

Application vs. Technology Driven Optimization

Nat Seshan

Distinguished Member Technical Staff
Advanced Architecture and Chip Technology Team
Digital Signal Processing Systems Division
nseshan@ti.com

Perspective & Objective

❖ My Perspectives and Biases

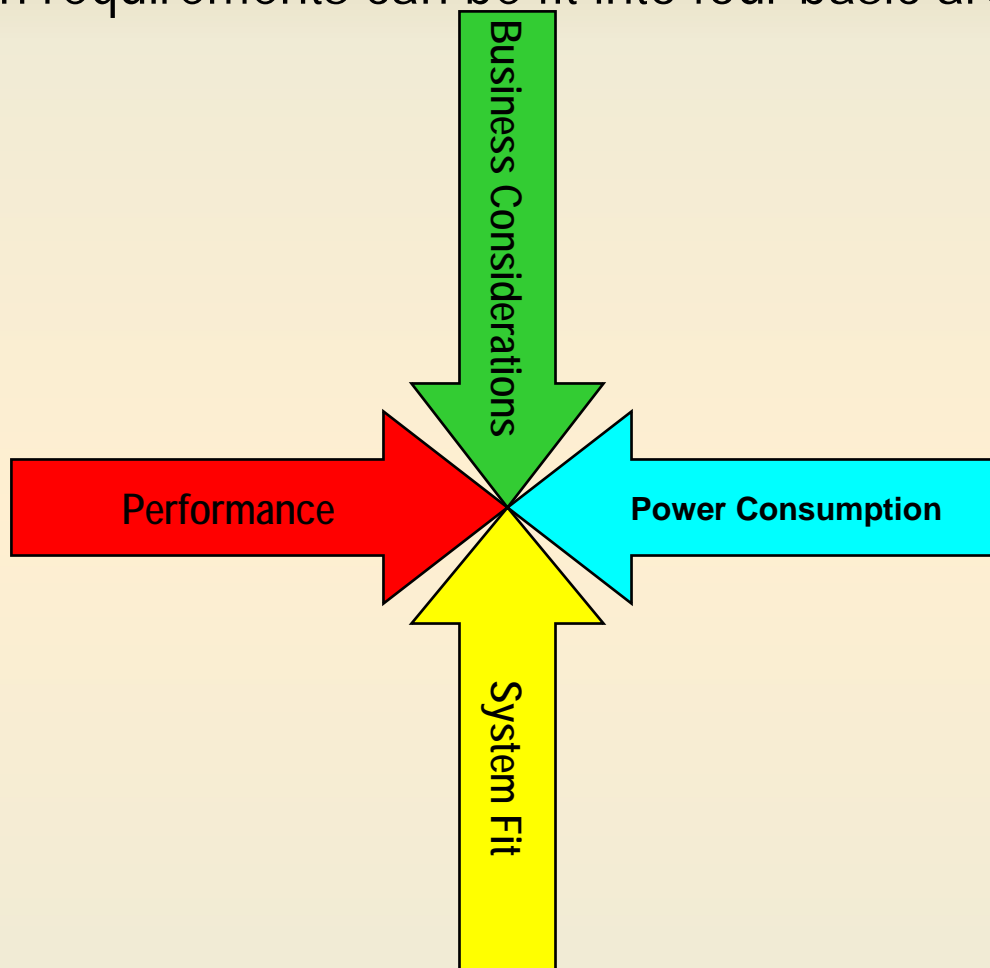
- ❑ My POV from CPU & SOC architecture work at Texas Instruments DSP for the past 18 years.
- ❑ Thus, my natural biases are towards programmable architectures, digital signal processors, flexible architectures that have headroom, breadth, and “longer legs.”
 - Silicon reticles and design teams are expensive and it is hard to get the detail of all uses of a device right the first time
 - A bit contrary to the maniacal SOC-for-one-application-socket focus.
- ❑ So...apply whatever filter or amplifier you need during this talk....

❖ Objectives

- ❑ To give you the full breadth of application requirements and trends in today's SOC's
- ❑ To look at various technology decision points and trends in SOC design
- ❑ To give examples of the diversity of SOC's today in terms of approaches to achieve applications requirements
- ❑ To suggest some problems that need solving

Application Requirements: The Ends

- ❖ The application requirements can be fit into four basic areas



- ❖ The next section discussions the requirements and their current trends
 - Note this is requirements trends ---- NOT implementation trends.

Application Performance Areas & Trends: Computational Diversity

❖ Characteristics of Computation: Diversity

- ❑ Multiple embedded functions in a single device
 - Concurrently
 - Multimode
- ❑ And General Purpose Processing
 - Richer user interface (display & entry)
 - Protocol processing (communications stacks)
- ❑ Throughput, Latency, Quality of Service
 - The multiple embedded functions have differing real-time requirements
 - Response time, deadlines, QOS
 - It's OK to drop sample or frame but don't drop the call or connection
- ❑ Example: Nokia 6680
 - Two integrated cameras: 1.3 megapixel and VGA
 - High-speed connections with 3G and EDGE
 - Two-way video call capability
 - Video sharing capability
 - Direct printing
 - Email access
 - High-resolution, 262,144-color display
 - Music player with stereo audio



Application Performance Areas & Trends: Obscene Performance

- ❖ The next performance drivers
- ❖ Smart Antennas, Software Defined Radio: Can the following functions be moved to software
 - ❑ For run-time adaptable, multi-standard systems
 - ❑ Spreading/de-spreading
 - ❑ Chip and frequency-rate hop
 - ❑ Forward error correction (Turbo Convolutional codes --- product codes?)
- ❖ HD Video Encode
 - ❑ 1080p (sets go on sale this year!)
 - (Don't expect content though)
 - ❑ Consider sports broadcast: capture, mix, overlay, encode
 - ❑ How to compress 3Gb/s with minimal quality loss into a < 2 Mb/s broadband link?

 - ❑ Ref: Sikora, T. "Trends and Perspectives in Image and Video Coding", Proceedings of the IEEE, Vol 93, No. 1, January 2005

 - ❑ *All of the books in the world contain no more information than is broadcast as video in a single large American city in a single year. Not all bits have equal value.*
 - Carl Sagan
US astronomer & popularizer of astronomy (1934 - 1996)

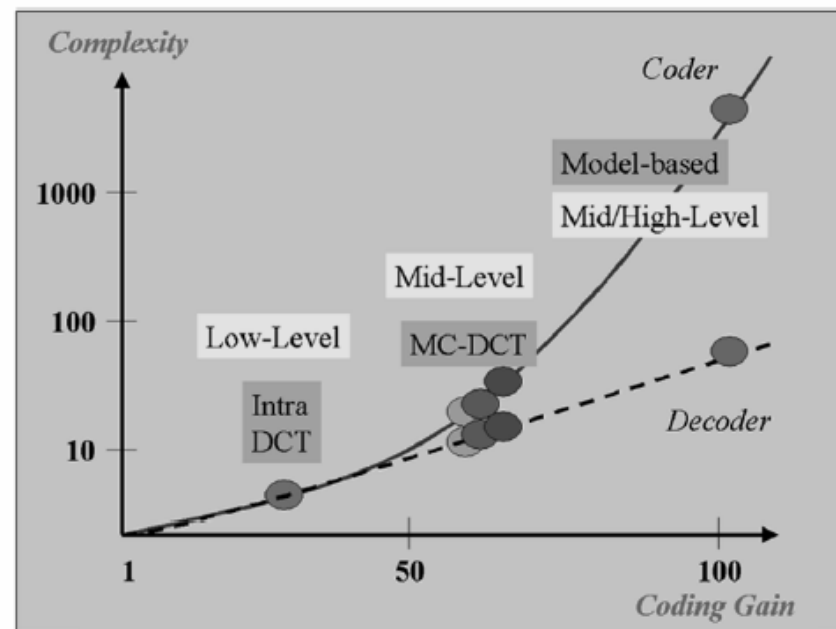
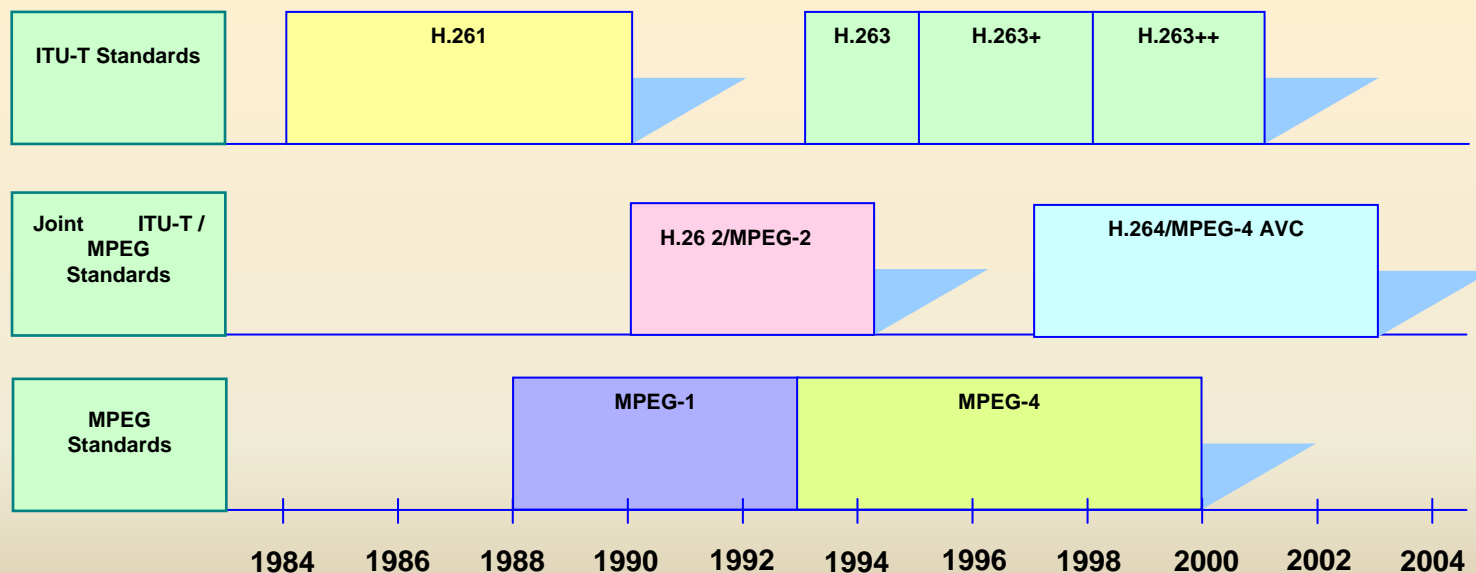


Fig. 1. Complexity of various video coding strategies versus coding gain.

Application Performance Areas & Trends: Evolving Standards

- ❖ First to market --- try to deploy with the release of the “Final Standard”
- ❖ Multiple drops of code during the definition
 - ❑ WMV9 had 4-5 revs
- ❖ H.264: Base, Main, and Extended Profiles
 - ❑ Doh! Suddenly there’s a high profile!!!
 - Added 8x8 DCT and quantization matrix
 - To compete with Windows Media in HD-DVD standard
 - ❑ And will there be a “fast” profile for wireless to reduce performance (and thus power) demands!!! – the application adapting to the technology
- ❖ Quality is not fixed
 - ❑ Consumer video and audio standards are not always bit exact
 - ❑ Tradeoff quality for performance



Power Consumption: 3 Sets of Concerns & Trends

- ❖ Infrastructure: Maximum Power Consumption Budget
 - ❑ Dictated by rack chassis with fixed board space and airflows
 - ❑ Ceiling not to be exceeded, must be evaluated *under worst case conditions*
 - ❑ Some benefit for lower average power (cost of service, electric bills, etc.)
 - ❑ Chassis remaining fixed with scaling performance requirements
 - Some equipments have more forgiving chassis (Advanced TCA)
 - ❑ More infrastructure equipment in hostile environments
 - on a telephone poll in Death Valley in a sealed box waiting to take a 911 call
- ❖ Portable/Battery Powered
 - ❑ Maximizing expected battery life under normal use conditions
 - ❑ Minimizing ENERGY *consumed...thus AVERAGE power consumption*
 - ❑ Battery advancements offset
 - Very compact, dense consumer equipment (tiny but heavy still cameras, cell phones, etc.)
 - More compact spaces (only smaller physical spaces allowed)
 - Increased performance expectations
 - More power consumed from non-SOC function: Richer UI
 - Brighter, larger displays --- portable speakerphones
 - Budget eaten up by mechanical components (hard disk drives, etc.)
- ❖ Non-Infrastructure but Plugged In (Set top box, steaming media boxes, etc.)
 - ❑ *System cost* (fans, heat sinks, power supply rating)
 - ❑ Desired form factor and available airflow
 - ❑ Thermal – don't let device get out of specified operating range
 - $T_{max} < T_{air} + \theta_{JA} * P$

System Fit Trends

- ❖ One slide to share our pain, won't discuss further to the interests of time
- ❖ Off-chip connectivity and level of integration
 - ❑ Increased peripheral integration
 - ❑ All: High speed memory (DDR, DDR2)
 - ❑ Consumer: consumer flash, USB 2.0, analog (video & audio ADC/DACs), storage connections ((Serial/CE)ATA), ethernet PHY, digital video
 - ❑ Infrastructure: networking (ethernet MAC for IP, Utopia for ATM), interprocessor (Serial RapidIO)
 - ❑ PC Chassis/peripheral I/F: PCI moving to PCI-express
 - ❑ Automotive (CAN & LIN)
 - ❑ MBs of SRAM for some infrastructure applications
- ❖ Form factor (physical dimensions, weight) & board manufacturing
 - ❑ Density/form factor driven: infrastructure and compact consumer
 - ❑ Non-space constrained consumer, automotive: driving for 1-2 layer boards and bigger, easier route
- ❖ Quality (DPPM)
 - ❑ More SOCs in automotive: 0 DPPM
 - ❑ Imagine if you had 100 DPPM per IC in a car with 1000 ICs
 - 9.5% of all cars would need warrantee work!!!
- ❖ Regulatory constraints
 - ❑ Trying to meet EMI constraints with high rate processing and many 1 MHz – 10 GHz signals
 - ❑ Lead Free

Business Considerations Trends

❖ Market window

- ❑ Performance, Power, System fit over market lifetime of SOC
- ❑ Time-to-market and market lifetime
 - Ability to hit and sustain desired market window
 - SOC development including software and board level integration
- ❑ Fast Entry
 - Nirvana: Customers would like 9 month concept to system ramp (products for next Christmas/year based on feedback from last years) --- now often SOC teams AND system teams EACH consume more than this
- ❑ Some very LONG market windows
 - Automotive: predict start of production 3 years out due to qualification cycles, lifetimes of 10 years
 - Telecommunications: often 3 years to deploy systems with 10 years of production lifetimes
 - Need factory and field upgradeability for bug fixes and new capabilities

❖ Price & Profit

- ❑ Broader proliferation of SOC's In a variety of consumer equipment
 - shoes, greeting cards, removable media, security tags, credit cards
 - "infotainment" in automobiles, handheld devices, refrigerators, bathroom mirrors
 - Printer cartridges (not just printers)
 - SOCs in sub < \$10 end-equipments
- ❑ More infrastructure with higher processing rates
 - Mbits on wireline and wireless to the consumer
 - Video delivery not just voice
 - Need lower cost of deployment to adopt new technologies
 - Costs of enabling equipment: basestations, set top boxes, etc.
- ❑ Driving lower price points

Technology: The Means

- ❖ Architecture
 - Processing Elements
 - Chip IO
 - On-chip interconnect
- ❖ Circuit Techniques
 - Transistor design (switching threshold, etc.)
 - Dynamic vs. static logic
- ❖ Manufacturing
 - Silicon Process
 - Packaging
- ❖ "Software"
 - Algorithm Realization
 - Code Generation
- ❖ Task Level Scheduling
- ❖ Tooling
 - Automation of software and silicon create
 - Standardized and configurable elements and development tools
 - Modeling
 - Ability for silicon and software engineers utilize tools and components: learning curve

Technology Trends: Complexity, Process, and Trivia

- ❖ Complexity
 - ❑ System and implementation
- ❖ Process
 - ❑ And issues hitting your performance & power goals
- ❖ A couple of other trends related worth mentioning but not diving into
 - Soft Error
 - Small geometry processes are more susceptible to soft errors (alpha-particles) flipping bits in flip-flops, register files, & RAMs
 - For high-reliability or life critical applications how do you add robustness and correction into your computation pipelines without hurting performance
 - Reference: Paper by Karnik & Hazucha: *IEEE Transactions on Secure Computing*, Vol. 1, No. 2. April-June 2004. P.128.
 - Noise – the boogey man
 - More circuits susceptible: (integrated analog, higher frequencies, thinner gate oxides)
 - The price of small dies
 - Harder to dissipate heat
 - Need advancements in package design and thermal architectures for no/low cost
 - So as we talk about power, it's not just watts and joules, you're also fighting temperature
 - SIP vs. SOC: SIP needed for drastically diverse process requirements
 - Integration of high density memories, some RF components (power amplifiers, SAW filters, RF switches, related passive elements) benefit from System on a package (SIP)
 - Krenik et al, *Proceedings IEEE 2004 Custom Integrated Circuits Conference*, P. 63-70



Complexity: Problems

❖ Component Level

- ❑ Potential for sloppiness give advances in silicon tooling
 - Inadequate partitioning to “right-sized” units for complete verification
 - Can create long interconnects where local ones would have been better
- ❑ Difficult to verify implementation meets specification

❖ System Level

- ❑ Complexity of applications and data flows driven by application trends
- ❑ Shear number of individual IP blocks...approaching 100+ in complex chips if you count all individual interconnect components (bridges, etc.)
- ❑ Makes it difficult to determine that you covered all SYSTEM corner-cases

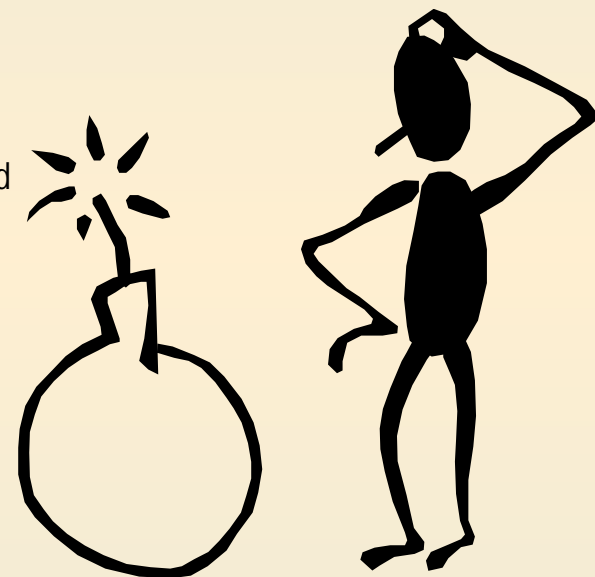
Complexity: Potential Solutions

❖ Component Complexity

- ❑ Formal Verification: Good for a well-designed IP block....
 - Depends on coverage definition
 - Get sparse coverage or infinite compute requirements if you don't verify hierarchically
 - Will electronic system level (ESL) methodologies automate this???
- ❑ FPGA Prototyping
 - Can system level components at board level out of context
 - Test chips useful for high amount of "analog content"
- ❑ System Modeling
 - Can explore dataflow issues and tradeoffs
 - Why system modeling here...to make sure performance is adequately modeled
 - Must be cross-verified with your DV infrastructure
- ❑ Re-use: of blocks & subsystems

❖ SOC Complexity: Do Early, in parallel with architectural development

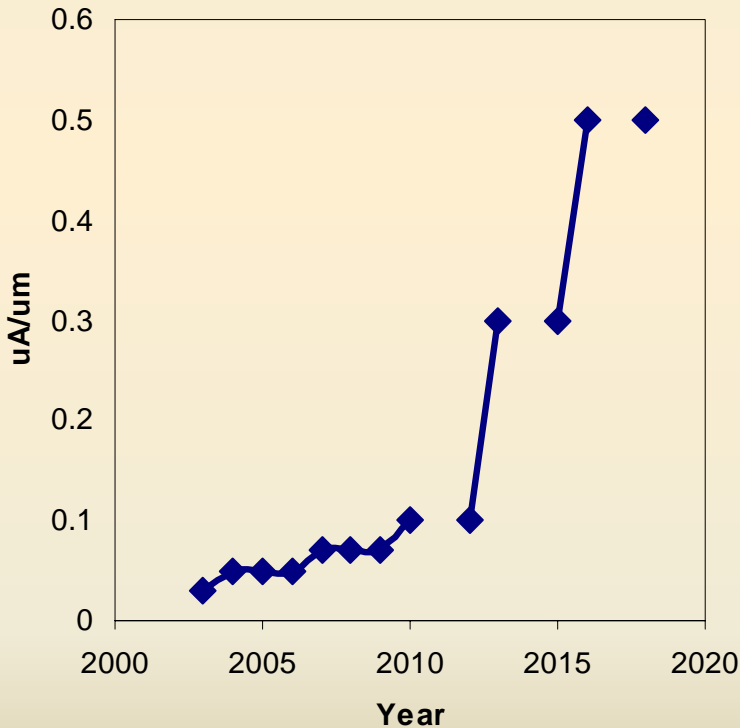
- ❑ CPU/ISA Simulation
 - Good for algorithm bring up, memory system optimization
- ❑ System Modeling perhaps moving to ESL design
 - At IP, Subsystem, and system level
 - To verify desired system operating mode
 - Actively look for breaking points in the system to evaluate "performance headroom"
- ❑ Emulation (Quickturn, etc.)
 - Full-chip system emulation of sufficient cases to validate the "real device" in system context
- ❑ Prototype in an FPGA
 - Small, synthesizable SO's



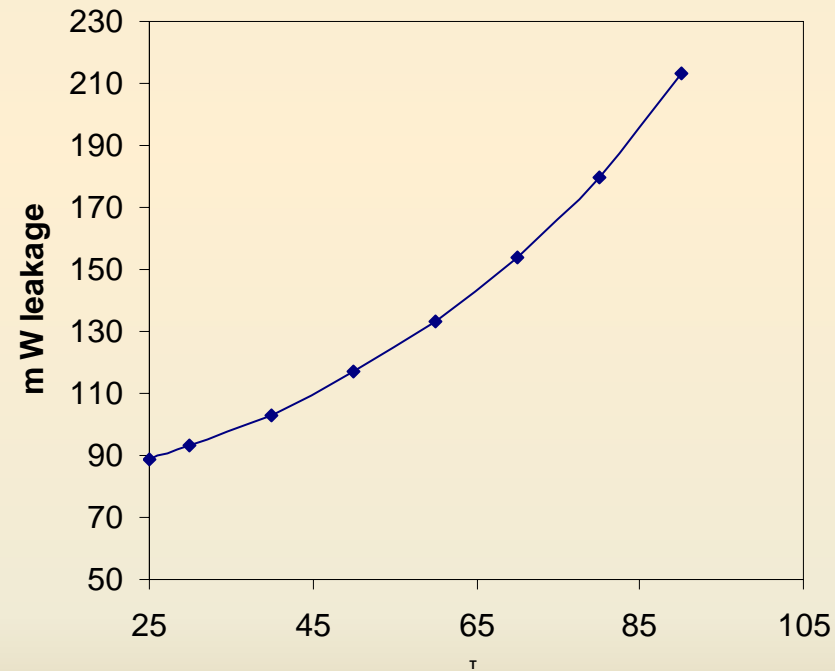
Process & Power

- ❖ Worth repeating the cliché
- ❖ Additional leakage
 - ❑ Ref; ITRS 2004 Update: www.itrs.org
- ❖ Additional leakage sensitivity to temperature (from a power spreadsheet for a 90 nm TI device)
- ❖ Prediction is very difficult, especially about the future.
 - ❑ Niels Bohr
Danish physicist (1885 - 1962)

I_{off} (uA/uM)

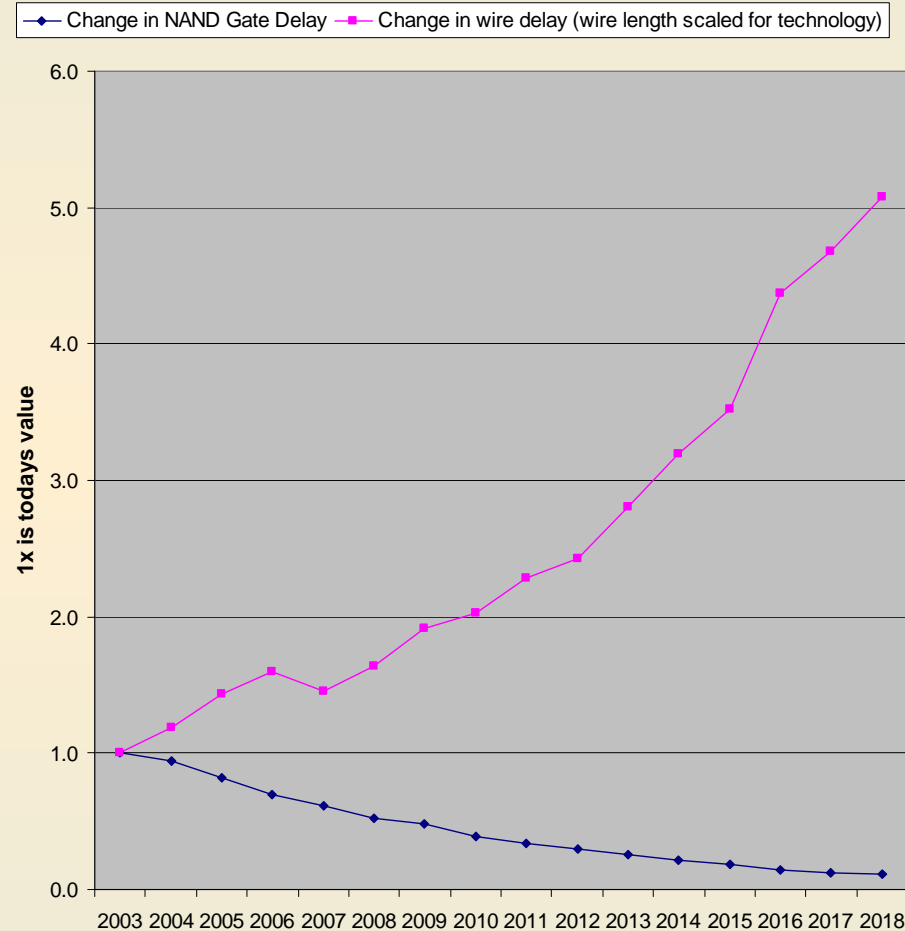


Leakage as $f(T)$



Process & Performance

- ❖ Another cliché
- ❖ Gates are faster
- ❖ Interconnect is slower
- ❖ Slight twist: look at wire delay scaled by the relative geometries of the process
 - Assume we will think more locally
 - Still it's a problem
 - Ref; ITRS 2004 Update: www.itrs.org

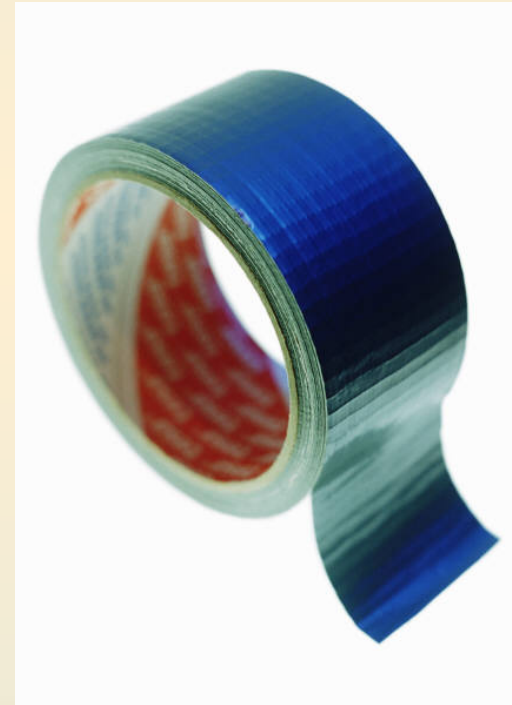


Now hold that thought

Switching back to how things are built today.....

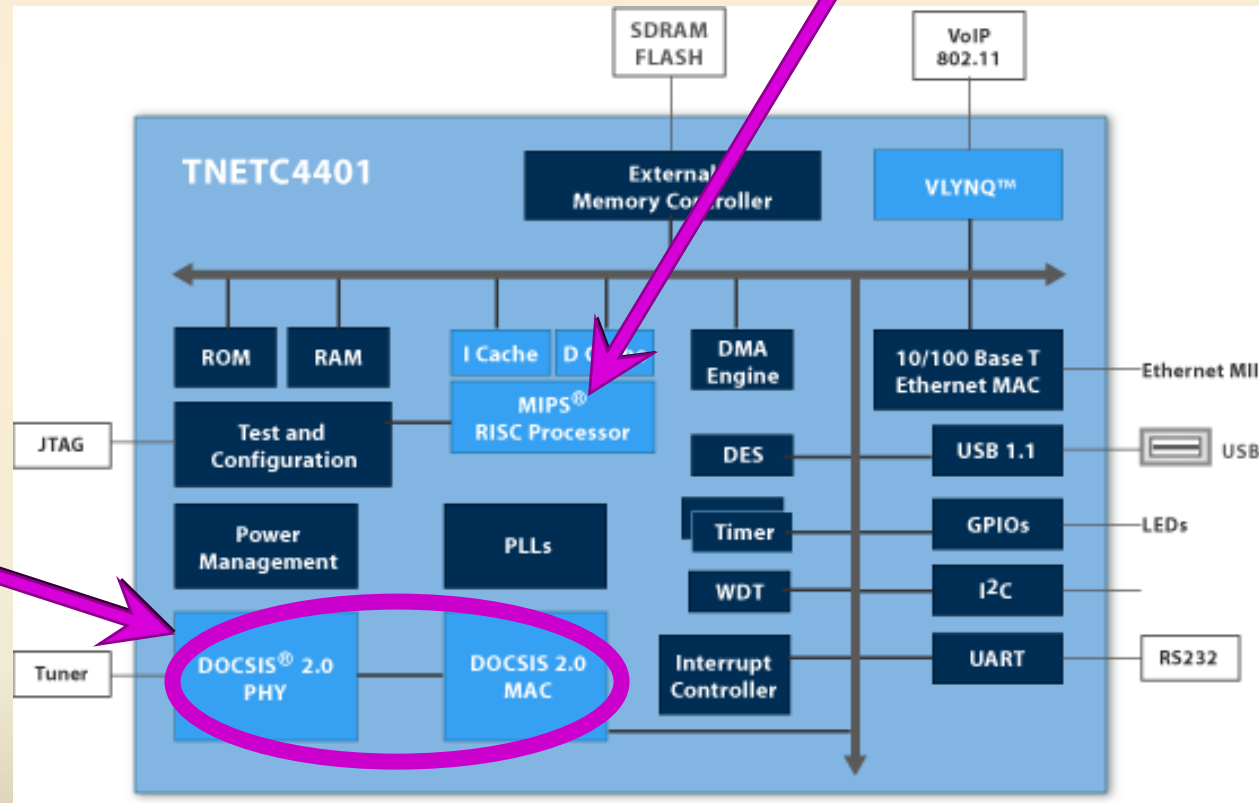
What to good engineers do?

- ❖ In the face of business realities
 - ❑ Availabilities and deployability of technologies in a timely fashion
 - ❑ People & Equipment
 - ❑ Variabilities in target application
- ❖ They find a way with whatever they have
 - ❑ *The ideal engineer is a composite ... He is not a scientist, he is not a mathematician, he is not a sociologist or a writer; but he may use the knowledge and techniques of any or all of these disciplines in solving engineering problems.*
 - N. W. Dougherty, 1955 (Dean, U. Tenn.)
 - ❑ Briefly...remember MacGuyver...think duct tape
- ❖ Power
 - ❑ Process selection, transistor/circuit design
 - ❑ Voltage, frequency, --- static and dynamic scaling
 - ❑ Level of integration
 - ❑ Memory architecture
 - ❑ Clock gating, rail isolation
- ❖ Development Time
 - ❑ Re-use Existing IP (no application tailoring)
 - ❑ High level language development for new code (or avoid changing code what-so-ever)
 - ❑ System Modeling for higher certainty
 - Dataflow, computation, power, area
 - ❑ Automation of design
 - Interconnect generation, IP configuration
 - ❑ Maintenance, upgradeability
- ❖ Leads to diversity of examples to follow



TNETC4401 DOC SIS 2.0 Cable Modem

- ❖ Cable Modem Function (DOCSIS PHY & MAC)
 - ❑ Can be performed by hard coded function
 - ❑ Processing & dataflow regular & fixed
 - Basically high rate/low latency filtering functions
- ❖ Use of low leakage process for power
 - ❑ Get needed performance out of parallelism in the DOCSIS PHY/MAC
- ❖ MIPS processor
 - ❑ Runs protocol stacks
 - ❑ General systems control



Digital Still Camera: DM310

❖ Multi-Modes

- Preview
- Still Image capture & Compress
- Live Video/Audio Capture & Encode
- Video/Audio Decode/Playback
- Still Image Decode/Playback
- Photo Printing

❖ Video Subsystem: Interface and "inline" processing

- Interface: Image Capture I/F: CCD, CMOS
- Inline processing
 - Saves processor/memory data flow & power
 - Including: Pixel conversion, digital zoom (scaling), faulty pixel conversion, optimal-black clamping, digital zoom, noise filtering, gamma correction, color-space conversion.

❖ C54x DSP Core

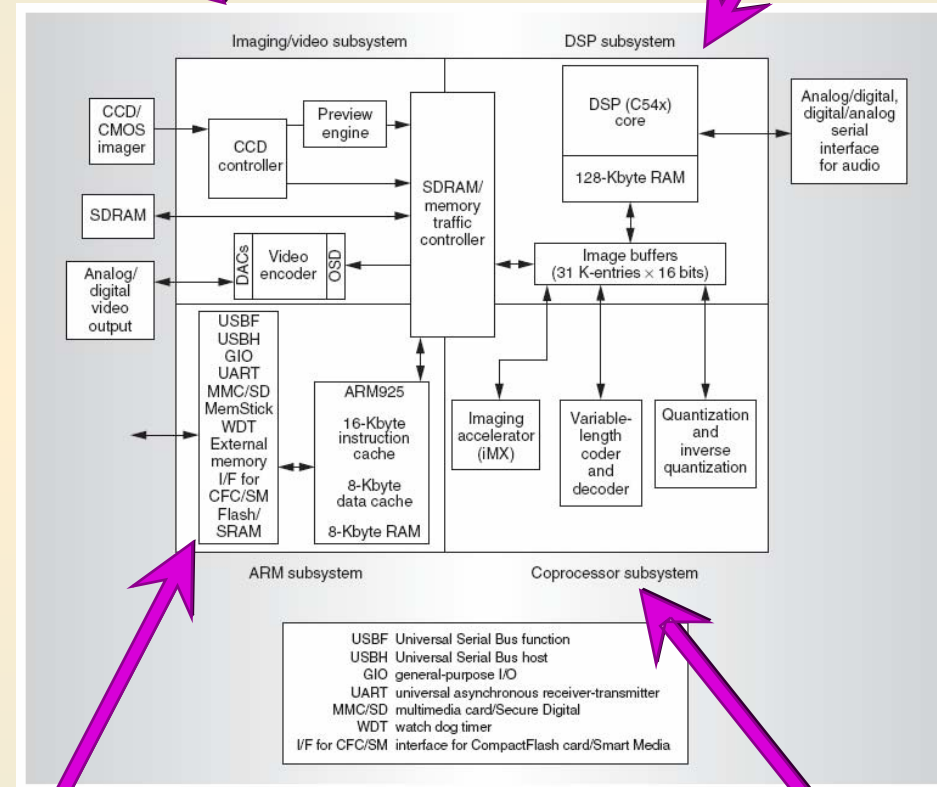
- Works in conjunction with imaging coprocessor to enable multiple standards

❖ Coprocessor Subsystem

- iMX - 8-way SIMD engine
- MAC, SADS, LUT, data re-organization (to organize the data for the iMX)
- QiQ (quantizer, inverse quantizer)

❖ ARM Subsystem

- OS, peripheral management (USB, serial, etc.), file systems (consumer flash)
- Reference: Talla et. al, Hot Chips 15, P 32-39



Single Chip CPE DSL Modem

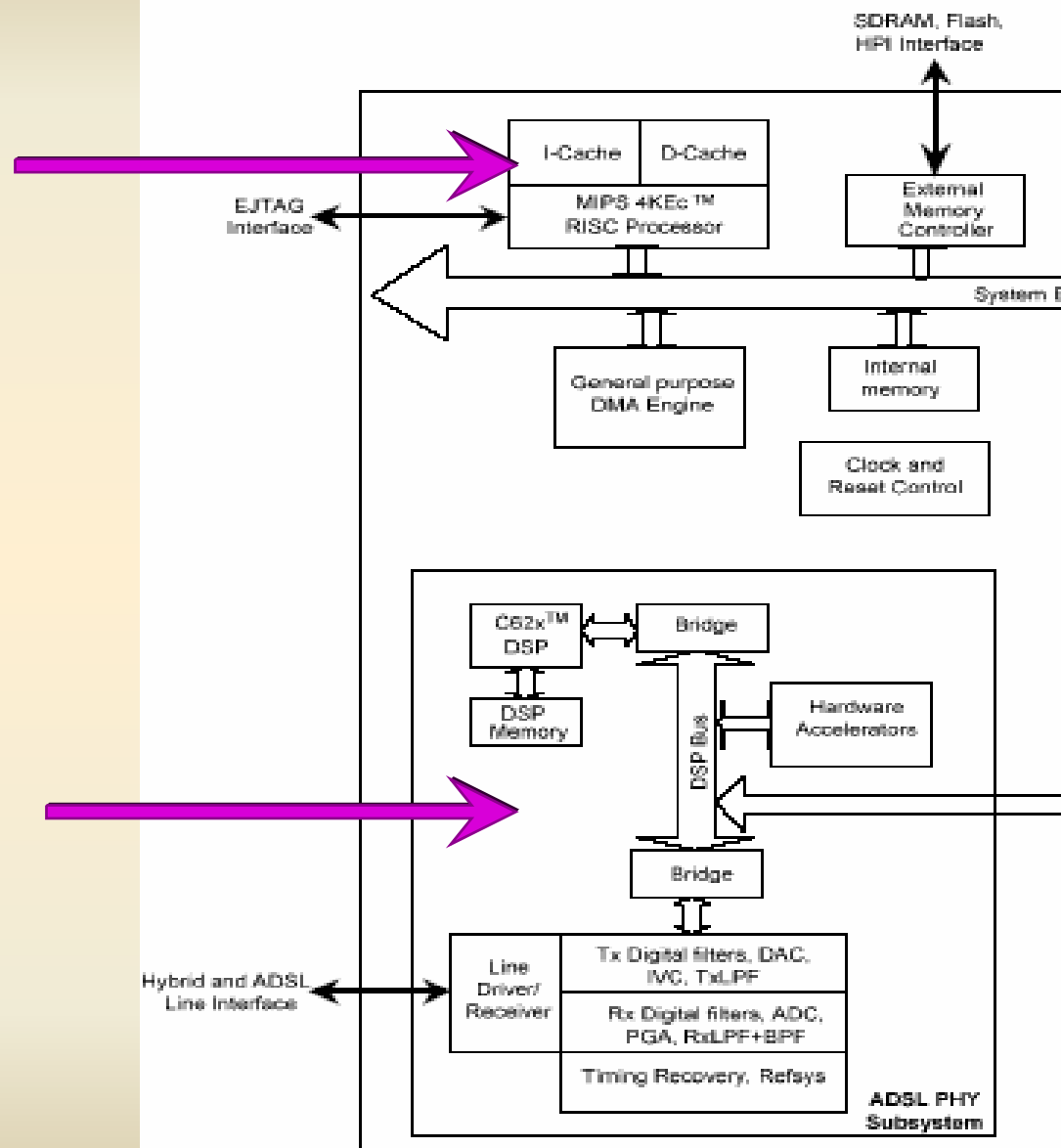
❖ DSL PHY Subsystem

- ❑ Time domain equalization, echo cancellation, FFT/IFFT operations, Trellis, Reed Solomon, decimation, interpolation, and filtering
- ❑ 200 MHz C62x DSP CPU
- ❑ Hardware accelerators
 - Some level of configurability
 - Hooks for dynamic adaptive equalization that enable higher data rates.
- ❑ The programmability of modules enables support of multiple modes like G.lite, G.dmt, and ADSL+ over POTS/ISDN lines, and interoperability with the various CO modems.

❖ MIPS Processor

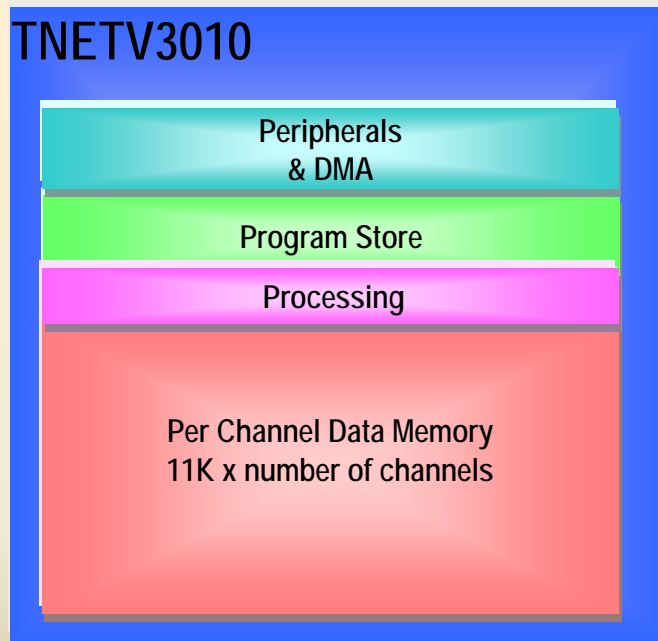
- ❑ Runs stack and manages overall network configuration

❖ Ref: Saha et al, *ISSCC 2004*, Session 18.7.



TNETV3010 SOC for High Density Gateway

- ❖ Application is driven by density
 - ❑ Channels per blade
- ❖ Core & process selection
 - ❑ Chose C55x core due to power optimization and existing C54x codebase
 - ❑ Process choice (130 nm)
 - Low leakage 1.5V 200 MHz
 - High Performance 1.2V 300 MHz
 - ❑ Chose High performance process for power reasons!!!
 - Voltage drop more than compensates for leaker process
 - Power is limited by rack supply so care about worst case power (not average like in portable app)
 - ❑ Can do 32 channels per 300 MHz C55x core
- ❖ Minimize Board Space & Power
 - ❑ Integrate all memory on chip
 - ❑ Memory Size driven by per channel data memory
 - ❑ 11K per G.711 channel
 - ❑ For T1/E1 granularity you want a multiple of either 24 or 32 channels per chip (LCM of 96 channels)
 - ❑ So about 1 Mbyte per 96 channels (or 3 cores)
- ❖ Limits on integration (and thus channels/device)
 - ❑ "Knee" in cost due to die size
 - ❑ Package power dissipation limits
 - ❑ This sets approximate limit on channels/device
 - ❑ All set the sweet spot at 192 channels & 6 cores
 - ❑ Added 384K (channel data plus some local store, stack etc.) per core
- ❖ No GPP --- typically a single one at board level
- ❖ Since all cores process same multiple standards have shared program memory



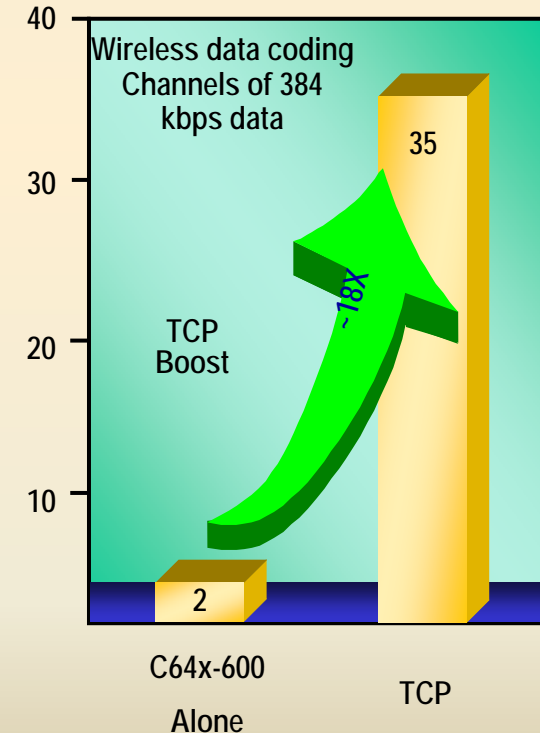
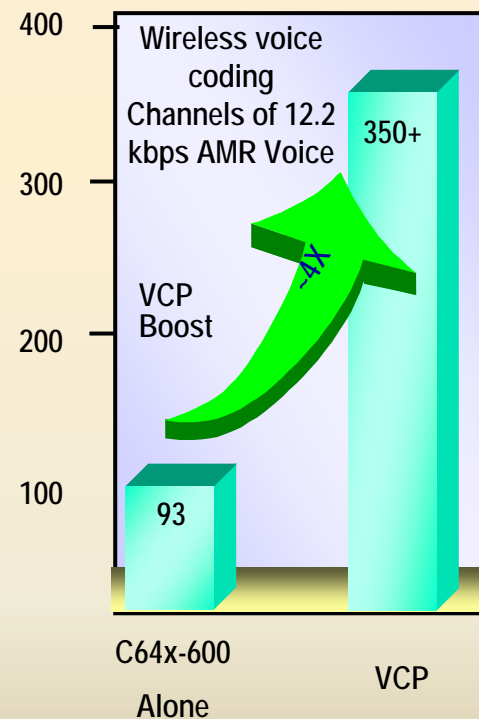
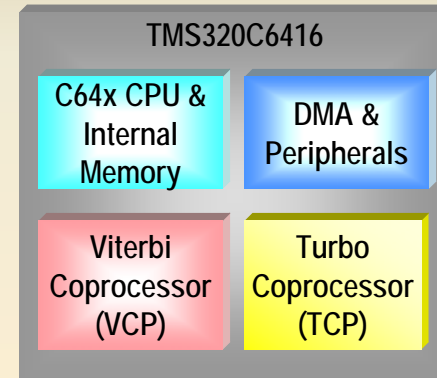
Wireless Infrastructure TMS320C6416

❖ C64x CPU

- ❑ ISA Enhancements for Wireless
- ❑ Quad MAC
- ❑ Galois Multiply
- ❑ But Insufficient performance for forward error correction for 3G standards

❖ Coprocessors

- ❑ Give boost AND free CPU
- ❑ Viterbi coprocessor
 - Gives 4x boost
- ❑ Turbo coprocessor
 - Gives 18x boost



SOC Solutions Summary and Observations

- ❖ GPPs: MIPS and ARM
- ❖ CPUs: C54x, C55x, C62x, C64x
 - ❑ Processor capabilities start to overlap
 - ❑ DSPs already run stacks in some set-top box and video-phone applications
 - ❑ GPPs already perform audio in some MP3 players and cell phones
- ❖ Reuse of some IP across apps (CPUs, standard peripherals, interconnects)
- ❖ Many unique coprocessors & accelerators
 - ❑ People find ways to exceed the performance of a single CPU core
 - ❑ However, the nextgen CPU's may be able to perform those functions
 - ❑ How to transition over

	Application	DSP	GPP	Loosely coupled coprocessors	Process (130 nm)	Inline accelerators
TNETC4401	Cable Modem	None	MIPs	None	Low Leakage	DOCSIS MAC/PHY
DM310	Digital Still Camera	C54x	ARM7	IMX, VLCD, IQ	Low Leakage	Video Processing Subsystem
	Single Chip CPE DSL Modem	C62x	MIPS	Yes	Low Leakage	TX, RX filters
TNETV3010	High Density Voice	6 x C55x	None	None	High Performance	None
TMS320C6416	Wireless Basestation	C64x	None	VCP TCP	High Performance	None
TMS320D610 (not discussed)	Audio Receiver	C67x	None	None	High Performance	None
TMSDRI300	HD Radio	C64x	None	None	High Performance	None
TM320DM642 (not discussed)	Set Top Box Video Phone Surveillance	C64x	None	None	High Performance	None

Challenges

(NP complete problems exercises left to the audience)

- ❖ Quote: Don't try to go where everyone is going, find out where everyone is going and help them get there. - I can't remember where I heard this.
- ❖ Tooling to achieve first pass success and quick time-to-market
 - ❑ In the face of
 - Complexity
 - Increasing process challenges
 - Concurrent IP and SOC development
 - ❑ System Modeling
 - For power, performance, system fit, cost
 - Accurately in the face of
 - ❑ More certain specifications (formal spec., executable spec, ESL), better partitioning from the beginning
- ❖ The ability to create ADDITIONAL compute hardware when you don't have enough....
 - ❑ Don't try to solve the whole problem
 - ❑ Look at emerging algorithms and methods beyond the means of current compute hardware
 - ❑ Look at ways of supplementing current compute hardware to get desired performance
 - WITH MINIMAL IMPACT TO THE REST OF THE SYSTEM
 - ❑ In a way that can be rapidly deployed, coexisting with a variety of other processing elements, with low risk

Thank you & Questions

