RTT TECHNOLOGY TOPIC
January 2006

The Bit Rate Race

The November and December Hot Topics talked about phones with extended voice, audio, image, video and data capture capability.

To start the New Year, we are going to look at the impact of these phones on future wide area wireless mobile/cellular radio systems.

The shift in phone form factor over the past three to 5 years towards more data centric applications has triggered a 'bit rate race'.

Initially the 'bit rate race' focussed on downlink data rates but there is now an increased recognition that many applications either require balanced bandwidth (uplink/downlink symmetry) or in some cases, uplink asymmetry.

There is also an increasing recognition that many applications are latency sensitive (conversational voice, conversational voice and video and interactive gaming). End to end delay and end to end delay variability are therefore important performance metrics that need to be accommodated in radio system link budgets.

Wide area wireless systems add substantial overheads in terms of channel coding and signalling which add to the 'end to end' cost of delivery.

The need for higher data rates and controlled end to end delay characteristics has led to ever more elevated marketing claims that new physical frontiers have been breached by a mix of revolutionary new techniques.

We attempt to put these claims into an engineering and historical perspective.

The cost of wide area in wireless networks

Superficially you might think that the larger the area covered by a cell site, the lower the cost.

This is determined however by whether the system is coverage or capacity limited. If the system is capacity limited, then the cost benefits of range become immaterial.

In cellular networks, some cell sites will be capacity limited some will be coverage limited.

The standard dictates the maximum cell radius.

For example GSM was specified in the late 1980's with a maximum cell radius of 35 kilometres (later extended by the Australians to 70 kilometres).

This provision introduced the need for a relatively complex and power and bandwidth hungry timing advance protocol (to take into account the round trip delay to and from the base station). Wide area also implies additional channel coding. Wide area therefore has a direct impact on power and spectral efficiency.

The cost of mobility in wireless networks

The same principle applies to mobility. The more mobile a user, the greater the signalling bandwidth needed (for power control, handover and cell to cell registration). Highly mobile users also need additional channel coding.

The standard dictates the maximum mobility supported by the radio system.
For example GSM was specified to support user mobility up to 250 kilometres per hour (later extended by the French to 500 kilometres per hour).

As a direct consequence of this wide application requirement, 62% of GSM channel bandwidth is absorbed by channel coding and signalling overhead - the 'cost' of wide area mobility.

Thus whenever wide area systems are to be compared in terms of spectral and/or power efficiency, it is always important to consider the dynamic range over which the network is expected to operate.

Higher data rates do not necessarily translate either into greater power or spectral efficiency.

Just as cell radius and mobility have a cost, data rate has a cost. Sending data at twice the rate does not halve the power consumption nor does it double the spectral efficiency.

It is a bit like running. Running twice as fast gets you there in half the time but overall more energy will have been expended.

In a radio system context, more complex modulation schemes may deliver a small net gain in terms of DC power efficiency versus throughput but the gain will be far less significant than sometimes suggested.

The Role of Standards in achieving an acceptable performance/power and spectral efficiency compromise.

It is one of the jobs of a standards making body to try and achieve an acceptable compromise between power and spectral efficiency, data rates and application 'dynamic range'.

The best technical option may not be the best commercial option and individual or group vendor interests may dominate the process particularly if IPR value is involved.

As you would expect, wide area wireless standards are not an exception to this rule and are driven by a mix of self interest moderated by common interest.

There are four standards making bodies/interest groups presently addressing higher data rate mobile wide area wireless.

The **Wimax** standard is overseen by the IEEE IEEE 802.16

The Wimax Forum www.wimaxforum.org exists to manage the marketing of WiMax and to address interoperability and verification issues.

802.16 e has developed out of 802.16 but adds in mobility management (and the related signalling and bandwidth overhead) to support mobility users. The 'headline' data rate is **15 Mbps** in a **5 MHz** channel with a maximum cell radius of between **2 and 5 kilometres.**

WiMax represents the first implementation of OFDM in bi directional wide area wireless. It builds on a number of year's experience of OFDM in low power WiFi and high power radio broadcast systems.

A separate but related standards initiative within the IEEE known as IEEE802.20 addresses IP protocol optimisation for high mobility applications (defined as up to 250 Km/h).

The **CDG** (Cellular Development Group www.cdg.org) promote **CDMA2000**. This is the original CDMA standard which has been in widespread use in wide area networks in the US, Latin America, and Asia (particularly Korea). It is used in the 800 and 1900 MHz cellular bands but is also being presently actively promoted for use in the 450 band in Europe (competing with GSM450) and potentially also in parts of the 700 MHz bands coming available as analogue TV disappears.

It was the first standard to add a code multiplex to the frequency and time division multiplex used in other 2G cellular systems. The code multiplex is built on 64 orthogonal Walsh codes (when multiplied together orthogonal codes produce a zero value). The Walsh codes are used to deliver additional sensitivity (spreading gain which translates into range/coverage gain) and selectivity (which translates into capacity gain).

CDMA 2000 is deployed in 1.25 MHz channels and uses QPSK to deliver data rates of **1.288Mbps**, 8PSK to deliver **1.8432 Mbps** or 16 level QAM to deliver **2.457.6 Mbps.** The present radio layer solution for non real time IP traffic is known as 1XEVDO Release 0 but with a migration path to a newer release known as Rev A.

Rev A increases the **forward link** peak rate to **3.1 Mbps** and the **reverse link** speed to **1.8 Mbps**. Rev A also includes Quality Of Service Support for low latency packet applications which will allow symmetric real time IP voice and IP video applications to be deployed.

Rev A is positioned as being the first iteration of a scaleable bandwidth proposition with channel bandwidths of **20 MHz** supporting a **forward link** of **46 Mbps** and a **reverse link** of **27 Mbp**s with the option of adding an OFDM multiplex to reduce channel symbol rate. The CDMA2000 standardisation process leads to ITU standards that are supported by ETSI (the European Standards Institute) under a work group known as 3GPP2.

3GPP1overseas the future evolution of the **W-CDMA** (wideband CDMA) standard presently deployed in 3G networks in paired frequency bands at 1.9 and 2.1 GHz using 5 MHz channel bandwidths.

Present WCDMA radio systems are Release 99 compliant. Release 99 uses a relatively complex code structure to support variable rate voice and data channels. The 64 Walsh codes used in CDMA2000 are reordered into a 256-code structure, which allows data rates to be changed every 10 milliseconds. Effectively it is an ATM friendly radio layer that can be closely matched to ATM transport used in the IP RAN and IP core network.

There is nothing wrong with the Release 99 code structure though it does depend on quite aggressive power control to maintain code orthogonality. This in itself absorbs a significant amount of channel bandwidth and channel power. Call set up is also rather over complex.

Release 5 simplifies the code structure to a 16 code implementation of the original 256 codes and adds 16 level QAM as a modulation option. This is the basis of HSDPA (high speed data packet access) and is positioned as an 'IP friendly' radio and MAC layer solution.

The 10-millisecond frame used in Release 99 is split into 5 sub frames of 2 milliseconds. In practice this means that the radio layer is simpler. The MAC layer is more complex.

Release 6 is similar to Rev A 1XEV in that some of the issues of end to end latency are addressed in addition to an enhanced uplink (HSUPA).

The highest achievable data rate in HSDPA (the headline data rate) is **14.4 Mbps** though **10.8 Mbps** is generally considered to be more generally realisable (and assumes the user is close to the base station).

HSUPA supports an uplink with a maximum headline data rate of **5.76 Mbps.**

Release 7 introduces an OFDM multiplex (initially on the downlink only) and 64 level QAM. This is to be known as **HSOPA** (O as in OFDM) and includes the option of using **20 MHz channels.** The 2 millisecond sub frames used in HSDPA will (probably) be split into **.5 millisecond sub sub frames** in order to increase admission control flexibility and decrease admission control latency (important for conversational and interactive applications using a contention based MAC).

Release 7 is sometimes now described as Super 3G but really just represents a logical migration from the existing standards base. **100 Mbps** data rates are suggested as being possible at some future stage (close to a base station).

W-CDMA is part of a wider standards initiative known as UTRAN-LTE (**U**niversal

Mobile **T**errestrial System **R**adio **A**ccess **N**etwork **L**ong **T**erm **E**volution) which aims to optimise future radio layer functionality with advanced network admission and traffic control protocols.

The fourth standards initiative was put forward to the ITU in 1998 by the China Wireless Telecommunications Group and is called **TD SCDMA.** (www.tdscdmaforum.org)

It was accepted by 3GPP as a candidate 3G wide area mobility access standard in March 2001. It uses time division duplexing combined with dynamic channel allocation combined with smart antennas combined with joint detection combined with uplink synchronisation. Note that these are all radio system features and functions that are supported in the other standards. It is way the techniques are used together which forms the basis of the claimed differentiation.

TDSCDMA lacks the implementation footprint of CDMA2000 and WCDMA but has strong political support in China. Mobile operators in China presently lead the rest of the world in terms of annual growth rate and margin. This, together with some determined and capable vendor involvement suggests that TD SCDMA cannot be lightly dismissed and indicates that China will be progressively more involved in future wide area radio standards.

Wireline Broadband as a benchmark

Ideally all four wide area wireless options would deliver a user experience that is similar to or comparable to a wired internet connection.

This implies a need to consider next generation broadband, specifically VDSL or even more specifically, VDSL2 (the latest flavour of VDSL) as a performance benchmark.

VDSL2 Profiles (Bi-directional bit rates)		
High Profile	Medium Profile	Low profile
30 MHz	12 MHz	8 MHz
100 Mbit/s	30 Mbit/s	25 Mbit/s
350 metres	900 metres	1.8 kilometres

There are three profiles for VDSL2, high, medium and low. All three profiles use discrete multi tone modulation (OFDM) and are differentiated by the frequency cut off point of 30 MHz, 12 MHz or 8 MHz.

The higher the frequency, the higher the bit rate and the shorter the distance (given that attenuation in copper increases with frequency).

For example the high profile 100 M/bit/s service needs to be within 350 metres of a fibre node.

Some observations;

This is bi directional bandwidth and marks a clear shift of emphasis from the downlink asymmetry model used in ADSL.

As with ADSL, the 'gold standard' (used to calculate the reach/range figure) is based on a maximum allowable bit error rate of 1 in 10 to the 10. These are good quality bits. Bit quality can be accurately specified because copper is a predictable and relatively stable communications medium (particularly when compared to wide area wireless).

There is probably some way to go before copper runs out of steam so it is likely that future new VDSL options may emerge that trade additional processor bandwidth for higher end to end bit rates.

Wireline VDSL therefore provides wireless (including wide area wireless) with a continuously moving performance benchmark.

Commonalities, Capacity, Coverage, Cost and Complexity Comparisons
So of the four wireless wide area candidates, which is the best option for providing a 'wireline equivalent' broadband end user experience?

The answer is all of them or none of them.

All four candidates are by necessity compromise solutions.

All four solutions have 'headline' data rates that are only achievable in ideal operational conditions.

Real user data throughput rates are often significantly lower.

With all wide area systems, channel bandwidth is absorbed by channel coding and signalling overheads.

Signalling overheads are a consequence of mobility and the need to implement power control and handover algorithms.

Channel coding is needed to counteract the effects of a fading channel (high error rates and burst errors). Traditionally, cellular networks have been designed around bit error rates of 1 in 10 to the 3, well short of the ADSL/VDSL benchmark. Channel coding can control error rates and burst errors but the 'cost' is reduced spectral and power efficiency (and an increase in the latency budget).

For internet protocol traffic, IP address overhead and TCP/IP send again protocols introduce additional bandwidth 'cost factors' which in radio engineering terms directly hit the radio system link budget (IPV6 address overheads in IP voice for example take about 3.5 dB off the link budget).

To an extent, radio processing and signal processing techniques can be used to hide these effects. These techniques are common across all four standards. They include adaptive modulation and coding, fast automatic repeat mechanisms, smart antennas and MIMO (multiple input/multiple output) techniques, multi user detection and an active mix of frequency, time and code multiplexing.

These techniques are really not a significant basis for differentiating one standard from another.

A more significant performance and cost differentiator is the frequency at which the standards are deployed. For example a network deployed at 1900 MHz will have a propagation loss at least 8 dB higher than a network deployed at 900 MHz. As a rough rule of thumb, each dB translates into a 10% increase in network density.

Additionally, wide area cellular networks have traditionally been deployed in paired frequencies with a duplex separation between transmit and receive channels of 45 MHz (at 900 MHz), 95 MHz (at 1800 MHz), 80 MHz at 1900 MHz and 195 MHz (at 1900/2100 MHz).

This duplex separation delivers a significant sensitivity gain over the non-paired band implementations typically used in TDD schemes.

Consider that a handset in a wide area cellular system may be transmitting at between 20 or 30 dBm (100 mW to 1 watt) and receiving signals at -110 dBm (.01 of a picowatt) - a dynamic range difficult to realise in non-duplex spaced systems.

TDD schemes however have some increasingly relevant advantages. For example, a simpler power control requirement and the ability to support rapid changes in uplink and downlink asymmetry.

These are practical deployment considerations that have a tangible and describable impact on cost and are generally more significant than the difference between competing candidate radio technologies.

Claims of step function gains in performance for any particular technology or technique are almost always hard to justify. Improvements tend to be incremental and are constrained by processor bandwidth limitations. For example, OFDM when combined with optimised packet control can deliver some useful performance advantage. A recent White Paper from Flarion 'Trains, Planes and Automobiles' sets out some of these benefits in terms of mobility performance.

OFDM does however require additional linearity (which generally involves additional processor bandwidth in the transmit and receive chain) and a Fourier transform (*see the October 2005 Hot Topic for a more detailed review of the processor load implicit in OFDM). Similar constraints apply to smart antennas, MIMO systems and active interference cancellation techniques.

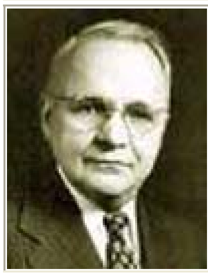
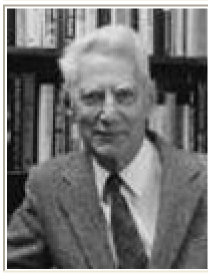
One solution is to have handsets that can access any or all of the wide area wireless systems depending on whichever happens to be available and offering the highest data rates (strongest /cleanest signal).

Unfortunately this requires a complex process of IPR ownership to be resolved across all four standards. The most likely outcome here would be that a phone would have an insupportable IPR component cost.

Someone also has to decide on the admission/connection algorithms that drive access policy. Is 'best connect broadband' best connect for the user or the operator?

The Laws of Physics in Radio System Engineering
All wireless systems are subject to the laws of physics. The laws of physics are however subject to reinterpretation over time and depend on the context within which they are applied.

Which brings us to consider the collective work of three men, Mr Agner Krarup Erlang (1878-1929), Mr Harry Nyquist (1889-1976) and Mr Claude Elwood Shannon (1916 to 2001).

Information theorists of the Month		
Agner Krarup Erlang **1878-1929**	**Harry Nyquist** **1889-1976**	**Claude Shannon** **1916-2001**
		

AK Erlang's two papers, 'The Theory of Probabilities and Telephone Conversations' in 1909 and 'The solution of some Problems in the Theory and Probabilities of Significance in Automatic Telephone Exchanges 'in 1917 provided the mathematical basis for calculating offered traffic and blocked call rates in telephone networks for the next 80 years. Erlangs are still used to calculate offered voice traffic and the mathematical approach remains relevant in present work on data Erlangs, image and video Erlangs.

Erlang based traffic modelling however increasingly needs to be considered in the context of Nyquist and Shannon's subsequent work on information and transmission theory.

Harry Nyquist produced two seminal papers 'Certain Factors Affecting Telegraph Speed' in 1924 and in 1928 'Certain Topics in Transmission Theory'

These papers set out the principle that the wider the frequency band of the signal to be transmitted, the higher the signalling rate required to represent the information in that signal. More specifically, Nyquist proved that the signalling rate needed to be exactly twice the bandwidth in hertz of the signal to be transmitted. He went on to study the time domain characteristics of pulsed signals, specifically pulse spreading and the concept of inter symbol interference. These concepts remain directly applicable today.

Claude Shannon's famous contribution was to take Nyquist's work and to study the behaviour of digital transmission systems in the presence of noise. He started from the assumption that a message is selected from a set of possible messages and that the 'alphabet' needed for the messages is dependent on the statistics of the source signal. As an example, he studied the statistics of English texts. This prompted later work by Huffman on lossless coding techniques.

Claude Shannon was thus the first person to clearly describe the concept of information entropy and redundancy. He also clearly set out the principles of channel coding and error control. Both were described in 'A Mathematical Theory of Communication' published in the Bell System Technical Journal in July 1948.

Shannon's contribution to telecommunications theory is fundamental and unique. However although he explicitly described the theory of source coding and the principles of channel coding and error control, he considered the two processes to be separate and subject to separate optimisation.

As our understanding of voice coding, audio coding, image coding and video coding has improved, so we have come to understand that source coding and channel coding are intimately linked. Source coding reduces transmitted bandwidth; channel coding increases transmitted bandwidth. Both processes however share a common objective - to improve power and spectral efficiency.

So what does this have to do with wide area cellular radio system design?
Wide area implies multiple users all firing offered traffic at a base station.

Let's assume these users, or at least some of them, have a new generation of phones with extended voice, audio, image, video and data **capture** capability.

These devices use source **encoders.** Source encoders use a mix of lossless (Huffman and arithmetic), perceptually lossless and lossy source coding techniques (AMR voice encoders, AAC audio encoders, JPEG image encoders and MPEG video encoders).

These encoders are variable rate and can respond either to changes in the source entropy or to changes in available or allocated bandwidth. They are context based adaptive encoders optimised to work with adaptive channel coding schemes and adaptive admission control algorithms.

The encoders however require certain properties from the radio channel other than simple bit rate. For example, as compression ratios are increased in a lossy source coder, the source code becomes less resilient to bit error rates and sensitive to burst errors (temporal compression schemes like MPEG are for example particularly sensitive to burst error distribution).

Our four wide area wireless candidates are all IP friendly or at least they are all going in an IP friendly direction but this misses the point.

We need wide area wireless systems to be **JPEG and MPEG friendly.**

The capacity and coverage available from any wide area radio system is a direct consequence of the noise floor at the base station receiver, the noise floor in the handset and the transmitter phase accuracy in both directions.

The transmit and receive paths are both exercised not only by the amount of traffic flowing in each direction but the properties of the traffic flowing in each direction.

This may include multiple simultaneous encoded bit streams supporting simultaneous voice, video, audio and text, any or all of which can be variable rate.

This implies that coverage and capacity and cost calculations can only be attempted once some clear upper bound/lower bound parameters have been established as to the mix of traffic in both directions.

This mix includes a combination of voice, audio, image, video and data some of which will be conversational, some of which will be interactive, some of which will be streamed, some of which will be best effort. This applies whether we are dimensioning radio bandwidth or network (IP RAN/ core) bandwidth.

Surely faster bit rates deliver a system efficiency gain?

Well yes and no.

Faster bit rates can either be used to send a file faster or to support higher bandwidth real time services such as voice, voice and video, real time audio (captured in a handset and shared in real time with other users) or real time still imaging. A 'fatter pipe' (faster composite channel rate) will improve multiplexing efficiency.

However for real time traffic, end to end latency has to be carefully managed and there should be no end to end delay variability. This implies that send again protocols if used at all, must be invisible to higher layers in the protocol stack. It also implies that flexible routing cannot be used in the end to end channel.

The error resilience techniques developed for video streaming depend on frame to frame (inter frame) comparisons which require buffering. These techniques are not compatible with real time bi-directional communication.

So why do we need fast real time bi-directional bit rates?
Partly technology, partly business. The technology story is that we are adding enhanced image, video and audio capture capability to next generation handsets. Still images, video and audio objects can be stored locally in the phone. This helps to sell embedded and plug in memory but does not intrinsically add to the operator's revenue stream. Somehow, 'snap and store' has to be turned into 'snap and simultaneous send'.

Real time still imaging as an example
Let's take real time still imaging as an example.

The figure below shows a scaling model for imaging which for illustration goes from 8 Megapixel resolution one frame per second still photography to 70,000/300,000

QVGA/VGA video at 30 frames per second.

Still Image		Moving image
8 Megapixel 1 frame per second		70,000/300,000 pixel 30 frames per second
JPEG	Motion JPEG	MPEG

You could scale this more dramatically if you wanted. There are CMOS sensors available today, which support frame rates up to 3000 frames per second.

But let's stick with individual still images for the moment.

An 8 Megapixel 24 bit colour depth uncompressed image makes a 24 Megabyte file. This can be compressed to a megabyte or less without a catastrophic loss of quality so it's not unreasonable to suggest that it is quite feasible to support the real time uploading of these images to an on line web site or the **real time** sharing of images with other mobile phones.

Note the image can be blended with an audio tag and/or voice tag and/or text tag and any other context relevant EXIF/XML metadata available from the phone, for example time, place and temperature.

It combines the benefits of a lower cost handset (less memory) and a new revenue stream for the mobile operator but (we would argue) the application has to be real time to be compelling.

Bi directional bit rate requirements come from the assumption that some of this bandwidth will be coming in the other direction from other mobile phones.

So in summary
We still tend to compare radio system technologies, in this case wide area system technologies using 'old fashioned' or at least inappropriate modelling and measurement techniques which are no longer valid in a mobile multi service network supporting mobile multi media phones some of which will have extended voice, audio, video, image and data capture capabilities.

The hardware and software form factor of the phone determines the offered traffic in both directions (encoder capabilities on the uplink, decoder and display capabilities on the downlink). The offered traffic has to be analysed in terms of composite bit rate and the properties that the offered traffic requires from the radio interface.

Much attention is focussed on the radio side on 'silver bullet' solutions such as MIMO and smart antennas, which promise step functions in capacity gain and or coverage. These promises are hard to achieve in practice.

As we shall see in future Hot Topics, one of the most significant areas for improving radio system performance is admission control and load management. In essence this is system level Multiple Input Multiple Output in which offered traffic is shared amongst multiple base stations.

These techniques draw on vendor experience in existing soft handover schemes combined with a developed understanding of radio layer and network behaviour when exercised with highly asynchronous symmetric or uplink asymmetric traffic, a significant percentage of which may be real time dependent.

Some, indeed most of this traffic will be compressed which in turn implies a need to fully understand JPEG and MPEG behaviour (particularly JPEG2000 and MPEG Part 4 Part 10 SVC) in wide area radio transmission systems (the need to be JPEG and MPEG friendly)

These skills are largely resident within the traditional telecoms vendor community and will continue to offer a competitive advantage to companies that can constructively combine these multiple disciplines.

Headline data rates are misleading. It's not the number of bits that's important it's the **quality** of bits that counts particularly when highly compressed real time content is being exchanged over a bi-directional wide area wireless radio system.

Erlang (hurrah for Denmark) Nyquist (hurrah for Sweden) and Shannon (hurrah for the USA) between them created a body of knowledge which is still directly applicable to the future optimisation of wide area wireless systems.

Their example is directly relevant today to design teams presently engaged in the wide area wireless 'bit rate race'.

The real bit rate race will be won by the design teams who most successfully integrate the still emerging science of advanced source coding (including scaleable wavelet based perceptually lossless and lossy image and video encoding), channel coding, admission control and end to end latency management.

About RTT Technology Topics

RTT Technology Topics reflect areas of research that we are presently working on.

We aim to introduce new terminology and new ideas to clarify present and future technology and business issues.

Do pass these Technology Topics on to your colleagues, encourage them to join our Push List and respond with comments.

Contact RTT

RTT, the Shosteck Group and The Mobile World are presently working on a number of research and forecasting projects in the cellular, two way radio, satellite and broadcasting industry.

If you would like more information on this work then please contact

geoff@rttonline.com

00 44 208 744 3163