the course of the mid to later 2000s, U.S. carriers began to introduce 3G networks that could increase data speeds by an order of magnitude from what 2.5G customers experienced. Both Sprint and Verizon rolled out cdma2000 networks, while AT&T and T-Mobile deployed Universal Mobile Telecommunications System (UMTS) networks that followed the GSM evolution.⁶⁵ Initially, the take-up of mobile data services was modest, but then in 2007 the Apple iPhone was introduced and device competitors quickly followed with their own smartphones. Now tens of millions of customers had devices with advanced operating systems and high resolution color screens that could interact with general websites. The only operator solution was to quickly complete deployment of upgraded 3G networks and to deepen them, fast.

UMTS was a radical change from its GSM ancestor that employed a combination of FDM to create multiple 200 kHz frequency blocks and TDMA to divide these into multiple voice channels. Instead, UMTS used FDM to divide available spectrum into large 5 MHz

⁶⁴ Because D-AMPS networks generally began to be phased out in the U.S. in favor of GSM during the early 2000s, this paper does not consider them further.

⁶⁵ UMTS was a creation of the *3rd Generation Partnership Project* (3GPP) – a group of telecommunications associations seeking to evolve GSM into a more capable mobile system. See <http://www.3gpp.org/>. cdma2000 was created by the *CDMA Development Group* (CDG), primarily by member Qualcomm. See <http://www.cdg.org/>.

frequency blocks, but then used wideband CDMA (WCDMA) to create its multiple access communications channels within these blocks. While initially this technology could support only a potential 384 kbps downlink channel, later High Speed Downlink Packet Access, High Speed Uplink Packet Access and High Speed Packet Access implementations (HSDPA, HSUPA, HSPA) were able support potential downlink traffic rates from 7 to 14 Mbps within these 5 MHz carrier blocks – and a 4G implementation called HSPA+ is able to support downlink rates up to 21 Mbps.⁶⁶ These higher rates are achieved both by evolving the modulation protocols used for UMTS to ones with greater spectral efficiency as well as from implementing more efficient transmission control protocols.⁶⁷

3G cdma2000 was designed to be a less radical evolution from 2G cdmaOne than UMTS was from GSM. cdma2000 retained the same 1.25 MHz FDM-created blocks as 2G cdmaOne technology but employed a more efficient implementation of CDMA to place 128 separate voice channels within each 1.25 MHz carrier. While the initial 1xRTT implementation of cdma2000 was only capable of handling packet speeds up to 144 kbps, this was upgraded substantially when the 1xEV-DO (Evolution-Data Optimized) variant was installed.⁶⁸ The EV-DO upgrade dedicated separate 1.25 MHz carriers solely to transport data packets (no voice). This, coupled with the use of more spectrally-efficient 16QAM modulation (rather than QPSK used with 1xRTT) enabled theoretic data rates for EV-DO, Rev. A to reach about 3.1 mbps in the downlink.⁶⁹

⁶⁶ While 4G HSPA+ technology has spectral efficiency similar to that of 3G technologies, because of its higher performance capabilities it has been recognized as a 4G technology by the International Telecommunications Union (ITU, 2010). Note that future enhancements to 4G HSPA+ may be able to push data rates even above 21 Mbps. It is unclear, however, whether all of these enhancements will be implemented. But for the purposes of this paper, which focuses on the efficiency with which the different technologies carry volumes of data within a given bandwidth (and not their speed), we list 4G HSPA+ in the “3G” efficiency category.

Actually achieved downlink rates generally are generally only about a third of potential rates. Potential uplink data rates are generally only about a quarter of these downlink rates. See Rysavy Research (2012), pp. 22-23, 41-45.

⁶⁷ UMTS initially used QPSK (quaternary phase shift keying) modulation that could encode 2 bits per transmitted symbol. HSPA evolutions used 16QAM and 64QAM (quadrature amplitude modulation) modulation to encode 4 or 6 bits per symbol, respectively.

⁶⁸ Because of the relatively slow data speed of 1xRTT, some observers have termed cdma2000 1xRTT to be a 2.5G or 2.75G technology – not achieving full 3G status until 1xEV-DO was implemented. See Rysavy Research (2012), p. 16.

⁶⁹ Uplink channels, which continued to use more robust QPSK modulation, were limited to about half the speed of the downlink. EV-DO, Rev. A is the 3G technology currently used by both Verizon Wireless and Sprint in the U.S.

Fourth generation LTE technology

It soon became clear that advanced mobile devices were only going to become more functional and would become increasingly larger consumers of mobile data services. Further, it was also evident that vast increases in wireless spectrum would not be forthcoming, so there would be a critical need for still another technology evolution. This newer technology needed to be more spectrally-efficient in wringing increased data carrying capacity from each MHz of spectrum, and needed both to have lower roundtrip latency to support two-way interactive applications as well as be able to support the extremely high data speeds required by advanced (largely video) applications displayed on mobile screens of increasingly high resolution.⁷⁰

Thus, in late 2008 the 3GPP began issuing specifications for a 4G technology known commonly as Long Term Evolution (LTE).⁷¹ This technology is intended to be flexible in its use of different size frequency blocks (with wider blocks permitting increasingly fast transmission speeds), to be significantly more spectrally efficient than earlier technologies and to offer substantially less latency to better support real time interactive data applications – including applications such as Voice over Internet Protocol (VoIP). Indeed, LTE’s specifications are robust enough to permit it to handle seamlessly all evolving use applications and to match the speed performance of most wireline broadband services. To achieve these goals, it would be necessary for LTE to:

- Be a data-optimized packet technology without separate or extraneous overheads and protocols designed specifically for the carriage of voice services;
- Adopt a new multiple access technology called Orthogonal Frequency Division Multiple Access (OFDMA);
- Support Multiple Input Multiple Output (MIMO) technology that employs multiple antennas on cell towers and within user devices to enable two or more parallel radio links connecting a tower to a user device;

⁷⁰ The initial Apple iPhone had a screen resolution of 320x480 pixels. The iPhone 4 with its Retina display has a resolution of 640x960 pixels – a fourfold increase. But even these resolutions pale relative to that offered by the iPad. iPads 1 and 2 both had screen resolutions of 1024x768 pixels. But with the iPad 3, this has grown far larger to a resolution of 2048x1536 pixels – which exceeds even the 1920x1080 pixel resolution of HDTV.

⁷¹ Another 4G technology, called *WiMax*, was introduced several years prior to LTE. While it was adopted by Clearwire in the U.S., it failed to gain sufficient worldwide acceptance, and is now being discontinued in favor of LTE. Qualcomm also pursued a 4G cdma2000 successor technology called *Ultra Mobile Broadband* (UMB), but abandoned this effort to focus on LTE.

- Operate over a range of several different-sized channel bandwidths (from 1.4 MHz to 20 MHz), and also to operate over channels that are asymmetrically sized between uplink and downlink.⁷²

Section 4 of this paper discusses in greater depth some of the several techniques that LTE employs to carry the greater amounts of traffic that customers are expected to offer to mobile networks. While each of these techniques is very promising and may make a substantial contribution to mobile network capacity, even together they will likely fail to expand capacity at a rate that stays ahead of anticipated levels of demand growth.

⁷² Because downlink data flows are generally much heavier than uplink flows, it may be more efficient to dedicate a greater portion of available spectrum to downlink channels than uplink channels. This is in contrast to the equal division of spectrum between downlink and uplink transmission channels that has been characteristic of earlier mobile wireless technologies.