Spectrum Issues: A Different Perspective

Brett Glass, *LARIAT.NET*
P.O. Box 1693
Laramie, WY 82073-1693
http://www.brettglass.com/mailbrett.html

Abstract

The allocation and licensing of portions of the radio frequency spectrum has engendered much controversy – more so than ever in this “wireless age”. This informal paper reframes the issue of spectrum allocation in terms of its roots in classical signal processing theory, and explains how accidents of history, physics, mathematics, engineering, and 19th Century entrepreneurship have caused the spectrum to be regarded as the most scarce, and hence valuable, “turf” of the 21st Century. It suggests ways in which a different view of the problem may lead to more sensible regulation and more harmonious sharing of the airwaves.

1 Introduction: Sines of the Times


Figure 1: Alexander Graham Bell. (Public domain photograph)

In the late 1850s, inventors Alexander Graham Bell and Elisha Gray – in an effort to increase the number of transmissions which could be sent simultaneously over scarce and expensive telegraph wires – simultaneously invented the “harmonic telegraph.” Their invention used a set of tuning forks coupled to coils similar to modern day guitar pickups. Each tuning fork generated a distinct signal – a sine wave – on a pair of telegraph wires. This signal caused a similarly pitched tuning fork – but not others which were coupled to the same wire – to vibrate. Multiple telegraph messages could thus be sent at one time over one pair of wires.

While this invention was eclipsed by the development of the telephone – on which both inventors also worked – it was significant in that it was the first use of *frequency division multiplexing* – a technique in which multiple signals can be transmitted simultaneously via the same communications medium without interfering with one another. Not long thereafter, the technique was later applied to telephony with the development of the *analog channel bank*, which allowed multiple telephone conversations to be transmitted over the same coaxial cable. Like the harmonic telegraph, the system used sine waves, at different frequencies, to carry each conversation.

When the first radio broadcasts began, they were received by radio sets whose designs evolved from “crystal sets” to super-regenerative circuits to superheterodyne circuits. All of these designs used simple tuned circuits constructed from coils of wire and capacitors. And to function, each of these circuits relied on the mathematical properties of – you are probably starting to detect a pattern here – sine waves.


Figure 2: RCA Superheterodyne Radiola (Early marketing photo)

Sinusoidal motion occurs in many places in nature. If you look, edge-on, at a stick attached to the spoke of a wheel as the wheel rotates, it will move up and down continuously in a sinusoidal pattern. The displacement of a simple pendulum – at least when the oscillations are relatively small – is likewise sinusoidal.

2 Sine Waves and Classical Signal Processing Theory

Introductory signal processing classes – often titled "Signals and Systems," "Linear Systems, " or "Introduction to Signal Processing Theory," usually begin with the introduction of Fourier's Theorem, which bears the name of French mathematician Jean-Baptiste-Joseph Fourier. This theorem states that any signal – no matter what its characteristics – can be expressed as the sum of a (possibly infinite) series of sine waves.

Figure 3: Jean-Baptiste-Joseph Fourier (19th Century engraving)

The proof is then generalized; students are shown that any arbitrary signal can be decomposed into one or more signals selected from a set of "proto-signals" called an *orthogonal basis set*. The word "orthogonal" refers to a special and useful property of a basis set: to wit, that any signal can be represented by one, and only one, linear combination of the signals in the basis set. There is no ambiguity; all listeners will decompose the same signal in the same way.

The instructor demonstrates that the set of all sine waves – at all frequencies and phases – forms an orthogonal basis set. He or she discusses what the results are when one decomposes certain common signals into a set of sine waves, and introduces the Fourier transform as a way of performing this decomposition.

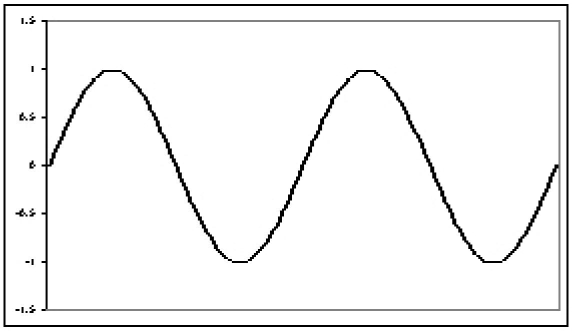

Figure 4: A sine wave. Fourier's Theorem states that any function – periodic or not – can be represented as a combination of sine waves of different periods, phases, and amplitudes.

This is usually the last students hear of the concept of an orthogonal basis set. All of the rest of the material will most likely take it as read that all signals are to be analyzed – as per tradition – in terms of sine waves.

As the course continues, students learn that a radio receiver does its work by identifying signals of interest and tuning out the rest. It starts with everything it "hears," filters out undesired portions of the frequency spectrum (that is, the basis set of sine waves), then decodes the desired portions to recover the information that has been sent.

To this day, all radio receivers still work by decomposing what they receive into sets of sine waves. Some still employ filters and resonant circuits which are electrical analogues of the tuning forks which Alexander Graham Bell used in his harmonic telegraph; others use digital circuitry to accomplish the same thing.

3 Reframing the Problem

The concepts described in the preceding section allow us to reframe the issues of spectrum allocation, radio interference, and spectrum scarcity in a very simple and useful way. The key insight here is that s*pectrum allocation can be thought of, very simply, as a way of divvying up an orthogonal basis set.*

Because radio signals, or electromagnetic waves, are changing electric and magnetic fields which can be detected at a distance, the right to use a particular portion of the radio spectrum is – in essence – the right to wiggle electric and magnetic fields in a distinctive way. If all receivers decompose what they hear in the same way, they can "tune in" your wiggles and "tune out" everyone else's. To put it another way, if one is keeping to one's section of the spectrum, the signals emitted –

when decomposed – all fall within a particular portion of the basis set consisting of all possible sinusoidal radio waves. The portion of the set which is allocated is often called a “channel” or “band.” (In many cases, a user is allocated two paired channels some distance apart – one for receiving and another for transmitting. This facilitates *full duplex* operation – transmitting and receiving at the same time – because it’s easier to filter out one’s own outgoing signals when listening for an incoming one.)

The notion of sine waves as the only conceivable choice of basis set is so pervasive that many – even scientists – neglect to consider that, but for the peculiar way in which human technology evolved, we might all be using an entirely different one. This prejudice is, unfortunately, built into the world’s largest distributed computing project: the Search for Extraterrestrial Intelligence, or SETI. According to the project’s Web site, SETI “uses radio telescopes to listen for narrow-bandwidth radio signals from space. Such signals are not known to occur naturally, so a detection would provide evidence of extraterrestrial technology.”

Alas, there is no reason to assume that a non-human civilization would just have happened to evolve as ours did, moving from acoustic tuning forks to vacuum tubes to transistors to Von Neumann-style computers performing fast Fourier transforms. Perhaps the alien counterpart of Fourier happened upon an entirely different orthogonal basis set for electromagnetic signals. Given the infinite number of possibilities, it seems highly unlikely that an alien civilization would have just happened to pick sine waves as a basis set, tussle politically over frequency allocation, and allocate narrow ranges of frequencies to particular users whose signals we could detect from light years away. It seems far more probable that the aliens sought by SETI’s antennas, if they exist, are doing something altogether different. And if they did indeed use a different set of basis functions – let’s call them “gleeps” – we would need to understand the design of that set of signals before we could even recognize, much less decode, what they were transmitting. Thus, the project’s chances of success would be small even if we abandoned our anthropocentric, sine wave-oriented perspective. It simply is not possible to know which of the infinite number of possible views of electromagnetic signals we would have to adopt to recognize an alien transmission.

4 Pushing the Envelope: Alternative Basis Sets

It has only been relatively recently – during roughly the past two decades – that Earthly scientists and engineers have begun to explore the possible uses of alternative basis sets for signal processing and analysis. These include, for example, the *Walsh functions* (endless trains of square pulses from which an arbitrary signal can be constructed). Among the most studied alternatives are *wavelets* – sets of signals which, unlike the endless sine waves used in Fourier analysis, are finite in duration. The basis set consists of similarly shaped pulses that are shrunk, stretched and shifted in time.

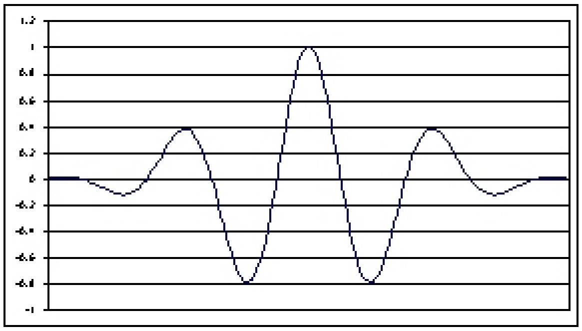

Figure 5: A Morlet wavelet is a pulse which is shaped like a sine wave (actually, a cosine wave) fitted inside a Gaussian “envelope.” Because it “dies off” quickly after a few periods of the sine wave on either side of the center, it has characteristics of both a time-limited pulse and a sine wave.

Decomposing a signal into wavelets can provide more useful information than a classic Fourier transform of the same signal. For example, it can be proven that a signal that is time-limited (a property of all the signals we actually use in the real world) cannot be frequency-limited; that is, its Fourier transform must go on forever. (In fact, the more “sharp edges” a signal has, the higher the frequencies one must use to reproduce it.) However, the wavelet decomposition of a time limited signal can be relatively compact while still capturing as much detail of the original signal as desired. Some wavelets – called "wave packets" – are intentionally shaped like waves which increase in amplitude and then die out. In this case, the wavelet decomposition of a signal may convey some information about its frequency as well as about its behavior over time. (The Morlet wavelet, shown above, is one example. It is, essentially, a "sine pulse" that reaches maximum amplitude at the origin and fades in the same way as the well known Gaussian "bell curve" on either side.)

5 Pros and Cons of Sine Waves as a Basis Set

Is it a good thing or not that mankind has, at least for the nonce, chosen to use sine waves as the basis set which we divide up so as to keep our signals straight? Let's look, briefly, at some of the pros and cons.

Pro	**Con**
• Easy to construct physical oscillators and filters (e.g. tuning forks, tuned circuits consisting of capacitors and inductors) • Corresponds, intuitively, to aspects of human perception such as color and pitch. (The cochlea is, in fact, a real time audio spectrum analyzer.) • Technology for separating and distinguishing signals by frequency is now well developed • Narrowband antennas are simpler to construct and better understood than wideband antennas (at least today) and "naturally" filter by frequency • Frequencies of signals are relatively easy to shift via mixing with sine waves ("heterodyning") and then filtering • Confining signals to a limited space (e.g. indoors) is easier if absorption spectra of materials can be exploited • It's tough to change now, due to massive investment in the existing regime!	• "Heisenberg uncertainty principle" dictates that frequency cannot be exactly determined at any point in time; therefore, no filter can be perfect and no slice of spectrum is 100% usable (there's always waste at the edges!) • A signal that's time limited cannot truly be frequency limited, again causing "slop" • Some frequencies have characteristics (e.g. "skip") which can cause unexpected interference at a great distance • Systems designed for a specific frequency are not easily "re-tunable" to others, making reallocation difficult (not only on the fly, but even with long notice) and expensive • Varying absorption spectra of materials (air, water, buildings) cause some frequencies to be "beachfront property" while others are of limited use. A different basis set could equalize the practical value of different allocations, eliminating some of the artificial scarcities caused by allocation by frequency • Some alternative schemes could make "sharing" of the airwaves easier due to increased receiver agility

Table 1: Pros and Cons of Allocating Transmission Rights by Frequency

6 The Shannon-Hartley Theorem (AKA Shannon's Law) and its Implications

The Shannon-Hartley theorem, first proposed by Ralph Hartley and proved by Claude Shannon in his classic paper *Communication in the Presence of Noise (Proc. Institute of Radio Engineers, vol. 37, no.1, pp. 10-21, Jan. 1949)*, specifies the theoretical limit on the amount of information that can be transmitted via a communications channel with a particular bandwidth and signal to noise ratio. It states that

$$\mathbf{C = BW \times \log_2(1 + SNR)}$$

where

C is the capacity of the channel in bits per second (also known as the Shannon Limit);

BW is the bandwidth of the channel in hertz;

SNR is the signal-to-noise ratio of the channel.

This theorem has important implications for all methods of sharing the airwaves. Some critics of spectrum allocation schemes – in particular, those who believe that there should be few or no rules governing the use of the airwaves – would have us believe that if only the rules regarding how one transmitted signals (be they based on frequency or any other scheme) were lifted, new technology could magically make interference a thing of the past. However, even if there is a very efficient way of distinguishing a desired signal from an undesired one (perhaps a better one than filtering by frequency), Shannon's Law dictates that the interference will reduce the capacity of the channel. That is, unless it's possible to build a perfect filter, interference will *always* limit one's ability to transmit information and successfully receive it at the other end. (Note that while the concept of "bandwidth" assumes allocation by frequency is not directly applicable to a basis set other than the set of sine waves, the proof in Shannon's paper can adapted to define an analogous quantity for other basis sets.)

There is thus no "magic technological bullet" that can eliminate all interference. However, it may be that allocation schemes not based on frequency could make it easier to minimize interference and eliminate the vast wastelands of unused and underused spectrum which we see under the current regime.

6 Implications

Understanding what we are doing when we distinguish our signals by frequency – and knowing that there are other options – is intellectually interesting. But does it help us to deal more gracefully with the technological, political, economic, and social problems which we face as we move to an increasingly wireless society? More to the point: Given the vast investment Mankind has already made in allocation of transmission rights by frequency, does it pay to consider other options?

Arguably, the answer is "yes." Some technologies – such as "Ultra-wideband" radio – successfully "fly under the radar" of the existing allocation scheme by causing only a small amount of easily overcome interference to any individual user who still conforms to a spectrum-based allocation scheme. Experimentation with this and similar technologies will allow us to weigh the cost of moving to a different allocation scheme against its benefits, and may also pave the way for a graceful transition to new ways of sharing the airwaves. (Of course, such systems must be designed not to interfere with one another. They thus must share a new basis set rather than simply relying upon the fact that no one else is using the same scheme that they are.)

In any event, it is helpful to recognize that when we lobby Congress or the FCC to change the rules or allocate a scrap of spectrum, we are contending more for a mathematical abstraction than for immovable turf. We thus have a better chance of engineering our way around the problem – insofar as we can – for everyone's mutual benefit.