Technical challenges to be in the race of the exploding Smartphone & Tablet market

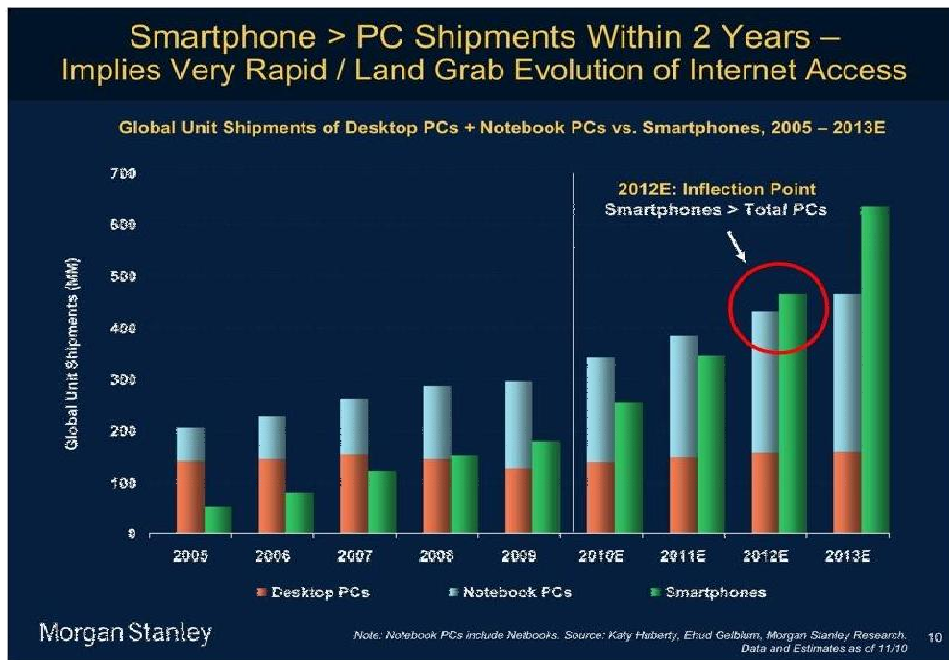Alain Artieri – Senior Fellow

MPSoC 2011, Beaune, July 6

ST ERICSSON

Abstract

We are seeing the most advanced computing concepts developed for mainframes and PCs over the past 6 decades, mind blowing multimedia & graphics capability and incredibly high speed wireless link coming into the small form factor Smartphone running on scarce energy resources.

The keynote speech will discuss the technical challenges and paradigm shift the industry has to face to drive this fantastic revolution.

ST ERICSSON

The Irresistible Rise of Smartphones

- Last November, Morgan Stanley Research predicted :

Smartphone > PC Shipments Within 2 Years –
Implies Very Rapid / Land Grab Evolution of Internet Access

Global Unit Shipments of Desktop PCs + Notebook PCs vs. Smartphones, 2005 – 2013E

2012E: Inflection Point
Smartphones > Total PCs

Smartphones already passed PCs …

Top Five Smartphone Vendors, Shipments, Market Share, and Year-Over-Year Growth, First Quarter 2011 (Preliminary Results) (shipments in millions of units)
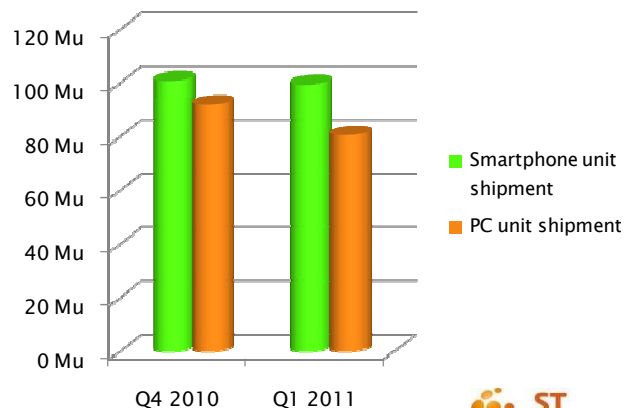
Vendor	1Q11 Shipments	1Q11 Market Share	1Q10 Shipments	1Q10 Market Share	1Q11/1Q10 Change
Nokia	24.2	24.3%	21.5	38.8%	12.6%
Apple	18.7	18.7%	8.7	15.7%	114.4%
Research In Motion	13.9	14.0%	10.6	19.1%	31.1%
Samsung	10.8	10.8%	2.4	4.3%	350.0%
HTC	8.9	8.9%	2.7	4.9%	229.6%
Others	23.2	23.2%	9.5	17.1%	143.7%
Total	99.6	100.0%	55.4	100.0%	79.7%

Source: IDC Worldwide Quarterly Mobile Phone Tracker, May 5, 2011

Top 5 Vendors, Worldwide PC Shipments, First Quarter 2011 (Preliminary) (Units Shipments are in thousands)

Rank	Vendor	1Q11 Shipments	Market Share	1Q10 Shipments	Market Share	1Q11/1Q10 Growth
1	HP	15,191	18.9%	15,624	18.8%	-2.8%
2	Dell	10,284	12.8%	10,469	12.6%	-1.8%
3	Acer Group	9,039	11.2%	10,733	12.9%	-15.8%
4	Lenovo	8,172	10.1%	7,028	8.4%	16.3%
5	Toshiba	4,809	6.0%	4,634	5.6%	3.8%
	Others	33,062	41.0%	34,712	41.7%	-4.8%
	All Vendors	80,557	100.0%	83,200	100.0%	-3.2%

Source: IDC Worldwide Quarterly PC Tracker, April 13, 2011

- ■ Smartphone unit shipment
- ■ PC unit shipment

From Feature Phones to Smartphones

Core Wireless Semiconductor TAM

$B (y-axis, 0 to 40)

Years (x-axis): 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016

Legend: ▪ Grey Market ▪ Connected Devices ▪ Smartphone ▪ Entry ▪ ULC

Source: ST-Ericsson Q211

Note 1: Core Wireless Handset TAM (Total Available Market) is cellular semiconductor content excluding imaging sensors, memory, optoelectronics, discretes, analog & standard logic

Note 2: Connected Devices include wireless game consoles, USB dongles, tablets, embedded modules in laptops/netbooks, M2M

TAM Value ($B)			CAGR
	2011	2016	5 Years
Handset	24	29	4%
Smartphone	12	23	14%
Connected Devices	3	7	22%

Smartphone unit shipment

2010	2015	CAGR
291M	718M	20%

Source: Strategy Analytics, April 2011

Smartphones represent more than 60% Core Wireless Semiconductors TAM in 2016

ST ERICSSON

From Components to Platforms

Component Approach
OEM designs and/or chooses and assembles various SC components to build mobile devices

Platform Approach
OEM uses pre-integrated chipset platform solutions to build mobile devices

2010

Platform 26%
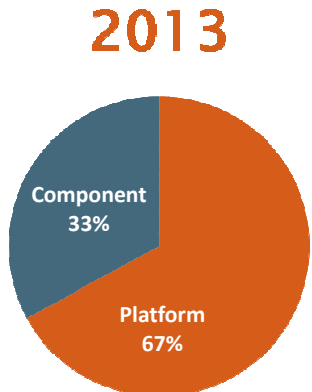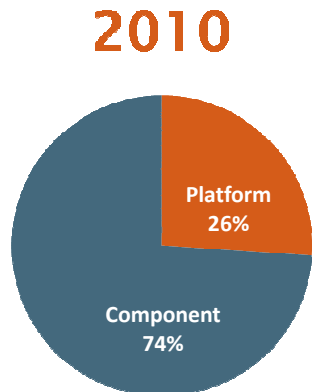
Component 74%

**Components vs. Platforms
(Smartphone volumes)**

2013

Component 33%

Platform 67%

Source: ST-Ericsson internal estimates based on market and analyst reports

**Very few SC companies can sustain platform R&D
… and fewer still know how to optimize wireless platforms**

ST ERICSSON

Convergence

Personal Computer

... Smartphones ...

Consumer Electronics
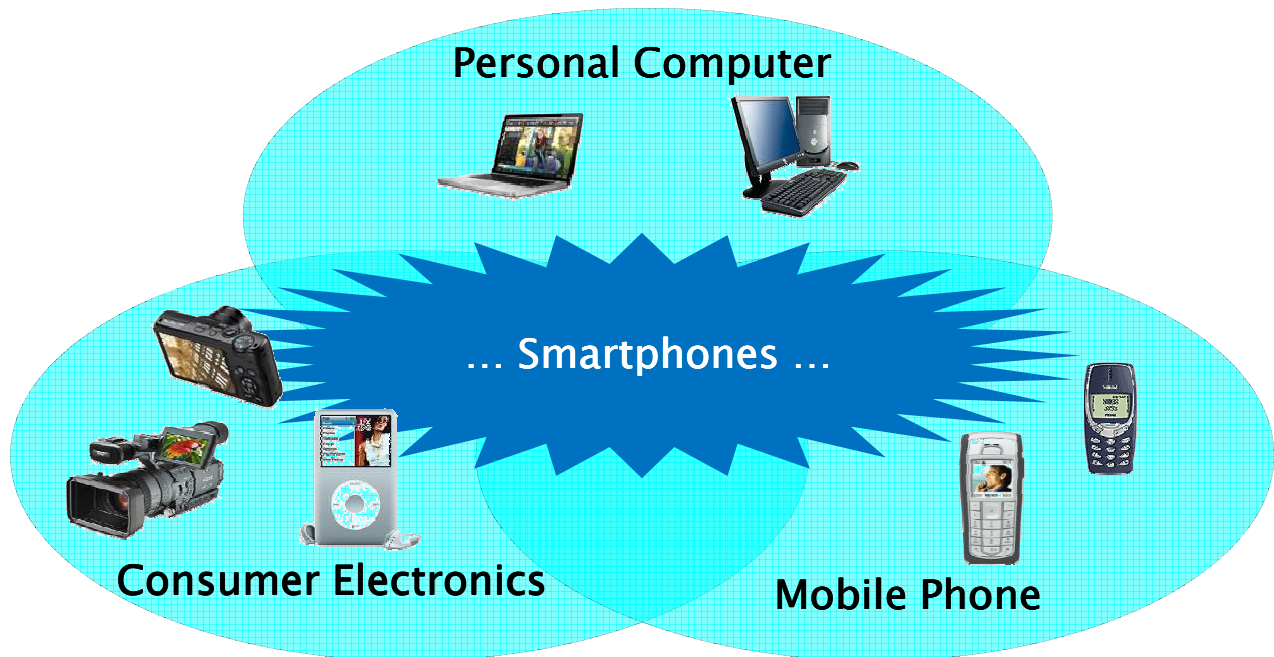
Mobile Phone

Convergence is about ...

- Convergence is not just about the capability to run PC applications and take pictures on a mobile phone ...

- It is actually a whole new way of interacting with the device and with the environment, thanks to mobility and the convergence of high-bandwidth communication, positioning, sensors, and new interfaces

  - "This is the first computer that is context aware, situation aware"
    Jen-Hsun Huang, NVIDIA CEO, GTC 2010 Keynote
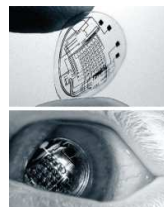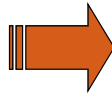

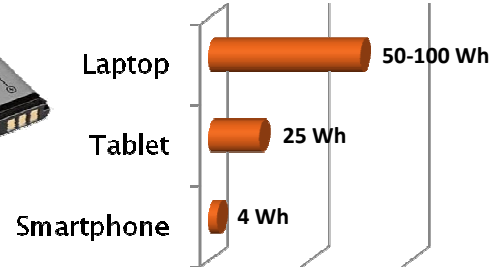Image: Raygun Studio

Convergence Dream ……… and Its Limit

Image: Raygun Studio

- The combination of **Sensor Reach Mobile** always connected to the **Cloud** is going to revolutionize our lives!
  - Efficiently connected to the infinite data storage & processing power of the cloud
  - Mobile and user-friendly Client
- No limits to imagination…
- Well, except for one: *POWER*…
  - Energy is scarce in a Smartphone
  - Low power is The technology driver
  - CMOS options, HW, SW, all is important
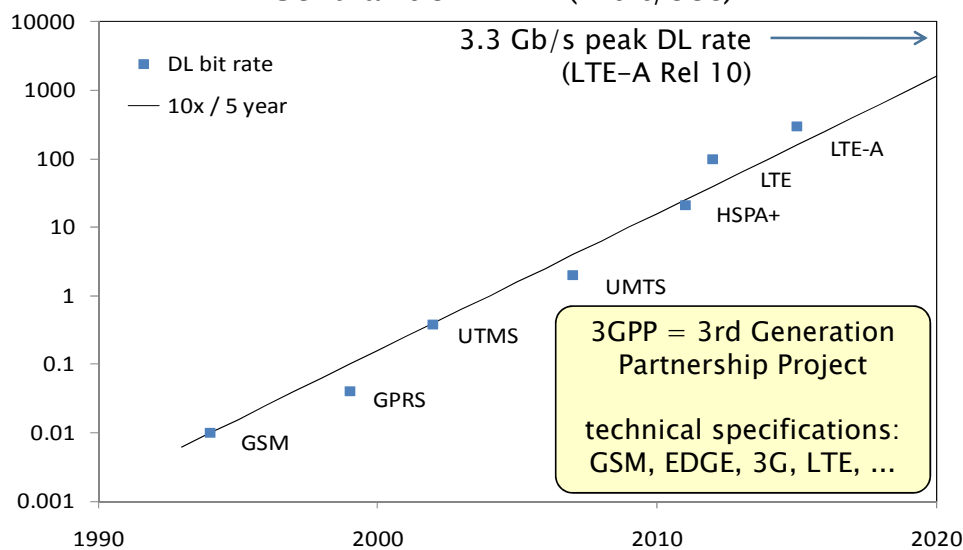
Typical LI-On Battery Energy Capacity

Laptop — 50-100 Wh
Tablet — 25 Wh
Smartphone — 4 Wh

Efficiently Connected …

Cellular downlink (Mbit/sec)

- DL bit rate
- 10x / 5 year

3.3 Gb/s peak DL rate (LTE-A Rel 10)

LTE-A
LTE
HSPA+
UMTS
UTMS
GPRS
GSM

3GPP = 3rd Generation Partnership Project

technical specifications:
GSM, EDGE, 3G, LTE, …

- 100X rate increase during the 2010-2020 decade
- Delivering 100X more GMAC in almost steady power is the challenge

Sensors

- Context Awareness & Multimodality
  - Adapt to the situation

- Detect Context 24/7

- Sensor Fusion

- Inference Algorithms

- And challenge is …

  ultra low power always on computing

Global Positioning
GNSS

Seeing
3D Camera

Listening
Microphones

Biomedical
Body Sensors

Ambient Sensing
Proximity   Ambient light

CMOS Technology for Mobile Computing

- Mobile computing is driving a large part of innovations for new CMOS process

- The need for high performance in a scarce energy budget command new compromise

- "Only comprehensive teamwork between Process Engineers, Circuit Designers, Chip Architects, and Systems Designers can make major advances." (ISSCC 2007 Keynote)

CMOS Technology Alliances
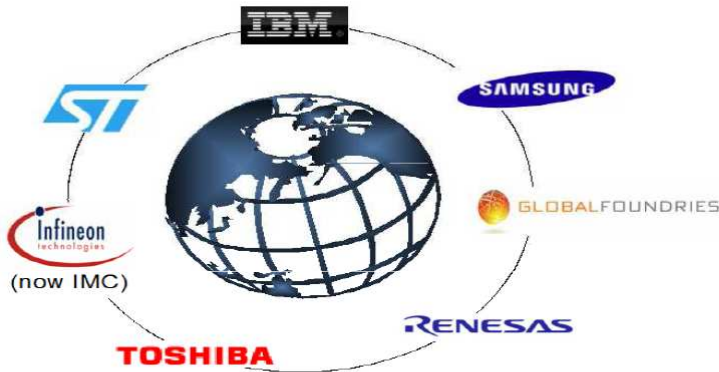
ST participates in these alliances:

- PreT0: advanced CMOS technology research
- ISDA: bulk/low-power (LP) technology development
  - 32/28nm
  - 20nm

Transistor Architecture Trends

32/28nm

28nm Low Power
(ST/IBM IEDM 2009)

Hybrid UTBB /Bulk
(ST/LETI IEDM 2009)

- Fully Depleted SOI
  with Hybrid Bulk
- Gate First
  Metal Gate
  High-K

22/20nm

25nm FDSOI
(ST/IBM VLSI 2010)

- Bulk w/ enhanced
  stressors
- 2nd Generation
  MGHK (Gate Last)
  - Improved
    junctions
- Fully Depleted SOI
  with Hybrid Bulk

16/14nm

15nm FDSOI
(ST/LETI VLSI 2010)

- Fully Depleted SOI
  with Ultra Thin BOX
  and Stressors

11/10nm

8/7nm

FinFET
(ST/IMEC VLSI 2006)

- FinFETs

To continue Moore's Law, 3 main disruptions are expected at 14nm:

Lithography: moving from 193nm immersion to EUV

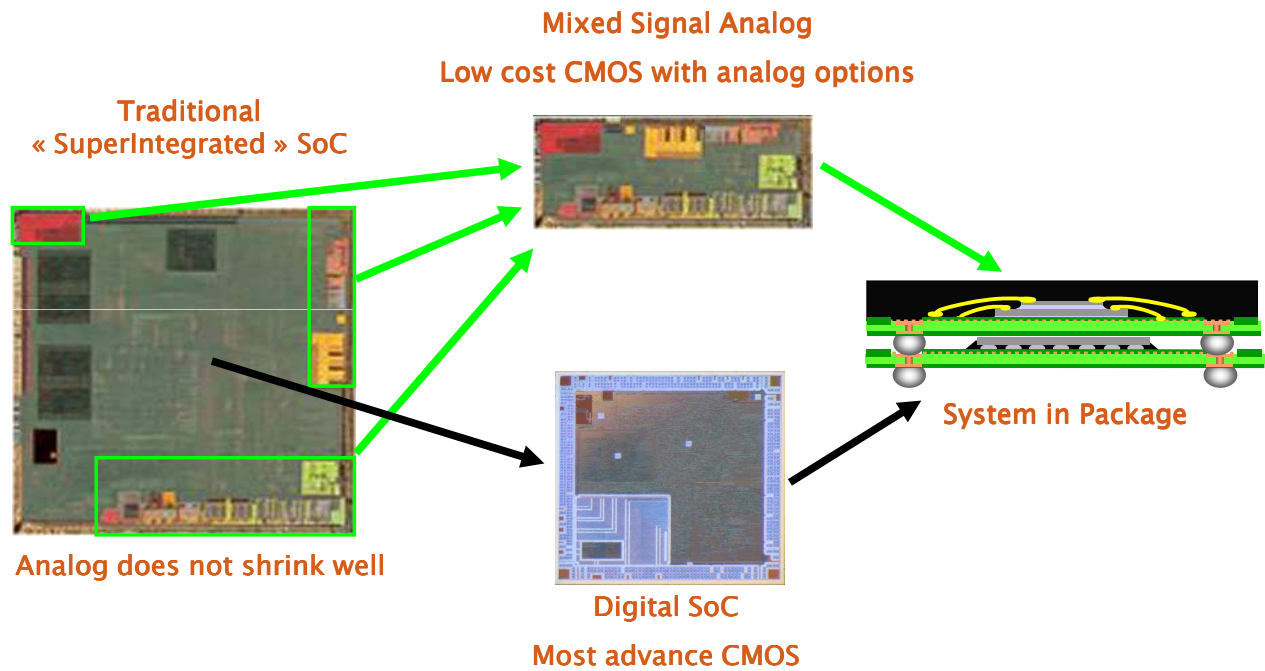Device: moving from bulk CMOS to Thin Silicon devices

3D integration

3D Integration

Mixed Signal Analog

Low cost CMOS with analog options

Traditional
« SuperIntegrated » SoC

Analog does not shrink well

System in Package

Digital SoC

Most advance CMOS

ST ERICSSON
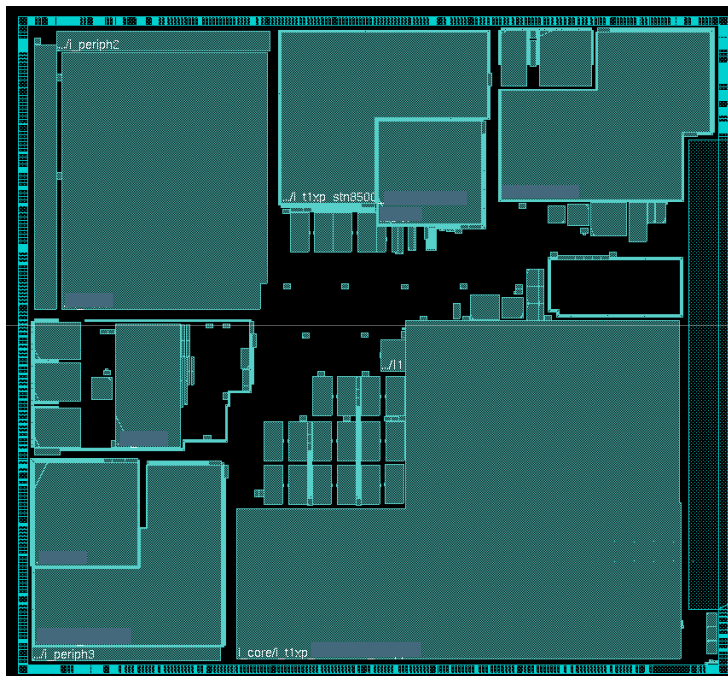

DB8500 Power Management

- Power management is dimensioning the IC design complexity
- Multiple voltage domains and power islands
  - 4 voltage domains
  - 17 switched power islands
- DVFS and AVS
- Forward and Reverse Body Biasing
- Dual Rails SRAM
- Thermal sensor
- Process Monitoring Box
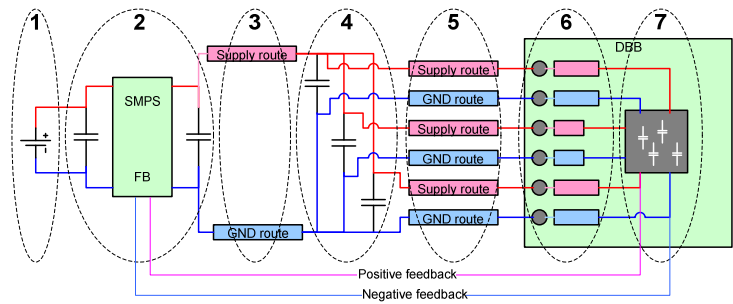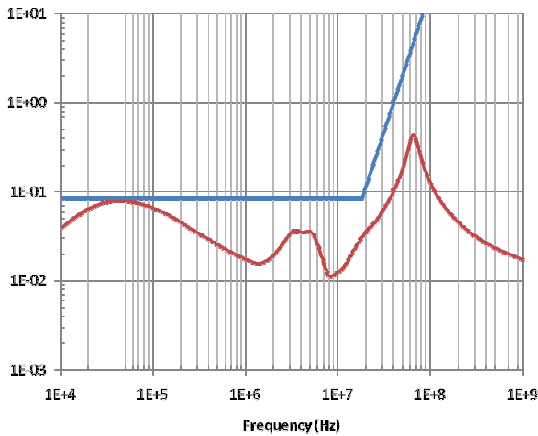- On chip 'decap' for power integrity
- Small Always On domain

ST ERICSSON

Power Integrity

- Power integrity analysis at platform level
  - SMPS Regulator
  - Board
  - Package

Target impedance
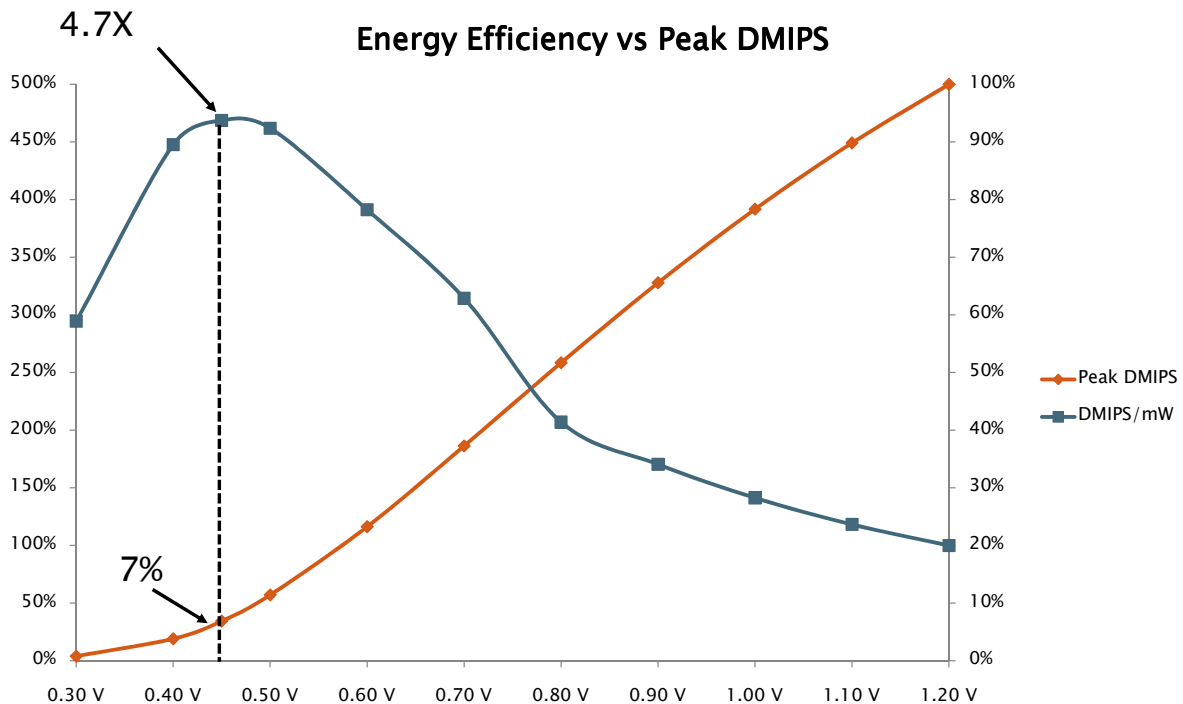

- More passive components embedded into die & package to achieve high clock speed for CPU

- Every mV counts

Energy Efficiency Sweet Spot
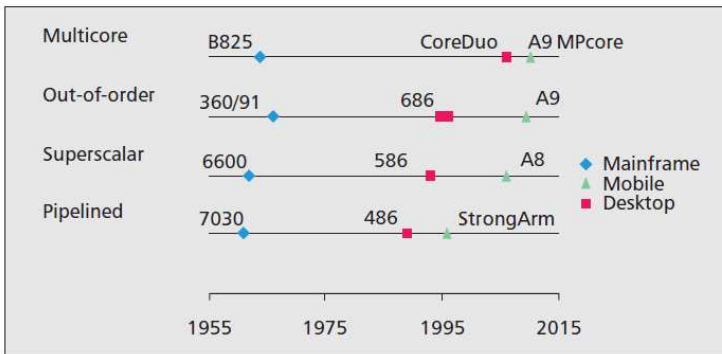

Energy Efficiency vs Peak DMIPS

4.7X

7%

Peak DMIPS
DMIPS/mW

Mobile Application CPU catch up :

Source : Steven Swanson and Michael Bedford Taylor, University of California
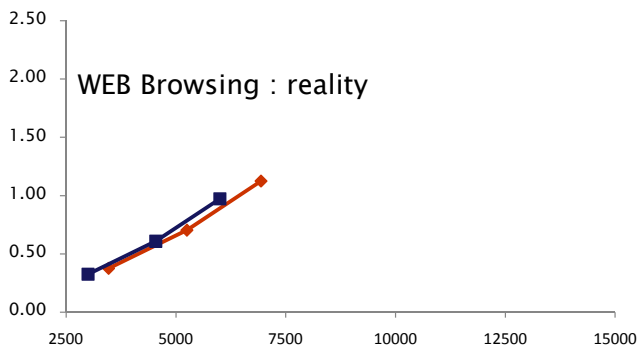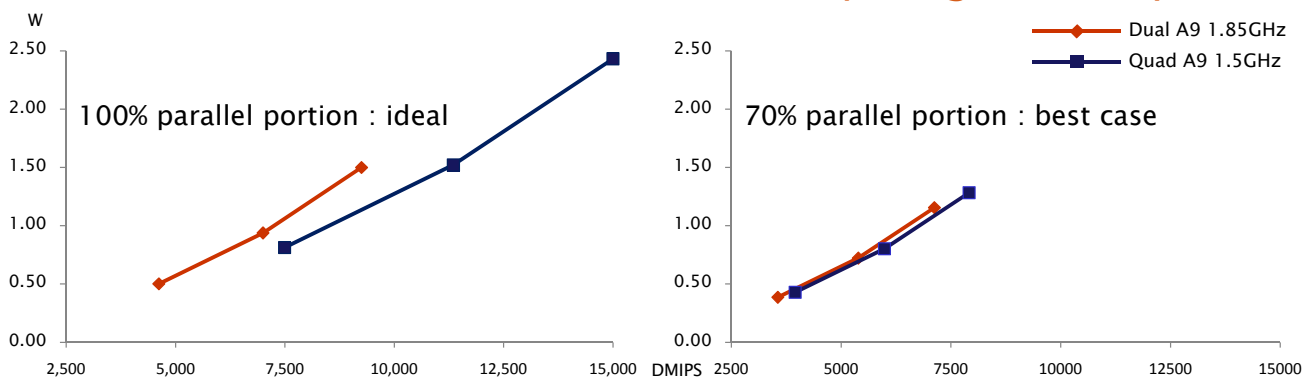
- Clock speed gap narrowing too :
  - Cortex A9 > 1.5GHz now common
  - Cortex A15 in 2GHz–3GHz range are coming

- ARM CPU offers now same features as desktop CPU

Feature	Benefit
Pipelined	Higher clock rate
Superscalar	Parallelization through multiple instruction dispatch
SMP Multi Core	Parallelization through multiple thread dispatch
Media Extension	Parallelization through vectorization
Out-of-order	Reduce idle time
Prefetch	Reduce idle time

ST ERICSSON

Amdahl Law Wall : Quad vs Dual Computing Efficiency

100% parallel portion : ideal

70% parallel portion : best case

WEB Browsing : reality

Source: ST-Ericsson Benchmark

- Ideal :
  - Quad core delivers 60% more DMIPS
  - Quad Core is 25% more power efficient
- Best Case :
  - Quad Core delivers 10% more DMIPS
  - Quad Core is marginally more power efficient
- Reality
  - Dual Core delivers 15% more DMIPS
  - Dual Core has better power efficiency

ST ERICSSON

SMP SW Exploitation

- System-level
  - High Level OS are exposing significant levels of concurrency, easily exploited by SMP :
    - Networking, multimedia frameworks, drivers, etc.
  - Multi-tasking increasingly used in Mobile
    - Multiple widgets, multiple concurrent applications
- Multimedia
  - SMP-friendly by nature: data partitioning, function partitioning and pipelining easily achievable
- Web Browsing
  - Today's browser cores are not SMP-friendly, but some important browser evolutions are very SMP friendly ,e.g.
    - Multi-tab, Multi-task, Out-of-Process Plugins
    - HTML5 Web Workers
- In addition, particularly critical applications may benefit from explicit SMP parallelization
  - Tough, but well known techniques
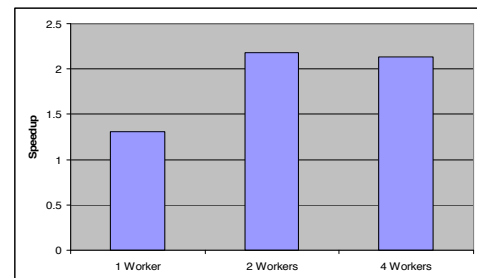 - Low-level SW and multimedia already optimized for SMP
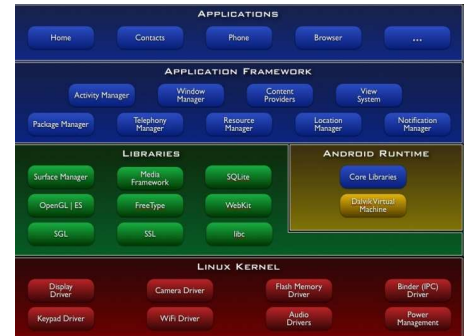
WEB Browsing SMP-Friendly Evolution

- HTML5
  - Multimedia: native support for Canvas and Video make SMP-optimized multimedia integration straightforward
  - WebWorkers: ability to spawn independent JavaScript threads
    - To avoid freezing the browser UI while processing, but also ideal for SMP

- Multi-processing
  - Multi-tab Multi-task – initially on Chrome, Mozilla follows, now WebKit2 (see later)
  - Mozilla Out-of-Process plugins

- Parallelization
  - Initial efforts towards SMP optimization started e.g. on WebKit, Mozilla, Google...
    - https://lists.webkit.org/pipermail/webkit-dev/2010-February/011637.html
  - Advanced research programs ongoing, e.g. Berkley parallel browser project and others (http://parlab.eecs.berkeley.edu/research/80)

      "Our goal is to make every website developer a parallel programmer"
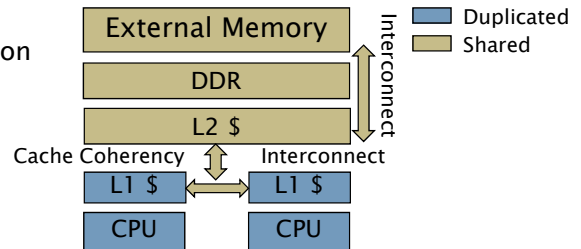
SMP: Scalability Limiting Factors

- Ideally N CPU's -> N times speedup and lower power than 1 CPU @ N*Freq

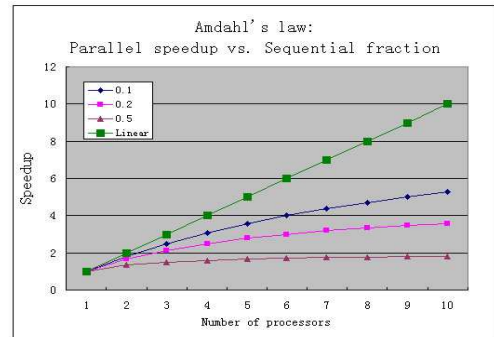- Scaling is limited by shared HW and SW serialization points

- HW
  - Cache Coherency
  - Resource Sharing
  - ⇨ Optimized CPU and System HW implementation

- SW Serialization Points
  - Application and/or system SW serialization points have strong impact on achievable scaling
    - Amdahl law
  - ⇨ SW

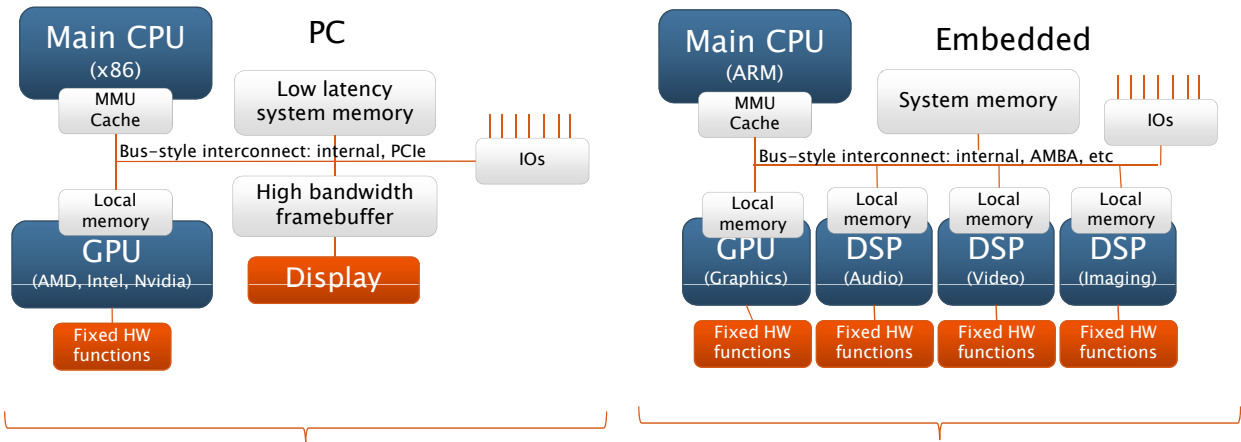- Extensive system characterization, benchmarking and optimizations required to define the right architecture

| | Duplicated |
| | Shared |

External Memory

DDR

L2 $

Cache Coherency | Interconnect

L1 $ | L1 $

CPU | CPU

Interconnect

Amdahl's law:
Parallel speedup vs. Sequential fraction

- 0.1
- 0.2
- 0.5
- Linear

Speedup

Number of processors

ST ERICSSON

From CPU to APU

- Moore's Law enabler for CPU performance somehow limited

CMOS Process	Expect a modest 20% per node
SMP Multi Core	Nice scalability but Amdahl Law Wall
CPU Architecture	Catch up done, not much breakthrough to expect

- Accelerated Processing Unit
  - Expose to programmers the full platform processing power
    - SMP CPU
    - GP-GPU
    - Fixed HW functions
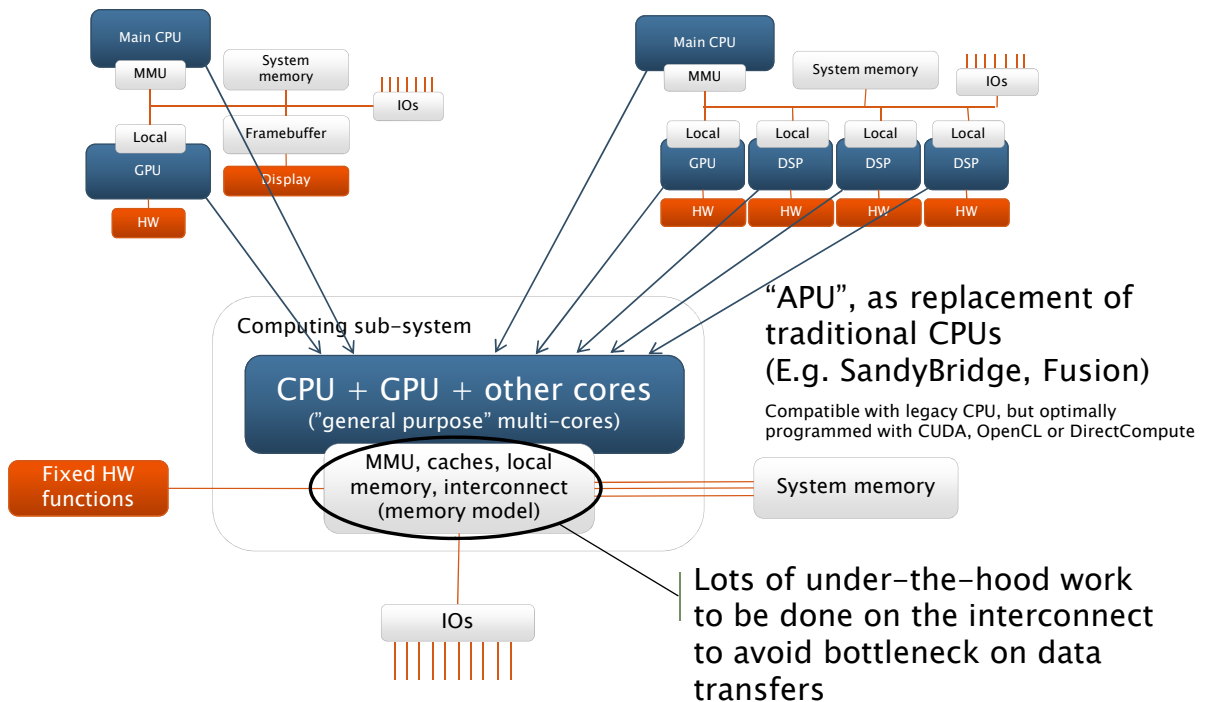 - Need framework for Heterogeneous APU programming
    - OpenCL

ST ERICSSON

Legacy Architectures

PC

Main CPU (x86)	Low latency system memory
MMU Cache	

Bus-style interconnect: internal, PCIe — IOs

Local memory	High bandwidth framebuffer
GPU (AMD, Intel, Nvidia)	**Display**

Fixed HW functions

Computing driven by general purpose programming + 3D gaming

Embedded

Main CPU (ARM)
MMU Cache

System memory — IOs

Bus-style interconnect: internal, AMBA, etc

Local memory	Local memory	Local memory	Local memory
GPU (Graphics)	**DSP** (Audio)	**DSP** (Video)	**DSP** (Imaging)
Fixed HW functions	Fixed HW functions	Fixed HW functions	Fixed HW functions

Computing driven by multimedia, and low-power / limited-resources

ST ERICSSON

Convergence Architecture

Main CPU
MMU

System memory — IOs

Local
GPU
HW
Framebuffer
Display

Main CPU
MMU

System memory — IOs

Local	Local	Local	Local
GPU	DSP	DSP	DSP
HW	HW	HW	HW

Computing sub-system

CPU + GPU + other cores
("general purpose" multi-cores)

Fixed HW functions

MMU, caches, local memory, interconnect (memory model)

System memory

IOs

"APU", as replacement of traditional CPUs (E.g. SandyBridge, Fusion)

Compatible with legacy CPU, but optimally programmed with CUDA, OpenCL or DirectCompute

Lots of under-the-hood work to be done on the interconnect to avoid bottleneck on data transfers

ST ERICSSON

ST-Ericsson/PGI Cooperation on GPGPU

- Portland Group Inc (http://www.pgroup.com)
  - Well known in HPC
  - Part of STMicroelectronics since 2000
- PGI GPGPU
  - CUDA-x86, OpenCL-ARM, Accelerator
- PGI Accelerator
  (http://www.pgroup.com/resources/accel.htm)
  - Pragma based set of directives (OpenMP-like) for C and Fortran
  - Automatic analysis of whole program structures and data, splits portions of the application into CPU and GPU as specified by user directives
  - Optimized mapping of loops to automatically use the parallel cores, hardware threading capabilities and SIMD vector capabilities of modern GPUs
  - Directives for fine-grained control over the mapping of loops, allocation of memory, and optimization for the GPU memory hierarchy
  - Automatic generation of CPU/GPU data movement
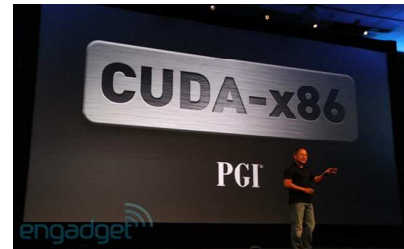
Jen-Hsun Huang Sept 2010 Keynote

Programming Challenge

- Re-writing Existing applications in CUDA or OpenCL can be a lot of work

  1. Split code between Host and GPU

  2. Manage data allocation & movement between Host and GPU memories

  Mechanical

  ----

  3. Tune GPU Kernel Schedules and Memory Usage

  Creative
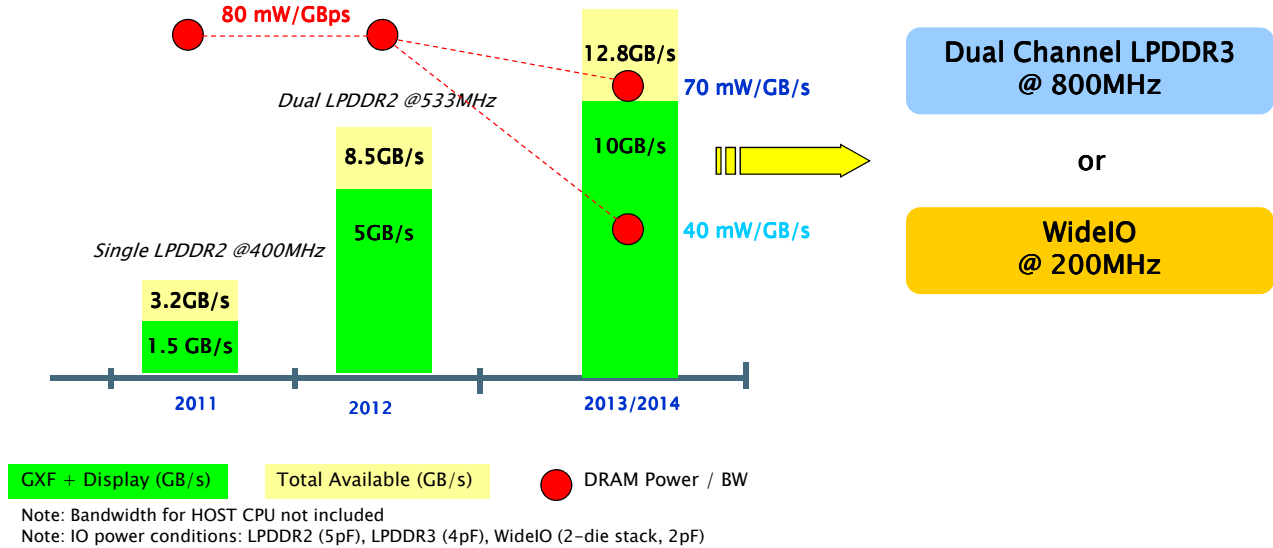
More HW Features to ease SW Programming

- System MMU
  - Virtualization of the address space for multimedia clients, no more constraining physical space reservation

- Hardware Virtualization
  - Multiple OSs and services efficiently and safely sharing reources without the burden of "para virtualization"

- Hardware Coherency
  - No more explicit cache management


Video & Display Resolution Driving Memory BW
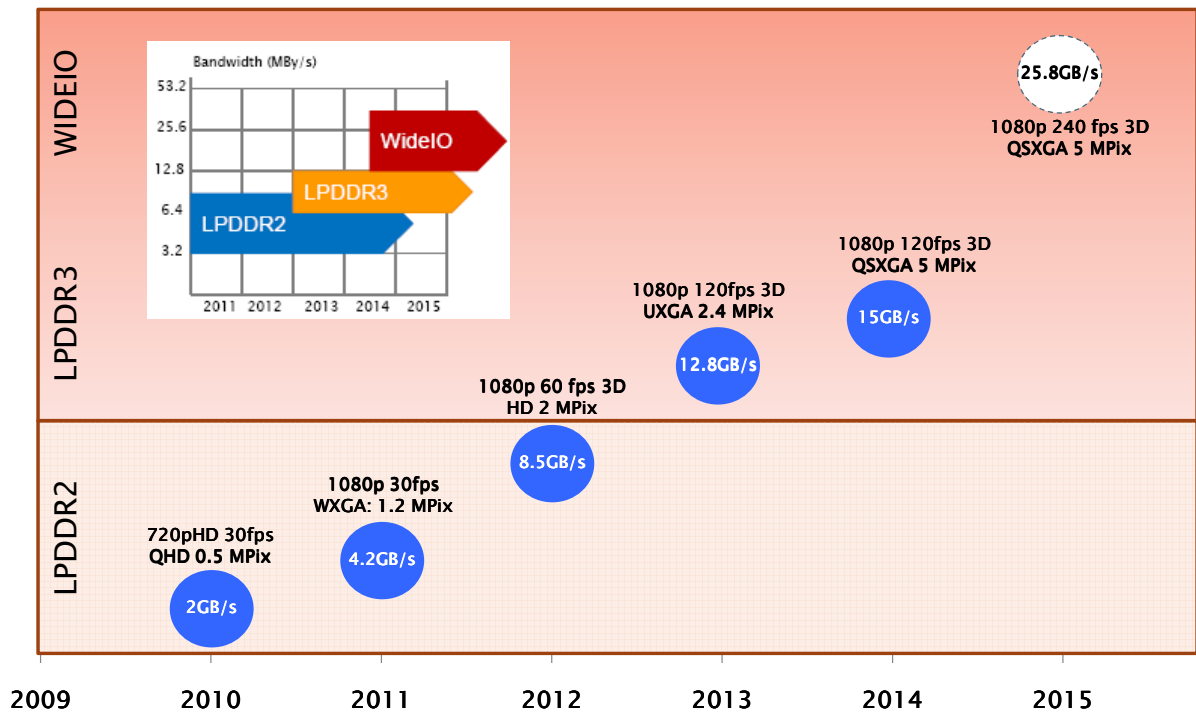
After LPDDR2: LPDDR3 and WideIO…

80 mW/GBps

Dual LPDDR2 @533MHz

12.8GB/s
70 mW/GB/s
10GB/s
40 mW/GB/s

8.5GB/s
5GB/s

Single LPDDR2 @400MHz

3.2GB/s
1.5 GB/s

2011   2012   2013/2014

Dual Channel LPDDR3 @ 800MHz

or

WideIO @ 200MHz

GXF + Display (GB/s)   Total Available (GB/s)   ● DRAM Power / BW

Note: Bandwidth for HOST CPU not included
Note: IO power conditions: LPDDR2 (5pF), LPDDR3 (4pF), WideIO (2-die stack, 2pF)

ST ERICSSON

3D Partitioning Strategies for Application Processors

1. Conventional die stacking

- Application board form factor and package cost reduction
- Move test ownership from application (board) to chip supplier (package)

DBB
NAND
DDR
Interposer
DDR
STn8815S22
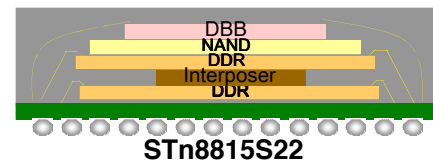
2. Wide-IO / Silicon Interposer

- Massive die interconnect
- Increase memory access bandwidth
- focus on power and performance

Wide IO memory
Host
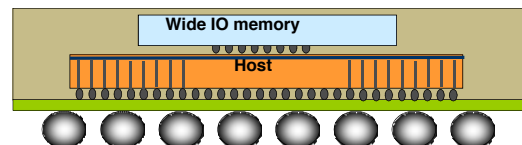
3. Split by process technology

- Implement functions in best suited silicon process technology
- Focus on cost and time to market

Digital
Analog
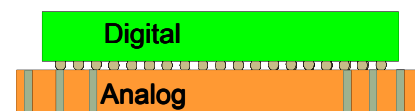
ST ERICSSON

Open Standards

- Smartphone platform heavily relies on industry standard

- With it's parent companies, ST-Ericsson has a strong position and is leading standardization in many areas

INTERFACES	CELLULAR	COMPUTING	MEMORIES

INTERFACES — mipi mobile industry processor interface

CELLULAR — 3GPP A GLOBAL INITIATIVE, 3GPP LTE

COMPUTING — KHRONOS GROUP

CONNECTIVITY — Bluetooth, WiFi, GALILEO, USB UNIVERSAL SERIAL BUS

MEMORIES — EMBEDDED REMOVABLE, JEDEC

ST ERICSSON

ST-Ericsson High-Performance Application Processor

Nova™ BY ST-ERICSSON

Optimized for Mobile

Industry-leading mobile computing, and multimedia performance and power

28nm A9600 Sampling 2011
- A15 dual-core @ 2.5GHz with 20,000 DMIPS
- 20x graphics improvement * with Imagination Rogue
- Next level of power innovation

32nm A9540 Sampling 2011
- A9 dual-core @ 1.85GHz
- 3x graphics improvement *
- 2x Memory bandwidth
- Lower power through process & architecture innovation

45nm A9500 Available
- A9 dual-core @ 1.2GHz
- 20% graphics improvement *
- 45nm

* vs U8500

ST ERICSSON

Conclusion

- A unique combination of sensors and high performance/low power computing always connected to the Cloud is going to revolutionize our lives. Smartphones becomes effectively an extension to our senses and brain.

- Very few companies can drive and combine all the necessary technologies
  - Cellular Modem, APU & Computing, Sensors, IC Packaging, Memories, HLOS, CMOS Process

- Low Power is the fundamental driver for innovation

- Smartphones is the new big driver for the semiconductor industry and ST-Ericsson with Its parent companies is in a unique position to take a leading position on this market

ST ERICSSON

THANK YOU

ST ERICSSON