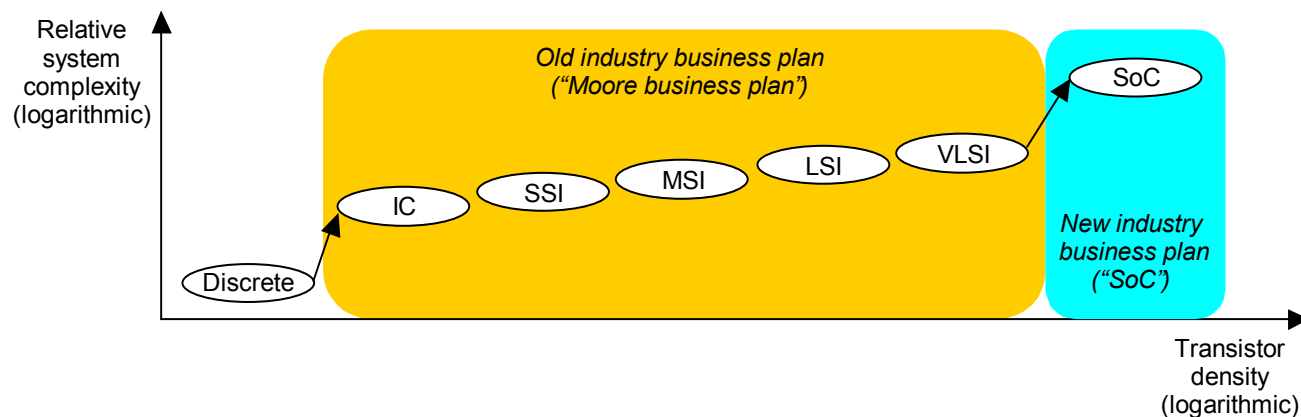


# Embedded Memories First Step Towards SoC

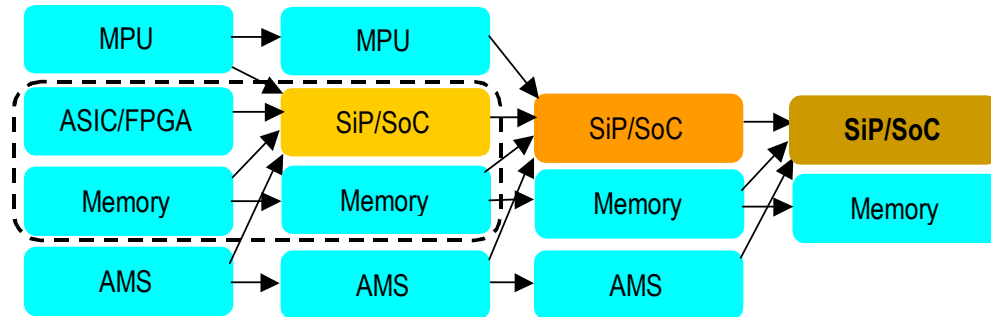
NVMTS 2002



- Semiconductor industry paradigm change
  - From: Moore business plan
    - Quantitative business objectives for the entire industry
  - To: SoC business plan
    - Supplements the quantitative objectives with qualitative ones
    - Driven by economics

- Product classes
  - Mature product classes – the result of the Moore business plan
    - MPU (all logic)
    - ASIC / FPGA (field programmability)
    - DRAM (all memories)
    - AMS (analog / mixed signal)
  - New product classes – the result of the SoC business plan
    - SiP (System-in-Package)
    - SoC (System-on-Chip)
- Design and manufacturing (d&m) technology requirements are driven by the economics of product class
  - Product class business cycles
  - Product class re-tooling cycles
- (Taxonomy consistent with SIA/ITRS view)

# The Convergence Process



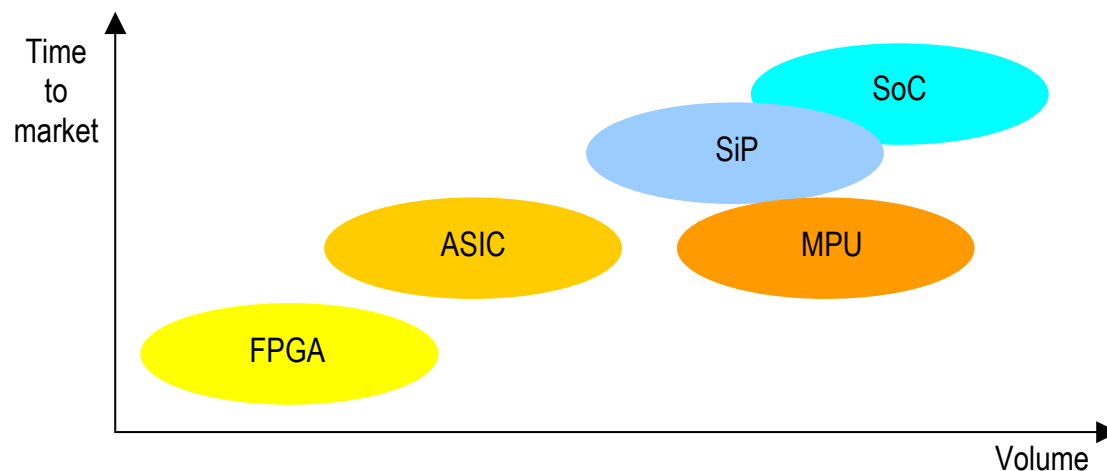
- MPU / ASIC
  - Evolve into SoC
  - Increasingly designed as cores
- SoC
  - Evolves from the ASIC category (as design and manufacturing methodology)
  - Results from the convergence of mature product classes
  - Predicted to drive future semiconductor industry business cycles
- SiP
  - Temporary alternative for SoC
- SiP/SoC
  - Incorporate self-consistent modules from different technology sources
  - Incremental process, each step involves all product classes
  - Memory + logic + AMS (limited) = “embedded memory”

## SoC objectives:

- a) Radical **system cost** improvement
  - Improving and harmonizing d&m technologies
  - Re-use of IP
  
- b) Radical **system performance** improvement
  - More processing power
  - Higher system integration
    - Use of heterogeneous technologies
    - Enables small device form factors
  - Higher system reliability
  
- Old industry business model
  - Objective: **Sub-system** cost and performance optimization
  - Results: High **sub-system** cost and performance improvement rate  
(leads only to marginal system performance and cost improvement)
  
- New industry business model
  - Objective: **System** cost and performance optimization
  - Expected results: High **system** cost and performance improvement rate
  
- System optimization – a fundamental challenge
  - Sub-system optimization does not necessarily lead to system optimization
  - System optimization takes precedence

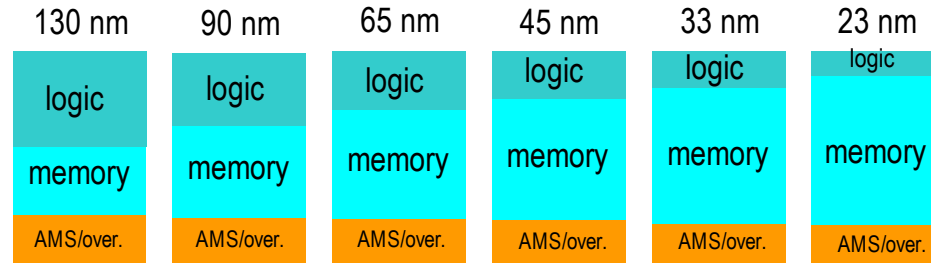
- “Embedded memory”
  - SoC sub-category
  - Driven by the harmonization of
    - Memory technologies
    - MPU technologies
- SoC vs SiP
  - Both drive the convergence of heterogeneous technologies
    - Within one package – SiP (System-in-Package)
    - Onto one silicon die – SoC (System-on-Chip)

# System Cost Considerations



- System costs vary continuously in a 3-dimensional space
  - System complexity
  - Time-to-market
  - Manufacturing volume
- Different regions of the cost space are best served by a particular system implementation
  - FPGA; ASIC; MPU; SiP; SoC
- Integration density limited by 2 key technical factors
  - **Die area**
  - **Power density**

## SoC structure evolution trend

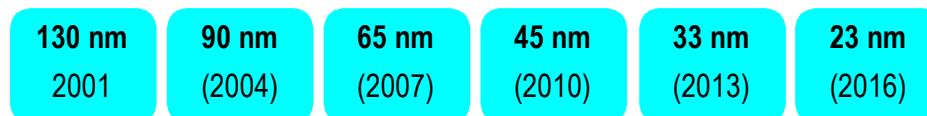


- Die size to stay flat
- Die sections (based on future system requirements)
  - Logic
  - Memory
  - AMS
  - Overhead
- Section sizes – variable
  - AMS + overhead                      – slightly decreasing share
  - Logic + memory                        – slightly increasing share
    - Memory share grows to the detriment of logic share
- “Processor in Memory”
  - Memory > 80% of die area

- Power density limitation – addition to the Moore business plan
  - New requirement – flat throughout product generations
- Extreme power density limitation – portable applications
  - More power constraints
  - Battery performance – only marginal improvement
- Factors that increase power density :
  - Voltage
    - Historic scaling rate: – 15% per node
    - Required scaling rate: – 30% per node
  - On-chip frequency
    - Historic scaling rate: – 200% per node
    - Required scaling rate: – 140% per node
  - Packaging
    - Pin count and pin pitch advances << integration density advances

- Methods to reduce power density
  - Die area limitation
  - Logic transistor limitation
  - Supply voltage reduction
  - Clock frequency limitation/reduction
  - More asynchronous than synchronous logic
  - More hard-wired than microprogrammed logic
    - Power efficiency: hard-wired vs microprogrammed logic (GOPs/mW)
      - Hard-wired logic is approximately 4 times more efficient than microprogrammed logic
    - Consequence
      - More: hard-wired special purpose cores
      - Less: microprogrammed general purpose cores

# System Performance Considerations

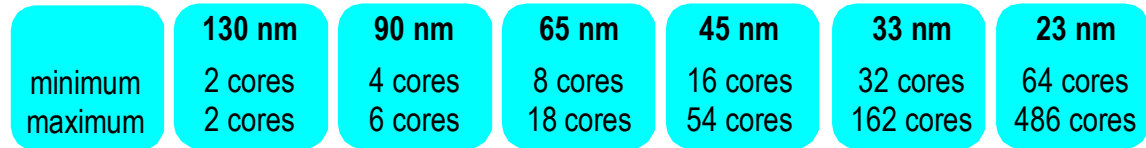


- System performance is paced by technology nodes, rather than absolute time intervals (years shown for orientation)
- Reason
  - Logic and memories have essentially constant layout densities (planar devices)
- One-off performance “discontinuities” might occur when new non-planar elements (dual-gate FET, FinFET), or new technologies (e.g. SOI) are introduced

- MPU architectures
  - General purpose
    - Higher flexibility, lower efficiency
    - Higher number of transistors
    - Examples – computer processors
  - Specific purpose
    - Lower flexibility, higher efficiency
    - Lower number of transistors
    - Examples – arithmetic, graphics, encryption, networking processors

- Core (ASIC terminology)
  - Functional modules that can be readily integrated into the system
  
- Performance improvement methods
  - a) Micro-architecture improvements (e.g. pipelining, scalarization, predictive algorithms)
    - Micro-architecture efficiency
      - Computing performance increase – 150%
      - Silicon area increase – 300%
  
  - b) Parallel processing (i.e. multiple processing units in parallel)
  
  - c) Task-specific hardwired cores (e.g. arithmetic, graphics, encryption, networking)
    - Parallel processing / hardwired core efficiency
      - Computing performance increase – 180%
      - Silicon area increase – 200%
  
    - Multi-core structures
      - Help improve processing performance in an efficient way
      - Future solution of choice

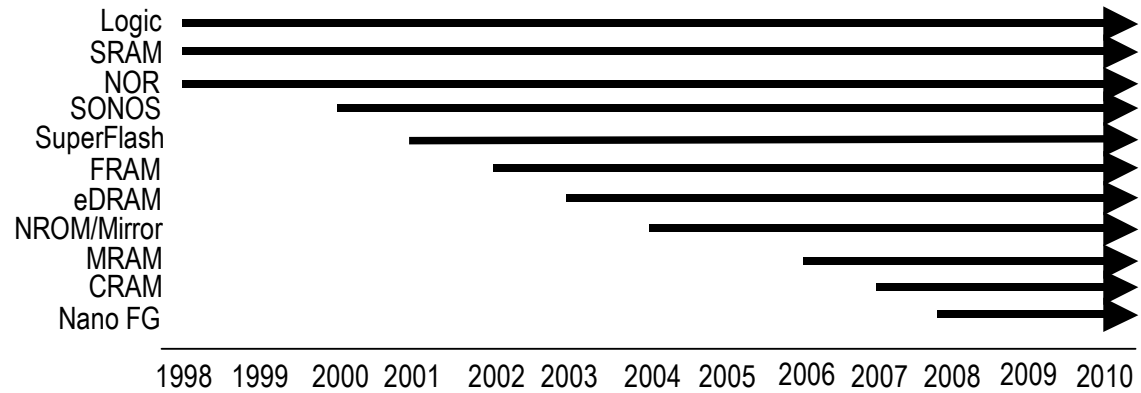
# System Performance Enhancing – Processing-3



- Future trend
  - Shift
    - From: general-purpose cores (high flexibility / low efficiency)
    - To: task-specific cores (low flexibility / high efficiency)
  - More emphasis on parallel processing and task-dedicated “helper-engines”
- The number of processing cores per chip will increase between two and three fold with every technology node
  - Individual configurations have specific growth rates
- Multi-core structure benefits
  - Power savings (selective activation)
  - Use of redundancy to improve manufacturing yields

- No need for inter-block standard interfaces
  - Optimized number of links
  - Optimized timing
  
- System busses can be replaced with high-speed point-to-point links implemented as transmission lines
  - Adapted for a certain load
  - Very high speeds
  - Optimized for both
    - Cost
    - Performance
  - AMS designs
    - Receiver, transmitter and timing recovery blocks require analog circuitry

# System Performance Enhancing – Memory



- First integration of the memory technology with logic in standard cmos process
- The integration process is increasingly modular
- Memory modules and system modules are not “stack-able” so far

# System Performance Enhancing – eMemory

- eMemory (eNVM) technology – optimized for embedded configurations
- eNVM – Technology requirements
  - Smallest memory block, rather than smallest cell size
  - Simplest manufacturing process
  - Smallest number of additional masks/steps
  - Less HV circuitry, if possible at all (advanced NVM technologies at an advantage)
- eNVM – System requirements
  - Access speed ~ processor bus speed
  - Symmetrical access cycles
- eNVM – Advantages
  - Better component economics
  - Higher system benefits
- eNVM – Examples
  - SuperFlash
  - SONOS
  - HIMOS
  - GreenFlash
  - eFalsh
- Note:
  - Discrete memory technologies have lower efficiency when integrated with logic

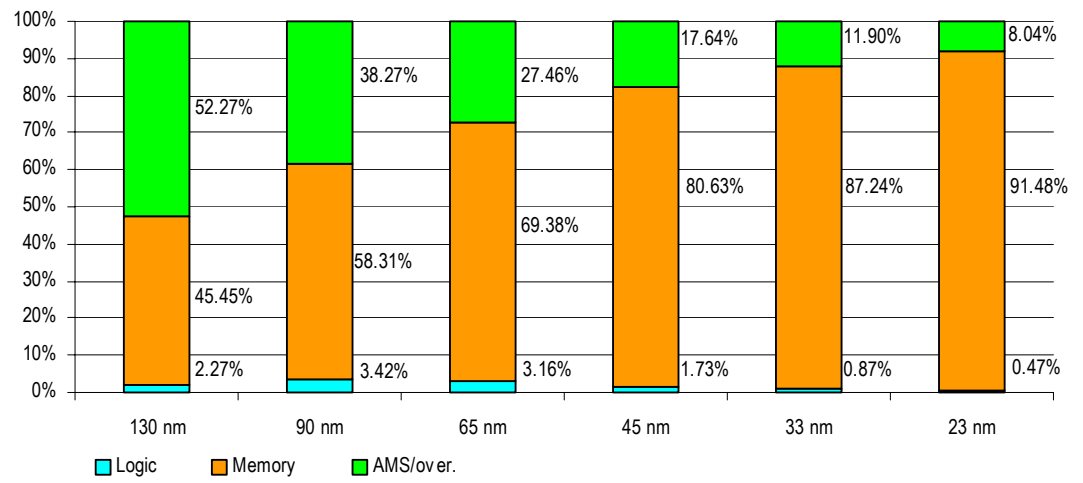
	Assumption	Units	PDA
MPU class			CE
Die size	constant	mm2	140
AMS/overhead size	constant	mm2	40
<b>Logic/memory size</b>	<b>constant</b>	<b>mm2</b>	<b>100</b>
Power density	constant	W	0.1

- Application class  
**PDA – with cost effective type logic (CE)**
- Configuration objectives per application and technology node
  - Adequate amount of logic – market requirements
  - Necessary amount of AMS / overhead – system requirements
  - Amount of memory – maximum
- Restrictions
  - Die size – constant from one product generation to the other
  - Power density – ibidem

- Professional market segment (e.g. PDA)
  - Average volumes, average margins
- Short product life cycle
  - Rapid system implementation
  - Short time-to-market
- Requires limited user programming capability
- Multiple cores / higher processing power driven by:
  - Scheduler
  - Web browser
  - Video communications
  - Voice-text / text-voice translation
  - Streaming video
  - Authentication (crypto engine)

Units	130 nm	90 nm	65 nm	45 nm	33 nm	23 nm
PDA cores	4	15	35	55	70	85

Million transistors	130 nm	90 nm	65 nm	45 nm	33 nm	23 nm
Logic	0.5	1.9	4.4	6.9	8.8	10.6
Memory (equivalent transistors)	10.0	32.0	96.0	320.0	880.0	2,048.0
AMS/overhead	11.5	21.0	38.0	70.0	120.0	180.0
Total	22.0	54.9	138.4	396.9	1,008.8	2,238.6



Mb	130 nm	90 nm	65 nm	45 nm	33 nm	23 nm
SRAM	2	n.a.	n.a.	n.a.	n.a.	n.a.
NOR	5	16	48	160	440	1,024
SONOS	5	17	n.a.	n.a.	n.a.	n.a.
SuperFlash	6	19	58	n.a.	n.a.	n.a.
FRAM	9	30	90	299	821	1,911
eDRAM	n.a.	72	216	720	1,980	4,608
NROM/MirrorB	20	64	192	n.a.	n.a.	n.a.
MRAM	n.a.	56	168	560	1,540	3,584
CRAM	n.a.	n.a.	224	747	n.a.	n.a.
Nano FG	n.a.	n.a.	384	1,280	3,520	8,192

- Size of the logic/memory block – 100mm<sup>2</sup>
- Power limitations – 0.1W (per die)
- NVM needed for
  - Work memory (code storage); requires high write endurance
  - External storage (data file storage)
- SiP can be utilized as an intermediate step towards SoC

- Economics of embedded memory configurations will improve with
  - Technology maturity
  - New applications
  - Higher volumes
  
- Memory technologies will be optimized for two types of applications
  - Discrete memory components
    - Work memory
    - Storage
  - eMemory – key competitive factor
    - Leading to SoC
  
- Strategic perspective – SoC and eMemory technologies
  - Less optional
  - Increasingly compulsory
  
- Most systems will need more memory than embed-able at any time
  - Solution: on-die and off-die work memory
  - On-die work memory has higher performance