The World's First Commercially-Available Stream Processor: Architecture, Algorithms and Benchmark Results

Simon McIntosh-Smith
ClearSpeed Technology, Ltd.
Email Address: simon@clearspeed.com

Ron Bell
AWE - Aldermaston
Email Address: ron.bell@awe.co.uk

Abstract:

This briefing describes new applications for ClearSpeed's CS301 device, the first commercially available stream processor. Launched in October 2003, the CS301 is an ultra-high performance next-generation Single-Instruction/Multiple-Data (SIMD) stream processor, delivering 25 GFLOPS and 12.8 GMACS at 1.8 Watts. The CS301's low-power, Multi-Threaded Array Processor (MTAP) architecture scales to hundreds and ultimately thousands of processing elements, each with both floating point and integer hardware, capable of data parallel processing on image and signal processing applications as well as for compression, encryption, search, and general sensor processing applications. The processor is supported by a flexible development environment, including assembly language and C-based language support, as well as a cycle accurate simulator, with plans to develop industry standard API Libraries such as L3 BLAS and FFTW. This new class of stream processor has been shown to provide ten to one hundred times the overall performance of PowerPC or Pentium-based architectures, especially when performing image and signal processing functions, such as FFTs or filters. In general, the architecture has been shown to provide significant throughput, size, and power advantages for embedded processing applications.

AWE Aldermaston has been investigating potential uses for CS301-class processors in its key algorithms and applications. AWE further optimised fast math library routines on the CS301 for SGEMM – a single precision floating point matrix multiply, verifying the CS301's record-breaking math performance. AWE took matrix mutiply from 5 GFLOPS sustained to over 12 GFLOPS sustained on a single CS301. AWE is performing ongoing work exploring the acceleration potential of the CS301 for several in-house and 3rd party scientific codes, such as DL-POLY.

CS301-based accelerator boards having been shipping since January 2003 and multiple algorithms have been ported, with more underway. A dual-processor PCI-based development card is available from ClearSpeed, providing a total of 50 GFLOPS of compute performance, for a total maximum power consumption for the board of 10 Watts. Single systems containing up to 5 boards have been demonstrated for a total compute of 500 GFLOPS and capable of 1 Million FFTs per second (1K complex single precision floating-point). This level of compute density has never before been commercially available, with the CS301 delivering more than 10 GFLOPS per Watt.

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 FEB 2005	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
The Worlds First Commercially-Available Stream Processor: Architecture, Algorithms and Benchmark Results	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **ClearSpeed Technology, Ltd.**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM00001742, HPEC-7 Volume 1, Proceedings of the Eighth Annual High Performance Embedded Computing (HPEC) Workshops, 28-30 September 2004 Volume 1., The original document contains color images.

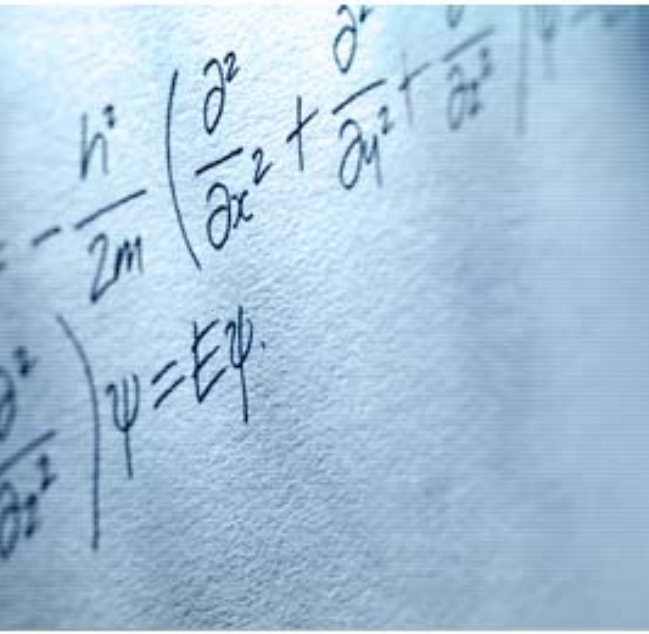
14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**15**	

In the first half of this briefing ClearSpeed will present performance results from the numerous algorithms and applications that have or are being ported to the CS301 stream processor. These include numerous sizes of FFTs and FIR filters which efficiently utilize the architecture and floating point per PE hardware to gain exceptional performance at very low power dissipation levels. We will include an update on improvements to work previously announced jointly with Lockheed Martin at HPEC03 on pulse compressions for radar (FFT – Complex Multiply – IFFT). The results to be reported are significantly higher than other industry standard processing and DSP platforms. New work on other transforms, such as DCTs, will also be presented. In addition, results from work to develop a Level 3 BLAS (Basic Linear Algebra Subprograms) library will be reported, including performance of certain vector and matrix operations, such as matrix multiplication and matrix inversion, including descriptions of the algorithms required on this high-performance, highly parallel architecture.

In the second half of this briefing AWE Aldermaston will present its work to benchmark the CS301. The briefing will include descriptions of optimizations to a fast matrix multiply algorithm for the MTAP streaming architecture, improving performance from 5 GFLOPS to over 12 GFLOPS on the CS301. AWE will also describe its investigations into using the CS301 to accelerate certain applications used in-house, such as the materials science code DL-POLY.

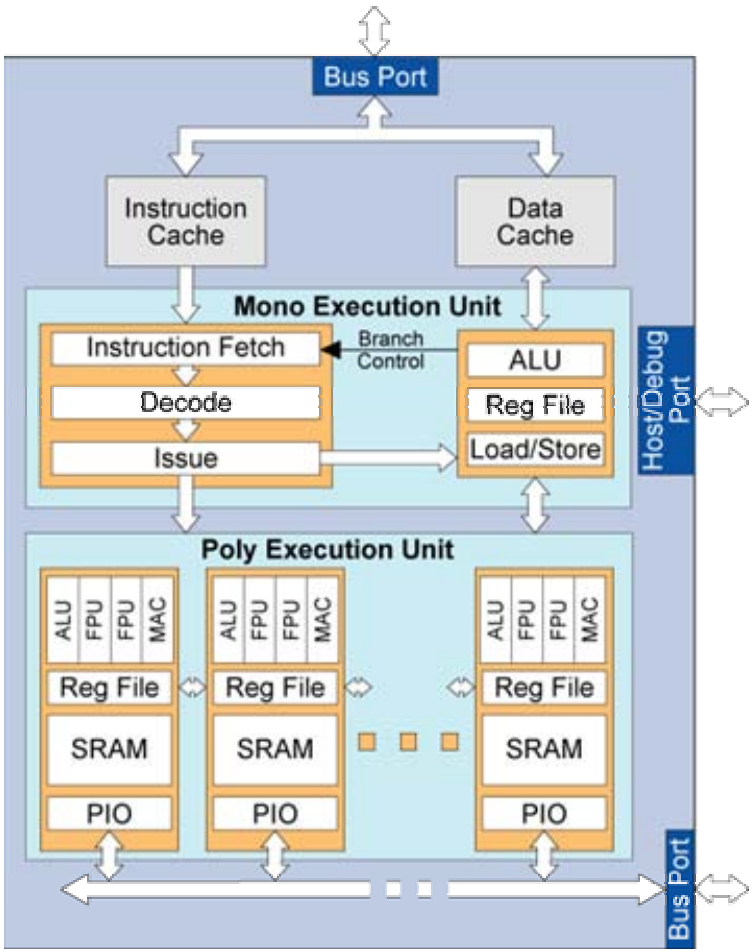


ClearSpeed's CS301:
The World's First Commercially-Available Stream Processor

Architecture, Algorithms and Benchmark Results

Simon McIntosh-Smith simon@clearspeed.com
Dairsie Latimer dairsie@clearspeed.com
Ron Bell ron.bell@awe.co.uk
Stephen Hudson stephen.hudson@awe.co.uk

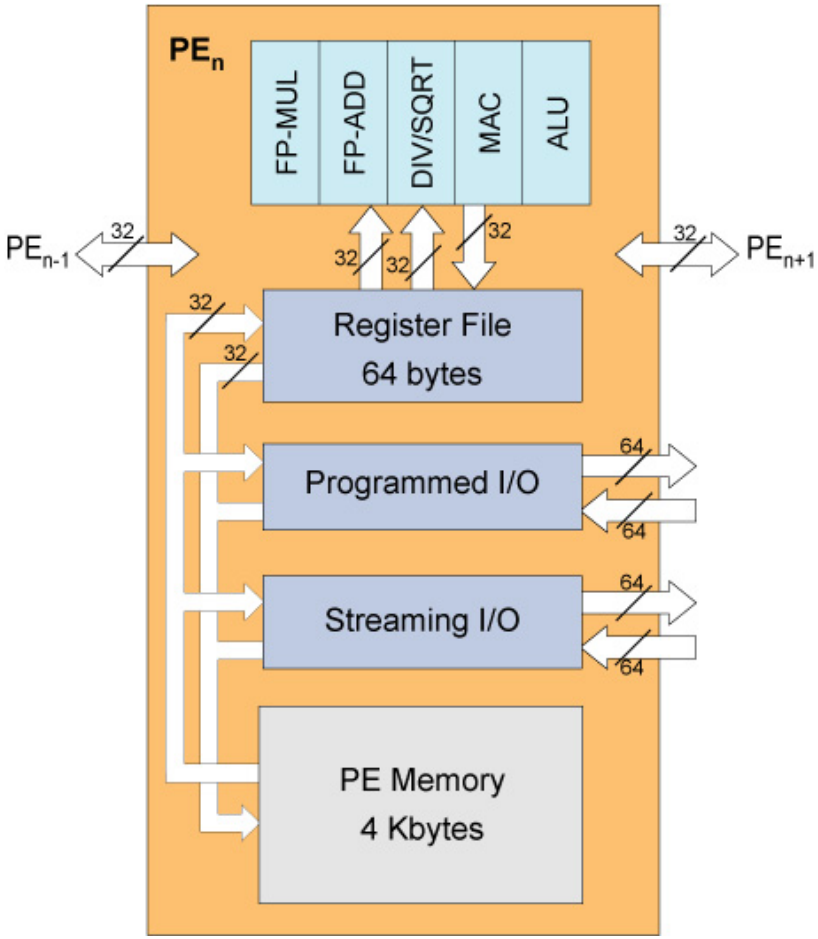
CS301 Processor



- Multi-threaded Array Processing
  - Programmed in high-level languages
  - Hardware multi-threading
    - Enables simultaneous data streaming and computation for latency tolerance
  - Run-time extensible instruction set
- Array of Processors Elements
  - PEs are VLIW cores
  - Flexible data parallel processing
  - Built-in PE fault tolerance, resiliency
- High performance, low power
  - 10 GFLOPS/Watt
- Multiple high bandwidth I/O channels

© *ClearSpeed* 2004 I www.clearspeed.com

CS301 Processing Elements



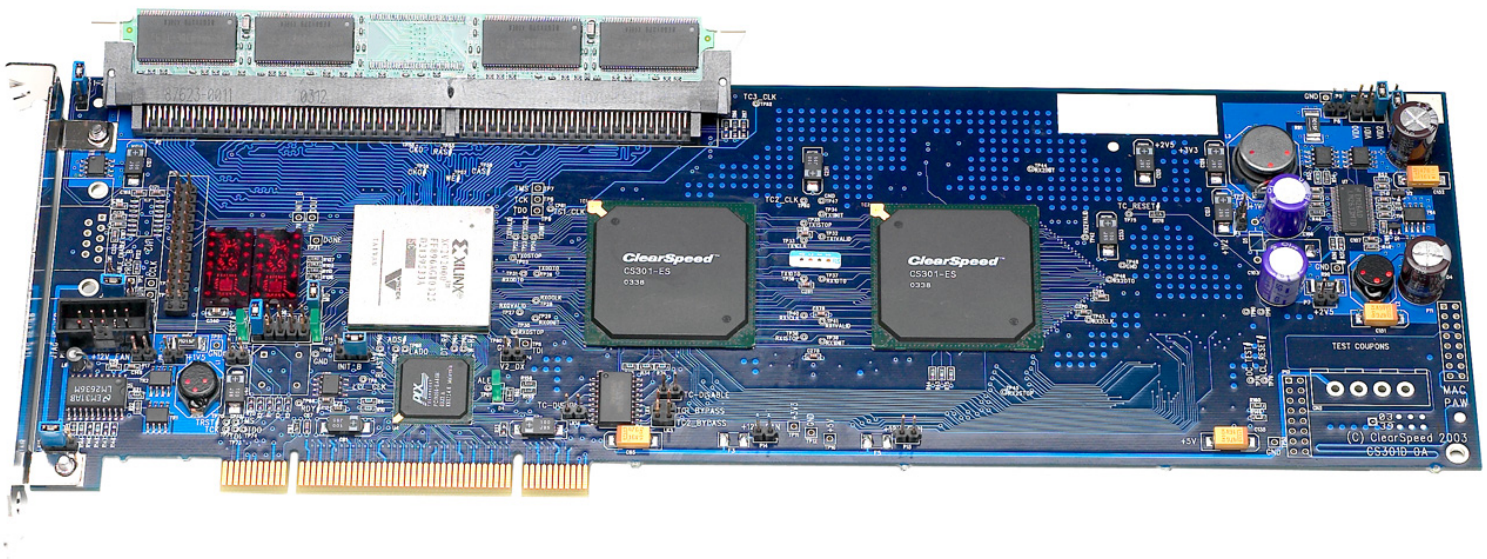
Each PE is a VLIW processor:

- Multiple execution units
  - Floating point adder
  - Floating point multiplier

  } 32-bit IEEE 754
  - Divide/square root unit
  - Fixed point MAC 8x8->16+48
  - Integer ALU with shifter
  - Load/store
- High-bandwidth, 5-port register file (3r, 2w)
- Closely coupled 4KB SRAM for data
- High bandwidth per PE load/store (PIO)
- Per PE address generator
  - Complete pointer model, including parallel pointer chasing and vectors of addresses

© *ClearSpeed* 2004 I www.clearspeed.com

CS301-based development board



- 2 chip board – 50 GFLOPS peak @ 10W total
- 200K FFTs/s (1K complex single precision IEEE754)
- Up to 1GB DRAM for local processing
- Shipping since 1Q04
- Single slot width full-size PCI card

© *ClearSpeed* 2004 I www.clearspeed.com

What Applications Can Be Accelerated?

Any applications with significant *data parallelism*:

- Fine-grained – vector operations
- Medium-grained – unrolled independent loops
- Coarse-grained – multiple simultaneous data channels/sets

Example applications and libraries include:

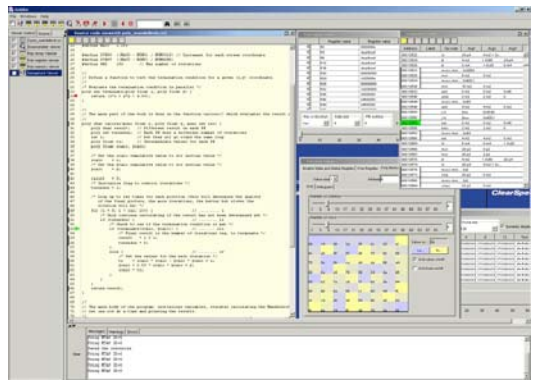
- Math libraries – BLAS, LAPACK (→ Matlab, Maple, …)
- Chemistry – GROMACS, CHARMM, BLAST, DLPOLY, …
- Computational finance – Monte Carlo, genetic algorithms
- Intelligent systems – artificial neural networks
- Signal processing – FFT (1D, 2D, 3D), FIR
- Simulation – CFD, N-body, Finite Element
- Image processing – filtering, image recognition, DCTs

© *ClearSpeed* 2004 I www.clearspeed.com

Software Development Environment

Software Development Kit (SDK)

- C compiler, assembler, libraries, visual debugger, etc.
- CS301-based development boards
- Available for Linux and Windows



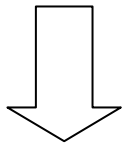
Applications and libraries under development

- Math – L3 BLAS, LAPACK
- DSP – FFTs (1D, 2D, 3D)
- Bio/Chemistry – GROMACS, DLPOLY, DockIt
- Financial – random number generation, Monte Carlo

© *ClearSpeed* 2004 I www.clearspeed.com

Porting Code

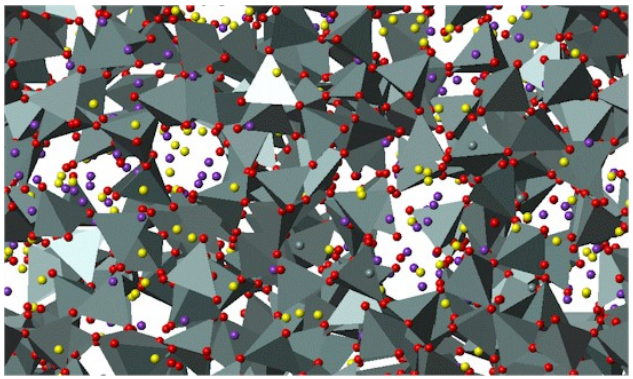

```
void daxpy(double *c, double *a, double alpha, uint N) {
  uint i;
  for (i=0; i<N; i++)
    c[i] = c[i] + a[i]*alpha;
}
```



```
void daxpy(double *c, double *a, double alpha, uint N) {
  uint i;
  poly double cp, ap;
  for (i=0; i<N; i+=num_pes) {
    memcpym2p(&cp, &c[i+pe_num], sizeof(double));
    memcpym2p(&ap, &a[i+pe_num], sizeof(double));
    cp = cp + ap*alpha;
    memcpyp2m(&c[i+pe_num], &cp, sizeof(double))
}
```


© *ClearSpeed* 2004 I www.clearspeed.com

ClearSpeed-AWE Joint Investigation



- Chemistry codes: DLPOLY (Molecular Dynamics)
  - Owned by UK Daresbury Lab, heavily used at AWE
  - Widely used in academia and industry
  - 91% of CPU in 5 relatively small routines
  - One of these (forces) calls the other 4 to compute forces on all atoms
  - “forces” called once per time step
  - Data needing to be returned by “forces” from CS to host relatively small
  - Calculation for each atom is independent

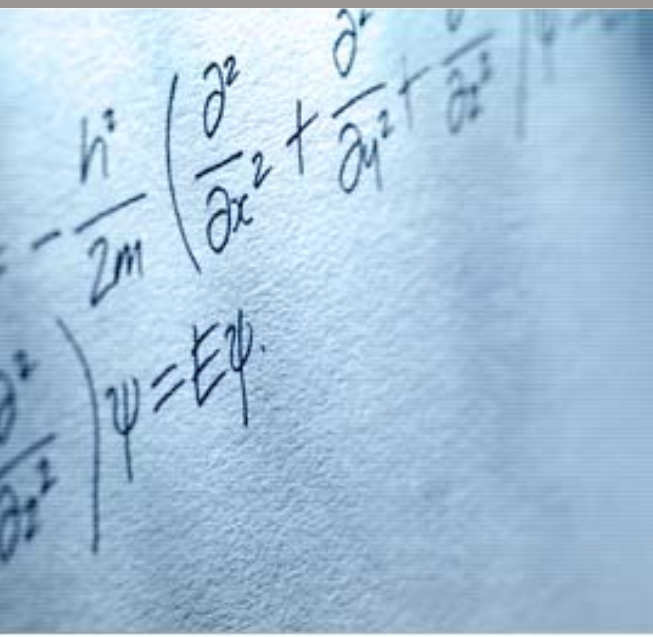
- Matrix Multiply Benchmark (SGEMM)
  - CS301 single precision code started at ~20% efficiency
  - AWE helped CS restructure code to give 12 GFLOPS – 47%
  - Performance verified by AWE on CS301 hardware
  - Next-generation processor from ClearSpeed significantly increases this performance – “Avebury”

© *ClearSpeed* 2004 I www.clearspeed.com

ClearSpeed

Poster

Agenda

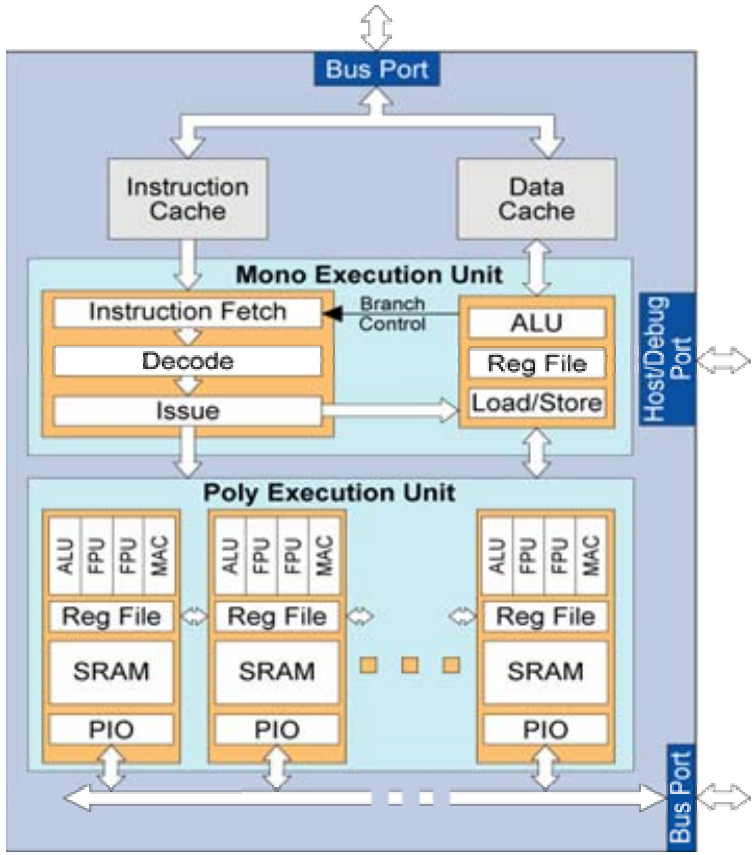


ClearSpeed's CS301:
The World's First Commercially-Available Stream Processor

Architecture, Algorithms and Benchmark Results

Dairsie Latimer dairsie@clearspeed.com
Simon McIntosh-Smith simon@clearspeed.com
Ron Bell ron.bell@awe.co.uk
Stephen Hudson stephen.hudson@awe.co.uk

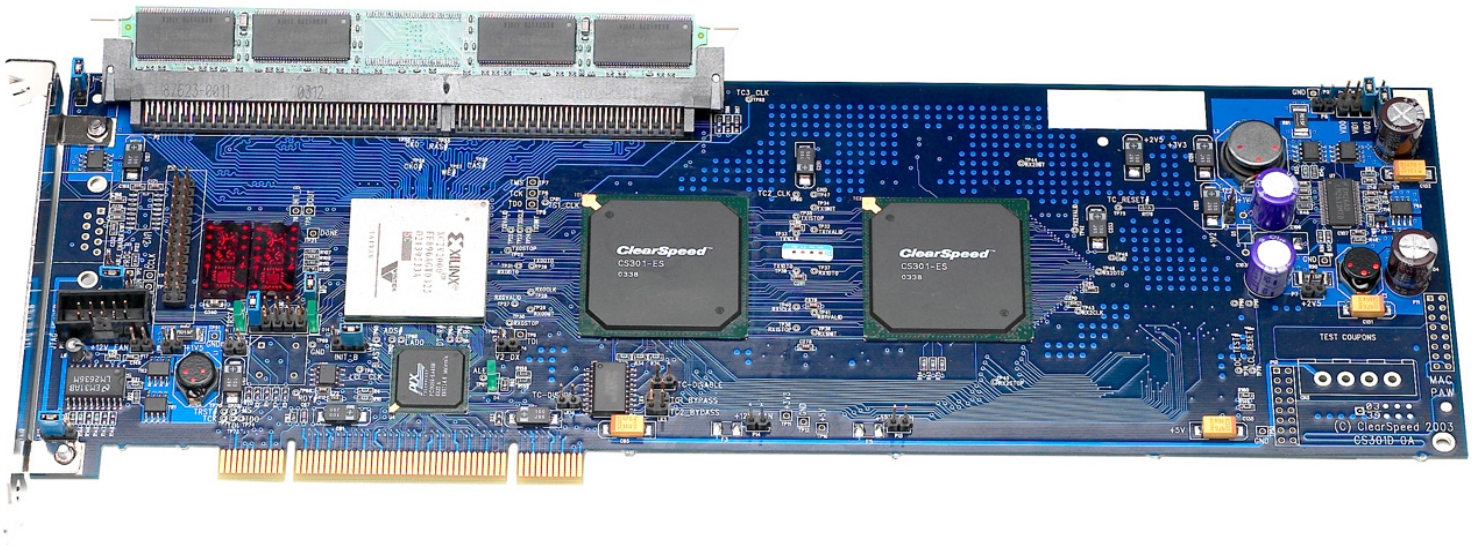
A Better Approach: ClearSpeed's CS301 Processor



- Multi-threaded Array Processing
  - Programmed in high-level languages
  - Hardware multi-threading
    - Enables simultaneous data streaming and computation for latency tolerance
  - Run-time extensible instruction set
- Array of Processors Elements
  - PEs are VLIW cores
  - Flexible data parallel processing
  - Built-in PE fault tolerance, resiliency
- High performance, low power
  - 10 GFLOPS/Watt
- Multiple high bandwidth I/O channels

© *ClearSpeed* 2004 I www.clearspeed.com

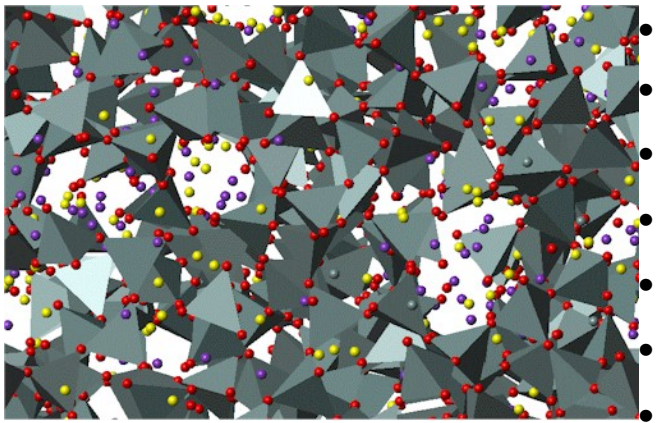
CS301 Based Development Board



- 50 GFLOPS peak @ 10W maximum
- 200K FFTs/s (1K complex single precision IEEE754)
- Up to 1GB DRAM for local processing
- Single slot width full-size PCI card
- In evaluation use since early 2004

© *ClearSpeed* 2004

ClearSpeed AWE Joint Investigation

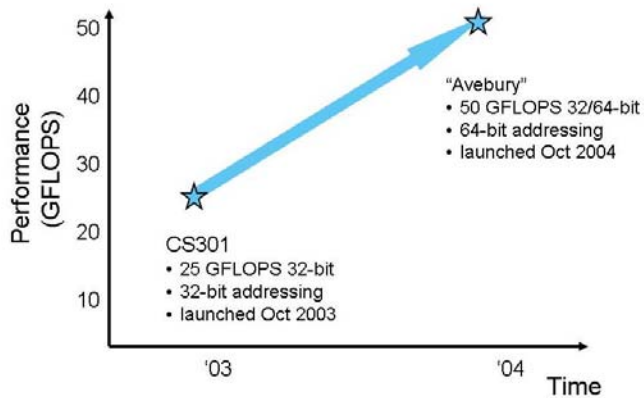


Chemistry codes: DLPOLY (Molecular Dynamics)

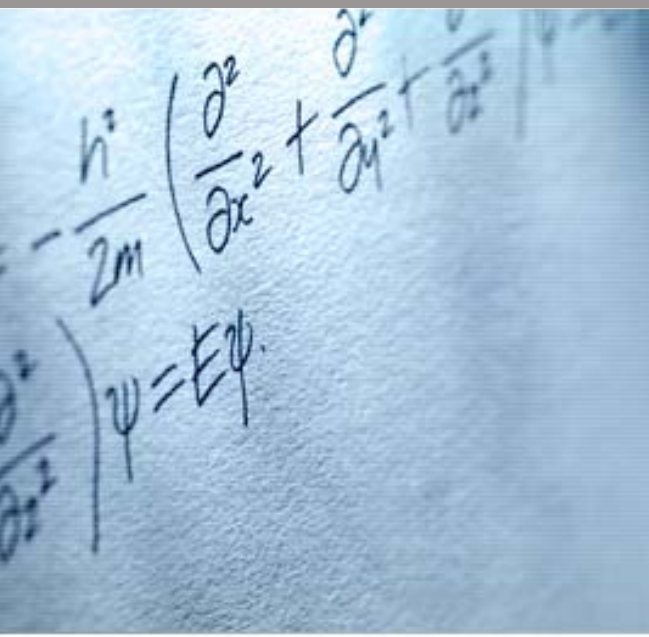
- Owned by UK Daresbury Laboratory
- Widely used within AWE, also academia & industry
- 91% of CPU time in 5 small routines
- One calls the other 4 to compute forces on all atoms
- Forces called once per time step
- Small amount of data returned by forces from CS to host
- Calculation for each atom is independent

Matrix Multiply Benchmark (SGEMM)

- CS301 single precision code started at ~20% efficiency
- AWE/CS code restructuring gave 12 GFLOPS – 47%
- Performance verified by AWE on CS301 hardware
- “Avebury” significantly increases this performance



© *ClearSpeed* 2004 I www.clearspeed.com



Extremely Cool, Extremely Fast

Intrigued? Visit our poster outside