TOSHIBA
Leading Innovation

Venezia: a Scalable Multicore Subsystem for Multimedia Applications

Takashi Miyamori

Toshiba Corporation

Outline

- **Background**
- **Venezia Hardware Architecture**
- **Venezia Software Architecture**
- **Evaluation Chip and Performance Results**
- **Summary**

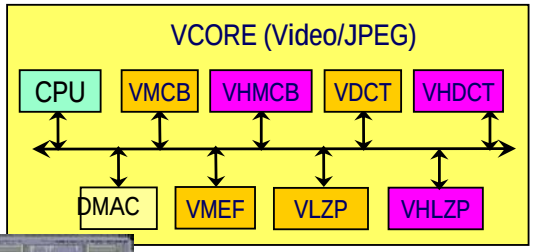
Outline

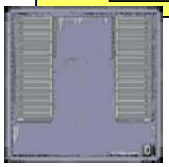
- **Background**
- Venezia Hardware Architecture
- Venezia Software Architecture
- Evaluation Chip and Performance Results
- Summary

Background

- **Today's mobile multimedia devices support many audio and video CODECs.**
  H.264, MPEG-4, VC-1, AC-3, MP3, WMA ‥etc.
- **The size of image processing is increasing rapidly.**
  QCIF, CIF, QVGA, VGA, 720p, 1080i, 1080p ‥etc.

Current SoC (T5V)

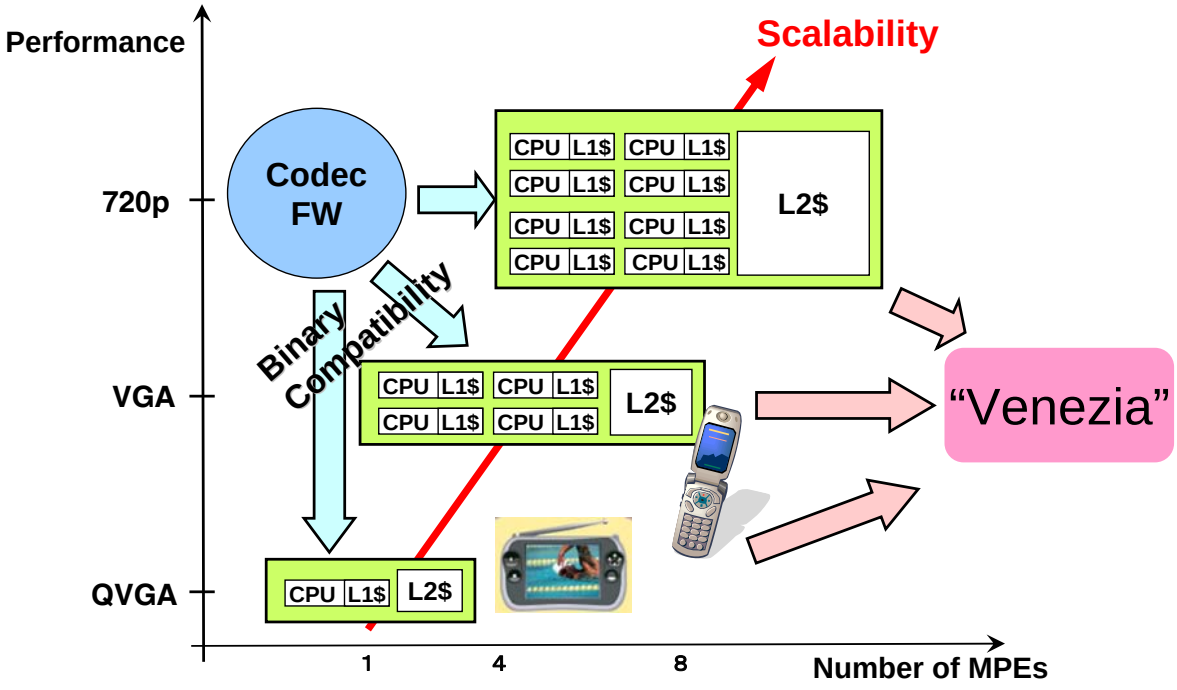




- 90nm CMOS (16Mbit eDRAM x2)
- H.264 decode @CIF 30fps,
  MPEG-4 encode/decode @VGA 30fps

HW Solution:
New designs are required for new CODECs.

Our Approach: Scalable Multicore Processor



"MPSoC Architecture Trade-offs for Multimedia Applications," MPSoC '07

Homogeneous vs. Heterogeneous

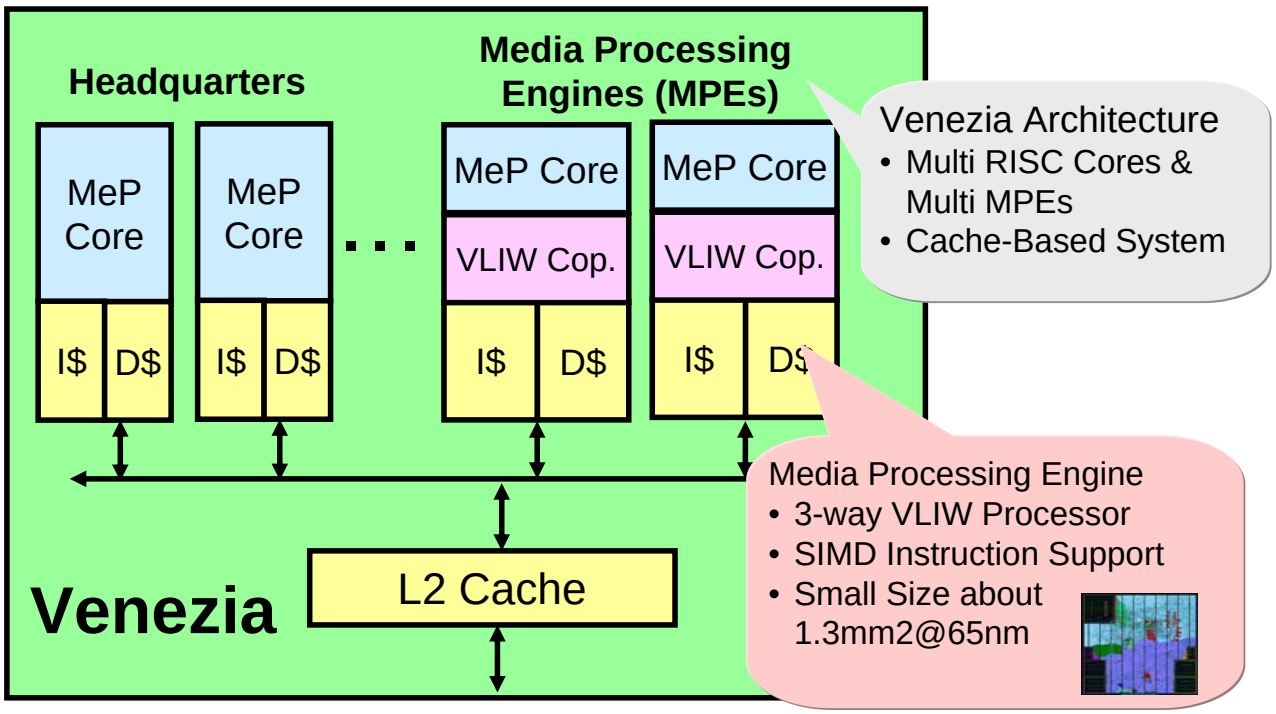
	Homogeneous	Heterogeneous		
Architecture	Mx	MDx	MDA	MDHx
	M M M M	M D D D	M D A	M D H H
Programmability / Scalability	Very good	Good		
Perf./Cost or Perf./Power	Fair	Good		
Examples	Core 2 (M_2, M_4), Xbox 360 CPU (M_3), MPCore(M_4), Niagara(M_8)	Cell (MD_8), SB3000(MD4), Philips Cake/Wasabi	Uniph	current SoCs

M: Main CPU
D: DSP/Media Processor
A: Accelerator
H: Hardware Engine

Outline

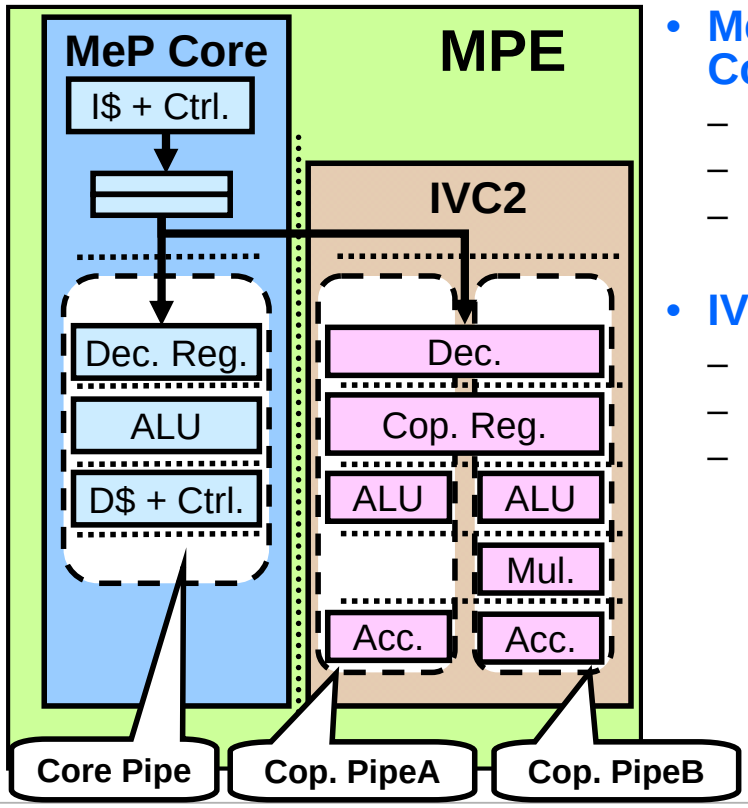
- Background
- **Venezia Hardware Architecture**
- Venezia Software Architecture
- Evaluation Chip and Performance Results
- Summary

Venezia Processor Architecture



MeP: Media embedded Processor

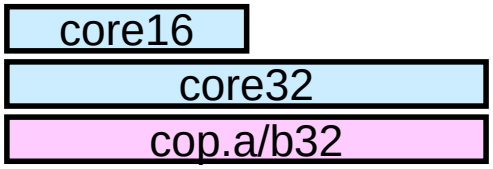
MPE (Media Processing Engine)



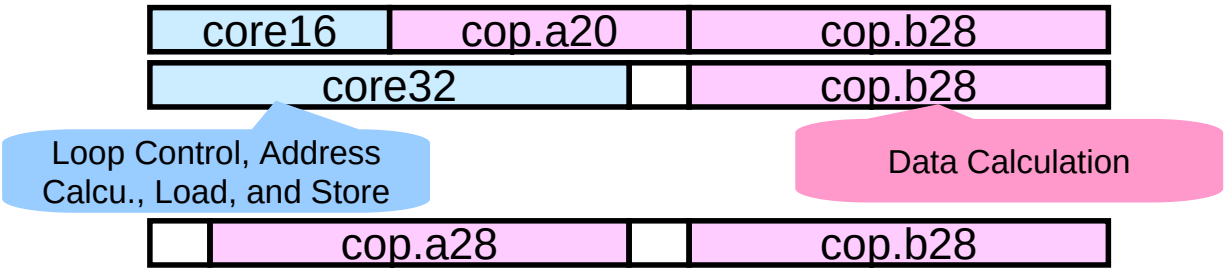
- **MeP (Media embedded Processor) Core**
  - 32-bit RISC Processor
  - 5 Pipeline Stages
  - 3 Low-power Mechanisms
- **IVC2(Coprocessor)**
  - Extension of MeP Core
  - 64-bit SIMD Operations
  - 3-way VLIW
    （Core + Cop.A + Cop.B）

MPE Instruction Formats

- 16b/32b Instruction Mode (Core Mode)

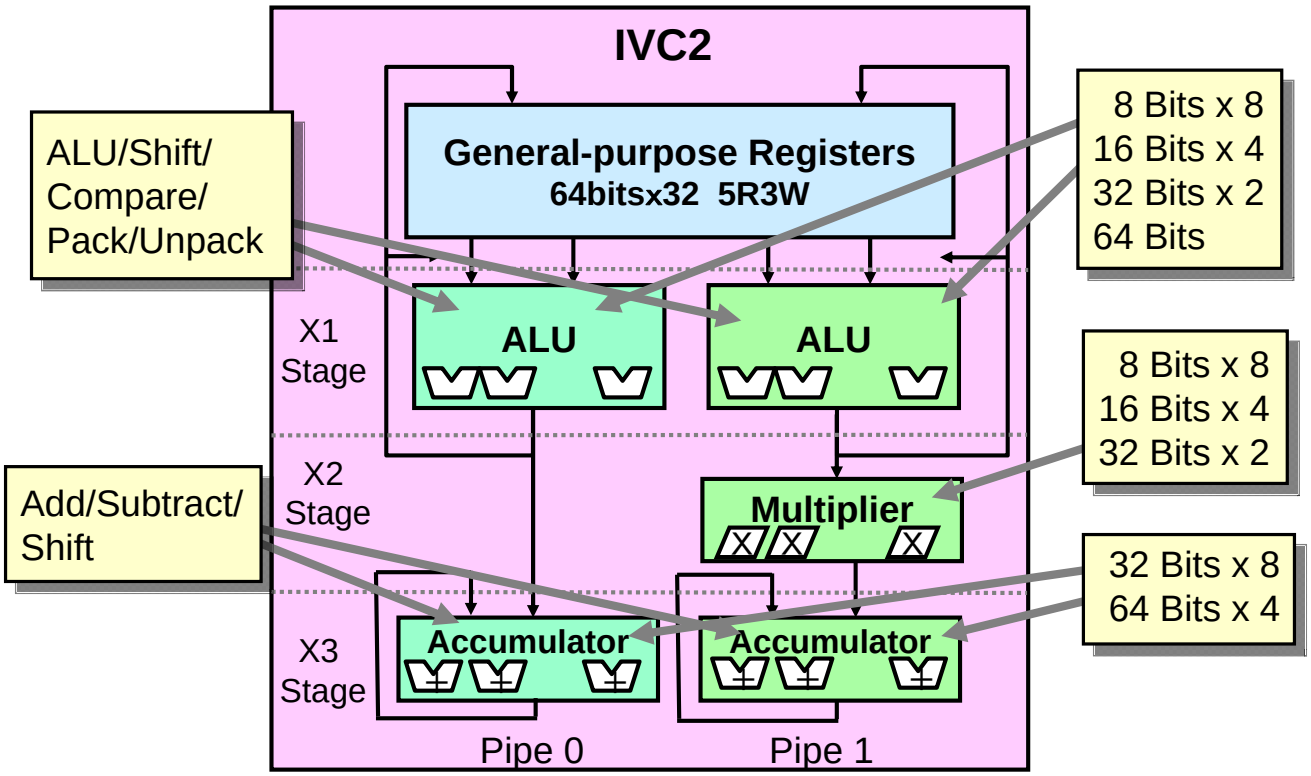


- 64b Instruction Mode (VLIW Mode)

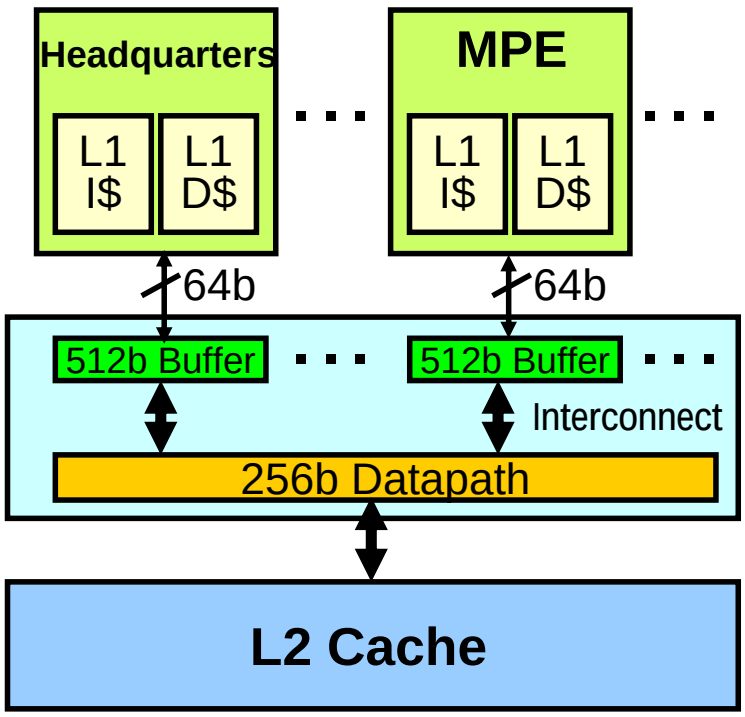


The instruction mode is switched by the special subroutine call instruction (bsrv).

IVC2 64-bit SIMD Datapath



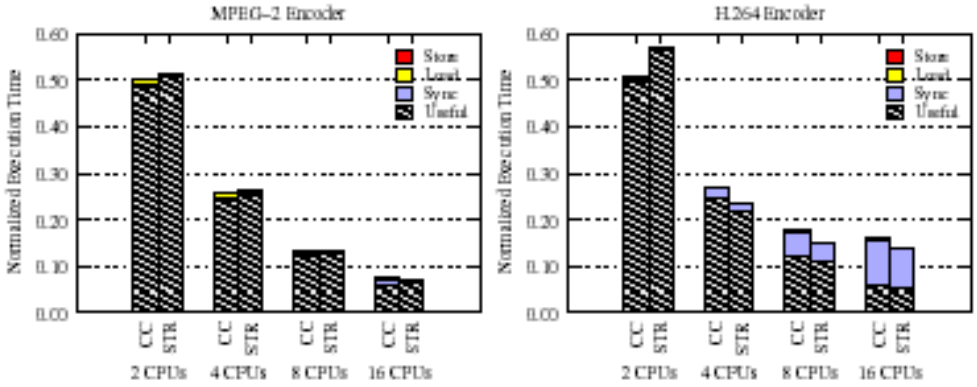
Cache Memory System



- **L1 Inst./Data Caches**
  - 2-way Set Assoc.
  - 8/16KB
  - 64B Line Size
- **L2 Cache**
  - 64/128/256/512KB
  - 4-way Set Assoc.
  - 256B Line Size
- **Prefetch Functions**
  - L1 I$ Auto Prefetch
  - L1 D$ Prefetch Inst.
  - L2 Interconnect Buffer
    - L2 $ → Buffer
  - L2 $ Prefetch
    - Main Mem. → L2 $

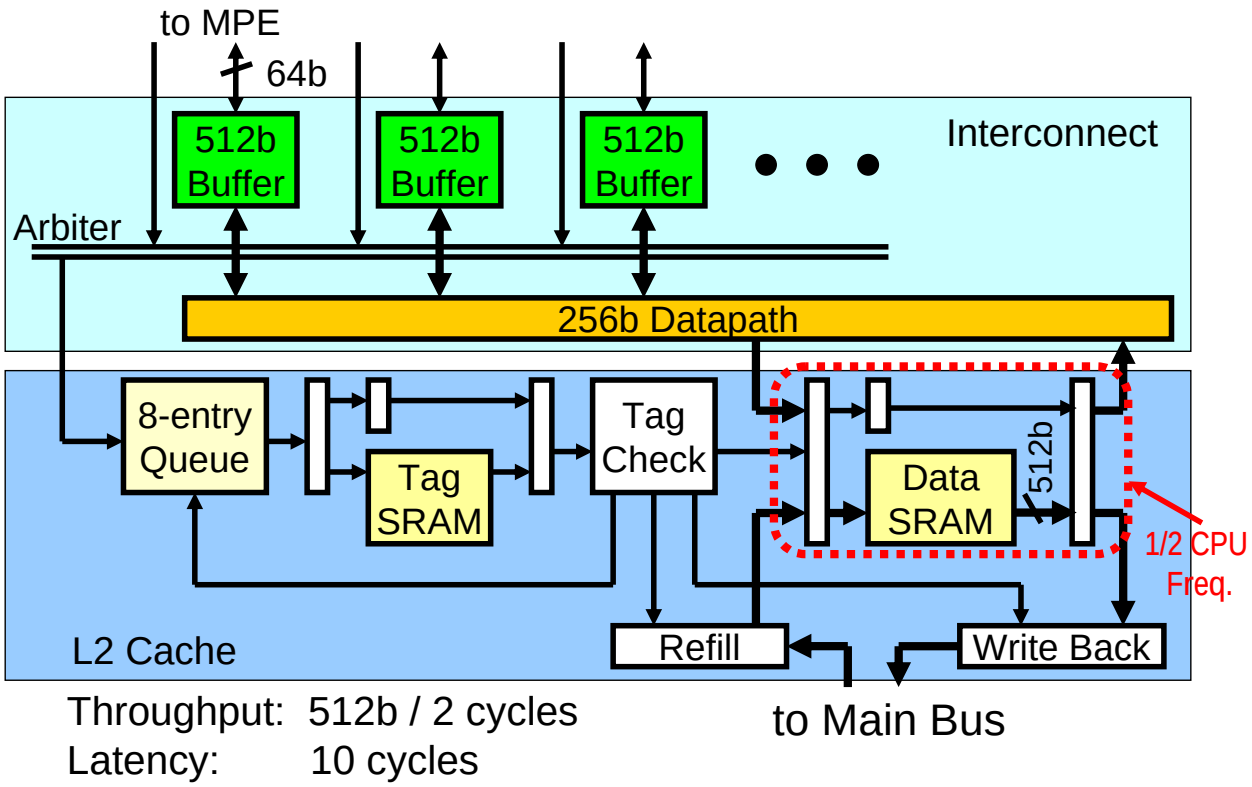
Comparing Memory Systems for Chip Multiprocessors
(Stanford Univ., ISCA'07)

- **CC: Cache-Coherent Model**
  - 32KB 2-way assoc. cache
- **STR: Streaming Model**
  - 24KB local memory and 8KB 2-way assoc. cache
- **512KB L2 cache**



- **Both models perform and scale equally well.**
- **"Non-allocate store to cache" can reduce memory traffic.**
  - Ex. Prepare for Store (PFS) instruction of MIPS32
  - MPEG-2: Traffic due to write miss was reduced 56%.
- **Streaming programming model, such as blocking and locality-aware scheduling, is efficient for cache model as well as streaming model.**

L2 Cache



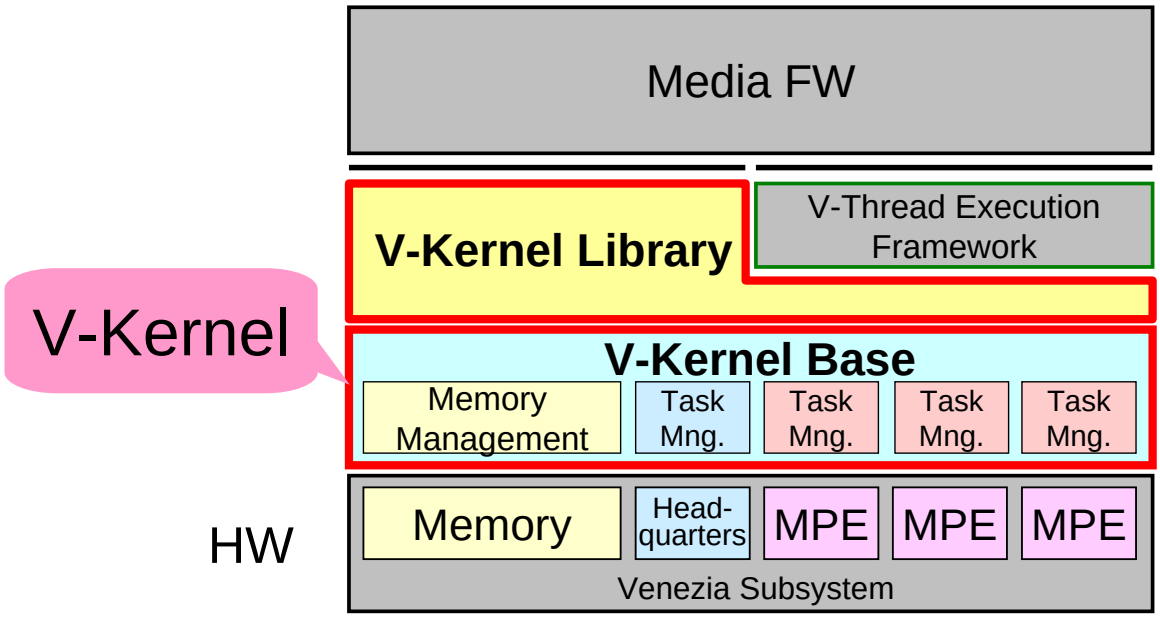
Throughput: 512b / 2 cycles
Latency: 10 cycles

Outline

- Background
- Venezia Hardware Architecture
- **Venezia Software Architecture**
- Evaluation Chip and Performance Results
- Summary

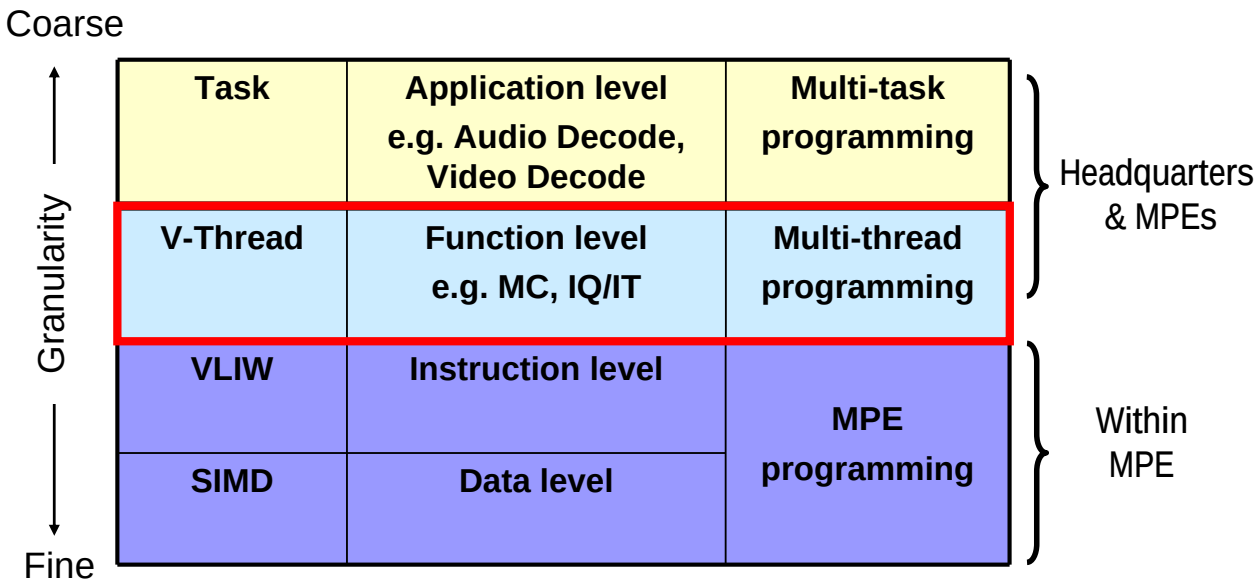
Software Hierarchy of Venezia

- **V-Kernel: simple and light operating system kernel**



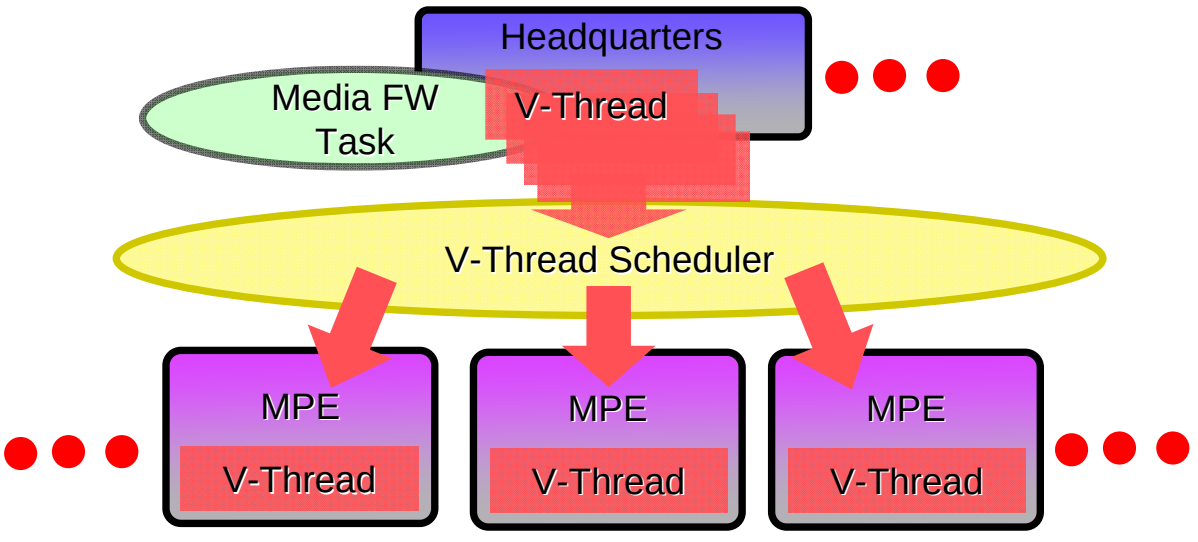
Multigrain Parallelism

- MPE Exploits Instruction / Data Level Parallelism
- Multicore Architecture Exploits Task and V-Thread Level Parallelism

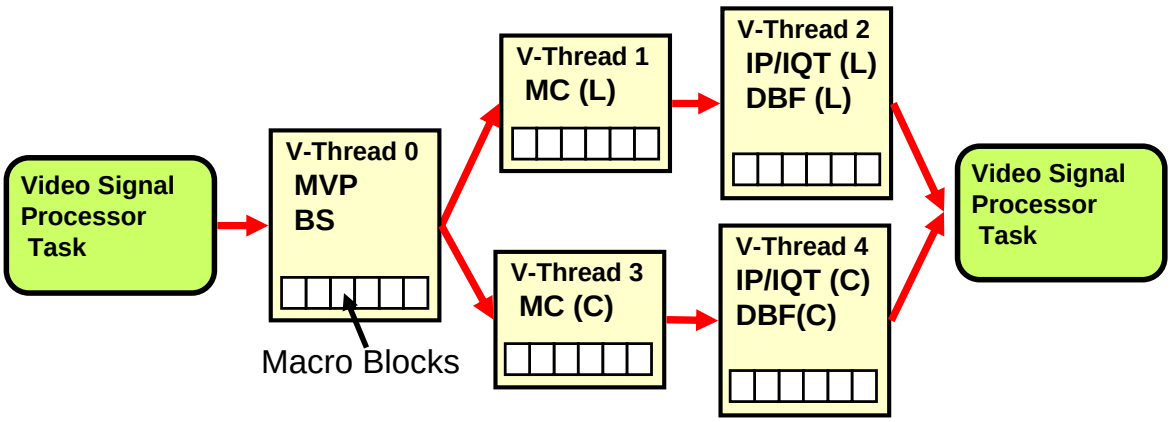


V-Thread Execution Model

- **Scalability and Compatibility by V-Thread**
  - Parallel Execution by MPEs
  - Abstraction of MPE Computing Resources



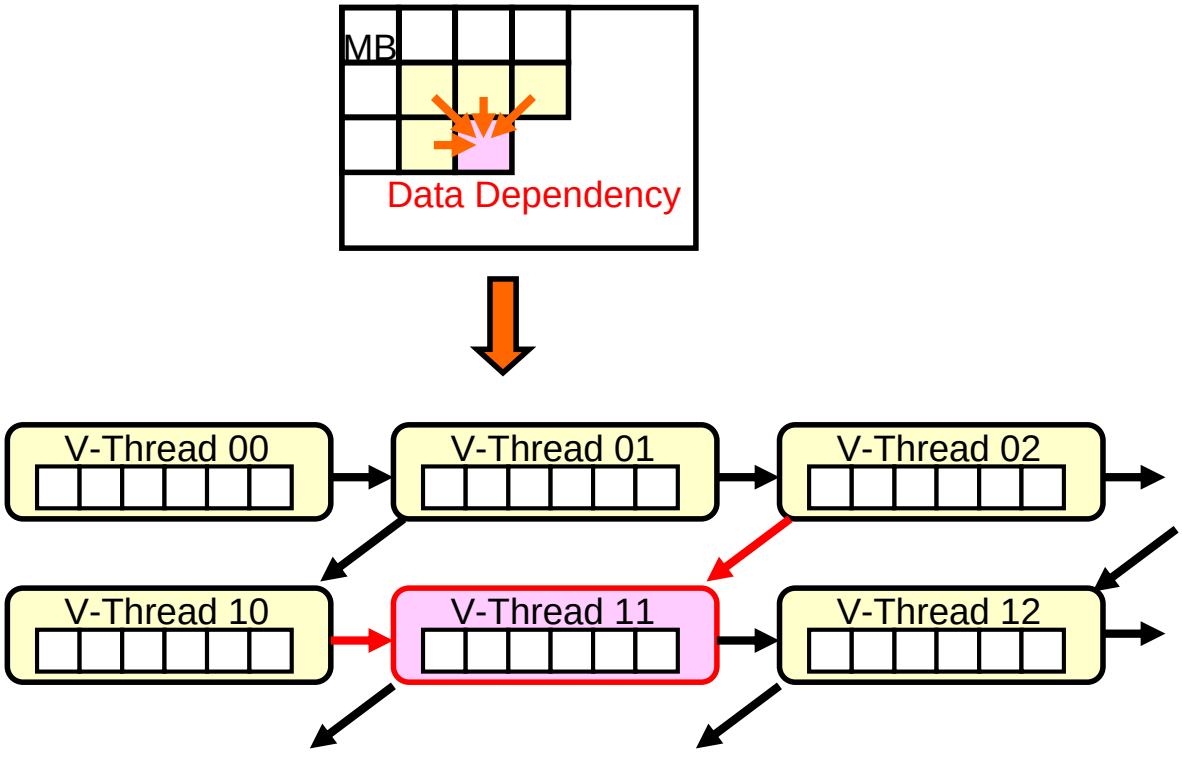
Granularity of V-Threads in H.264 Decode



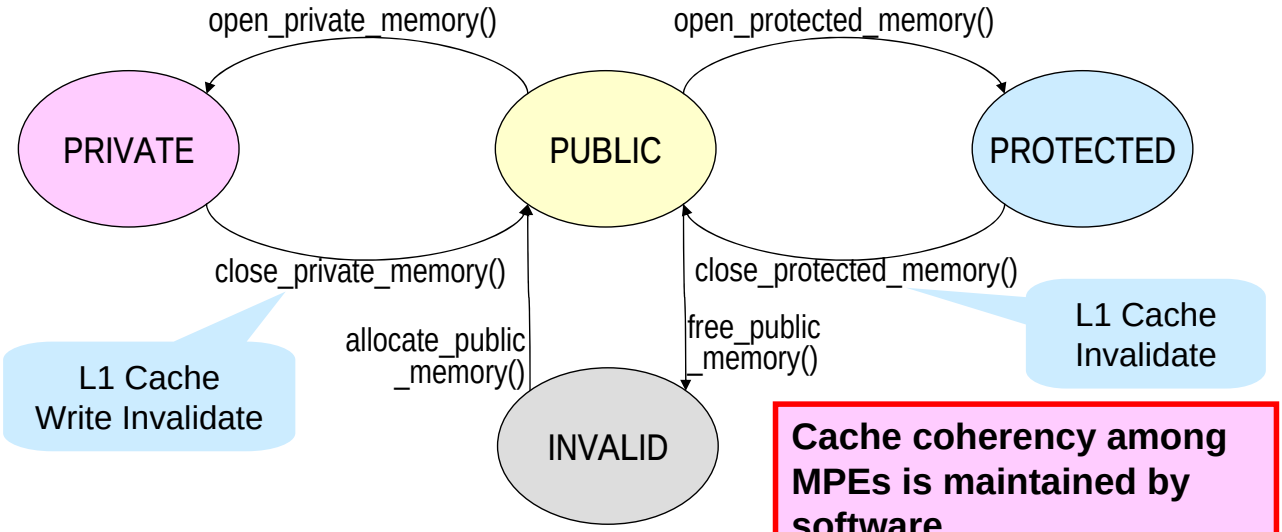
VGA: 1000 V-Threads / Frame
720p: 3000 V-Threads / Frame

MVP: Motion Vector Prediction
BS: Boundary Strength
MC(L): Motion Compensation (and Weighted Prediction) for Luma
MC(C): Motion Compensation (and Weighted Prediction) for Chroma
IP/IQT(L): Intra Prediction (and Inverse Quantization, Inverse Transform) for Luma
IP/IQT(C): Intra Prediction (and Inverse Quantization, Inverse Transform) for Chroma
DBF(L): De-Blocking Filter for Luma
DBF(C): De-Blocking Filter for Chroma
EoM : The end of the macroblock process

Spatial Dependency of V-Threads



V-Kernel Shared Memory Model

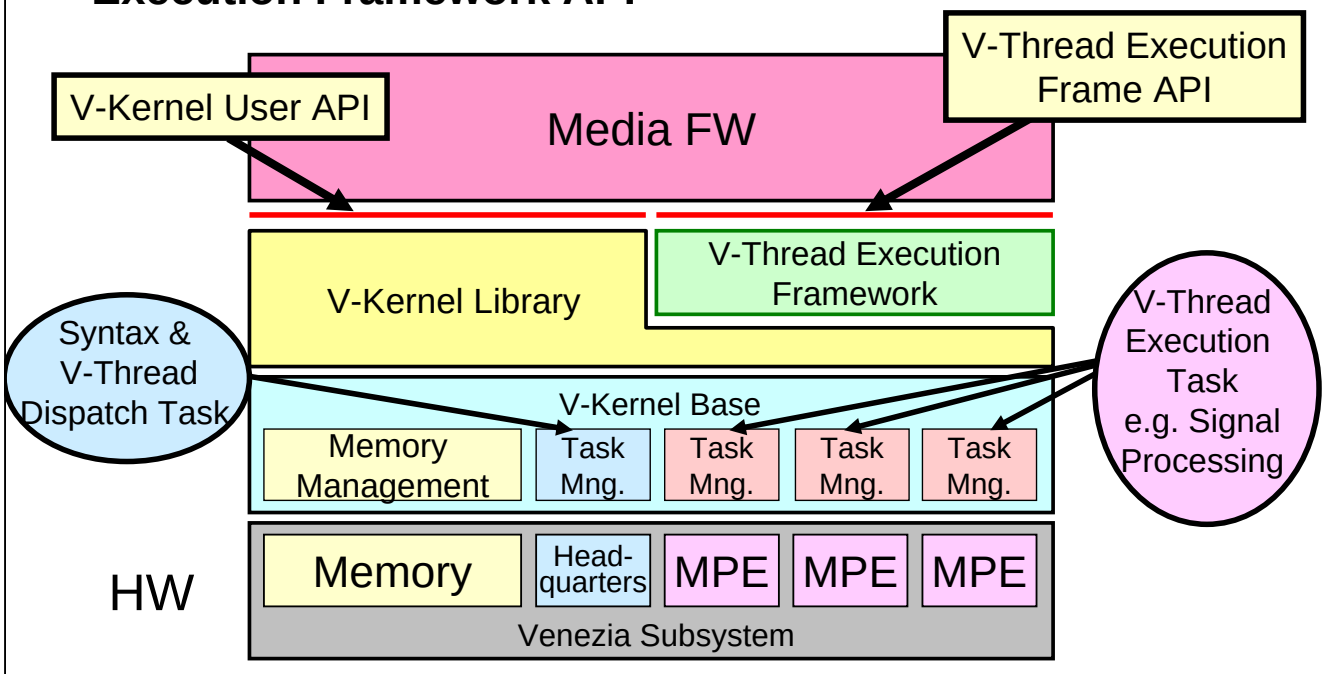


Cache coherency among MPEs is maintained by software.

	L1 $ read	L1 $ write	L2 Direct read	L2 Direct write
PUBLIC	NG	NG	OK	OK
PRIVATE	OK	OK	NG	NG
PROTECTED	OK	NG	(OK)	NG

V-Kernel and V-Thread Execution Framework

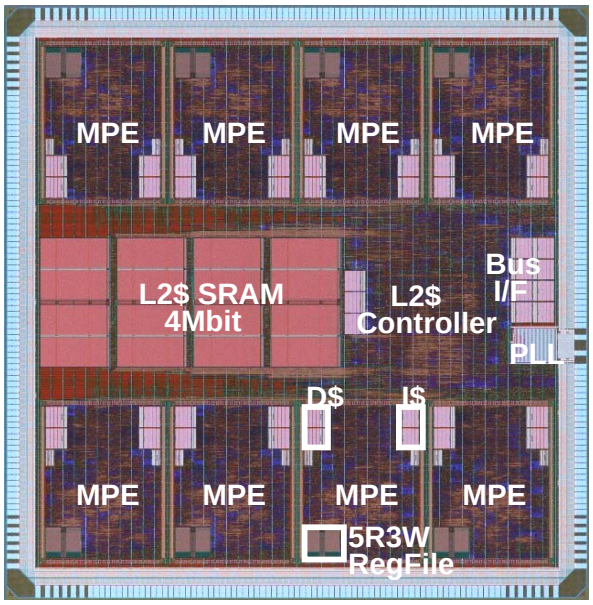
- **Media FW runs on V-Kernel User API and V-Thread Execution Framework API**



Outline

- Background
- Venezia Hardware Architecture
- Venezia Software Architecture
- **Evaluation Chip and Performance Results**
- Summary

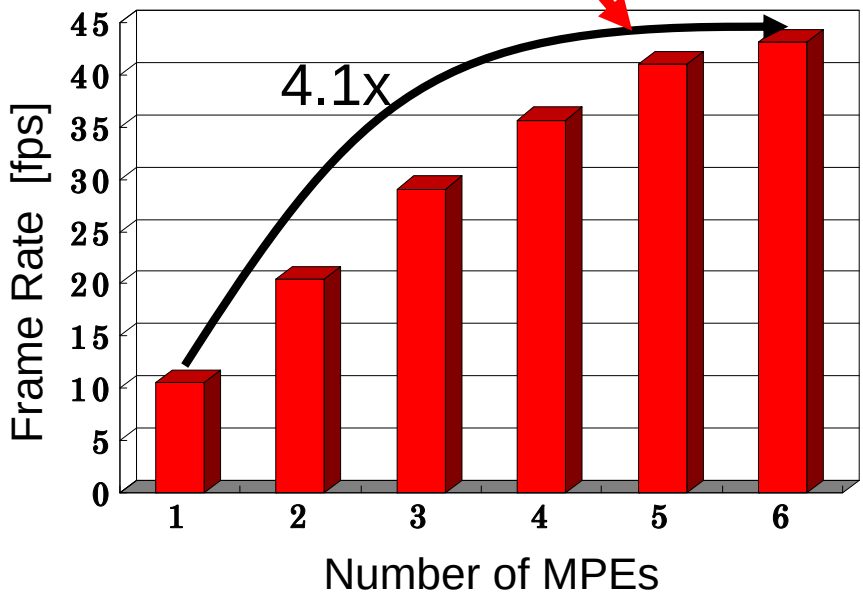
VeneziaEX: Evaluation Chip of Venezia Architecture



Technology	65nm CMOS, 8LM
Die Size	5.06mm x 5.06 mm
Frequency	333MHz (MPE, L2$ Logic) 166MHz (L2$ SRAM, Bus I/F)
Supply Voltage	2.5V (I/O) 1.2V (Core) 1.2V/0.95V/0V (SVC Output)
L1 Cache	8KB (Instruction), 8KB (Data) 2-way, FIFO, 64B Line
L2 Cache	512KB (unified), 4-way, LRU, 256B Line

Performance Scalability – H.264 720p Decode

Scalability Bottlenecks
- I-Picture: VLD
- P-Picture: L2 Cache Miss Penalty



Outline

- Background
- Venezia Hardware Architecture
- Venezia Software Architecture
- Evaluation Chip and Performance Results
- **Summary**

Summary

- **Venezia: Scalable Multicore Subsystem for Multimedia Applications**
  - RISC Cores (Headquarters), MPEs, and L2 Cache
  - MPE is 3-way VLIW Processor with SIMD Instructions
  - Cache Based Memory System to Realize Performance Scalability with SW Binary Compatibility
- **V-Kernel and V-Thread Execution Framework**
  - Focus on A/V CODECs
  - Video Task is Divided into a Large Number of Small V-Threads
  - Cache Coherency Is Maintained by Software
- **Performance Evaluation Results**
  - x4.1 Performance Improvement by 6 MPEs