

## White Paper

# **Telairity-1: A New Processor Architecture for High-Definition Video**

Telairity Semiconductor • 3375 Scott Blvd., Suite 300  
Santa Clara, CA 95054  
tel 408 764 0270 • fax 408 764 0271

[www.telairity.com](http://www.telairity.com)

## Challenges of handling HD video compression in silicon

The semiconductor industry has seen the development of numerous processors for digital entertainment applications over the last decade. In the late 1990s, as digital TV and DVD applications proliferated, vendors such as Philips (TriMedia), Motorola (StarCore), Equator, and BOPS developed multimedia processors for audio and video processing. General-purpose CPU vendors for CISC and RISC processors developed SIMD extensions to address multimedia processing. Broadcast and consumer equipment manufacturers started making the shift to standard-definition digital TV, and MPEG-2 became the video compression codec of choice.

Now the digital TV market is making a transition to high-definition (HD) video, and optical storage is moving to higher capacity with HD-DVD and Blu Ray DVD discs and players. The rapidly growing market for all things HD is creating demand for purpose-built HD processors and SoCs that can handle the staggering data-crunching requirements of HD resolutions and address new video compression standards such as VC1 and AVC.

This is the context in which Telairity Semiconductor has come to market with a new processor architecture specifically designed to handle the toughest HD video applications. Unburdened by any legacy architecture, Telairity Semiconductor started with a clean slate to design an optimal solution for HD encoding. Quickly, however, we realized that the best architecture for the job would need to meet some very specific and stringent requirements.

First, it had to be fully programmable, as this was the only way of keeping up with future refinements and extensions to the H.264 standard without having to redesign the hardware from scratch each time. We would thus need to enable OEMs to build on their existing software architectures and leverage their proprietary R&D and intellectual property. Second, we knew we needed to deliver the highest feasible multimedia performance in a single chip. Telairity looked at various DSP architectures and decided that a vector DSP architecture married well with the new generation of video codecs—VC1 and H.264—and would be particularly good for the intensive motion estimation computations required for HD video resolutions. Third, the design must be cost-effective.

Leveraging the considerable processor design expertise of Telairity’s founders, we opted for a full custom design methodology to achieve high clock frequency and silicon gate density. The design was targeted for Fujitsu’s 90-nm process technology.

Finally, we realized the design had to be completed in a short period of time to meet market windows for HD broadcasting. Therefore, Telairity decided on a multicore architecture—the quickest way to complete the design without compromising on overall performance.

### Building blocks of a multicore chip architecture for HD video

The result of this effort is the T1P2000, a video processor with five independent, identical, and fully programmable vector/scalar cores in a single chip. The T1P2000 processor operates at 668.25 MHz, running synchronously at a 9X multiple to the input video clock. The five loosely coupled processor cores

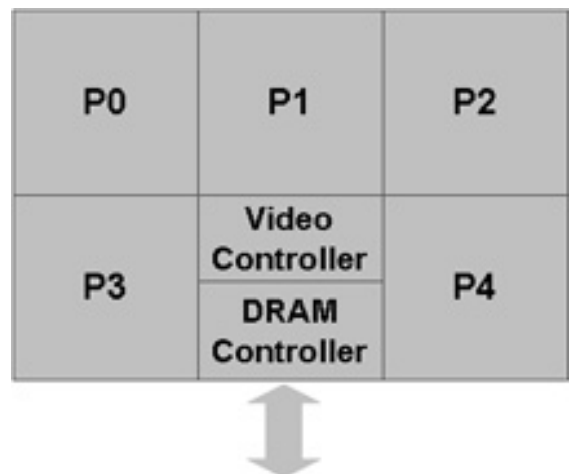


Figure 1

interact using interrupts and mailboxes to synchronize and pass messages, and share a large common DRAM space with 5.3 Gbps of memory bandwidth. The processor cores receive HD video data through a video controller that accepts a 20-bit parallel video input. Processed (compressed) video data can be output via a 5-bit SPI port. See Figure 1 below for the T1P2000 video processor block diagram.

### The TVP400 Processor Core

The fundamental architecture building block of the T1P2000, the TVP400 processor core, achieves maximum parallel performance by providing support for both long and short vectors. Vector instructions are particularly useful in image processing and video because these tasks involve performing the same operation on multiple data elements in an array. Short (or horizontal) vectors support single-instruction multiple data (SIMD), where operations are performed on four or eight data elements at the same time. Long (or vertical) vectors perform sequential operations on the elements (typically 32 or 64) in a vector.

Each TVP400 core is an SIMD machine with one 32-bit scalar pipeline and four 16-bit vector pipelines. A single instruction is dispatched to one or more of four execution pipelines. Each pipeline has its own register file structure, coupled to a function unit stack with separate load and store interfaces to the memory hierarchy. The memory hierarchy consists of 128 kbytes of vector SRAM, 4 kbytes of vector SRAM data cache, 8 kbytes of scratchpad memory and a 32-kbyte instruction cache. The available SRAM memory bandwidth is 32 GBps. The vector registers are designed to support 32 GBps to the functional units as well as 16 GBps to the SRAM for loads and stores. The external high-speed DDR2 DRAM memory interface supports up to 5.3-GBps to the SRAM via a DMA engine.

The Telairity-1 instruction set is a three-address architecture with two source register addresses and one destination register address. Many addressing modes are available for load and store instructions. The arithmetic operations support signed, unsigned, and saturating arithmetic. The scalar pipe supports byte, double, and quadlet data types while the vector pipes support byte and doublet data. The vector instructions have other programmable attributes such as starting vector register element and vector length. Vector load and stores have additional addressing attributes such as stride, repeat, and skip, which allows for complex address stream calculations. These instruction addressing features are a powerful engine for manipulating macro block structures in memory. Figure 2 shows the architecture of the TVP400 vector core.

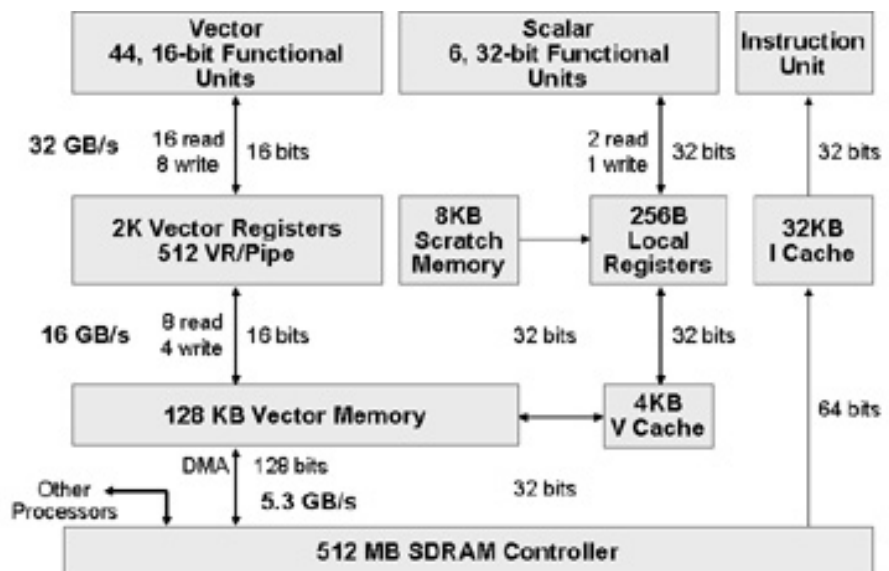


Figure 2

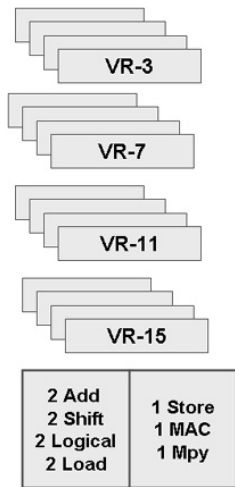


Figure 3

### Vector Unit

The vector unit has four identical but independent vector pipes, each of which has 16 vector registers. Each vector register consists of 32 16-bit elements for a total of 512 16-bit registers. The function units use 16 read ports for fetching operands and 8 write ports for committing results. Each vector unit or pipe can perform four loads, two stores, and two arithmetic operations on vectors with a vector length from 1 to 32. The function units contain two adders, one MAC, two shifters, and one logical unit. The adders support 24-bit accumulators and the MAC supports a 40-bit accumulator. Score-boarding of the register file is done to the element level, allowing for chaining and backfilling of register element values. Figure 3 shows a single vector pipe in the TVP400 core.

### Scalar Unit

The scalar unit contains the following features:

- Two register files, one for the supervisor and the other for the application program
- 32 of the company’s 32-bit registers per register file
- Each register file has two read ports and one write port
- A three-address architecture in which all scalar instructions issue in order and commit in order

The scalar unit has three memories as illustrated in Figure 4. The first is a 32-kbyte instruction cache that is fully associative with a 2-kbyte line size. Cache lines can be pre-fetched, locked, and invalidated under instruction control. The instruction cache gets instructions from the external DDR2 DRAM memory. A 4-kbyte vector data cache is organized as a four-way set-associative cache with a 16-byte line size, and sits between the scalar pipe and the vector SRAM. This cache has a write-through strategy to maintain consistency with the SRAM. In addition, all writes to the SRAM invalidate lines in the cache with a five-ported tag memory. The third element of the scalar unit is an 8-kbyte scratchpad memory used for storing local variables and implementing a system stack.

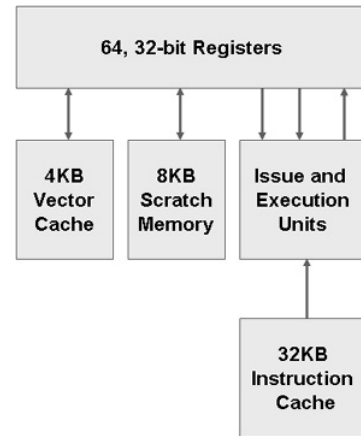


Figure 4

### SRAM

The flexibility of the vector SRAM is the key to achieving high performance in Telairity’s TVP400 vector core. Each TVP400 core has an SRAM which is a 12-ported, 128-kbyte memory structure. Eight 16-bit reads and four 16-bit writes can be initiated every clock cycle. The memory is built of 128 1-kbyte memory blocks with only a single read/write port. The banks are connected in a four-stage pipeline with 32 banks at each stage in the pipeline. The 12 ports are evaluated for bank hits and conflicts at each bank. A retry mechanism is in place to handle access collisions. Colliding

accesses continually re-circulate in this pipeline until retired. The load latency from the function unit to a write in the vector register file is seven clocks. Because of a separate 256-byte data shift register, DMA accesses steal the SRAM memory system for only one clock in order to efficiently read/write 256 bytes at a time. The scalar unit's 4-kbyte data cache has the lowest access priority into this SRAM memory-access hierarchy. It utilizes inactive access ports in order to write through to the vector SRAM and bring in a 16-byte line on a cache miss.

### I/O Controller

The video controller includes a 20-bit parallel input video channel as well as a 20-bit parallel video output channel that can be used for reconstructed video output or as a high-bandwidth data communication channel in multi-chip video applications. The input and output parallel channels support HD data at 74.25 MHz. The video controller hardware is set up to detect the SAV and EAV headers in the input HD video data. The video controller also includes five serial peripheral interface (SPI) channels for simple data communications between chips. There are three master and two slave SPI channels, and one of the master channels can be used to output compressed video data at either 74.25 MHz or 148.5 MHz. The parallel and serial channels assert interrupts to the processors based upon payload completions. Last but not least, the video controller also includes a bit packing unit with dedicated hardware that is used to efficiently process compressed output data.

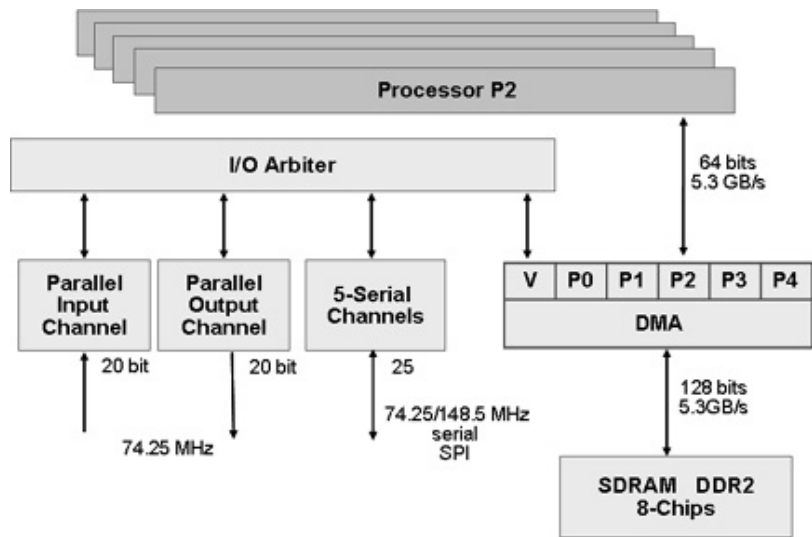


Figure 5

Each core processor has a DMA channel to the DDR2 memory space with access provided via a centralized DMA arbiter as shown in Figure 5. Programmable payloads in 256-byte increments up to but not including 64 kbytes are allowed. The 128-bit DDR2 interface operates in a burst of 8-mode-only and 256-byte-only boundaries

### Putting it all together for real-time performance

The Telairity-1 architecture is supported by a complete set of software development tools, including a compiler, vectorizer, assembler, linker, and debugger. Programs are written in C with intrinsics and can be developed on Telairity's T1P2000 EP Evaluation platform.

The TVP400 core can perform the following 21 operations simultaneously and within one clock:

- Eight 16-bit vector operations
- One 32-bit scalar operation

- Eight vector loads
- Four vector stores

With five processor cores per chip, this translates to a potential of 105 16-bit operations per cycle and a peak of 70 GOPS/s per chip, versus a sustained rate of 55.5 GOPS/s operating at 668.25 MHz.

## How Telairity-1 improves on previous multimedia processors

Notably, the Telairity vector architecture has a very powerful and well balanced memory system which avoids the typical problems encountered by cache-based processors. Many multimedia processors focus on achieving parallelism, but simply do not have the memory bandwidth to support it and mislead users by quoting only peak performance numbers that assume all function units are busy all the time (which is usually not the case).

The Telairity-1 architecture features an intelligent DMA controller that efficiently packs data from the external DRAM into the internal vector SRAM. The DMA controller has high-access bandwidth to the DRAM of 5.3GBpsec and the bandwidth between the vector SRAM and the registers is 16GBps. The multiport SRAM structure with eight reads and four writes per cycle can keep the vector function units humming all the time. Further, Telairity-1 avoids the issue of cache misses, which not only degrade performance significantly when dealing with sparse matrices, but also burden the processor with the operation of aligning data in the appropriate cache line boundaries. In Telairity's vector architecture, typical sustained performance is based on the issue rate of vector load instructions. The T1P2000 can perform a total of 666 operations in a 40-clock sequence, resulting in sustained chip performance of 55.5 GOPS/s—the highest performance for HD image processing in a single chip.

## Benefits for system designers: maximum flexibility and quality with fewer components

The multicore Telairity-1 high-definition, multicore video processor architecture offers broadcast equipment designers the greatest platform flexibility for multiple video encoding applications and for future standards migration while achieving the highest quality video encoding using state-of-the-art H.264 video compression. Fully programmable, it can be used for multiple HD video applications, such as encoding, decoding, and transcoding.

Another significant advantage relative to standard DSPs is that the Telairity-1 architecture produces very compact code. One instruction is capable of firing off a maximum of 128 operations—4x32—with four vector pipes per core and a vector length of 32. The video-specific instruction set, moreover, translates into efficient macroblock processing. A sophisticated multiport SRAM memory addresses video data efficiently. Unlike cached architectures, there are no penalties for non contiguous memory accesses. Compact scatter/gather operations pick all needed image data into a single memory load operation.

Implemented in 90-nm technology using Telairity's proprietary design methodology, the T1P2000 video processor achieves high silicon density of over 150,000 gates/mm<sup>2</sup>. The device's high level of integration, using a multicore architecture, minimizes the number of chips required for the application and also reduces the latency and bandwidth required for interchip communications, as all five cores share high-speed access to shared DRAM memory to maximize efficiency for video image processing.