Quintessence Architectures, Inc.

Architecting High-Performance SoC Video Processors

Sorin C. Cismas

http://www.quarc.com

Outline

- Digital Video and the SoC Challenges
- Architecture, Design Methodology and Tools
- Videris™ HD - MPEG2 4:2:2@HL Video Decoder
  - Multi-Threaded MPEG Decoder
  - Fused Multiply/Add/Subtract DCT
  - Tile based Super-Scalar Memory Controller
  - Videris™ HD Statistics
- Conclusions

Outline

- **Digital Video and the SoC Challenges**
- **Architecture, Design Methodology and Tools**
- **Videris™ HD - MPEG2 4:2:2@HL Video Decoder**
  - **Multi-Threaded MPEG Decoder**
  - **Fused Multiply/Add/Subtract DCT**
  - **Tile based Super-Scalar Memory Controller**
  - **Videris™ HD Statistics**
- **Conclusions**

More Performance @ Low Power

- **Higher Resolution, Higher Bitrate - Video quality**
  - Video CD 360 x 240 @ 1-3 Mbit/sec.
  - DVD 720 x 480 @ 4-10 Mbit/sec.
  - HDTV 1920 x 1080 @ 20-40 Mbit/sec.
  - Digital Cinema 4096 x 3072 @ ??? Mbit/sec.
- **Multiple Streams, Multiple Standards - Flexibility**
  - MPEG2 (DirecTV, DVD, DVB, ATSC, ISDB)
  - MPEG4, J2K, MJ2K
  - JPEG, DV
- **Wireless and Portable - Low power**
  - Limited and variable bandwidth
  - Scalable performance

SoC (Systems not Chips)

- System = Hardware + Software + Application
- Hardware/Software partitioning is crucial
- The SoC Challenges
  - reuse and easy integration
  - faster time-to-market and smaller circuit geometry
  - high performance and low power
  - all of the above @ low cost
- The Solution - Innovative Architectures, Design Methodologies and Tools
  - they will drive the SoC revolution
  - handcrafting to squeeze the last picoseconds and the last thousands gates will be a thing of the past

Outline

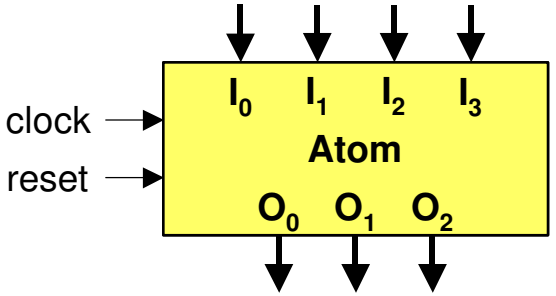
- Digital Video and the SoC Challenges
- **Architecture, Design Methodology and Tools**
- Videris™ HD - MPEG2 4:2:2@HL Video Decoder
  - Multi-Threaded MPEG Decoder
  - Fused Multiply/Add/Subtract DCT
  - Tile based Super-Scalar Memory Controller
  - Videris™ HD Statistics
- Conclusions

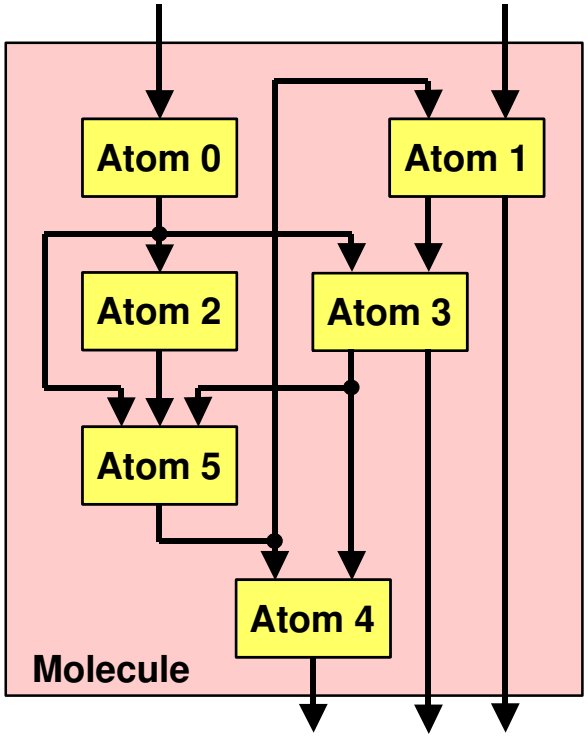
QuArc Architecture and Methodology

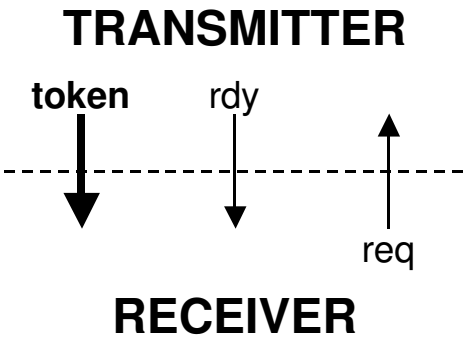
- "divide et impera"
  - Functional partitioning in manageable objects
  - Well defined interfaces
  - Independently testable objects, easy to integrate, and reuse
  - Automate most of the design, verification, and synthesis process
  - Enable engineers to work on the creative and fun stuff
- Key Features
  - Encapsulates algorithms in self-contained data-driven objects
  - No need for master controller or scheduler
  - Synchronous but self-timed (variable schedule, elastic pipelines)
  - Adapts to instantaneous variations in processing load
  - Split memory transactions
  - Stall tolerant - works well in systems with shared memory
  - Minimal interfaces to simplify the wiring complexity

QuArc Objects

- **Any design is a collection of Objects**
  - **Atoms: leaf objects (indivisible)**
    - **two global signals: clock and reset**
    - **one or more Input Interfaces**
    - **one or more Output Interfaces**
  - **Molecules: collection of Atoms and/or Molecules**
  - **Interfaces:**





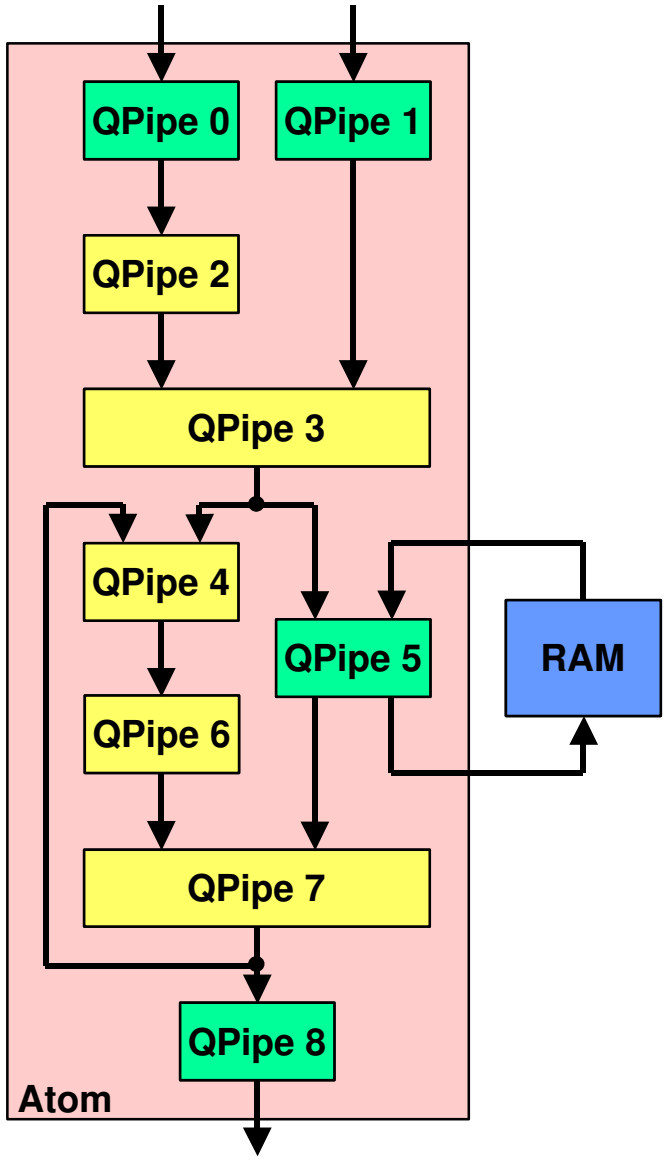


QuArc Interfaces

- Minimal set of signals
- Synchronous and uni-directional
- One transmitter and one or more receivers
- Sustained one token/cycle throughput
- Token bus:
  - At the physical level, like any other data bus
  - At the logical level, equivalent to a C-language data structure
  - Can be a collection of sub-busses, each with its own syntax
- Handshake signals:
  - Simple rdy/req handshake protocol
  - One rdy/req pair for each receiver
  - Data is handed over when both rdy and req are asserted
  - Transmitter and Receivers can stall the transaction in any cycle

QuArc Pipestages (Qpipes)

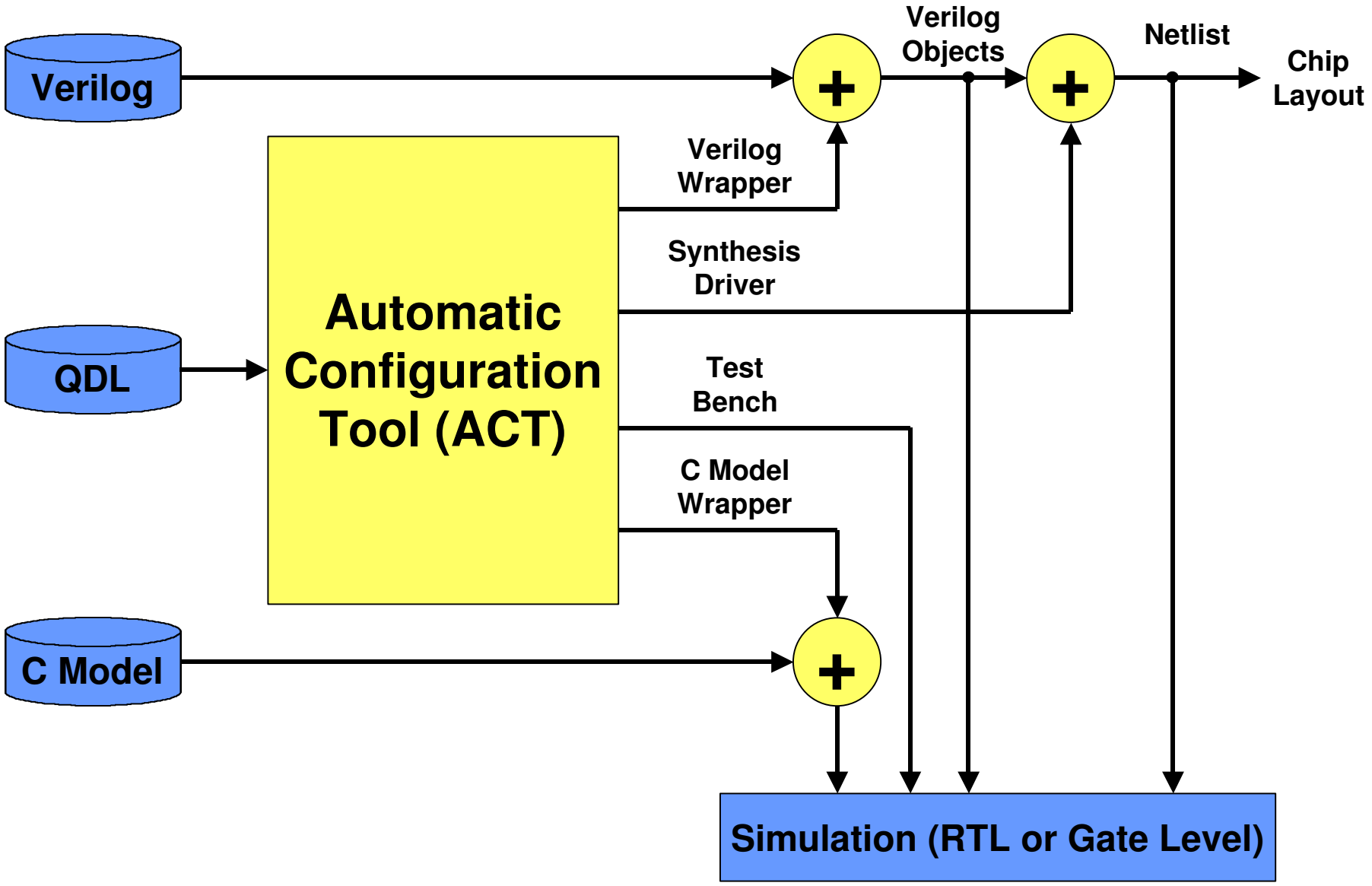
- Contains one or more registers, sometimes a memory
- Has it's own controller that keeps track of how many tokens are in the pipeline
- Atoms knows when valid data is in the pipeline
- Atoms can independently shut down the clock to save power
- Library of Qpipes
  - Hides variable schedule complexity
  - Simplifies design task
  - Designers can focus on algorithms, not on low level control



QuArc Design Language

- **Every Object has a .qdl file (Object Spec.)**
  - **What the Object is**
  - **How to use other Objects to build a system**
- **QDL files have four parts**
  - **Parameters - customizes Atoms and Interfaces based on the system requirements**
  - **Interfaces - describes their properties**
    - **Input/Output**
    - **Interface Type (Class)**
      - **the syntax is described in a QDL library**
      - **only Interfaces of the same type can be connected together**
    - **Prefix - to uniquely identify an Interface if an Object has more than one of the same Type**
  - **Instantiations (for Molecules only)**
  - **Register Description (for Atoms only)**

Automatic Configuration Tool



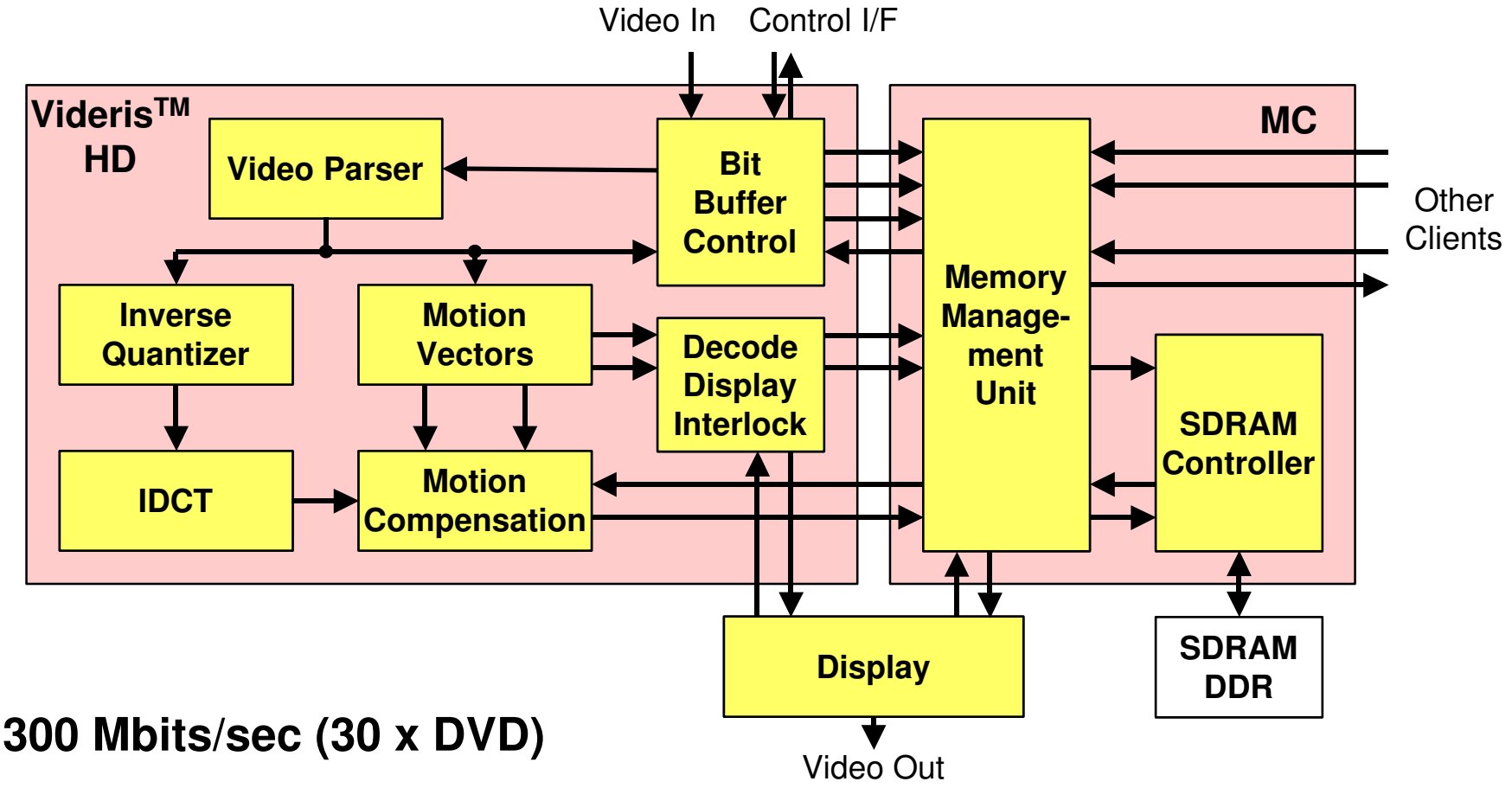
Design Style Rules for Easy Design Reuse

- Positive-edge triggered flip-flops, no latches
- Single clock
- Reset can be synchronous or asynchronous
- Control registers are always reset
- Low input set-up time (<25% of the cycle time)
- Low output delay time (<25% of the cycle time)
- Output Data comes directly from registers
- Input Data goes directly to registers
- No combinational paths from inputs to outputs
- Simple internal RAM (1-port or 1 read/1 write port)

Outline

- Digital Video and the SoC Challenges
- Architecture, Design Methodology and Tools
- **Videris™ HD - MPEG2 4:2:2@HL Video Decoder**
  - Multi-Threaded MPEG Decoder
  - Fused Multiply/Add/Subtract DCT
  - Tile based Super-Scalar Memory Controller
  - Videris™ HD Statistics
- Conclusions

MPEG2 4:2:2@HL Video Decoder



- 300 Mbits/sec (30 x DVD)
- All non-scalable profiles at all levels
- Dedicated, hard-wired units to guarantee high-performance at low power and low cost
- High level decisions and error recovery is under software control (few MIPS and not real-time critical)

Multithreaded MPEG Decoder

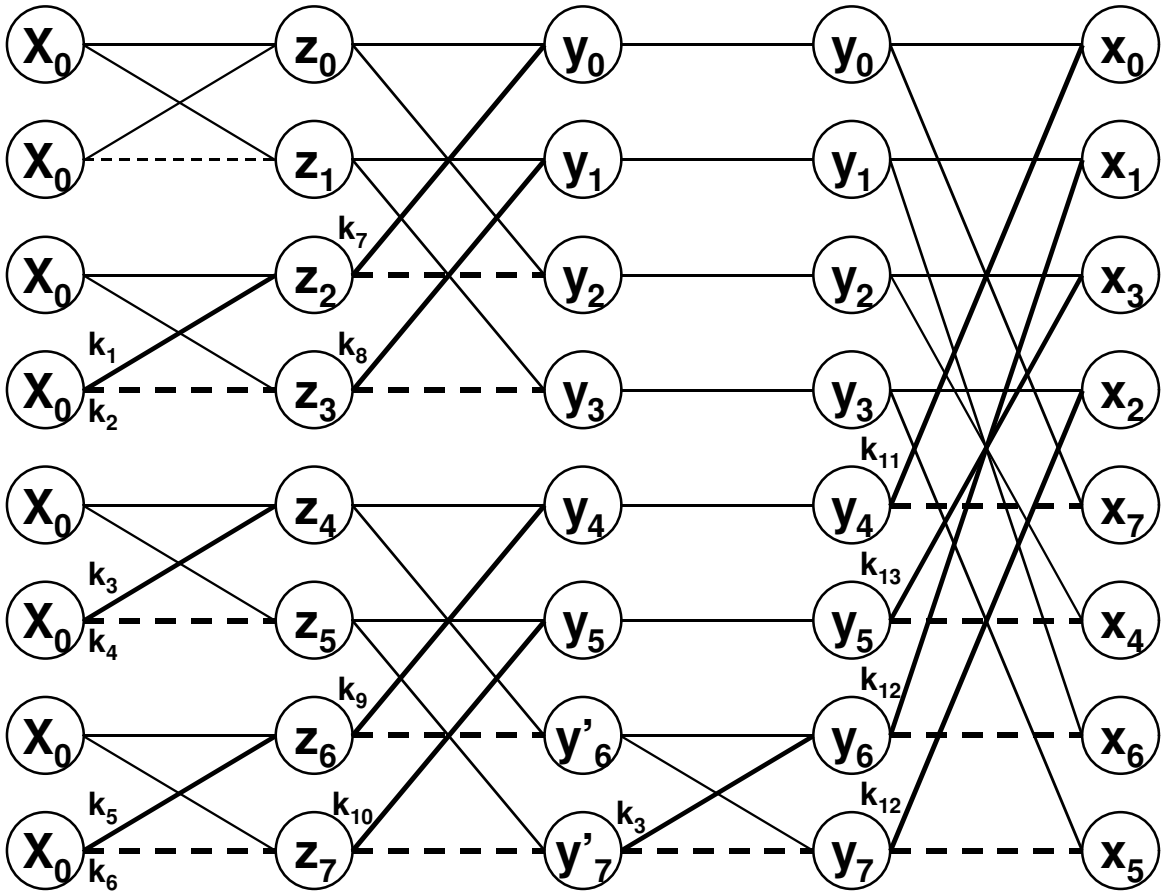
- **Video processors will evolve from dedicated, one-application-at-a-time to multithreading**
  - **In the '90ties, single stream decoders**
  - **Context switching and multithreading is a must in future visual communication and entertainment devices**
- **MPEG multithreading**
  - **Most HD decoders have the processing power to decode several SD bitstreams. After a picture is completely decoded, a new picture from a different bitstream can be decoded, but context switching takes far to many cycles and reduces performance**
  - **Videris™ HD can process up to 16 bitstreams simultaneously, each with its own different bitrate, resolution and frame rate, without any penalty in stall cycles**
  - **This is an important feature when many and relatively small MPEG textures need to be mapped on objects, as in games, multimedia and visual communications**

Video Context Switching

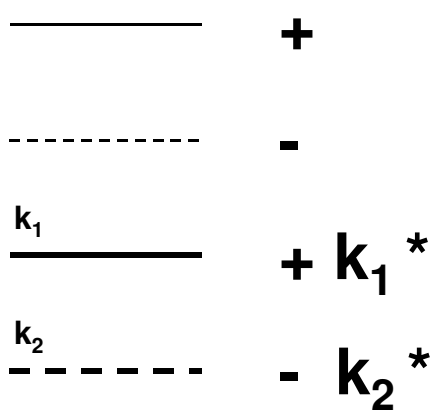
- **Context stored in special QuArc Pipestages distributed over the whole design**
  - **too much overhead to store/retrieve it to/from memory**
  - **very long pipeline compared to general purpose processors**
  - **context switch happens at different times in different objects**
  - **at any time, Videris™ HD can be processing several contexts**
- **Context switching is supported in QDL**
- **ACT configures the QuArc Pipestages for the requested number of contexts**
- **Independent Bit Buffer Control and Decode/Display Interlock for each bitstream**

Multiply/Add/Subtract IDCT Algorithm

- **Based on a DCT Algorithm by Elliot Linzer and Ephraim Feig (26 Multiply/Add)**
- **Our Algorithm (16 Multiply/Add/Subtract)**



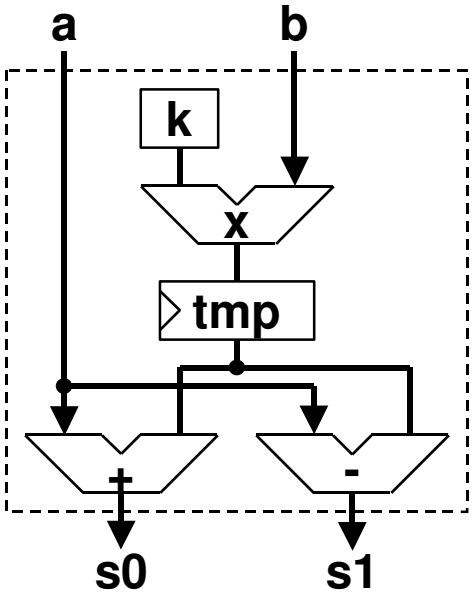
Legend:

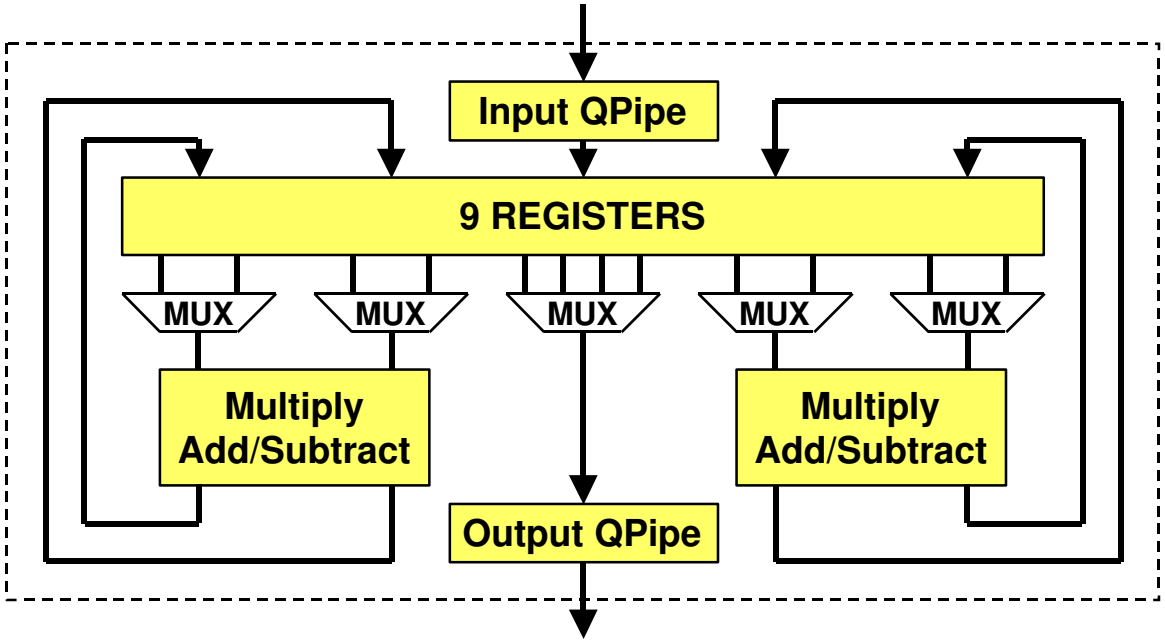


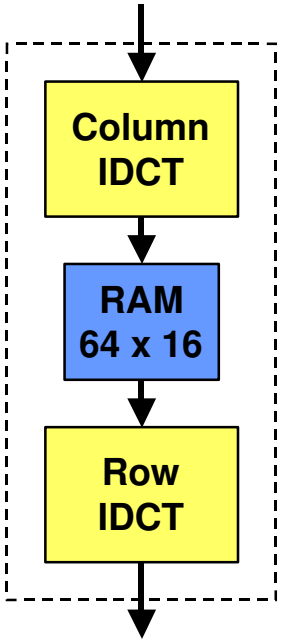
Multiply/Add/Subtract IDCT Architecture

MAS Operation:

tmp = k * b
s0 = a + tmp
s1 = a - tmp







Multiply/Add/Subtract IDCT Features

- Minimal hardware: 2 MAS, 9 registers and 5 muxes
- Throughput: 1 DCT coefficient/cycle - no stall cycles (94MHz for HD)
- Latency: 12 cycles - 2 or 3 IDCTs are processed at any time
- Narrow busses: 16 bits on the interfaces, 18 bits internally
- HD requires two IDCTs plus a 64 x 16 transpose memory
- Exceeds IEEE 1180 requirements
- Extra bit in data path to guard against large quantization noise

Precision	IEEE 1180	QuArc
Peak Error	1	1
Pixel Mean Error	0.015	0.0043
Overall Mean Error	0.0015	0.0001
Pixel Mean Square Error	0.06	0.0226
Overall Mean Square Error	0.02	0.0185

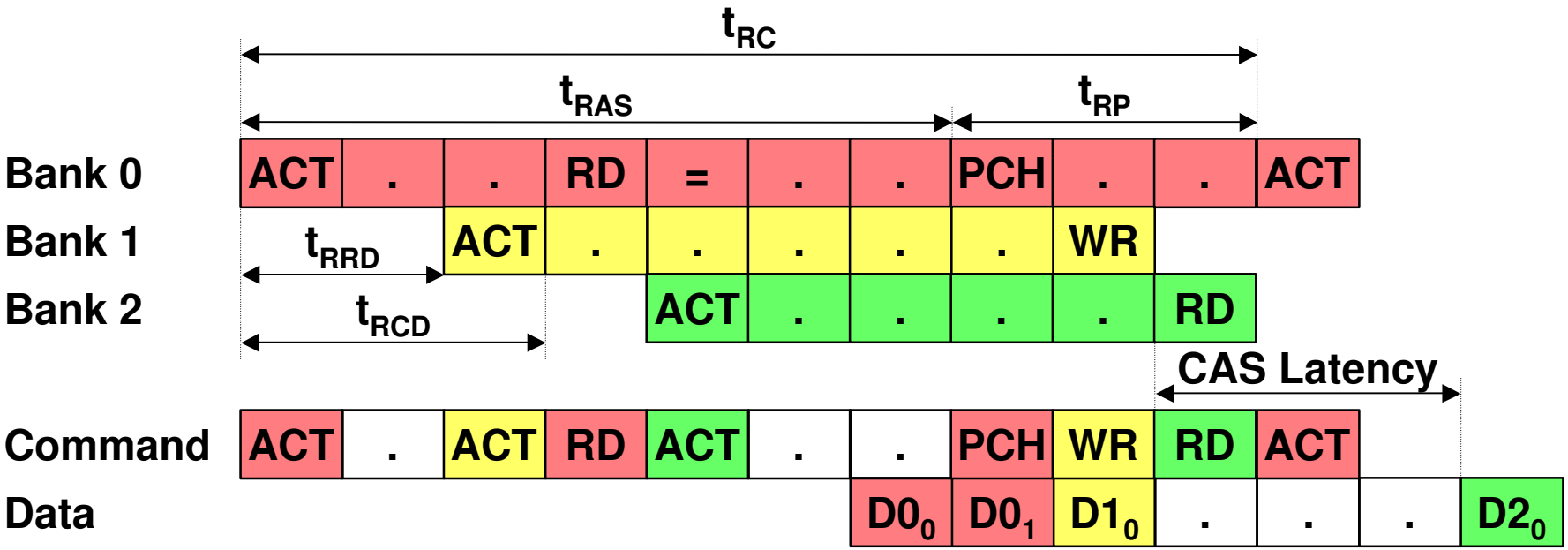
Memory Bandwidth

- Biggest limitation for high performance (microprocessors, 3D and video)
- Higher memory density makes them cheaper and deeper, but not faster
- Shared memory is a must for SoC
- MMU becomes the bottleneck that distributes bandwidth to many clients
- Caches do not necessarily help for video (random accesses for small blocks of data)
- Tile based memory organization, multiple banks and memory transactions reordering do help

Microprocessor vs. SDRAM Pipeline

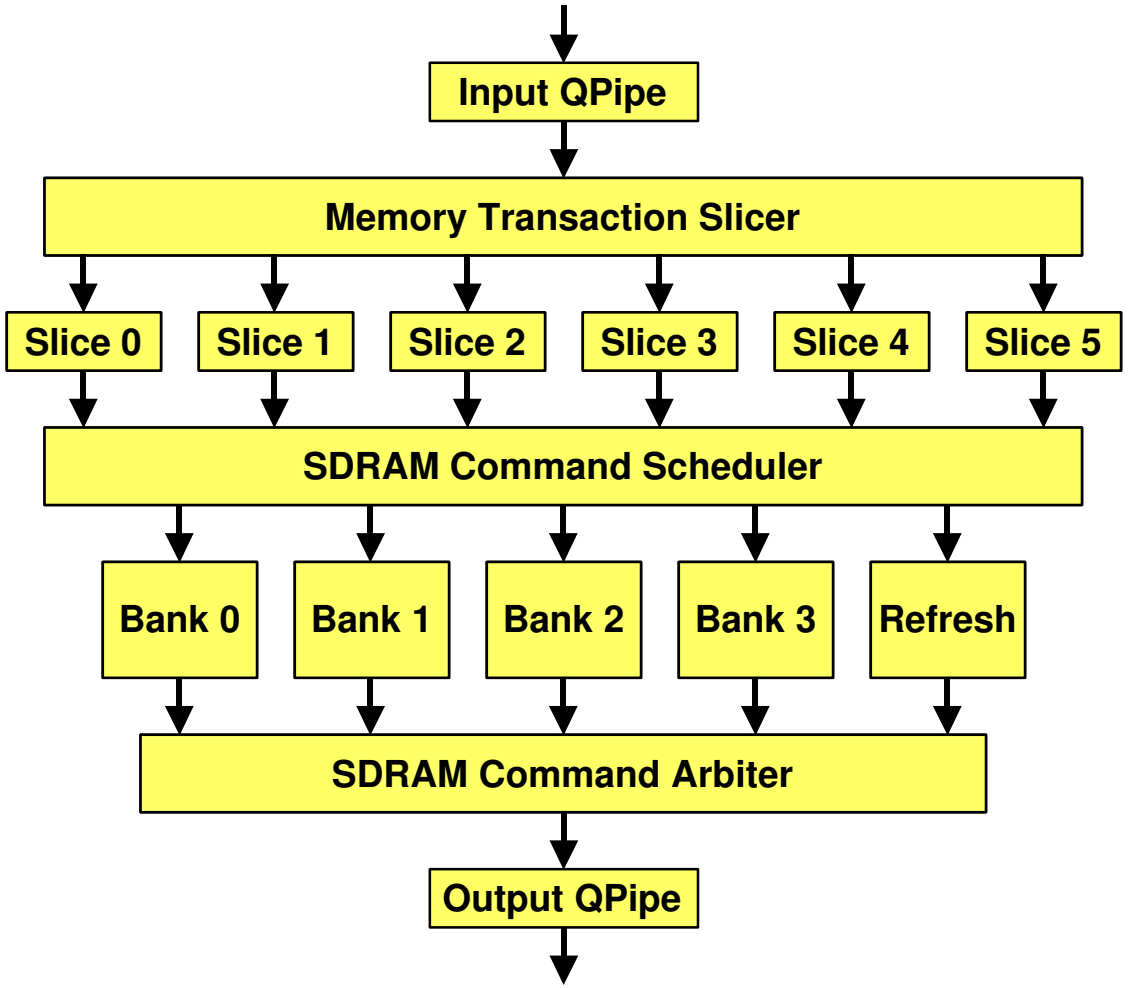
- **Microprocessor Pipeline**
  - Branches
  - Load delay

- **SDRAM Pipeline**
  - Variable length
  - Shared command bus
  - Write to Read penalty = CAS Latency



Super-Scalar Memory Controller

- **Complex Memory Transactions (array with implicit or variable stride) are translated to Simple Slice Instructions (access to consecutive addresses within the same bank and page)**



Complex Transactions

Simple Slice Instructions

6-wide Instruction Window

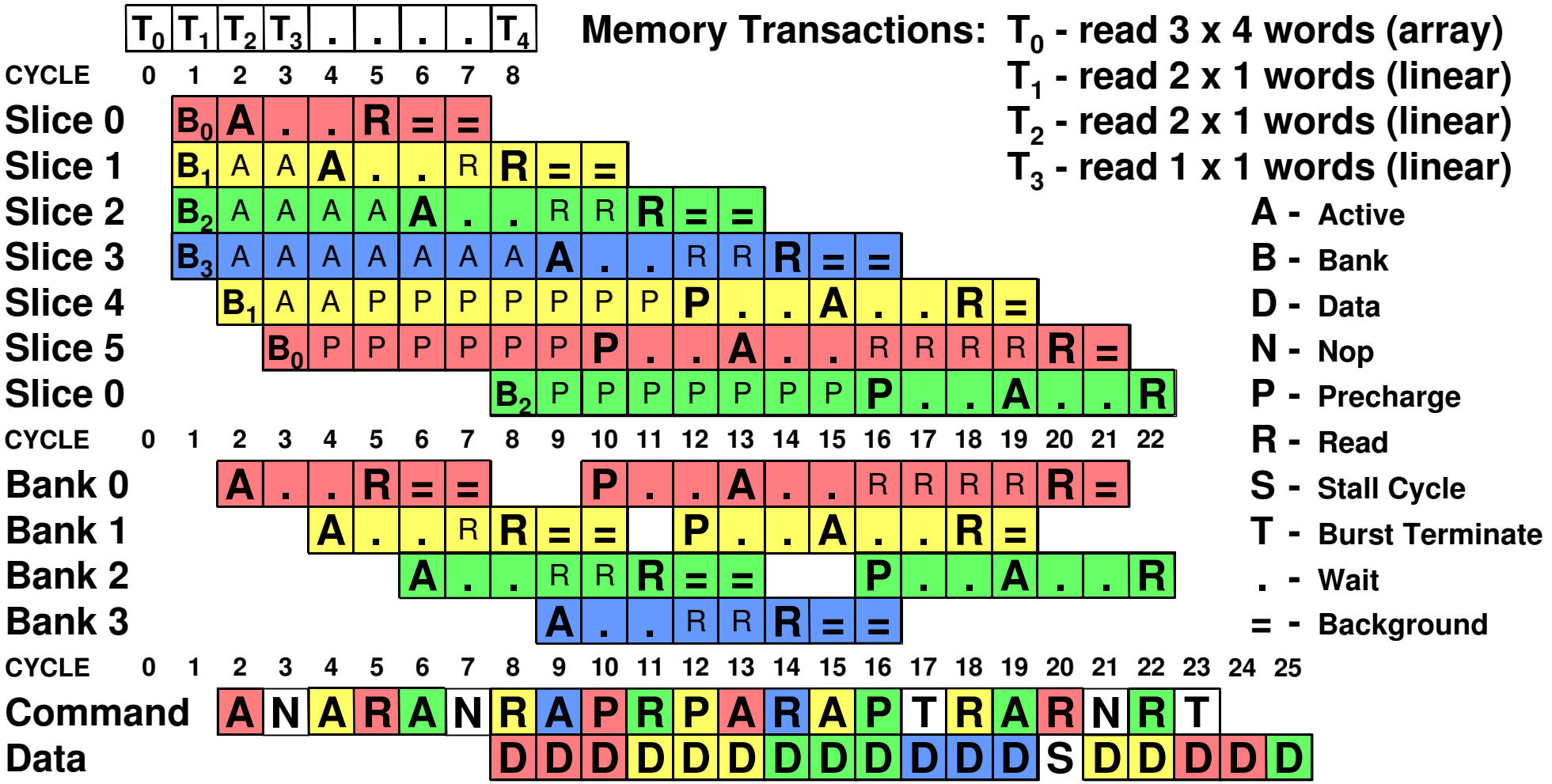
Out of Order Commands

(4+1) wide super-scalar

In Order Data

SDRAM Commands

Pipeline Operation



- Out of Order Commands: T_2 (Slice 5) is executed before T_1 (Slice 4)
- Stall cycle can be eliminated with Out of Order Data (requires on-chip memory to reorder the data for the clients)

Memory Controller Performance

- 90 - 94% memory utilization
- Needs memory transaction size longer than t_{RC} (Active to Active command period)
- MPEG prediction size is 3 x 9, 3 x 8, 3 x 5, and 3 x 4 words on a 64-bit bus
- This high efficiency enables Videris HD to use a 32-bit SDRAM or 16-bit DDR at 133MHz for HDTV
- 93 - 96% memory utilization with Out of Order Data
  - on-chip memories are needed for both read and write to reorder the data for the clients)
  - longer latencies

Videris™ HD Statistics

Videris™ HD Object	Base Area	Inc. Area	Base RAM	Inc. RAM
Video Parser	0.137 mm^2	0.004 mm^2	128 bytes	128 bytes
Inverse Quantizer	0.073 mm^2	0.005 mm^2	96 bytes	
IDCT	0.195 mm^2		128 bytes	
Motion Vectors	0.087 mm^2	0.005 mm^2		
Motion Compensation	0.148 mm^2		432 bytes	
Bit Buffer Control	0.154 mm^2	0.019 mm^2	512 bytes	
Decode/Display Interlock	0.045 mm^2	0.003 mm^2		
Videris™ HD Total	0.839 mm^2	0.035 mm^2	1,296 bytes	128 bytes
Memory Management Unit	0.340 mm^2	0.030 mm^2	1,280 bytes	
SDRAM Controller	0.193 mm^2			
Videris™ HD + Mem. Cntr.	1.372 mm^2	0.065 mm^2	2,576 bytes	128 bytes

- Logic Area based on TSMC 0.18um, 150MHz, worst-case operating conditions (V, T, P)
- up to 200MHz worst-case operating conditions (logic area increases by approx. 20%)
- Base Area (routing overhead not included) and Base RAM are for one context
- For every additional context, add Inc. Area and Inc. RAM
- Total size, including routing overhead for one HD + two SD (or eight SD) is 3-4 mm^2

Outline

- Digital Video and the SoC Challenges
- Architecture, Design Methodology and Tools
- Videris™ HD - MPEG2 4:2:2@HL Video Decoder
  - Multi-Threaded MPEG Decoder
  - Fused Multiply/Add/Subtract DCT
  - Tile based Super-Scalar Memory Controller
  - Videris™ HD Statistics
- **Conclusions**

Conclusions

- **Pure Hardware or pure Software are not the right answer for the future visual communication and entertainment devices**
- **SoC is a big challenge for all semiconductor companies**
- **Moore's Law can not be sustained just by the future progress in semiconductor processes**
- **Architecture, Design Methodology and Tools will drive the semiconductor industry**