RTT TECHNOLOGY TOPIC
January 2016

5G DSP

November's technology topic, LTE and 5G Public Safety, discussed the trend towards wider bandwidth channels from the present 5 or 10 MHz channels used in 3G and 4G systems to aggregated bandwidths of 100 MHz in LTE Advanced to 1 or 2 GHz channels in 5G supported within 5 GHz pass bands.

Vendor views of how rapidly the industry will move to these wider bandwidth channels and wider pass bands vary but there is some consensus that power efficient low cost digital signal processors capable of processing 100 MHz and 200 MHz channel bandwidths will be available by 2020 increasing to 500 MHz by 2025 and 1 to 2 GHz by 2030.

15 years ahead may seem a long time but for some of us the millennium does not seem long ago. In this month's technology topic we look at how DSP capability has progressed in the past 15 years and review the challenges that DSP engineers will need to grapple with over the next 15 years.

The rate at which digital signal processing and digital signal processors evolve determines user and IOT hardware availability and usability. The issue is not whether digital signal processors can handle channel bandwidths of 1 or 2 GHz but whether they can handle 1 or 2 GHz bandwidth channels power efficiently and at a cost that is at least equal and preferably lower than narrower bandwidth legacy systems.

Processing load may also move away from channel coding towards spatial processing suggesting a need for algorithmic and architectural innovation.

Read on

15 years back – the challenge of 5 MHz channel bandwidths

If we go back fifteen years to the year 2000, the industry was grappling with the challenge of implementing 3G base stations and 3G user devices. Although an increase in headline data rates was important it was not the only design requirement and there was an explicit expectation that voice channel costs would be reduced relative to legacy GSM.

The existence of two competing standards (CDMA 2000 and WCDMA) meant that it was advantageous to have at least some software configurability.

It is generally the case that when a new physical layer is introduced, some of the physical layer processing tasks, for example the Turbo Decoder and FFT, cannot initially be realised efficiently in a standard DSP and need to be implemented in an FPGA/ASIC, typically as a non programmable hardware accelerator.

This applies both for processors used in base stations and processors in user devices. The power constraints and heat rise limits and noise issues in a user device are key performance parameters but the DSP only has to extract the signal of interest of a single user or session whereas the DSP in a base station or access point will be handling tens or hundreds of simultaneous sessions across the whole pass band rather than just one channel.

In a 2001 3G base station, for example, the chip rate processing, spreading and despreading codes, acquisition and path delay estimation were too computationally intensive for a DSP.

Similarly the symbol rate processes including forward error correction and convolutional and turbo decoding were more efficiently processed in flexible semi programmable co processors.

A 64 user 3G base station typically had a chip rate processing overhead of 30 BOPS (billion operations per second) and required a clock speed of 1.1 GHz for symbol rate processing. By comparison, 2G modem processors in 2001 were running at a clock speed of 40 to 50 MHz. The channel bandwidth for WCDMA was (and is) 5MHz in a typical pass band of 35 MHz.

This distance between the relatively light processing load and flexibility of a legacy physical layer, in this example GSM, and the relatively heavy processing load and inflexibility of a new physical layer, in this example 3G WCDMA, had a fundamental impact on the user experience. This translated into slow market adoption.

4G DSP Today

Move on fifteen years and the story has not fundamentally changed. 4G LTE requires a fast Fourier transform and inverse transform but these are well defined algorithms that can be efficiently implemented when sized as powers of two hence the typical FFT sizes of 512, 1024 and 2028.

Clock speeds are not dissimilar to 3G with efficiency gains achieved through optimised combinations of fixed point (fast) and floating point (precise) operations. Floating point is generally used for MIMO equalizers where precision is important (for the matrix inversion). At least some of the efficiency gain has been achieved by combining native floating point support with fixed point architectures.

There are remaining challenges including getting good performance for lower data rate services in wider bandwidth channels, voice over LTE and 5G IOT being two specific and important examples.

The standards response has been to produce 17 categories of LTE user devices with another category, potentially Cat M in Release 13 also known as NB-IoT for low data rate low power machine type communication devices. This is evolved from Release 12 Category 0 to deliver better coverage and lower cost for smart metering and is based on 200 KHz channel bandwidths (the same as GSM) with OFDMA 15 KHz sub carriers on the downlink and either a 3.75 KHz FDMA or 15 KHz SC FDMA uplink.

The channel coding memory and processing overhead of a DSP is dimensioned by the bit throughput per transmission time interval (one millisecond). Category 14 supports the highest maximum downlink data rate at close to 4 Gbps. It is achieved with 8 MIMO layers and 256 QAM with five aggregated 20 MHz carriers (100 MHz channel bandwidth). The highest combined uplink and downlink data rate is supported on Category 8. As with Category 14 this requires 100 MHz of bandwidth.

LTE device bandwidth comparisons

DL Category	Max number of DL SCH transport block bits received within a one millisecond TTI	Max number of bits of a DL SCH transport block received within a one millisecond TTI	Total number of soft channel bits	Max number of supported layers for spatial multiplexing in DL	Support for 256 QAM in downlink
0	1000	1000	25344	1	No
8	2998560	299856	35982720	8	No
14	3916560	391656	47431680	8	Mandatory

The higher device categories have multiple users per channel so the maximum number of user specific user bits per transport block is a fraction of the overall shared channel bit rate. The maximum number of bits of a downlink shared channel transport block is 10% of the maximum number of downlink shared channel transport blocks received within a downlink shared channel.

The job of the channel decoder is to sort out the wanted user bits within the transport block or transport blocks (two blocks) carried within each TTI (transmission time interval). This determines the channel coding clock cycle and memory and buffer overhead of the DSP.

The soft channel bits determine the complexity of the convolutional decoder including decoder memory.

The modulation level determines the amount of noise that can be tolerated in the DSP, the higher the modulation level, the lower the resilience to noise.

The performance of a signal processor can be measured in MIPS (millions of instructions per second) or MOPS (millions of operations per second) but real life performance has to be benchmarked to take into account the efficiency of the instruction set which in turn is determined by the mix of tasks it is expected to perform.

The assumption is that Moore's Law will enable digital signal processing to keep up with standards driven data rate and channel bandwidth expectations. In practice algorithmic complexity is increasing faster than Moore's law, a two fold increase every two years versus a 1.5 times increase in processor performance. The need to support legacy baseband processing requirements and other physical layers including Wi-Fi adds additional processing load.

Gene's Law is the equivalent of Moore's Law but applied to power consumption. As with Moore's law the present assumption is that the power needed per computational MIP is halved every 18 months but the sustainability of that decrease beyond 2025 is presently questionable.

Latency

There is also a trade off between the amount of channel coding, the complexity of the channel coding, latency and silicon area (hardware cost). Latency is a function of the number of iterations per codeword and the clock frequency, more iterations per code word should result in better physical layer performance though with some throughput delay and additional die cost.

The DSP has been a workhorse of the industry for almost twenty years. Each new generation of technology has required some functions to be performed in hardware co processors but over the life time of each standard, programmable DSP's have managed to handle most Layer 1 channel coding and decoding functions and higher layer protocols.

However from standard to standard, Layer 1 and higher layer protocols have become more sensitive to delay and delay variability. This has meant that particular attention has to be paid to bus and memory architectures and bus and memory performance and memory location (on or off chip).

If algorithmic complexity increases at a faster rate than processor capability there can be no absolute guarantee that DSP performance will be able to keep pace with physical layer performance expectations over the expected 15 year 5G life span.

5G A to D as an additional constraint

The same caveat applies to A to D performance in the context of channel bandwidths increasing from the 10 MHz typical of a present system though to 1 GHz or 2 GHz channel in 2030.

The motivation of moving to wider bandwidth channels is partly to support higher user peak data rates and partly to deliver more multiplexing gain (more users on the same channel).

In LTE, the multiplexing gain is achieved by supporting devices on a varying number of resource blocks at varying power levels. The combination of wider channels and more users per channel increases the amount of dynamic range required in the ADC. Because LTE is a wide area network, the difference in received power is significantly higher than a Wi-Fi network. As a result, an ADC capable of handling a 20 MHz LTE channel needs at least 60 dB of dynamic range.

The dynamic range needed in the ADC determines the bit width of the ADC. The bit rate of the digital convertor is a function of the bit width and sampling frequency which must be at least twice the signal bandwidth. The bit rate of the ADC and associated signal processing together determine the amount of power consumed. The bit width has to allow for additional resolution to accommodate RFIC imperfections including direct conversion DC off sets and adjacent channel interference.

ADC options

The majority of ADC's to date are known as Successive Approximation Register (SAR) ADC's. The input analogue voltage is tracked and held and then compared with prior samples using a binary search algorithm. The power dissipation scales with the sample rate.

For LTE the two alternative options are either the sigma delta ADC or pipeline ADC. A sigma delta ADC produces a high resolution and low resolution signal and uses error feedback to compare the two signals. A pipeline ADC as the name implies produces a high resolution description of an analogue signal from a series of lower resolution stages with the first stage working on the most recent sample and the following stages working in analogue remainder voltages left over from previous examples.

All ADC's generate quantization noise and are sensitive to clock jitter. Noise and jitter becomes particularly important when demodulating 16 QAM, 64 QAM or 256 QAM signals. Power efficiency is therefore a composite of conversion efficiency and conversion effectiveness expressed as a signal to noise ratio which in turn determines error vector magnitude which in turn determines throughput efficiency.

The benefits of a well-designed well behaved front end can therefore be compromised by a poorly implemented ADC. Conversely a well-designed ADC can compensate for a poorly implemented (noisy and nonlinear) front end though the additional resolution required may result in unnecessarily high power consumption.

Specifying an ADC to handle the dynamic range required in a 20 MHz channel will however mean that the ADC will have substantial headroom when processing narrower band channels. This might allow for a relaxation of analogue filtering which in turn would reduce component count, component cost and insertion loss. Alternatively dynamically reducing the bit width of the ADC for narrower band channels reduces ADC power drain.

These performance trade-offs and costs scale to higher channel bandwidths and higher modulation options where additional noise can rapidly increase channel error rates.

A pipeline ADC is an open loop architecture with a latency of between 4 and 6 clock cycles. They are normally implemented in CMOS using switched capacitor discrete time circuitry. A relatively complex analogue anti-aliasing filter is needed which consumes power and silicon area. The pipeline DC is generally the most efficient option for bandwidth input signals of 10 MHz to 100 MHz.

Note that the receive path is usually split into two components I and Q which require individual ADC's generally known as IQ ADC's.

In wireless applications up to half of the effective number of bits can be needed to handle unwanted signals illustrating the typical trade-off between front end RF analogue filtering and ADC specification.

Just as a reminder, ADC's perform an amplitude quantisation of an analogue input signal into binary output words of finite length, a nonlinear process.

The non-linearity shows up as wideband noise in the binary output also known as quantisation noise. Quantisation noise can be reduced by over sampling and dithering but both options have an associated cost in terms of power consumption.

Is the A to D a fundamental constraint in 5G?

In the RF domain there is an advantage in moving to higher frequencies because additional gain can be achieved from compact short wavelength antennas. The bandwidth ratio also increases as frequency increases.

A 35 MHz pass band at 900 MHz equates to a bandwidth ratio of 3.8% (35 MHz = 3.8% of 900 MHz as a centre frequency). The same ratio at 9 GHz yields a channel bandwidth of 350 MHz and 3.5 GHz at 90 GHz.

This is not to say that RF filtering at 9 GHz or 90 GHz is easy. Designers have to deal with parasitics and matching and loss and noise but at least the resonance ratios remain similar.

In an A to D we have the fundamental constraint that usable system bandwidth is dependent on the A to D convertor's sample rate and system bandwidth cannot be greater than half the converters sample rate which suggests a sample rate of 4 GHz to digitise a 2 GHz channel at 90 GHz.

High performance 16 bit A to D's are available in 2015 with a sample rate of 200 M/sample/s giving a useable bandwidth of 100 MHz and a signal to noise ratio of 79 dB.

However with all high performance A to D converters a nonlinear charge is produced in the sampling process which is reflected into the input network each time the sampling switches close and there is always a risk that this will be resampled. Avoiding this requires a carefully matched (50 ohm) network.

Most A to D convertors are differential to provide good common mode rejection. Any loss of symmetry for example due to board layout and interconnects will show up as second order harmonic distortion. Differences in ground current on adjacent ground planes will add to this distortion.

The effects of direct sampling on the source of the analogue signal can be minimized by using an amplifier to absorb the charge from the sampling process. If the amplifier is located close to the converter the reflections can be reflected multiple times before the sampling period ends reducing the impact of glitches on the converter's spurious free dynamic range. However this requires amplifiers with a large gain bandwidth product.

It becomes apparent that the specified performance of an ADC on a specification sheet measured in laboratory conditions may not match a real life device and there will generally be some implementation loss.

The A to D and D to A must therefore be considered as one of the critical components that could potentially slow the implementation of wider channel bandwidth 5G radio systems.

The positive side to this is the market opportunity produced by the need to deliver innovative A to D architectures including optimised filtering and amplification.

It also points to an argument to keep 5G modulation options relatively simple in order to make the overall throughput of the radios more tolerant to phase noise and distortion.

Differentiating Digital Signal Processing and Digital Signal Processors

It is useful to consider digital signal processing and digital signal processors as two related but separate topics.

Digital signal processing has been fundamental to every generation of digital cellular technology from GSM onwards and has had arguably the single largest impact on spectral and power efficiency.

Initially spectral and power efficiency gains were achieved by exploiting the redundancy implicit in voice traffic with progressive improvements in speech encoding and decoding. Digital voice encoding combined with channel coding and digital error control techniques delivered additional performance improvements.

As data became progressively dominant in the traffic mix there was a need to implement wider bandwidth channels in order to realize higher per user peak data rates and to deliver multiplexing gain, particularly important with ‘bursty’ data exchanges.

Third and fourth generation cellular systems therefore implemented channel coding and channel equalisation techniques which allowed the introduction of 5 MHz channels in 3G systems and potentially 100 MHz aggregated channels in 4G LTE.

There has always been a lag between standards support for channel functionality and practical device availability. This is because digital signal processors initially struggle to support the required processing tasks and generally consume too much power or have memory constraints which introduce delay and delay variability in task processing. Given that many processing tasks are required to be strictly deterministic, this can be problematic.

The usual solution is to use hardware accelerators but this conflicts with an increasing need for software upgrades for new features both at the network side and in user devices.

Devices also need to switch between physical layer standards including legacy standards which place a premium on programmability.

ASIC hardware is also dependent on standards being stable but unless technology stops evolving this is unlikely to happen in the foreseeable future. Standards also need to achieve global scale for ASICS to be commercially viable.

The commercial success of a new physical layer is therefore generally dependent on having enough clock cycles available in a DSP at a sufficiently low cost and sufficiently low power budget.

This constraint also extends specifically to the A to D conversion process where there is an inherent trade-off between clock cycles and power consumption and channel quality.

The processing load in an A to D conversion can be lowered by reducing the bit width used to digitize the incoming waveform but this reduces dynamic range. Front end filtering and AGC can be used to manage this but adds cost and complexity.

Similarly it is possible to reduce processing load by reducing sampling rates but this introduces quantisation error.

The combination of these practical constraints suggest that the A to D and digital signal processor and digital signal processing are a crucial part of the critical path determining the rate at which a new 5G physical layer can be introduced particularly for applications where cost and power efficiency are dominant considerations. It is not a coincidence that the recruitment press is presently full of adverts for 5G DSP and 5G IOT design engineers.

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 help inform present and future technology, engineering, market and business decisions. The first technology topic (on GPRS design) was produced in August 1998.

17 years on there are over 200 technology topics **archived on the RTT web site.**
Do pass these Technology Topics and related links on to your colleagues, encourage them to join our **Subscriber List** and respond with comments.

Contact RTT
RTT, the **Jane Zweig Group** and **The Mobile World** are presently working on a number of research and forecasting projects in the mobile broadband, 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