NTIA Special Publication SP-03-401

Proceedings of the International Symposium on Advanced Radio Technologies March 4 - 7, 2003

J. Wayde Allen, General Chair
Timothy X Brown, Technical Program Chair

U.S. DEPARTMENT OF COMMERCE
Donald L. Evans, Secretary

Michael D. Gallagher, Deputy Assistant Secretary
for Communications and Information

March 2003

INTERNATIONAL SYMPOSIUM
ON
ADVANCED RADIO TECHNOLOGIES

Sponsored by:

IEEE Communications Society
IEEE Vehicular Technology Society

Institute for Telecommunication Sciences (ITS),
National Telecommunication and Information Adm. (NTIA)

National Institute of Standards and Technology (NIST)

University of Colorado, Department of Interdisciplinary Telecommunications

U.S. Department of Commerce, Boulder Laboratories

2003 ISART Technical Committee:

Wayde Allen, General Chair, NTIA/ITS

Prof. Timothy X Brown, Program Chair, University of Colorado

Ernst Bonek, Technical University of Vienna

Prof. Dale Hatfield, University of Colorado

Dewayne Hendricks, CEO, Dandin Group

Robert Matheson, NTIA/ITS

Dr. David Reed, Ind. Researcher & Consultant; Visiting Scientist, MIT

DISCLAIMER

Certain commercial equipment, components, and software are identified to adequately present the underlying premises herein. In no case does such identification imply recommendation or endorsement by the National Telecommunications and Information Administration, nor does it imply that the equipment, components, or software identified is the best available for the particular applications or uses.

PREFACE

The International Symposium on Advanced Radio Technologies (ISART) originated as a forum to explore the current state of the radio art and to identify directions in which the technology would likely evolve in the future. While the early ISART meetings tended to focus on specific areas of technology, it is clear that an understanding of the big picture requires a more inclusive point of view. After all, the development and deployment of radio technology is not driven solely by technical know-how. Certainly understanding the underlying technology clarifies what is theoretically possible, but the development of new products and services also depends on what existing policies and regulations allow, as well as what makes economic sense. Similarly, it would be a critical mistake to make business decisions or enact new regulations without a solid understanding of the radio science.

With this in mind, the ISART conferences have evolved to provide a more holistic view of the radio art. Rather than providing a purely technical focus, ISART now strives to bring together a diverse collection of people from academia, business, and government agencies to discuss these issues in a common forum. Assembling this group of business experts, technologists, and government regulators so they can share their points of view and debate issues during round table discussions makes a deeper exploration of the intricacies of the radio art possible.

Two specific changes were made to the ISART program this year. The first of these was adding a fourth day to the meeting. This was done to include more discussion on topics related to spectrum usage, management, policy and regulation. There are a number of controversial topics that will undoubtedly be discussed during these sessions. One of these is the question of how to create regulations that support the deployment of new technology such as ultrawideband. Another is the debate about whether the spectrum is a finite resource that should be treated as private property or if instead it should be open for relatively unregulated public use.

The other change establishes ISART as a place for speakers to publish papers. The new "ISART Proceedings" gives the radio community a resource in the form of an NTIA/ITS special publication where ideas spanning a number of disciplines can be recorded for peer review and reference. Not only does this allow speakers to provide more detail describing theoretical concepts, data, and/or methodology than is possible in an oral presentation, the call-for-paper process also gives the radio community a way to suggest presentation topics on subjects that the ISART Technical Committee might otherwise not have known about. Publication of this first issue of the ISART Proceedings is just the next step in growing ISART as a valuable community resource supporting the future of radio.

J. Wayde Allen, NTIA/ITS
General Chair

2003 ISART Symposium Agenda

March 4 – 7, 2003

Tuesday, March 4, 2003

8 am **Registration, Coffee, and Pastries**
9:00 **Welcome**
Al Vincent, Director, NTIA/ITS
9:15 **Keynote Address**
Nancy J. Victory, Assistant Secretary for Communications and Information, NTIA

Session I: Spectrum Usage and Management, Frederick Matos, Chair, NTIA
10:00 Paul Kolodzy, Federal Communications Commission
10:30 **Break**
10:45 Johnathon D. Blake, Covington and Burling
11:15 *Presentation of the People's Radio Spectrum Chart*,
Jim Snider, New America Foundation
11:45 Round Table Discussion
12:15 **Lunch**

Session II: Spectrum Policy and Regulation, Joe Gattuso, Chair, NTIA/OPAD
1:45 *Spectrum Policy Challenges for Industry and Policymakers*,
Diane Cornell, Cellular Telecommunications and Internet Association
2:15 *Looking to Spectrum for Network Utopia*,
Dewayne Hendricks, Dandin Group
2:45 **Break**
3:00 *Spectrum Buyouts – A Proposal for the Transition to Open Spectrum*,
Nobuo Ikeda, Research Institute of Economy, Trade and Industry, Japan
3:30 Bits Aren't Bites: Constructing a "Communications Ether" that can grow and adapt
David Reed, MIT Media Laboratory
4:00 *The Electrospace Model as a Tool for Spectrum Management*,
Robert Matheson, NTIA/ITS
4:30 Round Table Discussion

Wednesday, March 5, 2003

8 am **Coffee and Pastries**

Session III: Telecommunication in Developing Countries,
Dr. Robert Stafford, Chair, NTIA/ITS
9:00 Monique Maddy
9:30 *Mobile Paradigm in Developing Countries*,
Mark Jamison, University of Florida
10:00 **Break**
10:15 *Distance Education Everywhere (not "online" learning): Myths, Realities and Possibilities*,
Dr. Janet Poley, American Distance Education Consortium

10:45 *3G Licensing in Hong Kong: A Unique Approach*
Xu Yan, Hong Kong University of Science and Technology
11:15 *International Wireless Policies*, Sharon Black, CU Boulder
11:45 Round Table Discussion
12:15 **Lunch**

Session IV: Software Defined Radios, Brad Ramsey, Chair, RF Metrics Corporation
1:45 *Adaptive Spectrum Radio: A Feasibility Platform On The Path To Dynamic Spectrum Access*,
William D. Horne, Peter Weed, Daniel J. Schaefer, The MITRE Corporation
2:15 *Efficient Collaborative (Viral) Communication in OFDM Based WLANs*,
Aggelos Bletsas, Andrew Lippman, MIT Media Laboratory
2:45 *Frequency Selective IQ Phase and IQ Amplitude Imbalance Adjustments for OFDM Direct Conversion Transmitters*,
Edmund Coersmeier, Ernst Zielinski, Nokia Research Center, Germany
3:15 **Break**
3:30 *High-Efficiency, High-Gain Power Amplification for PCS*,
Dr. Chance M. Glenn, Ph.D, Syncrodyne Systems Corporation, Columbia, MD
4:00 *System Architecture for a Dynamic-Spectrum Radio*,
Allen Petrin, Paul Steffes, Georgia Institute of Technology
4:30 Round Table Discussion

Thursday, March 6, 2003
8 am **Coffee and Pastries**

Session V: Wireless Networking; Robert Achatz, Chair, NTIA/ITS
9:00 William Lehr, Massachusetts Institute of Technology
9:30 *Empirical Study of 802.11b Wireless Networks in the Presence of Bluetooth Interference*,
Cameron McKay, Fukumasa Masuda, NTIA/ITS
10:00 **Break**
10:15 *Relieving Spectrum Scarcity Through Real-Time Secondary Markets*,
John Peha, Carnegie Mellon University
10:45 *Stretchable Architectures for Next Generation Cellular Networks*,
S. Lakkavalli, A. Negi, S. Singh, Portland State University
11:15 *Standards Convergence*,
Dave Evans, Philips Research Laboratories, UK
11:45 Round Table Discussion
12:15 **Lunch**

Session VI: Antennas and Propagation, Katie MacReynolds, Chair, NIST
1:45 *On Coherency Requirements in Steered Beam Adaptive Antenna Systems*,
Sven Petersson, Aare Mällo, Ericsson
2:15 *Dependence of Radar Emission Spectra on Measurement Bandwidth and Implications for Compliance with Emission Mask Criteria*,
Frank Sanders, NTIA/ITS
2:45 *Comparison and Simulation of Digital Modulation Recognition Algorithms*,

Wei Su, John Kosinski, U.S. Army Research Development and Engineering Command, Intelligence and Information Warfare Directorate, Fort Monmouth, NJ

3:15 **Break**

3:30 *Evaluation of Two Site-Specific Radio Propagation Models*, Alakananda Paul, Paul McKenna, Frederick Najmy, NTIA

4:00 *Broadband Wireless Low-Power Transmission*, Joseph Hagerty, Zoya Popovic, University of Colorado, Boulder

4:30 Round Table Discussion

Friday, March 7, 2003

8 am **Coffee and Pastries**

Session VII: Wireless System Safety and Security, Glenn Ashton, Chair, NETdelivery

9:00 *Ad hoc Network Security*, David Johnson, Dept. of Computer Science, Rice University, Houston.

9:30 Seth Goldstein, Roving Planet

10:00 **Break**

10:15 *Public Safety and Emergency Services in Future Wireless Communication*, Douglas C. Sicker, Interdisciplinary Telecommunications Dept., Dept. of Computer Science, CU Boulder

10:45 Brian Levine, University of Massachusetts

11:15 *Anonymity in Wireless Services*, Dirk Grunwald, Dept. of Computer Science, CU Boulder

11:45 Round Table Discussion

12:15 **Lunch**

Session VIII: Wireless Broadband and 3G, Robert Matheson, Chair, NTIA/ITS

1:45 *Household Demand for Wireless Telephony: An Empirical Analysis*, Paul Rappaport, Temple University and James Alleman, CU Boulder

2:15 *LoL@ - A Prototype of a Network Independent Wireless Internet Service*, Harald Kunczier, Elke Michlmayr, Günther Pospischil, Hermann Anegg, Forschungszentrum Telekommunikation Wien, Austria

2:45 *Feasibility of Fixed Wireless Access*, Ali Eftekhari, R. Ryan Lewis, Scott Pollock, Tom Lookabaugh, Bryan Weber, Interdisciplinary Telecommunications Department, University of Colorado, Boulder

3:15 **Break**

3:30 *Broadband Demand and Wireless*, Tom Lookabaugh, Scott Savage, Douglas Sicker, Interdisciplinary Telecommunications Department, University of Colorado, Boulder

4:00 *Broadband Wireless Metropolitan Area Networks*, Roger B. Marks, National Institute of Standards and Technology

4:30 Round Table Discussion

Table of Contents

Page

Looking to Spectrum for Network Utopia

Dewayne Hendricks

What We Want

In network utopia, everyone will be connected across the digital divides of economics and geography. In network utopia, everyone will be connected with enough "bandwidth"—enough bits—that there will no longer be any impediment to innovation. Reaching network utopia may be possible by looking at where the most bits are: radio spectrum. Although spectrum has been treated like a scarce resource for almost one hundred years, today's emerging technologies are changing this situation. There is actually an abundance of spectrum—more than enough for everyone.

Where We Are

Today's communications technology is moving toward a world of all-digital transmitters and receivers. These advances in technology, combined with the swift evolution of mesh-based transmission and switching protocols, are opening up a new set of possibilities for unique new services utilizing intelligent wireless networks. These networks will contain smart transmitters, receivers, and switches. The Federal Communications Commission (FCC) has recently coined a term to describe these types of wireless devices: cognitive radios (CRs). Today's Internet is perhaps the best example of a self-regulating structure that embodies these new technological approaches to communications in the networking domain. However, to date, many of these innovations have not moved into the wireless networking arena.

Wireless networks of the future will likely involve a mixture of links and switches, of different ownership, that terminate at the end-user in the first mile via relatively short-distance links. What will then be required is a built-in, distributed, self-governing set of protocols to cause the network to make more efficient use of a limited, common shared resource—the radio spectrum. Creating such a self-regulating structure for the optimal sharing of spectrum will require much effort. A major problem that stands in the way of these new approaches today is the current FCC regulatory environment and the manner in which spectrum is managed and allocated under its rules.

One of the major hurdles encountered by a wireless entrepreneur who wants to develop innovative new communications products involving radio is access to the requisite amount of spectrum. Obtaining spectrum mandates the involvement of the wireless entrepreneur with the government, which immediately puts the entrepreneur at a disadvantage when compared with entrepreneurs in the computer sector, where government involvement is minimal. As a result, wireless innovation has occurred at a much slower pace, since the use of technologies such as spread spectrum (which is used in one of the Wi-Fi standards) requires more spectrum, not less, in order for their advantages to become apparent when used for high-speed data transmission.

The current regulatory approach to radio is based on the technology that was in use at the time the Communications Act of 1934 (the bill that created the FCC) was framed. Basically, this technology was what we would call today "dumb transmitters speaking to dumb receivers." The technology of that time reserved bandwidths to be set aside for each licensed service so that spectrum would be available when needed. Given this regulatory approach, many new applications cannot be accommodated because there is no available unallocated spectrum in which to "park" new services. However, given the new set of tools available to the entrepreneur with the advent of digital technology, what once were dumb transmitters and receivers can now be smart devices capable of exercising greater judgment in the effective use and sharing of spectrum. The more flexible the tools that we incorporate in these devices, the greater will be the number of uses that can be accommodated in a fixed amount of shared spectrum.

One of the most promising regulatory actions by the FCC in the last twenty-five years was its move in 1981 to permit the use of spread-spectrum technology in unlicensed devices. This proposal eventually resulted in a new type of wireless device that operates under the unlicensed Part 15 regulations and is deployed in what are called the industrial, scientific, medical (ISM) bands. More important, these devices are forbidden to operate at power levels greater than one watt, and their transmissions must be spread a minimum amount across the assigned spectrum.

Those restraints notwithstanding, the Part 15 regulations and later additions and changes to those rules have spawned the development, manufacture, and marketing of a wide range of "no license required" products. Because mass manufacturing has now started to occur, spread-spectrum products for data transmission from the several thousand current vendors no longer carry the premium price tags that had limited the technology mainly to large organizations, such as businesses, schools, and libraries. Today a radio that can handle Ethernet traffic (100 Mbps, suitable for high-speed computer communications) up to a distance of about 40 kilometers (25 miles) costs about $4,000. Devices with lower capability—with operation at T1 speeds (1.5 Mbps) to a range of 25 kilometers or so—cost less than $500. For very short ranges, such as for communications within a building, wireless local-area network (LAN) cards for PCs are priced as low as $50.

There is every reason to believe that these prices will drop further as manufacturing volumes increase to meet the growing market demand for higher bandwidth and secure wireless connections from PCs to the Internet. In the future, people may, for example, routinely rely on wireless transmission to reach a central system that would then connect to a traditional network of ground-based lines. Reliable, secure unlicensed data radios operating at T3 or higher speed to a range of more than 30 kilometers could soon cost less than $500 each. The current buzzword for the most prevalent of these new unlicensed devices is Wi-Fi, a shorthand for the various IEEE 802.11x standards. These devices can support communication at speeds of 2 to 100 Mbps, over distances from 50 feet to 30 miles. For example, in early January Apple introduced an 802.11g Wi-Fi base station that operates at 54 Mbps, supports up to fifty-five users, and costs $199.

Where We Can Go

The creation of a decentralized structure for the optimal sharing of the radio spectrum will require a substantial effort by a combination of telecommunications experts and entrepreneurs working with the various regulatory bodies around the world. The deployment and growth of such a system is achievable through increasingly "smart" wireless electronics with a built-in set of self-governing protocols. The early stages of this growth can be seen today as Wi-Fi device deployment is becoming more widespread.

This future is not from the realm of science fiction. The FCC is attempting to grapple with these issues in order to determine the regulations and policies that will affect the governance and use of spectrum for the balance of this century. In 2002 the FCC created a special task force on spectrum policy; the task force produced a report with recommendations on new approaches for spectrum policy and management, and the report has been released for comments from the general public. The comments that are submitted to the FCC on this report can play a significant role in guiding the hand of the FCC as it crafts a set of rules to implement the findings of the report.

Higher education has an important role to play as this process moves forward. Right now, in and around campuses across the country, these new wireless devices are being deployed in ever growing numbers—a trend that will only increase as the prices continue to drop. Already they are changing the way that students and faculty interact with one another. For instance, some schools are now considering banning the use of wireless networks in the classroom because they are often considered to be disruptive to the learning process. At other schools, these wireless devices are being debated as replacements for the wired telephone infrastructure currently being used, with the possibility of major cost savings.

As this regulatory process plays out, the higher education community needs to acquaint both the FCC and Congress with the uses to which these new wireless devices are being put and the benefits that can be realized as a result. Higher education will be on the "front lines" with many of these new wireless technologies. For this reason, as advanced radios are deployed, higher education can be in a position to best help the government tackle the crucial issues of incorporating both positive and negative incentives within the network infrastructure itself to make the best use of a shared common resource—the radio spectrum—and to move toward network utopia.

Dewayne Hendricks is CEO of the Dandin Group, located in Fremont, California. For the past four years he has been a member of the FCC's Technological Advisory Council.

Spectrum Buyouts

A Proposal for the Transition to Open Spectrum

Nobuo Ikeda
Research Institute of Economy, Trade and Industry (RIETI)
Phone: +81-3-3501-8286
E-mail: ikeda-nobuo@rieti.go.jp

Abstract

To remedy the stalemate of spectrum policy, regulatory reforms are being proposed to assign exclusive rights to spectrum. Such reforms are harmful because wireless LANs enable far more efficient communications by sharing a wide band. So it would be better to open the spectrum as *commons*, instead of dividing it into narrow bands. I propose a system of *spectrum buyouts,* by which the government would take back spectrum from incumbents and reopen it for use without a licensing requirement. The prices of these reverse auctions would be much cheaper than that of ordinary spectrum auctions.

1. introduction

In November 2002, the Federal Communications Commission (FCC) published a report written by the Spectrum Policy Task Force (SPTF). Summarizing a half year of extensive research and discussion, the report is indeed impressive in its deep understanding of digital wireless technologies and its call for bold reforms. Particularly noteworthy is the FCC's commitment to depart from the *command and control* approach that regulates usage by licensing. It is also remarkable that the FCC recognized the efficiency of the *commons* approach, which shares a wide band and enables the overlay usage of spectrum by different kinds of terminals.

However, the SPTF's conclusion is a half-hearted compromise between the commons model and the *exclusive rights* model, according to which incumbents can sell and buy their spectrum on secondary markets. SPTF insists that this "market-oriented" approach is more efficient than the commons approach for the band in which "scarcity is relatively high and transaction costs associated with market-based negotiation of access rights are relatively low" (FCC 2002: p.38). They claim that spectrum is scarce below 5 GHz because of its propagation characteristics and "high level of incumbent use."

Scarcity cannot be the justification for property rights. Roads, parks and streetlights, for example, are all supplied as commons even though they are scarce. In standard economics, a resource is efficiently allocated as private goods if its consumption is *rival* (the marginal cost of consumption is large) and its supply is *excludable* (externality is small). In principle, the externality of resources can be internalized by dividing common resources and assigning property rights to the parts, which can be priced and allocated efficiently by a market mechanism (Coase 1960).

The externality of common resources is often so large that they cannot be excluded without destroying value. Such an example is ordinary road traffic; the road is rival because traffic jams will take place if too many cars are on he road, but it would be inefficient to exclude each car by building fences on the road. It is enough for each car to drive while avoiding other drivers. The exclusive-rights model of spectrum management is similar to such a "fence-on-the-road" approach.

"Transaction (excluding) costs" are not given exogenously, but determined by market structure, regulation, and technologies adopted in the band. Exclusion is needed only in so far as the spectrum is separated by frequency. If terminals are intelligent, they can allocate spectrum dynamically by identifying each other's signals without exclusion, as we shall see later. In fact, alleged scarcity and transaction costs are *created* by old technologies and poor spectrum management based on exclusive rights. Therefore it would be circular logic to justify exclusive rights by the scarcity that is created by exclusive rights.

2. Open Spectrum

Is Spectrum a Scarce Resource?

The auction was hit upon as a mechanism for allocating spectrum efficiently, but it was based on the dubious assumption that the spectrum is a scarce resource that the government has the right to allocate. "[I]s the spectrum the government's to sell in the first place?" asks Noam (1998: p. 771), "Could the state sell off the right to the color red? To the frequency high A-flat?" He cited the licensing of spectrum as a violation of the freedom of the press. To understand this problem, it is necessary to distinguish *frequency* from *spectrum*. Frequency is not a resource but a *parameter* used to modulate original data (baseband) into radio waves, so it cannot be scarce any more than amplitude and phase are (Benkler 1999).

In radio communications, transmitters modulate basebands into airwaves by mixing them with carriers of a specific frequency and send the wave in radial form. Receivers identify radio signals by tuning in to the desired frequency and filtering out other frequencies. When basebands are modulated into radio waves, they are distinguished by the frequencies of their carriers. Sending multiple signals on the same carrier causes interference. Therefore interference is not a problem of scarcity but rather a result of *confusion* by receivers that cannot distinguish signals from noise (Reed 2002). So a frequency can be used by multiple users if their receivers can identify signals.

On the other hand, spectrum has limited capacity. According to Shannon's Channel Capacity Formula, the channel capacity *C* (bits per second) is limited by the bandwidth, *B* (Hertz):

$$C = B \log_2 (1+S/N),$$

where *S* is the power of the signal (in watts), and *N* is the noise level (W/Hz). In analog radio, as it is impossible to distinguish signals of the same frequency, spectrum should be divided into small portions to avoid interference. And, since *N* is given physically, the only way to do this is to magnify *S* to discern signals from noise. Thus radio signals are sent in narrow bands and at high power to large areas. If *B* is divided into small portions of equal size, $b_1, b_2,...b_n$ and allocated to each licensee, each licensee can get at most *C/n* of capacity. The inefficiency of this *high power and narrow band* radio system did not matter when radio equipment was very expensive and a small part of the spectrum was utilized, but it is posing serious problems today.

Cellular phones depend on the circuit switching in which each user occupies a band exclusively even if no signals are transmitted. A digital wireless technology called *packet radio* extends *B* by sending different packets in a band. Because packets are identified individually, interference can be avoided even if multiple signals are carried in the same frequency. Spectrum is used efficiently by *statistical multiplexing*, which levels traffic in a wide band. As average traffic usually represents a very small portion (less than 10%) of the maximum capacity, if 100 users share a bandwidth of 20 MHz, more than 2 MHz is available for each user on average. This is obviously more efficient than allotting 200 kHz across 100 users.

If *B* is large, it is not necessary to magnify *S* to increase *C*. Lowering power makes it possible to multiply spectrum by establishing many stations. This *low power and wide band* system makes digital radio more efficient than traditional broadcasting systems. The problem is thus not the *scarcity* but the *efficiency* of spectrum usage. Therefore, bandwidth can be better utilized as commons, shared by many WLAN terminals. If a wide band can be shared by many users identifying signals packet by packet, this will be much more efficient than dividing spectrum into narrow bands and selling them to individual users.

A packet radio technology called *spread spectrum* has been widely adopted to send various packets in a band while avoiding interference. In the direct-sequence spread spectrum (DSSS) adopted in WLAN, transmitters multiply original signals (baseband) by *pseudo-noise* (encryption key) and spread the resulting signals into thin waves over a wide band using weak power. Receivers decode the airwaves by inverse spreading, in which the signals are multiplied by the inverse pseudo-noise. By multiplying and dividing the baseband by the same number, this process recovers the desired data but scatters the noise thinly to allow its elimination by filters.

Thus it is not necessary to separate frequency to prevent interference. A number of users can use full bandwidth by multiplexing and identifying individual packets by their spread codes. Spread-spectrum technology was invented during World War II to prevent interception and electromagnetic jamming of military communications. It was later adopted for communications in the unlicensed band (2.4 - 2.5 GHz) to prevent interference from other devices such as microwave ovens. This band is called the ISM (Industrial, Scientific, and Medical) band, because it was originally released for unlicensed use by hospitals, factories, and so on, rather than for communication purposes.

WLAN technology, standardized in the 802.11 Committee of the Institute of Electrical and Electronics Engineers (IEEE), initially attracted little attention because its speed was only 2 Mbps. But after the enhanced mode IEEE 802.11b (Wi-Fi) was standardized in 1999, WLAN exploded; within a few years the number of users worldwide grew to more than 30 million (2002 figure). This is because 802.11b realized up to 11 Mbps (3-4 Mbps on average) by sharing the wide ISM band (22 MHz per channel)[1]. In contrast, the speed of data communications in current 2-G mobile telephones is around 10 kbps due to bandwidth limitations. For example, the PDC adopted in Japan allocates only 25 kHz (12.5 KHz in "half-rate" mode) per user.

Multiplexing by Space, Time, and Power

The method of multiplexing airwaves for many users is not limited to frequency. Shannon's Formula represents the limit of capacity in a given place, but it can be extended by multiplying stations because different users can use the same band repeatedly in separate places. This is the cellular technology by which mobile telephones enhanced bandwidth over traditional usage. The WLAN band is separated into a number of channels, which are allocated to each low-power station. As shown in Figure 2, channel A can be used repeatedly by dividing an area into many *microcells* in which each user can utilize full capacity without interference from other terminals. If the band is wide enough to allow division into many channels, theoretically, the capacity can be multiplied infinitely by dividing an area into an infinite number of cells.

[1] WLAN spreads the same signal several times, so the transmission efficiency per frequency of 801.11b stands at 11 Mbps/22 MHz = 0.5, similar to that of cellular telephones.

Of course, the overhead cost of connection between base stations will limit the number of cells in reality. But if they can be connected by wireless networks, this cost could be reduced. For WLAN terminals to be used as base stations in *ad hoc* mode, completely distributed multi-hop networks called *wireless mesh*, which link terminals to each other directly, can be built by WLAN terminals. If the price of WLAN chips falls to several dollars – as is likely in a few years – they will be incorporated into a wide range of devices that can communicate with each other.

In this regard, WLAN is even more revolutionary than wired Internet. TCP/IP is characterized by the architecture referred to as End-to-End (E2E), which means that the communication is controlled only by senders and receivers. In the wired Internet, however, routing and addressing are mostly performed by Internet Service Providers (ISP) because networks are built on the telephone-type topology. WLAN has deconstructed the centralized architecture and enabled completely decentralized E2E structures physically. Such *ad hoc networks* have been built throughout the world by volunteer organizations.

Public networks can be built by linking local wireless networks called *hot spots* in restaurants, hotels, airports, and so on. But the quality of the 2.4-GHz band is unsatisfactory. Industrial dryers, medical equipment, and different types of communication terminals such as Bluetooth interfere with WLAN. And the bandwidth (less than 100 MHz for 4 channels simultaneously) would not be sufficient if many operators built base stations in the same place. The quality of the 5-GHz band is higher than that of the 2.4-GHz band, although the higher the frequency (i.e., the shorter the wavelength), the heavier the attenuation, and the more vulnerable communication becomes to obstacles.

In the United States, 300 MHz is available within the Unlicensed National Information Infrastructure (U-NII) band at 5-GHz band. The European Union is planning to open 580 MHz for HiperLAN without a licensing requirement, which can be divided into more than 25 channels in which up to 54 Mbps can be transmitted in each channel with IEEE 802.11a. In Japan, however, there is no unlicensed outdoor band at 5 GHz; only 160 MHz is available by license and 100 MHz is available indoors without a license.

There is another dimension by which we can utilize spectrum efficiently: time. For example, meteorological radars occupy 5.25-35 GHz, but they use the band for only a few minutes per hour. If other terminals can sense the radar waves and stop using the channel while the radar working, they can work together in a channel. Such adaptive technologies, known as *agile radio,* have been standardized and implemented into some 802.11a chipsets. Dividing bandwidth by time, these technologies enables WLAN base stations to coexist with other terminals in a band and realize much more efficient use of idle spectrum. For example, 300 MHz of the UHF band is allotted to TV stations, but less than a half of it is used in Japan. So if WLAN terminals equipped with cognitive radio technologies can detect vacant channels and use them, more than 100 MHz of spectrum can be "created." If such *overlay* usage is allowed in all bands, available bandwidth will be so large that its allocation would not be necessary.

Software-Defined Radio (SDR) will make such adaptation even easier by changing physical layers by software, just like applications for PCs. And *smart antennas*, combining various antenna elements with a single processor, can change the transmission/reception mode in response to the communication environment. If a channel is occupied by 802.11a, other terminals can change its modulation to 802.11a by SDR. To deploy SDR, however, regulatory reforms will be necessary: the present Radio Act bans non-standardized communication devices by certification of equipment, but if communication is performed by software, it would make no sense to certify the equipment.

There is yet another dimension of multiplexing: power. Part 15 of the Code of Federal Regulations defines the admitted noise level for unlicensed devices. Ultra-Wide Band (UWB) is the technology to use such very weak signals that cannot be distinguished from the radio noise generated by TVs, computers, and hair dryers. In contrast to the conventional radio technology that modulates baseband with a carrier (sine curve), UWB modulates the baseband with very short pulses (less than a nanosecond). This technology realizes high-speed transmission (up to 500 Mbps) by emitting pulses in a wide band, over a frequency range of several GHz. Since their waveforms are completely different from those of conventional radio waves and are emitted at very low power levels, advocates of UWB claim, the system will make overlay use possible over all bands without interference. In fact, however, interference was found in experiments conducted by the FCC. In February 2002, the FCC authorized UWB with very conservative restrictions for its band (above 3.1GHz) and with weak power. Therefore, for the time being, use of UWB will be limited to indoor use.

Regulatory Reforms

These radically new technologies are demanding a "new spectrum policy paradigm" according to the FCC Chairman Michael Powell (Powell 2002). To cope with these changes, Noam (1998) proposed a reform named *open access*. If you allocate bandwidth dynamically, this will be far more efficient than the current system of static allocation. If demand is lower than capacity, everybody can access bandwidth freely. If demand exceeds capacity, a "clearing house" charges fees for wireless traffic, acting as a tollbooth. It is much harder to charge for airwaves than for cars because the former do not pass through specific gates, so this proposal has been regarded as unrealistic. However, digital technologies such as spread spectrum have now rendered this idea feasible.

If bandwidth is supplied to an extent greatly exceeding demand, open access will become possible without fees. Even if bandwidth did not exceed demand, the allocation of packets by spread spectrum would be more efficient than charging for packets. Packets in the wired Internet are stored and forwarded by routers without charge. Congestion leads to waiting, but this is not a very serious problem in data communications and can be overcome by widening the bandwidth. Already, we can reach up to 108 Mbps by using two channels of 802.11a together. UWB has realized 500 Mbps and its capacity will easily extend to more than 1 Gbps.

Thus rivalry of spectrum among multiple users can be eliminated using packet radio technologies, which increase capacity by adding stations and terminals. If a resource is neither rival nor excludable, it should be supplied as (pure) *public goods*. A typical example of public goods is national security. Since there are no economic problems of resource allocation with such services, a "market-oriented" approach does not make sense. Instead, in the long run, the spectrum should be maintained by public administration which makes rules and enforces them by monitoring abuses.

"Public" does not necessarily mean "governmental." The concept of commons is so old that it has been preserved by the social norms of the community without any government regulation. It is the destruction of social norms by Western companies that incurred the large-scale abuse of tropical rainforests. Today the commons of the Internet is preserved by hundreds of millions of users worldwide without any government control. Standardization of radio equipment by the government has ended with the failure of 3G. Today, such non-profit organizations (NPO) as IEEE and the Internet Engineering Task Force (IETF) have taken over the role of the ITU.

Of course, this does not mean that government regulation is unnecessary. Even if there is sufficient bandwidth, interference will occur between different physical layers. One way to prevent such interference is to fix a physical layer (modulation) for each band; for example, 802.11b for 2-3 GHz and 802.11a for 3-6 GHz. Some argue against unlicensed usage because such physical regulation will impede innovation (Hazlett 2001), but regulation is not necessary for this purpose. For example, if a channel is occupied by Bluetooth, WLAN can use another channel by sensing the carrier. If there is sufficient bandwidth and flexible technologies such as agile radio are deployed, various physical layers can coexist in different channels.

To coordinate various kinds of terminals to work cooperatively, regulating channels, powers, frequencies, and modulations of different terminals will be the important task of radio administration. Traditional regulation has focused on transmitters, but it is necessary to regulate receivers to control interference among different types of terminals. Since digital receivers are much more tolerant of interference than analog ones, there should be more flexible criterion *interference temperature*, according to the FCC's term, to enable different systems to coexist in a band.

Such regulation should be enforced not for operators but for manufacturers because communication terminals will exist as ordinary electronic appliances independent of operators and service providers. The standardization can be left to the NPO, but the certification of equipment and monitoring of abuse should be carried out by the government. Without such supervision, unlicensed bands tend to bring about a "tragedy of the commons" as recently evidenced by the 2.4-GHz band. Although it is most important to supply sufficient capacity to render abuse unnecessary and harmless, surveillance and enforcement will have to be intensified, at least transitionally.

3. Reverse Auctions

Strategy for Transition

During the transition period, licensed and unlicensed bands will coexist, but the criteria by which the spectrum rights are specified should be determined not by the so-called scarcity but by the excludability (efficiency of exclusion) of a band. Above 3 GHz, it is pointless to exclude spectrum because there is no new technology that depends on frequency division in that band. Exclusion might be justified in the extremely lower band (probably below 30 MHz) where high-power propagation is economical and no digital radio technology is likely to be implemented. In the intermediate band, the *easement* of overlay usage should be enforced.

Thus a strategy of transition to more efficient technologies is necessary. As SPTF insists, spectrum policy must "provide incentives for users to migrate to more technologically innovative and economically efficient uses of spectrum" (FCC 2002: p.15). To achieve the goal, however, it seems that the FCC is going to give spectrum away to incumbents as their private property and let them use it efficiently. At the same time when the SPTF report was published, economists at the Office of Plans and Policy of the FCC published a working paper to prescribe the "Big Bang auction" that would enable incumbents to sell and buy all spectrum freely (Kwerell-Williams 2002).

Indeed this system would be politically easy to accomplish because it is so advantageous to incumbents. However, there would exist a danger that exclusive rights would authorize incumbents to exclude other parties' more efficient usage. If spectrum were sold at a high price, the "owner" of spectrum would maximize its value by monopolizing it. This is rational behavior for individual users, but it would lead to socially inefficient outcomes. Even worse, such a policy is irreversible; once spectrum is given away to incumbents, spectrum commons would be lost forever because incumbents would never open it. Easement would be harder to enforce because incumbents would resist such "regulatory taking" of their private property.

Legally governments can take back the spectrum when licenses expire. The Ministry of Public Management, Home Affairs, Posts and Telecommunications (MPHPT) of Japan announced a plan for such ruling in November 2002. MPHPT is going to rule that, if licenses expire, licensees must return their spectrum with compensation for the remaining book value of their equipment. Because the term of license is five years and the term of amortization is six years, the average licensee's remaining value is very small. This is legitimate but difficult to enforce. If incumbents resist, it would take a long time to evict them by negotiation; MPHPT estimates that it would take 10 years to clear the 4-GHz band. Worse, many incumbents would refuse the "taking" of spectrum on which their business depends and regulatory nightmare would result.

Such a problem can be resolved by breaking it down into two parts; it is necessary to motivate incumbents to exit by compensation, but it is harmful to admit them exclusive rights for the spectrum. So it is advisable for the government to take back the spectrum through *reverse auctions* and then open the acquired spectrum without a license requirement[2]. This mechanism can be implemented as an ordinary procurement process by which the lowest bidder sells goods to the government.

Auction Design

The government should "clear" a band by taking back all the stations in the band nationwide, but it would not be necessary to open all spectrum because 1 GHz might be enough to supply the bandwidth for WLAN in current use. As it is difficult to compare the value of different bands, the government is advised to focus on specific bands. In Japan, the best candidates for WLAN band are the 3-5 GHz used for business-use communications and the backbones of mobile telephone networks. The procedure would be as follows:

1. The government announces the required minimum bandwidth, the target band, and the budget.
2. Bidders register their prices to sell their spectrum via computer network. The bids will be known to all bidders through the network.
3. As long as there are new bids, the government continues to lower the price.
4. If no bid is registered, the auction is over.

Because simple maximization of bandwidth may result in fragmentation of the band into many small pieces, there should be a requirement on bands; for example, the band should be continuous more than 50 MHz. Government can maximize the bandwidth per budget by buying the spectrum of all stations in the band in which aggregate price is the cheapest. Suppose the budget is 2 billion yen and aggregate bidding prices for each 10 MHz band in 4.00-4.10 GHz are as shown in the following Table:

4.00	01	02	03	04	05	06	07	08	09	10 GHz
.4	.3	.1	.8	.7	.3	.3	.4	.2	.4	.6 billion yen
←		→			←			→		

Table: Bids for the band groups

In this case, government can maximize the bandwidth per budget by buying the spectrum of all stations in 4.00-4.02 GHz with 0.8 billion yen and 4.05-4.08 GHz with 1.2 billion yen. Stated generally, the objective function of the government is to maximize (in a given range) the total bandwidth $W = \Sigma_i w_i$ ($i = 1,2,\ldots n$) that aggregates the individual bands w_i, within which the aggregate price $P_i = \Sigma_j p_{ij}$, where p_{ij} is the jth bidder's price within the ith band, subject to the condition

$$w_i \geq r,\ \ \Sigma_i P_i \leq y$$

where r is the required minimum continuous bandwidth and y is the budget. This specific procedure is due to the requirement that the unlicensed band should be opened in as large a block as possible. Other requirements are possible: for example, there should not be more than three fragmented bands or wider bands should be evaluated with some premium.

An important characteristic of this auction is that the *aggregate* price within each band group is compared. So the band that includes the least number of incumbents is likely to win even if the individual member's bid is higher than the other band. Conversely, if bidders know the number of participants in each group, they will minimize the aggregate price of the group to which they belong, instead of their individual prices. To avoid such a problem, the government can "normalize" the prices, for example, by distributing the average price per MHz to all incumbents in the winning bands.

With such adjustments in place, bidders will have little incentive to offer a higher price than the true value, because they bear the risk of losing the bid, while the gain will be equally distributed among all bidders in the group. For example, if the 4.05 GHz group raises its price by more than 0.4 billion yen, they will be outbid by the 4.09 GHz group. Collusion in this auction would be difficult because the boundaries of winning bands are variable. If a group of incumbents succeeded in lifting its aggregate prices, it could shift the boundary.

[2] As a complementary mechanism to patent licensing, Kremer (1998) proposed a mechanism called the "patent buyout" in which the government buys patents from inventors through auctions and opens the patents to everyone.

Because competition will be effective in a band where many incumbents are evenly distributed, governments are advised to have auctions in such well-organized bands. If these auctions are repeated, the least populated bands will be vacated and the next least populated will be the winner. Thus, incumbents in a densely populated band will not join the auction in the earlier stages. As a result, we can expect the bidding price to be roughly equal to the bidder's own valuation of spectrum, though it would be safe to conduct experiments along these lines before execution.

Some have argued that such an auction would be extremely costly, referencing the prices of PCS auctions, but this would not be the case. In an ordinary spectrum auction, an *ascending* English auction, the equilibrium price is equal to the net present value (NPV) of the *most efficient* use of spectrum. On the contrary, in a *descending* English auction, the equilibrium aggregate price is approximately equal to the opportunity cost of the *least efficient* use of spectrum on average.

The opportunity cost for a bidder includes the remaining asset value of equipment plus the net present value (NPV) of the profit that would be gained by using the spectrum. Usually NPV is determined by future cash flow, terms of license, interest rates, tax, and so on. However, even if an incumbent returns the spectrum, it can do the same business over wireless Internet when the spectrum is opened. In such a case, the profit would be lower because the market for the same services is more competitive. Thus, the NPV is the discounted value of monopolistic rent that would be smaller than the usual NPV.

Ignoring the interest rate and tax, I denote the opportunity cost of the least efficient user k as $Q_k(x) = V_k(x) + z_k$ where V_k is the NPV, x is the term of expiration, and z_k is the remaining asset value (supposed as a constant). If the government can credibly threaten incumbents to return the spectrum, or at least if there is uncertainty as to the duration of the license, this buyout would be more effective. Suppose $V_k(x)$ is subdivided into the cash flow v_j in each term. If the rule of returning is enforced in the second term, $Q_k(x) = v_k + z_k$. If it is enforced at probability q every term, the opportunity cost will be

$$Q_k(x) = v_k + (1\text{-}q)v_k + (1\text{-}q)^2 v_k + \ldots (1\text{-}q)^x v_k + z_k.$$

If x approaches infinity, $Q_k = v_k/q + z_k$. So the equilibrium price p^* will be

$$p^* = v_k/q + z_k.$$

If enforcement becomes more likely, q will approach 1, then p^* will approach $v_k + z_k$; the one-term profit plus the remaining book value of the least efficient user. Moreover, as v_k is the NPV of the most poorly operating incumbent's monopolistic rent, its value will be very small if reverse auctions are properly designed. Even if an inefficient incumbent refuses to join an auction, its monopolistic rent will deteriorate when entrants do the same business over WLAN using the opened band. Therefore, if sufficient bandwidth is opened without a license and incumbents are rational, we can suppose $v_k = 0$. This result coincides with the plan of MPHPT, which takes back spectrum while compensating for the remaining book value, that is, $p^* = z_k$. The difference is that the idle spectrum can be taken back right now in our mechanism. Thus, I recommend this reverse auction as an *optional* mechanism that incumbents can choose, together with a strong commitment that, after the government acquires the spectrum, it will open enough spectrum to wipe out monopolistic rents.

On the other hand, once the band is made private property, as is planned by the FCC, q will approach zero and v_k will increase because this becomes profitable; therefore p^* will be much higher and a buyout will become more difficult. That is, the "privatization" of spectrum makes it difficult to open it without a license. If it were possible to suppress the rent and make spectrum commons by regulation in the end, as Faulhaber-Farber (2002) claims, no one would buy the spectrum that would eventually be worthless. In other words, as it is inevitable that the value of spectrum will disappear if the wireless Internet prevails, the NPV of every spectrum would approach zero in the long run. It would be pointless to buy such a worthless asset or sell it at auction.

Public users cannot be bidders, but they should be compensated for the cost of converting equipment or of exiting. Their bands should be evaluated as the average of the nearby bidders. Another problem would be posed by whether or not a public band should be sold; for example, the band used by air traffic control could not be sold by the market mechanism.

4. Discussion

Spectrum buyouts may arouse controversy. Some would oppose this reform as an unfair income transfer for incumbents who are underutilizing allocated bands. I argue that, following the Coase Theorem, it is much more efficient to "bribe" incumbents to return their idle spectrum than to negotiate with them over a long time. The opportunity cost of wasting bandwidth and time would be much more expensive than the cost of buying the band back. In my scheme, the government does not have to negotiate with incumbents and politicians but only has to announce the reverse auction. Incumbents will bid and reveal their valuation of spectrum, and winners will give back their bands even if they are using them, as they would be reimbursed for the cost of replacing their old stations and terminals with WLAN.

Some argue that there is no need for buyouts if the overlay use of spectrum is admitted; if every terminal could use all idle bands dynamically, it would make no sense to reallocate spectrum at all. While this is true, agile radio is so complicated and expensive that it has not yet been implemented in portable terminals. To avoid interference, agile radio terminals should store detailed data of other equipment's characteristics such as the power, direction, and timetables of radar. This represents not just a technical but a regulatory challenge. Overlay use of different modulation systems in the same band requires complicated regulation of devices, which regulators are reluctant to enforce. UWB was at last authorized by the FCC in 2002 after 20 years of negotiation.

It is naïve to suppose that incumbents will admit the easement of overlay if it does not interfere the incumbent's communication. Since massive entrance will threaten their monopoly profits, incumbents will resist easement in their spectrum under the pretext of interference, as evidenced in the case of UWB. In such cases, the FCC can have "overlay auctions" to compensate incumbents for allowing easement of overlay usage. In the long run, this would be equivalent to the reverse auctions because incumbents will renew their equipment that can be used as overlay. So rational incumbents would be willing to sell their spectrum and change their stations and terminals with the auction fees.

The reverse auction is, as stated above, not a substitute for overlay use but a complementary strategy to facilitate transition in the band required for WLAN. Opening a clean band is obviously better than easement, so the problem is which is the faster and cheaper method for opening spectrum. This will depend on various factors such as progress in radio technology, the political power of incumbents, and so forth. My guess is that, at least in the band above 3 GHz over the next 10 years, buyouts will prove to be the faster way. It would not be cheaper, but it could buy precious time by "bribing" incumbents. This might work as a middle-of-the-road solution between the commons approach, which is economically efficient but politically difficult, and the exclusive rights approach, which is inefficient but easy. Both incumbents and entrants can benefit from this buyout, and we can open spectrum as commons through a market mechanism.

Financing might be the most difficult part of this reverse auction because the fee would be much larger than in usual procurement cases. A simple solution would be to finance the auction through general government accounts, in view of the fact that governments have made a great deal of money by auctioning off spectrum to private parties. This would cure the problem of spectrum auctions raised by Noam (1998): auctions "tax" the communications industry and suppress investment. Through such repayment the government can revitalize wireless operators, which lost a great deal of money in the collapse of the bubble. It is, in effect, a collective auction by WLAN users, so its cost is equivalent to that of an ordinary spectrum auction, in which the winner will pass on its costs to consumers.

Another solution, probably better suited to Japan, would be to compensate the government's cost of reverse auctions through *spectrum usage fees*. This would be more neutral to public finance, and raising the fee would press incumbents to use bandwidth more efficiently or to sell out. The present tariff of spectrum usage fees in Japan, however, is a disincentive for efficient use of bandwidth: because the fees are charged in proportion to the number of radio stations, more efficient users are charged more. If the fee is charged for bandwidth, this will offer incentives for efficient bandwidth use. As these financing methods are complementary, governments could use them in combination.

The greatest risk might lie in having the government conduct such a gigantic auction. It is possible that irrational behavior (as was seen in 3-G auctions) might lead to extreme behavior and unexpected results. If sellers rushed to sell their bands as soon as possible, the price would be near zero, but such mistakes would be harmless for the government. If sellers were to collude to keep the bidding high, the government would have the option to quit. Rent seeking and collusion would be most effective because the stakes are so high. It would thus be necessary to perform preliminary experiments before the buyout and to keep the procedure transparent. Further, governments would need to exit from spectrum management after all spectrum had been opened.

References

Benkler, Y. (1999) *Harvard J. of Law and Tech.* http://www.law.nyu.edu/benklery/agoraphobia.pdf.

Coase, R.H. (1960) *J. of Law and Econ*. 3:1-44.

Faulhaber, G. and Farber, D. (2002) http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6513282647

FCC (2002) *Spectrum Policy Task Force Report,* http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-228542A1.pdf

Hazlett, T. (2001) *Harvard J. of Law & Tech.* http://www.aei.brookings.org/publications/working/working_01_02.pdf

Kremer, M. (1998) *Quarterly J. of Econ*. 113: 1137-67. http://papers.nber.org/papers/W6304

Kwerell, E. and Williams, J. (2002) FCC working paper, http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-228552A1.pdf

Noam, E. (1998) *J. of Law and Econ*. 41:765-790.

Powell, M.K. (2002) "Brodband Migration III" http://www.fcc.gov/Speeches/Powell/2002/spmkp212.html

Reed, D.P. (2002)) "How Wireless Networks Scale" http://www.its.bldrdoc.gov/meetings/art/art02/slides02/ree/ree_slides.pdf

3G Licensing in Hong Kong: A Unique Approach

Xu Yan

Department of Information and Systems Management
Hong Kong University of Science and Technology
Tel: (852) 2358 7640, Fax: (852) 2358 2421
E-mail: xuyan@ust.hk

[Abstract] *The auction of 3G licenses in Europe has raised tremendous controversies. However, it is not the auction itself that should be blamed because commercialisation of spectrum might be the only solution to manage this increasingly scarce resource. What matters is the scheme of commercialisation. By taking the late-mover advantage, Hong Kong took n effective and unique approach in issuing its 3G licenses.*

On 19 September 2001, four 3G licenses were finally issued by the Office of Telecommunications Authority (OFTA) - the regulator of telecommunications in Hong Kong. Instead of auctioning the upfront payment for the spectrum to be used for 3G services, as did by regulators in many other economies, OFTA took a unique approach by auctioning the royalty, e.g. the percentage of 3G revenues that the bidders are willing to pay. Licensees are also obliged to open at least 30% of their network capacity to Mobile Virtual Network Operators (MVNOs).

Although response to the auction was not enthusiastic due to the slowing-down of economy and the burst of IT bubbles, this scheme has raised attention from regulators in many other economies. Importantly, the 11-month-long policy consultation on 3G licensing framework has trigged intensive debate among all interested parties. This paper, based on the author's research report for International Telecommunications Union[1], will provide a comprehensive review of these debates and give an introduction to the overall 3G licensing scheme in Hong Kong.

I. Policy Objectives of Hong Kong Government in 3G Licensing

To provide an efficient telecommunications infrastructure for the local economy and the public by means of deregulation and liberalisation has been a consistent policy stance of Hong Kong government since the mid-1980s. According to the Information Technology and Broadcasting Bureau (ITBB) - the policy-making body of Hong Kong government, these policy objectives remain unchanged for 3G licensing. [2]

[1] Xu Yan (2001) "3G Licensing Policy in China and Hong Kong SAR – A Study for the International Telecommunications Union" http://www.itu.int/3G

[2] Press Release of ITBB on 13 February 2001, *http://www.itbb.gov.hk*

The policy-making process of 3G licensing in Hong Kong is relatively open, transparent and subject to full debate. On 21 March 2000, OFTA issued an industry consultation paper on the licensing framework for 3G mobile services. This first consultation paper identified and discussed several key issues on the licensing of 3G mobile services and consulted with the industry and interested parties on these issues.

Based on the first round public consultation, OFTA issued its second consultation paper on 3 October 2000. This triggered off another round of debate and discussion. OFTA also organised an industry workshop on 5 January 2001 explaining its "Open Network" regulatory framework. Questions such as the definition of MVNO and the measurement of network capacity to be made available were furiously debated. On 13 February 2001, the government finally announced the licensing framework for 3G services in Hong Kong SAR.

Key issues that were discussed during the 11 month long consultation include the following:

- Choice of technical standards,
- Allocation of radio spectrum
- Licensing options
- Open network requirement

Each of these issues is considered in turn as follows.

II. Choice of Technical Standards

Hong Kong's economy is service-based rather than manufacturing-based. Using the telecommunications market to enhance the development of the domestic manufacturing industry is therefore not necessarily a policy priority. This has enabled OFTA to adopt a technology-neutral approach in licensing since 1996, when seven PCS mobile communication licenses were first issued.

Due to the fact that IMT-2000 is a family of 3G standards (WCDMA, CDMA2000, and TD-SCDMA) instead of a single standard, the question was whether Hong Kong should adopt a commonly agreed 3G standard or multiple 3G standards. On the one hand, the government realised that the mandatory use of a single 3G standard would reduce the flexibility of operators in evolving their 2G networks to 3G and limit consumers' choice of terminal equipment. On the other hand, the government is interested in promoting the adoption of technical standards that are compatible with each other from the users' point of view, so that subscribers are not locked-in by any single operator. As with number portability, this will have the effect of reducing the switching cost when subscribers migrate from one operator to another. Another objective is to maximise the convenience of users in accessing roaming services without having to change handsets when they travel to places outside Hong Kong, particularly to the more popular destinations. This question is especially critical given that the time scale of the commercial application of software-defined radio is still uncertain[3].

Based upon these considerations, OFTA intends to offer to prospective operators the use of any 3G standards within their assigned 3G frequency bands, subject to OFTA being satisfied that various technical standards are compatible with each other from the user's point of view. The main consideration is to ensure that customers can easily switch from one network to another and obtain similar services, and to maximise the ease and practicality of roaming services without having to change mobile terminals.

The principle proposed by OFTA has received the support of those parties who have commented on the issue. New World PCS Limited, for instance, emphasised the following:

> "…with 6 existing operators operating 11 networks of different standards in Hong Kong, a technology neutral approach allows a smooth migration from the existing 2G services to 3G services in the future … The technology neutrality approach, which is consistent with the policy adopted by OFTA in previous licensing process, also encourages operators to explore new services available under respective standards"[4].

However, former Cable & Wireless HKT (CWHKT) suggested that the adoption of multiple standards in Hong Kong might result in the requirement for guard bands between bands used by networks of different 3G standards, thus reducing the amount of usable spectrum. Also, it is unlikely that mobile terminals which are compatible with more than one 3G standard adopted by the ITU would be available in the initial years. The CWHKT strongly suggests that only the Universal Mobile Telecom Service (UMTS) of Europe, i.e., W-CDMA, should be adopted in Hong Kong[5].

Currently, there are a total of four 2G standards being used by the six operators in Hong Kong. It is expected that from the launch of 3G services, 3G mobile terminals would have to incorporate a dual mode design to enable backward compatibility with 2G networks. Thus, the availability of backward compatible equipment with existing 2G networks would be a constraint in the selection of 3G standards.

In this case, OFTA affirms that the prospective operators should be permitted to use any 3G standard adopted by the ITU within their assigned frequency bands for 3G mobile services, subject to standard compatibility from the user's standpoint. OFTA expects that the operators would take these requirements into consideration when choosing a 3G standard[6].

III. Allocation of Radio Spectrum

The allocation of radio spectrum is relatively a simple issue in Hong Kong, as most of the IMT-2000 defined spectrum is not presently being occupied. Only a portion of the spectrum is now being used for the operation of Microwave Multipoint Distribution Systems (MMDS) and microwave fixed links. At OFTA's request, users of this band have agreed to relinquish the spectrum by the end of May 2001 for the introduction of 3G mobile services.

Perhaps the major concern lies in how to handle spectrum which is currently used for 2G services. According to World Radio Conference in Istanbul (WRC-2000), spectrums in 800/900 MHz (for GSM) and 1700/1800 MHz (for PCS) bands are allocated for implementing 3G services. In this case, there is a necessity to clarify how to migrate current 2G services to 3G services for current 2G license holders.

---

[3] OFTA (2000) *Licensing Framework for Third Generation Mobile Services – An Industry Consultation Paper*, 21 March

[4] Submission of New World PCS Limited in response to OFTA's first consultation paper. March 2000

[5] Submission of CWHKT in response to OFTA's first consultation paper. April 2000

[6] OFTA (2000) *Licensing Framework for Third Generation Mobile Services – Analysis of Comments Received, Preliminary Conclusions and Future Industry Consultation*, 3 October

In line with its technology neutral policy stance, and in order to allow existing 2G mobile operators to evolve their networks to 3G, OFTA decided to allow existing 2G operators, whether they are successful or not in obtaining 3G spectrum, to use any 3G standard within their assigned 2G frequency bands for 3G services when equipment becomes commercially available. This would be subject, once again, to standard compatibility and the assurance that the interest of existing 2G consumers is adequately safeguarded[7].

However, this has raised concerns from the industry, especially for potential applicants who did not hold a 2G licence, like New T&T Hong Kong Limited. This is because the incumbent 2G operators would in effect be "guaranteed" a right to offer at least some if not all 3G or 3G-compatible services even if they failed to obtain a 3G license or decide not to apply for a 3G license. This is unfair to the other license applicants and the problem would be exacerbated in the initial phases of the service launch, as the full range of services would not yet be available. Therefore, as suggested by New T&T, incumbent 2G mobile operators should not be given any undue advantage or "guarantee" of right[8].

Another concern was that 3G spectrum would be auctioned whereas 2G spectrum was assigned through different procedures and in most cases for free. This might put potential 3G licensees in an unfair position, especially those operators that are without 2G licenses. One submission suggested that the incumbent operators should pay a premium, to be channelled through a mechanism similar to the variation of land leases, in return for the right to provide 3G services in the 2G bands. There was also a view that, as 3G services would not be mature until 2005, the incumbent operators could wait for equipment availability to provide services in their 2G band. If 3G spectrum were granted to them, they should return their 2G spectrum to OFTA[9].

OFTA believes that allowing the incumbents to make use of 2G spectrum for 3G services would be consistent with its policy of technological neutrality. It is up to individual operators to decide which technology is more efficient to use. Therefore, OFTA does not have any objection in principle to existing 2G operators making use of their own 2G spectrum for 3G services until their licenses expire. The resolution of the spectrum issue in the future will be the subject of another round of consultation prior to the expiry of current 2G licenses[10].

IV Licensing Options

Next to 3G standards and 2G spectrum migration, the important issue is licensing, notably the selection of operators. Hong Kong used to assign spectrums to telecommunications licensees on the merits of applications. The major concern in the past, according to OFTA, is providing the opportunity to a maximum number of operators to enter the market and allow market forces to determine the optimum number of operators, as long as the spectrum is available[11]. Although the auction methods was suggested for spectrum allocation for 2G licenses, OFTA did not adopt this method due to the concern that operators will eventually pass the cost on to individual subscribers.

However, spectrum is said to be a scarce resource and as such, must be used effectively. There should therefore be some financial pressure on operators to encourage the efficient utilisation of spectrum. OFTA's preferred method was to set up a performance bond. When issuing the license, the regulator defined an array of milestones that the licensee should meet, mainly on network coverage. Periodically, OFTA will review the licensee's progress with reference to the defined milestones. If licensees fail to reach the milestone, they are liable for the performance bond. Before the license is issued, the bank will evaluate the financial strengths of the applicant and guarantee the ability of the applicant to pay for the performance bond if required. In this manner, licensees are subject to financial pressure to rollout their networks aggressively, while at the same time avoiding the need for a lump sum upfront payment. This reduces the financial burden on operators, particularly new entrants. The method has worked reasonably well over the past years.

In 2000, the 3G licensing generated a spectrum auction fever in Europe. The US$47.5 billion licence fee in Germany and US$33 billion license fee in the UK encouraged governments in other countries to follow the same approach in the expectation of obtaining similar windfalls. However, the negative reaction from the stock market has led to a decline of auction fees in these later-mover countries (Figure 1).

---

7 OFTA (2000) *Licensing Framework for Third Generation Mobile Services – An Industry Consultation Paper*, 21 March

8 Submission of New T&T Hong Kong Limited in response to OFTA's first consultation paper. 22 May 2000

9 OFTA (2000) *Licensing Framework for Third Generation Mobile Services – Analysis of Comments Received, Preliminary Conclusions and Future Industry Consultation*, 3 October

10 *Ibid.*

11 Au, M.H. (1998) Public Mobile Services in Hong Kong – Updates and Future Development, *Proceedings of Asia-Pacific Mobile Communications Symposium*, pp. 33-38

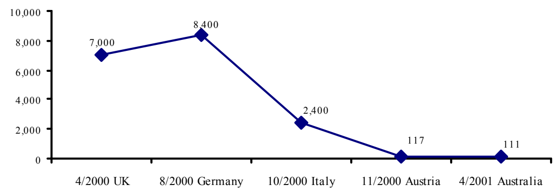


Figure 1 Global Trend of 3G Auction Price (Euro, million)

Nevertheless, the Pound Sterling 22.4 billion 3G auction fee that surpassed the expectation of the UK authority has cast doubt on the appropriateness of the comparative selection approach of OFTA. Emily Lau, Legislative Councillor, commented on Radio 3 of Radio Television HK (RTHK) on 14 May 2000 that

> "Although auctions could bring significant revenue to the government coffers, OFTA is concerned that such additional upfront costs to 3G operators would greatly increase their business risk. This in turn could adversely affect the development of the next generation of telecommunications and Internet services in Hong Kong. I don't think this argument is very sound. Regardless of how the licences are to be awarded, operators are likely to price their services according to how much the customers are willing to pay, and not according to the cost of providing such services … Companies which have over-stretched themselves will suffer, some may even go bankrupt. But other companies will bid for their licences and carry on. This is capitalism at work and should not harm the development of the industry".

In addition to economic considerations, there are also some technical considerations to take into account if OFTA follows its original approach to allocate spectrum for 3G services. In his submission to OFTA on the consultation paper, Xu expressed his concern that "it is very difficult to define the milestone for 3G service due to the fact that there is not a definite date that this service will be commercially available. Nor did the business model. Even the GPRS (2.5G) is still under trial, how can the regulator clarify the rate of deploying 3G services? Wrongly defined milestones will leave the regulator in the embarrassment, but not the operators". Xu was also concerned that if operators are granted licenses for free in Hong Kong, some of them might use Hong Kong revenues to cross-subsidise their customers in the market where they have paid a huge license fees. Indirectly, Hong Kong's consumers would become the victims of huge license fees in other overseas markets[12].

[12] Xu, Y. (2000) Penal discussion at Telecom Asia summit, Hong Kong

Some economists were in favour of auctions but pointed out that the conditions need to be justified once again, in order to ensure that spectrum fees are not artificially inflated. It was suggested that all of the spectrum should be open for the auction and there should not be a pre-determined number of licensees[13].

However, there are some concerns about this. As was suggested by some WRC-2000 delegates, allocating all spectrums to a single immature technology might be risky and harmful for the development of other innovative services[14]. 3G is not the end of mobile communications, 4G even 5G might appear in coming years. In this case, if all spectrums are allocated to 3G, then operators might feel reluctant to upgrade their technology when the new technology is available. For example, when GSM was available, no licensees in Hong Kong reacted to it. It was OFTA who issued a new license to SmarTone and clearly designated that GSM had to be used that digital network began to appear in Hong Kong. As a result, all incumbents had to change their system from analogue to digital. Keeping some spectrum aside might give more flexibility and bargaining power to OFTA in the future[15].

Argument against auctioning the spectrum pointed out that, unlike some other economies, Hong Kong does not need to raise revenue through auctions. Proceeds from spectrum auctions imply a one-off revenue and do not provide a steady base for ongoing government revenue. Also, the spectrum auction process would only favour operators with deep pockets. As such, companies that are more likely to propose high-quality innovative services may fail to obtain licence[16].

Ure worries that the large upfront payments increase the cost of financing, and this acts as a disincentive to roll out networks to more marginal areas. Indeed, it raises the margin. He also worries that the huge upfront payment incubates industrial restructuring leading to oligopoly and/or collusion. In his viewpoint, the upfront payment also weakens the authority of the regulator, especially when large multinational companies are involved who are under enormous pressure to address

[13] See Melody, W.H. (2001) Spectrum Auction and Efficient Resource Allocation: Learning from the 3G experience in Europe, *Info*, Vol.3 No. 3, pp. 5-10

[14] ITU, Newsletter of WRC-2000

[15] Xu, Y. (2000) The debate between beauty contest and auction, *Hong Kong Economic Journal,* 19 May

[16] Au, M.H. (2000) The implementation of third generation mobile services in Hong Kong, *Proceeding of Asia-Pacific Mobile Communications Symposium: 3G and M-Commerce*, pp. 11-17

the interests of their shareholders and financers, and keep an eye on the price of their stock[17].

Liu argued that 3G is a high-tech industry - its social benefits is much larger than the commercial benefit and will generate tremendous externalities. In an environment not favourable to high-tech investment, the success of 3G might be enhanced if the government could share part of the risk with investors. As this is a revenue risk but not a cost risk, the government's total risk is close to nil. In this case, the government is probably the most appropriate party to undertake the risk[18].

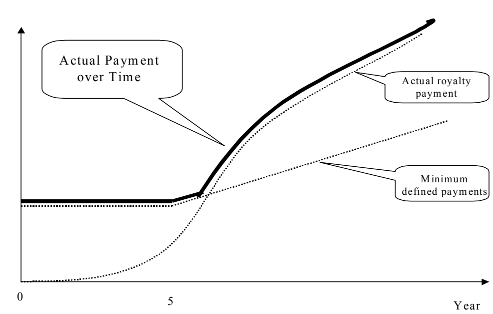
Taking into consideration of all of the views from the different sectors, the government formally released its 3G licensing framework on 13 February 2001. The decision was taken to issue four licenses by way of auction after conducting a pre-qualification exercise. The pre-qualification process is intended to be relatively light, but will involve setting certain minimum criteria on investment, network rollout, service quality, financial capability, etc. As to the auction process, the government has chosen a royalty-based proposal that requires the bidder to pay a certain percentage of their annual 3G revenue turnover determined by the auction. The royalty payment will be subjected to a guaranteed, minimum payment. The government believes this method will encourage market entry while keeping the financial burden on operators at a manageable level[19].

The particularities of the royalty auction are as follows:

> "Bidders would be asked to bid for a level of annual royalty by way of percentage of turnover from their 3G services network operations. Successful bidders who win the 3G licences at a certain royalty percentage bid would do the following:
>
> (a) for the first five years of the license: They will pay a guaranteed, minimum royalty payment fixed by the government. They will pay the same fixed amounts for this period regardless of their actual turnover. This is because it will be difficult in these initial years of the 3G licences to distinguish between second generation mobile service (2G) and 3G network revenues, if the 3G licensee is also an existing 2G operator;
>
> (b) from year six to the end of the licence period: They will pay royalties to the Government according to the royalty percentage determined by the auction. The same royalty percentage will apply to all licensees. The actual royalty payment will differ from licensee to licensee as their 3G revenue turnover will be different. However, the royalty paid by each licensee should not be less than the guaranteed, minimum royalty payment fixed by the Government. In other words, the Government collects the royalty based on actual turnover, or the guaranteed, minimum royalty payment, whichever is the higher; and
>
> (c) throughout the whole licence period: They will need to provide a 5-year rolling guarantee for each of their guaranteed, minimum royalty payment."[20]

According to the Government, this proposed method best meets its policy objectives. It is "pro-entry" as it alleviates the burden on successful 3G licensees, and allows the government to share the upside of the future 3G services market. It is also an efficient method of allocating licences to those bidders with the best business case, as the payment will be in the form of royalty and therefore will depend on the actual performance of each licensee. The guaranteed, required minimum royalty payment will minimise credit risks for the Government, and reduce the costs that may be passed on to consumers[21]. Figure 2 illustrates the royalty payment scheme.

Figure 2. Royalty Payment Scheme



[17] Ure, J. (2000) Response to OFTA's consultation paper, *http://www.ofta.gov.hk*

[18] Liu, C.W. (2001) Payment method of 3G licensing: new taxing paradigm of new economy. Hong Kong Economic Journal Monthly, Vol.24, No.10, pp.32-34

[19] Press release issued by Information Technology and Broadcasting Bureau on 13 February 2001: The licensing framework for third generation mobile services

[20] *Legislative Council Brief: licensing Framework for third Generation Mobile Services,* File Ref: ITBB CR 7/23/10(01)

[21] *Ibid.*

This proposal has been very well received, despite the fact that there are some controversies on the details of the auction design. For example, the government would like to settle on the final price when the fourth bidder from the bottom leaves the auction process, while the industry insists that the final price should be established when the fifth bidder from the bottom leaves[22]. In this way, the price should be lower and the cost of licensees will be reduced. The Government finally reaches a compromise and agrees to establish the price when the fifth bidder from the bottom leaves.

On 19 September 2001, the auction was conducted. Four applications for 3G licenses in Hong Kong were received and in accordance with the rules set out in the Information Memorandum which provided guidance for the 3G spectrum auctioning. The four bidders have each been awarded a license at reserve price, i.e. five per cent royalty subject to a minimum payment of HK$50 million for each of the first five year, and rising minimum payments from year six onwards.

V. The Open Network Requirement

Due to the limitation on spectrum, only four 3G licenses have been issued. To enhance competition, OFTA proposed the concept of separating service provision from network operation. This means that a service provider need not at the same time be network operator, although that possibility is not excluded. This is comparable to the European concept of the Mobile Virtual Network Operator (MVNO). In this case, the MVNO would not own a license to use radio spectrum, but would have access to the radio networks of one or more of the mobile network operators and be allowed to build and operate parts of the networks not requiring the use of radio spectrum (e.g. elements of an intelligent network). These service providers would then be able to offer 3G services to customers without actually operating the radio networks. The regulator suggests 3G licensees to open 30-50% of the network to MVNOs[23].

Almost all incumbent operators expressed their concern about MVNO. The major concern lies in the potential technical and financial difficulties for physical network operators to meet this requirement. It was argued that opening the networks to all competitors is technically inefficient for the network operators to rollout and maintain networks with an unknown requirement in terms of capacity just for the purpose of offering it to service providers. The business viability of the network operators would be jeopardised. The incumbent also argued that the future 3G value chain would allow customers to access service and content providers. The need for MVNO might not even exist[24].

Some analysts argue that opening the network to MVNOs is akin to permitting "free riders". As a result, some investors may not join the bidding for 3G licenses; however, once the risk is undertaken by initial spectrum bidders and 3G is proven to be successful, they would apply for a MVNO license and begin providing services. This scenario is seen as patently unfair to the 3G licensees[25].

Supporters of the open network requirement stressed the importance of non-discretionary access to networks by MVNOs and minimising collusion among big players who were both network operators and service providers. One submission suggested that as many as 20 MVNOs might be licensed by way of auction, enabling a large number of companies in different domains to enter the 3G services market[26].

Based upon feedback, the government proposed a regulatory framework for its open network:

(a) Successful bidders must open at least 30% of their 3G network capacity for use by non-affiliated companies to operate as MVNOs and/or content providers. More capacity could be opened up if they wish to do so by commercial agreement. However, to preserve the commercial incentive of 3G network operators to develop their networks, the regulator would not intervene for a MVNO or content provider if that operator/provider already has access to capacity equivalent to 30% capacity of a network operator.

(b) The wholesale prices for MVNOs' access should be negotiated commercially with the 3G licensees. However, if commercial negotiation fails, the regulator reserves the right to make a determination based on fair interconnection principles. A sufficient return on cost of capital will be allowed, reflecting the higher risk of 3G service investment, when regulator makes the determination to ensure that the investment incentives are preserved.

---

[22] The time for 3G auction might be postponed, Hong Kong Economic journal. 30/05/2001

[23] OFTA (2000) *Licensing Framework for Third Generation Mobile Services: An Industry Consultation Paper*, 21 March

[24] OFTA (2000) *Licensing Framework for Third Generation Mobile Services – Analysis of Comments Received, Preliminary Conclusions and Future Industry Consultation*, 3 October

[25] Hausman, J. (2001) Guest speech at Bloomberg Auditorium, Hong Kong, 23 May

[26] See http://www.ofta.gov.hk/frameset/home_index_eng.html

(c) Content providers will buy capacity at tariffs set by the 3G licensee, reflecting all relevant costs and the above-mentioned cost of capital. The regulator would only intervene in cases of unfair, discriminatory treatment or on anti-competitive grounds.

(d) Measurement by the regulator of the capacity sold to non-affiliated companies will not be necessary unless the 3G licensees refuse to supply the requested capacity. The licensee should then provide evidence to the satisfaction of the regulator that 30% of their capacity has already been opened up. The regulator is prepared to accept alternative methods of measurement proposed by operators including the simplest documentary proof of the total capacity sold, e.g. in the contracts or agreements with non-affiliated companies[27].

In addition to the open network requirement and ensuring a level playing field for both new entrants and incumbent operators in the 3G services market, the government proposed a final requirement on certain successful 3G licensees. Those licensees that are also 2G operators are to provide roaming to successful non-2G licensees in locations not yet covered by the latter's 3G networks during the initial years. This mandatory requirement is crucial, because it takes time for new entrants to roll out a network with coverage comparable to that of a mature 2G network. These new entrants are expected to have rolled out their own network and to satisfy the mandatory requirement over time. The requirement will end on a "sunset" date, intended to be five years after licences were issued[28].

By the end of January 2003, there are six companies have obtained MVNO licences in Hong Kong. These companies are: Trident Telecom Ventures Ltd., China Motion Telecom (HK) Limited, China Unicom International Limited, CITIC Telecom 1616 Limited, i100 Wireless (Hong Kong) Limited, China-Hongkong Telecom Limited. They are going to targeting different market niches. For example, both China Unicom and China Motion has begun to use its MVNO license to provide economic and seamless services for frequent travelers between Hong Kong SAR and mainland China, while Trident is targeting overseas visitors. They are currently using 2G technology for service provision, and are planning to migrate to 3G technology once the networks are ready.

VI. Summary

This study provided a comprehensive review of the debate in the process of formulating 3G licensing policy in Hong Kong. It also gave an introduction to the 3G licensing framework in Hong Kong.

3G imposes some challenges on Hong Kong SAR. Due to limited spectrum and hence limited number of licenses, it is unclear whether the current competitive 2G market in Hong Kong can be successfully transferred to a competitive 3G market in the future. Also, OFTA's traditional licensing and regulatory scheme for 2G services has been threatened by economic and technical constraints of 3G services.

OFTA's 3G licensing framework, to a certain extent, relieved the licensees from a heavy burden of paying huge upfront spectrum fee. This might be able to accelerate the paces of rolling-out 3G networks in Hong Kong. The open network obligation of the licensees may also enhance competition in the 3G market in the fact that the spectrum can only accommodate limited network operators.

Additionally, the commercialisation of spectrum may also pave the way for free trade of spectrums among operators, although this issue has not been on the top agenda of OFTA yet. In fact, as spectrums were assigned for free in 2G services, OFTA has had to be fair by assigning the spectrum to all licensees in an equal way, although this equally assigned portion may be over-sufficient for small operators but insufficient for large operators. Several applications for more spectrum by large operators have been rejected by OFTA in the excuse of fairness, and these operators have to increase the efficiency of spectrum by investing on spectrum compression system, which has affected the service quality, increased the operation cost and raised the price of services. In the meantime, the small operators have huge surplus spectrum due to their small customer base but they have been restricted to trade the spectrum, as they have obtained the spectrum for free. This dilemma scenario maybe changed under the new 3G licensing framework, as the licensees, to some extent, owns the spectrum during the defined period of the licenses.

3G mobile is not just an innovative technology and service, but it also brings about new economic and regulatory dynamics. Whether or not the innovative 3G licensing framework and open network requirement of the Hong Kong Government will be successful remains to be seen.

[27] *Legislative Council Brief: licensing Framework for third Generation Mobile Services,* File Ref: ITBB CR 7/23/10(01)

[28] *Ibid.*

Adaptive Spectrum Radio: A Feasibility Platform On The Path To Dynamic Spectrum Access

William D. Horne*, Peter Weed, & Daniel J. Schaefer

The MITRE Corporation
7515 Colshire Drive, McLean, VA 22102
* Phone: 703-883-6198; Fax: 703-883-7068; Email: whorne@mitre.org

Abstract

A new paradigm for managing and allocating the electromagnetic spectrum is now possible due to advances in technology and a receptive environment within regulatory agencies. New radiocommunication technology provides a key component to this possibility as it increases the flexibility for radio transmissions to dynamically adapt and access the spectrum. Such dynamic spectrum access, however, requires revised policies and regulations since traditional spectrum management allocates a band of frequencies to specific services and assigns a single frequency or group of frequencies to specific radios.

In order for the policy community to gain confidence in the possibilities of the new paradigms, the technology developers need to demonstrate the capabilities of the systems that enable dynamic spectrum access. The MITRE Corporation has developed a feasibility radio platform that demonstrates the principles for dynamically accessing the spectrum as it adapts its frequency and modulation to exploit spectrum gaps both in frequency and time. This paper discusses both a path to demonstrating the technology for the policy community as well as the MITRE developed Adaptive Spectrum Radio (ASR) testbed.

Introduction

The era for overhauling the framework and processes by which governments manage the radio frequency spectrum is upon us. Government policy initiatives, such as the United Kingdom's reorganization of its spectrum policies based on Martin Cave's report [1] and the United States Federal Communications Commission's (FCC) Spectrum Policy Task Force [2], along with private efforts at various "think tanks" indicate that regulatory agencies and the private sector are seriously reviewing options for spectrum management reform. While traditional economic arguments for reform have been around since Leo Herzel's proposal [3] and Ronald Coase's seminal paper [4] in the 1950s, the key factor impelling this current review is recent technological advances.

The ability to dynamically and adaptively access the spectrum is one of the most important emerging technologies that enable spectrum policy reform [5]. Advances in digital signal processing enabled by increasing capabilities of microprocessor and related components allow radios to dynamically change transmission parameters including data and coding rates, modulation, and even frequency. Digital signal processing-based radios, typically referred to as Software Defined Radios (SDR), provide the platform for these dynamic radio systems.

Forms of adaptive spectrum access have been implemented in the past including current and next generation mobile systems, but the key difference in the emerging adaptive spectrum radios is the increased flexibility, both in speed and variety of parameters, to alter transmissions that dramatically increase the ability to adapt to a changing radio frequency environment. A rudimentary form of adaptive spectrum access implemented in some systems operating in unlicensed bands is the selection of a frequency channel based on measurements of activity. More advanced implementations of adaptability include the various methods for changing data rate, code rate, and modulation order implemented by second (2G) and third generation (3G) mobile standards [6-7]. For example, Enhanced Data Rates for GSM Evolution (EDGE) adapts the transmission parameters modulation (GMSK or 8-PSK) and coding rate. This type of adaptability is well suited for packet data and in situations with slow variations in fading and distance loss over the cell coverage area.

The increased capability of emerging adaptive spectrum access systems requires changes to the methods for managing the spectrum since they enable new systems to opportunistically access the spectrum and to change their transmission characteristics, challenging notions of fixed frequency assignments and rigid certification–pillars of the current spectrum regime. Before changes can be implemented in spectrum management, the technology community needs to present demonstrations and associated studies showing the capabilities and possibilities. This paper reviews factors necessary to demonstrate the technical principles for dynamic spectrum access and the interrelationship between technical demonstrations and policy changes. One element of this review is the use of prototype systems to investigate options under consideration by the policy community. Demonstrations are not intended as final indicators of future applications or design, but they do provide a means to explore policy options.

To illustrate the use of prototype systems, this paper also presents the MITRE Adaptive Spectrum Radio (ASR) testbed [8]. The testbed architecture supports SDR development using Application-Specific Integrated Circuit (ASIC), Digital Signal Processor (DSP), and Field Programmable Gate Array (FPGA) components, but with MITRE developed software. Success thus far demonstrates not only the feasibility of dynamic spectrum access but also allows future exploration of policy considerations.

Dynamic & Adaptive Spectrum Access

Forms of dynamic spectrum access already exist, but the full concept for adaptive spectrum access whereby a system can adjust its transmissions in a cooperative or even an un-cooperative environment is only now becoming possible. The definition of a fully "dynamic" or "adaptive" system is not clearly delineated, but the capabilities of such a system would include:

- Sensing the radio frequency environment;
- Controlling its transmissions based on measurements and other a priori information in an autonomous, opportunistic, and real-time fashion;
- Adjusting multiple transmission parameters including, but not limited to, frequency, power, modulation, signal timing, data rate, coding rate, and antenna; and,
- Operating in cooperative networked systems and/or environments with non-cooperating systems (i.e., opportunistically accessing spectrum).

One key benefit for adaptive spectrum access is the potential to improve the efficiency in spectrum use. By adjusting transmissions, adaptive systems can utilize unused frequencies even if they vary over time. In addition, adaptive systems may maintain a quality of service in a changing environment while also adjusting emissions to reduce interference to other systems.

Numerous research organizations and companies are investigating adaptive spectrum access systems. One notable effort is the Defense Advanced Research Projects Agency (DARPA) neXt Generation (XG) program [9]. This program is developing an access technology to increase the ability of military systems to access spectrum adaptively and to ensure the operation of new systems, without the extensive, frequency by frequency, system by system, coordination now required in each nation where a system will operate.

Demonstrations Path to a New Spectrum Policy

The impact of adaptive systems to policy is extraordinary as it challenges the current basis for spectrum management. At present, mechanisms for accessing spectrum and limiting interference typically assume "fixed" or statically-assigned frequency assignments for a system within a band allocated to a defined radio service sharing characteristics of the system under consideration. Often, additional spectrum is set aside to provide a guard band or minimum frequency separation between neighboring spectra, and a set of technical or operational rules is defined for usage in cases where a common spectrum is shared or where neighboring systems require protection from interference. In addition, regulatory agencies require certification of systems that they operate in accordance with technical parameters governing frequency and emissions, both in-band and out-of-band.

Adaptive spectrum access systems do not operate in such a rigid manner, and, indeed, their primary benefits are derived from the flexibility afforded them by operating in a dynamic way. Since the current methods of assigning fixed frequencies inhibit the advantages of adaptive systems, spectrum management policies and procedures need updating to accommodate them.

The FCC recognizes this need in their Spectrum Policy Task Force Report [2]:

> "Because new, smart technologies can sense the spectrum environment and because they have the agility to dynamically adapt or adjust their operations, increasing access to the spectrum for smart technologies, such as software-defined radios, can improve utilization, through more efficient access, of the radio spectrum without detriment to existing spectrum users." (p. 15)

The report goes on to recommend that "the Commission develop access models that take this new technological potential into account." While the FCC report identifies these needs, it does not define specific rules or "rights" to accommodate these technologies nor does it present a specific path by which to delineate them.

One path that can assist the FCC and other regulatory bodies as they begin to define new spectrum access models is the use of demonstrations of technology. In order for the policy community to gain confidence in the possibilities of the new regulatory paradigms, technology developers need to show the capabilities of the systems that enable dynamic spectrum access. The two principal objectives for demonstrations are:

- Inform the policy and regulatory community of the feasibility for adaptive spectrum access; and,
- Identify and investigate considerations for policies using adaptive radio demonstration platforms.

By understanding the basic operations of adaptive radios, the policy and regulatory community can better revise and update rules and procedures. In addition, reviewing technology in its early development phases enables regulators to inform developers ways in which the technology may need to evolve to accommodate policy and regulatory concerns. This two-way process can mutually benefit all sides by identifying issues early and by preventing contentious proceedings as has been seen in recent regulatory deliberations involving emerging technologies like ultra-wideband (UWB).

A good example of a new technology awareness program is the DOD's Defense Information Systems Agency's (DISA) Emerging Spectrum Technology program, led by the Defense Spectrum Office (DSO). This project intends to proactively understand policy ramifications of new technology having military benefits to ensure that policies do not inhibit their introduction. Other organizations, including the FCC, are also enhancing technology awareness efforts.

The demonstration component for programs reviewing policies occur at different stages of both the technology's and the policy's development. Figure 1 illustrates the phases in which demonstrations are applied to support policy revisions. The key to this paradigm is the evolutionary nature where policy is prepared in conjunction with and influenced by technology development.

The principle objective for demonstrations during the conceptual phase is to show the notion of new technologies, like adaptive spectrum access. Demonstration testbeds used in later phases can identify and explore policy considerations and options. Three areas of policy that may require updating to accommodate adaptive spectrum access systems are: certification, spectrum access, and interference mitigation. Although adaptive radio platforms could be used to explore certification issues (e.g., integrity of the software), such issues are not unique to adaptive systems so they may be better explored using other SDR platforms. Because of the nature of adaptive systems, issues involving spectrum access and interference mitigation are inextricably linked. Consequently, the design and demonstration of feasibility platforms may be used to identify and explore spectrum access and interference mitigation policy considerations including:

- **Band/Channel Blocking**. Can adaptive systems operate with policies that prohibit the transmissions in specific channels, even non-contiguous ones? Can different emission levels be set for different channels?
- **Interference Temperature**. The FCC Spectrum Policy Task Force [2] introduced the concept of "interference temperature" as a means for defining the environment in which systems must operate. Specifically, the Task Force recommended that "the Commission shift its current paradigm for assessing interference–based on transmitter operations–toward operations using real-time adaptation based on the actual RF environment through interactions between transmitters and receivers." Consequently, adaptive systems may be an integral part of policy.

  In such a regime, the defined "temperature" levels need to be defined. Should policies define permitted limits for transient interference to account for imperfections in the algorithms and protocols for accessing the spectrum? What are the achievable sensitivity levels for sensing the environment?

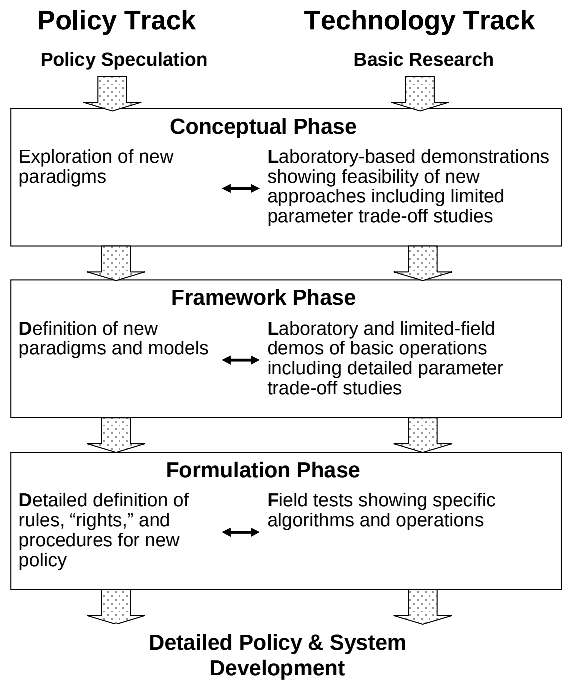


Figure 1. Policy/Technology Development Phases

- **Technical Parameters**. Should policies define acceptable parameter values associated with the adaptive algorithms (e.g., time required to sense environment, latency between "acceptable" measurement and transmission, etc.)?
- **Databases of Existing Use**. Do policies need to be adopted that require the availability of databases to assist the operation of adaptive systems?
- **Environments**. In what environments can adaptive systems operate (e.g., dense voice traffic, radar, etc.)?
- **Refarming**. Can the introduction of adaptive systems improve the efficiency in certain bands so that legacy systems may continue operation rather than relocate?
- **Secondary Markets**. Can adaptive spectrum access systems enable secondary market trading?

The design and demonstration of feasibility testbeds can provide preliminary answers to the above questions. Although analysis and simulation can also assist in answering these questions, the design and demonstration of a feasibility platform provide additional assurance during the early phases of policy definition. Such a demonstration radio does not need to be complete or fully functional, since its basic operation can stimulate discussion and identify the policies that need to be considered. The next section discusses the MITRE Corporation's Adaptive Spectrum Radio testbed which can be used for these early explorations of policy changes.

MITRE's Adaptive Spectrum Radio: Overview

The MITRE Corporation, in partnership with its government sponsors, has studied adaptive spectrum utilization as a potential way to ease congested spectrum. The level of unused capacity in currently deployed communication systems gauged by previous work [8] indicates a potential for the adaptive spectrum utilization concept relying, in part, on technology identifying momentarily unused capacity within a channel and then adapting a waveform to make use of it. Gaining access by sensing the communication media before transmitting has been used effectively in local area networks, including Ethernet's CS-CDMA scheme. Like CS-CDMA, an adaptive spectrum radio senses channel (spectrum) occupancy before transmitting, but unlike CS-CDMA, the system adapts the transmit waveform to take advantage of unused spectrum.

In general, an adaptive system must form its transmit waveform to exploit unused channels. The basic methods of separating channels are frequency division (FDM), time division, and code division. From a general perspective, an adaptive radio must be able to alter waveform characteristics pertaining to one or more of these separation techniques. For example, in the case of FDM, an adaptive system waveform must exploit momentarily unused frequency channels. Additional efficiencies accrue if an adaptive FDM waveform simultaneously exploits a discontiguous set of unoccupied frequency channels because it would, on average, support higher bandwidths and availability.

Viable communication using adaptive systems requires another important capability; namely, "opportunistic" media access control (MAC). The adaptive spectrum utilization concept requires that individual radios inform each other where idle channels exist and when to use them—without using fixed control channels. Since adaptive systems in different locations may identify different sets of unoccupied channels, there needs to be a means of selecting a mutually agreeable subset of channels, a "joint occupancy vector," and when to use it. Thus, an adaptive system must conduct a control dialogue over an unoccupied channel without first knowing what, if any, unoccupied channels each radio senses. In the case of FDM, the adaptive system cannot rely upon a fixed frequency channel for control because unoccupied frequencies may change from one moment to the next and from one location to another.

Realizing all three of these key adaptive functions, sensing the spectrum for unoccupied channels, adapting the transmit waveform, and designing an opportunistic MAC, presents many engineering challenges, but MITRE research indicates that current technology makes such communications possible. Advances in micro-electronics has led to FPGA cores and DSPs that can compute correlations and Fast Fourier Transforms (FFTs) fast enough to support timely channel occupancy estimation. Expanding research into ad hoc networks is likely to overcome many challenges similar to those inherent in implementing an opportunistic MAC. As discussed below, digital signal processing makes adaptive waveform synthesis possible. To tie these three key functions together, Figure 2 depicts the overall spectrum utilization process for an FDM system; in the figure, two adaptive systems make estimates of spectral occupancy and then negotiate to use a common subset of unoccupied channels. Although it is not explicitly shown, the process in Figure 2 constitutes a single burst cycle and would repeat itself at a rate dictated by requirements to control interference.

MITRE developed an adaptive spectrum approach to operate in the midst of an FDM communication system such as advanced mobile phone service (AMPS). Figure 2 visualizes an adaptive waveform that transmits data over the unoccupied frequencies. Such a waveform requires non-contiguous modulated carriers with excellent roll-off characteristics outside its passbands. One method could sum a series of individually modulated and filtered carriers to fill each unoccupied gap in the spectrum, but a more integrated solution improves computational efficiency. The following section discusses work that successfully implemented an adaptive waveform synthesis using a signal similar to orthogonal frequency division multiplexing (OFDM). The core IFFT process provides a potential advantage since, as the spectrum occupancy estimator for an FDM system uses an FFT, it might be possible to integrate the sensing and demodulation processes—or at least reduce DSP loading.

MITRE ASR: Architecture & Design

MITRE developed a testbed to demonstrate feasibility of the adaptive spectrum radio concept with the intention of providing four basic capabilities: (1) periodic estimation of a channel's occupancy state, (2) periodic adaptation of a time-limited waveform in response to occupancy state estimates, (3) periodic "joint occupancy vector" negotiation with subsequent burst data transfer, and (4) measurement of impairments to primary users. The testbed architecture reflects the desire to support these capabilities as well as a

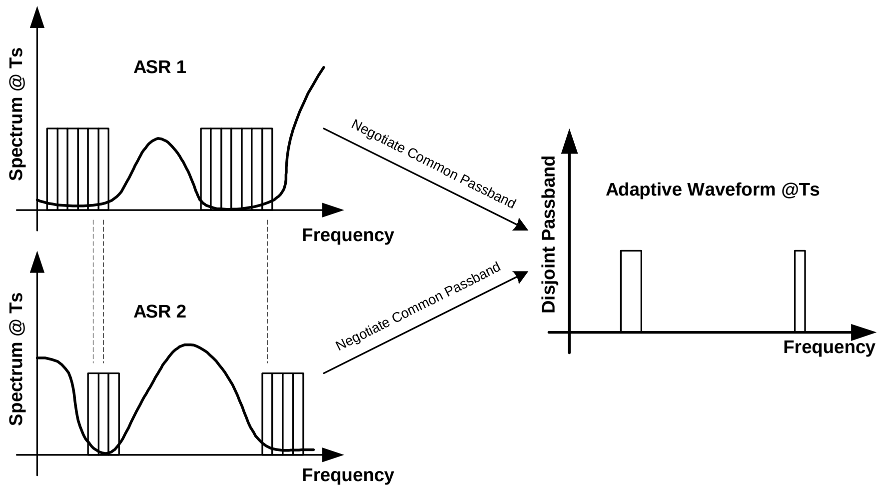


Figure 2. Adaptive Spectrum Approach

compromise between flexibility and performance. Flexible signal processing was important because experience with control schemes and algorithms could evolve to new code, favoring an architecture centered on programmable devices (DSP). Alternatively, a more realistic proof-of-concept demonstration would involve wideband signals, generally demanding application specific solutions. Figure 3 depicts an architecture that blends the speed of ASICs with the flexibility of DSPs.

The MITRE testbed uses commercial-off-the-shelf (COTS) signal processing hardware that implements the balanced approach shown in Figure 3 with a significant portion of its processing occurring within the DSP element. Thus, the testbed is a software defined radio platform. However, it is important to realize that high-speed processing within the ASICs allow the testbed to demonstrate the adaptive spectrum concept over a much wider bandwidth than would be possible without them.

Figure 4 shows the high-level hardware setup that constitutes the testbed. Two small VME chassis and a host computer make up most of the hardware, excluding test equipment like signal generators. Comparing it with Figure 3, the mezzanine boards contain the ASIC and FPGA elements while the quad-DSP carrier board provides programmability. Development within the testbed involves understanding the strengths and weaknesses of the architecture shown in Figure 3 and then mapping the required signal processing into the appropriate element. As noted earlier, promoting flexibility was prudent so most design activity focused on creating real-time code for the DSPs. COTS software running on the host computer (refer to Figure 4) provides a typical integrated environment for writing, running, and testing embedded code.

MITRE has demonstrated capabilities (1) and (2) listed above by implementing the Adaptive FDM Burst Modulator. Figure 5 shows how the signal processing maps into the hardware. As the diagram shows, the system implemented three principle functions: test signal generation, channel occupancy estimation, and adaptive FDM burst modulation.

The **Test Signal Generator** cycles through 15 different “spectrum occupancy” scenarios to create a test signal for the Adaptive FDM Burst Modulator. Each scenario has a different spectrum occupancy state defined by the on/off state of 256 modulated carriers. The AMPS forward control channel (FOCC) RF specification served as the basis for synthesizing each of the 256 modulated carriers that make up the test signal. AMPS specifies a 30 kHz carrier spacing, so if all the carriers were on, the test signal would spread across 7.68 MHz.

Two separate processes make up test signal generation. First, communication systems analysis software creates a pre-processed sample sequence for each spectrum occupancy scenario by simulating FOCC modulation on 256 carriers. In each simulation, unique sets of frequency-shifted carriers sum to form one of the spectrum occupancy scenarios. Each of the 15 complex sequences merges to form a sample-indexed file. The second process performs arbitrary waveform generation using the pre-processed file as input. Key functions include high-speed interrupt-driven data buffering, digital upsampling, and digital-to-analog conversion.

The **Channel Occupancy Estimator** periodically produces an “occupancy vector” (OV) that tells the Adaptive FDM Burst Modulator what waveform to synthesize. The OV tracks the test signal so that the burst modulator avoids using spectrum occupied by the simulated primary user.

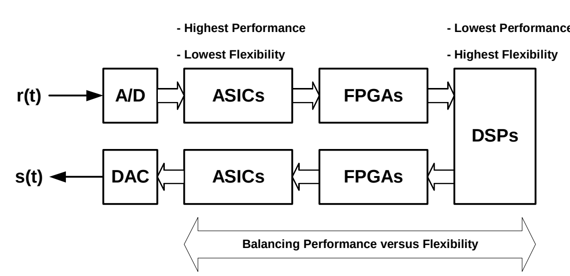


Figure 3. Adaptive Spectrum Radio Architecture

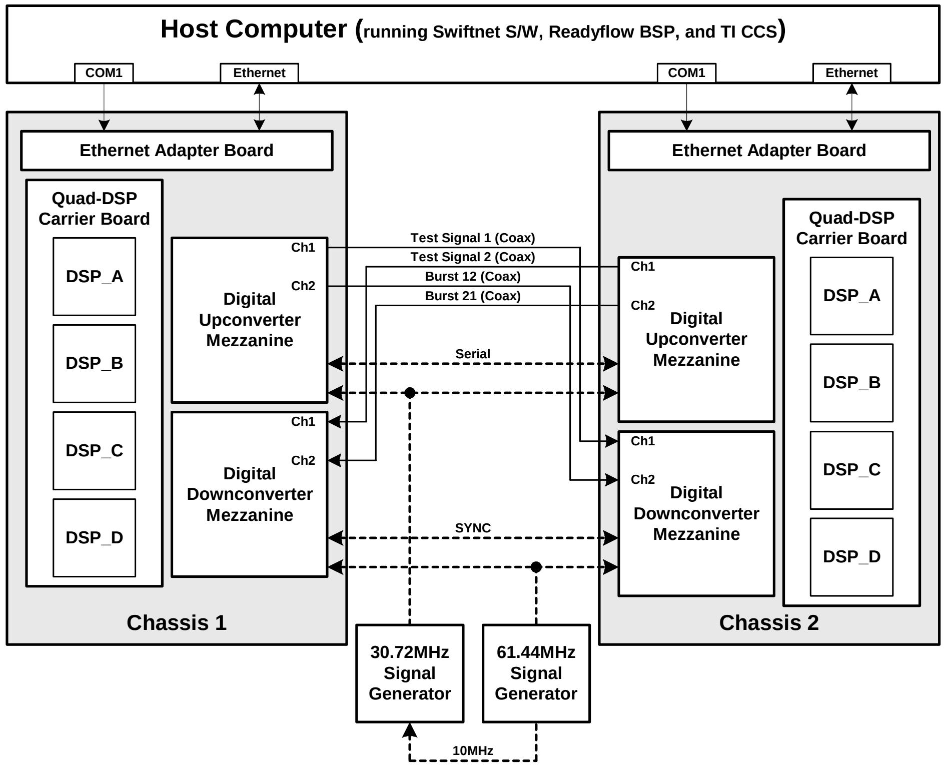


Figure 4. High-Level Testbed Architecture

Six functions make up channel occupancy estimation: analog-to-digital conversion (ADC), channelized digital downsampling, power spectrum estimation, occupancy determination, data fusion, and guard band insertion. ADC periodically captures a block of wideband samples. Hardware constraints make it necessary to channelize the digital downsampling process, so each block of samples passes through 8 digital downconverters that output downsampled blocks of data. After downsampling, the processing used an FFT to transform the time domain data into data representing estimates of power spectral density (PSD). To discriminate between relatively high-power and low-power signals, a threshold test converts the PSD data into binary variables representing the results of occupancy/no occupancy decisions. Data fusion merges the channelized data into an OV containing 256 binary decision variables; one variable for each of the carriers within the test signal. Finally, the Adaptive FDM Burst Modulator has frequency roll-off limitations between its populated and unpopulated carriers, so certain unoccupied channels are treated as if they were occupied to prevent their use. In essence, the Channel Occupancy Estimator alters the OV to avoid primary user interference by accommodating filtering limitations within the burst modulator.

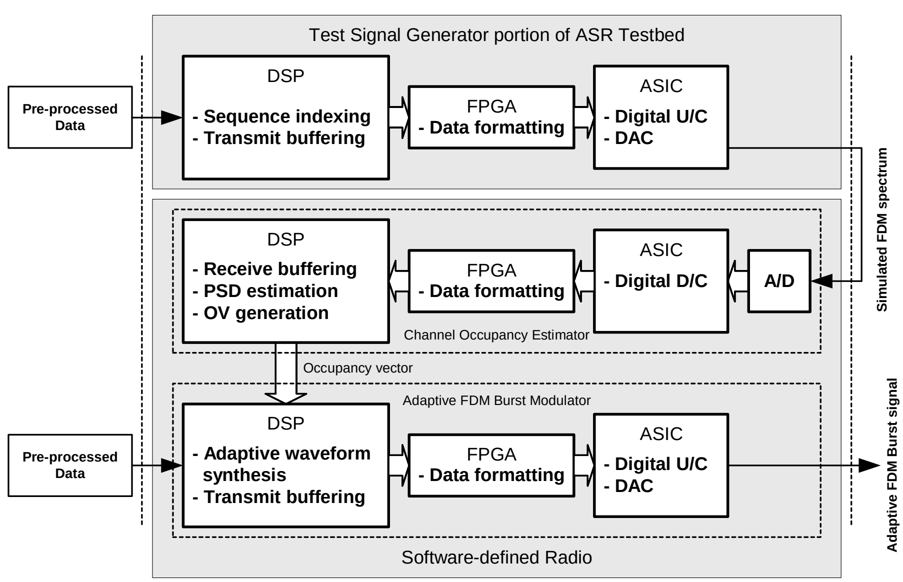


Figure 5. Adaptive Spectrum Radio Transmit Functional Allocation

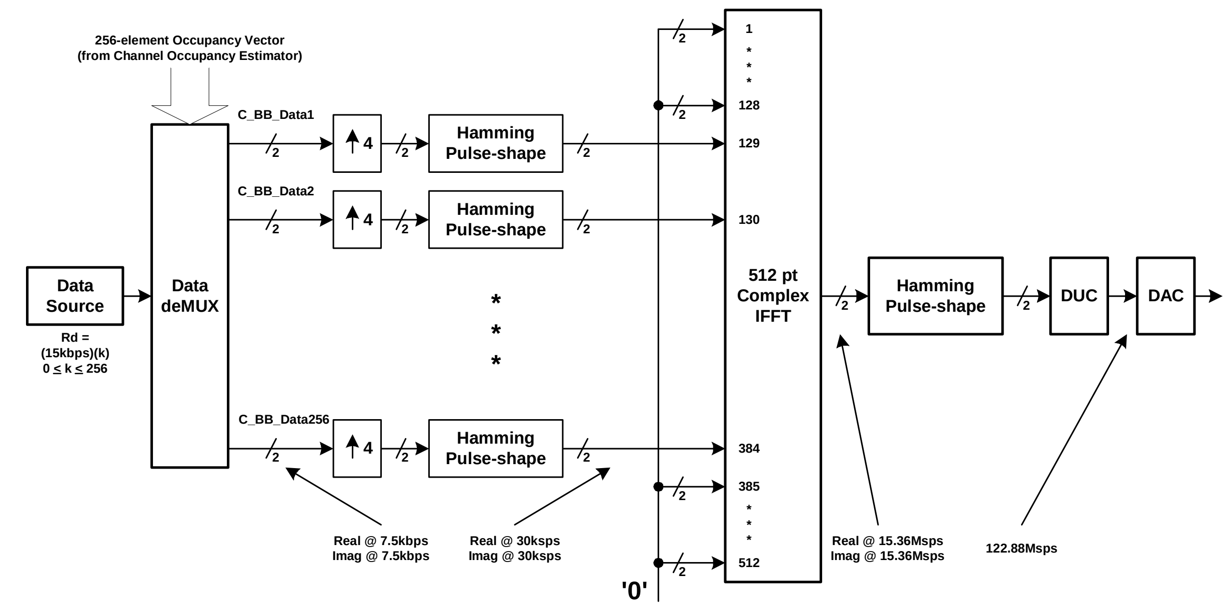


Figure 6. Burst Modulator Signal Processing

The Channel Occupancy Estimator periodically interrupts the **Adaptive FDM Burst Modulator** with an updated OV. After each OV-update interrupt, the burst modulator responds by synthesizing a waveform tailored to the OV. Figure 6 depicts the high-level signal processing carried out by the burst modulator. Variable-rate data enters on the left of the block diagram and a low-IF signal with a maximum bandwidth of 7.68 MHz exits on the right. The data de-MUX uses OV updates to determine which of the 256 frequency bins should receive data during the current burst cycle. All operations are complex, so every unoccupied channel should receive 2 bits from the de-MUX—assuming there is enough data in the transmit buffer. An upsampler allows Hamming pulse-shaping using 4 samples per real and 4 samples per imaginary data bit. Prior to performing a complex IFFT, pulse-shaped data fills the central 256 frequency bins of the IFFT. Null data fills 128 outlying frequency bins on each side of the IFFT input vector. After the IFFT transforms data samples from the frequency domain to the time domain, Hamming pulse-shaping forms the block of complex samples. Finally, upsampling with quadrature upconversion, followed by digital-to-analog conversion (DAC), completes a single burst cycle within burst modulator processing.

MITRE intends to add additional capabilities in future work. While the MITRE testbed provides a feasibility platform, it is not intended as the only architecture or waveform for future adaptive radio systems.

Summary

As the policy and regulatory community considers dramatic changes in the methods and processes used to manage the spectrum, technology demonstrations will prove useful. Platforms, like the MITRE Adaptive Spectrum Radio testbed, provide an opportunity to not only indicate the feasibility of adaptive spectrum access techniques but to also explore policy options involving this emerging technology.

References

[1] M. Cave, “Review of Radio Spectrum Management,” Department of Trade and Industry-Her Majesty’s Treasury, Mar 2002

[2] Spectrum Policy Task Force Report, FCC, ET Docket No. 02- 135, Nov 2002

[3] L. Herzel, Note, ‘Public Interest’ and the Market in Color Television, 18 U. OF CHIC. L. REV. 802, 1951

[4] R. Coase, “The Federal Communications Commission,” 2 J. L. & ECON. (1959), 1.

[5] R. Berezdivin, R.Breinig, & R. Topp, “Next Generation Wireless Concepts, Communications and Technologies,” IEEE Commun. Mag., Mar 2002

[6] S. Nanda, K. Balachandran, & S. Kumar, “Adaptation Techniques in Wireless Packet Data Services,” IEEE Commun. Mag., Jan 2000

[7] S. Catreux, V. Erceg, D. Gesbert, & R. W. Heath, Jr., “Adaptive Modulation and MIMO Coding for Broadband Wireless Data Networks,” IEEE Commun. Mag., Jun 2002

[8] D. Schaefer, “Wide Area Adaptive Spectrum Applications,” MILCOM 2001, Vienna, VA, Oct 2001

[9] P. Marshall, “Spectrum Management Protocols in the DARPA Next Generation Communications Program,” SDR’02 Tech Conf, San Diego, Nov 2002

Efficient Collaborative (Viral) Communication in OFDM Based WLANs

Aggelos Bletsas and Andrew Lippman

Media Laboratory
Massachusetts Institute of Technology
20 Ames St, E15-495, Cambridge, MA 02139, USA
{aggelos,lip}@media.mit.edu

Abstract

In this work we investigate the coordinated transmission and processing of distributed radios employing OFDM signals similar to those employed in 802.11a. After showing that OFDM can be viewed as a set of parallel Gaussian channels with different frequency gain for each subcarrier, we design receiving schemes that exploit both the direct transmission between transmitter and receiver as well as the assisting relayed signal of a radio that overhears the communication between transmitter and receiver and acts as an analogue repeater. The special structure of 802.11a OFDM signals allow the coordination between transmitter, receiver and intermediate relay to happen on the same channels and the collaboration results in substantial energy savings compared to the traditional single transmitter-single receiver case, providing efficient resource (energy, bandwidth) utilization.

1 Introduction

Decentralized wireless communications has been in the scientific spotlight for several decades. Recent information theoretic papers have shown that cooperative communication between network nodes can increase the overall transport capacity of the network [2]. Therefore, there is a strong interest for practical algorithms and schemes for collaborative wireless communication.

Inspired by [3], [4] where diversity techniques were exploited over two channels, one for the direct transmission between transmitter and receiver and one for the relayed transmission from an intermediate relay node, we propose a *single channel* collaborative communication scheme, exploiting the special properties of OFDM signals, in 802.11a based WLANs.

We call these communications systems viral because they use opportunistic cooperation among the nodes, exploiting any intermediate ones to propagate the messages. The *viral* communication scheme proposed achieves significant energy savings compared with the direct, non-collaborative case. Since the collaboration happens within the same channels as the direct case, we have an improvement that directly leads to more efficient resource utilization (energy, bandwidth) in a network sense, that potentially can lead to better scalability and therefore higher overall network capacity.

In section 2 we present the properties of the OFDM signals, necessary for our viral scheme, in section 3 we provide the details of the transmitters and receivers as well as theoretical performance bounds and finally in section 4 we provide the experimental results. We conclude in section 5 with some comments on how this system can be used.

2 OFDM Properties and Viral Communication

In this section we are going to explain the suitability of Orthogonal Frequency Division Multiplexing (OFDM) signals for collaborative communication in ways we nominate them as ”viral” and present in the rest of this paper.

$$x(t) = \frac{1}{N}\sum_{-k}^{k}(X_k e^{+j2k\pi\Delta f t}) \tag{1}$$

The baseband OFDM signal (in equation 1) is the sum of $N \geq 2k+1$ orthogonal carriers modulating N information symbols. For 802.11a N is 64 even though 48 carriers are used for data, 4 are used for frequency synchronization and the rest are zeroed obtaining spacing in the frequency domain [5]. The symbols X_k are set according to BPSK, QPSK or QAM modulation schemes and generally are complex numbers [5]. The orthogonality between the sub-carriers is ensured by setting the spacing between them equal to the OFDM signal $x(t)$ duration T :

$$T = \frac{1}{\Delta f} \tag{2}$$

Sampling this baseband signal $x(t)$ and acquiring $N = T/\delta t$ samples leads to the familiar IDFT formula (4), which implies that the baseband signal could be constructed by passing the $N\{x_n\}$ samples over fast IFFT modules and then through efficient analogue-to-digital converters (ADCs).

$$x_n = x(n\delta t) = \frac{1}{N}\sum_{-k}^{k}(X_k e^{+j2k\pi n/N}) \quad (3)$$

$$\Rightarrow x_n = IDFT\{X_k\} \quad (4)$$

It is interesting to see that oversampling this ideally constructed OFDM baseband signal with double frequency ($\delta t' = \delta t/2$) and consequently observing $x(t)$ at half of its duration T ($N\delta t' = N\delta t/2 = T/2$), can still lead to the observation of the original N information symbols X_k :

$$x(n\delta t') = x(n\delta t/2) = \frac{1}{N}\sum_{-k}^{k}(X_k e^{+j2k\pi\Delta f\ n\ \delta t/2})$$
$$= \frac{1}{N}\sum_{-k}^{k}(X_k\, e^{+j2k\pi n/2N})$$
$$= \frac{1}{N}\sum_{-k}^{k}(X_k\, e^{+j2k\pi n/2N}\, e^{+j2k\pi n/2N}\, e^{-j2k\pi n/2N})$$
$$= \frac{1}{N}\sum_{-k}^{k}\{(X_k\, e^{-jk\pi n/N})\, e^{+j2k\pi n/N}\}$$
$$\Rightarrow x'_n = x(n\delta t') = IDFT\{X_k\, e^{-jk\pi n/N}\}$$

From the last relationship we can see that observing the baseband signal $x(t)$ at half duration, *by means of oversampling*, results in a phase shift of the original information symbols by a known factor for every symbol X_k. Therefore, provided that we have an ideally constructed OFDM symbol $x(t)$ the following first property occurs:

- **Property 1**: Observation of an OFDM signal at half its duration could provide for the estimation of the information symbols X_k.

Another important characteristic of the 802.11a signal structure is the cyclic prefix added to each signal $x(t)$. That prefix is derived from the last 1/4 part of $x(t)$ and its duration of 0.8 μs ($x(t)$ duration T=3.2 μs) is larger than the maximum delay spread of the wireless channel (for the transmission power levels and reception ranges specified in the protocol [5]) ensuring zero inter-symbol interference (ISI) regardless the location between transmitter and receiver. That means that node 1 and 2 simultaneous transmissions in figure 1 could coincide at node 3 since the difference in the propagation times is absorbed by the cyclic prefix of the transmitted signals. Therefore the second important property occurs:

- **Property 2**: Simultaneous transmissions in 802.11a networks could lead to simultaneous receptions with zero ISI regardless the topology of the network.

The above properties enable the *Viral* wireless communication scheme presented in the following section. Before proceeding, it is useful to see that OFDM can be viewed as a set of N parallel channels (one for each sub-carrier), with different channel gains for each sub-carrier:

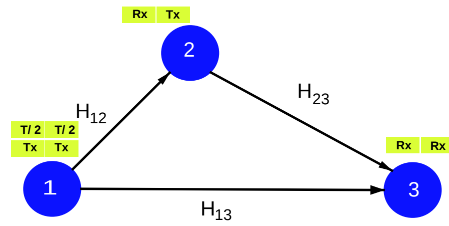


Figure 1: The simple topology examined in this work. Node 1 transmits, Node 3 receives and Node 2 receives during the first half of the signal duration and re-transmits during the second half of the signal duration.

Theorem 2.1 *The convolution of a baseband OFDM signal $x(t)$ with the wireless channel impulse response $\sum_{m=0}^{L-1}(h_m\, \delta(t-\tau_m))$ results in a new baseband signal $r(t) = x(t) * h(t)$ with information symbols $\{H(k)\ \ X_k\}$, where $\{H(k)\} = DFT\{h_m\} = DFT\{h_0, h_1, \cdots, h_{L-1}, \underbrace{0, 0, \cdots, 0}_{N-L}\}$.*

The proof is provided at the Appendix.

The analysis is simplified since the convolution with the channel multiplies each information symbol X_k with a complex number $H(k)$ that represents the channel frequency response at the corresponding sub-carrier.

3 Viral Communication

3.1 System Model

In the previous section it was shown that observation of the OFDM signal at half of its duration (through oversampling) results in a known phase shift of the information signals $\{X_k\}$. It was also shown that the wireless channel filters the OFDM signal by a multiplication of the information symbols $\{X_k\}$ with a frequency response per sub-carrier $H(k)$.

Having in mind the above, we devised a simple scheme for collaborative communication, part of a greater theme of collaborative radio communication: node 1 transmits an OFDM signal $s(t)$ (figure 1) to node 3. The intermediate node 2 "observes" the signal for the first half $T/2$ of its duration (through oversampling), amplifies it (amplifying also its own noise) and retransmits during the second half of the signal duration $s(t)$. Node 3 also over-samples and receives during the first half the direct transmission of $s(t)$ while during the second half, node 3 receives the direct (from node 1) as well as the relayed transmission (from node 2).

For each information symbol X_k (out of the N in the baseband signal $x(t)$) it can be analytically derived (the derivation has been omitted due to space restrictions) that node 3 during the first half of signal duration receives v_{31}:

$$v_{31} = H_{13}X_k + z_{31} \quad (5)$$

and during the second half of signal duration node 3 receives v_{32}:

$$v_{32} = H_{13}X_k + \beta H_{23}\hat{X}_k + z_{32} \quad (6)$$

where the phase shift of $\{X_k\}$ due to oversampling has been incorporated in the channel coefficients H_{ij} which are modeled as iid, complex zero-mean Gaussian random variables with variance $\sigma_{ij}^2/2$ per dimension. Each variance σ_{ij}^2 could be further modeled as $\sigma_{ij}^2 = \frac{G}{d_{ij}^v}$ where G represents the antenna gains of the nodes i and j, d_{ij}^v is the distance between i and j and v is the propagation coefficient of the medium, with $2 \leq v \leq 4$ [7]. The variables z_2, z_{31} and z_{32} represent internal noise of the receiver 2, 3 (first half) and 3 (second half) and can be modeled as iid, complex zero-mean Gaussian random variables with variance $N_0/2$ per dimension. Under those assumptions the magnitude $|H_{ij}|$ of the channel coefficients is distributed according to a Rayleigh distribution and the square magnitude according to an exponential distribution with $E[\,|H_{ij}|^2\,] = \sigma_{ij}^2$.

We assume quasi-static fading model where the channel coefficients H_{ij} remain constant through several consecutive transmissions of the OFDM signals, an assumption which is valid since the symbol duration is on the order of μsecs while the channel tends to change within msecs [7]. It is also important to note that the channel coefficients are computed using special known training sequences. The above model is a baseband model, therefore we have omitted the challenging tasks of frequency and timing offset estimation, in a network sense [1], between nodes 1, 2 and 3. Future protocols for collaborative radio communication should incorporate network channel estimation and network time and frequency synchronization (as opposed to point-to-point estimation techniques currently practiced).

Equations 5, 6 refer to the first and second half of the signal $s(t)$ transmitted from node 1 (figure 1), therefore the collaboration between nodes 2 and 3 is happening within the same signal duration (*the same channel*), as opposed to the non-OFDM schemes proposed in [3]. That is possible only because of the properties of OFDM explained in section 2.

In equation 6, $\hat{X}_k$ represents the signal that node 2 is relaying (figure 1). As described in [3], there are two options: either decode using standard estimation techniques (like a ML receiver) and regenerate or just amplify and forward (act as an analogue repeater). The second technique was more efficient as we experimentally verified and that was the technique followed in this work:

$$\hat{X}_k = \beta(H_{12}X_k + z_2) \quad (7)$$

$$\text{eq.} \quad (5),(6),(7) \Rightarrow$$

$$v_{31} = H_{13}X_k + z_{31} \quad (8)$$

$$v_{32} = (H_{13} + \beta H_{12}H_{23})X_k + \beta H_{23}\, z_2 + z_{32} \quad (9)$$

As we have said, the information symbol X_k is generally a complex number which is set according to the modulation scheme practiced (BPSK, QPSK, QAM). We will compare the above collaborative scheme with the non-collaborative point-to-point case, using BPSK. We will denote as E the energy per bit used in the point-to-point case, $E1$ the energy per bit used at the transmission of node 1 and $E2$ at node 2 in the viral (collaborative) scheme.

For BPSK, $X_k = \pm\sqrt{E}$ in the non-collaborative case and more particularly:

$$'0' : v = +H_{13}\sqrt{E} + z_3 \quad (10)$$

$$'1' : v = -H_{13}\sqrt{E} + z_3 \quad (11)$$

and the ML receiver for equi-probable bits is the following:

$$Re\{v\, H_{13}^*\sqrt{E}\} \lessgtr_{'0'}^{'1'} 0 \quad (12)$$

The probability of error for this receiver, given knowledge of H_{13} ($\gamma_{13} = |H_{13}|^2 E/N_0$) is

$$P_{e/H_{13}} = Q(\sqrt{\frac{|H_{13}|^2 E}{N_0/2}}) = Q(\sqrt{2\gamma_{13}}) \quad (13)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{+\infty} e^{-x^2/2}$. Integrating the above relationship over the exponential distribution of γ_{13} we obtain the following probability of error [6]:

$$P_e = \frac{1}{2}(1 - \sqrt{\frac{\bar{\gamma_{13}}}{1+\bar{\gamma_{13}}}}) \simeq \frac{1}{4\,\bar{\gamma_{13}}} \qquad \bar{\gamma_{13}} \gg 1 \quad (14)$$

where $\bar{\gamma_{13}} = \sigma_{13}^2 E/N_0$ and σ_{13}^2 as defined above.

The last equation (equation (14)) will be used for the comparison of the direct, non-collaborative scheme with the collaborative (viral) one. Before proceeding to the performance results, we need to layout specific details about the viral transmitter and receiver.

3.2 Viral Transmission

In section 2 we showed that oversampling of an ideally constructed OFDM signal $s(t)$ results in a known phase shift $(e^{+jk\pi n/N})$ of the information symbols $\{X_k\}$. Since the intermediate node 2 (figure 1) over-samples during the first half signal duration and destination node 3 over-samples during the first and second half of the signal duration, the signal to be transmitted should be an ideally constructed OFDM signal exhibiting the above property after oversampling.

Therefore, the transmitted signal during the first half should be constructed using the N modified symbols $\{X_k e^{+jk\pi n/N}\}$ through an IDFT procedure. The same process should be followed during the second half of the signal duration at node 1 as well as during the transmission of the intermediate node 2. With that simple modification (since that involves a simple phase shift in software) "ideal" OFDM signals for viral communication are constructed.

3.3 Viral Reception

From equation (7) it is easy to see the relationship between the amplification β of the relayed symbol X_k with the transmission energies E1, E2 at node 1 and 2 respectively:

$$\text{eq. (7)} \Rightarrow \beta^2 |H_{12}|^2 \, E1 + \beta^2 N_0 = E2 \qquad \Rightarrow$$
$$\beta^2 = \frac{E2}{|H_{12}|^2 \, E1 + N_0} \qquad (15)$$

For the simple BPSK system presented below z_1, z_2 are iid complex, zero-mean Gaussian random variables with variance $\frac{N1}{2}$, $\frac{N2}{2}$ per dimension respectively and H_1, H_2 are known, iid complex coefficients, according to a Gaussian distribution:

$$\begin{aligned} y_1 &= H_1 X + z_1 && (16) \\ y_2 &= H_2 X + z_2 && (17) \\ '0' : X &= +\sqrt{E} \\ '1' : X &= -\sqrt{E} \end{aligned}$$

the (optimum) ML receiver is the following *Maximum Ratio Combining* receiver, which is simply a straightforward consequence of detection of vector observations:

$$\frac{4\sqrt{E}}{N_1} Re\{y_1 \, H_1^*\} + \frac{4\sqrt{E}}{N_2} Re\{y_2 \, H_2^*\} \lessgtr_{'0'}^{'1'} 0 \qquad (18)$$

The probability of error for this ML receiver, given knowledge of H_1, H_2 is:

$$P_{e/H_1,H_2} = Q(\sqrt{\frac{2|H_1|^2 E}{N_1} + \frac{2|H_2|^2 E}{N_2}}) \qquad (19)$$

Integrating the above probability of error over the exponential distribution of $|H_1|^2$, $|H_2|^2$ and using the inequality $Q(x) \leq \frac{1}{2} e^{-x^2/2}$ we have the following upper bound on the unconditional probability of error for the ML receiver above:

$$P_e \leq \frac{1}{2} \frac{1}{\frac{|H_1|^2 \, E}{N_1} + 1} \frac{1}{\frac{|H_2|^2 \, E}{N_2} + 1} \qquad (20)$$

We apply the above results in our system model described in equations 8, 9 having in mind that for BPSK $X_k = +\sqrt{E1}$ for '0' and $X_k = -\sqrt{E1}$ for '1'. The Maximum Ratio Combining receiver is the following:

$$Re\{\frac{\sqrt{E1} \, v_{31} \, H_{13}^*}{N_0}\} +$$
$$+ Re\{\frac{\sqrt{E1} \, v_{32} \, (H_{13}^* + b \, H_{12}^* H_{23}^*)}{b^2 |H_{23}|^2 N_0 + N_0}\} \lessgtr_{'0'}^{'1'} 0 \qquad (21)$$

Based on equation 20 and the system model equations 8, 9, we provide the following upper bound on the unconditional error probability of our receiver, which was experimentally proved a tight bound for locations of the intermediate node 2 alongside the line between transmitter 1 and receiver 3:

$$P_e \leq \frac{1}{2} \frac{1}{(\overline{\gamma_{13}} + 1) \frac{\overline{\gamma_{13}} + \frac{\overline{\gamma_{12}} \; \overline{\gamma_{23}}}{\overline{\gamma_{12}} + N_0}}{N_0 + \frac{\overline{\gamma_{23}} \, N_0}{N_0 + \overline{\gamma_{12}}}}} \qquad (22)$$

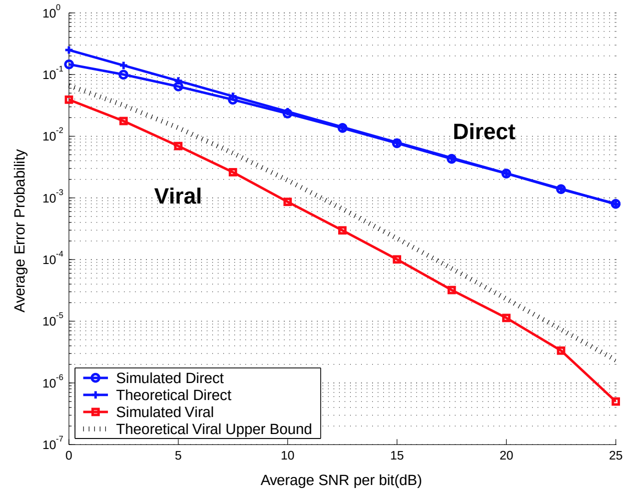


Figure 2: Error Probability for E1=E2=E/2, $d_{12} = d_{23} = 0.5$, $d_{13} = 1$ and v=4.

where $\overline{\gamma_{ij}} = \sigma_{ij}^2 Ei/N_0$ and $\sigma_{ij}^2 = \overline{|H_{ij}|^2} \simeq \frac{G}{d_{ij}^v}$ as defined above.

We are ready now to proceed to the experimental results.

4 Performance

After normalizing the distance between nodes 1 and 3 to $d_{13} = 1$ we evaluated our *single channel* viral receiver (equation 21) according to the system model (equations 8, 9) and compared with the direct point-to-point transmission (equations 10, 11). Initially, we kept the total transmission energy E1+E2 of the viral scheme equal to the transmission energy E of the direct, single hop case and we plotted the error probability after the simulations as well as the theoretical upper bounds calculated from the equations 14, 22 for the direct and viral case respectively, versus the average single hop SNR per bit $\frac{E}{N_0}\sigma_{13}^2$ where $\sigma_{ij}^2 \simeq \frac{G}{d_{ij}^v}$.

In figure 2 where the intermediate node is half way between transmitter and receiver, the probability of error is one to three orders of magnitude smaller than that of the direct case, for the same energy used. Similar remarkable behavior is observed in figures 3, 4 where the intermediate node is either closer to the transmitter or closer to the receiver. Similar behavior was observed in the experiments in [3] where two channels were used, one for the direct transmission and one for the relay.

It is interesting to quantitatively find out the energy gains due to collaboration using the above single channel viral scheme. Using equations 14 and 22 which describe the error probability for the direct and the viral scheme respectively, we performed the minimization of energies E and (E1, E2) given a specific target error probability and plotted the computed ratio $\frac{E}{E1+E2}$. From figures 5, 6 we can also see an approximately two orders of magnitude energy saving for an

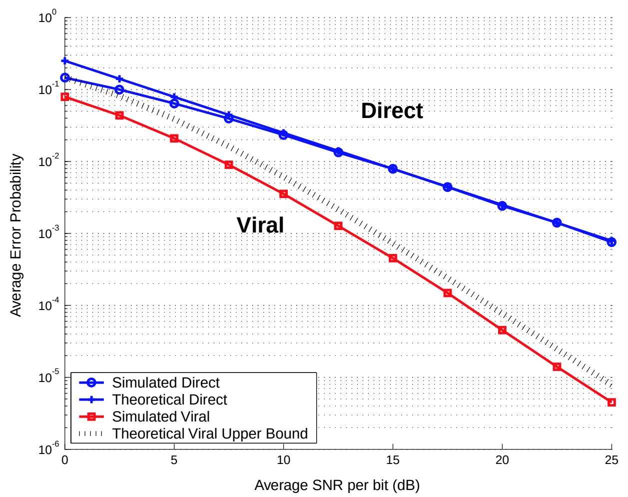


Figure 3: Error Probability for E1=E2=E/2, $d_{12} = 0.1$, $d_{23} = 0.9$, $d_{13} = 1$ and v=4.

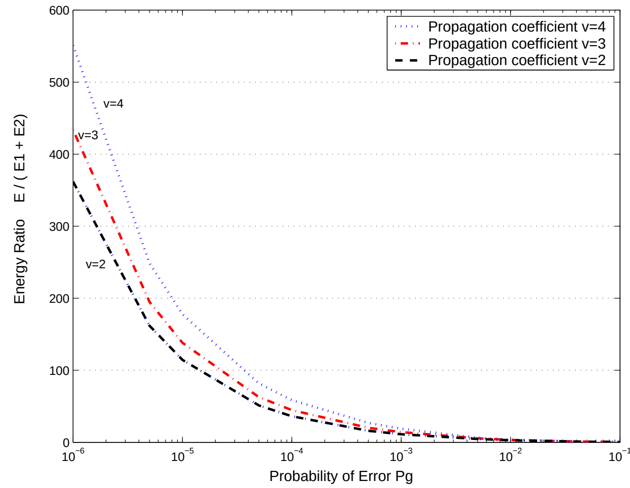


Figure 5: Energy ratio between the energy used in direct transmission E and energy used in viral scheme E1+E2 versus target error probability with $d_{12} = d_{23} = 0.5$ and $d_{13} = 1$.

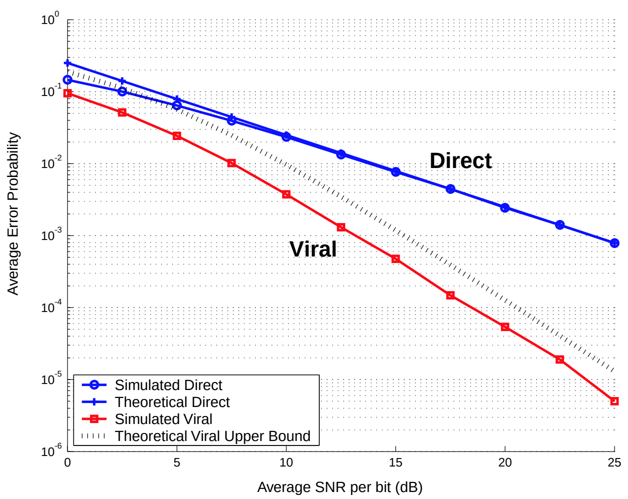


Figure 4: Error Probability for E1=E2=E/2, $d_{12} = 0.9$, $d_{23} = 0.1$, $d_{13} = 1$ and v=4.

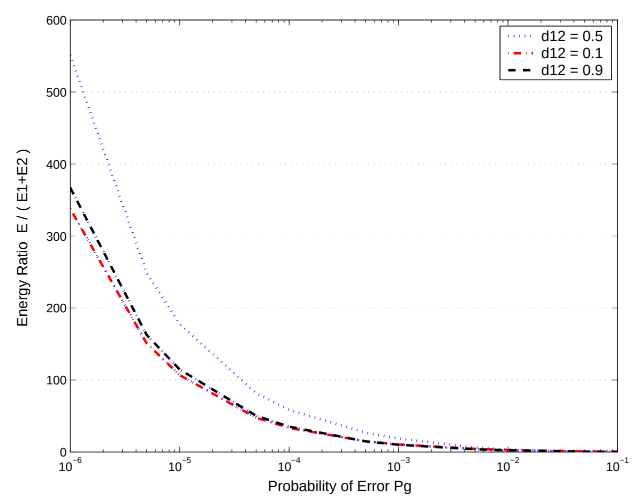


Figure 6: Energy ratio between the energy used in direct transmission E and energy used in viral scheme E1+E2 versus target error probability with $d_{13} = 1$, v=4.

intermediate relay across the connecting line between transmitter and receiver, for all tested values of the propagation coefficient v. Savings with this viral scheme were also observed for a relay in the vicinity of transmitter and receiver, however precise topology bounds were not possible to derive due to the complexity of equation 22 (and its inaccuracy at regions away from the area between transmitter and receiver). We will address the problem of suitable topologies for viral communication in future work.

5 Conclusion

In the above sections we demonstrated a "viral" scheme for single channel, collaborative, wireless communication. The collaboration resulted in energy savings or in higher reliability (smaller error probability) for the same energy used, compared with the direct, non-collaborative case. Since the collaboration happens within the same channels as the direct case, we have an improvement that directly leads to more efficient resource utilization (energy, bandwidth) in a network sense, that potentially can lead to better scalability and therefore higher overall network capacity.

In addition to the ability of such a system to scale through the re-use of the channel in the same region, this system propagates the message through relay nodes with a delay that is small compared to the information rate. Thus, it can be used for realtime applications such as telephony. The essence of the plan is that the information is passed from node to node without decoding and remodulating the messages.

Nevertheless, there are important technical challenges to be explicitly addressed like network frequency/phase/time synchronization and distributed channel estimation as well as important sociological issues like privacy/security and economics to be discussed. We have just scratched the surface and exciting research horizons are now opened.

References

[1] A. Bletsas. *"Evaluation of Kalman Filtering for Network Time Keeping"*. IEEE Pervasive Computing and Communication Conference (Percom 2003), March 2003.

[2] P. Gupta and P. R. Kumar. *"The Capacity of Wireless Networks"*. IEEE Transactions on Information Theory, March 2000.

[3] J. N. Laneman and G. W. Wornell. *"Energy-Efficient Antenna-Sharing and Relaying for Wireless Networks"*. Proc. IEEE Wireless Communications and Networking Conference (WCNC-2000), September 2000.

[4] J. N. Laneman and G. W. Wornell. *"Distributed Spatial Diversity Techniques for Improving Mobile Ad-Hoc Network Performance"*. Proc. of ARL FedLab Symposium on Advanced Telecommunications and Information Distribution (ATIRP-2001), March 2001.

[5] IEEE Standard for Information Technology. *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification (802.11a)*. 1999.

[6] J. G. Proakis. *Digital Communications*. McGraw-Hill, 4th edition, 2001.

[7] T. S. Rappaport. *Wireless Communications: Principles and Practice*. Prentice-Hall, 2th edition, 2002.

APPENDIX

Theorem: *The convolution of a baseband OFDM signal $x(t)$ with the wireless channel $\sum_{m=0}^{L-1}(h_m\,\delta(t-\tau_m))$ results in a new baseband signal $r(t) = x(t) * h(t)$ with information symbols $\{H(k)\ \ X_k\}$, where $\{H(k)\} = DFT\{h_m\} = DFT\{h_0, h_1, \cdots, h_{L-1}, \underbrace{0, 0, \cdots, 0}_{N-L}\}$.*

Proof :

$r(t) = x(t) * h(t) =$

$= \frac{1}{N}\sum_{k=-K}^{K}\sum_{m=0}^{L-1} h_m X_k e^{+j2\pi k(t-t_m)/T}$

$= \frac{1}{N}\sum_k X_k e^{+j2\pi kt/T}\sum_m h_m e^{+j2\pi k t_m/T}$

$\simeq \frac{1}{N}\sum_k X_k e^{+j2\pi kt/(N\delta t)}\sum_m h_m e^{+j2\pi km\delta t/(N\delta t)}$

$= \frac{1}{N}\sum_k X_k e^{+j2\pi kt/(N\delta t)}\underbrace{\sum_m h_m e^{+j2\pi km/N}}_{DFT\{h_m\}=H(k)}$

$= \frac{1}{N}\sum_k H(k) X_k e^{+j2\pi kt/(N\delta t)} \Rightarrow$

$r(n\delta t) = \frac{1}{N}\sum_k H(k) X_k e^{+j2\pi kn/N} =$

$= IDFT\{H(k)X_k\}$ **Q.E.D.** $\Box$

Frequency Selective IQ Phase and IQ Amplitude Imbalance Adjustments for OFDM Direct Conversion Transmitters

Edmund Coersmeier, Ernst Zielinski
Nokia, Meesmannstrasse 103, 44807 Bochum, Germany

edmund.coersmeier@nokia.com, ernst.zielinski@nokia.com

Abstract *– Low power and low cost analog front-end architectures are required to provide competitive applications based on e.g. the IEEE802.11a Wireless LAN standard. Direct conversion analog front-end architectures can fulfill these demands but may introduce unwanted imperfections like IQ amplitude and IQ phase errors. This paper presents a fully digital solution to eliminate frequency selective IQ phase and IQ amplitude imbalance errors caused by the analog modulator in conjunction with low cost analog base-band filters. This paper covers additionally the relevant hardware and software partitioning of the related error detection and error correction blocks.*

Index Terms *– OFDM, direct conversion, frequency selective IQ phase imbalance, frequency selective IQ amplitude imbalance*

I. Introduction

Fully digital compensation techniques for IQ phase and IQ amplitude imbalance errors provide advantages, because they offer a cheap solution with regards to the overall radio architecture. The IQ imbalance errors can appear in the case of a direct conversion analog front-end architecture. These architectures are recommendable if the radio application requires a low cost solution. This is very often true for portable or wireless devices such as IEEE802.11a Wireless LAN applications. To provide anyhow the required high signal accuracy one has to guarantee that the analog direct conversion front-end IQ imperfections will be eliminated satisfactory.

The use of cheap analog front-end components might introduce analog filter amplitude ripple and group delay, which make the IQ imbalance errors frequency selective. The frequency selectivity can be removed successfully by employing frequency selective IQ phase and frequency selective IQ amplitude adjustment algorithms. The here presented time domain adjustment techniques employ a decision directed IQ amplitude and filter pre-equalizer. The non-decision aided IQ phase adjustment provides a pre-equalizer-like architecture and processes a certain amount of time domain coefficients, too.

Both imperfections, the frequency selective IQ amplitude and IQ phase imbalance errors, can be assumed to vary very slowly over a certain period of time. Hence the IQ error detections are required during the tracking phase relatively seldom. In contrast to that the IQ error corrections are needed all the time to correct all outgoing samples continuously. Hence it is possible to implement the IQ error detection algorithms via software on a Digital-Signal-Processor (DSP) and the IQ error corrections via hardware in an ASIC or FPGA.

Based on an example for a direct conversion OFDM transmitter this paper will present the mathematical equations for an IQ phase and IQ amplitude imbalance error generation and the corresponding frequency selective IQ phase and IQ amplitude error adjustments. For both algorithms there will be given an implementation overview and simulation results provide visual comparability.

II. Direct Conversion Front-End

A possible direct conversion analog front end for an IEEE802.11a OFDM transmitter is given by figure 1). Starting from the left upper corner in figure 1) the digital IQ symbols are generated and the IFFT block converts them from the frequency domain to the time domain. The next block takes care about correcting the IQ amplitude and analog filter imperfections via a pre-equalizer. The IQ phase error is pre-compensated by the IQ phase pre-equalizer. Both blocks operate on the ideal incoming time domain symbols and pre-modify these with the corresponding correction values.

After the IQ pre-processing has been finished the digital signal is handed over to the analog domain, where the analog low-pass filters might insert filter amplitude ripple and group delay. The analog base-band I-branch and the analog Q-branch differ physically and hence the analog filters most probably do not equal exactly. Hence the frequency selective I- and Q-imperfections in both branches differ and the correcting pre-equalizers operate as real-number devices in both branches independently. Complex-number pre-equalizers would not be able to provide the required corrections.

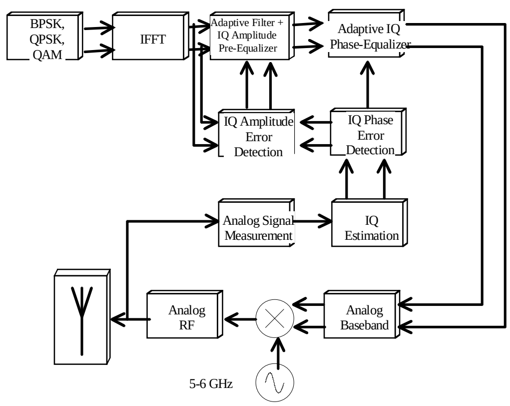


Fig. 1. Example of an IEEE802.11a wireless LAN direct conversion analog front-end.

The analog IQ modulator will add the IQ phase imbalance imperfections. Finally the outgoing signal will be measured at the antenna input port. An analog envelope measurement is fed back to the digital base-band transmitter domain. The envelope can act as the signal source to find out the required information about the IQ imperfections. A digital base-band block provides estimates of the virtual analog I- and Q-values at the antenna input port without a down-modulation process. Because the RF signal is real and not complex, one can only estimate base-band equivalent I- and Q-branch samples. In case of possible estimation errors it can be assumed that the wrong values are negligible during the loops' tracking phase. During the acquisition mode the IQ estimation in conjunction with the IQ feedback loops provide reliable estimates, too. During that time wrong estimates will be low-pass filtered through the feedback loop architecture.

After the analog I- and Q-symbol estimation has been calculated successfully, the frequency selective IQ phase and IQ amplitude error detection take place. The IQ phase error detection is a blind algorithm and requires only the IQ symbol input from the estimation block. The IQ amplitude and filter error detection algorithm needs both, the analog IQ estimates and the corresponding ideal IQ symbols. All information is necessary, because the IQ amplitude and filter pre-equalizer is decision aided. After all errors have been calculated and the correction coefficients for both pre-equalizers have been provided, the feedback loops close with the IQ correction blocks.

The next sections will provide mathematical descriptions, block and implementation diagrams for the IQ phase and IQ amplitude error generation and the error adjustments.

III. IQ Phase Imbalance Error

For the analog mixing process two signals, a sine- and a cosine-signal, have to be provided. Because of technical reasons precise orthogonal functions cannot be guaranteed and hence an additional offset angle $\varphi \neq 0°$ will be measurable between the sine- and cosine-functions. This phenomenon will be called non-frequency selective IQ phase imbalance error.

Additionally the analog base-band components like low-pass filters might add frequency selective imperfections like amplitude ripple or group delay. Hence the non-frequency selective IQ phase imbalance imperfections will result in frequency selective IQ phase imbalance inaccuracies. The non-frequency IQ phase imbalance error is formulated in equation (1).

$$\begin{aligned} s'(t) &= I'(t) + j \cdot Q(t) \\ &= I(t) + Q(t) \cdot \sin(\varphi) + j \cdot Q(t) \end{aligned} \quad (1)$$

One can see, that a certain part of the Q-branch, which is proportional to sin(φ), has been added to the I-branch. Hence the ideal, un-correlated I- and Q-branches will be correlated after an IQ phase error has been inserted. Before the IQ phase error insertion in equation (1) takes place imperfect analog filters might introduce frequency selectivity. Hence at the antenna input port it is not possible anymore to divide between the IQ phase imbalance and the analog filter effects. Hence digital pre-compensation techniques need to employ a pre-correcting signal that provides frequency selectivity. This can be achieved by a pre-equalizer-like compensation architecture, which will be presented in this paper.

The IQ phase error insertion normally introduces at the same time a small IQ amplitude error. This additional IQ amplitude error has been neglected in equation (1).

IV. IQ Amplitude Imbalance Error

The IQ amplitude imbalance error is formulated in equation (2).

$$s'(t) = I'(t) + j \cdot Q'(t) = a \cdot I(t) + j \cdot b \cdot Q(t) \quad (2)$$

Each of the two branches can have an own amplification factor, because analog I-branch and Q-branch components might differ in terms of their amplification. Because the analog filters will introduce amplitude ripple in the pass-band the IQ imbalance errors might become frequency dependent.
A time domain pre-equalizer handles the frequency selective imperfections [4].

The following two sections introduce the frequency selective IQ phase and the frequency selective IQ amplitude adjustment loops.

V. Frequency Selective IQ Phase Imbalance Adjustment

In this section it is assumed that one or both analog base-band filters provide imperfections. These imperfections could be one or more items like amplitude ripple or non-linear filter phase behavior. Because of these additional imperfections a non-frequency selective IQ phase imbalance adjustment loop locks to a wrong error value.

Hence it is necessary to implement an IQ phase imbalance error detector, which is frequency selective and able to cover analog filter imperfections. The $\tilde{I}$ and $\tilde{Q}$ syntax defines I and Q samples which are afflicted by the frequency selectivity of the analog filters.

Equation (3) describes the mathematical operations for the error detection $e_i[n]$. N different IQ phase errors are calculated.

$$e_i[n] = \tilde{I}[n-(N-1)/2] \cdot \tilde{Q}[n-(i-1)] \quad (3)$$
$$i = 1,2,\cdots,N$$

In this paper it is assumed that N is an odd number.

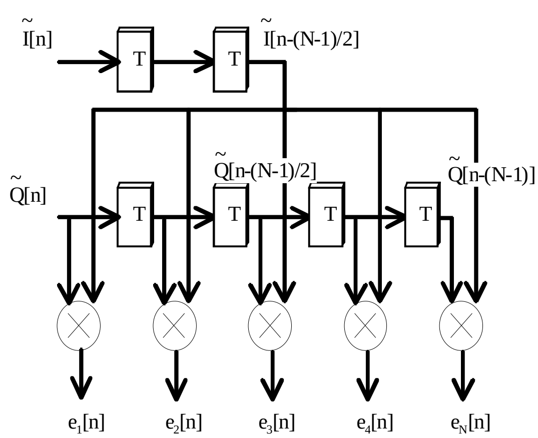


Fig. 2) Frequency selective IQ phase imbalance error detector. N = 5.

Figure 2) presents a possible implementation setup. The center-tap (N-1)/2 of the I-branch will be multiplied with N different values from the Q-branch. Equation (4) shows that each error value $e_i[n]$ will be low-pass filtered by its own integrator.

$$c_i[n] = \mu \cdot \int_{k=0}^{n} e_i[k] \ , \quad i = 1,\cdots,N \quad (4)$$

The constant μ describes the step width of the adaptation loop. The final frequency selective IQ phase imbalance error correction will be done by equation (5) and figure 3).

$$I'[n-(N-1)/2] = \begin{bmatrix} I[n-(N-1)/2] - \\ \sum_{i=1}^{N} c_i[n-m] \cdot Q[n-(i-1)] \end{bmatrix} \quad (5)$$
$$i = 1,2,\cdots,N \ \ and \ \ m > 0$$

The variable m describes the implemented loop latency. Similar to a normal pre-equalizer the Q-branch values from the tap-delay line are multiplied with the corresponding correction coefficients c_i and are summed up. This result is subtracted from the perfect I-branch center tap. Hence the digital I-branch values start to provide a frequency selective IQ phase imbalance error, which will be compensated after the analog IQ phase imbalance error has been inserted. Frequency selectivity is needed to pass the analog base-band filters correctly.

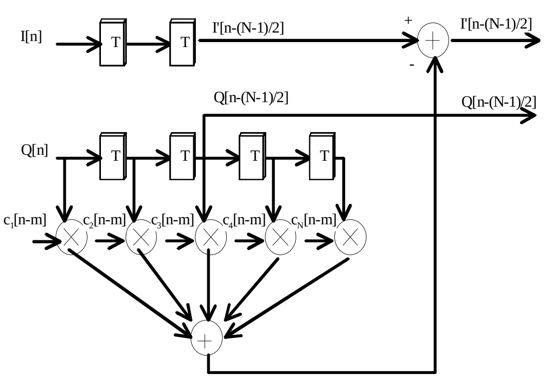


Fig. 3) Frequency selective IQ phase correction. N=5.

VI. Frequency Selective IQ Amplitude Imbalance Adjustment

In this section there will be introduced an LMS based equalizer [4], which does not operate with complex coefficients, but with real ones. This is unusual but makes it possible to handle I-branch and Q-branch imperfections independently. The I-branch and Q-branch filter imperfections are generated by the analog base-band filters, which are two real filters. The IQ amplitude error detection will be done via equation (6).

$$e_I[n] = I[n] - \tilde{I}[n] \qquad (6)$$
$$e_Q[n] = Q[n] - \tilde{Q}[n]$$

From the ideal transmitted symbols there will be subtracted the imperfect estimated symbols. The calculated errors of both branches need not to be the same values. Hence there have to be calculated for both branches independent correction coefficients. This is described by equation (7). The new coefficients at the time n+1 will be calculated from the current coefficients at the time n and an additional addend.

$$\underline{c}_{I,Q}[n+1] = \underline{c}_{I,Q}[n] + \mu\, e_{I,Q}[n]\, \underline{D}_{I,Q}[n]\, \underline{h}^{\#}_{I,Q}[n] \qquad (7)$$

The addend consists out of four factors. First the constant μ describes the step width. The step width defines the loop accuracy, loop adaptation speed or loop bandwidth. Because the expected IQ amplitude imbalance errors will not change over a very long period of time the loop bandwidth needs not to be large and hence the loop accuracy can be high. The second factor is the calculated error from equation (6). After that the product of the ideal input data matrix $\underline{D}$ and an approximation $\underline{h}^{\#}$ [3] of the analog filters $\underline{h}_{I,Q}$ follows. The approximations of both analog filters will be simple tap-delay lines, providing the same latency as the analog filters contain. After the update of the coefficient vectors the correction takes place in programmable FIR filters.

VII. HW and SW Partitioning

To enable a flexible radio design it is useful to implement some functionality via software and not only hardware. The advantage is that software code can be optimised in an existing radio environment. A disadvantage might be that a Digital-Signal-Processor (DSP) has not enough processing power to calculate high data rate functions in time.

Both new adjustment loops require three different steps. The first one takes care about the IQ sample estimation from the envelope. The second provides the error detection of the IQ phase and IQ amplitude imperfections. The last step handles the error corrections. Assuming that the IQ imperfections are stable over a long period of time and they will not change their values rapidly it is possible to employ for the IQ sample estimation and the IQ error detection algorithms a DSP-based software approach. This is possible because the low rate of error value changes does require only a low rate of error detection updates and a low rate of the IQ sample estimations. Hence a DSP, which might not handle the same operations on symbol rate in time, will be able to process now a limited number of operations based on a lower sample rate.

Assuming that the DSP processing power is not enough to handle the symbol rate based error corrections it is still necessary to employ hardware for both IQ correction blocks. Figure 4) provides a hardware-software partitioning overview. Three blocks from figure 1) have been replaced by the DSP. In this new architecture the envelope measurement results and ideal IQ samples will be transferred through the data bus to the processor. The instruction memory provides the software-based algorithms and the correction values will be transferred by the data bus to the IQ correction blocks.

The advantage of the software-based approach is the chance to change the IQ sample estimation and error detection after the overall system implementation has been finalized. The success of the IQ compensations is strongly depending on the algorithm development based on the modelling quality of the analog components. Might the analog components change with regards to the used simulation models, it is an advantage to be able to change the IQ sample estimation and IQ error detection, too. Changes of the analog components might appear because of technical or commercial reasons.

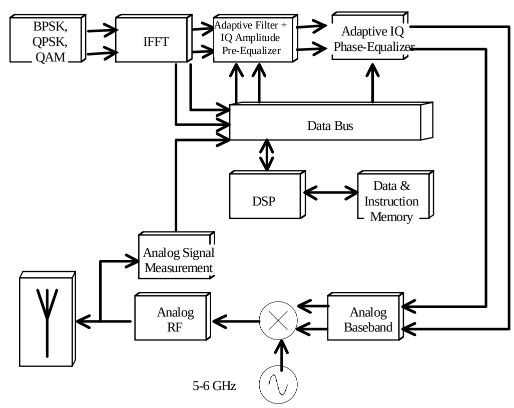


Fig. 4) DSP-based IQ estimation and error detection.

VIII. Simulation Results

Figures 5)-8) show different IQ diagrams in an IEEE802.11a OFDM environment.

Figure 5) provides the imperfect IQ symbols after ideal down-modulation. The imperfections were caused by IQ phase and IQ amplitude errors. Frequency selectivity has been added by imperfect analog base-band filters. Both base-band filters have slightly different transfer-functions.

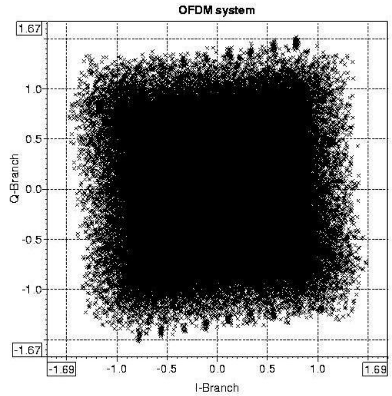


Fig. 5) Imperfect IQ diagram.φ = -10°.a = 0.7,b = 1.0.

Figure 6) provides the first corrected results. A 19-coefficient filter pre-equalizer and a non-frequency selective IQ phase adjustment do not provide full IQ compensation.

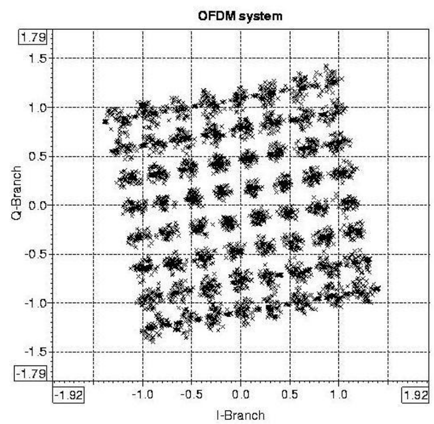


Fig. 6) 1-coefficient IQ phase adj., 19-coefficient IQ amplitude adjustment. φ = -10°. a = 0.7, b = 1.0.

Figure 7) differs from figure 6) only because the IQ phase pre-equalizer coefficient number equals 3.

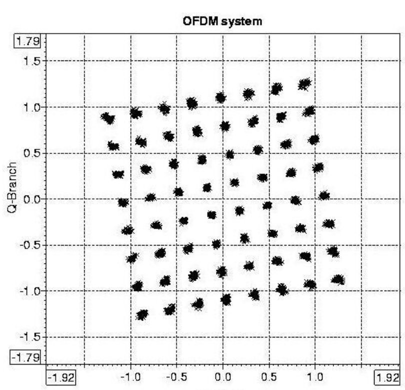


Fig. 7) 3-coefficient IQ phase adj., 19-coefficient IQ amplitude adjustment. φ = -10°. a = 0.7, b = 1.0.

Figure 8) provides the required accuracy. Both adjustment loops employ 19-coefficients for each correction loop. The constellation points provide the required accuracy.

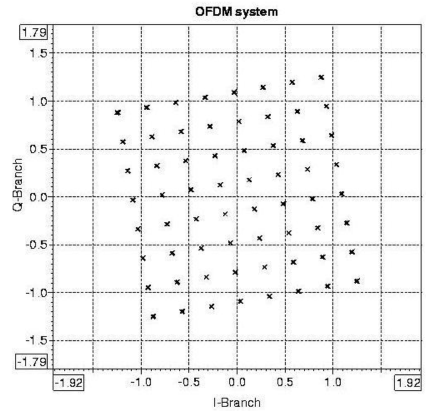


Fig. 8) 19-coefficient IQ phase adj., 19-coefficient IQ amplitude adjustment. φ = -10°. a = 0.7, b = 1.0.

The remaining phase shift of about -10° has been generated by the IQ phase error and will not be compensated.

IX. Conclusion

This paper provides two different time-domain algorithms to handle frequency selective IQ phase and IQ amplitude imbalance errors in an OFDM transmitter. By partitioning the algorithms into hardware and software it is possible to adapt critical parts of the digital algorithm based on new realizations of the analog components. In the case of a software-based approach it is important to consider the loop update rate. The simulation results show that the algorithms will provide significant improvements to the IQ samples in case of frequency selective IQ imperfections.

References

[1] A Novel IQ Imbalance Compensation Scheme for the Reception of OFDM Signals, A. Schuchert, R. Hasholzner, IEEE Trans. on Consumer Elec., Vol. 47 Is. 3 , Aug. 01

[2] Frequency Selective IQ Phase Imbalance Adjustment in OFDM Direct Conversion Receivers, E. Coersmeier, IEEE ISCE 02, Erfurt, Germany

[3] Comparison Between Different Adaptive Pre-Equalization Approaches For WLAN, E. Coersmeier, E. Zielinski, IEEE PIMRC 2002, Lisabon, Portugal

[4] Adaptive Filter Theory, Simon Haykin, Prentice Hall Third Edition, 1996

High-Efficiency, High-Gain Power Amplification for PCS

Dr. Chance M. Glenn, Sr. – *Founder and CTO*
Syncrodyne Systems Corporation
Columbia, MD
410-964-3326 – phone
410-964-0364 – fax
chance.glenn@syncrodyne.com - email

Abstract *–The next generation of mobile telecommunications systems will continue to integrate higher level computing functions, thus increasing the power demand of the device by virtue of increasing the throughput, and increasing on-time. The development of battery technology has not kept up with needs of the industry, therefore OEMs can only maintain their competitive edge in terms of cost, size, functionality, and features, by utilizing the source power more efficiently. In this paper we describe a new concept in power amplification, called syncrodyne amplification, which uses fundamental properties of chaotic oscillators to provide high-efficiency, high gain amplification of standard communication waveforms. We show results of this system providing nearly 60-dB power gain and greater than 70% PAE for communications waveforms conforming to GSM modulation. Finally we show results from a modeled syncrodyne amplifier design operating in the 824 -850 MHz (PCS) band utilizing heterojunction bipolar transistors (HBTs).*

Introduction

In this paper we introduce formally and demonstrate experimentally, a new method of high-gain, high-efficiency power amplification of standard communication waveforms using a chaotic process. We show that this process, called *syncrodyne amplification*, is capable of operating on communication waveforms such as GSM and CDMA and capable of providing gains in excess of 60 dB and power added efficiencies greater than 70%.

In a general sense, the goal of this work is to provide evidence of the notion that the application of chaotic dynamics to engineering technology can provide significant advantages over traditional design. Our goal is to provide a solid example of such an advantage.

We will show experimental results for a 2 MHz prototype, show modeling results for a 150 MHz design, and finally show results for a chaotic oscillation at 850 MHz with the goal of developing a circuit viable at PCS frequencies.

The Future of the Telecommunications Industry

The communications industry has squarely trained its focus on wireless technology. Even though there has been a significant down-turn in the past year or so in this industry, there still remains significant optimism as consumer demand for new features and the integration of current features continues to grow [1]. Table 1 below shows world mobile handset subscribers, past, present, and projected through 2005, having a compound annual growth rate of 16.5%. Regardless of the commercial climate, there is a large, ever-evolving market for wireless communications.

Table 1. Mobile Handset Market: Mobile Subscriber Forecast (World), 1999 – 2005.

Year	Subscribers (Million)	New Additions (Million)	Subscriber Growth Rate (%)
1999	477.5	159.5	-
2000	722.0	244.5	51.2
2001	943.3	221.3	30.7
2002	1,151.5	208.2	22.1
2003	1,363.6	212.1	18.4
2004	1,560.1	196.5	14.4
2005	1,739.0	178.9	11.5
CAGR			16.5%

CAGR = Compound Annual Growth Rate (2001-2005)

One of the greatest challenges for the mobile wireless communications industry has been the provision of mobile power sources capable of meeting the growing demand by users. Improvement and integration of features in mobile handsets increase on-time, and processor requirements, all placing higher demand on the battery. According to Frost & Sullivan, the development of battery technology, specifically its energy storage capacity, has not kept up with the demand by the mobile wireless industry [2].

The solution to this dilemma lies in the efficient utilization of the power provided. In mobile communications the power amplifier uses the bulk of the supply power. Table 2 shows a survey of the leading power amplifier products and the performance of their products. For a power amplifier, power-added efficiency (PAE) is defined by $PAE = \frac{P_o - P_{in}}{P_{dc}}$, where P_o is the output power, P_{in} is the input power, and P_{dc} is the power delivered by the dc source. As shown by the table, efficiencies tend not to exceed 40% in practice. This wasted energy, often in the form of heat, is due to employment of inefficient linear design techniques in order to meet the strict spectral requirements of the industry.

Table 2. Summary of performance parameters for the leading PA products.

	Raytheon	RF Micro Devices	IBM	Sirenza
Part	RMPA 0951A-102	RF2162	2018M009	SPA-2118
Gain	30 dB	29 dB	28 dB	32.5 dB
PAE	30%	35%	34%	38%

Chaotic Dynamics

Since Ott, Grebogi and Yorke's paper in 1990 there has been a tremendous push for the application of chaotic dynamics to technology [3]. Like other endeavors of the 90's many a student, professor and entrepreneur rushed to this area to find what gems lie there. Technology companies have been established, employing researchers in chaotic dynamics in order to find important links to commercial technology. Applications ranging from the control of fluid dynamics, weather prediction and control, spacecraft guidance, sensors and detection, and control of lasers, to various facets of communications have been studied and in some cases put into practice.

By far, the most intriguing and sought after application of chaotic dynamics is in the area of communications. In 1993 Hayes described a formal linkage between chaotic dynamics and information theory [4], showing that the *symbolic dynamics* of a chaotic system could be controlled, thereby paving the way for the direct encoding of digital information into a chaotic oscillation [5]. Figure 1 shows the oscillations of a typical chaotic oscillation encoded to produce a pre-described digital sequence.

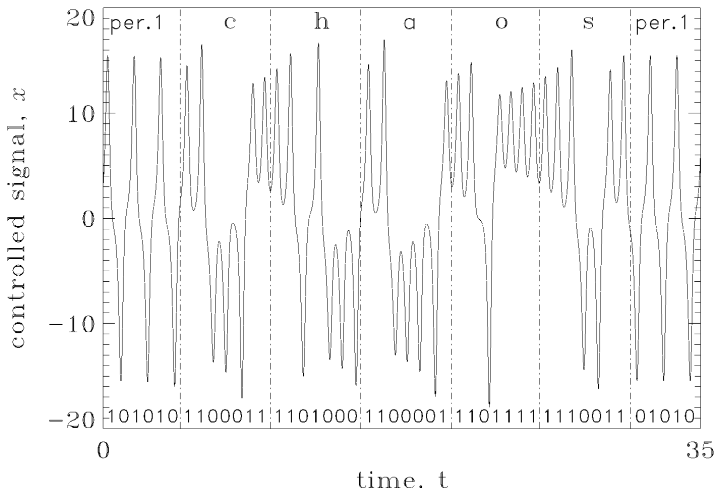


Figure 1 Lorenz oscillations encoded to carry a digital sequence which produces the ASCII text 'c-h-a-o-s'.

It is well known that the operation of electronic circuits and devices in the strongly nonlinear regions yields higher power conversion efficiency [6]. The heart of chaotic dynamics is its operation in the nonlinear regions of the system. Figure 2 shows a typical I-V characteristic curve for a typical transistor. Traditional designers bias their circuit such that the transistor will operate in the linear

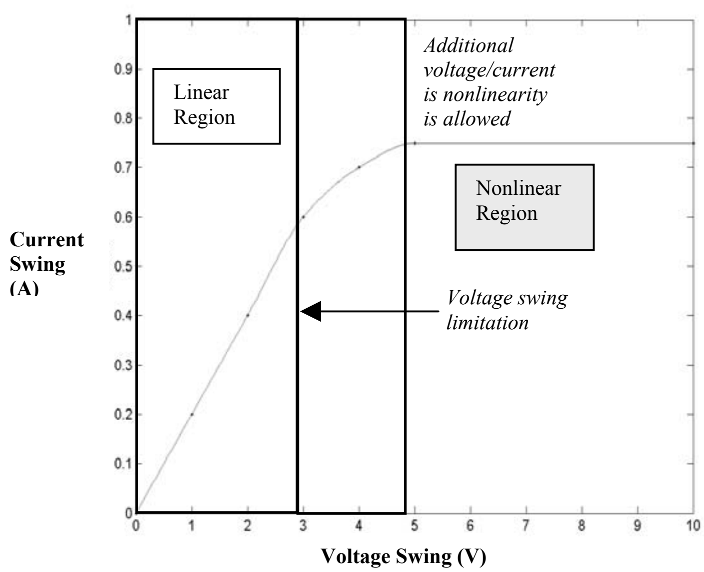


Figure 2 A typical I-V characteristic curve for a transistor describing the linear and nonlinear regions of operation.

region. This type of operation limits the output voltage swing of the circuit. Although the circuit may be capable of producing output voltages outside of the linear region of the transistor, the circuit is purposely designed to avoid this in order to suppress the generation of harmonics. Circuits that are designed to operate in the nonlinear region of the transistor's operation are capable of broader output voltage swings and even higher current, thus capable of higher output power, thus capable of higher power conversion efficiency.

Chaotic oscillations occur as a result of operating the system in its nonlinear state. This is not to say that all nonlinear operation results in chaos, only that the system must be nonlinear in order for chaos to exist.

In the following section we will consider the Colpitts oscillator as the basis of our implementation of syncrodyne amplification. There are three fundamental reasons why we choose this particular oscillator. The first reason is that the Colpitts configuration has been a staple of communications electronics for years. Most analog electronic circuits that require sinusoidal signals employ Colpitts-type circuits. Figure 3 is the Colpitts circuit used for this analysis. Note the feedback is a tank circuit consisting of an inductor and two capacitors. The resonant frequency is given by $\omega = \sqrt{\frac{C + C_e}{LCC_e}}$.

The circuit equations are,

$$L\frac{di_L}{dt} = V_{CC} - v_c - (R + R_L)i_L$$

$$C_e\frac{dv_e}{dt} = i_L - \frac{v_e - V_{EE}}{R_e}$$

$$C\frac{dv_c}{dt} = C\frac{dv_e}{dt} + i_L - i_c$$

where i_c is the forward transistor collector current defined by $i_c = \gamma\left(e^{-\alpha v_e} - 1\right)$, γ and α are empirically derived factors for the transistor and R_L is the series resistance of the inductor.

The second reason for choosing the Colpitts oscillator is apparent from both the circuit mathematical expression and the circuit schematic diagram. The Colpitts circuit is a simple circuit easily modeled, easily realized, and scaleable in frequency. These are critical factors in considering this type of architecture for practical, commercial technology.

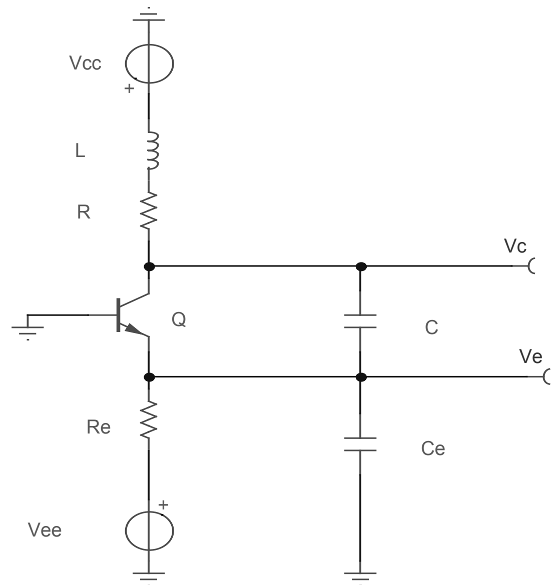


Figure 3. Transistor based Colpitts oscillator circuit used for this analysis.

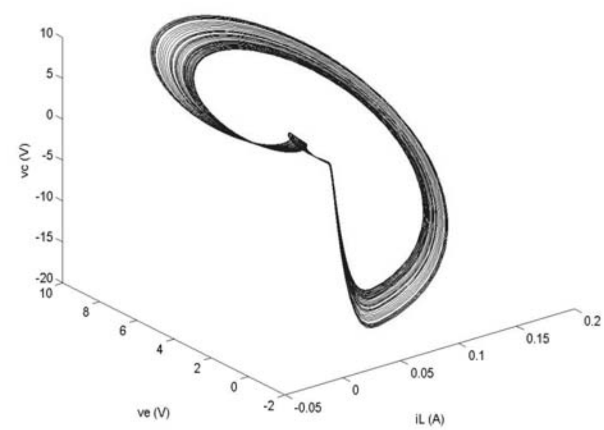


Figure 4 Three-dimensional representation of the state-space trajectories for the Colpitts oscillator illustrating the

The third reason is that in general the chaotic dynamics produced by this oscillator are well understood. There are parameter sets that produce chaotic oscillations of a Rössler type. Once such set of parameters are: [V_{cc} = 5V, V_{ee} = -5V, C = 1.6 nF, C_e = 1.8 nF, L = 6.8 μH, R = 62.5 Ω, R_L = 2Ω, R_e = 260Ω, γ = 1.06 x 10^{-15}, β = 41.2]. These parameters produce chaotic oscillations that have a dominate frequency about 2 MHz. Figure 4 shows a three-dimensional plot of the solutions of the state equations for the circuit utilizing these parameter values. The object formed is what is termed a *state-space attractor*, and is a *strange attractor*, in that it has *fractal* dimension [7].

Figure 5 shows the power spectrum for v_c. Here we see the broad spectral content typical of chaotic oscillations. The calculated resonant frequency for the circuit, using the equation above, was 2.1 MHz.

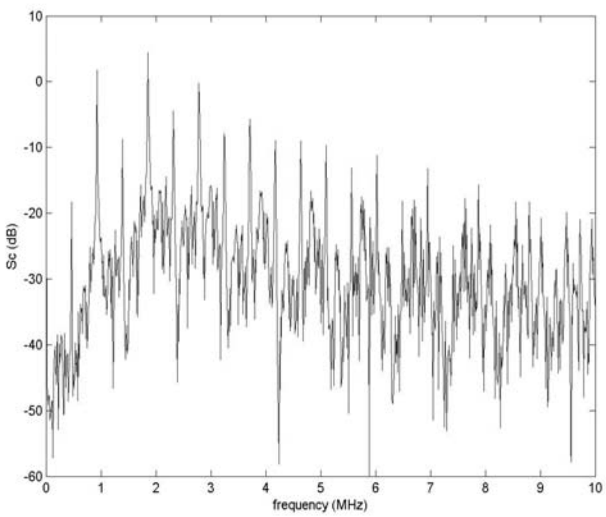


Figure 5 Power spectrum of the collector voltage for the Colpitts oscillator in a chaotic mode of operation.

Syncrodyne Amplification
The syncrodyne amplifier based on the concept of signal amplification using **synch**ronous **dyn**amics. Syncrodyne amplification is the process of locking a chaotic oscillator, larger in power, to a smaller, continuous-time, information-bearing oscillator through a process called *synchronization* [8]. The smaller oscillator is called the guide signal. The continuous-time guide signal has the advantage of only needing to guide one state variable. The guide signal need only supply a small amount of power to the "amplifying oscillator" in order to stabilize its dynamics. Figure 6 shows a simple block diagram of a syncrodyne amplifier. As synchronization occurs, the error signal, $e(t)$ goes to zero.

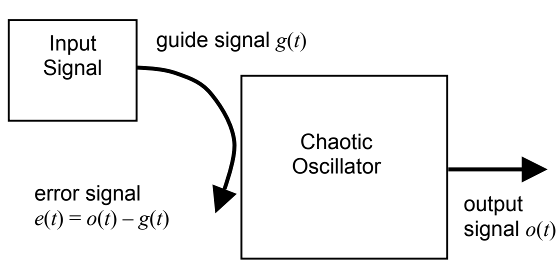


Figure 6 Block diagram of a syncrodyne amplifier system.

This error is directly related to the current flow from the guide system to the output oscillator. As the current goes to zero the power flow from guide to output goes to zero and power amplification occurs.

Since the chaotic system is operating in the nonlinear region of operation it is more efficient. Further, we found that this high-efficiency, high-gain process worked for guide signals that were non-chaotic. The guide signal can be phase modulated sinusoidal oscillations such as phase shift keying (PSK), quadrature phase shift keying (QPSK), minimal shift keying (MSK) and even more sophisticated communications signal formats applicable to the global system for mobile communication (GSM) and code-division multiple access (CDMA). This leads to the application of this technique to standard digital communications technology, offering the possibility of high-gain, high efficiency power amplification.

Development
In the broader development plan for the syncrodyne amplifier we have approached it in the following manner:

(a) Series-2: Low frequency (2 MHz) mathematical model.
(b) Low frequency circuit fabrication and testing.
(c) Series-3A: Mobile, mobile-satellite band (150 MHz) SPICE model and analysis using a SiGe heterojunction bipolar transistor (HBT).
(d) Series-3A circuit design, fabrication and testing.
(e) Series-3B: Land-mobile band (450 MHz) SPICE model and analysis.
(f) Series-3B circuit design, fabrication and testing.
(g) Series-3C: PCS frequency (824-850 MHz) model.
(h) Series3C frequency circuit design, fabrication and testing.
(i) Series-4: Microwave frequency development.

Our goal is to produce devices in the relevant frequency bands identical in form, fit, and function to present power amplifier products. It is important to determine the frequency scaling characteristics so to provide a complete picture of this concept. Viable markets exist for the Series-3A and 3B devices as we prepare to push this technology to industry.

This paper reports on the results of (a), (b), and (c) while showing a basis for (e), (g) and beyond.

Series-2 Results
We fabricated a Colpitts-based syncrodyne amplifier and drove it with a 650 bps, 0.3 GMSK modulated

waveform with a 2 MHz carrier frequency. This was done to mimic the specifications of an 824 MHz – 850 MHz GSM waveform [9]. Figure 7 shows the frequency spectrum of the input waveform.

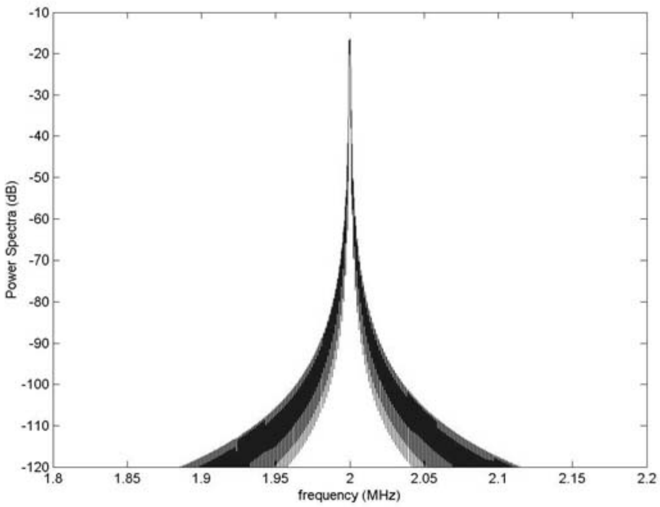


Figure 7. the power spectrum of the 2 MHz GSM input waveform.

We took measurements of the maximum power gain and the power-added efficiency and compared it with the anticipated results from the computer model. Figure 8 shows the comparison of the modeled and measured relationship of the power gain and the PAE.

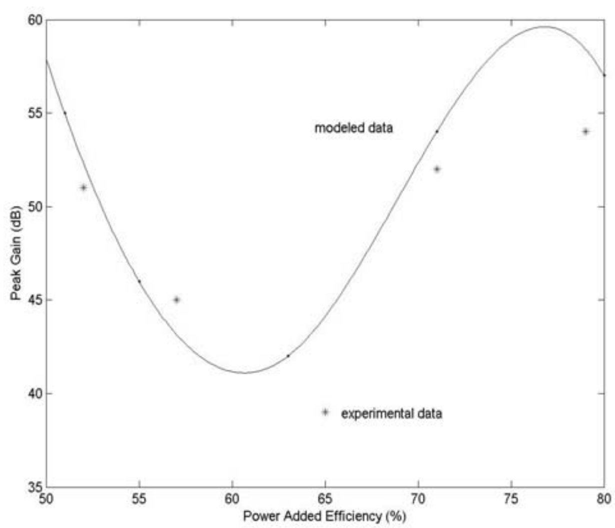


Figure 8. Comparison of experimental data from the 2 MHz syncrodyne amplifier with modeling results.

We see a device capable of delivering enormous power gain and very high efficiency. There was good agreement with the results of the model as compared with the experimental results. We found both experimentally and through the model, that the capacitor C provided an excellent means for tuning the circuit in order to maximize the gain and efficiency when input conditions change. For example, when the peak-to-peak voltage of the input signal changes the circuit must be tuned. Figure 9 shows this response. This provides design guidance in order to achieve maximum performance. Figure 10 shows the output frequency response for a 2 MHz sinusoidal input used to characterize the system. Note that the harmonics occur below -45 dB.

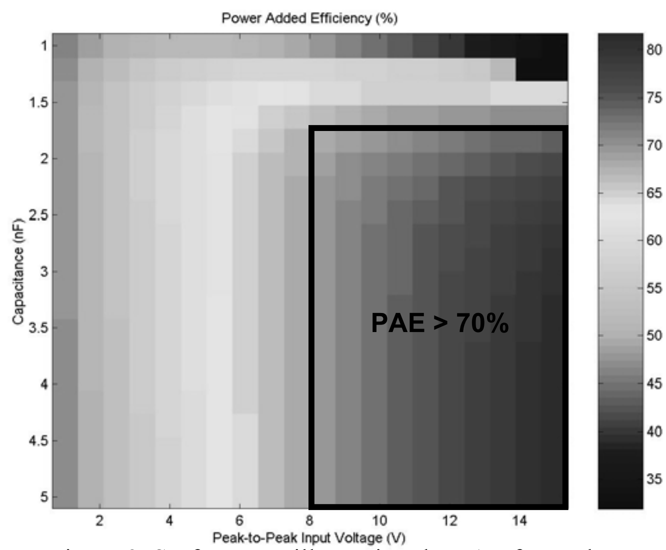


Figure 9. Surface map illustrating the PAE for peak-to-peak input voltages and tuning capacitance values.

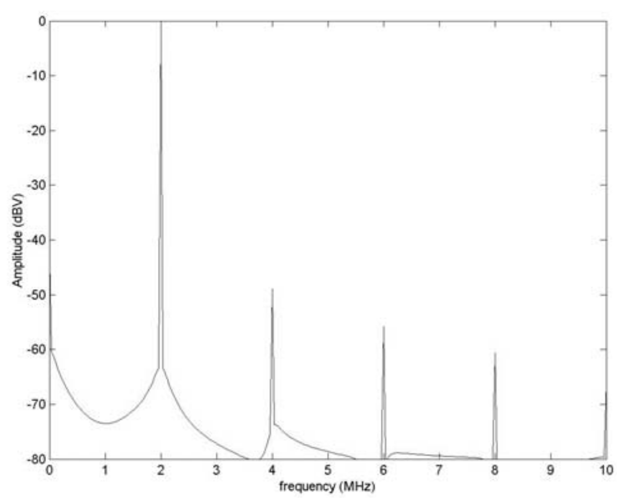


Figure 10. Output frequency response for a 2 MHz sinusoidal input. Note harmonics below -45 dB.

Series-3A SPICE model

Figure 10 shows a general block diagram for a syncrodyne amplifier. An important component of this system is the optimization/stabilization block. In practice, this is a feedback mechanism that senses the error between the input and the output signals and applies perturbations to a circuit element to minimize this error in real-time. The series-2 modeling stage showed that minimization of this error can be achieved by tuning a capacitance, and that this worked to maximize the efficiency and gain.

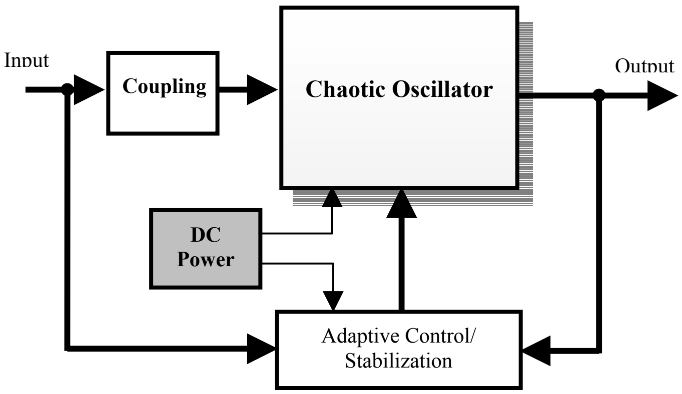


Figure 10. General syncrodyne amplifier block diagram.

The key design goal for syncrodyne amplification is to first produce chaotic oscillations at the frequency band of interest. We use the Colpitts oscillator as a basis for circuit architecture. Figure 11 shows the schematic diagram for the circuit model. We were able to incorporate modeling parameters for a Sirenza sga-8343 HBT while building this model to incorporate specific models for each obtainable circuit element. Some of the devices are tunable so that true operating points could be attained

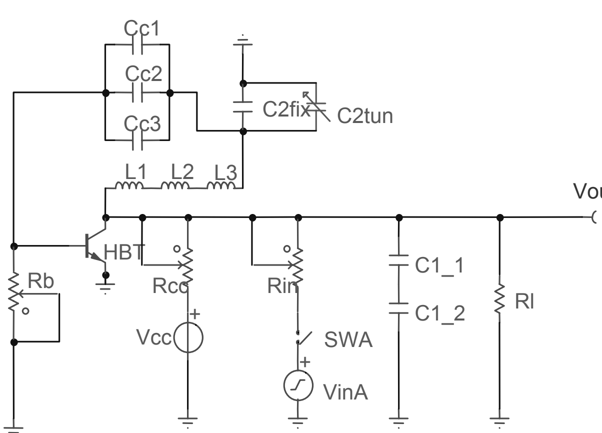


Figure 11. Series-3A syncrodyne amplifier schematic diagram for SPICE model

Figure 12 shows (a) the output frequency spectrum of the oscillator when the switch is in the off position and (b) when the switch is turned on. When switched off, the circuit is a free-running chaotic oscillator with oscillations centered about 154 MHz having about a 25 MHz bandwidth. This broad-banded behavior is typical of chaotic oscillators. When the switch is turned on the oscillator locks to the sinusoidal drive oscillation. This system produced a gain of a over 50 dB and was about 76% efficient. We believe that these numbers can be improved upon by judicious choice of circuit elements and the balancing of the operating point of the circuit for stronger chaotic oscillations.

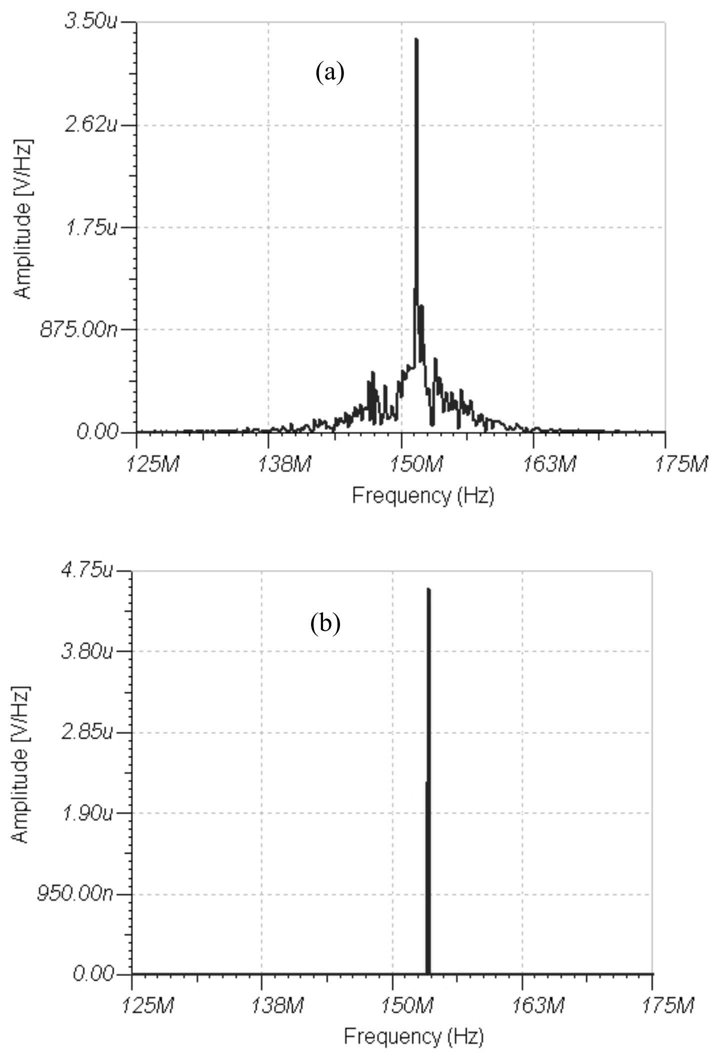


Figure 12. Frequency spectrum of the output signal of the Series-3A syncrodyne amplifier model when (a) the switch is off and (b) on.

The optimization/stabilization portion of this circuit will replace the tuning capacitance, C_{2tun}, and will incorporate a varactor diode balanced for the proper capacitance tuning range.

Series-3B/3C Design Preparation

We have begun this process by identifying circuit architectures and parameter values that lead to chaotic oscillations around 824 MHz. In figure 13 we show a state-space plot of the voltage across one of the circuit capacitances versus the inductor current. This is a typical method of displaying a chaotic oscillation, called a state-space plot [11]. This is only a 2-dimensional projection of the true state-space. The Colpitts-based transistor circuit was modeled with a SPICE simulator using the Sirenza sga-8343 HBT. The oscillation frequency was 849 MHz with a 50 MHz bandwidth.

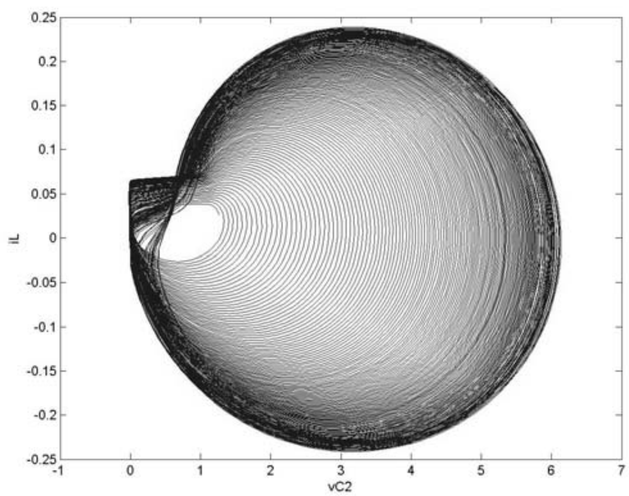


Figure 13. 2-dimensional projection of the state-space of an 850 MHz chaotic oscillation.

Conclusions

Syncrodyne amplification is a new concept in power amplifier implementation. It is not amplification in the traditional sense. It uses fundamental aspects of chaotic dynamics, sensitivity to small changes, synchronization, and the natural efficiency that can result from operating a device in it's nonlinear region, to derive its operation. Essentially, linear amplification is provided through nonlinear means.

Our development path is taking us into the realm of PCS frequencies and beyond. Chaotic dynamics has been observed and reported on in the microwave frequency range [10]. It is obvious the many benefits that can result from more efficient power utilization. Devices become smaller, cheaper, operate longer, and transmit further and faster. These and other benefits are critical to the communications industry in both commercial and military sectors.

We have been able to report on successful results up to 150 MHz and preliminary results towards PCS frequencies.

Chance M. Glenn, Sr. is a co-founder and chief technology officer for Syncrodyne Systems Corporation in Columbia, MD. He received his bachelors degree in electrical engineering from the University of Maryland at College Park, his Master's and Ph.D. degrees in electrical engineering from The Johns Hopkins University.

References

[1] Frost & Sullivan, *2001 World Handset RF-Semiconductor Markets*, January 2002.
[2] Frost & Sullivan, *Mobile Telecommunications Markets – 2nd Quarter 2001*, July 2001.
[3] E. Ott, C. Grebogi, J. A. Yorke, Phys. Rev. Lett. **64**, 1196 (1990).
[4] S. Hayes, C. Grebogi, E. Ott, Phys. Rev. Lett. **70**, 3031 (1993).
[5] S. Hayes, C. Grebogi, E. Ott, A. Mark, Phys. Rev. Lett. **73**, 1781 (1994).
[6] S. M. Sze, ***High-Speed Semiconductor Devices***, AT&T Bell Laboratories, John Wiley & Sons, New York, 1990.
[7] Martin J. Hasler, *Electrical Circuits with Chaotic Behavior*, Proceedings of the IEEE, vol. 75, no. 8, August 1987.
[8] L. M. Pecora and T. L. Carroll, *Synchronization in Chaotic Systems*, *Phys. Rev. Lett*. **64**, 821 (1990).
[9] Theodore S. Rappaport, ***Wireless Communications: Principles & Practice***, Prentice Hall, New Jersey, 1996.
[10] C. M. Glenn, S. Hayes, *Observation of Chaos in a Microwave Limiter Circuit*, IEEE Microwave and Guided Wave Letters, vol. 4, no. 12, December 1994.
[11] Edward Ott, ***Chaos in Dynamical Systems***, Cambridge Univ. Press, Canada, 1993.

System Architecture for a Dynamic-Spectrum Radio

Allen Petrin[(1)], Paul G. Steffes[(2)]

[(1)] Georgia Institute of Technology, 324341 GA Tech Station, Atlanta, GA, USA, 30332-1005
me@allenpetrin.com; 404-894-5280 (phone); 404-894-5935 (fax)

[(2)] Georgia Institute of Technology School of Electrical and Computer Engineering, 777 Atlantic Dr., Atlanta, GA, USA, 30332-0250; ps11@prism.gatech.edu;
404-894-3128 (phone); 404-894-5935 (fax)

Abstract:

The increasing demand for wireless services necessitates a reassessment of how radio spectrum is allocated. This paper will propose the system architecture of a dynamic-spectrum radio. Such a radio system will seek out underutilized spectrum, and for a short period of time operate on these frequencies. With adequate available temporary spectrum, a realizable, moderate-cost, high-data rate radio network can be operated over paths in the 10 km range.

Finding spectrum that is not being used either passively or actively in a propagation environment that is non-homogenous and time varying with a high degree of certainty is the leading requirement for the system's design. An error in this process could cause the dynamic-spectrum radio system to generate disruptive interference to other spectrum users. The preliminary results of a 500 MHz to 6 GHz spectrum search in urban Atlanta, Georgia demonstrates the existence of usable temporary spectrum. This search examines radio spectrum in time, frequency, polarization, and azimuthal direction.

The operation of a wideband (one or more octave) dynamic-spectrum radio is limited by its low-noise amplifier (LNA) and power amplifier (PA). Intermodulation from the LNA reduces the sensitivity of the receiver; concealing the presence of other spectral users and requiring more transmit power to achieve the required carrier-to-noise ratio. Non-linearity from the transmit PA generates intermodulation that pollutes the spectral environment. Radio frequency (RF) filtering, amplifier predistortion, and intelligent control software can mitigate the impact of intermodulation. The resulting architecture will consider these effects, in addition to results from the spectrum studies.

Introduction:

The unrelenting demand for new allocations of radio spectrum for communications services necessitates a more efficient method of spectrum sharing.

In the United States, the Federal Communications Commission (FCC) is responsible for controlling commercial use of spectrum and the National Telecommunications and Information Administration (NTIA) has this responsibility for the government's use of spectrum. Currently the FCC assigns blocks of frequencies for a specific type of use in bands and then licenses a part of these bands to users as channels. By being licensed, one is guaranteed the exclusive right to a channel in a geographic region. The owner also has obligations placed on them by the FCC, both technical and policy-based.

The banding of frequencies for licensed and unlicensed channels allows for standardized media and telecommunications uses. This allows for TV, radio, terrestrial microwave, terrestrial cellular, satellite communications, radio navigation, and wireless networks to have fixed frequency ranges, regardless of locality. The main benefit of this is lower cost equipment and interoperability [1].

Radio spectrum can be shared in time, frequency, polarization, and space. Present allocation methods only consider frequency and geographic area. In free space propagation, electromagnetic waves can be polarized linearly (horizontal and vertical) or circularly (right hand and left hand) [2]. For environments where waves are reflected and scattered there are six possible states of polarization and angle separation that can be reused without interference [3]. Point-to-point microwave links and satellite communications reuse linear or circular polarization to double system capacity. The space variable is broader than just geographic area; it includes both location and the direction of the radiation for both the emitters and receiver.

Studies by the NTIA, FCC and preliminary results of research conducted at the Georgia Institute of Technology show that radio spectrum is not optimally utilized [4, 5]. Several spectrum surveys by the NTIA in different locations around the U.S. have shown that significant unused radio spectrum exists even in major metropolitan areas. The FCC has recently published limited data on the temporal use of spectrum, and this data does demonstrate the time dynamic nature of spectrum usage with significant periods of inactivity. Research being conducted by the authors examines radio spectrum usage in time, polarization, and azimuthal direction for the 500 MHz to 6 GHz frequency range. This study is also designed to find usage differences in urban, suburban and rural areas by an independent survey of each region.

System Architecture:

The paramount goal of a frequency agile smart radio system is to utilize dormant radio spectrum without interfering with incumbent spectral users in any harmful way. This can be accomplished with control algorithms and knowledge of the susceptibilities of protected users. Protected users encompass all spectral users either passive or active that have rights to allocated spectrum. Figure 1 shows a smart radio system deployed in an environment populated with other spectral users. For this example system, three users surround a single base station. Figure 2 shows a high-level block diagram for the transceiver present in each element of the example system.

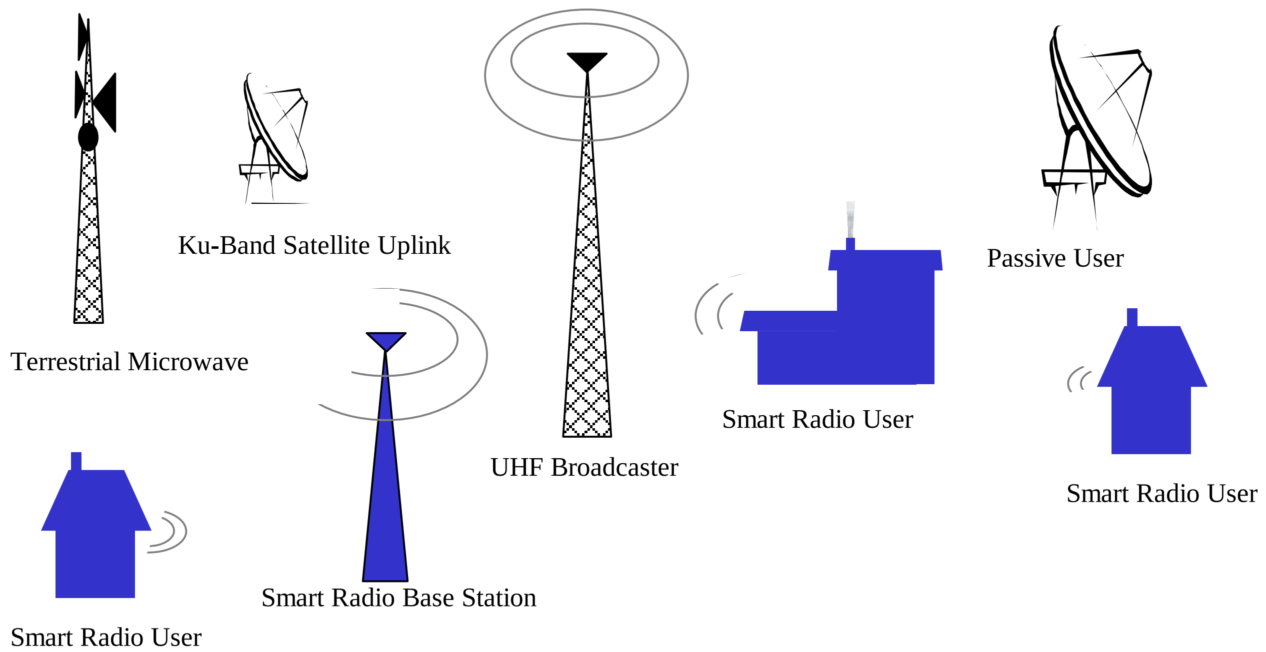


Figure 1 Spectrum Operation Environment

This transceiver includes a global positioning system (GPS) receiver, system controller, and RF subsystem. The antennas attached to the transceiver can be omnidirectional, directive, or a steerable array, depending the desired system cost, flexibility, and spectral efficiency.

The GPS receives provides location information and precise time. The location information is used for the propagation protection model and removes the possibility of device operation in a radio quite zone or near sensitive passive users. The time information is used to synchronize the system and could be used for some transmission methods. Indoor use of this system is not precluded by the use of GPS, given that the demand for E911 capabilities for cell phones has lead to the development of low cost and low power GPS receivers that are capable of operation in an attenuated indoor environment [6].

The system controller in the base station and the user terminals use their collective spectral environment information to supply information to the base station's propagation protection model. The base station is responsible for all dynamic assignment of spectrum by the smart radio system. For a larger implementation requiring several base stations, a higher-level controller would execute the propagation protection model, and subsequently assign dynamic spectrum.
When the user terminal is first activated, it would sweep the range of potential operating frequencies, measuring the power flux density (PFD). The data from this spectrum survey is used to determine frequencies that are unused. Subsequently the user terminal will look for the control channel from a base station.

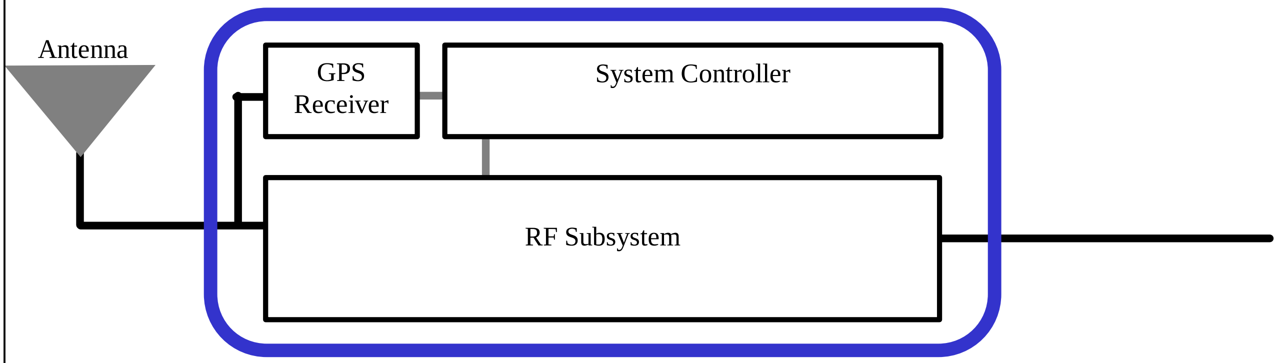


Figure 2 Transceiver Block Diagram

When this channel is found, the user terminal will combine the results of its spectrum search with a list of available channels broadcast by the base station to find a channel that both it and the base station have approved for use. With a communication channel now selected, the user terminal will send its location and spectrum survey data to the base station. The base station then composes its propagation protection model with the data from this user terminal and the other user terminals operating in the area along with the spectrum surveys it has performed, knowledge of the limitations of the detecting equipment, a digital elevation model (DEM) of the area, past spectral knowledge of that area, known protected users in that area, and an up-to-date knowledge of transmitting restrictions so as to assign the user terminal frequencies and transmitting method.

It would be naive to assume that protected users have adequate front-end filtering for dense spectral reuse. Thus even though a channel in a band is unused this does not necessary relate to *usable* spectrum for a smart radio system. Additionally unused spectrum adjacent to band is not always *usable* spectrum. To determine when unused spectrum is available for sharing requires knowledge of attributes of the protected user's receiver. Such information can be used in propagation protection model.

Since spectral occupancy varies with time due to usage patterns and propagation variability, the smart radio system will check several times per second in a "receive-only" mode, to see if the environment has changed, necessitating it to cease using that frequency to avoid creating interference. This would be done immediately to prevent disrupting the protected user.

Figure 3 shows the results of the propagation protection model for one of the user terminals in the system. This terminal is precluded from operating in areas of spectrum that it perceives as unused because of the information gathered by the rest of the system.

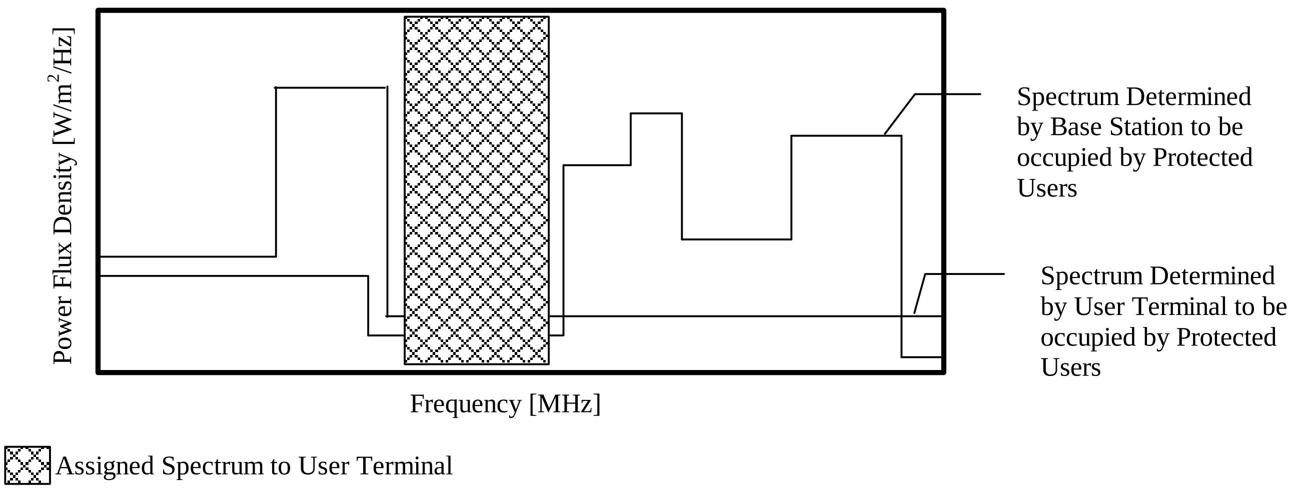


Figure 3 Spectrum Assignment Map for User Terminal, with Usage Profiles

Another advantage of this system architecture is the ability to conduct data mining of spectrum usage patterns; this reduces the probability of interfering with protected users and improves the smart radio's ability to manage its communication links.

RF Subsystem:

The RF subsystem comprises the most costly part of the transceiver, and places the greatest constraints on system performance.

Figure 4 shows the configuration of the RF subsystem. The low-noise amplifier (LNA) in this example is multi-octave to cover an extended bandwidth; such LNA's are commercially available [7]. If there is no filtering before the LNA, intermodulation (IM) from the polluted spectral environment will likely create IM products. IM products occur at frequencies $mf_1 \pm nf_2$ (where m,n = 0, 1, 2, …), with m+n being the order of the IM [8, 9]. In the Atlanta, GA area it was found that a 500 MHz to 8 GHz LNA with a high third order intercept point (IP3) of 27 dBm suffered an increased noise floor from additive IM, in areas of the spectrum adjacent to high usage. This increase in noise floor was measured to be about 10 dB. This rise in noise floor was in addition to discrete IM products. These discrete IM products falsely represent user signals, which reduce the perceived amount of spectrum available for reuse. Hence filtering before the LNA is essential for a smart radio's effective performance. Several switched microelectromechanical systems (MEMS) filter banks and tunable filters have been developed which could provide the necessary front-end filtering requirement [10-14].

The power amplifier (PA) can also generate undesired signals. Unlike the LNA these signals are emitted into the spectral environment. Several techniques for PA linearization have been demonstrated to reduce the filtering requirement needed to suppress these emissions [15-17].

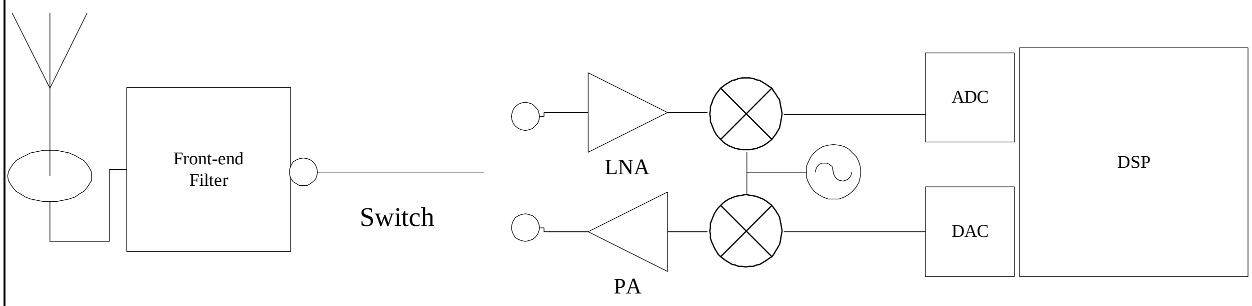


Figure 4 Smart Radio RF Subsystem

Time division duplexing (TDD) has been chosen for the nominal design since it allows reuse of a single front-end filtering system, and allows for insertion of a “dummy” time interval for all the network components to listen for protected user activity.

Conclusion:

This paper has proposed the architecture for a frequency-agile smart radio that honors the spectrum rights of present spectral users. The RF subsystem proposed incorporates several developing technologies to improve system performance.

Frequency-agile smart radios by their nature can adapt to the spectral environment surrounding them. This ability can allow for the establishment of smart radio networks without exhaustive spectrum availability studies and subsequent static reallocation. This could reduce barriers to entry in the wireless marketplace.

The authors are performing a radio spectrum usage study in the 500 MHz to 6 GHz frequency range. This study improves on ones previously conducted by measuring spectrum in time, polarization, and azimuthal direction. Furthermore all the data collected will be retained for post processing to find usage patterns. Information from this data mining could result in improvements to the smart radio system’s architecture.

References:

[1] J. M. Peha, “Spectrum Management Policy Options,” IEEE Communications Surveys, vol. 1, no. 1, pp. 66-68, Fourth Quarter 1998.

[2] S. Ramo, J.R. Whinnery, and T. Van Duzer, Fields and Waves in Communications Electronics, Third Edition, New York: Wiley, 1994.

[3] M.R. Andrews, P.P. Mitra, and R. deCarvalho, “Tripling the capacity of wireless communications using electromagnetic polarization,” Nature, vol. 409, pp. 316-318, January 18, 2001.

[4] R. Engelman et al, “Report of the Spectrum Efficiency Working Group,” FCC Spectrum Policy Task Force, November 15, 2002, available at http://www.fcc.gov/sptf/report.html.

[5] F.H. Sanders, B. J. Ramsey, and V.S. Lawrence, “Broadband Spectrum Survey at San Francisco, California May-June 1995,” NTIA Report 99-367, July 1999.

[6] F. van Diggelen, “Indoor GPS theory & implementation,” IEEE Position, Location, and Navigation Symposium, pp. 240–247, 2002.

[7] http://www.jcatech.com.

[8] K. Chang, I. Bahl, and V. Nair, RF and Microwave Circuit and Component Design for Wireless Systems, New York: Wiley, 2002.

[9] B. Razavi, RF Microelectronics, Upper Saddle River, NJ: Prentice Hall, 1998.

[10] H.J. De Los Santos, R.J. Richards, “MEMS for RF/Microwave Wireless Application The Next Wave – Part II,” Microwave Journal, July 2001.

[11] D. Peroulis, S. Pacheco, K. Sarabandi, and L.P.B. Katehi, “Tunable Limped Components with Application to Reconfigurable MEMS Filters,” IEEE MTT-s Digest, pp. 341-344, 2001.

[12] C.T.-C. Nguyen, ”Transceiver front-end architectures using vibrating micromechanical signal processors,” Meeting on Silicon Monolithic Integrated Circuits in RF Systems, pp. 23–32, 2001.

[13] J.B. Muldavin, and G.M. Rebeiz, “X-band tunable MEMS resonators,” Meeting on Silicon Monolithic Integrated Circuits in RF Systems, pp. 116–118, 2001.

[14] S. Diamantis, M. Ahmadi, G.A. Jullien, and W.C. Miller, ”A programmable MEMS bandpass filter,” Proceedings of the 43rd IEEE Midwest Symposium, vol. 1, pp. 522-525, 2000.

[15] C.S. Aitchison, “The Current Status of RF and Microwave Amplifier Intermodulation Performance,” IEEE Radio Frequency Circuit Symposium, pp. 113-116, 2000.

[16] J. C. Pedro, J. Perez, “An MMIC Lineraized Amplifier Using Active Feedback,” Microwave and Millimeter-Wave Monolithic Circuits Symposium, pp. 113-116, June 1993.

[17] S.G. Kang, I.K. Lee, K.S. Yoo, “Analysis and Design of Feedforward Power Amplifier,” IEEE MTT-S Digest, pp. 1519-1522, 1997.

Empirical Study of 802.11b Wireless Networks in the Presence of Bluetooth Interference

Cameron McKay and Fukumasa Masuda
Institute for Telecommunication Sciences
303-497-4418
cmckay@its.bldrdoc.gov

Abstract: Two complementary wireless networking standards, Bluetooth and 802.11b, operate in the 2.4 GHz Industrial, Scientific, and Medical (ISM) band. Although they use different methods to modulate and transmit data, significant interference can occur. Under certain conditions, a Bluetooth-enabled device can render an 802.11b connection almost useless. This paper presents measurement results from a study on the throughput of an 802.11b link when one end of the link is subjected to interference from Bluetooth devices.

I. Introduction

Wireless networks are rapidly becoming commonplace in businesses, homes, and schools. These networks have made Internet access more available than it has ever been in the past — a user can roam up to 100 m away from an access point and still remain online. One of the most popular medium-range wireless networking standards is 802.11b, which operates in the 2.4 GHz band.

Another fast-growing wireless technology is Bluetooth, which comes standard with many new cellular phones and handheld computers. It is intended for short-range (generally 10 m or less) cable replacement. Although its goals differ from those of 802.11b, it shares the same 2.4 GHz Industrial, Scientific, and Medical (ISM) band with 802.11b devices. This unlicensed band extends from 2400 to 2483.5 MHz in the United States.

Although Bluetooth and 802.11b use the same frequency range, their signals are distributed over this range in completely different ways. Bluetooth is a Frequency-Hopping Spread Spectrum (FHSS) standard, where data is transmitted over a 1-MHz-wide band. Bluetooth normally 'hops' to a different frequency 1600 times per second. The scheme used by 802.11b is Direct Sequence Spread Spectrum (DSSS), where a network will occupy a fixed 22-MHz-wide frequency band. The power in this band is not constant — it is weighted so that the frequencies near the center of the band carry more information.

The performance of 802.11b, though tolerant of some fixed narrowband interference, can be devastated by an FHSS scheme such as Bluetooth's. If the Bluetooth transmitter hops into the frequency band occupied by an 802.11b network, it is likely that the current 802.11b packet will be corrupted, requiring a retransmission. Because of the fast hop rate, there is ample opportunity for these collisions to occur.

Several excellent coexistence mechanisms have been developed — either by using traffic-shaping or other means — but none are in widespread use [1, 2].

Although many papers have been published about the interference between Bluetooth and 802.11b systems, few have presented any empirical data. Much of the available data was gathered in idealized environments [3], where Bluetooth and 802.11b devices were connected via coaxial cable and not permitted to radiate RF energy. The aim of this paper is not to endorse or refute any proposed interference models, but simply to present the results of our real-world study in the hopes that it may be useful to other researchers.

II. Test Setup

The focus of our tests was to gather sufficient data to draw basic conclusions about the robustness of 802.11b links. We did not attempt an exhaustive analysis of the mechanics of the interference between the two protocols, as this type of work has already been carried out several times before [4, 5]. Rather, we examined one specific real-world setup.

We configured a simple wireless network — one 802.11b-equipped laptop associated with one access point. The access point was connected to a 100 Mbps ethernet switch. Also connected to the switch was a desktop computer running an FTP server. (This is similar to the setup used in [6].)

In order to characterize the performance of the 802.11b link, we tested its speed at nine evenly-spaced Signal-to-Noise Ratios (SNRs) between 10 and 50 dB. At each signal level, we ran speed tests without Bluetooth interference and with one nearby Bluetooth unit transmitting. At SNRs of 50, 30, and 10 dB, we also measured the speed with two and three Bluetooth interferers.

To measure the SNR of the 802.11b network, we used a piece of software that displayed the current received signal and noise levels reported by the 802.11b card. (All SNR measurements were conducted before Bluetooth interference was introduced, as Bluetooth activity raises the noise floor and greatly affects the SNR.)

Our equipment was set up in a relatively complex environment, electromagnetically speaking. There were many metal surfaces present in the test area, so multipathing was inevitable and signal paths were difficult to predict. We feel that this is typical of the situations in which 802.11b and Bluetooth are used.

To test the speed of the link, the laptop computer would download five large files from the FTP server, and the transmission time would be recorded for each file. The files were either 50 MB or 20 MB, depending on the predicted speed of the link.

For our interferers, we acquired six Bluetooth transceivers and installed them in computers. A file transfer was set up between one pair of transceivers to simulate Bluetooth activity and to create a constant source of interference. The transmitting radio was placed 1 m away from the laptop's 802.11b radio. Up to three interferers could be created in this manner.

The power output of the 802.11b radios in the access point and the laptop was 15 dBm, and the Bluetooth interferers transmitted at 10 dBm. All power-saving and encryption features were disabled to allow the radios to perform at their full capacity.

III. Measurements and Results

Without any Bluetooth interference, the throughput of the 802.11b network stayed relatively constant for SNRs above 10 dB (Fig. 1). At 10 dB, the network speed dropped by approximately 50%. This is consistent with how 802.11b functions — for strong signals,

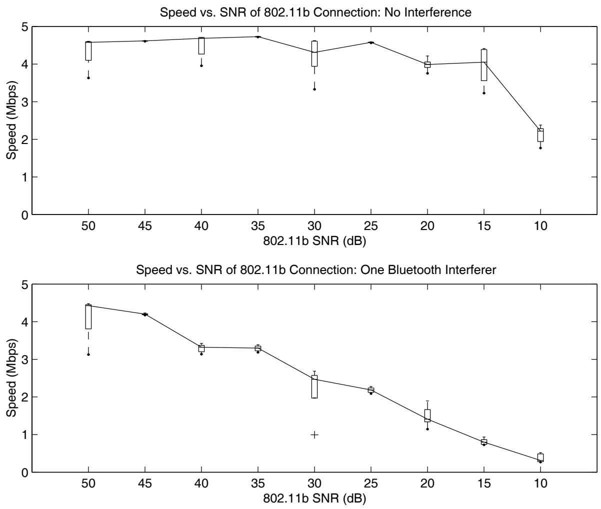


Fig. 1: Speed vs. SNR of 802.11b

the network sends data at 11 Mbps. However, when the signal level drops below a certain threshold (-82 dBm for our particular wireless card), the rate is rolled back to 5.5 Mbps. The signal level was -86 dBm when the SNR was 10 dB, so we believe that this feature was the cause of the large discontinuity in our results.

When one interferer was introduced, the throughput graph of the network changed drastically. The highest SNRs were nearly unaffected, but the speed of the network decreased almost linearly with the SNR — even after the network rate change that happened between 15 dB and 10 dB.

(Note that the single-interferer measurements performed with an SNR of 40 dB are anomalous, in that they are almost identical to the measurements taken at 35 dB. It is possible that the tester inadvertently changed the SNR during the test by moving too close to the 802.11b laptop's radio, or simply by changing his position slightly and attenuating a reflected signal. Also, an intermittent source of 2.4 GHz interference, such as a leaky microwave oven, may have been active near the lab during that particular test.)

By computing the median percentage drop in link speed caused by one Bluetooth interferer, one can see that the plot is almost perfectly linear between 15 dB and 45 dB, except for the 40 dB point discussed previously (Fig. 2). This result suggests an important notion. It appears that there is no 'magic SNR' at which a Bluetooth device will suddenly start crippling the throughput of 802.11b networks — an increase of the SNR by a few decibels will not significantly change a device's susceptibility to Bluetooth interference.

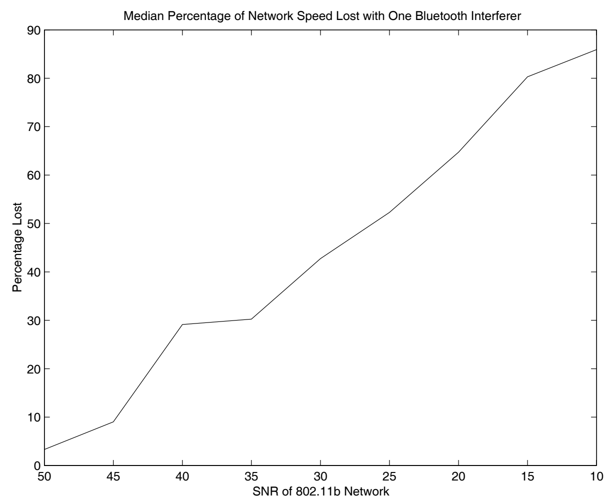


Fig. 2: Percentage of Speed Lost vs. SNR of 802.11b

The effects of additional Bluetooth interferers were heavily dependent on the SNR of the 802.11b connection (Fig. 3). At 50 dB, the network operated at a median speed of 3.8 Mbps, compared to 4.6 Mbps without interference. At 30 dB, the speed transfer test only successfully completed with one or two interferers (the FTP connection could not be sustained with three Bluetooth transmitters present). When subjected to two interferers, the network speed dropped to less than 1 Mbps. With an SNR of 10 dB, the speed test would fail if more than one Bluetooth device was active. The median data rate with one interferer was 0.312 Mbps.

Although the degradation of the network speed appears linearly correlated with the number of interferers, more testing is needed to confirm this relationship. The slope of the speed-versus-interferers graph is heavily dependent on the SNR of the 802.11b network, and further data is required in order to determine how the slope changes with the SNR.

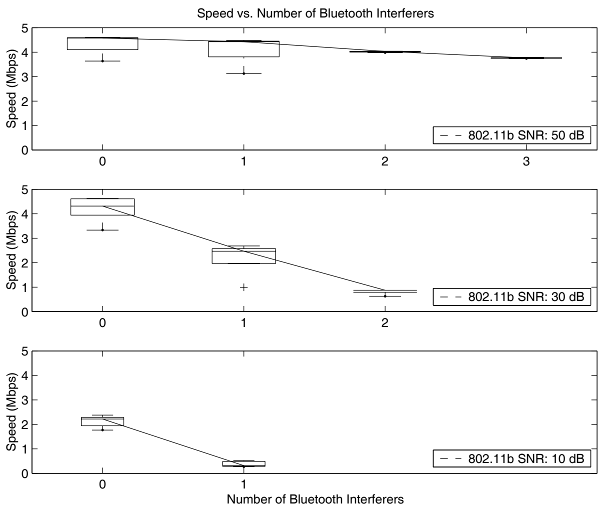


Fig. 3: 802.11b Speed vs. Number of Interferers

IV. Conclusions

Although Bluetooth and 802.11b devices can coexist under many circumstances, some Bluetooth setups can cripple an 802.11b link. At SNRs below 20 dB, over two thirds of the network speed can be lost by introducing one nearby Bluetooth interferer. Two or more interferers at an SNR of 30 dB or less can also severely impact performance.

The most surprising result is the linear correlation between the SNR of the 802.11b network and the speed loss caused by one Bluetooth interferer. We see that the 802.11b link is predictably affected by Bluetooth interference, and that there is no point at which Bluetooth devices suddenly start affecting performance — they always have some effect, even at high SNRs. Although the effects of two or more Bluetooth interferers seemed to follow a linear trend as well, this relationship requires more research.

With more empirical study, it may be possible to evaluate the existing interference models, and find one which can be used to reliably predict the effects of Bluetooth-802.11b interference under a variety of conditions.

V. References

[1] J. Lansford, A. Stephens, R. Nevo, "Wi-Fi (802.11b) and Bluetooth: Enabling coexistence," *IEEE Network*, vol. 15, no. 5, pp. 20-27, Sep/Oct 2001.

[2] C.F. Chiasserini and R.R. Rao, "Performance of IEEE 802.11 WLANs in a Bluetooth environment," in *Proc. Wireless Communications and Networking Conference*, vol.1, pp. 94-99, 2000.

[3] R.J. Punnoose, R.S. Tseng, D.D. Stancil, "Experimental results for interference between Bluetooth and IEEE 802.11b DSSS systems," in *Proc. Vehicular Technology Conference*, IEEE VTS 54th, vol. 1, pp. 67-71, Fall 2001.

[4] I. Howitt, "IEEE 802.11 and Bluetooth coexistence analysis methodology," in *Proc. Vehicular Technology Conference*, IEEE VTS 53rd, vol. 2, pp. 1114-1118, Spring 2001.

[5] M. Fainberg and D. Goodman, "Analysis of the interference between IEEE 802.11b and Bluetooth systems," in *Proc. Vehicular Technology Conference*, IEEE VTS 54th, vol. 2, pp. 967-971, Fall 2001.

[6] J. del Prado and Sunghyun Choi, "Experimental study on coexistence of 802.11b with alien devices," in *Proc. Vehicular Technology Conference*, IEEE VTS 54th, vol. 2, pp. 977-981, Fall 2001.

Stretchable Architectures for Next Generation Cellular Networks

S. Lakkavalli(lakkavas@cs.pdx.edu), A. Negi, and S. Singh
Department of Computer Science, Portland State University

***Abstract*—This paper examines the technique of conserving energy in 3G cellular systems by allowing calls between a mobile and the base station to be forwarded by some other mobile in the cell. This "stretched" call has the benefit of reducing overall energy cost for the call in cases where the calling mobile is NLOS (Non Line of Sight) to the base station but is LOS to the intermediary and the intermediary is LOS to the base station. The main contributions of the paper are twofold: first, we show how this stretched call model can be implemented in UMTS and CDMA2000; second, we show that this stretched call model results in overall energy savings in a cell of between 2x and 7x.**

I. INTRODUCTION

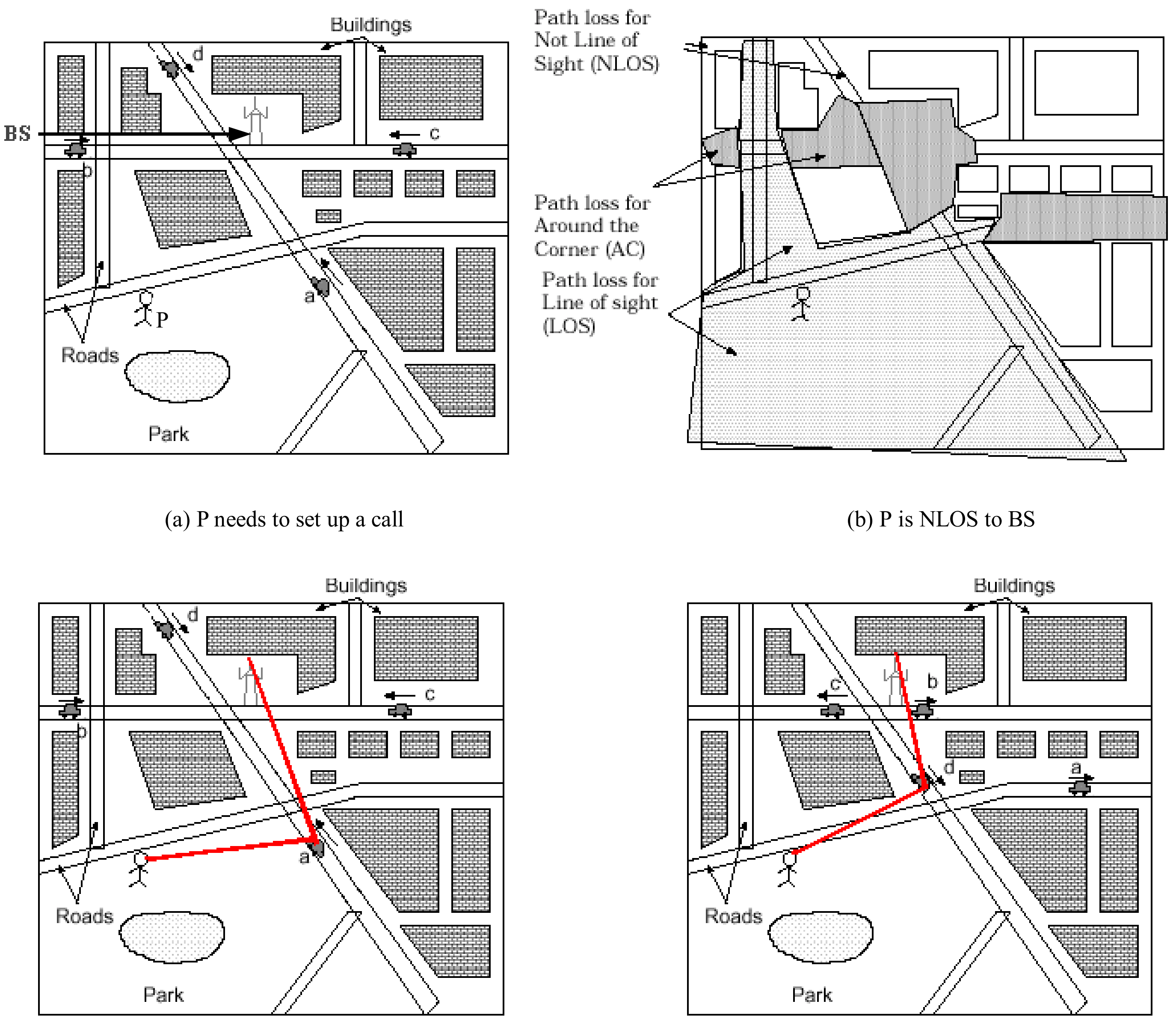
Approaches for saving energy at a mobile terminal (MT) in cellular networks include, dynamic transmission power control whereby the MT increases or decreases its transmission power in response to requests received from the base station (BS) [1][2], terminating calls when the interference becomes too high [3], or putting the MT into deep sleep state and waking it up periodically to receive data [4] (as is done in the paging protocol). In this paper we examine a new technique to enable power savings at the MT. The key idea is to split the connection between the MT and BS at an appropriate intermediary (another MT). Therefore, on the reverse link, the MT transmits to the intermediary and the intermediary forwards the call to the BS. In Fig.1, say the mobile **P** needs to set up a call to the base station (BS). However, as shown in Fig.1 (b), the **BS** is NLOS to this mobile. Let us assume that there are other mobiles present in the cell that are idle. Mobile **a** is idle and, as shown in Fig.1(c), it is LOS to the **BS** as well as to the mobile **P.** The call is therefore set up as **P-a-BS**. After some time, **a** is no longer LOS of the BS and the call is handed off to another mobile **d** as shown in Fig.1(d), who is LOS to both **P** and the BS. This *stretched* call model ensures that the total energy used for the call is less than that of a direct call if the intermediaries are selected appropriately. The problems involved in deploying such a stretched call model include intermediary-to-intermediary handoffs, preserving end-to-end encryption, maintaining closed-loop power control at the MT and the intermediary, and efficient algorithms for intermediary selection. In this paper we describe our stretched call model in detail and explain how it fits into CDMA2000, Universal Mobile Telecommunications System (UMTS), and Time Division Synchronous Code Division Multiple Access (UMTS TD-SCDMA) standards for 3G cellular networks. We also show that unlike Opportunity Driven Multiple Access (ODMA), our scheme ensures bounded end-to-end delay enabling voice calls to be carried. Finally, we present results from detailed simulations of our model and show that energy savings of as much as 2x to 7x is possible in many cases even when using a simple greedy algorithm for intermediary selection.

There are two barriers to implementing the stretched call model in 3G systems: first, the problem of additional hardware complexity in the MTs to enable them to become intermediaries capable of carrying calls, and second, convincing users to allow their MTs to be used as intermediaries when idle. Our work assumes that each MT has *one transmitter and two receivers, which already exist in 3G cellular phones*. Furthermore, the complexity of handoffs between intermediaries is also minimal because, as in the case of repeaters [12], the MT treats the signals from the intermediary as multipath of the BS's transmission. The second challenge, that of convincing users to allow their MTs to carry calls is more problematic because being an intermediary means that the intermediary is allowing its battery to be used for someone else's call! We believe that a differential pricing model (where stretched calls are charged at a higher rate than direct calls) with appropriate credits (either as cash or free call minutes) being provided to intermediaries will encourage more users to allow their cell phones to be used as intermediaries when idle. While interesting, in this paper we focus only on the technical problem of implementing stretched calls and leave a discussion of the pricing model to a later paper.

Section II describes other related work in the area. We describe the stretched connection model in section III and in section IV we provide how power control, handoff and security aspects of 3G networks can be handled in our model. We describe our simulator and the experimental design in detail in sections V & VI and present the main results in section VII.

II. RELATED WORK

The idea of using relays in cellular systems is not a new one and has been used to provide coverage in dead spots [5] or to improve capacity in dense areas such as shopping malls etc. In these systems, repeaters allow MTs to use lower transmit power as well as provide spectrum reuse within the same cell. It is important to note, however, that our scheme differs conceptually from the idea of using repeaters in two ways: first, intermediaries in our model are themselves MTs and thus mobile, and second, our scheme provides overall energy efficiency in the cell regardless of location of MT, whereas, in the case of repeaters, only mobiles in range of the repeaters can avail themselves of the reduced transmit power opportunity. It is also important to point out that using repeaters incurs an additional infrastructure cost while our scheme has no such capital cost.



(a) P needs to set up a call

(b) P is NLOS to BS

(c) Call setup via intermediary ‘a’

(d) Call handoff from intermediary ‘a’ to ‘d’

Figure 1. Stretched Call Scenario

The 3G standards of Universal Mobile Telecommunications System (UMTS) [6] specify a adhoc mechanism called ODMA for multi-hop connections between the BS and MTs. The intermediaries here are other MTs and each MT maintains a connectivity table of other MTs within range. Each entry in the table contains the identity of the MT, the data rate possible to that MT and the transmit power level required to reach that MT. This information is refreshed via the broadcast of periodic probe packets, and so every MT regularly updates its routing table. In our case, where power conservation is the goal at the MT, the overhead of periodic updates and additional processing required for maintenance of the routing tables will decrease the already depleting battery resources at the MT. For routing mechanism, we use the repeater principle of passive handoff, where the MT thinks that signals from the intermediary are like multipath from the BS. So here, the MT need not know the identity of the intermediary, and this saves the overhead of routing table maintenance at the MT. Therefore, our two-hop relaying mechanism is different from the adhoc network mechanism like ODMA. The responsibility of choosing the intermediaries is with the BS, which has a database of all the nodes in its domain. The advantage of letting the BS choose the intermediary is that it has knowledge of the speed and direction of motion of both the MT and the intermediary, based upon which the BS can optimize and select the best intermediary (in terms of low energy as well as stability to minimize the need for handoffs). Another important advantage of our stretched connection model is that, if a stretched connection can be maintained for the duration of the call, the delay is not variable as in the case of ODMA because we only interject one node between the MT and the BS. In ODMA, on the other hand, the number of intermediaries can vary, and therefore the end-to-end delay is variable making it unsuitable for real time applications like voice and video conferencing.

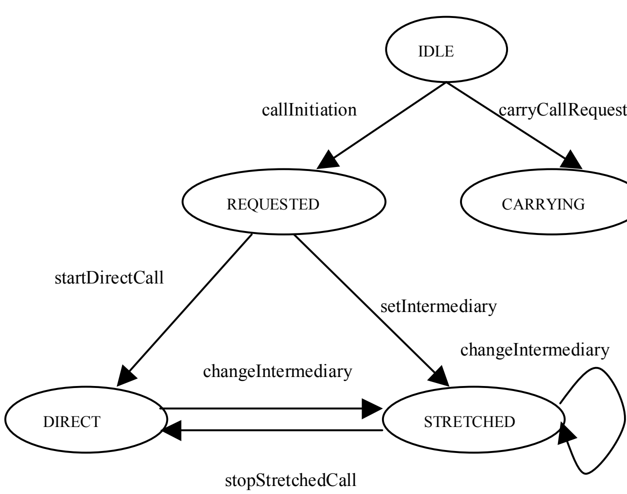


Figure 2. Finite State Automata of a Mobile

III. OVERVIEW OF STRETCHED CONNECTION MODEL

A stretched connection has an intermediary, which carries the call from mobile to BS and vice-versa. The connection between the mobile and the intermediary is called as Lower Arm (LA) and the connection between the Intermediary and BS is called as Upper Arm (UA). The connection between mobile and BS is called a Direct Connection. An intermediary can be stationary or mobile. Mobile phones with sufficient battery power or a car phone are two examples of intermediaries. The reverse link as well as the forward link can be split at the intermediary or we can have an asymmetric case where only the reverse link is split at the intermediary. The intermediary in our system is assumed to have *only one transmitter* and therefore it has to time multiplex its transmission between the two arms of the stretched connection. To increase the usability of the intermediary, it is *assumed to have dual receivers*, so that it can receive transmissions from both the mobile and the BS simultaneously.

A. State Machine

The node has five call states – IDLE, REQUESTED, DIRECT, STRETCHED, CARRYING as shown in Fig. 2. The node is in IDLE state when it is not calling. It goes to REQUESTED state when a call initiation has been notified to the BS (BS). The BS will assign an intermediary to the Mobile Terminal (MT), in which case it will go to STRETCHED state. If the BS was not able to assign an intermediary it goes to DIRECT state. The node can toggle between DIRECT and STRETCHED states within the duration of a call. Once the call is terminated, it goes back to IDLE state. If the node is IDLE and the BS requests it to carry a call, it goes to CARRYING state. From CARRYING state, the node can transition to either IDLE state or to REQUESTED state, if the intermediary itself wants to initiate a call.

In Fig. 3, we have the BS's view of mobile states. It maintains the current state of the MT, both its call state and mobility state. The BS maintains two additional call states for the node. They are CHANGING-INTERMEDIARY and CHECKING-INTERMEDIARY states. CHANGING-INTERMEDIARY state of the node is maintained between intermediary handoffs. This state is required to represent the state when the BS is finding a suitable intermediary for the MT. The CHECKING-INTERMEDIARY state is used for the possible intermediary when it is being sent a request to carry. The mobile has two mobility states – MOVING and STATIONARY. The call states and mobility states are updated to the BS, whenever they change.

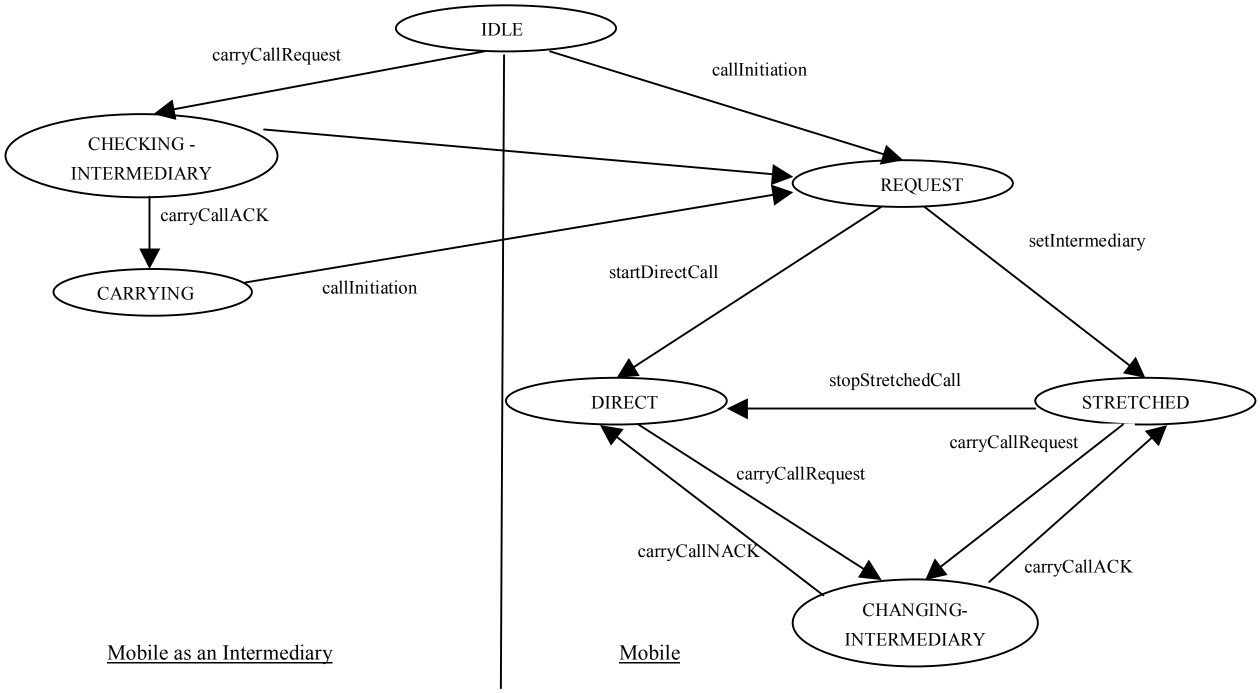


Figure 3. Finite State Automata of a Mobile in the Base station

B. *Finding An Intermeaıary*

Since the number of hops in the radio link is increased by just one, the BS can continue to control the handoffs between intermediaries. The BS can frequently broadcast the forward link orthogonal codes and the reverse link phase offsets of the long PN code in case of cdma2000 or the scrambling code in case of UMTS Radio Terrestrial Access (UTRA)[1].

The mobiles in IDLE mode on reception of these broadcasts can adjust their PN generator hardware and their demodulators to synchronize with the reverse and forward link traffic channels. It is assumed here that the BS transmits required information in the broadcast channel to enable the intermediaries to synchronize with the mobile's traffic channel.

The intermediaries maintain an active list of mobiles in DIRECT state, whose signal SNR exceeds the threshold setup for the connection. The intermediaries use the same mechanism of soft handoff to maintain active, candidate, neighbor and remaining sets. But here, they maintain these sets containing mobiles wishing to get their calls carried, instead of the BS'. The intermediaries repeatedly report the active list members in their dedicated control channels to the BS. The BS, which will receive similar information from IDLE mobiles in the vicinity of the DIRECT state mobile, will run optimizations to select the best intermediary. Algorithms to determine the best intermediary are still under research. In this paper we use a simple *greedy* approach to find the appropriate intermediary – the intermediary that minimizes the total power for the connection is selected.

It is important to note that the upper and lower arms of the stretched connection require different frequencies, so as to prevent the transmitter at the intermediary from overwhelming its own receiver. Though two frequencies are required, which might seem to double the required bandwidth and therefore half the capacity, lower transmitting powers of the MT and intermediary decrease intracell and intercell interference, which increases the capacity of the cell. This topic is under research.

IV. 3G STANDARDS RELATED ISSUES

So far, we have seen that intermediary-to-intermediary handoff can take place using the soft handoff mechanism that is part of 3G specifications. We have also seen that single transmitter and dual receivers required for a stretched connection are already a part of 3G mobiles. As in ODMA, it can be assumed that uplink spreading and modulation can be made similar to downlink spreading and modulation, so that minimal hardware changes are required at the intermediary. In this section, we describe how delay, power control, handoff and security aspects of 3G networks can be adopted in our model. Before we discuss the above issues, however, we need to observe that the

[1] In cdma2000, the phase offsets are determined using Equipment Serial Number (ESN) number of the mobile. To generate the phase offset, the ESN number is required. Therefore, this has a security risk. But in UTRA, the orthogonal codes and the scrambling codes do not have any security use and therefore, they can be broadcasted.

intermediary can operate in one of two modes when carrying a call – *translating and non-translating* modes.

In translating mode, the whole frame has to be received and regenerated, while in non-translating mode, the intermediary has to deal only with demodulation, despreading and decoding the power control bits from every Power Control Group (PCG). The power control mechanism used for both the modes is similar. In the translating mode, the intermediary receives the signal from any arm, say for example the upper arm in the case of forward link, demodulates with the carrier of the receiving arm, reads the power control bit from each slot – PCG so that it can modify its transmit power in that arm, de-spreads, de-scrambles, de-interleaves and decodes the data. It calculates the Frame Error Rate (FER) and compares it with the one in the frame so that it can modify the required value of Signal to Noise Ratio (E_b/N_0). Also, for non-real time data, the Radio Link Protocol (RLP) in the link layer uses this information for controlling retransmissions. While transmitting the frame, the intermediary recalculates the FER, appends it to the data, encodes, interleaves, scrambles, spreads and then modulates the frame with the carrier of the other arm. It also punctures the power control bit in every PCG to control future transmissions from that arm. Similar is the case for the reverse link.

In the non-translating mode, the intermediary has to only process the PCG in a frame before forwarding it. There are 15 PCG in a UTRA frame [2] and 16 PCG in cdma2000 [7] and depending on the frame duration, the duration of the PCG can vary between 0.667ms to 1.25ms. For the forward link for example, for each slot, the intermediary only extracts the power control bit the BS sends to it and inserts its own power control bits for the mobile. So, the intermediary has to save just a PCG before forwarding it. Similar is the case for the reverse link. But, to verify the FER, the intermediary has to still save the whole frame.

If it is a translating intermediary, due to the regeneration time, the regeneration will include interleaving and de-interleaving delays and processing delays. Shorter the frame, lesser is the delay. So, if TD-SCDMA is used, where the frame duration is 5ms, the delay is minimized.

The security mechanisms in cdma2000 and UMTS differ. Voice privacy is provided in cdma2000 in the physical layer, whereas data privacy is in the MAC layer. Because voice privacy is implemented in the physical layer in cdma2000, and the stretched model in translating mode needs do the entire physical layer processing, the stretched model cannot be used for voice. Data calls, however, can use the stretched model. On the other hand, since voice as well as data privacy is implemented at the MAC in UMTS, the stretched model can be used for both types of calls without affecting privacy.

V. SIMULATOR DESIGN

We have developed a very detailed standards-compliant cell-site discrete event simulator to evaluate the stretched connection model. The propagation model implemented is a recursive model specified in [10] and explained in [11]. The arrival of calls at the BS is modeled as a Poisson distribution, with inter-arrival duration modeled as an exponential distribution. The node mobility is modeled as a Gaussian distribution. The mean velocity is a variable factor with the standard deviation equal to 10% of the mean. The mobility is assumed as a Gaussian distribution because most of the vehicles travel at the mean velocity, with few traveling at higher or lower velocities.

There are two types of events – external events and internal events. External events dictate the time of call initiation, termination and node mobility. These are controlled using the call and mobility probabilistic models used in the simulator. In response to these external events, internal events are generated. These internal events occur within the duration of a call. External events have higher priority over internal events.

The simulator creates a grid-based map with user-specified grid sizes. The user can add obstructions to simulate buildings as in a Manhattan environment. The user also specifies location of roads and intersections. Directional probabilities are assigned to each of the four roads originating from the Intersection. A mean velocity is assigned to each road with a standard deviation of 10% of the mean velocity. When a MT reaches an intersection, the direction and velocity are assigned using the above probabilities. The BS can be located anywhere in the map and this is also a user specified parameter. The size of the map is limited by the maximum allowable path loss for the terrain that the map is trying to simulate. For example, for speech in a pedestrian environment like the Manhattan model, the maximum reverse link path loss is 148.4dB [6].

Path loss between any two points in the terrain is calculated and stored for repetitive experiments. For improved accuracy, we use the inverse square weighted interpolation formula. This formula takes the path loss between centers of the current and next grids of both nodes. Then, by calculating the offsets of the distance that both the nodes have moved (zero in case of BS), we can find out the exact path loss.

VI. EXPERIMENTAL SETUP

The size of a grid is fixed for our simulations at 20meters. The link budget values are taken from [10] for a pedestrian terrain. The *factors* considered here are *numbers of nodes, call rate* and *cell antenna gain*. The numbers of nodes is varied from 5 to 30 in steps of 5. The call rate is 1 and 2 calls per hour per MT. The call duration has a mean value of 120seconds. The antenna gain was set at either 6dB or 10dB. The simulations are done for a single cell with only one BS. Two sets of experiments were conducted. In the first, the *BS was kept in the center of the cell* and in the second case, *the BS was kept in the corner of the cell*. The mean node speed is 1.5meters/sec. The BS finds the best intermediary whenever the call state changes or when a MT (either the MT itself or its intermediary) changes grid location.

For the link budget of the lower arm of the reverse link, there is no BS gain because it involves communication between mobiles and the mobiles are assumed to have isotropic antennas. Similarly, for the forward link in the lower arm of the stretched connection, there is no BS gain. For the upper arm of the reverse link the link budget is similar to a direct connection. The mobile antenna gain is assumed to be zero for all cases.

VII. RESULTS AND DISCUSSION

For an evaluation of our stretched call model, we used the following metrics:

- Total energy used during a run of 1000s (*direct calls* between MT and BS or *stretched calls* where the intermediary is selected using the greedy algorithm described in section II.B). We also show the energy used per node.
- Number of handoffs between intermediaries.
- Percentage of time the calls were stretched as a function of number of MTs.

In the remainder of this section, we discuss only the results obtained using a 10dB antenna gain. The results for 6dB show a similar trend and have been left out for reasons of space.

The first set of figures shows results for the case when the *BS was located in the center* of the cell. Fig. 4 shows the total system energy as a function of the number of nodes in the cell for two different call rates. Fig. 5 shows the same information on a per node basis. As can be easily seen, higher call rates consumes more energy. However irrespective of call rates the stretched model consumes less energy as compared with the direct model. *For a call rate of 1, the savings are greater than 50%, while for a call rate of 2 the savings vary from 3x to more than 7x.* Also the spread in confidence interval is less for stretched system.

Fig 6 shows the number of handoffs during periods of length 1000s. As we can see, the number of handoffs is higher at higher call rates. This is because at higher call rates, more MTs are actively placing calls and thus, if they are serving as intermediaries when a new call request arrives, the carried call will need to be handed off to another idle MT. This figure shows the handoff rate for the case when the BS is at the center of the cell. We observe similar numbers for the case when the BS is in a corner of the cell.

We ran similar experiments when the *BS was in a corner* of the cell (this is a realistic scenario where the BS covers the cell using sectored antennas). Fig 7 shows the direct and stretched mode energies. On comparing with Fig. 4 when the BS was at center of the cell, it can be seen that for higher number of nodes, more energy is being spent. This is because the distance to the BS has increased thus directly impacting the energy of the direct or stretched call. In terms of savings, we see a *energy savings of up to 4x* when using the stretched call model.

In Fig 8 for BS at center, the percentage of time spent by a node with stretched connections increases with the number of nodes. Similar is the case for BS at corner scenario. We see that as the number of nodes increase, the percentage of carrying time of a node does not increase as much as the stretched time. Thus the overhead on the intermediaries is minimal.

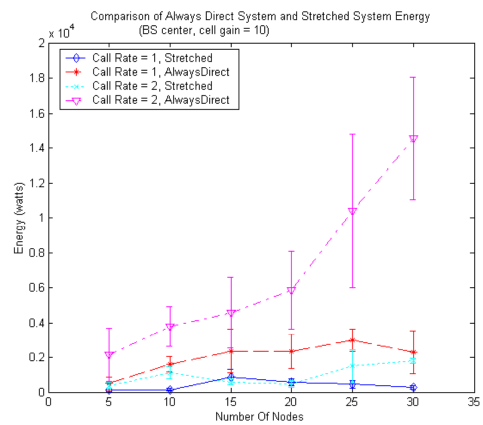


Figure 4. Total system energy for always-direct system and stretched system with BS at center

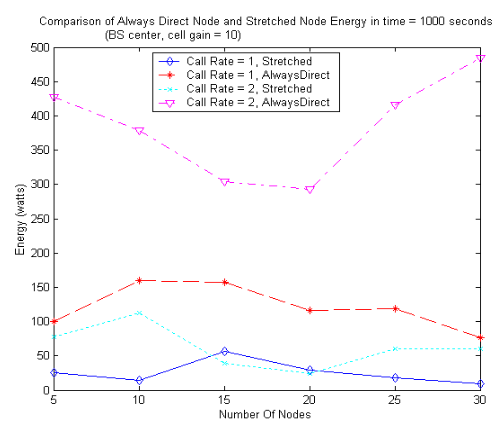


Figure 5. Per node energy for always direct system and stretched system with BS at center

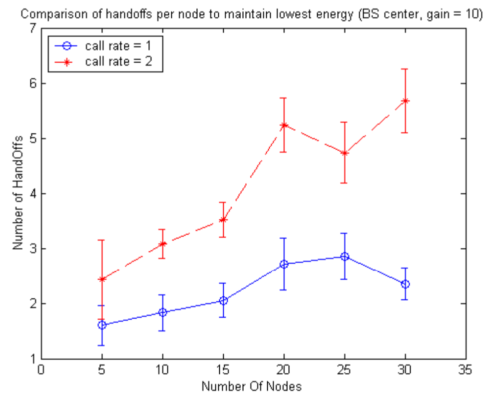


Figure 6. Number of handoffs required per node with BS at center

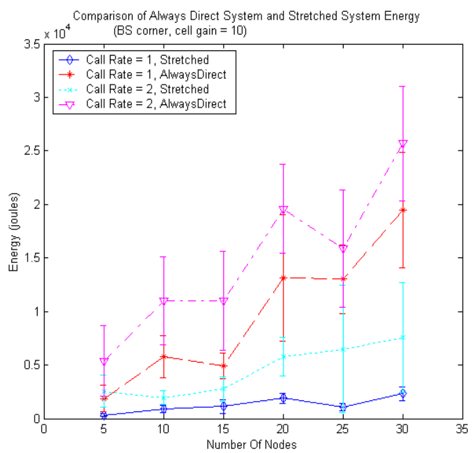


Figure 7. Total system energy for always-direct system and stretched system with BS at corner

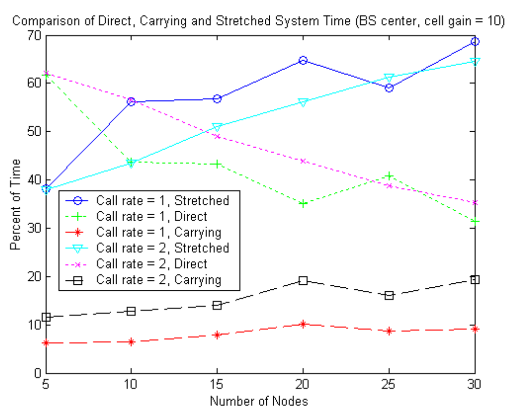


Figure 8. Percentage of direct, stretched, carrying duration with BS at center

VIII. CONCLUSIONS

In this paper we presented a stretched call model for 3G systems in which high energy direct calls between mobiles and BS could be split at another mobile in order to reduce the overall energy used for the call. *We described how this call model could be implemented in UMTS and CDMA2000 standards and guidelines for selecting different implementation parameters.* Finally, we evaluated the stretched call model in a simulator to determine the extent of energy savings obtainable. As we show, energy savings of up to *7 times* are possible! This huge energy reduction is a good reason for further exploration of the stretched call idea in next generation cellular systems.

IX. REFERENCES

[1] Jhong Sam Lee and Leonard E. Miller, *CDMA Systems Engineering Handbook*, Artech House, Boston, 1998.

[2] Vijay K. Garg, *Wireless Network Evolution, 2G to 3G*, Prentice Hall PTR, NJ, 2002.

[3] John M. Rulnick and Nicholas Bambos, "Mobile power management for wireless communication networks", *Wireless Networks*, vol. 3, pp. 3-14, 1997.

[4] Tajana Simunic, "Dynamic management of power consumption", *Power Aware Computing*, Edited by R. Graybill, R. Melhem, Kluwer Academic Publishers, 2002.

[5] George N. Aggelou and Rahim Tafazolli, "On the relaying capacity of next-generation GSM cellular networks", "*IEEE Personal Communications*", pp. 40-47, Feb 2001.

[6] Concept evaluation, Universal Mobile Telecommunications System (UMTS), *UMTS Terrestrial Radio Access (UTRA), UMTS 30.06 version 3.0.0*, 1997.

[7] Physical Layer Standard for cdma2000 Spread Spectrum Systems, Release A, *3rd Generation Partnership Project 2 (3GPP2),* 3GPP2 C.S0002-A, 2001.

[8] Services and service capabilities, *Universal Mobile Telecommunications System (UMTS),* 3GPP TS 22.105 version 5.1.0 Release 5, 2002.

[9] M.Haardt, A.Klein, R.Koehn, S. Oestreich, M. Purat, V. Sommer and T. Ulrich, ``The TD-CDMA based UTRA TDD mode,'' *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1375-1386, Aug. 2000.

[10] Selection Procedures for the choice of Radio Transmission Technologies of the UMTS, *Universal Mobile Telecommunications System (UMTS)*, UMTS, 30.03 version 3.1.0, TR 101 112, 1997.

[11] Jan-Erik Berg, "A recursive method for street microcell path loss calculations", *PIMRC*, 1995.

[12] Repeater Technologies, Inc., http://www.repeaters.com, July 10, 2002.

On Coherency Requirements in Steered Beam Adaptive Antenna Systems

Sven Petersson* and Aare Mällo
Antenna Research Center Radio Network Development,
Ericsson AB, SE-431 84 Mölndal, Sweden.
*Phone: +46-31-7473583. Fax: +46-31-7473993. Email: sven.petersson@ericsson.com

Abstract: *The effect of array errors (amplitude-, phase- and time errors) on the ensemble average directivity of an adaptive antenna array is studied using Monte-Carlo simulations. It is shown that the effect of the errors can be reduced by introducing a so-called Butler matrix close to the radiating elements of the antenna.*

1 Introduction

Adaptive antenna arrays is a means to improve the performance of mobile communication systems by decreasing the cochannel interference, resulting in an increased spectral efficiency and an increased trunking efficiency [2]. However, to achieve the desired performance gain it is necessary that the entire signal chain is properly calibrated, certainly during production, and typically also during normal operation

In this paper Monte-Carlo simulations are used to investigate the effect of calibration errors (amplitude-, phase- and time errors) on the average azimuthal directivity pattern of a steered beam antenna system. The directivity is chosen as a suitable performance measure since it does not require us to specify a complex traffic model. Also, since time errors are included in the present model, simulations are chosen as the simplest tool of evaluation. The effect of amplitude- and phase errors only can be described analytically as shown in reference [1], but with time errors present the analytical approach becomes more complicated.

To illustrate the effect of array errors three model systems, A, B and C, are considered here (Figure 1). System A is an ideal array without any errors. System B is a non-ideal array with errors present, and finally, system C is a non-ideal array with a Butler matrix placed close to the radiating elements in the antenna. It is assumed that the Butler matrix together with the antenna elements forms a perfectly coherent subsystem. The array errors are assumed to be independent, and introduced in the feeder cables as shown in Figure 1.

The purpose of the Butler matrix in system C is to reduce the effect of array errors. This is indeed the result, and, as will be shown later in this report, in some directions the presence of the Butler matrix will even nullify the effect of array errors.

The outline of this paper is as follows. In section 2 analytical expressions for the *instantaneous*, azimuthal directivity (IAD) of the three model systems in Figure 1 is derived. In section 3 these analytical expressions are taken as a starting point for calculating a time and ensemble average of the IAD using Monte-Carlo simulations. Finally, in section 4 a short discussion and a summary of the results are given.

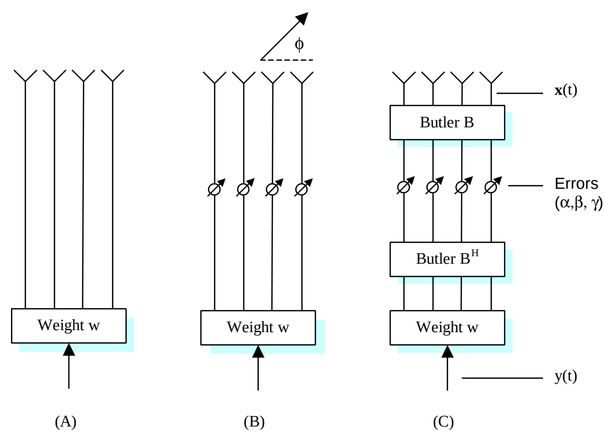


Figure 1. The three model systems considered in this paper. (A): Ideal antenna array without errors, (B): Non-ideal array with errors, and (C): Non-ideal array with errors and with a Butler matrix "B" close to the radiating elements. In the figure f denotes the azimuth angle, $y(t)$ is the baseband signal, $\mathbf{w}$ is an adaptive (complex) weight vector that is used to shape the antenna pattern, and $\mathbf{x}(t)$ is the antenna excitation vector. Parameters a, b and g represent the errors in amplitude, phase and time, respectively.

2 Model

2.1 Instantaneous directivity

The instantaneous, azimuthal directivity (IAD) for an antenna array at time t is defined as

$$D(q,f,t) \equiv 2pF(q,f,t)/P_{rad}(q,t) \quad (2.1)$$

where $F(q,f,t)$ is the radiated power per unit solid angle in direction $\{q,f\}$, and $P_{rad}(q,t)$ is the total radiated power per radian in elevation angle q,

$$P_{rad}(q,t) \equiv \int_0^{2p} F(q,f,t)df \quad (2.2)$$

The radiated power F can be written as

$$F(q,f,t) = |G(q,f,t)|^2 \quad (2.3)$$

where G is the complex, antenna far-field pattern,

$$G(q,f,t) = g_e(q,f)g_a(q,f,t) \quad (2.4)$$

and g_e and g_a are the element- and the array-patterns, respectively. The element-pattern itself is not of particular interest here, and we simply assume that it has a sinusoidal shape, and that it is directed in the positive y-direction,

$$g_e(q,f) = \begin{cases} \sqrt{2/p}\,\sin(f), & 0 \le f \le p \\ 0, & \text{else} \end{cases} \quad (2.5)$$

The array-pattern can be written as

$$g_a(q,f,t) = \mathbf{s}^{\mathrm{H}}\mathbf{x}(t) \quad (2.6)$$

where $(\cdot)^{\mathrm{H}}$ denotes the Hermit conjugate, $\mathbf{x}(t)$ is the L-dimensional antenna excitation vector, and $\mathbf{s}$ is a short-hand notation for a steering vector in direction $\{q,f\}$,

$$\mathbf{s}^{\mathrm{T}} = [s_1, s_2, \cdots s_L] \quad (2.7)$$

$$s_k = s_k(q,f) = \exp\{i2pf_c t_k(q,f)\} \quad (2.8)$$

In eqs. (2.7-8) $(\cdot)^{\mathrm{T}}$ denotes matrix transpose, L is the number of antenna elements, f_c is the carrier frequency, and t_k is the time delay for a plane wave traveling from antenna element number k, measured relative to the origin,

$$t_k(q,f) = \mathbf{r}_k \cdot \hat{n}(q,f)/c \quad (2.9)$$

where c is the speed of light, $\hat{n}(q,f)$ is the unit vector in direction $\{q,f\}$, and $\mathbf{r}_k$ is the position of antenna element k.

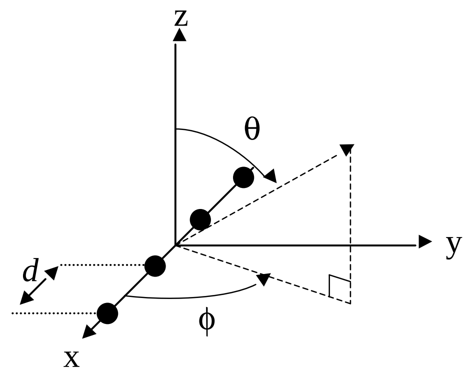


Figure 2. Co-ordinate system. Filled circles indicate the antenna element positions.

Assuming that the antenna elements are located on the x-axis with equal separation d (see Figure 2), the position vector is given by

$$\mathbf{r}_k^{\mathrm{T}} = [d(k-(L+1)/2), 0, 0] \quad (2.10)$$

For system C in Figure 1 the antenna excitation vector in eq. (2.6) can be written as

$$\mathbf{x}(t) = BM(t)B^{\mathrm{H}}\mathbf{w} \quad (2.11)$$

where B is the $L \times L$ Butler matrix, and M is a time-dependent, diagonal, $L \times L$ matrix that describes the baseband signal, including the array errors,

$$B_{kl} = \frac{1}{\sqrt{L}}\exp\left\{i\frac{2p}{L}\left(k-\frac{L+1}{2}\right)\left(l-\frac{L+1}{2}\right)\right\} \quad (2.12)$$

$$M_{kl}(t) = P_{kl}G_{kl}(t) \quad (2.13)$$

where P and G are two diagonal matrices,

$$P_{kl} = (1+a_k)\exp\{ib_k\}d_{kl} \quad (2.14)$$

$$\begin{aligned} G_{kl}(t) &= y(t-g_k)d_{kl} \\ &= a\sum_{m=-\infty}^{m=+\infty}\exp\{ij_m\}h(t-mT_c-g_k)d_{kl} \end{aligned} \quad (2.15)$$

In eqs. (2.14-15) a_k is the relative amplitude error in feeder cable number k, b_k is the phase error, g_k is the time delay in signal, d_{kl} is the delta-function, a is an amplitude factor, T_c is the chip period time, j_m is the phase in the I-Q diagram for the chip at time $t = mT_c$, and $h(t)$ is the impulse-response for the pulse-shaping filter.

In the present work it is assumed that the pulse-shaping filter is a root-raised cosine filter. Then, for $t = 0$,

$$h(t) = 1 + a\,\frac{4 - p}{p} \tag{2.16}$$

for $|t| = 1/4a$,

$$h(t) = \frac{a}{\sqrt{2}p}\left[(p - 2)\cos\left(\frac{p}{4a}\right) + (p + 2)\sin\left(\frac{p}{4a}\right)\right] \tag{2.17}$$

and otherwise,

$$h(t) = \frac{\sin(p(1 - a)t) + 4at\cos(p(1 + a)t)}{p\left(1 - (4at)^2\right)t} \tag{2.18}$$

Note that in eqs. (2.16-18) a is a filter design parameter, not to be confused with the amplitude error a_k in eq. (2.14).

Assuming QPSK modulation, the phase angle j_m in eq. (2.15) can take values

$$j_m \in \left[\frac{p}{4}, \frac{3p}{4}, \frac{5p}{4}, \frac{7p}{4}\right] \tag{2.19}$$

The weight vector $\mathbf{w}$ in eq. (2.11) can be designed in many ways to create a specific antenna pattern [3]. However, to keep things simple it is here assumed that it is designed only to create a plane wave in a given target direction $\{q_0, f_0\}$, that is,

$$\mathbf{w} = \frac{1}{L}\mathbf{s}_0 \equiv \frac{1}{L}\mathbf{s}(q_0, f_0) \tag{2.20}$$

where $\mathbf{s}$ is defined by eqs. (2.7-8).

The so-called array-factor for system (C) in Figure 1 can now be written as

$$|g_a(q, f, t)|^2 = \mathbf{s}^{\mathrm{H}} BM(t)B^{\mathrm{H}}\mathbf{w}\mathbf{w}^{\mathrm{H}} BM^{\mathrm{H}}(t)B^{\mathrm{H}}\mathbf{s} \tag{2.21}$$

Setting the matrix B in eq. (2.21) equal to the identity matrix we obtain the array factor for system (B),

$$|g_a(q, f, t)|^2 = \mathbf{s}^{\mathrm{H}} M(t)\mathbf{w}\mathbf{w}^{\mathrm{H}} M^{\mathrm{H}}(t)\mathbf{s} \tag{2.22}$$

and finally, by setting the array errors equal to zero in eq. (2.22), $a_k = b_k = g_k = 0$, we obtain the array factor for the ideal antenna array, (A) in Figure 1,

$$|g_a(q, f, t)|^2 = \mathbf{s}^{\mathrm{H}}\mathbf{w}\mathbf{w}^{\mathrm{H}}\mathbf{s}\, y(t)y^*(t) \tag{2.23}$$

From eq. (2.23) one can see that for the ideal antenna array both F and P_{rad} on the right-hand side of eq. (2.1) will have the same time-dependence, and hence the IAD will be time-independent. This is also true for system B and C, as long as we only have amplitude and phase errors. However, with time errors present, the IAD will in general be error-, data-, and time dependent. That is, the IAD will depend on the specific array errors that are present, it will depend on the specific chip sequence that is being transmitted, and it will also depend on the sampling time-position at the receiver.

2.2 Average directivity for a specific array

The time and data dependence in the directivity can be removed if the directivity is calculated as an average over a given type of chip sequence distribution, and as a time average over one chip period. In a spread-spectrum communication system the chip sequence will be pseudo-random, and one may therefore assume that the phase angles j_m are randomly picked from the allowed values in eq. (2.19). The average directivity for a *specific* antenna array with a *specific* set of array errors is defined as

$$\overline{D}(q, f) \equiv \overline{E_j\{D(q, f, t)\}} \tag{2.24}$$

where E_j denotes the expectation with respect to a random chip sequence.

Since $E_j\{D(q, f, t)\}$ is a periodic function in time, with a period time equal to the chip period T_c, the time average can be calculated as

$$\overline{E_j\{D(q, f, t)\}} = \int_{-T_c/2}^{+T_c/2} E_j\{D(q, f, t)\}dt \tag{2.25}$$

2.3 Ensemble average directivity

The exact values of the array errors are in general unknown, and in many cases they can only be treated in a statistical fashion. From a practical point of view it is therefore useful to define the directivity as an *ensemble average*, that is, as an average calculated over a hypothetical collection of antenna arrays. The array errors are then described by their statistical distributions. In the present work we assume that the errors are independent, zero mean normal variables with standard deviations s_a, s_b, and s_g, respectively.

The ensemble average azimuthal directivity, $\tilde{D}$, is defined as

$$\tilde{D}(q, f) \equiv E_{abg}\{\overline{D}(q, f)\} \tag{2.26}$$

where E_{abg} denotes the expectation with respect to the array error distributions.

3 Monte Carlo simulation results

In this section the ensemble average azimuthal directivity $\tilde{D}$ in eq. (2.26) is estimated using Monte-Carlo simulations. The directivity is estimated for the three model systems A, B and C in Figure 1, and for three different directions of the main lobe in the antenna pattern.

As an example we consider a linear array with $L = 4$ antenna elements with equal separation $d = l/2$, where l is the carrier wavelength. The intrinsic Butler matrix beams in system C are directed at angles

$$f_k = \arccos\left(\frac{l(L-2k+1)}{2dL}\right), \quad k = 1,2,\ldots L \qquad (3.1)$$

that is, in this case at angles 41.4°, 75.5°, 104.5° and 138.6°. In figures 3, 4 and 5 the envelope of the Butler matrix beams is shown as the uppermost, dotted curve. The pulse-shaping root-raised cosine filter $h(t)$ is described by a roll-on-roll-off factor $a = 0.22$, and the summation over index m on the right-hand side of eq. (2.15) is truncated at $|m| = 3$.

The magnitude of the array errors used in the simulation is chosen to be fairly realistic, and the numerical values are shown in Table 1 below.

The Monte-Carlo estimation of $\tilde{D}$ is straightforward. For a given target (main lobe) direction the weight vector $\mathbf{w}$ is first calculated using eq. (2.20). For each Monte-Carlo run n the matrix $M(t)$ in eq. (2.13) is then calculated for a random chip sequence j_m, $m = -3,-2,\ldots+3$, for a random set of normally distributed array errors, and at a random time t picked from a uniform distribution on the interval $[-T_c/2, +T_c/2]$. The instantaneous directivity D_n for system B and system C is after that calculated using eqs. (2.21-22) and (2.1-4). The above procedure is repeated $N = 1000$ times, and the ensemble directivity is finally estimated simply as

$$\tilde{D} = \frac{1}{N}\sum_{n=1}^{N} D_n \qquad (3.2)$$

Table 1. Standard deviations for the array errors.

Parameter	σ_α [-]	σ_β [degrees]	σ_γ [seconds]
Value	0.25	15	$0.1 \times T_c$

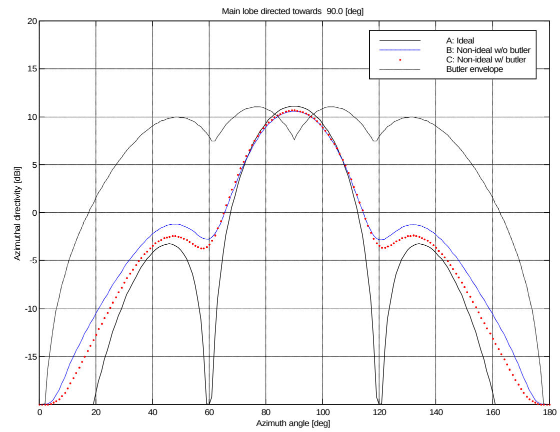


Figure 3. Main lobe directed at 90 degrees, between the second and the third Butler matrix lobe.

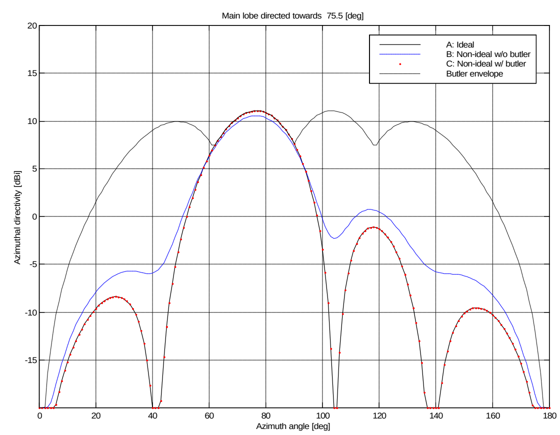


Figure 4. Main lobe directed at at 75.5 degrees, on top of the second Butler matrix lobe.

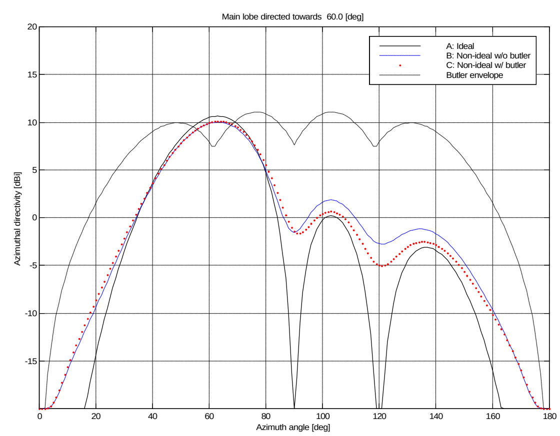


Figure 5. Main lobe directed at 60 degrees, between the first and the second Butler matrix lobe.

In figures 3-5 is shown the ensemble average directivity, estimated using eq. (3.2), and with the variance of the array errors given by Table 1.

In the first plot, Figure 3, the main lobe is directed straight ahead at 90 degrees, in the middle between the second and the third Butler matrix intrinsic lobe. In the figure one can see that the reduction in the main lobe level (MLL) is 0.55 dB for system B, and slightly less, 0.45 dB, for system C. However, the increase in the first side-lobe level (SLL) for system B - roughly 2 dB - is significantly higher than the corresponding value, for system C, 0.85 dB.

In Figure 4 the difference between a system with and without a Butler matrix is shown even more clearly. In this figure the main lobe is directed at 75.5 degrees, on top of the second intrinsic Butler matrix lobe. System C here is effectively immune to the array errors, and displays the same directivity pattern as the ideal array, system A. System B on the other hand shows the same behavior as in the previous figure, the MLL is decreased by 0.5 dB, and the first SLL is increased by 1.8 dB.

Finally, in Figure 5 the main lobe is again placed in the middle between two of the Butler matrix lobes, now at 60 degrees. The overall picture from Figure 3 is repeated for both systems. For system B the MLL is decreased by 0.6 dB, and the first SLL is increased by 1.6 dB. For system C the corresponding numbers are 0.5 dB for the MLL, and 0.4 dB for the first SLL.

Another way to illustrate the benefits of the Butler matrix in system C is shown in table 2, 3 and 4. In those tables we compare the first SLL for system B and C (both measured relative to the SLL for the ideal system) for different magnitudes of the amplitude-, phase-, and time errors. In each table one of the errors is set equal to zero, and the SLL is tabulated as a function of the standard deviation of the two other errors.

The simulations are done in a worst-case situation for system C, with the main lobe directed straight ahead, in the middle between two of the intrinsic Butler lobes.

For the beam direction and the error magnitudes used in table 2-4, the maximum difference in the MLL between system B and C is quite small, around 0.1 dB, and the MLL numbers are therefore not shown here.

As an example of how to read the tables, consider Table 4 on this page. For a phase error $s_b = 20$ degrees and a time error $s_g / T_c = 0.1$, the increase in the SLL is 2.0 dB for system B, and 0.5 dB for system C. That is, the presence of the Butler matrix has reduced the SLL for system C by 2.0 - 0.5 = 1.5 dB.

Table 2. Side-lobe level (SLL) increase in dB for system B and C as a function of amplitude and phase error magnitudes (no time errors present).

			Phase error σ_β [degrees]			
			0	10	20	30
Amplitude error σ_α [-]	0.0	B	0.0	0.4	1.2	2.4
		C	0.0	0.0	0.1	0.4
	0.1	B	0.2	0.4	1.3	2.4
		C	0.1	0.1	0.2	0.5
	0.2	B	0.5	0.7	1.5	2.7
		C	0.3	0.3	0.4	0.7
	0.3	B	1.0	1.3	2.0	3.1
		C	0.6	0.7	0.7	1.0*

Table 3. SLL increase in dB for system B and C as a function of amplitude and time error (no phase errors).

			Time error σ_γ / T_c [-]			
			0.0	0.1	0.2	0.3
Amplitude error σ_α [-]	0.0	B	0.0	1.0	2.1	2.9
		C	0.0	0.3	0.7	1.1*
	0.1	B	0.2	1.1	2.2	3.0
		C	0.1	0.3	0.8	1.2*
	0.2	B	0.5	1.3	2.5	3.2
		C	0.3	0.6	1.0	1.1*
	0.3	B	1.0	1.8	2.9	3.3
		C	0.6	1.0	1.4	1.7*

Table 4. SLL increase in dB for system B and C as a function of phase and time error (no amplitude errors).

			Time error σ_γ / T_c [-]			
			0.0	0.1	0.2	0.3
Phase error σ_β [deg]	0	B	0.0	1.0	2.1	2.9
		C	0.0	0.3	0.6	1.1*
	10	B	0.4	1.2	2.3	3.1
		C	0.0	0.3	0.8	1.2*
	20	B	1.2	2.0	2.8	3.5
		C	0.1	0.5	0.9*	1.3*
	30	B	2.4	3.0	3.8	4.2
		C	0.4	0.8*	1.2*	1.9*

(*) No distinct side-lobe peak found for C – stated value is system C directivity at system B peak position.

4 Summary and conclusions

In this paper we have shown that it is possible to reduce the effects of array excitation errors (amplitude, phase, and time errors) by placing a Butler matrix close to the radiating elements in an antenna array in a steered beam system. In the system with Butler matrix we introduce the errors in beam space, while they are introduced in element space for the system without Butler matrix.

Changes in directivity in the main beam direction as well as for the first sidelobe are studied in systems with and without Butler matrix. Performance for the system with a Butler matrix significantly depends on the main beam direction (whether it is pointed in the direction of a beam defined by the butler matrix or not) while the system without Butler matrix does not show this dependency.

When errors are introduced, the system with Butler matrix shows the largest performance reduction when the main beam is directed in between two intrinsic beams. The gain in directivity in the main beam direction for the system with Butler matrix is for this case very low, on the order of 0.1 dB, compared to the system without Butler matrix. For the side lobe level, the gain using a Butler matrix is more significant, on the order of 1-2 dB for the errors levels considered in this study.

In the most favorable direction for the system with Butler matrix, i.e., when the beam direction equals one of the intrinsic beam directions, the performance gains are significant for main beam directivity as well as for side lobe level.

In conclusion, a steered beam adaptive antenna system with a Butler matrix beamforming network offers potential gains. These can be used either to improve antenna pattern performance for a given error level, or to reduce the coherency requirements for a given pattern performance.

References

[1] B. D. Steinberg, *Principles of Aperture and Array System Design*, New York: Wiley, 1976, ch. 13, pp. 301-316.

[2] L. C. Godara, "*Applications of Antenna Arrays to Mobile Communications, Part I: Performance Improvement, Feasability, and System Considerations*", Proc. IEEE, vol. 85, no. 7, pp. 1031-1060, July 1997.

[3] L. C. Godara, "*Applications of Antenna Arrays to Mobile Communications, Part II: Beam-Forming and Direction-of-Arrival Considerations*", Proc. IEEE, vol. 85, no. 8, pp. 1195-1245, August 1997.

Dependence of Radar Emission Spectra on Measurement Bandwidth and Implications for Compliance with Emission Mask Criteria

Frank Sanders
Institute for Telecommunication Sciences

303.497.5727; fax 303.497.3680
fsanders@its.bldrdoc.gov

Abstract. *Radar transmitter emission criteria normally include the specification of frequency-dependent emission masks. These masks specify the amount by which unwanted radar emissions (both out-of-band and spurious) must be suppressed relative to the power levels emitted at the radars' fundamental frequencies. Compliance with emission masks is determined through measurements of emission spectra. The measured levels of radar unwanted emissions and fundamental-frequency emissions both vary as a function of measurement system bandwidth,* B_m*. But the variation with* B_m *differs between the unwanted emissions and the fundamental-frequency emissions. Moreover, the variation of unwanted emission levels varies as a function of frequency as well as* B_m*. This creates a problem for radar emission mask-compliance measurements.*

The National Telecommunications and Information Administration (NTIA) Institute for Telecommunication Sciences (ITS) has explored this problem by performing emission spectrum measurements on a maritime surface search (navigation) radar. In the spectrum data that are presented, the radar unwanted emission levels are found to vary between 12 log(B_m*) and 20 log(*B_m*), depending upon frequency. But the measured power of the radar fundamental frequency is found to vary as 20 log(*B_m*) for all bandwidths that are less than or equal to 1/(radar pulse width). The result is that the level offset between the unwanted emissions and the fundamental-frequency emission level depends upon the measurement bandwidth and the frequency of the unwanted emissions. This result implies, at a minimum, that measurement personnel must take the effect of* B_m *into account when performing radar emission spectrum measurements for the purpose of determining emission mask compliance. Based upon the results of these maritime radar spectrum data, some technical strategies for measuring radar emission spectra for emission mask compliance are proposed. Possible technical implications for future development of radar emission masks are also discussed.*

1. Introduction

Pulsed radar transmitters produce broadband spurious emissions at relatively high levels compared to emissions from narrowband communication systems. As documented in [1], radar spurious emission levels are typically tens of decibels higher than theoretically predicted sinc^2 spectra of pulsed emissions. These emissions in the frequency domain probably correspond to short-term transient behavior in the time domain characteristics of the rising and falling edges of the radar pulses.

For spectrum management purposes, radar out-of-band and spurious emissions are limited in the U.S. by Government regulations such as the radar spectrum emission criteria (RSEC). RSEC spurious emission masks are specified in terms of amplitude suppression relative to the power produced at radars' fundamental frequencies [2]. (E.g., suppression might be required to be at least 60 dB below the fundamental at frequencies of 100 MHz or more away from the fundamental.) Mask-compliance limits are computed on the basis of a theoretically perfect pulsed emission plus a margin that allows for realistic performance of an economical transmitter design.

Compliance with emission masks is determined through measurements of emission spectra. The measured levels of radar unwanted emissions and fundamental-frequency emissions both vary as a function of measurement system bandwidth, B_m. But the variation with B_m differs between unwanted emissions and the fundamental-frequency emissions. Moreover, the variation of unwanted emission levels varies as a function of frequency as well as B_m. This presents a problem for radar emission mask-compliance measurements.

In a peak-detected power measurement, completely non-coherent (noise-like) emissions should vary as $10 \log(B_m)$, while coherent emissions should vary as $20 \log(B_m)$. Since the bandwidth progression coefficients of 10 and 20 represent limiting cases, radar spurious emissions should have a coefficient somewhere between these two values. But no study is known to have been undertaken to determine the actual coefficient for radars.

The National Telecommunications and Information Administration (NTIA) Institute for Telecommunication Sciences (ITS) has begun to address this problem by performing emission spectrum measurements on a maritime surface search (navigation) radar. The purpose of the measurements has been to determine the actual variation of spurious emission levels of a radar relative to the measured fundamental level as a function of the measurement bandwidth, B_m. With this value known, NTIA will be better able to specify appropriate bandwidths for RSEC-compliance measurements, as well as any post-measurement correction factors that might be required.

2. Approach

A magnetron-based X-band maritime surface search radar was set up with an antenna height of approximately 4 m on a mast at a prairie location free of local obstructions. A measurement system contained in the RF-shielded enclosure of the NTIA Radio Spectrum Measurement System (RSMS) was positioned at a distance of 105 m from the radar. The RSMS received radar emissions via a 1-meter diameter parabolic dish antenna with a linear, matched-polarization log-periodic feed. Figure 1 shows the radar transmitter and the RSMS during the measurement.

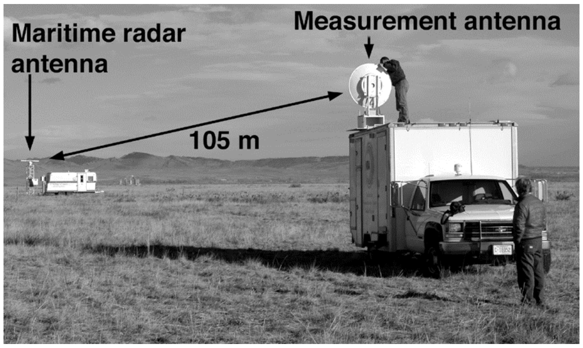


Figure 1. Maritime radar and measurement system.

The measurement system is shown in block-diagram form in Figure 2. Following the antenna, a variable (0-70 dB) attenuator was adjusted on a frequency-dependent basis to keep the received signal level from the radar within the dynamic range of the measurement system. That is, the attenuation level was zero when radar emissions were close to measurement system noise, but was gradually increased (in 10-dB steps) to as much as 70 dB as the measured frequency approached the radar fundamental. The attenuation was gradually decreased at frequencies above the radar fundamental, being finally reduced to zero at the upper end of the measured spectrum.

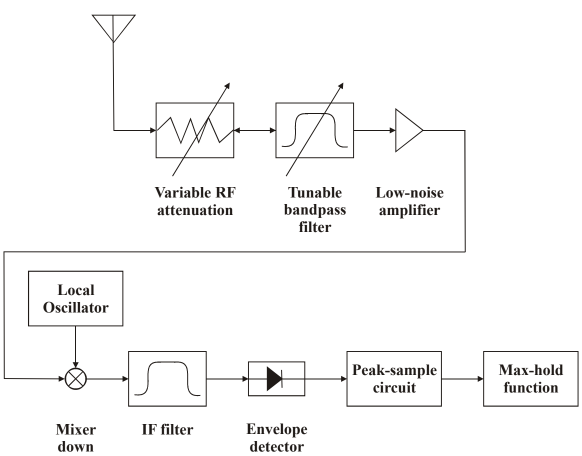


Figure 2. Measurement system functional block diagram.

Following the attenuator, a tunable bandpass filter based on yttrium-iron-garnet (YIG) technology isolated the system's broadband low-noise filter (LNA) from the high-power radar fundamental energy as the spectrum measurement progressed. The LNA provided the sensitivity required to measure out-of-band and spurious emission levels as much as 100 dB below the measured fundamental power level.

The LNA output was fed into a spectrum analyzer. The critical spectrum analyzer stages are shown in Figure 2, including frequency downconversion; intermediate frequency (IF) filtering; envelope and peak detection; analog-to-digital conversion; and final output to a data-recording computer.

As noted above, the requisite dynamic range of the measurement (about 100 dB) demanded that the RF attenuator setting be varied as a function of frequency. Therefore, the measurement system could not be operated in a swept-frequency mode. Instead, the measurement system was fixed-tuned to a single frequency with a single attenuator setting. A peak-detector circuit was operated at the single frequency for a period of time (3 sec) slightly in excess of the radar's antenna rotation period (2.7 sec). This ensured that the radar beam maximum output would be sampled at each measured frequency.

With the emission amplitude measured at a single frequency, the measurement system was tuned to another frequency, the RF attenuator was adjusted (if necessary) and the measurement process was repeated for another 3 sec. This stepped-frequency measurement process was used to acquire all data presented in this paper.

With 100 dB dynamic range available in the measurement, the radar spectrum was measurable across 4.2 GHz of spectrum, from 7300 MHz to 11500 MHz. Because the out-of-band and spurious emissions are a continuum, lacking discrete carrier frequencies, those emissions could be sampled at arbitrary spacings between frequency steps. For this measurement, the step interval was set at 6 MHz, for a total of 700 steps (and 701 measured frequencies) across the range 7300-11500 MHz. Each spectrum run therefore required about 35 minutes for completion.

Because the 6-MHz step spacing could result in the omission of the radar fundamental frequency from the measured spectrum, a supplemental measurement was made at the radar fundamental for each spectrum. This ensured that radar peak power was accurately included in each spectrum output.

The radar can transmit multiple pulse widths, ranging from 80 nsec to 800 nsec. The radar spectrum was measured in two pulse modes (80 nsec and 800 nsec) in four IF bandwidths (where B_m=300 kHz, 1 MHz, 3 MHz, and 8 MHz). Thus a total of eight spectrum measurements were performed.

3. Results

The measured emission spectra for short-pulse and long-pulse radar modes are shown in Figures 3 and 4. Measurement system internal noise occurs as a flat floor at frequencies between 7300-8000 MHz, and as a flat area between 11000-11200 MHz. It is observed that the spurious emission levels are changing with a progression that is somewhere between 10 log and 20 log bandwidth, depending upon frequency.

Figure 5 shows measured power at the radar fundamental frequency (approximately 9410 MHz) as a function of measurement IF bandwidth. The data in this figure confirm that the radar fundamental power follows a 20 log(B_m) rule for values of B_m that are equal to or less than (1/pulse width).

In the short-pulse mode, the 80-nsec pulse width results in a predicted fundamental-frequency 3-dB emission bandwidth of 12.5 MHz, which is wider than the maximum measurement IF bandwidth of 8 MHz. Consequently, the measured power increases as 20 log(B_m) for all points. But in the long-pulse mode of 800-nsec pulse width, the 3-dB emission bandwidth is predicted to be 1.25 MHz. As a result, the 20 log(B_m) progression breaks down for 3 MHz and 8 MHz B_m values.

Figures 6a through 6f quantify the decibel differences between spectra measured in successive values of B_m. These are shown as deviations from a 20 log progression, computed as follows:

$$\Delta = \left[\frac{P_x - P_y}{\log\left(B_x / B_y\right)} \right] - 20 \qquad \text{(unitless)}$$

where:

Δ = deviation from 20 log(B_m) progression;
$P_{[x,y]}$ = log power measured in B_x and B_y;
$B_{[x,y]}$ = measurement bandwidth;
[x,y] are subscripts for successive measurement IF bandwidths (e.g., 3 MHz and 1 MHz).

The differences should lie between –10 (corresponding to a noise-like 10 log progression) and 0 (for the 20 log rule that coherent emissions should follow). Some points lie outside these bounds. Deviations outside this range could result from the interaction of spectrum peaks and valleys with a particular measurement bandwidth; uncertainty in measured values; or variation in shape factors between filters.

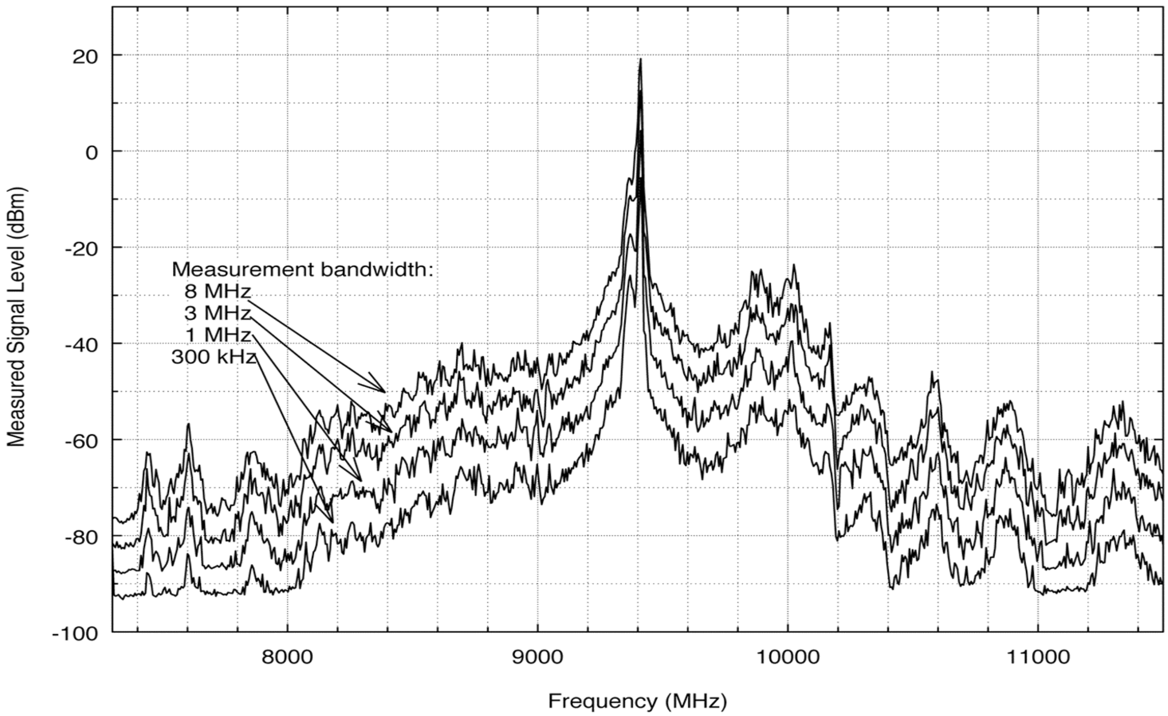


Figure 3. Maritime radar emission spectrum measured in four bandwidths with transmitter operating in short-pulse mode.

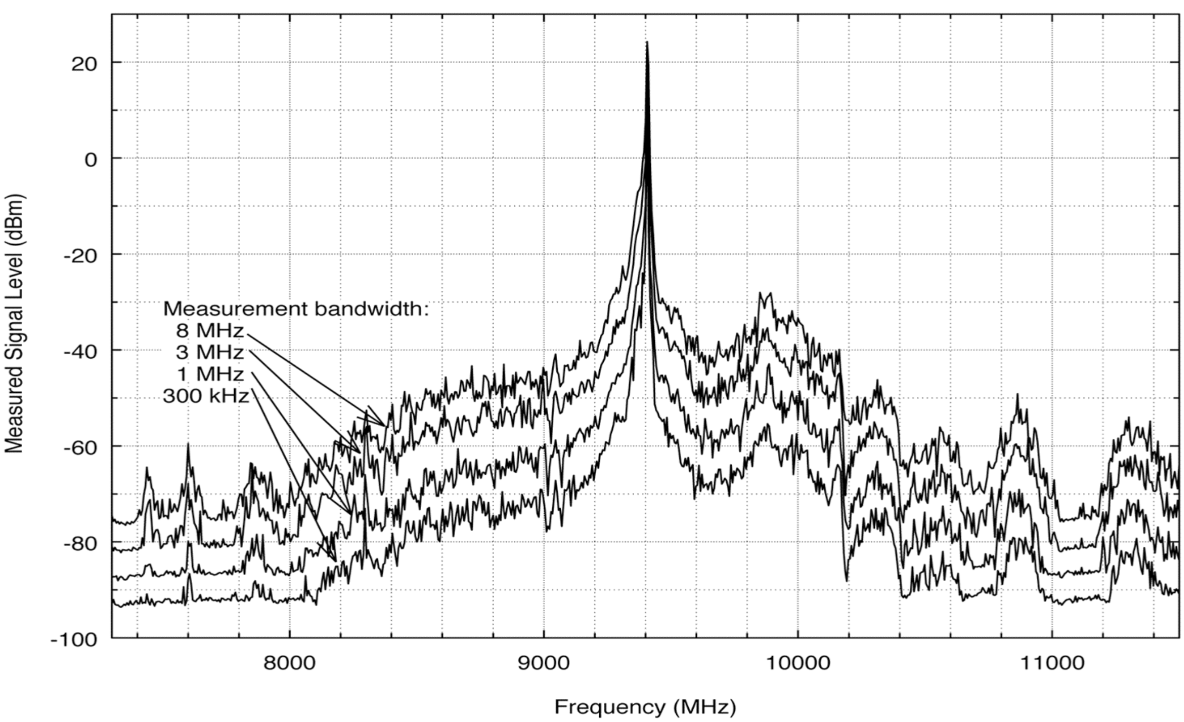


Figure 4. Maritime radar emission spectrum measured in four bandwidths with transmitter operating in long-pulse mode.

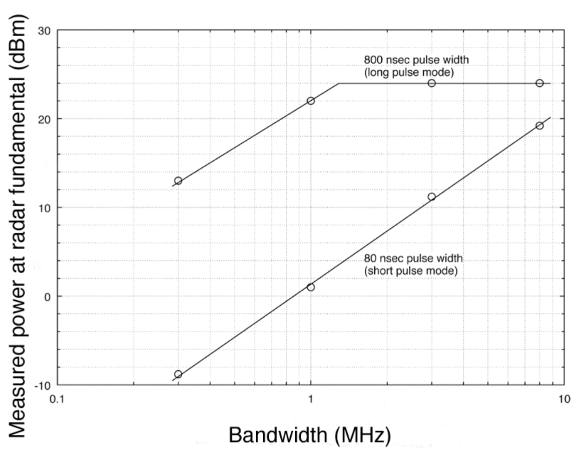


Figure 5. Variation in measured power at the radar fundamental as a function of measurement bandwidth and pulse mode.

Some trends are clearly observed in the values for the curves in Figure 6. The values of the deviation curves typically range between −4 and −2, corresponding to variation coefficients between 16 and 18. Neglecting the points that exceed zero, the extreme values range between −8 and 0, corresponding to variation coefficients between 12 and 20, respectively.

4. Summary and Conclusions

In summary, spurious and out-of-band emissions measured from a maritime navigation radar were found to typically vary at a value between $16 \log(B_m)$ and $18 \log(B_m)$. Extreme values ranged from a low of $12 \log(B_m)$ to a high of $20 \log(B_m)$.

These results indicate that the radar spurious emissions did not vary as would be predicted for thermal noise (10 log progression). Thus radar spurious emissions should not be characterized as "noise-like," at least for the radar measured in this study.

Radar emission spectrum measurements that are performed for the purpose of determining the suppression of out-of-band and spurious emissions relative to radar fundamental-frequency power will show a bandwidth dependent variation. Measurement of spectra in multiple bandwidths is recommended until this phenomenon is better understood.

It may be desirable for future emission masks to accommodate this effect. Such masks might include a recommended IF measurement bandwidth, and possibly a method for computing correction factors for measurements made in other bandwidths.

5. References

[1] R.J. Matheson, J.D. Smilley, G.D. Falcon, V.S. Lawrence, "Output tube emission characteristics of operational radars," NTIA Report 82-92, Jan. 1982.

[2] *Manual of Regulations and Procedures for Federal Radio Frequency Management*, revised Jan. 2000, Revisions Jan./May/Sep. 2001, NTIA Office of Spectrum Management, U.S. Government Printing Office, Stock No. 903-008-00000-8.

Acknowledgments

The author wishes to thank the following individuals for their valuable assistance: Richard N. Statz of ITS for software development; Bradley J. Ramsey of RF Metrics for assistance in getting the field measurement system operational and in data analysis; and the technical and editorial reviewers who commented on this paper on short notice.

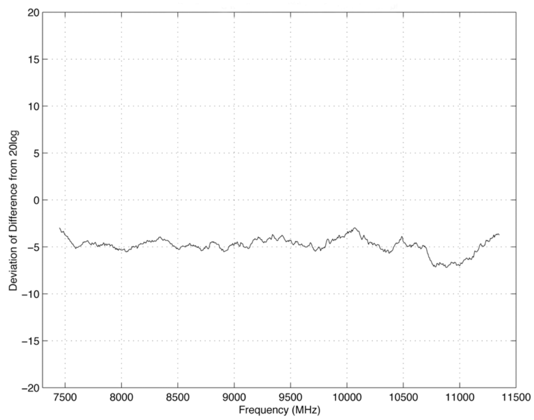


Figure 6a. 8 MHz – 3 MHz smoothed difference, short pulse radar mode.

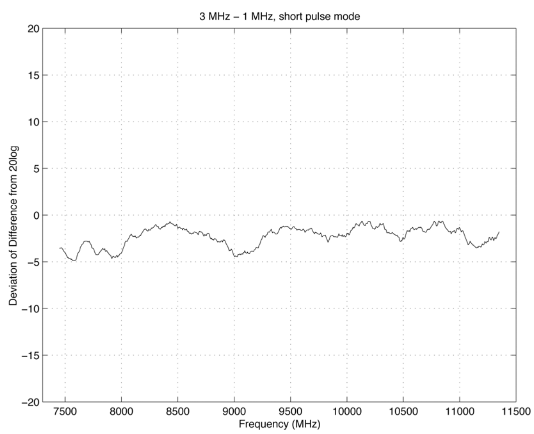


Figure 6b. 3 MHz – 1 MHz smoothed difference, short pulse radar mode.

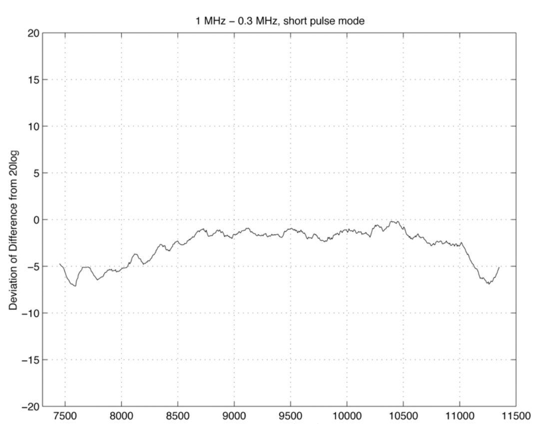


Figure 6c. 1 MHz – 300 kHz smoothed difference, short pulse radar mode.

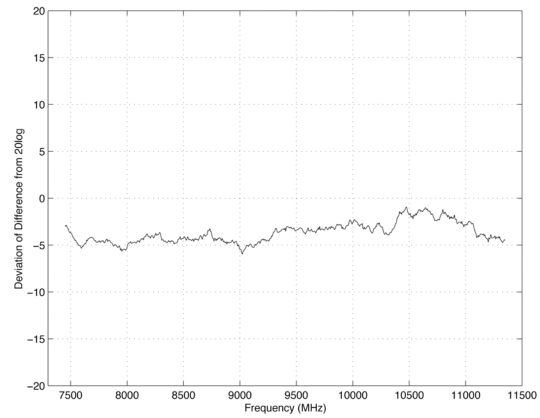


Figure 6d. 8 MHz – 3 MHz smoothed difference, long pulse radar mode.

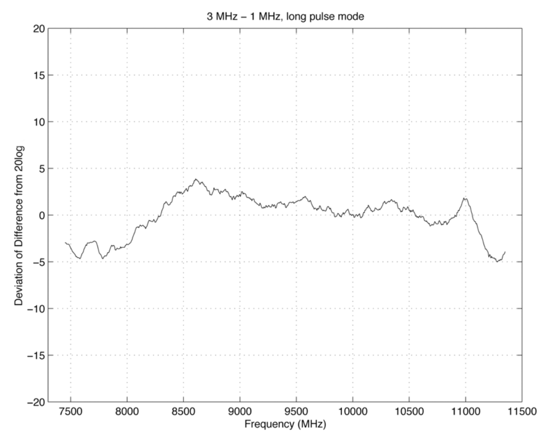


Figure 6e. 3 MHz – 1 MHz smoothed difference, long pulse radar mode.

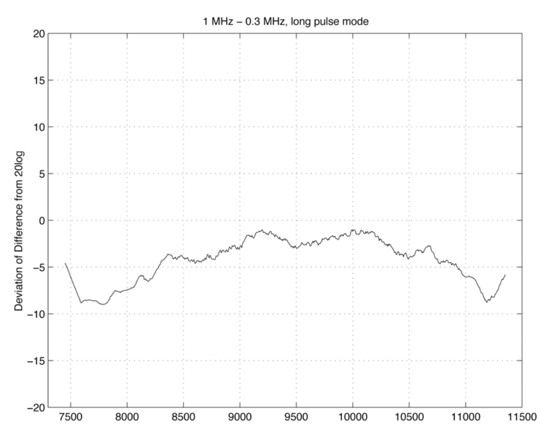


Figure 6f. 1 MHz – 300 kHz smoothed difference, long pulse radar mode.

Comparison and Simulation of Digital Modulation Recognition Algorithms

Wei Su
U.S. Army CECOM RDEC
Intelligence and Information Warfare Directorate
AMSEL-RD-IW-I
Fort Monmouth, NJ 07703
1-732-427-6332
wei.su@us.army.mil

John Kosinski
U.S. Army CECOM RDEC
Intelligence and Information Warfare Directorate
AMSEL-RD-IW-I
Fort Monmouth, NJ 07703
1-732-427-5605
john.kosinski@us.army.mil

Abstract - Modulation recognition has been studied for decades with numerous amounts of papers published. Modulation classifiers are developed under assumptions and may not be robust in some applications. This paper studies and simulates several popular digital modulation recognition methods and discusses the pros and cons of those algorithms based on the robustness of the modulation feature and algorithm fundamentals.

Index Terms – electronic warfare, communication, modulation recognition, signal classification, algorithm comparison, modulation feature.

I. INTRODUCTION

Modulation recognition is an important subject not only in commercial application but also in classifying emitter types for military electronic support cases [1-4]. Modulation recognition is a non-cooperative communication practice which, in general, starts with signal processing to remove center frequency, re-sample the signal, synchronize baud rate and carrier phase, and equalize the channel distortion. It is followed by modulation feature extraction to obtain unique information related to amplitude, phase, and frequency. Then, feature recognition is applied by using logic analysis to match features to known templates, or using statistical analysis to find a solution based on probabilities. Therefore, based on the techniques being used, a successful modulation recognition technique may depend not only on factors such as: signal bandwidth, available signal length, digitization method, number of samples per symbol, modulation types, transmission environment, signal noise ratio, frequency stability, processing power, processing time, implementation cost, and dimension of confusion matrix, but also on qualities of baud rate estimation, pulse synchronization, pulse re-sampling, carrier synchronization, etc. Many publications assume to have perfect knowledge of center frequency, baud rate, and pulse shape so that a fair comparison of different algorithms becomes a challenge. To compare algorithms based solely on signal to noise ratio (SNR) and probability of success may have little value. Fundamental similarities and differences between the algorithmic bases should be explored.

II. ALGORITHM COMPARISON

The modulation recognition includes converting the analog RF signal to a digital IF signal, extracting modulation features, and recognizing modulation types. Some classifiers can extract modulation features directly from an IF signal. But, in most cases, a coarse estimation is needed to convert IF signal to I and Q components, and extract modulation features with the presence of pass-band signal residuals such as center frequency offset (CFO) or timing errors. Many classifiers extracts features by assuming to have perfect base-band symbols since those features are very sensitive to pass-band disturbances. The modulation recognition is conducted by searching the best match between modulation features and given templates. The result of modulation recognition could be a confusion-matrix, which is a table of statistical values obtained with a specified signal-to-noise ratio (SNR). This table provides the values of probability-of-success in respect to a list of candidate modulation types. The result of modulation recognition could also be a set of curves representing candidate modulation types. Each curve gives the probability-of-success versus SNR for a given modulation type. Since modulation classifiers are developed under various assumptions and objectives, the algorithm comparison is nontrivial. Probability-of-success will be a performance measurement only if two classifiers are developed under the same assumptions. Probability-of-false-alarm should also be included to measure the failures that an unknown type is forced to a know template. Therefore, fundamental differences between the algorithms should be studied.

A. Phasor Variation Analysis Classifier

Phasor analysis approach utilizes the phase variation and amplitude variation as features in modulation recognition. In this approach, the IF signal is down-converted to base-band and the symbols are extracted for analysis. Many algorithms assume the unknown signal is transmitted through an ideal channel so that the white Gaussian AWGN is the only concern in estimation error, some of the algorithms, such as Azzouz and Nandi's modulation classifier [5,6], also assume a perfect recovery of signal symbols. Azzouz and Nandi's classifier recognize both analog and digital modulation types. The digital modulation includes: ASK2, ASK4, FSK2, FSK4, PSK2, and PSK4, Figure 1. The standard-deviation of modulation parameters: nonlinear component of the phase, the absolute value of the nonlinear component of the phase, the absolute value of the normalized-centered instantaneous amplitude, and the absolute value of the normalized-centered frequency are used as features. These standard deviation features are used together with the amplitude power spectrum density in a logic flowchart to match the predetermined feature thresholds to determine the modulation type. Reference [5] is a good tutorial for this type of general modulation recognition. However, the variance of the nonlinear component of the phase and the variance of the absolute value of the nonlinear component of the phase require the removal of the center frequency and linear phase components from the carried signal. Since most communication systems employ some type of filtering prior to transmission for shaping the signal for bandwidth efficiency, as shown in Figure 2, the pulse function of a PSK signal will have a smooth transition and the pulse shape will not be as rectangular as shown by the dashed-line of Figure 2. The linear component of the phase will not be removed easily. In this case, the performance of phase variation and absolute phase variation tests will be failed.

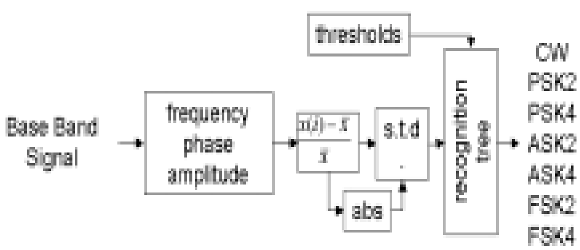


Figure 1. Feature Variation Analysis Modulation Classifier

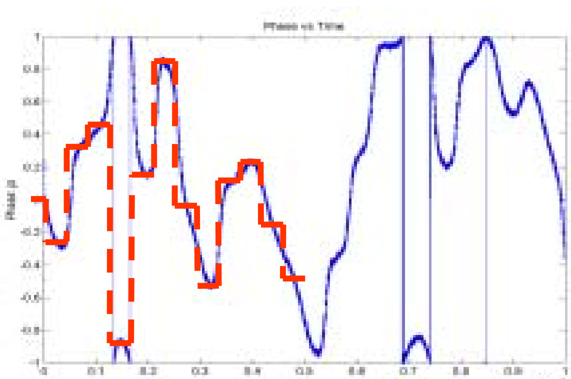


Figure 2 Phase Plot of A PSK Signal

Furthermore, if the center frequency is not removed perfectly, the phase samples $\phi(i)=\phi(i-1)+2\pi\frac{f_r}{f_s}$ will drift in time as shown in Figures 3 for a BPSK modulation and Figure 8 for a PSK8 modulation, where f_s is the sampling frequency and f_r is the center frequency offset. Carrier timing is another issue in this approach, if one of the PSK phase states is located close to the constellation point π as shown in Figure 4, the phase plot may wrap between -π and π, as shown in Figure 5, due to the random noise. This wrap effect may generate large phase variations to fail the standard deviation test.

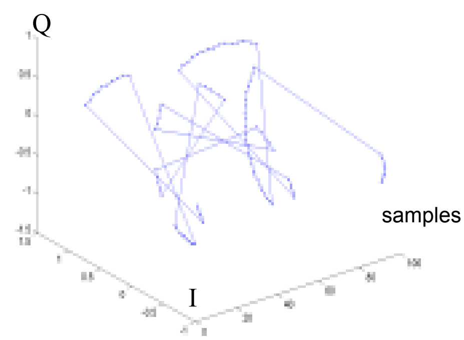


Figure 3 Phase Plot of A BPSK Signal with CFO

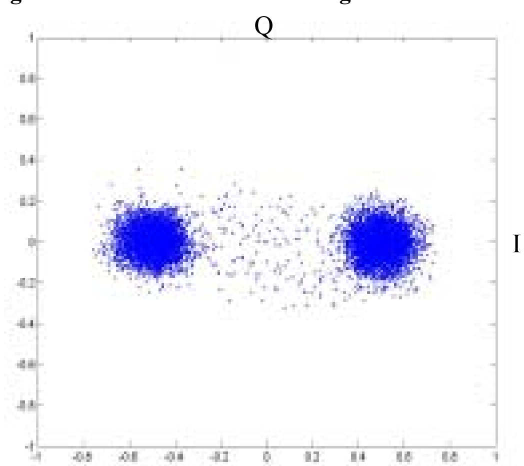


Figure 4 Constellation Diagram of BPSK

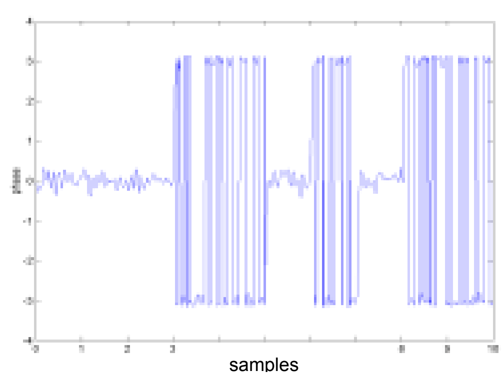


Figure 5 Phase Plot of A BPSK Signal with Phase Wrap

This approach may be extended [7] by adding a frequency estimator, a symbol phase tracker, and a timing device as shown in Figure 6. The frequency estimator provides the instantaneous center frequency of the IF signal. The frequency variance can be used to determine if it is a single tone modulation type. If the frequency variance is large, the standard deviations of the estimated frequency and the absolute value of the frequency are sent to recognition block. Otherwise, the signal will be down modulated to base-band using the estimated center frequency. A timing-recovery circuit is applied to extract amplitude, phase, and frequency symbols. Since the carrier estimated by the frequency estimator may not be accurate and the center frequency may be unstable, a phase tracking and correction block has to be used to remove the residual carrier frequency and prevent phase warp. Figure 7 illustrates that the phase drift of the PSK2 signal in Figure 3 is corrected by using a blind carrier phase tracking algorithm [8].

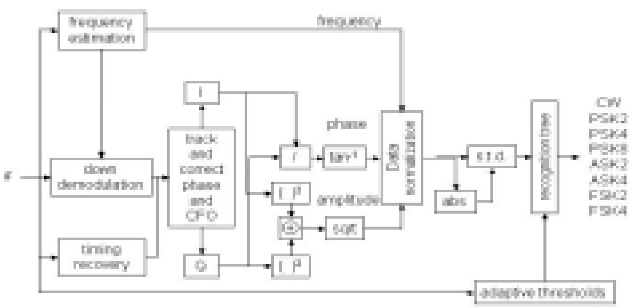


Figure 6. Modified Feature Variation Analysis Classifier

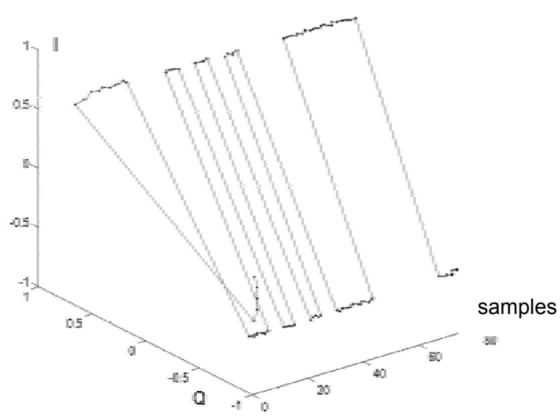


Figure 7. Corrected Phase Plot of BPSK with CFO

To solve the center frequency offset problem, Liedtke [9] adapted delta-phase (the phase difference between two adjacent phase symbols) as a feather rather than phase itself. Figure 8 illustrates the constellation diagram of a PSK8 signal with center frequency offset.

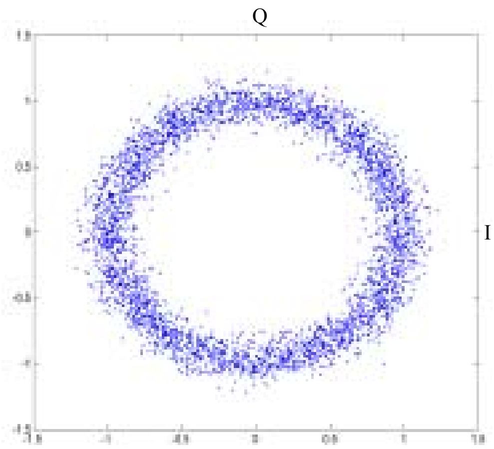


Figure 8. Phase Constellation Diagram of PSK8

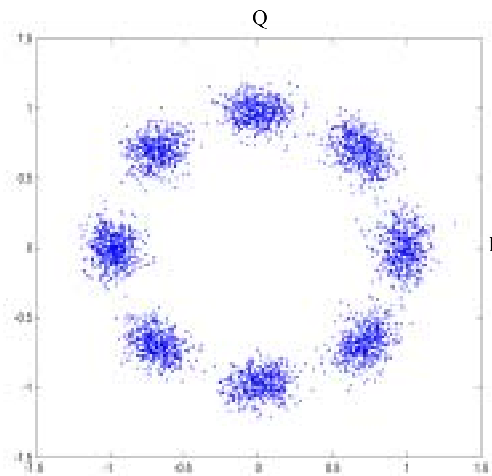


Figure 9. Delta-Phase Constellation Diagram of PSK8

The PSK8 phase feature cannot be observed due to the phase rotation caused by center frequency offset. Figure 9 shows the same signal plotted with delta phase. The clusters of PSK8 are recovered. Liedtke recognizes modulation types: ASK2, FSK2, PSK2, PSK4, PSK8 and CW, Figure 10, by utilizing the histograms of delta-phase. This algorithm only assumes to roughly know the center frequency of the signal and symbol rate. The signal is converted to almost base-band with the residual center frequency by using a concentric FIR filter bank centered at the center frequency. A symbol recovery circuit is employed to extract a sinusoidal waveform with an appropriately tuned narrow band pass filter centered at the symbol rate. This waveform is used to recover the PSK and FSK symbols. Then, the symbol amplitudes and the delta-phase are obtained. The frequency measurement is conducted by taking the phase difference of two time samples. The amplitude variance and frequency variance are used as features for separating modulation types among PSK, ASK, and FSK. The delta-phase histogram is used for probability density analysis to separate PSK2, PSK4, and PSK8. An addition-only sub-optimal histogram separation process is used to recognize PSK modulation types automatically. The advantage of using the delta-phase as a feature is that the center frequency offset will be eliminated in PSK recognition. Figures 11 and 12 show the deference between phase sample histogram and the delta-phase symbol histogram of a DQPSK signal. Furthermore, with the timing recovery circuit, all feature parameters are observed at the Nyquist sampling rate, and pulse shaping will no longer significantly affect the recognition result.

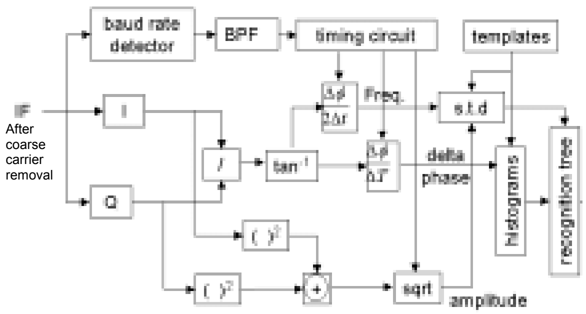


Figure 10. Universal Demodulator

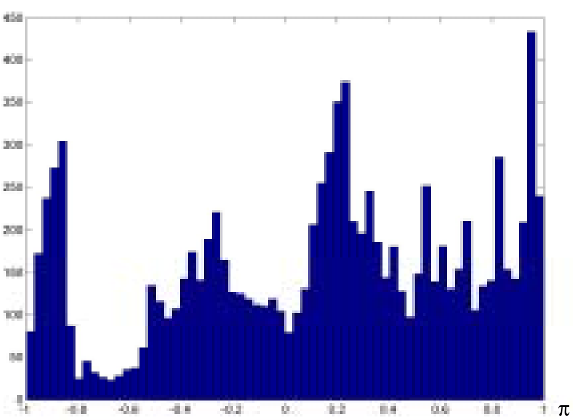


Figure 11 Histogram of Phase Samples of DQPSK

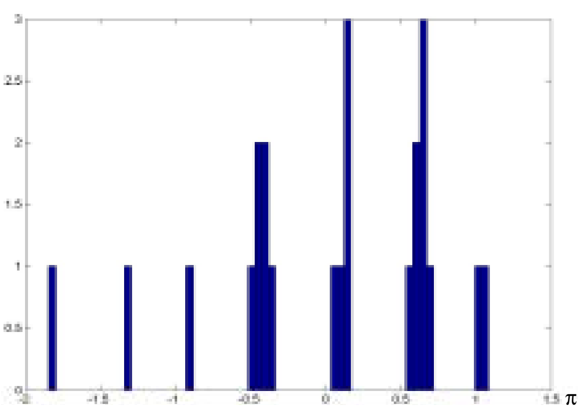


Figure 12 Histogram of Delta-Phase Symbols of DQPSK

This approach may be extended [10] by adding a frequency estimator and an automated timing-recovery device to replace the manual tuning. Since the delta-phase is used as the phase feature, the residual carrier frequency will be eliminated and the phase tracking and correction operation is not needed. If the frequency variance is larger than the threshold value, FSK modulation recognition will be considered. Otherwise, the signal will be down modulated and CW/PSK/ASK modulation recognition will be processed. The modulation feature recognition of the phasor variation analysis approach could be conducted by comparing the feature variances to thresholds [5, 6, 9, 25, 26]. Since the variance is quite sensitive to SNR, this method is better being used for top-level classification. Histograms analysis [9, 25, 26] is another frequently used technique and it is usually use for PSK modulation types. If the perfect symbol recovery is possible, maximum likelihood classifiers [11-21] will be an optimal approach for MPSK and QAM modulation recognitions. It may also be possible to treat constellation plot as image pixels so that many image recognition algorithms [22-24] can be useful in classifying the PSK and QAM modulation types.

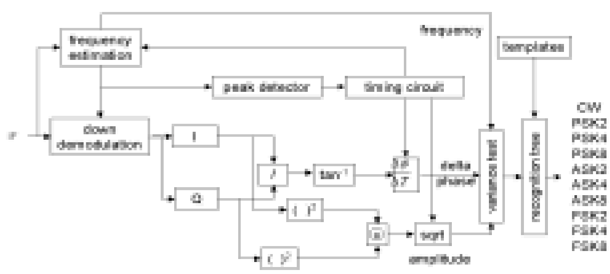


Figure 13 Modified Universal Demodulator

B. Zero-Crossing Modulation Classifier

Hsue and Soliman [25,26] introduced a modulation recognizer for PSK2, PSK4, PSK8, FSK2, FSK4, and FSK8 modulation types based only on the zero-crossing characteristics of signals as shown in Figure 14. The modulation classification procedure extracts the zero-crossing interval sequence for frequency estimation. The estimated frequency is then used together with the zero-crossing sequence for phase estimation. Similar to Liedtke's method, variances are used to separate FSK from single tone signals, and delta-phase histograms are used for parameter variation estimation of PSK signals. Variance thresholds, histogram templates, and likelihood ratio tests are employed for making the modulation decision. Frequency histogram is used for FSK recognition although the classification of a frequency histogram is not trivial. Unlike Liedtke's method, Hsue and Soliman use zero-crossing carrier estimation instead of manual tuning. Although the estimation is not accurate enough to provide a precise carrier frequency, the delta-phase approach will overcome the center frequency offset in PSK recognition. However, the accuracy of zero-crossing frequency estimation is very sensitive to the SNR as shown in Figures 15 and 16. When SNR is low, the noise may produce additional zero-crossing points. A good resolution of zero-crossing measurement requires a very high sampling rate as shown in Figure 15 for a BPSK signal, where the solid-line is has 80 samples per symbol and the dashed-line has only 4 samples per symbol. The low sample rate will reduce recognition performance greatly, but the high sample rate will pick up thermal noises. Since some data collecting devices requires only two samples per symbol, zero-crossing modulation classifier has to up-sample the data before processing. Our simulation also shows that the rising-edge/falling edge zero-crossing estimation is less noisy than the either-edge zero-crossing estimation as shown in Figure 16, where the smooth dashed-line stands for the frequency estimation of a FSK signal using rising-edge zero-crossing and the noisy solid-line stands for the frequency estimation using either-edge zero-crossing. Furthermore, zero-crossing frequency estimation may yield a single frequency result for an unknown signal with multiple instantaneous frequencies.

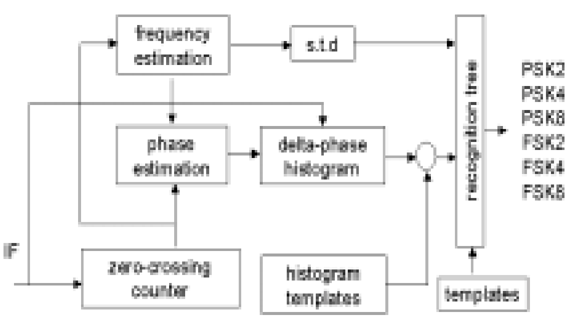


Figure 14. Zero-Crossing Modulation Classifier

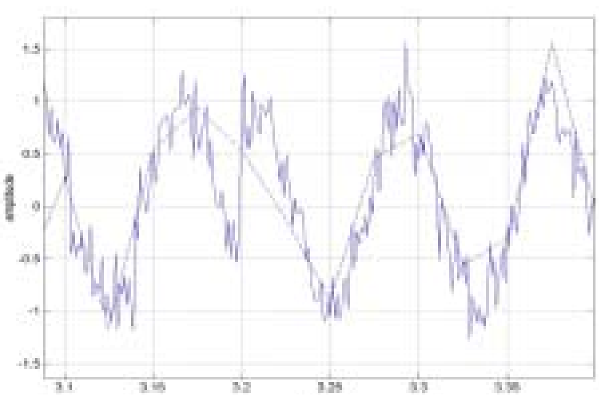


Figure15. BPSK signal with Two Different Sampling Rates

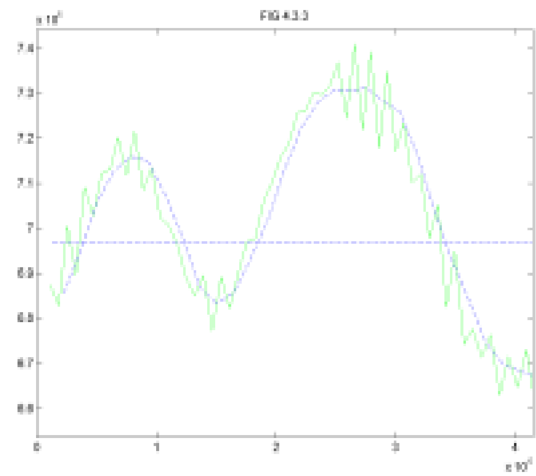


Figure 16 Zero-crossing Frequency Plot of A FSK Signal

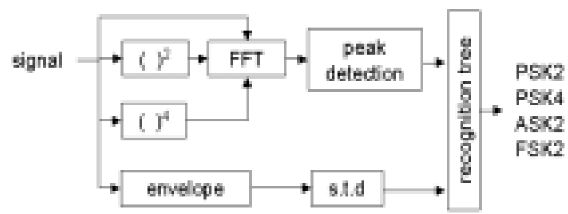


Figure 17. Power-Law Modulation Classifier

C. Power-Law Modulation Classifier

DeSimio and Glenn [27] introduced a technique to classify digital modulations: ASK2, PSK2, PSK4, and FSK2, as shown in Figure 17. In this recognizer a power-law classification is conducted to obtain the squared and the fourth power of a signal. Therefore, the key features are magnitude and location of the spectrum peaks in the frequency domain. That is, the magnitude of the spectral component at twice the carrier frequency of the signal raised to the second power, or the magnitude of the spectral component at four times the carrier frequency of the signal raised to the fourth power. For example, Figure 18 is the frequency spectrum of 100 symbols of a QPSK signal with center frequency f_c at 12 KHz, sampling frequency at 160 KHz, and SNR=10dB. Figure 19 is the spectrum of the same signal after taking the forth power. A peak of $4f_c$ is shown at 48 KHz, which indicates the QPSK modulation type. This peak will not be seen in Figure 18 or in the spectrum of the squared signal. The mean and variance of the signal envelope are used for ASK recognition. A decision tree is used to recognize modulation type based on the values of above features. The concept of power-law classification is based on the fact that squaring of an MPSK signal is another PSK with M/2 phase states. The power-law approach has the convenience of recognizing modulation type without converting the signal to base-band. Although band pass filtering of the spectrum peak requires knowledge of the center frequency, the exact carrier frequency and pulse shape are not necessary. However, this method is limited by the sampling rate. The signal must be sufficiently over-sampled (for example: 2X for PSK2 or 4X for PSK4) in order to satisfy the sampling theorem. Figure 20 shows the same signal sampled at a lower rate of 80 KHz. Since the sampling theorem limit the frequency range to 40 KHz, the peak at 48KHz will not be displayed at Figure 20. Although the image of $4f_c$ exists at 32 KHz, it is too noisy to be properly detected. The peak of $4f_c$ can be detected by correlating the spectrum of the signal with a $sinc^2(x)$ reference function [27], but a high resolution spectrum may be needed. Our simulation also shows that the detection of spectrum peak becomes more difficult if the pulse shaping is used. The capability in FSK modulation recognition is limited in this approach.

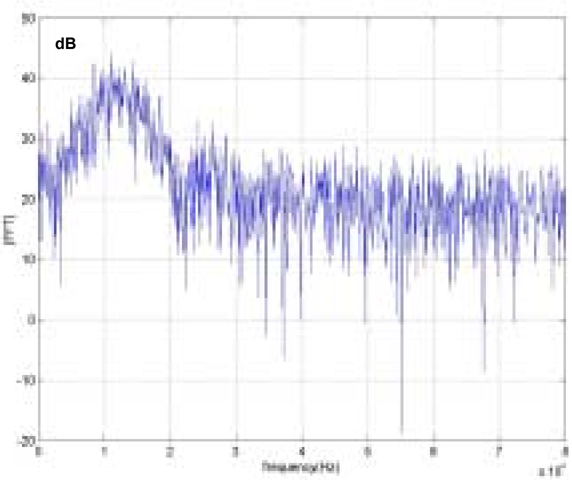


Figure 18 Frequency Spectrum of A QPSK Signal

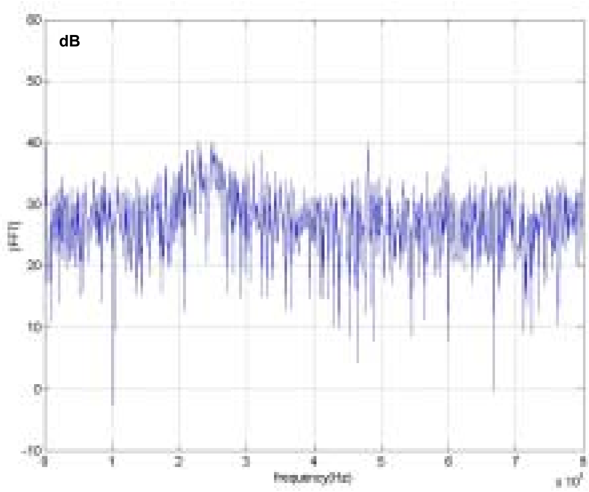


Figure 19 Spectrum of A QPSK Signal with Forth-Power

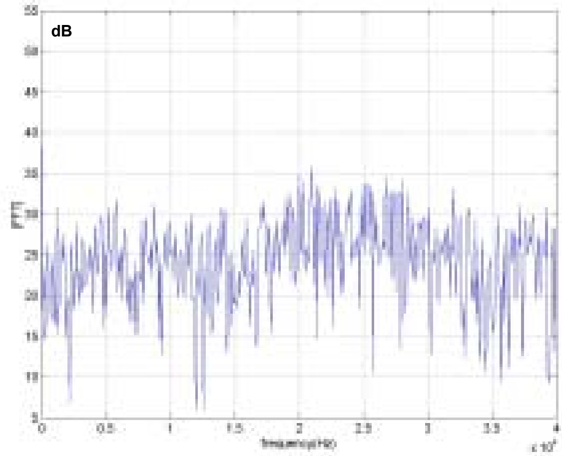


Figure 20 Under-Sampled QPSK Signal with Forth-Power

III. CONCLUSION

Some well-known modulation recognition algorithms are studied and simulated. The comparison of modulation recognition algorithms is not straightforward since algorithms are developed under different motivations and they are all good in solving certain problems. Our analysis is based on our application requirements, which may not apply to other cases. All signal plots used in this paper are simulated by MATLAB software and are not associated with any commercial or military devices/systems. Modulation classifiers may be sensitive to center frequency offset, and pulse shaping, symbol or carrier timing, and sampling frequency. The assumptions, limitations, and the fundamental similarities and differences between the algorithms should be explored in algorithm comparison.

IV. REFERENCES

[1] W. Su, "Computing and simulation of optimal location for RF emitters over rough terrain," in *Proc. of the 46th Joint Electronic Warfare Conference*, May 2000.

[2] B. Abbe et al., "Use of unconventional battlefield systems for IO," in *Proc. of the 45th Annual Joint Electronic Warfare Conference*, May 1999.

[3] W. Su and T. Huyuh, "A novel method in automated map based intelligence support systems," in *Proc. of 1998 Joint Electronic Warfare Conference*, April 1998.

[4] T. Huynh and W. Su, "Analysis of a back-propagation approach for automated jammer deployment," in *Research and Development Technical Report*, CECOM-TR-99-1, April 1999.

[5] E. Azzouz and A. Nandi, “Automatic modulation recognition of communication signals,” *Kluwer Academic Publishers*, 1996.

[6] E. Azzouz and A. Nandi, “Automatic Identification of Digital Modulations,” *Signal Processing*, Vol. 47, No. 1, Nov. 1995, pp.55-69.

[7] W. Su and J. Kosinski, "Comparison and modification of automated communication modulation recognition methods," in Proc. of MILCOM’02, October 2002.

[8] Belouchrani and W. Ren, “Blind Carrier Phase Tracking with Guaranteed Global Convergence,” *IEE Trans*. on Signal Processing, Vol.45, No.7, July 1997.

[9] F. Liedtke, “Computer simulation of an automatic classification procedure for digitally modulated communications signals with unknown parameters,” *Signal Processin*g, Vol 6, No. 4, pp. 311-323, August 1984.

[10] W. Su and J. Kosinski, “Application of COTS algorithms to automated modulation recognition and signal classification method,” *2001 AOC Intelligence Electronic Warfare Conference*, August 2001.

[11] K. Kim and A. Polydoros, “Digital modulation classification: the BPSK versus QPSK case”, in *Proc. IEEE Conf.*, 1988, pp. 431-436.

[12] C. Y. Huang and A. Polydoros, “Likelihood methods for MPSK modulation classification”, *IEEE Trans. Commun*., vol. 43, No. 2/3/4, pp. 1493-1504, Feb./March/April 1995

[13] C. Long, K. Chugg and A. Polydoros, “Further results in likelihood classification of QAM signals,” in *Proc. MILCOM* ’94, Oct. 1994.

[14] K. M. Chugg, C. S. Long and A. Polydoros, “Combined likelihood power estimation and multiple hypothesis modulation classification”, in *Proc 29th Asilomar Conf.*, pp. 1137-1141, 1996

[15] J.A. Sills, “Maximum-likelihood modulation classification for PSK/QAM”, in *Proc. IEEE MILCOM*, vol.1, Oct. 1994, pp. 57-61.

[16] W. Wei and J. M. Mendel, “Maximum-likelihood classification for digital amplitude-phase modulations”, *IEEE Trans. Commun*., vol. 48, No. 2, pp. 189-193, Feb. 2000.

[17] P. C. Sapiano and J. D. Martin, ”Maximum likelihood PSK classifier”, in *Proc. IEEE ICASSP*, 1996, pp. 1010-1014.

[18] D. Boiteau and C. Le Martret, ”A generalized maximum likelihood framework for modulation classification”, in *Proc. IEEE ICASSP*, vol. 4, pp. 2165-2168, May 1998.

[19] P. Panagiotou, A. Anastastasoupoulos and A. Polydoros, ”Likelihood ratio tests for modulation classification”, in *Proc. IEEE*, 2000, pp. 670- 674.

[20] A. Polydoros and K. Kim, “ On the detection and classification of quadrature digital modulations in broad-band Noise”, *IEEE Trans. Commun*., vol. 38, No. 8, pp. 1199-1211, Aug. 1990.

[21] M. Moeneclaey and G. D. Jonghe, “ML-Oriented NDA carrier synchronization for general rotationally symmetric signal constellations,” IEEE Transactions on communications, Vol 42, No.8, pp. 2531-2533, August 1994.

[22] Y. Shi et al., “A Shift-Invariant Modified Kohonen Neural Network,” in *Proc. of 2000 International Conference on Signal Processing Application and Technology*, October 2000.

[23] W. Su, “Method and Apparatus to Process Drawing Images,” *US Patent* Number 5,946,415, August 31, 1999.

[24] W. Su, “A Simple and Quick Approach in Processing Drawing Images,“ in *Proc. of 1995 International Conference on Signal Processing Applications and Technology,* August 1995.

[25] S-Z. Hsue and Samir S. Soliman, “Automatic Modulation Recognition of Digitally Modulated Signals,” in *Proc. of MILCOM’89*, 1989.

[26] S-Z. Hsue and Samir S. Soliman, “Automatic modulation classification using zero-crossing,” *IEE Proceedings*, vol.137, Pt. F, No.6, pp.459-464, December 1990.

[27] M. DeSimio, E. Glenn, “Adaptive Generation of Decision Functions for Classification of Digitally Modulated Signals,” in *Proc. Of NAECON’88*, pp. 1010-1014.

Evaluation of Two Site-Specific Radio Propagation Models

Alakananda Paul, Paul McKenna and Frederick Najmy
Office of Spectrum Management | Institute for Telecommunication Sciences | Office of Spectrum Management
National Telecommunications and Information Administration
U.S. Department of Commerce
Phone: (202) 482-0110, Fax: (202) 482-4595, Email: apaul@ntia.doc.gov

Abstract: *This paper discusses evaluation of site-specific propagation models used in the VHF and UHF range of frequencies which are needed for prediction of coverage and interference, especially for wireless communication applications. It describes two ongoing tasks, one at the National Telecommunications and Information Administration (NTIA) and the other at the International Telecommunication Union (ITU). In the United States, two major deterministic site-specific propagation models have been used for a long time: the Terrain Integrated Rough Earth Model (TIREM) developed for the Joint Spectrum Center (JSC) of the Department of Defense, and the Irregular Terrain Model (ITM) developed by the Institute for Telecommunication Sciences (ITS) of NTIA. About two years ago, the Office of Spectrum Management (OSM) of NTIA started a task for comparison and harmonization of the two models (TIREM and ITM). Both ITS and OSM are working on this task in cooperation with JSC. Predicted propagation losses from both models have been compared with large numbers of measured data. The first order statistical results, such as mean prediction error and its standard deviation, are similar for the two models. However, errors for individual paths between the two models sometimes differ by 20 dB or more. Some of the results of the comparison task and possible explanations for the discrepancies are presented. At the ITU Radiocommunication Study Group 3 (ITU-R SG 3) on propagation, Working Party 3K decided to proceed with a Preliminary Draft New Recommendation (PDNR) on a method for path-specific propagation prediction. An outline for this document, developed in 2002, is also discussed.*

1. Introduction

Site-specific propagation models are often needed for more accurate predictions of coverage and/or interference for wireless communication systems than would be available from site-general models. Two major site-specific propagation models have been used in the United States for a long time. These are: the Terrain Integrated Rough Earth Model (TIREM), developed by Alion Science and Technology (formerly the Illinois Institute of Technology Research Institute (IITRI)) for the Joint Spectrum Center (JSC) of the Department of Defense (DOD); and the Irregular Terrain Model (ITM) developed at the Institute for Telecommunication Sciences (ITS) of the National Telecommunications and Information Administration (NTIA) of the U.S. Department of Commerce (DOC). Both of these models have been used at NTIA, and occasionally they have given different results for the same path and same physical and electrical input parameters.

ITM has an area or site-general prediction mode as well as a point-to-point (or site-specific) mode, whereas TIREM has only a point-to-point mode. Therefore, the comparison studies reported here utilized ITM in its point-to-point mode. Speaking very generally, each model is deterministic and subdivides the propagation path analysis into line-of-sight, diffraction, and troposcatter ranges, and each will calculate the dominant median excess loss contributions (i.e., relative to the free space loss) due to three possible mechanisms, line-of-sight (LOS), diffraction, and troposcatter, based on the path analysis. Both models' path analyses will obtain identical results for the terminals' radio horizon distances and the horizon elevation angles. The models differ in an important aspect in the path analysis, however, in that ITM uses it to infer the additional intermediate quantities of each terminal's effective height above its dominant reflecting plane and the interdecile range of the terrain elevations, or terrain irregularity parameter, while TIREM does not.

Both models use a two-ray approach, i.e., direct ray plus ground reflected ray, for the LOS path, though they compute the ground reflections differently. However, it appears that, when the effective heights are close to the antenna structural heights and polarization effects on the ground reflection coefficient are unimportant, both predictions are close. The same is true for the troposcatter losses, because the algorithms used in the models are similar. For diffraction paths whose lengths are less than the total smooth earth horizon distances, both models compute the loss to the radio horizon at each end of the path using the two-ray algorithm. However, the losses between the horizons are calculated differently. ITM computes the weighted combination of double knife-edge and smooth earth diffraction losses and adds a clutter loss factor based on the actual terrain. TIREM computes the loss based on the sum of diffraction losses that would result from multiple knife-edges along the path. In all of these cases, the effective antenna heights have a significant effect on ITM's predictions. Also, neither model currently supports the prediction of additional losses due to land-use/land-cover (LULC) variations along a given path.

One way to compare the models is to compare their predictions to the same measured radio propagation data. In the measurement datasets described in the next section, the excess loss relative to free space was either available or derivable, and this quantity was used for comparison to the models' predictions. In ITM, the (median) computed reference attenuation or excess (i.e., relative to free space) loss is given. In TIREM, this quantity is the difference between the calculated values of the median basic transmission loss and the free space loss. To predict losses for a given measurement, both models require the distance and terrain profile between the two terminals, heights and polarization of the antennas, frequency of operation, surface refractivity, and ground constants. TIREM also requires the atmospheric humidity.

Tasks were initiated at NTIA (the Office of Spectrum Management (OSM) and ITS) and JSC to compare the predictions given by the two models with the measured data for specific paths and to harmonize the models, if possible. This paper will discuss the results of this partially completed task.

Section 2 describes the results of the preliminary comparison of ITM and TIREM with several measurement data sets. Section 3 includes the comparison of the measured data and predicted losses using factor analysis and augmented data. Possible improvements to the models are discussed in Section 4. ITU-R Study Group 3 also has initiated a study on site-specific propagation models that will result in a *Preliminary Draft New Recommendation* (PDNR) on a method for calculating propagation losses over specific paths at VHF, UHF and SHF frequencies. The Sub-group 3K-1 has been studying different methods for two years and has defined the outline for the PDNR. This work is briefly described in Section 5. Finally, the conclusions derived from the above tasks are discussed in Section 6.

2. Univariate Comparison of ITM and TIREM to Measurement Datasets

A study [1] was performed at NTIA to determine the accuracy of predictions by the models in an effort to harmonize them. It also was performed to complement a similar study done at JSC. Thirteen different datasets containing more than 41,000 measurements were included in the study of the 20 – 10,000 MHz frequency range with various types of terrain and antenna heights. Eight of the datasets are the same as those used in the JSC study. The thirteen data sets can be classified into five groups or measurement campaigns:

i) The Phase I [2,3,4] data consisting of three datasets at HF and VHF frequencies, with various polarizations and receive antenna heights, measured at Colorado Plains and Mountain locations and in Northeast Ohio;

ii) The Phase II [5,6,7,8,9] data consisting of five datasets at VHF, UHF and SHF (as many as seven) frequencies and 13–24 discrete receive antenna heights measured at four locations in Colorado and at the Virginia Piedmont region;

iii) The Low Antenna data [10] consisting of three datasets measured in Idaho, Washington and Wyoming at 230 and 416 MHz at two low transmit antenna heights and four low receive antenna heights;

iv) The Fort Huachuca data provided by JSC and measured at Fort Huachuca, Arizona for 60 MHz and fixed transmitter and receiver antenna heights of 10 and 2 m respectively;

v) The TASO (Television Allocations Service Organization) data [11] from an FCC website consisting of measurements of FM radio and TV broadcast signals for a variety of locations in the continental U.S. in the UHF and VHF bands.

In all of the measurement campaigns except iv) and part of v) above, many paths were observed and multiple measurements were attempted per path. In consequence, though many measurements are considered here, a very large amount of the data is correlated; hence, care must be taken when using the data in a statistical context. However, the data provide useful information about height gain, frequency dependence, clutter losses, etc.

The preliminary results of the models' prediction errors (i.e., $L_{predicted} - L_{measured}$) are summarized in Table 1 in terms of the overall datasets' means, standard deviations, skewnesses, and excesses, along with the standard errors of these quantities. The values in parentheses are the corresponding results from the JSC study. In general, there were reasonable agreements between the statistical results of the two studies. In view of the correlated nature of the measurements and the predictions, however, further analysis is required.

3. Comparison of Prediction Errors Using Factor Analyses

As its title suggests, the analysis described in Section 2 is univariate, i.e., it assumes that the prediction error associated with each measurement is independent of every other measurement and identically distributed as the population of the universe of tropospheric radio circuits. However, when the data/prediction errors are correlated, it is necessary to generalize the concepts of mean and variance statistics and do a multivariate analysis [12]. The correlation model adopted here is that data and predictions for different paths are assumed independent, while data and predictions for the same path are not.

In this phase [13], eleven datasets containing more than 18,000 measurements were used. Five obviously erroneous TASO paths were excluded, and missing data were augmented with maximum likelihood estimates using the E-M algorithm. The augmented datasets were subjected to factor analyses based on the eigenvalues and their corresponding eigenvectors of the datasets' prediction errors covariance matrices, formulated on the correlation model described above. In general, it was found that the factors in the measurement datasets correspond to the propagation path, antenna heights, frequency, and polarization. The factor analysis determines how much of each of these eigenvalues contributes to the total variance (i.e., the trace of the covariance matrix) of the prediction errors of the models. The weighting of each of the elements of the corresponding eigenvector determines how the factors contribute to the prediction errors. Appendices I-III of [13] summarize the results of the models' prediction error mean vectors and covariance matrices, and the corresponding ordered eigenvalues and eigenvectors for the Phase I, Phase II (VA only), Low Antenna and TASO datasets. Appendices I, II, and III of [13] summarize the results for both models using 100 m, 200 m and 450 m extraction intervals respectively, using both available and augmented data. In most cases, the total variances of both models decrease with increasing extraction intervals. Appendices Ia, IIa, and IIIa of [13] similarly summarize the results for the augmented data for the same three extraction intervals.

Table 2 gives a summary of the percentages of the datasets' total variances of the errors due to the two largest factors. For every dataset and both models, the first factor always corresponds to the propagation path. The second factor varies among datasets and models; it is sometimes frequency and sometimes antenna height. However, for both models, the propagation path accounts for a large percentage of the error variance. For the results reported here, ITM predictions have been obtained using effective antenna heights above the effective reflecting surface, whereas TIREM uses structural antenna heights.

4. Improvement and Harmonization of the Models

Results of the two studies done at NTIA indicate that sometimes ITM's predictions are more accurate when the effective antenna heights are used, and there are other cases where ITM's predictions are more accurate if the structural antenna heights are used. However, ITM predictions are most erroneous when effective antenna heights are much larger than the structural antenna heights. In the case of TIREM, use of structural heights gives good predictions in some cases. However, in other cases, the use of effective antenna heights may be called for.

In the study mentioned above, the ITM program was used to examine the effective antenna height behavior. Propagation loss predictions were made by using the effective antenna height calculation currently in use and by using structural antenna heights. Both of these methods fail to give low prediction errors in all cases. In fact, in many cases, the optimum value of the effective antenna height is somewhere between the structural antenna height and the effective antenna height calculated by the current ITM algorithm. This is indicated by the prediction errors being positive in one path's measurement and negative in another's. Since the propagation path is the primary factor contributing to the prediction errors, and the effective height calculations depend on the path's terrain profile, it is intuitive that an optimum way of calculating the effective antenna height will most likely reduce ITM's prediction errors. In addition, greater prediction accuracies could accrue from more information for the propagation path, such as LULC, vegetation, buildings, clutter, etc.

5. ITU-R Work on Site-Specific Propagation Models

At the ITU-R Study Group 3 on propagation, there have been discussions regarding site-specific propagation models. In 2000, the U.S. Administration submitted ITM to be considered as the source of a new recommendation. However, a model was also submitted from Germany which has been used for mobile communications in Europe. Therefore, ITU-R Working Party 3K formed an international correspondence group to study these models and compare them to results obtained from measured data as well as to results obtained from ITU-R Rec. P.452, which is used to predict site-specific interference between terrestrial stations between 700 MHz and 40 GHz.

After considerable deliberations over the next two years, the Subgroup 3K-1, responsible for path-specific propagation prediction methods, decided to initiate a new PDNR which will provide guidance for the prediction of path loss and field strength for terrestrial propagation over specific paths from 1 to 1500 km at frequencies of 30–5000 MHz. The models will be deterministic with empirical adjustments to account for clutter and variability. Delay spread also will be considered. Two dimensional terrain profiles will be used with minimum resolution of 1 km and maximum resolution of 50 m.

6. Conclusions

Evaluation of path-specific models both in the U.S. and at the ITU has offered much insight into propagation prediction methods. Important factors for this evaluation include the availability of 'good' measurement data and understanding the nature of its underlying regression, when it exists. For the U.S. measurement datasets considered here, the variances of the two models' prediction errors predominantly depend on how well or poorly the path is modeled. Therefore, good terrain data and better information about the particulars of each path are important for reducing models' prediction errors, in addition to improving the quality of the physical approximations used in the models. Additional information regarding vegetation, buildings, urban clutter, ground constants, atmospheric refractivity, etc., seems also to be needed. For ITM and, perhaps, for TIREM, an improved/optimized effective antenna height algorithm holds some promise of effecting some of this reduction. The work at the ITU will be based on many of these factors plus contributions from other countries based on their data and experience.

7. References

[1] P. McKenna, A. Paul and F. Najmy, "On the Comparison of ITS's Irregular Terrain Model (ITM v1.2.2) and JSC's Terrain Integrated Rough Earth Model (TIREM v3.14) Several Measurement Datasets," NTIA, Nov. 2000.

[2] M. E. Johnson, M. J. Miles, P. L. McQuate and A. P. Barsis, "Tabulations of VHF propagation data obtained over irregular terrain at 20, 50 and 100 MHz, Part I: Colorado plain data," ESSA Technical Report, IER 38-ITSA 38-1, 1967.

[3] M. E. Johnson, M. J. Miles, P. L. McQuate and A. P. Barsis, "Tabulations of VHF propagation data obtained over irregular terrain at 20, 50 and 100 MHz, Part II: Colorado mountain data," ESSA Technical Report, IER 38-ITSA 38-2, 1967.

[4] M. E. Johnson, M. J. Miles, P. L. McQuate and A. P. Barsis, "Tabulations of VHF propagation data obtained over irregular terrain at 20, 50 and 100 MHz, Part III: Ohio data," ESSA Technical Report, IER 38-ITSA 38-3, 1967.

[5] P. L. McQuate, J. M. Herman and A. P. Barsis, "Tabulations of propagation data over irregular terrain in the 230- to 9200-MHz range, Part I: Gunbarrel Hill receiver site," ESSA Technical Report, ERL 65-ITS 58, 1968.

[6] P. L. McQuate, J. M. Herman and A. P. Barsis, "Tabulations of propagation data over irregular terrain in the 230- to 9200-MHz range, Part II: Fritz Peak receiver site," ESSA Technical Report, ERL 65-ITS 58-2, 1968.

[7] P. L. McQuate, J. M. Herman M. E. McClanahan and A. P. Barsis, "Tabulations of propagation data over irregular terrain in the 230- to 9200-MHz range, Part III: North Table Mountain site," ESSA Technical Report, ERL 65-ITS 58-3, 1968.

[8] P. L. McQuate, J. M. Herman and M. E. McClanahan, "Tabulations of propagation data over irregular terrain in the 230- to 9200-MHz range, Part IV: Receiver site in a grove of trees," OT/TRER 19, 1971.

[9] G. A. Hufford and F. K. Steele, "Tabulations of propagation data over irregular terrain in the 230- to 9200-MHz range, Part V: Virginia," NTIA Report 91-282, 1991.

[10] L. G. House, F.G. Kimmet and J. M. Harman, "UHF propagation data for low antenna heights," ESSA Technical Report, ERL 134-ITS 93, Volumes I and II, 1969.

[11] FCC Website, ftp://www.fcc.gov/pub/Bureaus/Engineering_technology/Databases/mmb/fm/model/taso Created by D. Ring and associates.

[12] P. McKenna, "A comparison of radio propagation measurements and predictions at VHF and UHF using univariate and multivariate normal statistics," presented at ISART 2002.

[13] P. McKenna, N. DeMinco and K. Allen, "ITM and TIREM Comparison and Improvement," NTIA, Jan. 2003.

Table 1. Comparison of Overall Dataset Prediction Error Statistics for ITM & TIREM

Data	No. of meas.	ITM mean (dB)	TIREM mean (dB)	ITM std. dev. (dB)	TIREM std. dev. (dB)	ITM skew-ness	TIREM skew-ness	ITM excess	TIREM excess
CO. MTNS.	550 (286)	-17.1 +/- .7 (-22.8)	-4.4 +/- .6 (-4.5)	16.2 +/- .5 (12.1)	13.7 +/- .4 (15.0)	.6 +/-.1	.0 +/-.1	.5 +/-.2	-.1 +/-.2
CO. PLNS.	1983 (1983)	-14.9 +/- .2 (-16.7)	-4.4 +/- .2 (-5.6)	10.2 +/- .2 (10.3)	9.9 +/- .2 (9.8)	0.0 +/-.1	.1 +/-.1	.4 +/-.1	.4 +/-.1
NE OH.	1787 (1787)	-10.1 +/- .2 (-12.7)	0.0 +/- .2 (-.2)	9.2 +/- .2 (8.7)	9.6 +/- .2 (8.7)	0.0 +/-.1	.1 +/-.1	.4 +/-.1	.1 +/-.1
R-1	6780	2.0 +/- .2	1.2 +/- .1	13.9 +/- .2	12.0 +/- .1	.8 +/-0.0	-.2 +/-0.0	2.5 +/-.1	.8 +/-.1
R-2	2458	-7.5 +/- .5	-18.4 +/- .4	25.7 +/- .3	20.8 +/- .3	-.2 +/-0.0	-.6 +/-0.0	-.3 +/-.1	.0 +/-.1
R-3	5149	1.9 +/- .2	2.8 +/- .2	11.6 +/- .2	11.4 +/- .1	.9 +/-0.0	.2 +/-0.0	3.1 +/-.1	.4 +/-.1
R-4	9498	-12.8 +/- .2	-14.1 +/- .2	16.6 +/- .2	16.6 +/- .1	-.4 +/-0.0	-.8 +/-0.0	1.3 +/-.1	.3 +/-.1
VA.	1655 (1871)	-.9 +/- .3 (-3.7)	-.2 +/- .4 (1.8)	13.2 +/- .3 (9.6)	15.6 +/- .3 (10.8)	-.2 +/-.1	-.2 +/-.1	1.5 +/-.1	.2 +/-.1
ID.	435 (435)	-17.5 +/- .7 (-15.4)	-10.9 +/- .6 (-8.7)	14.5 +/- .4 (12.7)	11.9 +/- .4 (11.3)	-.3 +/-.1	-.4 +/-.1	-.4 +/-.2	-.2 +/-.2
WA.	892 (892)	-2.4 +/- .4 (-5.7)	5.1 +/- .4 (4.7)	12.8 +/- .3 (11.7)	12.0 +/- .3 (13.2)	.3 +/-.1	.1 +/-.1	.1 +/-.2	.2 +/-.2
WY.	704 (704)	-11.9 +/- .6 (-15.8)	-6.8 +/- .5 (-5.5)	14.6 +/- .4 (12.6)	12.5 +/- .3 (9.7)	-.1 +/-.1	0.0 +/-.1	0.0 +/-.2	-.2 +/-.2
Ft. Hua.	372 (420)	-3.0 +/- .6 (-5.4)	11.4 +/- .3 (7.8)	11.5 +/- .3 (11.3)	6.0 +/- .2 (5.9)	-.5 +/-.1	-.2 +/-.1	-1.2 +/-.3	-.9 +/-.3
TASO	8865	-3.2 +/- .1	-1.2 +/- .1	12.5 +/- .1	14.0 +/- .1	-.2 +/-0.0	-.5 +/-0.0	2.3 +/-.1	.7 +/-.1

statistic +/- std. error (JSC statistic)

Table 2. Percent of Error Variance due to First and Second Factors

	ITM		TIREM	
Dataset	**First**	**Second**	**First**	**Second**
Phase I, CO mountains	85	7	78	12
Phase I, CO plains	84	6	79	8
Phase I, NE Ohio	72	9	67	11
Phase II, Virginia	81	8	60	19
Low Antenna, Idaho	89	5	93	4
Low Antenna, Washington	93	3	93	4
Low Antenna, Wyoming	93	4	87	6
TASO, t2_nb	91	9	82	18
TASO, t8	87	7	74	17
Average	**86**	**6**	**79**	**11**

Public safety and emergency services in future wireless communication

Douglas C. Sicker
Interdisciplinary Telecommunications Department
Department of Computer Science
University of Colorado at Boulder
303-735-4949 (phone)
303-492-1112 (fax)
douglas.sicker@colorado.edu

As various forms of wireless communications become more pervasive, the expectation for this communication to support public safety and emergency functionality will arise. The public switched telephone network presently supports many such functions. The most familiar of these functions is that of Enhanced 911 (E911), for the provision of emergency response capability. A lesser known function is that of Telecommunication Service Priority (TSP), for the prioritization of service provisioning and restoration for national security or emergency preparedness missions. Other functions, such as Government Emergency Telecommunication Service (GETS) and Wireless Service Priority (WSP) provide what the Federal Communications Commission refers to as Priority Access Service (PAS). GETS provides a means of increasing call completion probabilities during times of heavy congestion, which may result from natural or man-made disasters or other emergencies. Likewise, WSP is now being deployed in the wireless space to provide increased probability of accessing resources during times of congestion. WSP will serve to complement GETS for end-to-end connections from wireline or wireless environments. Together these services (and a few not mentioned) form the primary basis for public safety communications.

Most of these public safety services rely on traditional telephony technology. For example, Signaling System 7 (SS7) and Intelligent Network (IN) allow priority services, specialized routing and network management capabilities for such services as GETS and WSP. However, this type of approach may not apply to emerging wireless systems, in that there is a fundamental shift away from centralized control environments, such as SS7 and IN. In such an environment, an authority (e.g., the service provider) controls the network by monitoring and assigning resources to the end users. This approach is rather antithetic to many of the emerging wireless architectures, where control and resource assignment is decentralized. A question this raises is what problems this shift creates and whether such services can still be provided in a reasonably reliable manner? Maybe a more fundamental question to consider is if these networks should support public safety and emergency services?

In this paper, we begin by considering a number of existing public safety and emergency services. Next, we consider the technical problems and requirements associated with providing these services. We next examine an existing wireless technology with respect to these requirements. To end, we consider the applicability of such services.

Keywords: public safety, emergency services, wireless, E911, priority, GETS, WSP

Introduction

As we have seen in the commercial wireless space, there is high expectation from the public, government agencies and public safety officials for communications networks to support emergency services. [1] This expectation has only increased with the recently heightened concerns regarding terrorism and the accompanying public safety readiness. The range of national security or emergency preparedness (NS/EP) events that might warrant the use of public safety and emergency services (hereafter referred to as public safety service) is broad.[1] It can include responses to requests made under heavy traffic loads (such as during a national disaster) to responses to requests made under normal traffic conditions (such as most E911 calls). However, we will not debate the virtues or the need for such services; nor we will define what warrants the invocation of such services. Rather, we consider the technical requirements for public safety and emergency services in light of changing network architectures and the implications this change has on the provision of such services.

In this paper, we examine public safety services with respect to future wireless networks. The intention is to outline the functions required to provide public safety services in future wireless networks. We begin with a brief discussion of how wireless network architectures are changing. Next, we provide a brief background on

[1] While the term 'public safety' is often perceived as applying only to communication services among public safety officers, we use the term more loosely to include such things as priority access, priority provisioning, emergency services (E911) and more.

existing telecommunications public safety services. From this we extract the basic functionality needed to provide these services. We then consider these basic functional requirements as they relate to existing wireless technology. This paper will not focus on the steps required to make existing wireless systems public safety capable. Rather, we will focus on describing the functions required of these networks. Finally, we will consider whether such public safety services are even applicable to these emerging wireless networks.

Architectural Change

We begin by briefly considering how networks are changing and what this change means to public safety services. We briefly examine a number of functional differences that exist between traditional and future networks, and what this might mean to providing such services. When we say future networks we mean packet based networks, namely Internet Protocol based networks.

Control: In the circuit switched world (including cellular), connections among end points are made by a provider controlled signaling system, namely SS7/IN. This function exists separate from the voice channel and provides a means of querying, setting up, altering, and tearing down connections. Packet networks, such as the Internet, do not make use of such centralized systems to control resources and connections. Rather, information such as routing data is contained within the packet as header information. Route updates and other such information are propagated as packet data along with other service data. This is not to suggest that signaling does not exist in the Internet, but rather that the design approach is not one of strict provider control.

As indicated, most public safety services rely on traditional telephony technology. For example, the SS7/IN allows priority services, specialized routing, and network management capabilities for such services as GETS and WSP.[2] It is the centralized coordination of this signaling system that allows for many of the public safety services discussed earlier. However, this approach does not apply to emerging wireless systems. For example, there is a fundamental shift away from control environments, where providers control the network by monitoring and assigning resources to the end users. This is antithetic to many of the emerging wireless architectures; where control is highly decentralized.

[2] Interestingly, most E911 systems still depend on antiquated technology and do not make use of SS7 and IN, which incidentally have been in existence and use for decades. However, rather than argue for the move toward another highly centralized control environment (such as SS7/IN), we must consider the public safety implications of moving toward distributed control environments.

Layered models: The layered model of Internet Protocol (IP) based networks separates the application from the transport, allowing for decentralized control. This decentralized control means that the end point may be responsible for providing the intelligence required to offer services. However, some public safety services cannot simply be pushed into the application layer and assumed to operate as required. For example, an E911 call requires that location information be passed to public safety officials. This location information will likely involve some radio signal information for position determination. The application layer will depend on this lower layer service to provide location capability.

Delay: Unlike a circuit switched network, most packet networks do not dedicate resources to the exclusion of other services. Various protocols exist to address the problem of resource reservation and prioritization for packet networks, but few of these are deployed. As such, packets may be delayed an unacceptable period of time, which may cause an E911 call to drop or be unintelligible. End-to-end delays of as little as 250ms can make conversation difficult and destroy video. Similar issues with packet networks include delay variation and packet drop.

A question that remains is how to allow highly distributed, highly decentralized networks to support services that have traditionally depended on highly centralized structures. Another question is how to deploy these services in emerging wireless networks in a reasonably reliable manner.

Existing Public Safety Services

In this section, we will examine a number of existing public safety and emergency services. While this will serve as a general template for future services requirements, it should be realized that the expectation of what defines a public safety service (and its requirements) will evolve just as networks evolve. To that end, we begin small with a core set of services.

The services we consider in this section include Government Emergency Telecommunications Service (GETS), Wireless Priority Service (WPS), Cellular Priority Services (CPS), Telecommunications Service Priority (TSP) and Emergency Services. While this is not exhaustive of all public safety services, it represents an important set of services.

GETS: GETS is the wireline Priority Access Service (PAS) service. It provides a means of increasing call completion probabilities during times of heavy congestion, which may result from natural or man-made disasters or other emergencies. It allows an authorized user to access services through the use of a universal access number and a PIN. GETS provides enhanced routing and various priority treatment of calls to improve completion rates. [2]

WPS: WPS provides a wireless counterpart to GETS by offering an enhanced probability of completing calls.

This technology has been provided through channel reservation or priority access queuing techniques. [3]

CPS: CPS, a generalized term for cellular wireless PAS, provides end-to-end cellular priority for NS/EP users and can make use of various technologies to provide this priority function. [4]

TSP: TSP provides for more timely installation and restoration of public safety services. This is a means of being placed on the top of the repair and install order process. [5]

Emergency Services: This provides a means of requesting emergency services. Enhanced 911 (E911) extends basic 911 by adding automatic number identification (ANI) and automatic location information (ALI) technology to emergency calls. The requirements involved in providing E911 services include; use of a universal emergency number ('911'), locating the user, determining the appropriate Public Safety Answering Point (PSAP), determining the number of the calling party and corresponding location and routing the call. [6]

From the above we will now extract a set of basic and specific functions.

Functionality

In this section, we examine a set of functional requirements necessary to support public safety services. To generalize, we might divide the above services into 1) prioritization of access and restoration and 2) emergency. Prioritization of access provides a means of increasing call completion probabilities during times of heavy congestion, which may result from natural or man-made disasters or other emergencies. Prioritization of restoration is simply the ability to offer a more timely installation and restoration for public safety users. Emergency services provides a means of requesting emergency response.

GETS, WPS, TSP and E911 represent a set of Public Switched Telephone Network (PSTN) oriented public safety services, and as such, depends on the architecture and functionality of the PSTN. However, this architecture will change substantially as we move more services toward packet based networks. Therefore, we need to generalize these requirements; uncouple them from the architecture in which they are provided.

In this following section, we consider the building blocks required of each service, with the goal being to create functions that are independent of the architecture.[3]

Addressing and numbering: The network must support the addressing/numbering required to route a universal number. A common example of this function is that of an emergency service number '911' or a directory service number '411'. This functionality will depend on the existence of an agreed upon session functionality to provide such an address.

Location: The network must support the ability to locate calling parties. Various methods may be used to locate a user including; base stations triangulation, Global Positioning Satellite (GPS), assisted GPS, manual registration and others. The accuracy of such systems depends on a host of technical and operational issues. The details of how this operates is outside the scope of the paper.

Querying: The network must support the ability to query the appropriate database in the appropriate manner. For example, a determination must be made to which public safety point a call should be routed. This process will depend heavily on directory services and access to these directories, as well as on the validity of these records. Many of the more difficult problems within this space may come back to policy and regulation, to resolve the who, what, where, when and how of accessing these directories. In that much of this functionality is at higher layers of the protocol stack, this requirement may be abstracted away from the underlying wireless networks.

Mapping: Arguably distinct from the location and querying system, public safety systems require geographic information systems. This provides the means of mapping and associating the various elements required in location dependent systems.

Quality of service: The network must provide the appropriate Quality of Services (QoS) required of the service. While there are many solutions for this QoS differentiation in packet networks, few are actually deployed. When we consider this issue in terms of a future wireless network, we must consider how best to integrate such capabilities with the public safety service. If a service requires support for voice or video, QoS may be an important issue. However, there are other services, such as messaging, that may not require special QoS treatment.

Prioritization of service: The notion of Prioritization of Service (PoS) relates to that of QoS. The idea is that not only does a service require special treatment (such as QoS for video), but that certain users of that service require additional specialized treatment (such as preemption). This issue arises when many players are trying to make use of a highly constrained resource, such as during times of disaster.[4] At such times, public safety officials may be required to preempt other players. These services may be origin dependent (based on the device) or user dependant (based on the user). A primary difference is the need to authenticate a user on any line, versus a line being specified for such services. [8] There

[3] For example, traditional PSTN public safety functions include things like trunk prioritization, emergency override and calling line restrictions override. While it may be that such services exist in future networks, these functions are closely tied to the PSTN.

[4] One provider, T-Mobile, a Global System for Mobile (GSM) provider, has taken advantage of a GSM feature referred to as enhanced Multi-Level Precedence and Preemption, which allows calls to queue while waiting for the next available radio channel. [7]

are a number of other PSTN specific functions used in PoS, such as priority routing, automatic repeat request and last trunk reservation. Some of these may apply to future services and some may not. Rather than focus on each of these functions, we merely describe the basic requirement.

Power: An easily overlooked issue is that of power. Any device that provides its own power has an additional availability constraint. Many rely on the availability of their wireline phone when their power is out. This sets an expectation for future networks. However, users of cell phones have come to realize the need to charge batteries on a regular basis and this same expectation will likely translate to other wireless devices. Nonetheless, this creates a possible availability issue for wireless devices with respect to NS/EP services.

Reliability: The network must provide the appropriate level of reliability. The ability for a packet network to provide the 99.999% reliability of the PSTN requires considerable effort. The public Internet provides around 99% reliability at a macro scale. While this falls short of the expectation of the PSTN, the Internet was not designed to provide similar reliability measures. Error detection and correction provides improved reliable delivery of data services but does not work well with real time services. Reliability raises a number of questions with respect to public safety services. What are the reliability and availability requirements expected of the system? Should these requirements differ among systems and how? How stable and usable is the design?

Security: The network must provide a broad level of security including the typical areas of confidentiality, integrity, availability and non-repudiation. Public safety services may require secure communications to avoid message interception. Certain communications might require strict authentication and authorization schemes to ensure that only appropriate users access the resources. This proposes an interesting problem in the wireless arena, where signals are broadcast into the ether for anyone to intercept. Wireless networks also raise interesting availability problems, such as jamming (a type of Denial of Service (DoS) attacks), where an attacker could block communications by transmitting a powerful and disruptive signal. Key management, the means of generating, distributing, storing, using, renewing and removing keys, is another area of concern for wireless public safety services. For example, over-the-air-rekeying, the means of renewing keys among wireless endpoints, requires additional security mechanisms unique to the wireless arena. Several of the above requirements may be fulfilled (or partially fulfilled) by the use of encryption technology. It is common to see Data Encryption Standard (DES) and triple DES employed in public safety services, and moving forward we can expect to see the use of Advanced Encryption Standard (AES). Much of the work is employing these schemes with the appropriate protocols.

One major concern that arises within the E911 space is the issue of false reports. The serious nature of this problem is easily realized by the consequence of a prank call tying up resources while a response to a real emergency is delayed. The methods of providing non-repudiation may be borrowed where applicable from existing security models. For example, a "handshake could occur where keys are exchanged in a manner that demonstrates the identity of the calling and called party (realize that the public safety answering point may also be spoofed). Presently in the PSTN, these false reporting events are addressed by law, in the form of felony crimes. A similar expectation is not unreasonable in future wireless space. The question that remains is whether the increase in the number of devices, the ease of spoofing, and other such differences, makes enforcement untenable.

Signaling: The network must provide a means of signaling session information. In the PSTN, this is provided by the SS7/IN and in the IP space this might be provided by Session Initiation Protocol (SIP).[5]

Order Processing: The network provider must support a means of prioritizing repairs and restorations. The complex Operational Support Systems (OSS) and Back-office Support Systems (BSS) that exist within the PSTN do not presently exist within most packet networks.

While not all of these functions are required to provide any one public safety service, together they represent a core set of requirements for the services we discussed in the previous section. With this said, it may turn out that only a subset of services (and therefore functions) will come to be expected of future networks. Further, new methods of performing these functions may evolve. Therefore the intent of the above list is to describe functionality regardless of how it is ultimately provided. Next, we will consider these functions with respect to a number of existing and emerging technologies.

Technologies

In this section, we will briefly examine an existing wireless set of standards, namely 802.11. This is not meant to be an exhaustive examination, it is only meant to be illustrative (and cursory at that).[6] We also briefly discuss the role of software defined radios and ultra wide band technologies with respect to public safety services.

---

[5]See [9]

[6]For example, we do not cover 802.16, Bluetooth, 3G/4G or Local Multipoint Distribution Service (LMDS) and Multichannel multipoint Distribution Service (MMDS) services. There are a number of other wireless options to consider for public safety services. This includes satellites, that while costly, can provide ubiquitous coverage and good diversity alternatives in disaster relief efforts.

802.11: 802.11 describes a family of wireless LAN standards.[7] These networks make use of unlicensed spectrum, which raises concerns of crowding and congestion. While these networks have been shown to be surprising resilient under heavy traffic loads [10], they do not support the QoS or PoS functionality described previously nor do they provide the level of reliability generally associated with public safety services. Furthermore, while 802.11 has been shown to support real-time services, [11] unlike 3G and 4G systems they were not designed for this purpose.[8] Another common complaint is the weak security protocols associated with 802.11 networks. These networks do not provide the security requirements discussed in the previous section. A final functionality to consider is that of location. The limited range over which 802.11 networks operate limits the geographic scope and subsequent location range of a user. However, there is no means of determining this location. Furthermore, even this limited range is too large to ensure a timely response to an emergency call. There are a number of possible solutions to the location problem.

Before leaving the topic, it might be interesting to consider a few experiments using 802.11 for public safety services. While 802.11 networks were designed for the portable environment, it turns out that they support high mobility as well. An experiment by the California Department of Transportation showed that an 802.11b network could support service to a vehicle moving at 70 mph. [13] This suggests a role for their use in highway patrol cruisers. Other experiments have shown 802.11b useful in supporting novel emergency services. For example, a new medical messaging service has been developed to receive data on FM sub-carrier (including RDS, DARC, or SuperDARC formats) and retransmit on 802.11b. In this service, the ambulance essentially becomes an 802.11 access point, allowing medical messages to be transmitted up to a mile. [14]

Two other technologies, Software Defined Radio (SDR) and Ultra Wideband (UWB), while not network protocols will likely have a profound impact on future wireless networks. Rather than consider the ability of this technology to support all of the afore mentioned functions, we describe a few specific applications that they may enhance.

Software Defined Radio: SDR provides a multiband, multimode radio technology. This allows users to change modulation type, operation mode, radiation power, and air interfaces. This brings such benefits as interoperability among dissimilar services and novel security services. SDR technology could support interoperability with existing wireless services and integration with new services. It could also provide the ability for public safety users to quickly negotiate into less congested spectrum during times of emergency. [15]

Ultra Wideband: UWB is defined as a wireless technology that makes use of more than 1.5 GHz of spectrum. The concept is to make use of broad ranges of spectrum often at low power. Such technology can co-exist on other service spectrum without inference and thereby make large amounts of spectrum and high rate services available. UWB brings the ability to reduce spectral congestion, minimize interference among devices, reduce power requirements, improve security and minimize interference from multipath interference. [16] The UWB signal provides a low probability of interception and detection, thereby increasing security and creating novel alternate path communications. This could be a useful security tool for public safety users. It is worth mentioning that a number of concerns do exist regarding the potential for interference of UWB with other services. The FCC recently addressed this problem in a Report and Order. [17]

Considerations

In this section, we examine the applicability of public safety services in future wireless networks. The question is whether these services are relevant and/or appropriate to future wireless networks. Rather than propose an answer to this question, we raise of number of important considerations. While we will consider several policy related issues, we will not examine policy specifically in this paper.[9]

Expectation: It is important to realize that future public safety services will likely be quite different from those provided in the PSTN space. For example, voice might be replaced with text messaging or it might be expanded to full voice, video and telemetry. While emerging wireless services will create new problems they will also offer new services (video), alternate devices, increased penetration, neutral platforms, flexibility and additional information. There may be a tradeoff among the inability to offer traditional services and the ability to offer new services (limited service support over broader coverage). The base expectations for what qualifies as a public safety service will likely change as new platforms and devices emerge.

Offering: One point worth considering is the distinction that could be made between a service that is held out as commercial "for fee (e.g., cellular service)

---

[7] While the standards include 802.11, 802.11a, 802.11b and 802.11d, we will look only at 802.11a and 802.11b.

[8] One advantage of 3G is that its PSTN-based architecture should rather readily support the network control traditionally applied to public safety services. 4G service is expected to provide much higher rates and support such services as voice and video. In that this is a very nascent technology there is good opportunity to integrate new public safety services into 4G. [12]

[9] Numerous federal actions have recently occurred regarding wireless public safety issues. Many of the actions focus on spectrum issues concerning public safety spectrum, while others have included such issues as priority access services waivers. While these issues are quite relevant to the topic of this paper, we do not have the space to address them.

and one that is not. The point being that a commercially offered service might have a higher expectation for providing public safety services. This notion of "holding out of a service may have some relevance in the legal and regulatory space, which is beyond the scope of this work.

Cost: To understand the appropriateness of public safety services requires some form of cost/benefit analysis. This cost analysis can be based on the direct, indirect or exogenous costs, with varying implications. Cost will be a major driver in the success or failure of such services. Note that the cost for such services may also be forced upon carriers by regulatory policy.

Diversity: The diversity of access options is a key reason to consider wireless support for public safety communications. Wireless networks would provide alternative communications channels in the event wireline networks were rendered inoperable. They also provide coverage in areas that might not otherwise be reached. In this sense, it would provide reliability and availability through diversity.

Accessibility: Should public safety functions allow a user to log into a network on which they do not have an account in order to allow that user to report an emergency event? Will network operators block interoperability with a possible negative impact on public safety services? In this way accessibility may be viewed as an access control problem.

Deployment: To avoid false starts, a number of initial requirements could be defined and considered for each service. Consideration must be made regarding the degree to which these requirements are implemented. Care should be taken not to spend too much time creating a complete implementation, without interim deployment cycles. For example, an interim IM based 911 service for 802.11 based devices could be implemented well in advance of a full-blown voice based E911 service. Coordination of local, state, federal and industry players will play an important role in the short and long term deployment and integration of public safety services.[10]

Interoperability: Interoperability with legacy public safety systems will be crucial. The PSTN can serve as a common platform on which other networks interconnect. However, the PSTN was created to support voice and thereby might be viewed as limiting emerging functionality such as video and high rate data. A question to consider is how will the new system integrate with existing systems and other new systems? Are there international interoperability issues? Technology will both provide solutions and additional problems for interoperability among so many disparate technologies and devices. Interoperability will have to include higher layer specifications, such as new extensible markup language (XML) formats and other application layer functionality, as well as the operational and back office support systems. Independence of the application from the transmission mode may be key to interoperability and backward compatibility with other public safety services, but this does not ensure that the application can necessarily interoperate or provide the functions needed for public safety functions.

Openness: The openness of these technologies may be relevant to their success in supporting public safety services. Openness is a broadly used term and may include aspects of design, agreement, process or more. Some questions to consider include: Will this system be based on "open, non-proprietary standards? Will the interfaces to the operating system be available for developers? Will the process require regulatory intervention? What, if any, security issues does open or closed design create?

As we have demonstrated, there are many factors to consider regarding the applicability of public safety services to future wireless networks and each of these factors has a set of questions to ponder. A difficulty arises in defining expectation and functionality of devices that do not yet exist. Another difficulty that exists is that of moving away from traditional expectations set by previous technology that might not be applicable in the future.

Conclusions

In this paper, we examined the functionality required of future wireless networks to support public safety services. We began by examining some of the architectural changes that are making this assessment necessary. We also examined existing technologies in order to determine the extent to which these technologies might support public safety services. We closed by discussing the applicability of public safety services in future wireless networks.

There are several issues that should be considered regarding public safety in the development of new wireless technologies. The first is that the architectural changes in emerging wireless networks will impact the manner in which public safety services are provided. The second issue to consider is that existing public safety services require a significant set of functions. The third issue to consider is that existing wireless networks may not provide the functionality associated with PSTN based public safety services. This brings us to consider the applicability of public safety services to future networks

---

[10] Within the public safety community various groups are looking at the role of wireless networks. One effort is the Public Safety Wireless Interoperability National Strategy (WINS). They are involved in promoting and interoperability efforts in wireless public safety networks. Project Mobility for Emergency and Safety Applications (MESA) [18] and Capital Wireless Integrated Network (CapWIN) provides a focus on the administrative, operational, and interoperability issues of existing wireless services. ITU E.106 represents new standards for emergency telecommunications services. [19] While this work is developing useful output, they are not presently focusing on the next generation wireless services issues.

and the notion that some NS/EP services may be provided in a different way on emerging networks and additional NS/EP services that are not currently provided on legacy networks may be provided by emerging networks. As we have attempted to demonstrate, it can be difficult to determine how applicable certain public safety services are for new technology, particularly since these technologies are so fundamentally different from existing technology.

While we do not recommend that government dictate public safety requirements for future wireless services, neither do we perceive this as an excuse to step away from supporting such services. By allowing new devices to support a general public safety functionality, we will increase public safety connectivity and availability. It might be prudent from the start to make a cursory analysis of what issues might arise for a particular technology implementation of public safety. This type of general analysis has precedence. Consider the 'security' section of an Internet Draft or efforts underway within the disabilities community to raise awareness of accessibility problems. The outcome of such an analysis could be as simple as 'it is not presently feasible or applicable'.

References

[1] *Revision of the Commission's Rules to Ensure Compatibility with Enhanced 911 Emergency Calling Systems*, CC Docket 94-102, RM-8143, Notice of Proposed Rulemaking, 9 FCC Rcd. 6170 (1994) (E911 NPRM).

[2] *GETS*, http://gets.ncs.gov

[3] *WPS,* http://www.ncs.gov/N2/wps

[4] *CPS*, http://www.ncs.gov/N2/Wps/cps1.html

[5] *TSP*, http://www.ncs.gov/tsp/briefing_hi_body1.html

[6] *A Report on Technical and Operational Issues Impacting The Provision of Wireless Enhanced 911 Services,*http://gullfoss2.fcc.gov/prod/ecfs/retrieve.cgi?native_or_pdf=pdf&id_document=6513296239

[7] *WPS*, http://63.121.95.245/wps/info_body.html and www.ncs.gov/N2/wps

[8] *Emergency PSTN Features*, http://www.ietf.org/internet-draft/draft-otty-ieprep-pstn-features-00.txt

[9] "Providing Emergency Services in Internet Telephony Schulzrinne H. and Knarig, A, www.computer.org/internet, June, 2002.

[10] *Testbed-based Performance Evaluation of VoIP over Wireless & Wired LANs*, http://www-sop.inria.fr/planete/qni/VoIP_wireless.pdf

[11] *Voice over 802.11,* http://www.80211-planet.com /news/article.php/1015241

[12] *4G Applications and Services*, http://www.pswn.gov/library/pdf/wireless_technologies_one.pdf

[13] *Towards Distributed Data Collection and Peer-to-Peer Data Sharing*, http://www.its.uci.edu/its/publications/papers/AS-WP-02-4.pdf

[14] *Roadside Telematics and Cue to Deliver Emergency Medical Messaging Over 802.11 and Bluetooth,* http://www.roadmedic.com/news/may0601.htm

[15] *Software-Enabled Wireless Interoperability Assessment Report,* http://www.pswn.gov/admin/librarydocs9/software_defined_radio_report_final.pdf

[16] *Ultra Wideband Communications,* http://www.pswn.gov/admin/librarydocs8/Emerging_Wireless_Technologies-Part4.pdf

[17] *Docket 8-153*, http://wireless.fcc.gov/publicsafety/2002docs.html and http://www.pswn.gov/admin/librarydocs10/uwb_gps_final.pdf

[18] http://www.projectmesa.org

[19] *E.106*, http://www.itu.int/ITU-T/studygroups/com16/ets/

LoL@ - A Prototype of a Network Independent Wireless Internet Service

Harald Kunczier[1], Elke Michlmayr[1], Günther Pospischil[2], Hermann Anegg[1]

[1]ftw. Forschungszentrum Telekommunikation Wien, Donau-City-Str. 1, A-1220 Vienna
Email: {kunczier, michlmayr, anegg}@ftw.at,
Phone: +43-1-5052830-0, Fax: +43-1-5052830-99

[2]Mobilkom Austria AG & Co KG, Obere Donaustrasse 29, A-1020 Vienna
Email: g.pospischil@mobilkom.at, Phone: +43-1-33161-0

ABSTRACT

In this paper we introduce LoL@, the Local Location Assistant, a mobile tourist guide for the inner city of Vienna. The project was carried out by several industry and research partners including Vienna University of Technology and the Telecommunications Research Center Vienna (ftw). Since upcoming mobile services are likely to utilize Internet technology in addition to existing telecommunication infrastructure, our vision was to create such a wireless Internet service prototype. We designed and implemented a proof-of-concept demonstrator of a flexible service architecture for network-independent wireless Internet services. LoL@ offers the look and feel known from the Internet combined with additional network functionality like location determination and call control. Different IP-bearers like LAN, WLAN, GPRS, or UMTS are supported. LoL@ features interactive map-based navigation by guiding a visitor along a tour through Vienna's inner city, offers terminal positioning and the provision of multimedia information related to the sights. This information, as well as an online tour diary with personal entries and digital photos, is also accessible in advance for tour planning or after the tour for later use.

I. INTRODUCTION

The increasing data rates in mobile communications as well as the convergence between telecommunications and data communications like the Internet has amplified the creation of mobile applications. The WAP-based [1] application environment, already state-of-the-art, with its binary HTML like language has been a first step towards a more flexible application creation. MExE classmark 2 and 3 [2] introducing Java in the mobile world has paved the way for further improvements of mobile applications and allows the usage of familiar Internet technology. However, there are still numerous challenges for the design of mobile applications. In this paper we will present LoL@, a wireless Internet service, implemented as a proof-of-concept demonstrator to point out possibilities, challenges and problems in the intersection between the Internet and the telecommunications world. The issues described are divided into three sections. After this introduction the second section will introduce the service LoL@ including a short use case. Section III will then identify several challenges like to access different network capabilities and some security issues related to the interworking between network operator and service provider. Possible solutions are presented on the basis of LoL@. Section IV will emphasize LoL@'s application design concepts trying to overcome challenges arising from limited device capabilities like small display sizes and challenges resulting from the characteristics of wireless networks like low bandwidth or frequent disconnections. We will conclude in the last section with ideas for future work and a short summary.

II. LoL@ - A PROTOTYPE MOBILE INTERNET SERVICE

Developing applications for the mobile Internet is a challenging task because of technical and user interaction constraints. We will describe this in more detail in Sections III and IV. However, these constraints should be considered already during the design phase of a mobile service. Figure 1 gives an overview of which capabilities and limitations had been most decisive to the design process of LoL@. The result is a prototype of a mobile tourist guide offering

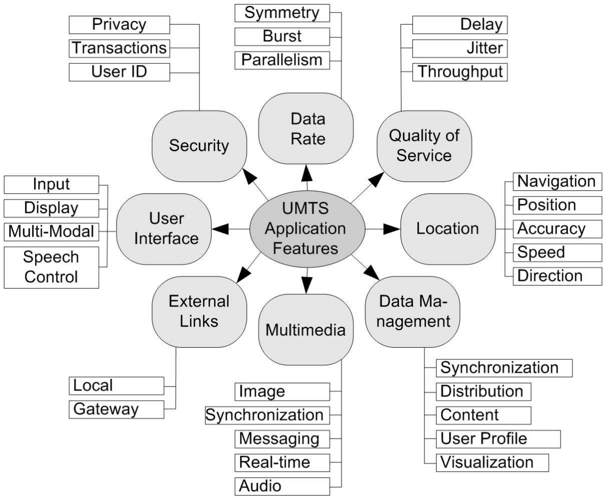
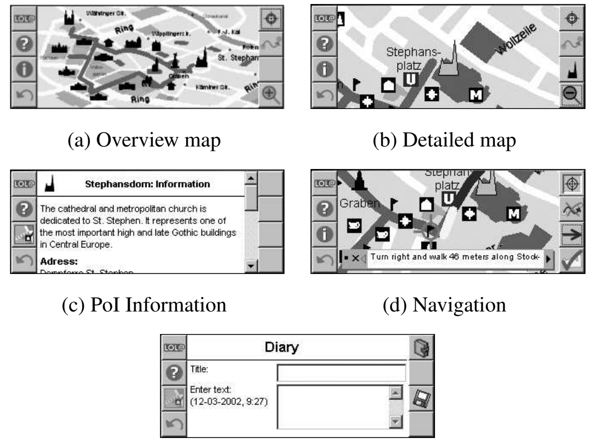


Fig. 1. UMTS Service Features



(a) Overview map

(b) Detailed map

(c) PoI Information

(d) Navigation

(e) Tour Diary

Fig. 2. LoL@ GUI

- guidance to user selectable points of interests (PoIs)
- interactive multi-media information (pictures, video, text) about the selected PoI and
- a flexible tour diary with automatic and manual entries, allowing to substitute currently used textbooks.

The application offers a map-centric graphical user interface with additional hypertext information screens and is optimized for the limited screen size of mobile devices. A sample usage scenario looks like this: When the user starts LoL@ she/he is automatically authenticated (see Section III for further details) and logged on. Thus, the user does not have to enter an additional password. After choosing a tour she/he is presented with the overview map as shown in Fig. 2(a). A region of interest can be selected, the user is able to zoom into the detail map (Fig. 2(b)) and then select a PoI for which to get information (Fig. 2(c)). If the user decides to visit the PoI, LoL@ can determine her/his current position and navigate the user through the city of Vienna to the desired PoI. The navigation is assisted visually (Fig. 2(d)) as well as acoustically. When the PoI is found (or whenever the user wants to) textual notes or photos can be stored in the tour diary (Fig. 2(e)). Additional information like the admission fee of an exhibition or some photos to get a first impression can be requested to support the decision whether to visit or rather find an even more interesting PoI.

III. Architecture

Portable devices and mobile computers bring up a new set of problems the entire value chain has to cope with. Point-to-point and end-to-end solutions have been developed to become more independent of the traditionally slower development of the network domain. The open system model of the Internet has enabled everybody who has access to the network to act as an application service provider and has thus speeded up the service development process. However, in the Internet this has also paralyzed business by providing more opportunities for Denial of Service (DoS) attacks or spreading of viruses. Telecommunications in contrast has chosen the contrary way of standardizing the network as well as services in detail by the International Telecommunication Union (ITU) and other dedicated standardization bodies. The network is not constructed to just provide data bearer capabilities, but also contains intelligent elements, providing a set of well-defined telephony services, like call forwarding or barring. Consequently ”dumb” terminals can be used allowing easier interoperability between different telephony networks. The disadvantage of this network model is the cost and delay that occurs upon introducing new services [3]. Our approach to design LoL@ was thus to find an architecture-based solution as a compromise between the open Internet world and the strongly standardized telecommunications domain.

Figure 3 shows the architecture we use for LoL@. LoL@ may use GSM networks (utilizing a Network Access Server (NAS) and circuit switched communication) or packet networks (GPRS, UMTS). A CORBA-based OSA/Parlay API [4], as defined for UMTS, was implemented for localization and call control, allowing a simple connection to other OSA enabled networks (including 3G) without using network specific protocols. Thus, we can easily include features like click-to-dial and user localization into browser based third party Internet applications like LoL@.

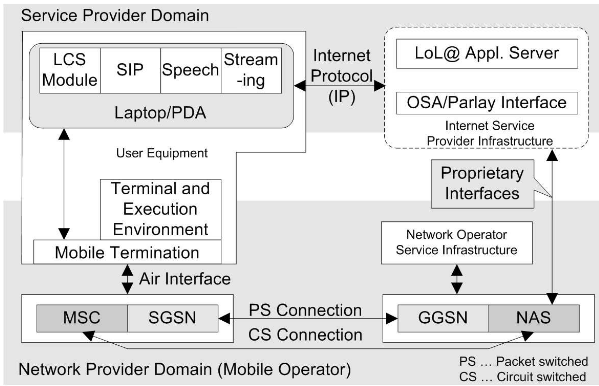


Fig. 3. LoL@ Architecture

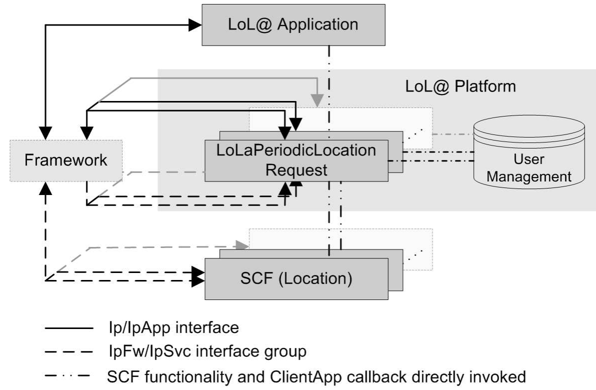


Fig. 4. LoL@ Platform

The implemented architecture of LoL@ represents the entire OSA/Parlay architecture (except the framework) consisting of the Parlay services (service capability features) for localization and call control embedded in service capability servers hosted by the network operator. The main benefit of this type of architecture is the standardized API a third-party service provider can access. In the network layer we use SIP for the call control and a proprietary implementation for the localization capabilities. This was necessary since no localization technology, except cell ID, was available in the operator's network. The Parlay Application Programming Interfaces (APIs) are currently defined in a programming language independent Interface Definition Language (IDL) and will also be available in XML [5]. This allows the creation of a connection between the Parlay service and the Parlay service user via mutually calling distributed objects.

Additionally our architecture shows a platform between the OSA/Parlay API and the third party provider for two purposes. First, we offer simplified interfaces, similar to the currently running Parlay X efforts [6] since the majority of location based applications will require only a subset of the API's functionality to request the location information. Second, we enhance the offered service functionality while still keeping the interfaces compliant with the standard.

We have reduced the set of datatypes by avoiding 5 user defined datatypes and have also reduced the number of parameters (6 instead of 18) significantly. Web programmers, not familiar with accessing telecommunications services have reported less trouble during the implementation of LoL@ and thus the concept ensured a faster and more efficient implementation of the application.

The platform offers a simple third party user management, which coordinates the OSA/Parlay requests, and can compute the speed of a mobile station based on consecutive localization requests. This was implemented for demonstration purpose only, showing that this enhanced service can again provide APIs compliant to the Parlay standard and thus being offered to potential customers in a standardized way. This equals the ”VAS” (Value Added Service) concept described in [5]. The concept and a sample interface (for a simple periodic location request) is shown in Fig. 4. The advantages for a mobile service like LoL@ are (see also [5]):

- Network functionality can be extended, e.g. to offer velocity information. Even the role of a virtual network operator is imaginable.
- Interfaces can be adapted to meet specific constraints of service customers.
- Application users can be managed by the platform. In LoL@'s case this is restricted to the storage of a service access permission, but could be enhanced to a complete user management.

A. *Location Subsystem*

In the traditional world, where computers are not mobile, services do not have to be capable of delivering content based on the current position. In the mobile world however, the space domain can be used to prepare content individually according to the requirements of specific environments. Some can imagine location dependent information support like getting an on-line map showing the nearest restaurant or a push based advertisement showing special offers of miscellaneous stores. Requirement is a location enabled network infrastructure. Two important aspects we will discuss here: (1) the single user log in and (2) a Session Initiation Protocol (SIP [7]) based push architecture to provide location data efficiently.

1) LoL@ Single User Log In: The user has to be authenticated and the application has to be authorized by the network operator to prevent misuse. In the Telco world the SIM (Subscriber Identity Module) is used for

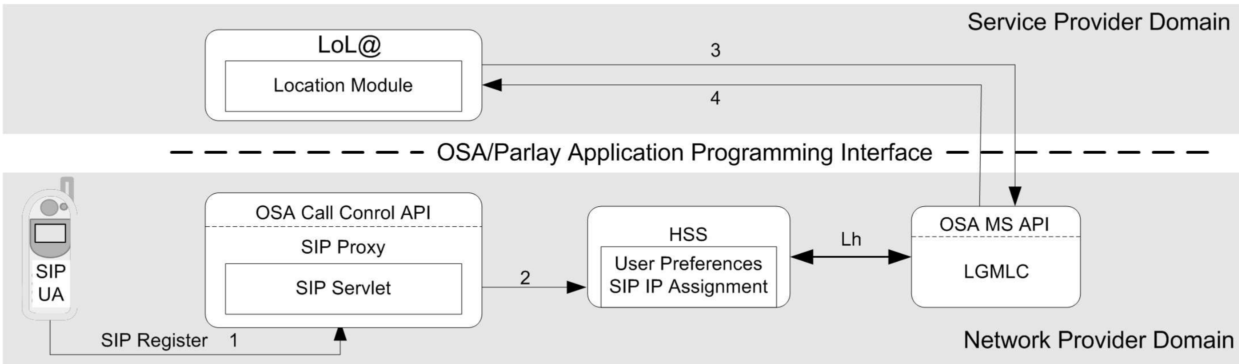


Fig. 5. LoL@ single log in concept

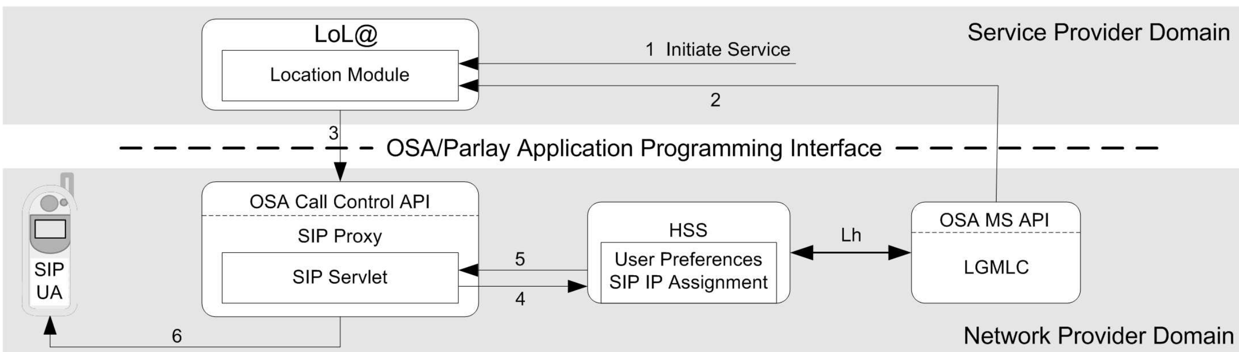


Fig. 6. LoL@ push

authentication and the authorization is stored in a central server, the HLR (Home Location Register). The core network itself does not provide any security mechanisms [3]. Instead it is simply protected by shielding it from the outside world, i.e. only a telephony operator can offer services or connectivity to the network. To allow access to the LoL@ service the user must thus be authenticated and authorized. In contrast to the Internet with repeated username-password log-ins we have implemented a SIP-based sign-on mechanism allowing a user to access all subscribed services after a single log-in when turning on the mobile device. The mechanism is shown in Fig. 5.

A SIP register message including the current network address (IP address) is automatically sent to a SIP proxy (step 1) after the mobile SIP UA (user agent) is turned on. This ensures that the UA is reachable in the corresponding network. Furthermore the currently assigned IP address is stored in the HSS (Home Subscriber Server) (step 2). The HSS thus contains the current SIP-IP assignment and personal settings of the specific user. Whenever a mobile service tries to access network functionality, e.g. for user localization, the HSS is checked and determines whether the service is authorized. Thus the user does not need to log on separately for LoL@. This task is performed automatically when the SIP UA is launched. The user can then access the entire service portfolio of LoL@.

2) LoL@'s SIP based push architecture: Push services use the limited bandwidth of mobile networks efficiently because communication only occurs if there is information available [8]. Thus, push services are especially suitable for the notification of asynchronous events. LoL@ uses this concept for the delivery of location data. Whenever new location data becomes available to the server (via the OSA API) the information is "pushed" to the mobile device. The control flow (Fig. 6) is as follows: The user starts LoL@ and turns on the location detection, which activates the periodic location request (1). If a new location estimate is available, the GMLC notifies the LoL@ application (2). The application determines whether the movement of the user was big enough to send a notification to the LoL@ client running in the user terminal. In this case a SIP push is issued (3). The SIP proxy identifies the request and activates the push servlet which checks the push settings in the HSS (4,5). If push is allowed, the push request is sent to the SIP UA running in the terminal (6). The UA is responsible to forward the request to the local LoL@ applet which updates the display. It is important to note that this push request itself is not location based because the user location is the content of the request. Even though this situation might look like a standard SIP session activation at the first glance, there are three important differences:

1) Push preferences are checked
2) The SIP session is automatically accepted and closed after data delivery.
3) Session data is directly sent to the appropriate application.

IV. Application design concepts

For the LoL@ application itself, we rely on standard Internet protocols and technology, like HTTP, HTML and Java. The network is treated as a bit pipe extended by call control, single sign-on, and location functionality accessed via the OSA API.

Compared to the fixed Internet and its comfortable interfaces, formidable constraints are imposed by mobile devices and mobile networks.

The displays of mobile devices are small compared to desktop computers' displays. The input possibilities are limited. Application design for mobile devices is influenced by these characteristics. The use of the limited input and output media must be optimized to ensure a positive user experience. For the LoL@ demonstrator we assume a screen size of 320x120 pixels and pen input. A multimodal user interface was designed. A detailed discussion of usability issues concerning LoL@'s user interface as well as a documentation of the user interface design process can be found in [9] and [10].

Even if the network link is slow, the application's response time must correlate with the response times and attention spans of human beings. Users are impatient and do not want to wait for the information they have requested. In contrast to Internet applications, where the business logic resides completely at the server, we decided to split the application's business logic between terminal and server. Simple interactions (e.g. the pressing of a button) that cause only small changes in the state of the application can be handled by the terminal part. This minimizes network connections and hence improves the response time of the application. In addition, putting logic at the terminal makes it possible to optimize utilization of network resources. Whenever possible in LoL@'s data flow, we do not rely on the standard HTTP request/reply communication scheme, which sets up a network connection each time the user requests information, but transfer larger information items. Presentation of these data items to the user is then controlled by the terminal part of the business logic. This turned out to be especially useful for the interactive routing process. Given the user's source

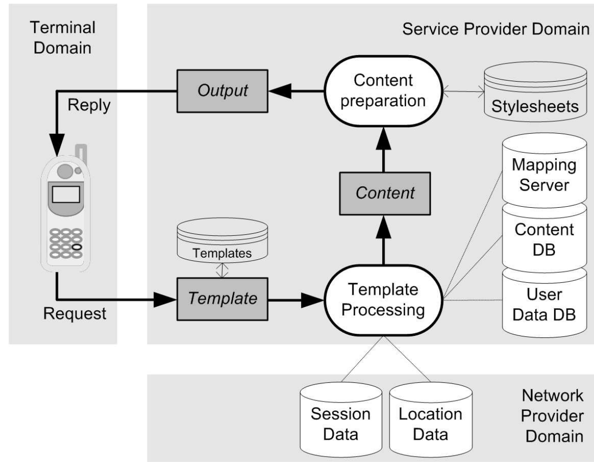


Fig. 7. Content preparation and content delivery

and his/her desired destination position as parameters of the request, all required information is gathered from the server in one chunk of data, which is subdivided and presented subsequently in the vector map and in text boxes at the terminal. The terminal part of LoL@'s business logic is implemented as Java applets on top of a Web browser to minimize the memory footprint of the application. The map viewer component as well as the application's service control facilities (service control buttons, see Fig. 2(a)), which are stored locally at the terminal, can either be downloaded at first use or installed permanently. In the latter case, version management of local components must be considered. This is subject for future work.

To overcome problems with network disconnection (caused either by the user or by the network) we implemented a suspend/resume functionality that allows a user to continue using LoL@ in the same state as before the interruption. Relevant application state information is stored at the LoL@ server, allowing a resynchronization of client and server states upon a client restart. Using SIP addresses for user tracking is secure compared to the unreliable mechanisms available in Internet applications (hidden HTML form fields, URL rewriting or cookies) where data about the client's state is transferred between client and server in a way that can be manipulated by the client. Since state synchronization may not happen permanently because of the limited bandwidth, only major changes in user state completely done at the client (e.g. switch between map and text view) are reported to the server for resume. Minor changes are not reported and therefore lost when resume occurs.

For the server part of LoL@'s business logic, it is necessary to build a scalable architecture that is able to adapt and deliver content to a wide variety of mobile devices. A content delivery system to support different Java-based terminal types was designed (see Fig. 7). Device independence is achieved by uncoupling application data and presentation of that data. The design strictly separates

- content retrieval (business logic) and
- content presentation (presentation logic).

The *business logic* component is responsible for collecting content from different data sources within the service provider domain, e.g. a database with tourist attractions, a routing server, or a video streaming server. The LoL@ user interface including its data flow is completely defined using templates in XML [11] format. The template-based approach results in an extensible and easy-to-maintain system that supports adding new data sources and subsequently defining a user interface for these new features. Changing the application's user interface by editing the templates is a straightforward task.

The templates do not define the design of the screens, but rather their structure, i.e. the data items necessary. A template contains commands, links, and parameters. Commands are used to define which data is needed and from which data source it has to be fetched. Links describe LoL@'s data flow, i.e. hypertextual structure. The HTTP parameters of the request are needed for processing. Template processing consists of three steps:

1) Substitute the parameter placeholders in the template with their actual values.
2) Execute commands, e.g contact the various data sources to to fetch the data.
3) Check links: If no content is available for a link, it is removed.

The results of template processing are filled-out templates which contain all data to satisfy the user's request.

In the next step, the *presentation logic* component prepares the content data according to the viewing device's capabilities. Different presentations of the same data are delivered to clients depending on which presentation is suitable for being displayed on the terminal. XSL [12] stylesheets define which layout information is added to content data during the XSL transformation process. Different output formats, like WML [13] or HTML for different display sizes, can be supported by creating appropriate XSL stylesheets. Integration of a content negotiation system like CC/PP or UAProf is subject for future work. More information about the content delivery system can be found in [14].

V. Conclusions and Future Work

In this paper we have introduced LoL@, the Local Location Assistant, a mobile tourist guide. We have demonstrated that the combination of standardized internet technologies on top of wireless networks is a promising concept for future wireless services. Standardized interfaces to create a functional split between the network and service providers, but open enough to keep it flexible for new ideas, are necessary to allow fast and efficient service development. In the near future more powerful mobile devices will allow a wide spreading of multifunctional services exceeding common ones like simple data download, voice traffic or MMS. LoL@ demonstrates a promising concept, but further work will be necessary to allow a higher personalization regarding content as well as the presentation of the content. Network functionality must become easily accessible to service providers while security e.g. for push services must be ensured.

References

[1] Wapforum, "Wireless application environment specification," 1999, http://www.wapforum.org.

[2] 3GPP TSG Services and System Aspects, "TS 22.057: Mobile Execution Environment (MExE), Service description, Stage 1 (Rel. 4)," April 2001, http://www.3gpp.org/ftp/specs/2001-03/Rel-4/22 series/22057-400.zip.

[3] G. Pospischil, "Application development for UMTS," Ph.D. dissertation, Vienna University of Technology, Department of Communications and Radio-Frequency Engineering, October 2002.

[4] 3GPP TSG Services and System Aspects, "TS 22.127: Stage 1 Service Requirement for the Open Service Access (OSA) (Rel. 4)," April 2001, http://www.3gpp.org/ftp/specs/2001-03/Rel-4/22 series/22127-410.zip.

[5] S. Bessler and J. Zeiss, "Eine Offene Service Plattform für Telekomanwendungen," *Praxis der Informationsverarbeitung und Kommunikation*, vol. 4, pp. 202–207, 2002.

[6] Parlay X Working Group, "Parlay APIs 4.0, Parlay X Web Services White Paper - Version 1.0," December 2002, http://www.parlay.org.

[7] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg, "SIP: Session Initiation Protocol," *RFC 2543, IETF*, March 1999, http://www.ietf.org.

[8] G. Pospischil, J. Stadler, and I. Miladinovic, "Location based push architectures for the mobile internet," November 2002, to be published in R. Prasad ans S. Dixit (ed.): Wireless IP Evolution, Artech House, Norwood, USA.

[9] G. Pospischil, M. Umlauft, and E. Michlmayr, "Designing LoL@, a mobile tourist guide for UMTS," in *Proc. 4th Symp. Human-Computer Interaction with Mobile Devices*, September 2002.

[10] M. Umlauft, G. Pospischil, G. Niklfeld, and E. Michlmayr, "LoL@, a mobile tourist guide for UMTS," *Journal of Information Technology & Tourism*, 2003.

[11] T. Bray, J. Paoli, C. M. Sperberg-McQueen, and E. Maler, *Extensible Markup Language (XML) 1.0*, 2nd ed. World Wide Web Consortium, October 2000, http://www.w3.org/TR/2000/REC-xml-20001006.

[12] J. Clark, *XSL Transformations (XSLT) Version 1.0*. World Wide Web Consortium, November 1999, http://www.w3.org/TR/1999/REC-xslt-19991116.

[13] *Wireless Markup Language version 1.3 Specification*, WAP Forum, http://www.wapforum.org/.

[14] E. Michlmayr, "Flexible Content Management for the LoL@ UMTS application," Master's thesis, Vienna University of Technology, May 2002.

Feasibility of Fixed Wireless Access

Ali Eftekhari, R. Ryan Lewis, Scott Pollock, Tom Lookabaugh, Bryan Weber

Interdisciplinary Telecommunications Department
University of Colorado
Campus Box 530, Boulder, CO 80309
Corresponding Author: rlewis@colorado.edu
+1 (303) 492-3686 (phone) +1 (303) 492-6772 (fax)

Submitted: January 21, 2003
International Symposium on Advanced Radio Technologies
4-7 March, 2003, Boulder, Colorado USA

Topic: Fixed Wireless Broadband

***Abstract*- Fixed wireless access has the unrealized potential to play a major role in the adoption of broadband in the U. S. We conduct a case study of the feasibility of business in a suburban U. S. community and conclude that with a preference for the use of licensed spectrum, such a small scale, grass roots business would be feasible. This opens the question of promoting the economic availability of licensed spectrum in areas smaller than a Basic Trading Area (BTA).**

I. INTRODUCTION

Fixed wireless access (FWA) has not been particularly visible or successful to date as an option for providing broadband Internet access; it contributes a very small fraction of such lines and most potential subscribers are simply unaware of it [1]. Notwithstanding some novel business and technology efforts in unlicensed spectrum, attempts to launch businesses in licensed spectrum have been subject to numerous setbacks and reconsiderations, and are under the constant threat that spectrum will be repurposed for other applications such as backhaul for mobile telephony networks or directly for mobile use [2,3]. Nonetheless, fixed wireless has the potential to play various important roles in the adoption of broadband and in fulfilling steadily increasing demand [1].

In this paper we attack in particular the question of business feasibility of fixed wireless access through a case study. We carefully examine the technology, market, and business drivers for fixed wireless access in a particular suburban Denver community. We demonstrate a preference for the use of licensed spectrum if available, and show that a successful small but scalable local business with limited capital requirements could be established by entrepreneurs in areas like the subject of our study. The possibility of encouraging grass roots bootstrapping of fixed wireless access businesses is an exciting one, but our choice of licensed spectrum presents a critical conundrum for spectrum policy: how to promote regulation or practice in spectrum licensing that allows the economic allocation of spectrum in areas geographically much smaller than Basic Trading Areas.

II. METHOD AND SCOPE

The method of this paper is to analyze the background information of the regulatory, technical, economic and interdependent factors pertinent to FWA. After creating a common understanding of the FWA environment, a feasibility study of each driving factor is provided. To generate additional insight for the project, a broadband market survey and technical (physical) site survey specific to a particular location were conducted. We supplemented our analysis with interviews with experts in the FWA sector.

We are interested in suburban residential U. S. markets and to for-profit businesses (rather than co-ops, government run networks, or other organizations). Superior, CO, a suburban town of about 9,000 located 20 miles Northeast of Denver, was used as the case study for the research project. Although this case study may be relevant to many other suburban areas in the United States, it is important to note that every market area will have its own nuances.

III. ANALYSIS OF THE FIXED WIRELESS ACCESS ENVIRONMENT

Regulatory Factors and Analysis

The Federal Communication Commission (FFC) allows FWA services to operate in three licensed frequency bands: Multipoint Multichannel Distribution Service (MMDS), Instructional Television Fixed Service (ITFS), and Local Multipoint Distribution Service (LMDS). Each of the service licenses are distributed throughout the U.S. in 493 Basic Trading Areas (BTA). Tables 1-3 briefly describe each of the licensed frequency bands.

Table 1. MMDS Services Information [4]

Multipoint Distribution Service	
MDS is a commercial service generally used to provide multichannel video entertainment programming and is often referred to as "wireless cable." Wireless cable is similar to land based cable television.	
Also Known As	MMDS (Multichannel) Wireless Cable
Established	1970
Service Rules	C.F.R., Part 21
Spectrum / Frequencies	
Band	2.1-2.2 GHz, 2.5-2.7 GHz
Channels	13
Auctions	
6 - Multipoint/Multichannel Distribution Services 11/13/1995 - 3/28/1996	

Table 2. ITFS Services Information [4]

Instructional Television Fixed Service	
ITFS is used to provide educational instruction and cultural and professional development in schools and other institutions. ITFS may be leased to companies offering subscriber-based services, provided usage requirements are met.	
Established	1963
Service Rules	C.F.R., Part 74
Spectrum / Frequencies	
Band	2.5-2.7 GHz
Channels	20

Table 3. LMDS Services Information [5]

Local Multipoint Distribution Service	
A local multipoint distribution system (LMDS) is capable of offering subscribers a variety of one and two-way broadband services, such as video programming distribution; video teleconferencing; wireless local loop telephony; and high speed data transmission, e.g. internet access. Because of its multi-purpose applications, LMDS has the potential to become a major competitor to local exchange and cable television services.	
Spectrum / Frequencies	
Band	Block A: 27,500-28,350 MHz; 29,100-29,250 MHz and 31,075-31,225 MHz Bands Block B: 31,000-31,075 MHz and 31,225-31,300 MHz Bands
Auctions	
17 – Local Multipoint Distribution Services February 18, 1998 - March 25, 1998	

Currently the two unlicensed frequency bands available are: Industrial, Scientific and Medical (ISM) and Unlicensed National Information Infrastructure (U-NII) "support spread spectrum operation on a non-interference unlicensed basis" [6].

The [FCC] has provided for the operation of low power unlicensed devices under part 15 of the rules. Devices operating under part 15 must meet technical standards that are designed to control harmful interference to radio communications services. Users must correct any harmful interference that may occur and must accept any interference that is received [7].

The U-NII spectrum is subdivided into three sections of which the "5.725 - 5.825 GHz portion of the band is intended for community networking communications devices operating over a range of several kilometers. The FCC permits fixed, point-to-point U-NII devices to operate with up to 1-W transmitter power and directional antennas with up to a 23-dBi gain" [6].

Policies that Supersede State and Local Policies

The Telecommunications Act of 1996 contains important provisions concerning the placement of antenna structures or towers that provide personal wireless services[1] and most local communities have worked closely with personal wireless service providers to place such facilities within their localities. Section 704 oversees federal, state, and local government issues regarding personal wireless service facilities. "It also prohibits any action that would ban altogether the construction, modification or placement of these kinds of facilities in a particular area and requires the federal government to take steps to help licensees in spectrum-based services …get access to preferred sites for their facilities" [8]. Section 704 of the Telecommunications Act of 1996, "also directs the [FCC] to offer assistance to state and local governments in resolving wireless facility sitting issues" [9]. Section 207 of the Telecommunications Act of 1996 pertains to the policy of attaching antennas to a residential property.

Technical Factors and Analysis

The technical factors involved in implementing a FWA solution must be examined to build a sound infrastructure. The predominant technologies available for a FWA solution are primarily based on the Institute of Electrical and Electronics Engineers (IEEE) standards 802.11a, 802.11b and 802.16. The essential assessment of characteristics such as: coverage, speed and capacity, reliability, and security can ultimately assist the determination of the effectiveness and efficiency of deployment.

Table 3 summaries the major technical characteristics of the four primary FWA solutions.

[1] "Personal wireless services" include commercial mobile services, unlicensed wireless services, and common carrier wireless exchange access services.

Table 3. Technical Characteristics of FWA Solutions [10][11]

Common Name	Wi-Fi	Wi-Fi5	MMDS	LMDS
IEEE Standard	802.11b	802.11a	802.16.3	802.16.1
Frequency Bands (US)	2.4 - 2.43825GHz	5.15 - 5.35GHz 5.725 - 5.835GHz	2.1 - 2.2GHz 2.5 - 2.7GHz	27.5 - 28.35GHz 31.0 - 31.30GHz
Speed and Capacity				
Data Rates (theoretical)	1-11Mbps	6-54Mbps	3Mbps-1Gbps	5Mbps-1.25Gbps
Channels and channel size	12 channel; 6Mhz 3 non-overlapping	8 non-overlapping 6Mhz	13 channels; 6Mhz 4 non-overlapping	up to 1.3GHz
Coverage				
Distance (max.)	10 miles	3-5 miles	35 miles	3-5 miles
Line-of-sight (LOS)	Near-LOS	Near-LOS	Non-LOS	LOS
Reliability				
Weather	Minimal Interference	Minimal Interference	Minimal Interference	Large Interference
Interference	Highly Congested	Minimal Congestion (new technology)	No Congestion; licensed spectrum	No Congestion; licensed spectrum

Multiple Access Techniques

At this time the two primary techniques that are used in non-line-of-sight (non-LOS) FWA solutions are Code Division Multiple Access (CDMA) and Orthogonal Frequency Division Multiplexing (OFDM). CDMA is highly dependent on "bandwidth spreading for both multiple access, as well as, coverage gains - the higher the spreading factor, the larger the cell area that can be covered" [10]. With CDMA, in order to deliver users with the same data rates throughout a cell, it must "claim spectrum reuse of one" [10]. By setting the spectrum reuse to one, CDMA effective coordinates all cells to use the same RF frequency. The underlying problem with CDMA is that as the number of users increase, the more the cell coverage (radius) decreases.

OFDM is an efficient modulation technique as it "takes a broadband data pipe and distributes it among many parallel bins, the exact number being a function of the Fast Fourier Transform (FFT) size [10]. OFDM is less susceptible to multi-path interference because this technique uses the parallel bins which allow the symbol duration to be longer, thus making it less susceptible to multi-path interference [10]. Table 4 summaries the major characteristics of OFDM and CDMA.

Table 4. Characteristics of Media Access Techniques [12]

	OFDM	CDMA
Spectral Efficiency	More efficient	Less efficient
Multipath	Handles larger number of paths	Some diversity benefit; performance sensitivity to number of paths coherently combined
Multiple Modulation Support	Downlink and uplink	Downlink only
Resistance to Narrowband Interference	Not as robust as CDMA	Spread spectrum provides protection
Network Planning	Cell overlays simple to do; frequency planning	Difficult for cell overlays; PN offset planning; cell-breathing complications
Power Control	Usually employed but not a fundamental requirement	Required function
Peak-to-Average Ratio	10-12 dB	Varies; can be as high as 11 dB
Standards Adoption	WLAN, IEEE, 802.16, 4G	WLAN, 3G

Economic Factors and Analysis

Broadband Industry Overview

The recent downturn of the US economy and more specifically the technology sector may perhaps lead one to the conclusion that the adoption of broadband adoption is also slowing. As for the first six months of year 2002 this conclusion appears to be false. Between January and June of 2002 the number of broadband subscribers has increased by 25 percent to nearly 24 million households [13]. Over the last two years, even with the economic slow down, broadband access has increased by four times the number of subscribers in 2000 [13]. This rapid growth rivals the CD player and PC while surpassing the growth rate of VCRs and color televisions [13].

The price for broadband access remains an important issue. Studies have shown that demand for broadband is very price elastic [1]. Recent price reduction by British Telecommunication (BT) caused demand for ADSL to double [14]. With current prices of cable and DSL access increasing in the US, this fact could be detrimental to broadband adoption. Most of these increases are small, 1.4 percent to 5 percent increase, so the effect might not be as drastic as in the BT scenario, 14 percent decrease [15]. With wireless Internet service providers (ISP) just entering the market, there is inconclusive data about its pricing. Generally broadband costs range from $45 to $55, regardless of transmission media.

Predicaments with Previous FWA Service Providers

While most of the recent efforts to provide a viable FWA solution were started by small, competitive companies, Sprint was one of the first to enter the market. The primary obstacles encountered were: scalability, LOS issues, expensive CPE costs, and the need for professional installation. These issues made the viability of the business model next to impractical with then current technology [16].

Interdependent Factors and Analysis

A principle argument for providing FWA service is whether to operate in the licensed or unlicensed spectrum. By analyzing the regulatory, technical, economic and interdependent factors, we make a subjective determination of preferred frequency band in Table 5 based on currently available equipment.

Table 5. Comparison of Licensed versus Unlicensed Spectrum

Metric	Weight	Licensed Spectrum		Unlicensed Spectrum	
		MMDS	LMDS	ISM	UNII
Regulatory Factors	**0.10**				
Acquiring Spectrum	0.05	**4** ±1	**4** ±1	**6** ±2	**6** ±2
Legal Protection from Interference	0.05	**9** ±1	**9** ±1	**2** ±1	**2** ±1
Technical Factors	**0.40**				
Performance (Speed and Capacity)	0.1	**5** ±2	**9** ±1	**5** ±1	**6** ±2
Line-of-Sight (LOS)	0.1	**7** ±2	**4** ±1	**5** ±2	**5** ±2
Coverage	0.05	**7** ±2	**3** ±1	**6** ±1	**5** ±1
Interference	0.05	**8** ±1	**8** ±1	**2** ±2	**4** ±2
Reliability	0.05	**7** ±2	**7** ±2	**7** ±1	**7** ±1
Security	0.05	**5** ±1	**5** ±1	**3** ±1	**3** ±1
Economic Factors	**0.20**				
CPE Cost (Variable)	0.10	**5** ±2	**2** ±1	**8** ±1	**6** ±2
Infrastructure Cost (Fixed)	0.05	**5** ±2	**3** ±1	**7** ±1	**6** ±1
Cost for Spectrum	0.05	**4** ±2	**4** ±2	**9** ±1	**9** ±1
Interdependent Factors	**0.30**				
Ease of Infrastructure Deployment	0.05	**5** ±2	**2** ±1	**5** ±2	**5** ±2
CPE Installation Ease	0.15	**8** ±2	**3** ±1	**7** ±1	**7** ±1
Scalability	0.05	**7** ±1	**3** ±2	**4** ±1	**4** ±1
Total (Range)	**1.00**	**(4.75 - 7.85)**	**(3.30 - 5.70)**	**(4.05 – 6.55)**	**(3.75 – 6.65)**
Total (Uncertainty)	**1.00**	6.30 ± 0.52	4.50 ± 0.40	5.30 ± 0.59	5.20 ± 0.64
Key: Quantitative Score: Poor – 0 … Excellent – 10 (Uncertainty)					

After analyzing the different factors when providing a FWA solution, at this time it is evident the best overall approach is to operate in a licensed MMDS spectrum.

IV. FEASIBILITY STUDY OF FIXED WIRELESS ACCESS

Regulatory Feasibility

Interviews with Experts

After conducting interviews with expert faculty members such as: Dr. Ray Nettleton, Professor Dale Hatfield, and Dr. Douglas Sicker at the University of Colorado at Boulder Interdisciplinary Telecommunications Department, one can conclude that there are no regulatory issues that are restraining FWA in the U.S. market.

As Professor Hatfield stated there are no regulatory issues because "all regulatory winds are on [our] back …[he] wouldn't worry about the regulatory" issues when considering the feasibility of entering the FWA market [17]. Dr. Nettleton agrees with this point of view and he also adds that from the regulatory point of view, one can do whatever is necessary to enter the market as long as the FCC guidelines are followed. All three interviewees also agree that FCC has done everything possible to try to bring small businesses into this market in order to promote competition. As Dr. Sicker mentioned in the interview, the FCC is still putting a lot of effort into promoting the entrance of small businesses into this market arena [18].

Local Policy Implications

The town of Superior, CO authorities were well aware that they do not have jurisdiction over the CPEs including antennas due to the FCC allocating authority to Home Owners Associations (HOA) across the US [19]. At this time there are three HOAs located in the Superior, each having distinctive policies regarding attaching CPEs or antennas to residential properties. In the Saddle Brooke community, any and all external antennas are prohibited [20]. In the Horizons community, external antennas are permitted as long as they are not attached to any part of the building and the antenna itself cannot extend beyond the balcony area [21]. As for the other residential properties within the Rock Creek area, there are no restrictions for putting up any kind of external antennas within the homeowner's property. The matters surrounding HOA policies and guidelines need to be considered when deploying a FWA solution that requires attaching a CPE or antenna to the residential property.

Technical Feasibility

To better determine the technical feasibility of a FWA solution, the software application PlaNET was used to model the LOS and propagation in the Superior, CO area. The model predictions illustrates that it is technically feasible as more than a majority of the Superior area would be covered with LOS transmission. In addition, a wireless site survey was conducted within the Superior area using unlicensed wireless transmission analysis tools Grasshopper, and AEROPEAK software, a wireless packet sniffer to determine a spectrum use and analysis.

Reliability

In order to effectively provide a reliable architecture, a site survey was conducted to determine the quality of service (QoS) at the receiver. QoS is determined by the signal-to-noise ratio (SNR), bit error rate (BER), and available bandwidth. The site survey concluded that the noise levels in the Superior region were extremely high, resulting in a low SNR. Furthermore, Superior has mass deployment of wireless Access Points causing tremendous amount of interference in the 802.11b bands. This in turn, increases the BER and decreases the available bandwidth.

Line-of-Sight Considerations

The PlaNET propagation model indicated that 90 percent of Superior is covered by LOS. This is acceptable for utilizing technologies such as MMDS, because it has the ability to offer near-LOS, wireless coverage. Near-LOS can only be obtained in lower frequencies, less than 10GHz, because the wavelength becomes greater as frequency is decreased; furthermore, near-LOS utilizes modulation techniques, OFDM or CDMA, that allow for enhanced signal propagation to overcome environmental and man made obstacles.

Weather and Climate Considerations

The geological positioning of Superior demonstrates some consideration resulting from the semi-arid mountainous region in which it resides. However, the yearly average precipitation rate for Colorado is 17 inches, resulting in minimal distortion of wireless signals throughout the year. The higher the frequency, the greater the distortion from rain fade and snow have on wireless signal [11]. The amount of attenuation from climate conditions dramatically decreases with decreasing frequency; therefore, the 10GHz range will have minimal attenuation from climate conditions; to illustrate, Rob Flickenger, author of "Wireless Community Networks", concluded that a one percent attenuation factor from rain must be applied to frequencies less than 10GHz [11].

Interference Considerations

Interference considerations have an enormous impact on the ability to provide a reliable connection. A site survey for the area of Superior was conducted to map the wireless concentrations with a "Grasshopper WLAN IEEE 802.11 Tester", and AEROPEAK software. The results from the grasshopper demonstrated that there was a large amount of noise interference being projected over the entire area. As noise levels increase, the signal degrades exponentially, as observed by the SNR. There are five reasons why the SNR would have such a dramatic effect on the signal quality: the transmitter input gain maybe too low, the transmitter distance maybe too far away to obtain a quality signal, environmental RF noise maybe centered near the receiver, multi-path phase cancellation is occurring at the receiver; and obstructions that reside between the transmitter and receiver [22].

In addition to PlaNET, we conducted an assessment by sniffing throughout the region for wireless AP's and traffic watches using "AEROPEAK Software". After conducting this survey, it was obviously apparent that the 802.11b spectrum was completely saturated; to illustrate, the tests indicated that there were over 100 Access Points being used causing high bit error rates and low signal to noise ratios. This is an early indication that the market price for 802.11 devices has reached the saturation point, and the situation will only get more congested from this point forward. Since QoS is highly diminished due to congestion, licensed spectrum is a more viable solution for the Superior area.

Economic Feasibility

Market Analysis

In order the accurately gauge the demand and willingness to pay for broadband in Superior we conducted a door-to-door survey. The survey focused on the customer's willingness to pay for monthly and startup costs. With 68 percent of the people surveyed willing to pay at least $30 for monthly service and 27 percent willing to pay over $50 a month for service, our survey findings show that there is a large market demand for broadband in this area. Even though incumbent DSL and cable broadband is not offered in this area, alternative forms of broadband are available. Currently 56 percent of people surveyed do not have broadband access. Of this group, 59 percent are willing to pay more than $30 for broadband access.

Besides the monthly cost of broadband, another important issue when starting an ISP is the customer startup cost. The startup cost includes equipment and installation charges. Our results show that 76 percent of the people surveyed are willing to pay more than $50 for installation. More importantly, 67 percent responded that they would rather purchase startup equipment for $200 than rent equipment for $10 a month. This helps to reduce the acquisition cost of a customer.

Acquisition Costs

For a wireless ISP, the cost of acquiring a customer will be lower than that of DSL and cable providers. This is caused by the reduced cost of network infrastructure when obtaining new customers [23]. Current acquisition costs for ISPs average about $400 [24]. This price can be reduced with customer's willingness to purchase startup equipment and self installs. Recently Bell South had a 90 percent success rate with DSL self installs [25].

Feasibility of Interdependent Factors

Scalability

One of the interdependent factors Professor Hatfield stated that must be overcome is the issue of scalability [17]. Scalability is predominately a technical and economic issue, but one cannot overlook the regulatory aspects as well. To overcome the technical issues of scalability, a microcell or cellular approach can be used. The mobile wireless communication industry is a proven model for applying the cellular concept because it can utilizes the spectrum efficiently with frequency reuse patterns and increase user capacity by employing sectoring and repeater techniques.

Backhaul is both a technical and economic issue when considering operating a FWA service. By utilizing LDMS for the backhaul from the base stations (microcells) to a Tier 1 ISP using the national transport infrastructure from Level3, both the technical and economic constraints can be overcome. LMDS has high data rates (155 Mbps) that could provide adequate speeds for backhaul and can be relatively easily deployed for a backhaul network infrastructure. More importantly it provides better cost savings compared to other backhaul access solutions such as T1, T3, fiber and satellite.

Franchise Business Model

When considering the feasibility of providing a FWA service the unsound business model is the primary economic issue that must be overcome. "This technology would seem the perfect choice for all WISPs. However, unless a service provider has a long line of investors behind them, there is very little opportunity to make a business case out of the MMDS proposition" [26]. This economic barrier can be overcome by using a franchise concept to provide FWA service. After establishing a sustainable business model, a franchise can share the startup costs and risks while expanding into different service or market areas. A prime example of a predominant service provider that effectively demonstrates the service franchise concept is McDonalds.

Multi-mode CPE

Another approach to solving the business case problem is the proposition of starting up the FWA service in the unlicensed spectrum. After establishing financial stability one could then switch to the licensed spectrum for potentially better long-term success. The key to making a successful migration from the unlicensed to the licensed spectrum is no end-user (CPE) intervention. The mobile wireless communication industry is another prime example that supports this theory, a mobile service providers requested equipment provider to develop and manufacture dual-mode and tri-mode mobile (cellular) phones.

Spectrum Availability

Having reviewed many technical, regulatory, and economic issues, and determined a preference for licensed spectrum, we come to the critical question of spectrum availability. MMDS and ITFS bands are currently underutilized in most urban areas and should nominally be available in almost any BTA at a discounted rate. Greg D. Widroe at Media Venture Partners states, "it is difficult to determine the price of buying or leasing spectrum in a service area, but a good estimate would be 25-50 cents per household" [27]. Dr. Nettleton noted the price two years ago was $3-5 per household [28], indicating the relative absence of other high value economic uses having emerged to sustain valuations of the technology bubble of the late '90s. Critically important, though, is that licensed spectrum trades in relatively large BTA's; the BTA relevant to Superior, CO, for instance, covers the entire Denver metro region and beyond. There is no evidence that current BTA license holders are willing to geographically sub-license spectrum. FCC spectrum reform policy is moving towards more aggressive sharing geographically and temporally of spectrum [29] and success in this activity and its application to MMDS frequencies is likely a pre-requisite for the success of grass-roots small fixed wireless access businesses within licensed spectrum such as the one envisioned in this paper.

V. CONCLUSION

To conclude, after analyzing the FWA environment and conducting a feasibility study on the regulatory, technical, economic and interdependent factors it, appears that a FWA solution for residential suburban areas is feasible and scalable if licensed spectrum is economically available in limited geographic areas. This report provides a definitive franchisable business solution using the MMDS licensed spectrum in a microcell approach with LMDS backhaul and multi-mode consumer premises equipment.

While we are encouraged by the evolution of technical solutions and market demand for broadband, we see this as calling for efforts on the part of policy makers and interested industry parties to foster the availability of licensable spectrum in more limited geographic areas than current BTA's.

VI. REFERENCES

1 "Background for Broadband Demand Project" Tom Lookabaugh, July 5, 2002 (Can be found at: http://itd.colorado.edu/lookabaugh/Documents/Background%20for%20Broadband%20Demand%20Project.pdf)

2 "Sprint Revives MMDS Plans," Telephony, May 13, 2002. Online at http://currentissue.telephonyonline.com/ar/telecom_sprint_revives_mmds/. Last Accessed on January 21, 2003.

3 "Will WorldCom Departure Mean Farewell to MMDS?" Wireless Week, May 6, 2002. Online at http://www.wirelessweek.com/index.asp?layout=story&articleid=CA215731. Lase Accessed on January 21, 2003.

4 FCC, "ITFS & MDS Radio Services", November 2,2002 (Can be found at: http://wireless.fcc.gov/services/itfs&mds/)

5 FCC, "Auction 17: Local Multipoint Distribution System (LMDS)", November 2,2002 (Can be found at: http://wireless.fcc.gov/auctions/17/factsheet.html)

6 "Band Information", November 4, 2002 (Can be found at: http://www.comsearch.com/broadband/design1.jsp)

7 "Spectrum Policy Task Force Announces Anelists For The August 1st Public Workshop On Unlicensed Spectrum And Experimental Licenses And The August 2nd Public Workshop On Interference Protection", Federal Communications Commission, July 26, 2002 (Can be found at: http://hraunfoss.fcc.gov/edocs_public/attachmatch/DA-02-1816A1.pdf).

8 "FACT SHEET", Federal Communications Commission, April 23, 1996. pp. 1-4 (Can be found at: http://wireless.fcc.gov/siting/fact2.pdf). Visited on 24 October 2002.

10 Shtrom, Victor, "CDMA vs. OFDM in Broadband Wireless Access", Broadband Wireless Business, Volume3 Number 5, August 2002 (Can be found at: http://www.nextnetwireless.com/assets/news/pdfs/CDMAvOFDM.pdf)

11 Rob Flickenger; "Wireless Community Networks"; November 2001

12 "A Comparison of Technologies for Fixed Broadband Wireless Access: OFDM and CDMA", March 22, 2002.

13 Collins, Jonathan, "US broadband shows strong growth", October 28, 2002 (Can be found at: http://www.vnunet.com/News/1132963)

14 Arnott, Sarah, "ADSL price cut doubles demand for broadband", October 28, 2002 (Can be found at: http://www.vnunet.com/News/1132346)

15 Richardson, Tim, "US suffers broadband inflation", October 28, 2002 (Can be found at: http://www.theregus.com/content/6/25106.html)

16 Rush, Sarah, "Sprint evaluates broadband wireless technologies", October 28, 2002 (Can be found at: http://www.broadbandweek.com/news/020506/020507_wireless_sprint.htm)

17 Hatfield, Dale. Private interview, October 8, 2002.

18 Sicker, Douglas Dr. Private interview, October 24, 2002.

19 William, Bruce E. Personal interview, October 24, 2002.

20 Small, Ron. Personal interview, October 24, 2002.

21 Walsh, Sue. Personal interview, October 24, 2002.

22 "Signal to Noise Ratio", November 4, 2002 (Can be found at: http://www.lectrosonics.com/wg/wgs-n.pdf)

23 Market Research Reports, "Fixed Broadband Wireless 2001-2006", October 29, 2002 (Can be found at: http://www.insight-corp.com/fixed.html)

24 Pierce, Lisa, "What the cost of customer churn means to you", October 29, 2002 (Can be found at: http://www.nwfusion.com/columnists/2001/1112eye.html)

25 Cap Gemini Ernst & Young, "The broadband impact", October 29, 2002 (Can be found at: http://www.cgey.com/tmn/pdf/Broadband_Impact.pdf)

26 Monro, Ken, "The Promise of the U-NII Bands", November 4, 2002 (Can be found at: http://www.shorecliffcommunications.com/magazine/volume.asp?vol=21&story=199)

27 Widroe, Greg. Private interview, October 24, 2002.

28 Nettleton, Ray. Private interview, October 16, 2002.

29 "Spectrum Policy Task Force Report," Federal Communications Commission, ET Docket No. 02-135, November 2002, Available Online at http://www.fcc.gov/Daily_Releases/Daily_Business/2002/db1115/DOC-228542A1.pdf. Last Accessed January 21, 2003.

Broadband Demand and Wireless

Tom Lookabaugh, Scott Savage, and Douglas C. Sicker

Interdisciplinary Telecommunications Department
University of Colorado
Campus Box 530, Boulder, CO 80309
Corresponding Author: Tom.Lookabaugh@colorado.edu
+1 (303) 492-0429 (phone) +1 (303) 492-1112 (fax)

Abstract- **Notwithstanding the inconclusive status of the public policy debate on intervention in support of broadband Internet access, wireless broadband can play an important role in the overall adoption of broadband in the U. S. Results from a recent consumer survey of broadband demand attributes informs here a discussion of wireless broadband, including the key role of reliability, wireless as an element of government competition policy, spectrum policy reform, fostering of novel decentralized access services, the role of wireless in closing the digital divide, the use of mobility to expand applications' user communities, and wireless broadband in public safety.**

I. The Broadband Policy Debate

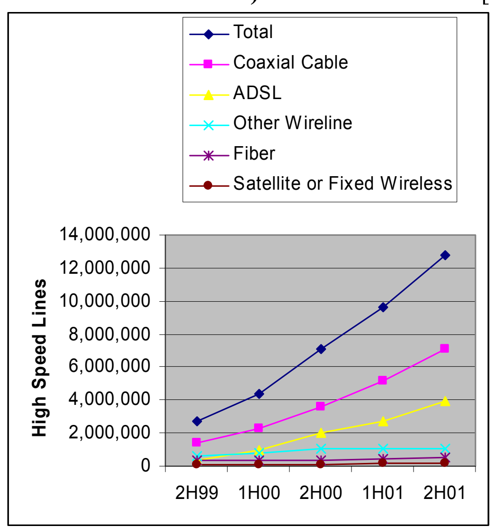
Broadband Internet access suffers from a multiplicity of definitions [1]. Usually, bit rate is taken as the defining characteristic with definitional rates ranging from 200 kbps to 100 Mbps. Alternatively, broadband is sometimes defined in terms of the applications it enables, or other characteristics than bit rate (such as "always on" connectivity or low latency). We will use the U. S. Federal Communication Commission's definition of "high speed lines" – at least 200 kbps in at least one direction. This definition fits many actual and proposed cable modem, DSL, fixed wireless, and satellite Internet access solutions.

Although exact capabilities vary by service provider and the type of service purchased, broadband Internet access is often used for web browsing with quick response times, transfer of crisp video images and CD quality audio files over the Internet, quick downloads of large files, playing real-time interactive games with people in different locations, and providing efficient access for others on the Internet to large audio, text and video stored on local PCs or other devices.

A majority of Americans use the Internet; however, a smaller fraction is using broadband, on the order of 15% of households [2]. The U. S. public debate on broadband demand policy revolves around two main positions: (1) broadband demand is growing too slowly for the nation's good (implying that intervention is in order) or (2) broadband demand is growing at a reasonable rate (no intervention is in order). Figure 1 shows the number of "high speed lines" from FCC data for various access technologies.

Clearly, subscription has been growing, particularly for cable modem and DSL technologies. Fixed wireless and satellite have relatively low penetration.

Figure 1. High Speed Lines (greater than 200 kbps in at least one direction) in the United States [3]**.**



The rate of growth is a subject of concern, however. The semester over semester growth rate in Figure 1 dropped from approximately 60% in both semesters of 2000 to approximately 35% in both semesters of 2001. The fact that growth rates (expressed as increase in subscribers divided by subscribers at the start of the period as in the above percentages) decline during the adoption of a technology should not by itself be a cause for alarm – indeed the most commonly used mathematical model for adoption, the logistic or "S-shaped" curve, is characterized by continually declining growth rates throughout the entire adoption process. However, the decline from 60% to 35% growth rate is substantial and exceeds any reasonable fit for simple diffusion models. The most popular explanation for this decline in growth is rising prices associated with a decline in competition at the end of the technology bubble of the late 90's [4, 5].

A second point in the wireless policy debate is the widespread availability of broadband relative to subscription. FCC data shows that broadband is available in 98% of zip codes [3]. Our own survey suggested that at least 76% of our respondents' households believed that they had access to broadband (Table 3 below). Yet a much smaller fraction (26%) subscribe. Again, though, this should not in itself be a source of concern; most consumer technology products are adopted at a slower rate than they are available to consumers since many other issues than simple availability pace adoption. The evolution of prices is surely a factor (many consumer technology adoptions are associated with falling prices over time, unlike the recent flat to increasing evolution of prices in broadband), as well as externalities such as network effects (the value of subscription rises as others subscribe) and complementary product and service effects (the value of subscription rises as products and services enabled by broadband become increasingly available).

A third point is the relative adoption rate of broadband in the early adoption period compared to consumer innovations such as telephone, television, cellular phones, compact disk players, etc, many of which took substantially longer to reach the 15% penetration level (suggesting that broadband adoption rates are, in fact, historically quite robust). These comparisons are challenging, though, since consumer technology adoptions have been accelerating and there are many specific factors that could differentiate the adoptions of these different technologies.

Overall, it is difficult to draw a strong conclusion about the adoption rate of broadband; subscription is steadily increasing although the decline in growth rate suggests some abnormality to the adoption process.

Regardless of the conclusion on whether the growth rate is "normal" or not, proponents of intervention [6, 7] point to the potential benefits of accelerating adoption in terms of economic growth (based on improved utilization of information technology), international competitiveness (some other countries have substantially higher penetration than the U. S. [8]), and a number of applications that are enabled or enhanced by broadband, and are of economic benefit or other societal benefit (e.g., security, health, education). Some express a need to identify a "killer application" for broadband [9, 10]; that is, one so compelling that by itself it can also motivate the purchase of broadband. But, its not clear that such a broad platform needs a killer application and indeed most adopters are satisfied while mainly using, to first order, the same applications as dial-up users, albeit with greater convenience (Table 4 below). The result, then, is a general consensus on the desirability of broadband adoption itself without a consensus around the desirability of expensive interventionist policies to accelerate adoption.

II. Broadband Usage and Attributes

To better understand the state of residential demand for broadband, we conducted a nationwide mail survey of US residences during September and October 2002 [1, 11]. In particular, we sought to investigate consumer awareness of Internet access service, profile residential Internet access and use, and gain insight into how important "always on" connectivity, cost, rate, installation and reliability attributes are in a household's choice of service.

Survey data provide a profile of the representative household respondent. The average respondent is a white, 50 year old male, with a two year degree at a college or technical school, who resides in a household with 1.7 other members. He was employed last month at a location outside of the home, and has annual household income of $71,934. A description of how Internet access varies by income, race, household size, age, education and employment status can be found in [11].

Here we summarize some of the results of this survey that either help describe the characteristics of actual and potential broadband users or illuminate issues of particular importance to wireless broadband.

A. *Social Disadvantage*

The "digital divide" debate refers to the perceived gap in computer and Internet use between high and low income households, educated and less educated populations, white and minority populations, urban and rural areas, etc. [12]. Table 1 shows type of Internet access by various measures of social disadvantage. The percentage of respondents with no access is relatively high for less educated, senior, and

Table 1. Internet access by measures of social disadvantage

Disadvantaged groups	No access	Dial-up	Broadband
Less educated (high school or less)	39.5	50.0	10.5
Senior (age greater than 65)	44.7	40.8	14.5
Lower income (income less than $20,000)	70.6	29.4	0.0
Minority (non-white)	22.2	63.0	14.8
Women (women head of household)	26.5	58.9	14.6
Total (all households)	28.3	51.5	18.8

Note. Cells are percent of respondents in the access category.

lower income groups. Broadband access is low for less educated and lower income groups, and no lower income respondents have broadband access to the home.

B. Computers and Telephones

78 percent of all respondents have at least one PC or laptop in the home, while 32.6 percent have two or more PCs or laptops in the home. Table 2 presents a cross-tabulation of the number of computers in the home and type of Internet access for respondents with Internet access. Broadband Internet access is positively associated with the number of computers in the home. Further, 88.7 percent of respondents have at least one telephone line from the home, and 24.1 percent have a second line. The most frequently sited reason for a second line is "for dial-up Internet access – to free up the primary telephone line for voice calls" (48 percent of homes provide this reason).

Table 2. Internet access by computers in the home

Computers in the home	Dial-up (*n*=214)	Broadband (*n*=77)
One (*n*=170)	79.4	20.6
Two or more (*n*=80)	72.5	27.5
Three or more (*n*=41)	51.2	48.8
Chi-square test for independence of the variables $\chi^2(2) = 13.55$*		
Correlation coefficient for linear association between variables $\rho = 0.205$*		

Note. Cells are percent of respondents in the access category. * is significant at the five percent level.

C. Awareness of service availability

Awareness of broadband service availability is relatively high for cable modem and DSL technology. Table 3 shows responses to the question "which ways of getting broadband access are available in your neighborhood." 15.3 percent of respondents replied "not sure" for cable modem, 30 percent for DSL, 68.1 percent for fixed wireless, and 64.7 percent for satellite. For fixed wireless, only 15.8% of respondents believe that fixed wireless access is available in their neighborhood; fixed wireless appears to suffer both from low availability and low awareness.

D. Internet Access

71.7 percent of homes connect to the Internet. 71.8 percent of these homes access the Internet with a dial-up connection, two percent use WebTV, and 26.2 percent use a broadband connection. Survey data suggest that 18.8 percent of the population have a

Table 3. Awareness of broadband Internet access

Access technology	Available	Not available	Not sure
Cable modem (*n*=352)	73.0	11.6	15.3
DSL (*n*=333)	57.7	12.3	30.0
Fixed wireless (*n*=311)	15.8	16.1	68.1
Satellite Internet (*n*=317)	27.4	7.9	64.7

Note. Cells are percent of respondents answering "available", "not available", or "not sure."

broadband connection at home. The most frequently sited reason for broadband Internet access is "speed is appropriate", followed by "I like the always on connection", and "to free up my telephone line for voice calls". The mean price per month for dial-up and broadband access, respectively, is $17.51 and $40.76, although to the extent a consumer does, in fact, surrender a second dial-up telephone line when switching to broadband, the cost of broadband becomes less onerous.

E. Internet Activity and Experience

Broadband users are more active with 19.44 hours of online activity per week compared to dial-up users with 12.55 hours of online activity. Including home, school, work and other locations, broadband users have been going online for 3.48 years compared to 3.22 years for dial-up users. When asked whether they use a broadband Internet connection at any location outside of the home, 94.6 percent of all respondents (i.e., those with and without Internet access at the home) indicate they have used broadband Internet at either a cyber café, library, place of employment, school, friend/relative's house, or other location.

Internet activity data are obtained by asking respondents "how often do you and other household members do each of the following activities: email and IM; use search engines or purchase products; play games or gamble; share music files or photos; banking, trading stocks, or bill payment; and download movies to view on the PC." Table 4 shows Internet activity for "many times a week." Email and IM, and search engines and product purchases are frequent activities for both dial-up and broadband users, which is consistent with findings from the BLS [2]. As suggested above by the number of hours online per week, broadband users are more active than dial-up users. The percentage of broadband users answering "many times a week" is higher for all Internet activities.

Table 4. Frequency of Internet activity – "many times a week"

Activity	All	Dial-up	Broadband
Email & instant messaging	71.4	68.5	79.7
Search engines & purchase products	37.8	29.5	60.8
Play games & gamble	21.0	17.7	30.8
Share music file or photos	10.1	6.8	19.5
Banking, trading stocks, or bill payment	9.8	7.4	16.5
Download movies to view on PC	1.0	0.9	1.3

Note. Cells are percent of respondents using the activity "many times a week."

F. *Internet Access Attributes*

The "Internet Access Attributes" section of the questionnaire describes and informs respondents about Internet access attributes. Respondents consider their preferences for different attributes when answering the question "how important is (or would be) the attribute of Internet access to you." A single answer is selected for each question from the following choices, "not important", slightly important", "somewhat important", "very important", and "extremely important."

Table 5 shows the percent of respondents who indicate the attribute is either a very important or extremely important part of their Internet access. Speed, reliability, and always on functionality are clearly important to broadband users. Interestingly, reliability of service is also quite important for respondents with no access, dial-up access, and broadband access. This latter finding supports anecdotal evidence that consumers desire a service they can count on being available whenever they want to use it, with consistent speed (that is as fast as advertised), and any problems that do arise are immediately handled by good customer service.

Table 5. Importance of attributes by Internet access

Attribute	All	No access	Dial-up	Broadband
Always on	49.1	40.3	39.0	87.3
Cost	59.6	48.6	64.6	58.3
Speed	53.6	38.1	46.2	92.4
Installation	33.6	34.2	26.0	54.4
Reliability	66.3	52.0	64.1	89.9

Note. Cells are percent of respondents who indicate attribute is "extremely important" or "very important."

In addition to asking respondents which attributes were important, we used conjoint analysis to estimate their willingness to pay for improvements in each of the attributes. The results were broadly consistent with their description of importance of attributes. In particular, willingness to pay for improvement in reliability was the largest across all the attributes and for both current users and latent users (dial-up users who live in areas where broadband is available).

III. WIRELESS AND BROADBAND

Wireless technology, applications, and business are a critical component of the overall broadband demand status and policy discussion. To date, wireless has not played a large role in broadband adoption but it can potentially be quite important.

A. *Awareness*

Aside from low actual availability, most potential users are simply unaware of fixed wireless (unsure as to whether it is or is not available). While wireline providers have been communicating the value and availability of their offerings for several years now, wireless options will have to spend substantial effort creating basic consumer awareness. On the other hand, communities which have been educated about broadband by wireline providers (spill over from generic promotion campaigns) but which are underserved should provide quick yield from education and promotion campaigns for wireless and represent obviously attractive initial targets.

B. *Attributes Relevant to Wireless*

Our survey on broadband demand provides some important information on broadband attributes for wireless. Not surprisingly, speed is a valued attribute of broadband. Interestingly, though, ease of installation is not highly valued by either current or potential users. Wireless businesses would be better off optimizing other attributes or their own cost to install over the perceived ease of installation on the user's part. Conversely, reliability is highly valued by current and potential users and is, in fact, more highly valued than speed. This has important and somewhat awkward ramifications for wireless systems and businesses. Wireless systems intrinsically share capacity among users; systems which take advantage of statistical usage patterns to increase user density will occasionally experience peak loading resulting in perceived reduction in reliability. This needs to be carefully managed. The situation is more pronounced in unlicensed spectrum; here there is little ability to control even the number of users let alone their particular usage patterns. Moreover, wireless systems can also suffer capacity degradation based on physical changes in the environment (e.g., moving objects, changing topography due to construction or vegetation, weather conditions). Care needs to be taken in engineering systems so that wireless does not become associated with appreciably lower reliability from the users' perspective than wireline alternatives, unless cost advantages permit substantially lower pricing to compensate for loss in this valued aspect of service.

C. Inter-Modal Competition Policy

The FCC and other government bodies generally support the idea that robust competition among telecommunications service providers is in the public interest. While some continue to advocate access to incumbents' plant by competitors within each access technology (intra-modal competition) [13], the FCC seems more oriented recently towards a focus on competition between competitors using different access technologies (inter-modal competition) [14]. As we noted in Section I, while there is general support for steps that would encourage more rapid broadband adoption, there is little consensus around an interventionist government policy, particularly an expensive one. Wireless can play an important role in facilitating intra-model competition and accelerating broadband adoption by creating a viable and widespread third platform (to compete with DSL and cable modems) without government subsidies. One possibility is through the emergence of innovative technology and business models in unlicensed spectrum (discussed further below). The other is through the effective use of traditional licensed spectrum; indeed notwithstanding the travails of various businesses attempting to offer Internet access in licensed spectrum, the continued evolution of technology and the presence of underserved communities suggest that fixed wireless access models in licensed spectrum could be viable, were it practical to license or sub-license spectrum in geographic regions more limited than Basic Trading Areas [15]. The viability of such businesses rests on policy and practice in licensed spectrum allocation.

D. Spectrum Policy

Reform of spectrum policy is being touted by the U S government administration as a key component in its general support for broadband deployment and is a subject of a comprehensive current effort within the FCC [16]. The general notions are to increase reliance on market-based mechanisms to allocate spectrum, both through exclusive transferable spectrum rights, and through unlicensed spectrum commons, and adoption and reliance on new technology approaches which allow reuse of spectrum without impinging on incumbent spectrum licensees. Increasing the flexibility of access to spectrum is likely to have a strong positive affect on the role of wireless in broadband adoption. While technology, such as Software Defined Radios (SDR) and UltraWide Band (UWB), become more mature, it may well be the flexibility in government policy that determine their success.

E. Decentralization and Diffusion of Power

Wireless, particularly in the form of IEEE 802.11b (Wi-Fi) networks operating in unlicensed spectrum, has shown the potential for novel, diffuse, and politically decentralized versions of network access that are hard to duplicate in wired configurations. Wi-Fi is most widely deployed as a local area networking technology, but public access is also expanding through creation of "hot spots." Although several traditional geographically diverse network service providers are emerging directly (for example, the recently announced Cometa Networks joint venture, www.cometanetworks.com) or as aggregators (e.g., Boingo, www.boingo.com and Joltage, www.joltage.com), a number of innovative alternative business models are also developing, including grass-roots cooperatives [17, 18] (albeit some such networks run into trouble with telephone or cable company backhaul [19]) and municipality installed networks, such as in Long Beach, California [20]. And technology and business models are co-evolving with the development of mesh-oriented 802.11 networks in which subscribers cooperate to carry their own signals and those of their neighbors [21]. Overall, these developments support the FCC's attempts to reform spectrum policy in favor of innovation and recent congressional moves to increase allocations of unlicensed spectrum [22]. But, challenges in unlicensed spectrum are also already emerging, as, for example, commercial hot spots and community grass-roots networks come into conflict [23].

Several commercial cellular providers are also looking at Wi-Fi as a means of extending their present cellular coverage and capabilities in hot spot areas. This suggests an interesting model; one in which a traditionally centralized service provider enters into a decentralized service. Such efforts could alter the grassroots networking efforts considerably, in that large service providers could come to dominate popular access areas. However, Wi-Fi, as an unlicensed service, requires that providers must contend with interference issues (important to perceptions of reliability as discussed above). Control of key hot spots through exclusive arrangements with property owners, such as a mall or airport, may limit the interference and provide essentially a default control of the spectrum. In this way, the battle might be won by property boundaries rather than spectrum ownership.

F. *The "Digital Divide" and Universal Access*

Although broadband is widely available, it is least available in rural and economically disadvantaged areas (see Table 1 regarding disadvantaged populations). Wireless can play a role in economically extending access to broadband to these communities. Satellite based access is obviously of importance to rural subscribers, although the combination of relatively high cost of satellite services and relatively lower adoption rates among this population mean that satellite based systems are having little impact so far on overall subscription rates. Fixed terrestrial wireless systems could play an important role in some rural communities, although perhaps requiring supportive government policy [24]. Altruistically motivated cooperatives and municipally supported networks may also help in reaching disadvantaged urban communities; an example of the former is NYCwireless in New York City [25]. The advent of more flexible Universe Service Funding (USF) measures could aid in the development of wireless services for rural and economically disadvantaged based on cross-subsidization between users.

G. *Mobility and Network Externalities*

Many broadband applications exhibit network externalities: they become more valuable the more users there are. All users benefit, then, if wireless can incorporate mobile users into an application's user community. In fact, each of the major types of applications usually touted as important to broadband [1] is advantaged by mobility in some way:

Entertainment (including games) – consumers have abundantly demonstrated their desire to consume entertainment content while mobile.

Education – some forms of education can substitute for entertainment for mobile consumption; in other cases wireless delivery of broadband might facilitate delivery of education to disadvantaged communities (as part of closing the digital divide).

National Security and Digital Government – some applications, such as public safety (discussed further below) benefit substantially from mobility.

Teleworking – like consumers, business users have historically demonstrated a high desire for mobile access.

Telehealth – some applications, such as emergency medical teams, would benefit from mobile broadband access, and others, as with education, could benefit from improved access to disadvantaged and rural communities.

H. *Public Safety*

Integrating wireless data access for public safety agencies has become a national priority [26]. At the same time, increasing the bandwidth of mobile data access to broadband levels can improve the effectiveness of public safety networking. Some private-public partnerships could help accelerate this; for example, Ricochet agreed to provide 1000 free modems to Denver city police cars as part of its application to reactivate its network locally after the bankruptcy of original parent Metricom [27].

Other examples of wireless public safety technology include, the development of a combined FM subcarrier/Wi-Fi enabled emergency medical services. This service makes use of FM subcarrier technology to deliver medical data, such as medical conditions and emergency contacts, to the ambulance. The ambulance then acts as a base station transmitting to paramedics up to a mile away.

As wireless services become more ubiquitous, it is likely that we will expect them to support a broad set of public safety communications. The scope of this coverage may vary depending upon the user, the provider, the type of service, and its cost. A basic emergency service ("911") may be among the first expected, but other lesser known services, such as priority access service and priority restoration, may soon follow.

IV. CONCLUSION

Given the state of broadband adoption and the broadband policy debate, broadband wireless' opportunity to play a substantial role in adoption rests on providing viable competition to wireline without government subsidization. The FCC and Congress have correctly understood that spectrum policy reform enabling technology and business innovation is likely necessary for this to happen. Beyond competition policy, though, broadband wireless can advance other public interest policy goals, such as closing the digital divide, enhancing public safety, and democratization and diffusion of network control.

Broadband wireless entrepreneurs will have to overcome very low awareness of wireless as an option and be cognizant in particular of the relatively low value consumers place on ease of installation relative to other attributes and, more importantly, the relatively high value placed on reliability. Careful engineering of broadband wireless will be necessary to position reliability attractively relative to wireline alternatives.

Finally, broadband wireless can uniquely address mobile access. This not only provides a unique market for wireless but also benefits all users of broadband by adding mobile users to each application's user population, promoting additional application development and economies of scale.

REFERENCES

[1] M. Jackson et al., *Broadband Demand Study Final Report*, November 15, 2002. Online at http://itd.colorado.edu/broadband_demand.html. Last Accessed: January 21, 2003.
[2] *Current Population Survey: Computer and Internet Use Supplement*, Bureau of Labor Statistics, U S Government, September 2001, Online at http://www.bls.gov/cps/ Last Accessed: January 21, 2003.
[3] "High Speed Services for Internet Access: Status as of December 31, 2001," U. S. Federal Communications Commission, July 2002, Online at http://www.fcc.gov/Bureaus/Common_Carrier/Reports/FCC-State_Link/IAD/hspd0702.pdf Last Accessed: January 21, 2003.
[4] "Behind the High-Speed Slowdown," BusinessWeek Online, September 17, 2002. Online at http://www.businessweek.com/technology/content/sep2002/tc20020917_2824.htm. Last Accessed: January 21, 2003.
[5] M. Pastore, "Broadband Prices Are Up; Adoption Rates are Down," CyberAtlas. Online at http://cyberatlas.internet.com/markets/broadband/article/0,1323,10099_957511,00.html Last Accessed: January 21, 2003.
[6] "Building a Positive, Competitive Broadband Agenda," Positively Broadband. October, 2001. Online at http://www.positivelybroadband.org/library/whitepapers.shtm Last Accessed: January 21, 2003.
[7] "A National Imperative: Universal Access of Broadband by 2010." TechNet. Online at http://www.technet.org/news/newsreleases/2002-01-15.64.pdf. Last Accessed: January 21, 2003.
[8] M. Jackson et al., *International Implications for Broadband Policy*, January 15, 2002. Online at http://itd.colorado.edu/broadband_demand.html. Last Accessed: January 21, 2003.
[9] M. W. Smith and H-T Leung, "Finding the killer application – the role of the Broadband Applications Laboratory," *BT Technology Journal*, vol. 20, no. 1, January 2002, pp. 11-21.
[10] N. Victory. "Creation of a Broadband Universe: A 'Big Bang Theory'", Speech delivered at the Alliance for Public Technology Broadband Symposium, Washington, D.C., February 8, 2002.
[11] D. Hatfield et al., "Broadband Internet Access, Awareness, and Use: Analysis of United States Household Data," *Pacific Telecommunications Council Conference 2003*, Honolulu, Hawaii, January 19-23, 2003.
[12] B. Compaine, *The Digital Divide: Facing a Crisis or Creating a Myth*. Cambridge, MA: MIT Press, 2001.
[13] F. Bar, et al., Access and innovation policy for the third-generation Internet, *Telecommunications Policy* 24 (2000), 489-518.
[14] "Remarks of Michael K. Powell, Chairman, Federal Communications Commission," Speech at the Goldman Sachs Communicopia XI Conference, New York, NY, October 2, 2002. Available online at http://www.fcc.gov/commissioners/powell/mkp_speeches_2002.html. Last accessed January 21, 2003.
[15] A. Eftekhari, et al. "Feasibility of a Fixed Wireless Access Business," *International Symposium on Advanced Radio Technologies*, 4-7 March, 2003, Boulder, Colorado USA.
[16] *Spectrum Policy Task Force Report*, ET Docket No. 02-135, Federal Communications Commission, November 2002.
[17] R. Flickenger, *Building Wireless Community Networks*. Sebastopol, CA: O'Reily, 2001.
[18] D. O'Shea, "Peace, Love & Wi-Fi," *Telephony*, March 18, 2002.
[19] "Cable provider wages war against Wi-Fi freeloaders," *Telephony*, July 15, 2002.
[20] J. Markoff, "More Cities Set up Wireless Networks," *New York Times*, January 6, 2003.
[21] "Ad Hoc Peer-to-Peer Routing Technology," Online at http://www.meshnetworks.com/pages/technology/intro_technology.htm Last Accessed: January 21, 2003.
[22] *Jumpstart Broadband Act: Staff Working Draft*. Office of Senator George Allen, November 19, 2002.
[23] "Wi-Fi Access Already Turns Nasty," *Communications Today*, v. 8, i. 161, August 21, 2002.
[24] Parker, Closing the digital divide in rural America, *Telecommunications Policy* 24 (2000), 281-290.
[25] "Unwiring the Web," *Technology Review*, v104 i10, December 2001, pp. 22-23.
[26] "FEMA targets public-safety wireless stovepipes." *Government Computer News*, July 1, 2002, v.21, i.17, pp. 14-15.
[27] B. Koerner, "The New Old Thing," Slate, December 6, 2002. Online at www.slate.com. Last Accessed: January 21, 2003.

The Electrospace Model as a Frequency Management Tool

Robert J. Matheson*
NTIA/ITS.M
Phone: (303) 497-3293, fax: (303) 497-3680
matheson@its.bldrdoc.gov

In this tutorial paper, the electrospace is described as a theoretical hyperspace occupied by radio signals, which has dimensions of location, angle-of-arrival, frequency, time, and possibly others. Because these dimensions are independent, a given radio signal has a unique descriptor in the electrospace. Signals having different electrospace descriptors can theoretically be separated by a suitable receiver. The electrospace model provides a good framework to define spectrum user rights that divide licensed spectrum into parcels that can be flexibly used in an independent and non-interfering manner, while allowing complete freedom to divide and aggregate spectrum parcels via a secondary market. Disadvantages of the electrospace model are that it assumes ideal receivers and it allows the specification of spectrum parcels that cannot practically be used in the real world. Additional rules can be added to account for non-ideal receivers.

1. A General Description of Radio Systems

For many decades, radio regulators have realized that multiple radio systems could be used without interfering with each other if the radio systems were operated at different frequencies, or at different locations, etc. Ways to squeeze in more non-interfering radio users have been a major concern of radio regulators for almost a century, though the specific techniques have changed greatly with the technologies in use at various times. Formal concepts of how the radio spectrum could be divided among users included the development of the electrospace model by Hinchman in 1969. Others have developed similar models, synonymously using terms like "spectrum space" or "spectrum utilization." This tutorial paper describes electrospace concepts and their application to spectrum management environments based on flexible-use user rights.

All radio systems are characterized by a transmitter that emits a radio signal, a transmission path, and a receiver that examines the signal at the receiving location. In communications systems, the purpose of the system is to move information (e.g., TV programming, cellphone conversations, or credit card data) from the transmitting location to the receiving location. In the case of sensing systems like radars, the purpose is to compare the received signal with the transmitted signal to gain information about the transmission path (inferring the presence of airplanes, tornadoes, or hidden underground pipelines). Although the above characterization is useful, it is significantly incomplete, because it omits interactions between different radio systems.

The most common form of interaction between different radio systems is interference. "Interference" is defined in this paper as any degradation of radio system performance caused by the presence of extraneous radio signals. The radio system being degraded is known as the "victim" system. The source of the extraneous radio signal is known as the "interferer." All interference occurs within receivers; no interference occurs outside of receiver. The term "receiver" means the whole receiving system, including the receiving antenna(s). By definition, interference is a classic example of an "externalized cost." It is always done to somebody else–at their expense–often without the slightest awareness on the part of the interferer that someone else is being caused interference. As an externalized cost, mechanisms must be established to control it, since there is no benefit to the interferer to control it.

Lacking a specific regulatory limit, there may be only a vague opinion as to when interference

*This tutorial paper represents the author's understanding of electrospace concepts and should not be construed to reflect current or future NTIA policies or procedures.

becomes harmful. All interference is unwelcome, even if the only cost is to consume some of the system gain margin or forward error correction budget. The victim receiver is always culpable in cases of interference, since a sufficiently capable receiver (possibly outrageously complex and expensive) could always be constructed to operate adequately in the presence of any unwanted signal that was radiated from a different antenna than the desired signal. Any instance of interference is *prima facie* evidence that the owner of the receiver didn't provide a good enough receiver. There is no technical basis to say that some interference is caused by inadequate receivers, while other interference is not. Regulations that protect against interference operate by allowing interference-free service using simpler or less expensive receivers. One major function of spectrum management is to maximize the value of spectrum by designing frequency allocations so that less expensive equipment can be used effectively.

In a command-and-control regulatory environment, the regulatory agency takes responsibility for the whole design of radio systems, licensing only specific services, over given service areas, using specific technical parameters. Receiver performance can be strictly enforced to ensure that interference will be caused only by higher-than-expected unwanted signal amplitudes.

In a flexible-use environment, the user has wide latitude to choose the type of service provided and many technical parameters. It is sufficient to provide receiver users with an expectation (guarantee) concerning the maximum levels of unwanted signals. Depending on specific details of the user's operation, the receiving system could achieve rejection of unwanted signals using antenna gain patterns, frequency separation, band-pass filtering, receiver location, forward error correction, etc. In a given set of circumstances, some of these solutions would be much more cost effective than others. Economical industry solutions would presumably be developed for many standard circumstances.

2. The Electrospace

The electrospace is a formalized description of the radio signal environment, as it applies to all types of radio systems and all regulatory environments. It can be used to describe the ways in which the radio environment can be shared among multiple radio systems. It is particularly useful in flexible-use environments, where it provides a straight-forward basis for aggregating and dividing multiple electrospace regions.

The electrospace describes radio signals, which means that it describes the domain of transmitters and transmission paths. The domain of receivers is totally separate from the electrospace description. Although any real radio system must consider all system components – i.e., the electro-space and the receiver– it is appropriate to divide the two components for regulatory purposes. The crucial regulatory difference is that the electro-space describes the ability of radio signals to cause interference to others–which is an externalized cost that must be strongly regulated. The receiver domain includes only system components that do not cause interference to others (i.e., no associated externalized costs that need to be controlled by regulations).

The use of an electrospace model in no way diminishes the role of receivers in the design and implementation of systems. In fact, it overtly recognizes the fundamental role that receivers play in the design of any radio system, and it completely frees the engineers and the business-men to optimize receiver performance without any regulatory entanglements. In Section 6, various assumptions about the limitations of real receivers to reject unwanted signals will be used to modify electrospace management rules (discussed later), giving interference rights that are more appropriate and efficient for real-world use.

The electrospace describes the radio field strength at a given electrospace "location" that is defined by the 7 electrospace dimensions. These seven dimensions are all independent of each other, which means that the electrospace can be considered to be a 7-dimensional hyperspace. A "location" in the electrospace can be described by assigning specific values to several independent variables. It should be noted that different investigators have sometimes included somewhat different variables in the electrospace. The set shown in Table 1 is a useful starting point, and probably no great harm is done by including or omitting some marginal variables. The cases for two of these marginal variables–polarization and modulation–are discussed in the next section.

Table 1 - Electrospace Dimensions

Quantity	Units	# Dim
Frequency	kHz, MHz, or GHz	1
Time	seconds, hours, or years	1
Spatial location (geography)	latitude,longitude, elevation	3
Angle-of-arrival	azimuth, elevation angle	2

The physical location of a test point or hypothetical receiver is defined by the three spatial dimensions. The field strength at that location is described by the remaining variables, including the frequency, time of occurrence and direction of arrival. In a frequency band whose licensing is based on the electrospace, a numerical limit will typically be established, e.g., X FV/m, such that field strengths in excess of X are considered to be signals and are not permitted outside of the user's licensed regions of the electrospace. A given signal is said to occupy an electrospace "region" consisting of all locations in the 7-dimension hyperspace where field strength is greater than X.

One characteristic of the electrospace is that receivers can theoretically separate any radio signals that differ by at least one of their seven electrospace dimensions. For example, two co-located radio receivers could function without interference if the signals were at different frequencies, or if the signals occurred at different times, or if the signals came from different directions. Radio signals using the same frequency, operating time, and angle-of-arrival could be separated without interference if they were present at different locations.

3. Electrospace Dimensions

Frequency. The frequency dimension of the electrospace has the standard meanings of the word, namely a description of the frequency or range of frequencies (pass band) at which field strength is being characterized. Frequencies can be divided over a wide range of increments, typically matching the channelization of particular services.

Time. The time dimension can be subdivided over a wide range of increments. Useful time divisions might include the several-year duration of a licence, an agreement to allow a particular user to transmit regularly during the midnight-to-5 AM time block (when bandwidth would be inexpensively available to update computer files for the following day), or a one-time use during a 3-hour special events broadcast. On a much smaller time scale, a user could use a particular time slot on a TDMA system, to broadcast for a 2.5-ms time slot that would be available once every 20 ms, or transmit data during the vertical blanking interval of an NTSC television signal 30 times every second.

Spatial location. The spatial dimensions represent physical locations. They can be problematic, because there is no practical way to confine radio signals within a desired region. In typical hilly terrain, there are many distant locations that have higher signal amplitudes than many closer locations. Therefore, although one might easily select an arbitrary spatial region, the selected region might be extremely inconvenient to use efficiently. In order to prevent excessive signal levels (larger than "X") outside the boundaries of the selected spatial region, it might be necessary to greatly diminish signal amplitudes at many useful locations within the spatial boundaries. Transmitter power, details of the terrain, and the use of directional transmitting antennas are operative in establishing the spatial boundaries of the electrospace associated with a given transmitter.

A geographical area could be subdivided into smaller geographical regions or aggregated as part of a larger region, keeping in mind that radio waves propagate according to the laws of physics. Sub-dividing regions may have to be done carefully, so that the new users can actually make use of the new smaller regions, without creating or suffering interference with their neighbors. A useful technique might be to license the spatial dimensions by areas matching real coverage regions, determined in advance by measurements or modeling. For very-short-range systems, the geometry could possibly even be subdivided vertically. For example, the same frequency might be used on every tenth floor of a skyscraper (e.g., floors 7, 17, 27, ...).

Angle-of-arrival. This factor describes the angle-of-arrival or direction of radio signals at a given location, including the possible effect of multipath components scattered from many objects in many different directions from the receiver location. Note that this factor is not created by physical antenna pointing angles. The pointing direction of transmitting antennas primarily affects the spatial dimensions of the electrospace, i.e., the geographical areas where signals are strongest. No aspect of receivers–including the pointing angle of receiving antennas–ever has any effect on the electrospace. Therefore, neither transmitting nor receiving antennas influence the angle-of-arrival factor. On the other hand, receivers that exploit the angle-of-arrival will often have directional antennas.

In traditional radio systems, the useful angle-of-arrival is limited to the direct path between transmitter and receiver, e.g., point-to-point, free-space propagation between high-gain, narrow-beam antennas furnishes the technical basis for terrestrial microwave networks. The multipath signals arriving from other directions are often a major detriment to traditional receivers, causing frequency-selective fading and other problems. Directional receiving antennas can very efficiently exploit the electrospace direction-of-arrival dimension, separating out individual signals at the same frequency from multiple microwave towers or geostationary satellite orbital slots. This technique is easily scalable; additional angular subdivision is possible by using more-directional receiving antennas to separate out additional signals arriving from new transmitters at different angles at the receiving site.

If there were an easy way for a transmitter to send a signal to the vicinity of a distant receiver and make it appear that the signal came from a different direction than from the transmitter, the distant receiver could use a directional antenna to preferentially receive that signal. If the transmitter could somehow generate multiple independent signals in the vicinity of the receiver that appeared to be coming from different directions, the receiver could use multiple directional receiving antennas to receive multiple independent signals from the transmitter. This would remain true even if all of the multiple signals were at the same frequency. Thus, the direction-of-arrival dimension could (theoretically) be used to increase the traffic that could be carried to a given receiver–using the general principle that a receiver can separate any signals having different electrospace descriptions. Unfortunately, there is no practical method for a transmitter to generate multiple signals that appear to a distant receiver to be coming from different directions.

Recently developed Bell Labs Layered Space-Time (BLAST) technology exploits multipath reflections and multiple transmitting and receiving antennas to generate independent transmission channels, somewhat like the postulated multiple beams with different apparent angles-of-arrival described in the previous paragraph. Instead of using a combination of properly-phased antennas to produce directional beams, however, BLAST processing represents a much more generalized approach to the vector addition of multipath signals received on multiple omnidirectional antennas. Under certain conditions, the signals received on the multiple receiving antennas can be mathematically processed to separate the independent signals that were transmitted by a set of multiple transmitting antennas.BLAST technology uses a mathematical combination of multipath signals reaching the receiver site from different directions to synthesize multiple radio channels between a pair of transmitter and receiver sites.

Polarization. Some investigators have included polarization as one of the electrospace variables, since radio waves can be polarized using two orthogonal polarizations–e.g., vertical and horizontal or clockwise and counterclockwise. Many satellite systems transmit two separate signals, one on each polarization. However, a variable that has only two possible values seems to provide a very meager basis for a generalized dimension, so polarization was not included in Table 1.

Modulation. Some investigators have suggested that the electrospace should include a parameter for modulation. They reason that a suitable receiver could separate out many independent signals based on orthogonal coding, e.g., CDMA codes. The argument against using a "modulation" dimension in the electrospace is mainly a practical consideration, namely that there are no obvious codes/modulations that are orthogonal to all other general user codes. This prevents users from independently choosing a code and knowing that it will remain unaffected by signals from other users employing different codes and modulations. Even if all other users are carefully constrained to use codes belonging to a specific set of codes, the other users will substantially affect the available S/I ratios. Thus, even the best choices of codes or modulations currently available do not guarantee independence among the corresponding signals.

4. Spectrum Use Rights from the Electrospace

The electrospace model can be directly applied to a flexible-use, market-based frequency management environment, since the model describes a way that all aspects of the use of the radio spectrum can be unambiguously divided (shared) among multiple users. The only significant regulatory principle is that a licensee has the right to radiate a signal within a licensed electrospace region. Outside the region, signals must be kept below a specified field strength limit, X FV/m. There are no restrictions on type of service, transmitter power, bandwidth, modulation, antenna height, number of sites, etc., as long as the signal is kept lower than X in all regions outside the licensed electrospace boundaries.

An electrospace region is permitted unlimited aggregation or subdivision along all of its dimensions. This means that electrospace can be freely repackaged and resold as a market-based commodity, distributing spectrum without requiring approval by a regulator. A 1-MHz bandwidth could be subdivided into 40 channels of 25 kHz or augmented with additional adjacent frequencies to make a 5-MHz bandwidth. A given channel could be subdivided into TDMA time slots of 10 ms occurring every second and rented to a hundred separate transmitters. A statewide geographic coverage area could be divided into much smaller geographical cells and rented to short-range neighborhood wireless ISPs. Multiple fixed transmitters could be allowed to radiate signals into a common receiver location, if the transmitters are arranged to provide signals that have different angles-of–arrival.

Although the electrospace model is critically based on a field strength value, X FV/m, which cannot be exceeded outside the licensed region, it is not obvious what value to choose for X. Presumably X will be chosen so that systems licensed outside the region will not receive interference from the signal. However, the minimum level of interfering signal for various types of systems varies over a wide range–perhaps 50-60 dB–depending on the system. Since all types of systems are assumed to operate in an electrospace-based band, which type of system should X protect? One answer is that the selection of a specific value for X might make that band particularly suitable or unsuitable for various types of services; multiple bands could use different values of X to efficiently accommodate various services.

5. Practical Limitations on Electrospace

Although the electrospace model is conceptually powerful and potentially very useful, there are a few important problems with its application to the real world. The major problem, non-ideal receivers, will be discussed in the next section.

The division of the electrospace along any selected dimensions–while theoretically possible– may or may not produce a useful division in the real world. Arbitrary spatial regions, for example, may not match easily achievable propagation/coverage areas. A more useful spatial division technique may be to determine easily achievable coverage areas and divide the electrospace regions in a corresponding way. The angle-of-arrival dimensions may be compromised by unintended scattering from the terrain or by lack of sufficiently narrow beamwidth receiving antenna performance (especially at lower frequencies). Division into very narrow time slots may produce systems that are difficult to synchronize properly. Division into very narrow frequency slots may produce unreasonable requirements for frequency stability and Doppler shift.

The spatial dimensions pose some other problems. The field strength at a particular location is often the vector sum of many multipath signals. These multiple signals can occasionally add up to a field strength that is larger than the average field strength in the general vicinity. Therefore, it may be desirable that the field strength limit contain a statistical description, which would allow the occasional presence of signals above the limit. The inclusion of a statistical limit might make it much more difficult to show that a licensee had violated the electrospace limits, since a single instance of excess field strength might not be sufficient proof of a violation.

One obvious application of spatial coordinates is to describe licensed regions using latitude and longitude. For many applications, radio signals will be attenuated by buildings, terrain, and the earth's curvature, which tends to give the greatest attenuation at ground level. Raising a receiving antenna farther above ground will usually increase the received signal level. Therefore, a transmitted signal that meets the field strength limit at ground level may not meet that limit at higher elevations above ground. Many radio systems have receiving antennas located on tall buildings, towers, or mountaintops. Therefore, the success of a given application may depend on an understanding of how field strength changes with all three of the spatial electrospace dimensions.

The frequency dimension can also cause problems. Although a transmitter can radiate any amount of power inside the licensed frequency band, the field strength outside the licensed band must be less than X. Presumably this condition must be met at all locations–even very close to the transmitting antenna, where the in-band field strength is very high. To meet this condition near a transmitter may require unreasonable amounts of filtering. Therefore, the out-of-region limit, X, may need to be suspended in the immediate vicinity of the transmitter.

6. Receiver Effects in the Electrospace Model

The most serious limitation on the practical application of the electrospace model to flexible-use spectrum management is that the electrospace model assumes that all receivers are "ideal." In this context, "ideal" means that the receiver has infinite rejection of unwanted frequencies (i.e., frequencies outside of the intended receiver band-pass) and infinite dynamic range (strong out-of-band signals will not cause intermodulation products or gain compression). Unfortunately, none of the receivers that are actually available to users are ideal. Even worse, the most popular and rapidly growing class of receivers–handheld multi-band cellphones–are especially non-ideal, with performance constrained by small size, low cost, and limited battery power.

The important characteristic of real (i.e., non-ideal) receivers is that they can generate interference even when no interfering signal is actually present at the tuned receiver frequency. Strong signals at close-in frequencies or very strong signals at frequencies further away from the tuned frequency will cause receiver distortions that are seen as interference. Therefore, the electrospace model may need to be supplemented with additional rules, if real receivers are going to be protected from interference while keeping a market-based, flexible-use environment.

The additional rules are needed mainly to control the presence of strong signals at frequencies near the receiver tuned frequency. Receivers will benefit from frequency bands that are carefully engineered into duplex band architectures, where base station receive frequencies are systematically separated from base station transmit frequencies. Therefore, some frequencies may be designated for base or mobile use. Maximum limits on transmitted power must be observed, to protect nearby receivers, even when increased power would not violate the geographical boundaries of the license. It may be desirable to limit the maximum field strength at ground level (where the receivers usually are) in the vicinity of transmitters. When bandwidths are aggregated or divided, the maximum transmitter power should aggregate or divide proportionally to the bandwidth. It may be useful to specify limits on how power can be distributed within the aggregated bandwidth, so that the total transmitter power cannot be shifted to one edge of the aggregated bandwidth, posing an impossible adjacent-band rejection problem for a receiver at the adjacent frequency. These additional receiver-based rules can be added to the electrospace rules, while still maintaining almost complete market-based flexibility with well-defined rights.

7. The Role of Receivers

Although this paper is primarily about the electrospace–which by definition does not include receivers–it may be useful to discuss the role of receivers in an electrospace environment.

In the command-and-control model, the receiver's role is well defined, usually having a required performance that is based on the technology that was available when the regulations were originally defined. Typically, the receiver is designed to operate in the absence of co-channel interference, often even requiring reduced amounts of adjacent channel interference. Operation under these specifications is expected to guarantee a given level of performance over a licensed area of service. Because service is guaranteed, there are tight definitions of harmful interference and receiver performance specifications. When service is guaranteed, using a specified receiver, the presence of a licensed receiver can constrain the operation of additional nearby transmitters. Therefore, both receivers and transmitters may need to be licensed and regulated.

The role of the receiver is completely controlled by the receiver owner in bands regulated by electrospace-based flexible-use rules. Although the maximum permitted levels of interfering signals are known, almost nothing else is. Nothing in flexible-use bands can be construed to specify any particular type of service, level of performance, or service area. However, since a receiver cannot cause interference to anyone else except the receiver user, and the receiver user is the only person who will benefit or suffer from the selected receiver, the user is the only party motivated to make proper decisions about receiver performance. Therefore, the receiver user can be entrusted to make all decisions about optimizing receiver performance for the job at hand. In bands where receiver limitations have been included in the electrospace rules, the expected receiver signal environments would be presumed to be somewhat more benign, allowing the use of less-expensive receivers.

The preceding paragraph should not be construed to mean that mandatory receiver standards are never desirable in a flexible-use environment. User groups may need to establish minimum receiver performance standards to assure interoperability or other shared aspects of system performance. Consumer groups may want to set standards by which the performance of receivers can easily be judged and compared. Such groups should be able to set whatever receiver standards seem appropriate to them. The point is that such receiver standards would be a proper concern of these groups, not spectrum managers.

In the flexible-use model, interference is not prohibited; instead, it is a system parameter that can be consciously optimized by the user to serve a particular function. The receiver must work well enough to provide the needed service, but not better (because better receivers will cost more, but provide only additional unneeded benefits). Since the flexible-use environment provides a maximum guaranteed level of interfering signals from external users, the whole system (including receiver performance) can be realistically engineered and optimized.

8. Summary

The electrospace model has been shown to have direct application to a market-based, flexible use spectrum management environment. In this role, it provides licensees very great freedom in providing services and distributing spectrum, while unambiguously providing rules for preventing interference to other users. The electrospace model becomes even more useful for frequency management when it is modified to account for normal receiver limitations.

As a conceptual tool, the electrospace directly suggests various methods to squeeze additional communications capacity into a given frequency band, since one can imagine subdividing any of the electrospace dimensions to allow extra users. Modern cellular systems have provided huge initial gains in spectrum capacity by dividing the spatial dimensions to get smaller cells and much higher geographical frequency reuse. The use of cellular base station adaptive antennas further subdivides the spatial dimension by the beam-width of a base station transmitting antenna array and exploits the different angles-of-arrival from individual mobile users.