PHILIPS

High Speed Low Cost Nexperia PNX1700 Super-Pipelined Media-Processor

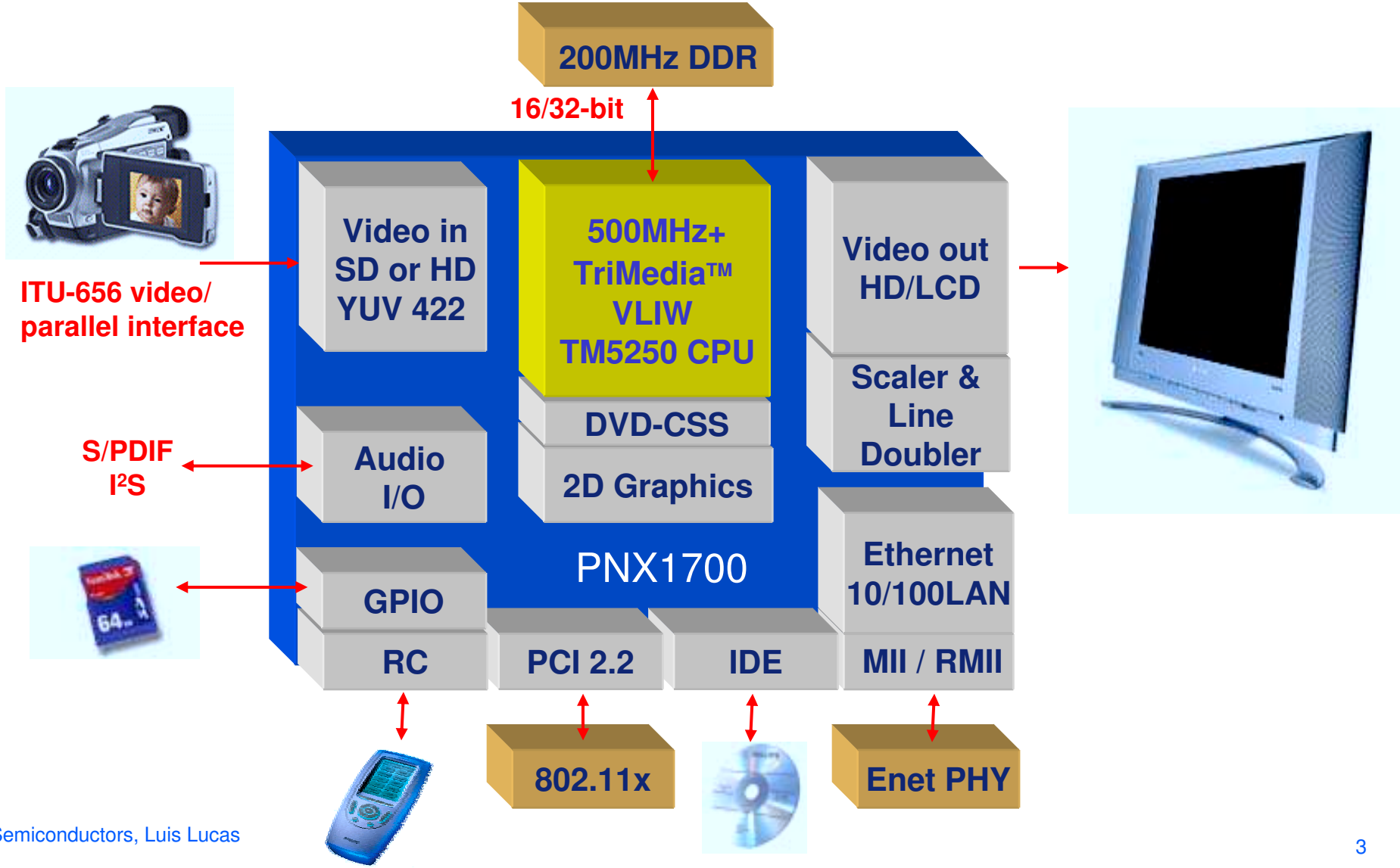
Luis Lucas
Philips Semiconductors

HOT CHIPS CONFERENCE, August 14-16, 2005

PNX1700 Markets

IP-STB
Digital media adapters
Videophones
Video security
Car infotainment
Video conferencing

PNX1700 System On Chip

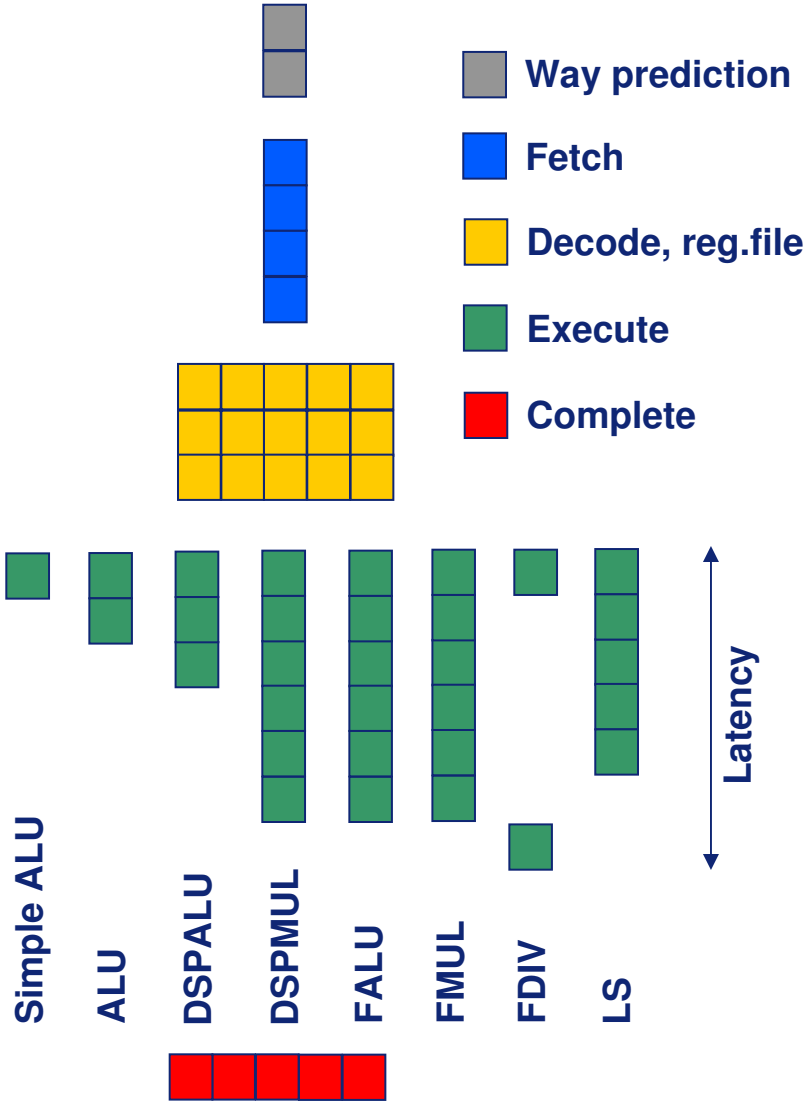


TM5250 Summary

Architecture	TriMedia DSPCPU32 source-level compatible Five-issue slot VLIW with 32-bit datapaths
ISA	All operations can be guarded (predicated) SIMD capabilities: 1 * 32-bit, 2 * 16-bit, 4 * 8-bit IEEE-754 floating point DSPCPU32 + additional MPEG4/H.264/WM9 pixel operations
Register file	128 x 32-bit unified register file 15 read ports, 5 write ports
Caches	64KB L1 instruction cache 16KB L1 data cache 128KB L2 data cache
Embedded peripherals	64 entry Vectored interrupt controller 8x Timers Instruction/data breakpoints
Clock frequency	450-500MHz

Deep Pipeline

- Depth
  - 2-cycle I$ way prediction
  - 4-cycle fetch
  - 3-cycle decode, access reg. file
  - Execute in 1–6 cycles
- Width
  - 31 execution units
  - 1 VLIW instruction contains up to 5 operations
- Branch unit
  - 3 branch operations per VLIW instruction
  - Branch prediction



TM CPU Architecture Levels

- Level 1: PNX1300 – TM32 CPU
  - Level 0
- Level 2: PNX1500 – TM3260 CPU
  - Level 1 + new operations (MPEG-4)
- Level 3: PNX1700 – TM5250 CPU
  - Level 2 + new operations (MPEG-4 / H.264)
  - + non-aligned load/store support
  - + Allocate-on-write cache miss policy
  - + memory region based pre-fetching

Functional Units Distribution

Functional unit	Slots	Latency/ recovery	Functional unit	Slots	Latency/ recovery
CONST	1,2,3,4	1/1	FALU	1,4	6/1
FAST_ALU	1,2,3,4	1/1	FTOUGH	2	17/16
ALU	1,4,5	2/1	FCOMP	3	2/1
SHIFTER	1,4,5	2/1	BRANCH	2,3,4	8/1
DSPALU	1,3,4	3/1	DMEM	5	5/1
DSPMUL	2,3	6/1	DMEMSPEC	5	5/1
IFMUL	2,3	6/1			

Level 3 Operations

- Provides ‘missing’ dual operations that are useful in vector processing and 16-bit pixel processing

Operation	Functional Unit	Slots	Latency/recovery
allocd_set	dmem_spec	5	-/1
clsame	dspalu	1,3,4	3/1
dualaddsub	dspalu	1,3,4	3/1
dualiadd	dspalu	1,3,4	3/1
dualisub	dspalu	1,3,4	3/1
dualimax	dspalu	1,3,4	3/1
dualimin	dspalu	1,3,4	3/1
duallsr	alu	1,2,3,4	2/1
swapbits	alu	1,2,3,4	2/1

Signed Dual 16-bit Median

- Median filter: the value in the middle

In ISA Level 2:

```
PACK16LSB(
      MEDIAN(SEX16(A), SEX16(B), SEX16(C)),
      MEDIAN(A>>16, B>>16, C>>16) )

With MEDIAN(X,Y,Z) defined as
      IMIN(
         IMAX(
            IMIN(X, Y),
            Z),
         IMAX(X,Y))
```

- 15 issue slots used
- 1+1+2+2+2 cycle latency

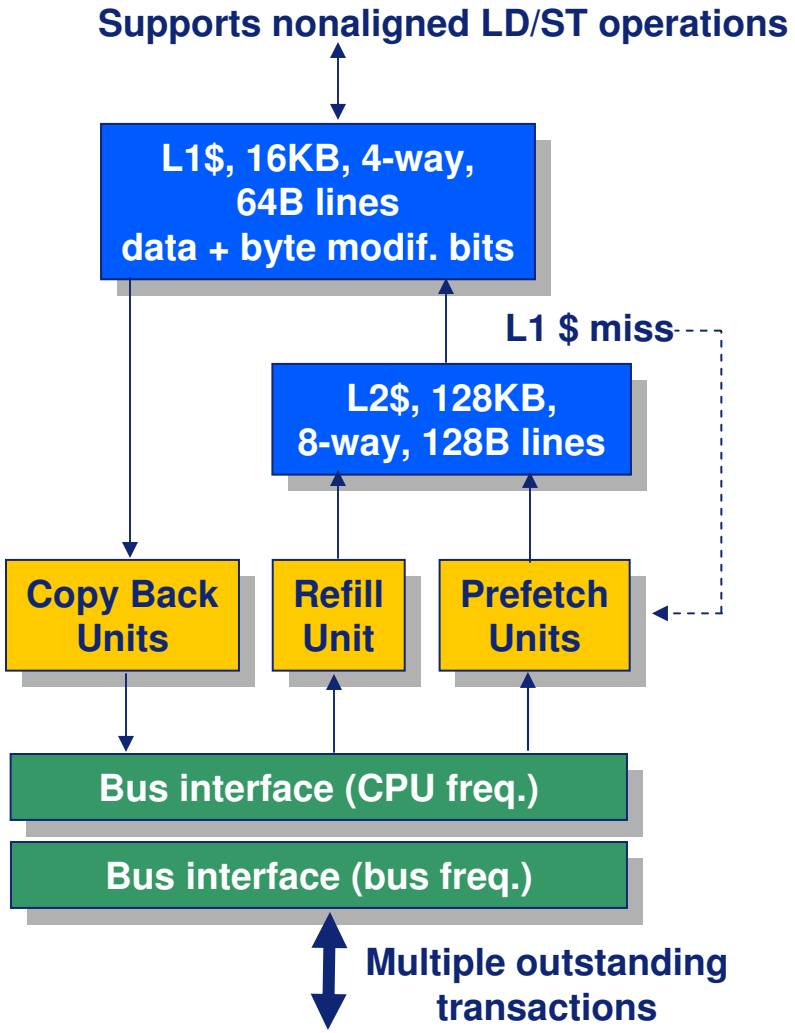
In ISA Level 3:

```
DUALIMIN(
      DUALIMAX(
         DUALIMIN(A,B),
         C),
      DUALIMAX(A, B)))
```

- 4 issue slots used
- 2+2+2 cycle latency

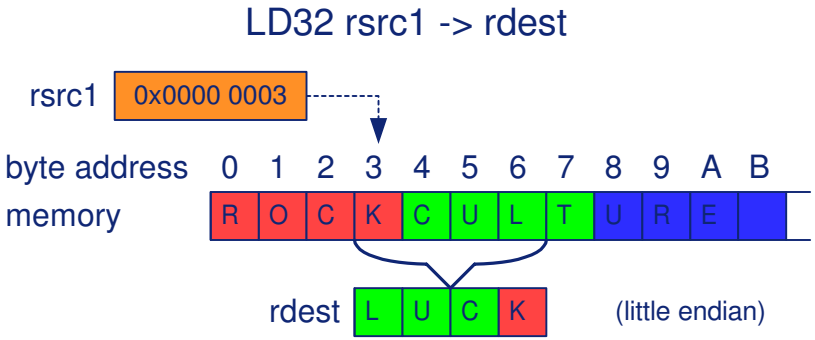
Data Pipe - Overview

- Patented L1 cache SRAM partitioning
  - Supports nonaligned LD/ST (no cycle penalty)
  - 4-way set associativity
  - Single-cycle L1 cache line refill/copy back (low cache-miss penalty)
- L1 cache keeps track of byte modifications
  - Allocate on L1 cache write-miss (no fetch!)
  - Copy back modified bytes only
- Asynchronous bus interface unit
- Multiple outstanding bus transactions



Unaligned Load/Store Operations

- Load 4 byte-sized pixels from non-aligned address A



In ISA Level 2:

```
FUNSHIFT3 (
    LD32D(0) A,
    LD32D(4) A)
```

- 3 issue slots used
- 4+1 cycle latency

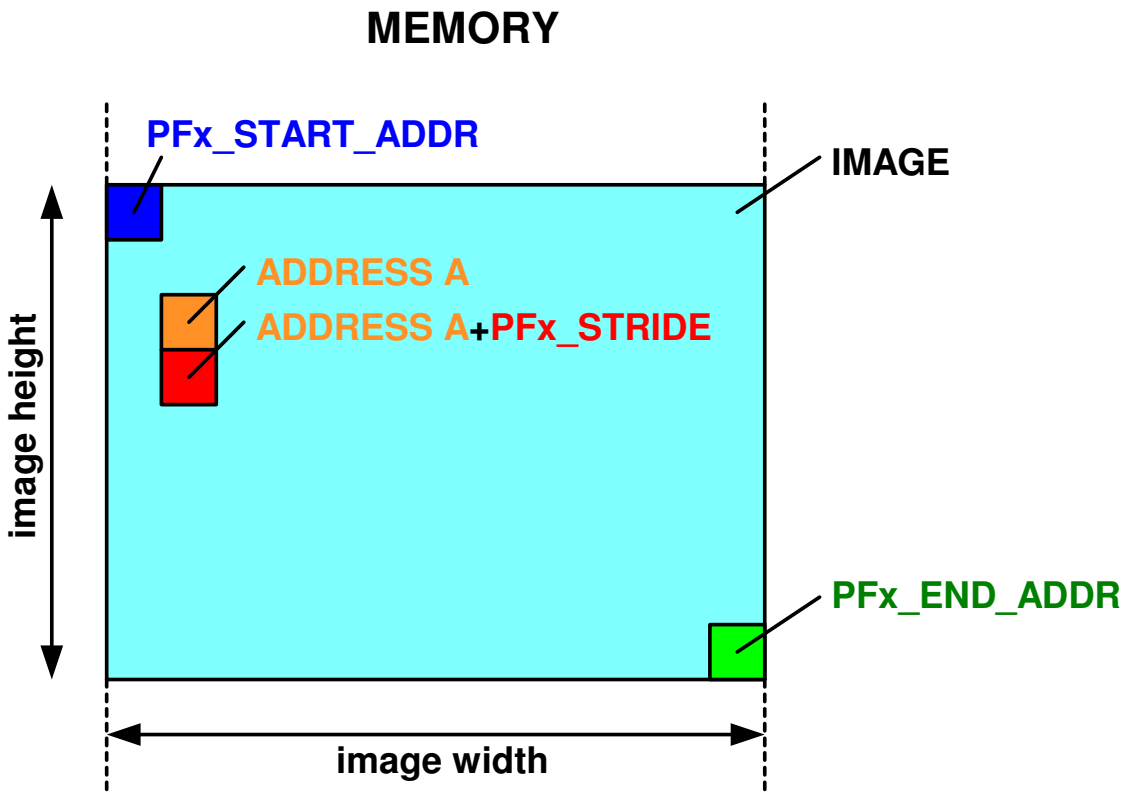
In ISA Level 3 and 4:

```
LD32D(0) A
```

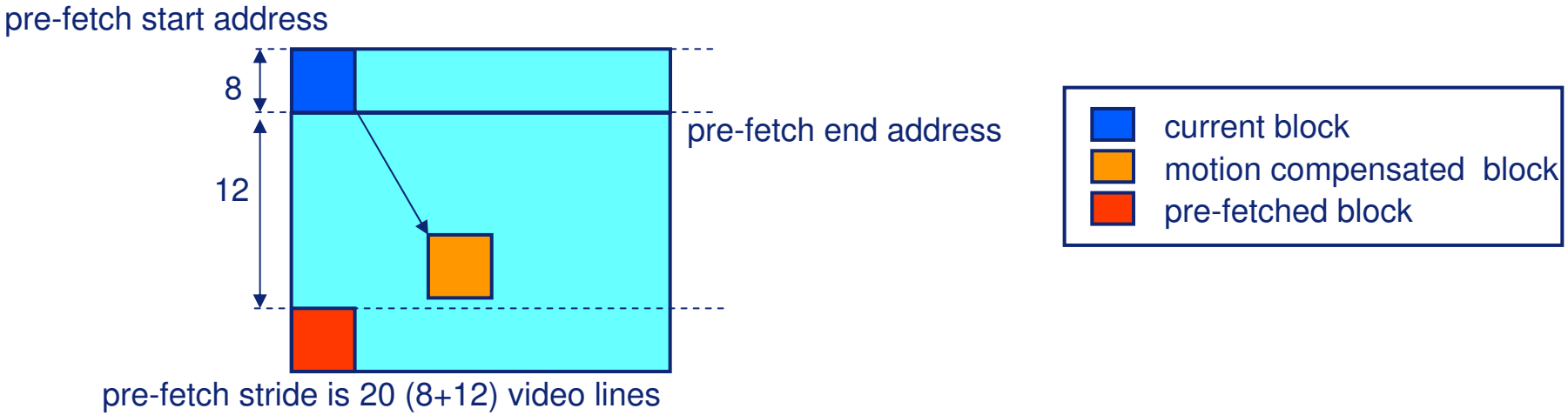
- 1 issue slots used
- 5 cycle latency

Hardware Prefetch

PRE-FETCH REGION x:
PFx_START_ADDR
PFx_END_ADDR
PFx_STRIDE



Prefetching Scheme for SD UPC



- Prefetching
  - with offset 20*stride for both frames
    - 8 lines being processed + 12 lines of maximum vertical vector range
  - start and end address are set to the slice of 8 lines that is being processed
    - to prevent prefetches on vector loads

SD UPC Performance Numbers

- 109 MHz load
  - 78 MHz instructions
  - 31 MHz stalls (29%)
    - **Same algorithm on TM3260 (PNX1500) has > 50% stalls !**

$$\text{Prefetch Efficiency} = 1 - \frac{\text{Read Miss Bandwidth}}{\text{Read Bandwidth}}$$

$$1 - \frac{\text{D\$ rd miss} * 128}{\text{Frame data} + \text{Vector data}} = 1 - \frac{1422 * 128}{2 * 720 * 480 + 4 * 720 * 480/64} = 0.74$$

TCS 4.6 Tools for TM5250/PNX1700

- Production Quality C/C++ Compiler, Instruction Scheduler, Assembler, Linker/Loader, C/C++ libraries, functional and cycle-true simulators, source level debugger, pSOS RTOS, performance analysis tools etc
- TCS 4.6 specifically optimized for TM5250 core, viz
  1. Deeper pipeline: longer load and jump latencies than previous cores
  2. Single load/store unit
  3. Unaligned load/store support

Compiler/Scheduler Optimizations

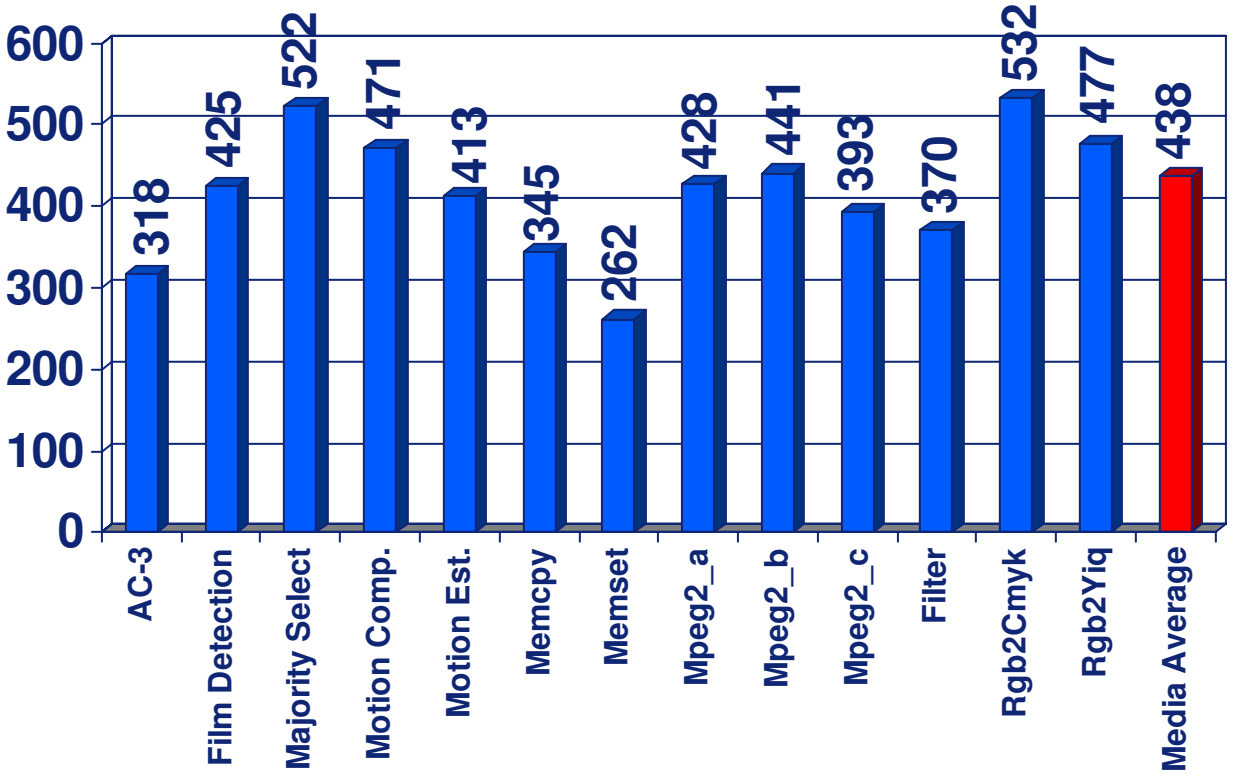
- Selective coalescing of adjacent 8/16 bit loads/stores
- Improved alias analysis to disambiguate integer vs pointer references
- Aggressive if-conversion, using guarded (predicated) operations to compute true/false branches simultaneously and eliminate jumps
- Grafting - replicate code blocks that follow a conditional and attach copies of original blocks both branches of a conditional, increasing opportunities for optimizations and ILP and reducing jumps
- New options to control amount (max increase in number of operations) and scope of grafting (statement, function, file scope)
  - Allows experimentation and trade off between code size increase and performance improvement

TCS 4.6 Compiler performance

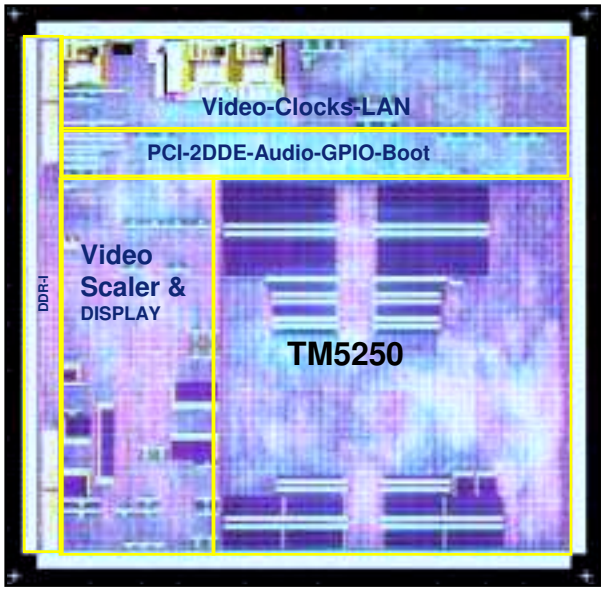
	TCS 4.51, -O3/-O5 -graft cycles	TCS 4.6, -O5 -graft cycles	Improvement %
Motion Estimation	2470554	2211024	**11.74**
unopt CABAC	9700313	6932530	**39.92**
UPC 8	1533750	1458558	**5.16**
UPC 10	2940622	2613998	**12.50**
Mediastone	15564520	14101820	**10.37**
Total / Avg	**32209759**	**27317930**	**17.91**
Code Size	9150208	9477884	**3.46**

TCS 4.6 Compilation uses different grafting parameters for some source modules to minimize code size increase

PNX1700 Performance



PNX1700 Silicon



- 56 mm^2
- 6 Metal Layer in CMOS 0.12 µm
- < 29 Million Transistors
- 3 Power Supplies:
  - 1.2/1.3 V core
  - 2.5/2.6V DDR-I
  - 3.3V I/Os (5V Tolerant)
- 3 W Max for Typical Application
- BGA456 (275 Functional I/Os)
- 500 MHz TM5250 at 1.3 V and Supports DDR400
- More Than 30 Different Clock Domains

Conclusion

- Data Cache Features Greatly Reduce Cycles Needed per Application:
  - Hardware Prefetching
  - Allocate-on-write cache miss policy
- Fine Sharing of the Available Bandwidth Between the Different Agents of the System On Chip.
- At Low Cost and Low Power


PHILIPS