Maxwell, Shannon, and Economics

Spectrum Policy in the Era of Spectrally Efficient Mass Market Wireless Communication

Dr. Arthur H. M. Ross

September 30, 1998

Introduction

As a participant in wireless standards activities, I welcome the FCC's request for comments on IMT-2000 spectrum issues in preparation for the upcoming World Radio Conference in the year 2000 (WRC-2000). The Commission will doubtless receive numerous responses to this NOI from industry, both operators and manufacturers. Those parties will naturally espouse US positions that favor their individual businesses. I believe, however, that the public interest will be best served by policy that takes collective advantage of the unique technical properties of the emerging SSMA air interfaces. This requires a new, possibly radical regulatory paradigm. I am attempting to act as a public-interest advocate for that new paradigm.

Please note that the opinions expressed here do not necessarily represent those of any of my clients, past or present. I would hope, however, that widespread support will develop for these views, as they represent, in my opinion, a win-win scenario for both the wireless industry and the communicating public.

The decisions facing the Commission encompass several very difficult technical problems and many imponderable policy issues. The most appropriate course for the Commission is not only to recognize the fundamental intractability and uncertainties of this multi-faceted problem, but also to actually *embrace* them. A new regulatory regime founded on *flexibility* rather than traditional carved-in-stone exclusive service licensing, is now possible due to the new spectrally efficient spread spectrum air interfaces. The primary regulatory consequence is that licensing need not be exclusive. Multiple operators can be permitted use of the same band in the same service area. Substantial benefits for vendors, operators, and consumers ensue from such a regulatory regime.

The title of the paper stems from the three foundations of the wireless business:

- Principles of radio physics that make cellular wireless service possible
- Information theoretic concepts that govern communication performance
- Business considerations

An enlightened synthesis of these three disciplines suggests the new regulatory paradigm. The remainder of this paper discusses how such a regulatory philosophy might work, why it is beneficial, and some of the technical issues.

Current US Situation

At least one industry group in the United States has attempted to project a requirement for IMT-2000 spectrum in the United States. The preliminary conclusion seems to be that an additional 300 MHz is needed for mobile (non-MSS) services. There is clearly a fundamental conflict between this estimate and what is achievable, given the realities of global electro-politics. There is almost no spectrum that does not have an incumbent constituency that will vigorously oppose its re-allocation. The opposition is often well-justified, particularly for applications like radars and aeronautical aids to navigation that have safety-of-life implications. Another problem exists in the Americas due to the PCS band plan, which conflicts with the WRC '92 identification of 1885-2025 MHz and 2110-2220 MHz as the IMT-2000 spectrum. Not only is the Region 2 PCS band plan different, it also overlaps the ITU-identified spectrum, so that deployments of both plans in one geographical area is problematic.

While one can sympathize with the good intentions of trying to rationalize a spectrum request for the WRC, the calculation that leads to the 300+ MHz number is not credible and cannot be taken too seriously. It postulates a demand number, a spectral efficiency per station, and some other system characteristics. From those numbers it calculates a bandwidth, using a straightforward formula. There are at least three problems:

First, demand is not a number; it is a relationship. The usage of a service will depend on the balance, as perceived by its consumers, between the price of the service and the benefits they derive from it. The price that the operator will charge depends on his costs, which certainly includes fixed and recurring infrastructure costs, and potentially includes a cost of spectrum, depending on the regulatory regime.

Second, the nature of the services is highly speculative. There are IMT-2000 services projected for which there is no precedent, and thus no credible market model.

Third, the spectral efficiency of the systems will be highly dependent upon the nature of the traffic. The putative very high data rate packet services, for example, will have very poor performance in low bandwidth allocations. On average, the number of simultaneous users that can be supported in a 5 MHz allocation at the highest data rate will be less than one. The law-of-large-numbers averaging, which is the hallmark of spread spectrum, is not effective when the number is not large.

In the best of all possible worlds, one might regard this as a very complex, classical resource allocation problem, to which a sophisticated linear programming model could be applied. Such a model would accept as inputs the many factors that influence the business, and would output the spectrum that should be allocated so as to optimize, in some sense, the effects on US industry and consumers. However such a calculation is quite intractable due to the many imponderables. These include market (read "human") behavior, business models, political considerations, and regulatory models. In the short term, we will have to make do with a sub-optimal or alternative approach.

Guiding Policy Principles

Some guiding principles can be identified, based on the current rough and tumble standards situation, and the difficulty of predicting future technology trends. For the most part, these are already US policy, but they perhaps bear explicit restatement.

- ***Technology-neutral national policy*** – This policy is already in place, and should be continued. The constraints on the operators should be those of a fair and orderly marketplace, not an arbitrarily imposed standard air interface. Although the ITU-R started with the goal of a single worldwide IMT-2000 air interface, it is quite clear that this is both undesirable and politically unachievable.

- ***Flexibility***
  The regulatory regime should not dictate service offerings. Within broad limits, the operators should be permitted to offer services over market footprints that are dictated by its business considerations, not an arbitrarily imposed service model.

- ***Application-independence*** – The rules of the road should permit a variety of applications, again for the operator to adjust service offerings in accordance with marketplace realities. Things like out-of-band emissions and maximum ERP should not be dependent on the nature of the service being offered.

- ***Traffic model independence*** – The nature of the traffic carried on these future networks is difficult to anticipate. While one is tempted to say that voice will be a large portion of traffic, even that may be carried via an IP-like packet bearer. The regulatory regime should accommodate widely varying traffic mixes, over both time and geography.

The overall regulatory goal should be a minimalist regulatory regime that permits a maximum of flexibility on the part of the operators as to their service offerings, and the air interfaces that they use. Adam Smith's invisible hand will dictate the evolution of the business within the broad regulatory constraints, whose primary purpose is to maintain a fair and orderly marketplace.

Underlying Technical Principles

Land mobile radio is both facilitated and limited by the peculiar phenomena that dominate near-earth propagation in the low GHz spectral bands. Application of spectrally efficient principles of information theory permit exploitation of this peculiar channel in near-optimal ways.

Maxwell – Radio Propagation in the Land Mobile Environment

The ugly communication channel that is the norm in the land mobile radio environment is both a blessing and a curse.

It is a blessing in the sense that the natural and man-made surface features result in larger-than-free-space path loss. The effective propagation law is generally modeled as a $r^{-\gamma}$ propagation law, with γ ranging from about 3 to more than 5, versus 2 in free space. Various explanations have been postulated, but regardless of the physical reason, it is a well-established empirical fact. It is a curiosity, in fact, that cellular radio would not work in an infinite plane, uniformly populated with users. The integral that represents the average interference power diverges like $log(r)$. The faster-than-free-space empirical law avoids this infinity.

It is a curse in the sense that the terrain and structures lead to multipath propagation. Multipath propagation underlies the familiar Rayleigh fading of the narrowband systems. It is particularly troublesome to analog FM because it has direct adverse effects on the recovered audio. The wideband CDMA systems are affected differently, but always adversely. While the manifestations of the multipath differ in detail, the overall result is a reduction in quality-of-service, or capacity, or both.

The multipath nature of the channel does have strong effects on the best choice of modulation and spreading. Roughly speaking, the correlation time of the signal should be of the same order as the correlation time of the channel. While advocates of various products will claim superiority for their particular choices, there really is no unique optimum answer because there is no such thing as a typical channel. The nature of the channel is a strong function of the physical environment. Dramatic differences are found between indoor applications (both subscriber and base station indoors), exterior urban, and exterior rural environments. What is optimum for one environment may not be so for another environment.

Later we argue for large bandwidths as an aid to achieving large aggregation of users. This must be done with due regard for the nature of the channel, as excessively large bandwidth in a direct-sequence spread signal will result in performance loss for a number of complex reasons. This is not to say that it cannot be done, but only to say that it must be done *carefully*.

Shannon – Fundamental Properties of the Noisy Communication Channel

Claude Shannon taught the world, fifty years ago, in 1948, that there is an information rate, usually measured in bits per second, associated with a communication channel, such that perfect communication can be achieved so long as the actual rate does not exceed this channel capacity. For a channel of bandwidth W, at fixed power, the channel capacity in bps is

$$Rate \;=\; W \cdot \log_2 \frac{P_S + P_N}{P_N}$$

where P_S and P_N are the signal and noise powers at the receiver. The subsequent 50 years of development have seen the development of coding schemes that achieve rates within a fraction of a dB of the Shannon limit. This result holds for a power-limited channel with additive white Gaussian noise.

Shannon further concluded that the optimum signals for communication in a mutual interference limited situation are noise-like. While there was no prescription given at the time of how to do this, subsequent developments have produced approximations that use digitally generated pseudo-noise sequences. These can come quite close to the ideal noise-like model. Under the assumption that all users arrive at the same power, and that there are M>>1 of them, then the Shannon capacity rate for each is

$$Rate \;=\; W \cdot \log_2\left(\frac{1+M}{M}\right) \;\approx\; \frac{W}{M\ln(2)}$$

or about $(\ln 2)^{-1} \approx 1.44$ bits per second per Hertz of bandwidth, split among the M users. Current systems actually achieve perhaps 10-15% of this, in an environment that is much more complex and difficult than simple additive white Gaussian noise. The lower figure includes all real effects, such as handover complications real-world mutual interference, and power control.

The Shannon rate corresponds to $E_b/N_0 = \ln(2)$, or about -1.6 dB. In this form it is more directly comparable to the performance metrics of realistic systems. Current practice achieves E_b/N_0 of the order of +3 dB to +12 dB, depending on the channel, coding, and modulation. Moreover, the forward link performance is a much more complex issue than this simple analysis assumes, primarily because of handover considerations.

But even with the several dB deficit with respect to the Shannon limit, the use of SSMA has improved spectral efficiency for the cellular voice services by a factor of at least 5 or more.

Probably the most dramatic improvements in air link spectral efficiency will come from clever uses of adaptive antenna technology. Use of the so-called "turbo" codes will provide some additional improvements, as these have been shown to come within a fraction of a dB of the Shannon bound in additive white Gaussian noise.

Economics Meets Maxwell and Shannon – How to best manage the scarce natural resource that is the electromagnetic spectrum

Although some improvement may still be possible, dramatic new coding and modulation techniques are *not needed* for greater spectral efficiency. The *simple* benefits that accrue from SSMA have not yet been fully exploited. Formerly we managed mutual interference by geographically sparse frequency reuse. In cellular practice this is typically 21-way reuse, that is, each channel assignment is used in only 1/21 of the stations (station meaning one angular sector of one site). SSMA replaces this with universal frequency reuse, but the mutual interference now appears in the E_b/N_0 budget. Overall, a gain of perhaps 5-7 is realized, equivalent to a frequency reuse of perhaps 3 to 4. The aggregation of interference (pooling) that is achieved by SSMA is the key feature of the technique. The law of large numbers has benefits for efficiency in the sense that the variances of load and interference become fractionally smaller as the pool of traffic is increased.

The law of large numbers is not effective for excessively large data rates, as the number of users per station will eventually come to be *not* large. This will be the case for *all* of the proposed air interfaces in the current ITU-R RTT evaluation at the highest data rates, especially in the smaller bandwidth channels. Variances in traffic, power, and other properties become very large fractions of the averages when the number of users is less than one.

Separation of *Operators* in Code Space

Current Commission rules allow any Commercial Mobile Radio Service (CMRS) licensee the flexibility to change its radio transmission technology without further approval beyond initial licensing. Only the most basic technical compatibility constraints are imposed. In the cellular and PCS services, a variety of air interfaces have in fact been chosen without regulatory intervention. Each operator deploys its chosen technology within its allocated frequency bands and geographic footprint. Coexistence is enforced by simple, physical layer criteria that are often redundantly imposed via both Commission regulations and air interface performance specifications. Frequency reuse is at the discretion of each operator, but only within his system.

The fact that even this partitioning works in practice is a consequence of the near-earth, faster-than-free-space propagation.

For current purposes, the key observation is that this is an *exclusive* licensing model, in the sense that each operator is granted an *exclusive* license to use a particular duplex band pair, in a particular geographical area.

But in the context of SSMA systems, *frequency* exclusivity is no longer essential. In the same way that each operator can (and should) employ universal frequency reuse for maximum system capacity, multiple operators can make use of a common frequency band. Not only can *users* be separated by spreading code rather than frequency, as we often like to say when describing CDMA technology; *operators* can be separated as well.

Why share spectrum over operators?

In a word, *aggregation*. The primary benefits that ensue from the use of SSMA arise from the aggregation of the mutual interference. It changes the design criteria from worst case interference, manifested in the frequency reuse patterns, to *average* interference. Aggregation of users within a single operator's system helps; aggregation over multiple operators *helps more,* due to the law of large numbers. In the case of some of the higher rate IMT-2000 services, the estimated channel capacities are less than one Erlang per station in a 5 MHz allocation. Aggregating load will lead to greater air interface and trunking efficiencies. It may also the viability of the business. It may be unprofitable for an operator to deploy high data rate services if there is so little demand that his spectral efficiency is poor. This could be a major influence in the decision to deploy the more exotic services or not.

The difficulty, of course, is that sharing over operators complicates the business and regulatory models. It is also not entirely obvious how it can be done technically, although some possibilities are outlined below.

Benefits

Neglecting for the moment the technical aspects of how the sharing is accomplished, one can identify some beneficial business consequences of sharing spectrum. They are, in a sense, all economies of scale due to the aggregation of traffic.

Resource Allocation

Carriers could fine-tune their support of spectrum in accordance with the distribution over time and geography of demand for their particular services. It would not be necessary to deploy support for the entire band at all locations. Capacity could be utilized in greater or lesser amounts, in accordance with demand.

Efficient Market

Amicable sharing agreements between operators could be implemented as a sort of real-time auction, where the operators bid for the use of available capacity, in accordance with their customer service agreements. An analogy of this might be found in the real-time trading of electric power that takes place among utilities in accordance with the shifting of supply and demand over geography and time.

Creative Service Offerings

Service pricing could be in accordance with the very real tradeoff between supply and demand. For example, low prices could be attached to services that utilize extra capacity during off-peak hours, or to use of business services used in residential neighborhoods, or vice versa. Conversely, high data rate premium services could be supported for those customers who demand it, with the extra cost of providing that service passed through in the form of higher prices.

Kinds of Spectrum

Several different classes of licensed spectrum can be envisioned, distinguished by their sharing status. They are:

Exclusive use, by frequency and geography – This is the current licensing model. The licensee gets exclusive use in the authorized band and market. It is at the discretion of the operator what air interface and services he offers. In the current US usage, these are all frequency duplex, with the duplex spacing specified by the license.

Multiple use, by frequency and geography – Two or more operators get use of the authorized band and market. It is up to those operators how they implement the sharing and the uses to which they put its capacity.

Uncommitted Pool – An uncommitted band, over a particular geographical area. Capacity could be purchased, as needed, in particular areas. That purchase could be by advertised price, or by real-time auction.

It could be anticipated that most operators would opt for a partially exclusive license, to support their core services, with a shared component to support other, more competitive, variable services.

Duplexing in the shared-use options is a question for debate. One possibility is that there might be two (or more) sub-categories: one with the same, fixed duplex spacing as the exclusive licenses, and another category where uplinks and downlinks are managed separately. Only the most basic technical regulatory constraints are imposed on the licensees, aimed only at other-service interference control, and health and safety issues. Asymmetric allocations are often mentioned, under the assumption that human typing data rates are much lower that video screen painting rates.

Implementation of Shared Allocations

The non-exclusive shared licenses could be managed in several ways. Some examples follow.

Dynamic Footprint (Negotiated Frequency-Geography Partitioning)

A frequency-time partitioning could be negotiated among the shared licensees. Functionally it would be similar to the exclusive license, but done privately rather than by regulatory action. It also might change with time, subject to a real-time, automated negotiation. Any particular band, in any particular station, would be used by only one operator at a time.

This presents a problem of what to do at the dynamic boundaries between operators. The simplest procedure, that of guard bands and guard zones is also the most wasteful. Other procedures, involving multiple support at the border cells can be envisioned, but it a technical challenge. For example, in IS-95-like systems, the border cells could be populated with pilot beacons, which would be used to trigger a frequency change for those subscriber stations that need to handover across the boundary. Those subscribers would be moved to the exclusive portion of his operator's band before executing the handover.

Non-orthogonal Overlay

The sharing participants could negotiate a technical overlay. Two different air interfaces would be supported through a common antenna in the same band. They must cooperate on issues of power control so as to properly share the available channel capacity. The capacity used by each could be either held constant, or could itself be negotiated real-time in accordance with a pre-arrangement.

There is a performance penalty for the non-orthogonal overlay, as the incompatible signal manifolds mutually interfere *as radiated* from the base station. Orthogonal overlay or a virtual system is preferable.

Orthogonal Overlay

An operating model in which orthogonality is imposed in the manifold of signals radiated from one station, but they are otherwise logically independent. There is no proposal corresponding to this kind of concept.

The virtual system is probably the optimal variant of this, although it poses regulatory and business model questions.

Virtual System

A virtual system is one in which a single, high capacity, common downlink is used to support multiple operators. The uplink could be either the same air interface protocol, or something completely independent, as the level mutual interference of the uplinks is not normally dependent on what air interface they use. The appearance of separate systems is synthesized via the network software. The network architecture, that is, whether traffic is separated by the base stations themselves, or it is carried out at a central facility (BSC or MSC) is an implementation issue.

This is the form of sharing that carries the least risk.

Kinds of Base Stations

Existing systems have been deployed largely independently, although in some cases antenna towers are shared. In a non-exclusive license scenario, however, there are strong incentives to collocate facilities, and for several reasons.

Non-collocated Base Stations

This is the current model in most cellular and PCS installations. It must be restricted to exclusive band plan operation because the near-far problem would created intolerable mutual interference in a common band. Their base stations and subscribers would severely jam one another in cross-system proximity situations.

Non-collocated base stations can use nominally exclusive bands simultaneously, but they will have to observe guard zones and guard bands in order to avoid incidental interference.

Collocated Base Stations

Collocated stations are technologically very helpful in all sharing schemes.

Collocation helps the out-of-band emissions problem in exclusive band plan operation because the spurious emissions will have a fixed relationship to the desired signal. The interference level is set at the transmit antenna and is insensitive to subscriber set location.

Collocation is *required* for all frequency overlay scenarios because of the power control issue. Achievement of maximum channel capacity requires that *all* in-band signals, desired and undesired, arrive at the receiver at the minimum necessary power. This cannot be achieved without collocation.

Receiver Intermodulation Problem

It has been found in the early deployments of CDMA systems that serious interference problems occurred in CDMA handsets when they were too close to AMPS stations operating in an adjacent band. This is caused by nonlinear mixing products of the AMPS carriers, generated within the receiver front end, that fall inside the IF passband. This is a problem that will always exist at some level: there always will be a "kill zone" around alien base stations due to such intermodulation products. The only question is the size of that zone. While collocation is not *required* in exclusive band plan situations, doing so will largely mitigate the co-channel intermodulation problem.

Most historical experience with intermodulation problems has been in AMPS-CDMA adjacent channel situations. However the problem is not restricted to narrowband-on-CDMA. It does not seem to have been encountered much by PCS operators yet, but as the D, E, and F block operators begin to turn on, CDMA-on-CDMA intermods can be expected to become an issue as well.

Secondary Benefits

The sharing of RF transceivers and antenna structures, while primarily motivated by considerations of communication performance, also has favorable side effects from the standpoint of public policy. Sharing facilities reduces the costs, over all operators, of the inevitable legal wrangling over local zoning, environmental issues, real estate, and maintenance costs. This is a savings in the cost of doing business, and should ultimately benefit the end consumer.

Global Control Channel

The recent discussions in the ITU over the global control channel, an agendum for WRC-2000, have missed the point. A global control channel of some kind is needed in order to carry out the dynamic spectrum management that we have outlined. It would broadcast the band plan, which might change over time and geography in accordance with the sharing agreements and procedures. It might serve, in fact, as a truth-in-advertising vehicle, i.e. it could offer a shopping list of services and billing rates from which the subscriber set could select the purveyor of choice, based on subscriber-defined service choices and price criteria.

What is the Ether[1] Worth?

While this is obviously a politically charged issue, a regulatory regime that *leases* spectrum, with use-based payments would appear to present an attractive alternative to the one-time auction process. But how would this work in the context of station-by-station, hour-by-hour changes in band configuration and load? What constitutes "use" of the electromagnetic spectrum for billing purposes?

Clearly "consumption" of ethereal capacity by communication links is related to the energy radiated and its geographical distribution. The radiation from each transmitter degrades the noise figure of all non-targeted receivers to some degree. The deleterious effect of a transmitter can be viewed as a form of receiver heating within its antenna coverage footprint. To the extent that the interference is equivalent to Gaussian noise over the signal bandwidth, the apparent noise figure of impacted receivers will rise due to the interference. The fee might be viewed as a sort of pollution charge based on the net degradation to those other potential users.

A plausible metric might be a time integral of radiant power flux, averaged somehow over the non-targeted receivers that are adversely impacted by the radiation. In practice this probably would take the form of a negotiated weighting factor derived by integrating the antenna pattern over a demography map. If all the radiation can be concentrated at the intended receiver, and no others, the adverse impact, and hence fee, is zero. The incentive is to implement the system with the minimum possible energy needed to communicate. And the metric is indeed *energy,* not power. A large power radiated for a short time is treated as equivalent to a small power radiated for a long time if their time integrals are equal. For any particular combination of modulation and coding, error rate performance can be characterized by the signal-to-noise ratio E_b/N_0.

Business model

The notion of shared spectrum licenses, while attractive technically for the reasons already cited, presents some interesting business model questions. Technical considerations very much favor implementation of shared or community facilities, encompassing primarily the base station transceivers and antennas. There are also economies of scale and favorable effects on zoning boards and local jurisdiction site licensing. However, from the standpoint of regulatory policy this would appear to be a step *backwards*, contrary to the trend toward deregulated utilities, transportation, and communication companies. Who might own and operate such shared facilities? Would it be a regulated, monopolistic entity akin to the old Bell System? Would it be a joint venture of the local operators? Or would it be a wholesale operator who would own and operate the radio infrastructure. The wholesaler would resell

---

[1] Ether: The pervasive, infinitely elastic, massless, medium postulated, erroneously, by 19th century physicists, as the medium of propagation of electromagnetic radiation. The term is used here as a picturesque placeholder for the asset value of electromagnetic spectrum.

capacity to other business entities, who would be the visible purveyor of wireless services to the consumers. Whether the wholesaler owns the network too, or stops at the base stations, is open to debate.[2]

It is possible here that there might be a role for some central, regulated monopolistic entity to operate what might be regarded as a national wireless information utility. This would be, obviously, a very controversial proposal in this era of world wide deregulation and privatization of government owned utilities. There are, however, strong analogies with other kinds of persistent utilities like water, sewer, postal service, electric power. Clearly any such proposal would be the subject of a spirited public debate.

Recommendations

The Commission, in its infinite wisdom, would do well to explore this concept in some detail, consulting with the expert communities in communication technology, public utility policy, business, and economics. I suggest the following concrete objectives for such consultation and rulemaking:

1. Determine the viability of such a creative new regulatory regime permitting shared spectrum licenses. This might include a study of the viability of the broadly tunable mass market subscriber sets that would be needed to support such a regime.

2. Determine the minimum set of technical issues that would need regulation.

3. Draft strawman regulatory policy and licensing procedures.

4. Search the current United States spectrum allocations for candidate bands suitable for a reassignment to this new service, with a preference for bands below 3 GHz. One of the primary purposes in suggesting this flexible policy is to permit a degree of ad hoc spectrum management, so these candidates need not necessarily align with the ITU identifications. This is particularly desirable given the current disconnect between ITU Region 2 and the rest of the world in the PCS bands.

5. Initiate a formal inquiry in accordance with Commission procedures.

---

[2] It is perhaps noteworthy that this is more or less the business model that was attempted by Nextwave Telecom in their C-block PCS system. While this failed for other business reasons, the concept is technically sound.