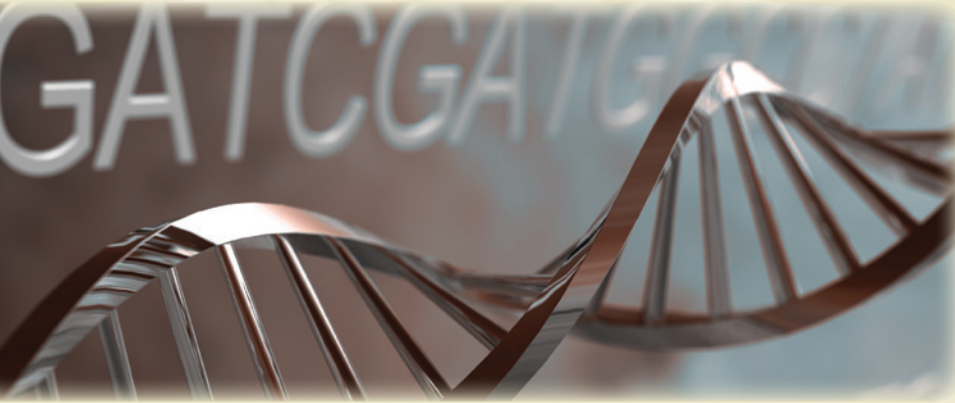


Hybrid-Core Computing for High Throughput Bioinformatics



Convey's groundbreaking hybrid-core technology is an elegant bioinformatics solution that dramatically increases throughput for genomics and proteomics applications. Many bioinformatics applications are uniquely suited to take advantage of Convey hybrid-core computing, achieving order of magnitude performance improvements.

Instruments for biology and medical research, such as next-generation sequencers, are generating vast amounts of data—so much that traditional high performance computing (HPC) servers can't keep up. As researchers and scientists demand more computational performance, computing centers are running out of the resources needed to obtain that performance: space, electrical power, and cooling.

The bioinformatics industry needs a more elegant solution—a new computing technology that significantly improves high throughput processing. Convey Computer Corporation has developed that solution: breakthrough hybrid-core technology that dramatically increases performance for high-throughput bioinformatics applications.

Hybrid-core computing extends a commodity instruction set architecture (ISA) with application-specific instructions to accelerate HPC applications. Convey's innovative hybrid-core solutions combine the economies and programmability of industry standard processors with the performance and efficiency of a hardware-based, application-specific design. Researchers can now dramatically increase application performance, while saving money on power and facilities costs.

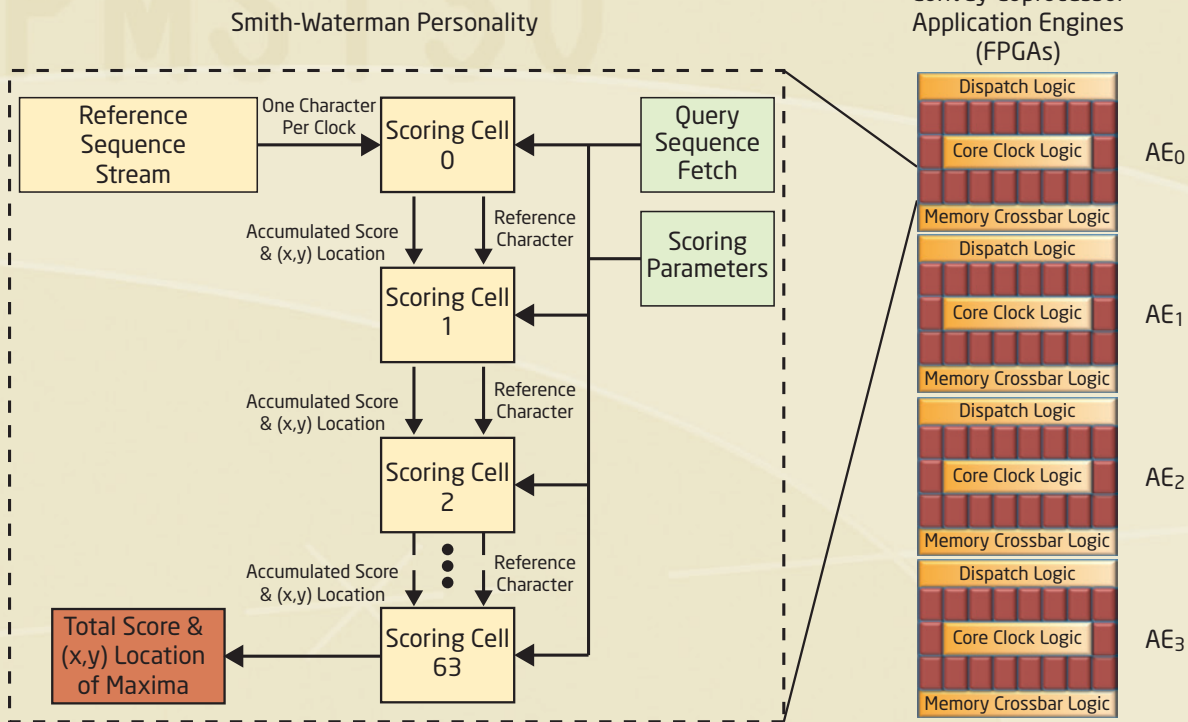
BIOINFORMATICS APPLICATIONS: A PERFECT MATCH

Most bioinformatics applications are uniquely suited for hybrid-core technology because they have low data interdependence, which greatly increases performance through parallelism. Furthermore, data types are small—four nucleotides can be represented in two bits, and all amino acids can be represented in five bits. Such small data type operations make more efficient use of logic gates than the data types dictated by commodity processors and conventional programming models. These characteristics allow bioinformatics applications to directly leverage the FPGA-based architecture of hybrid-core computing, achieving impressive speedups.

Many bioinformatics applications that randomly access memory, such as graph-based algorithms, commonly experience memory bottleneck performance limitations. Random memory accesses are especially troublesome for the traditional cache-based memory typically found on commodity servers. The highly parallel memory of Convey's cache-less approach allows application-specific logic to directly address individual words in memory, increasing effective memory bandwidth as much as thirty-fold. Algorithms such as de Bruijn short-read assemblers can greatly benefit from this type of memory architecture.

The World's First Hybrid-Core Computer.

LEARN MORE ABOUT THE WORLD'S FIRST HYBRID-CORE COMPUTER.
VISIT CONVEYCOMPUTER.COM OR CALL 1-866.338.1768



The Convey Smith-Waterman personality is constructed using many individually parallel scoring “tiles.” Each tile can be instrumented to search a unique data set, irrespective of its neighboring tiles. A tile contains a systolic array implementation of the linear Smith-Waterman scoring routine with sixty-four query cells. Each cell loads a character from the incoming search as well as the application’s requested scoring parameters. Once loaded, a tile begins streaming reference characters into the first cell of the tile. Once the cell has completed its scoring comparison, the accumulated score and reference character is passed to the next cell [N+1]. The last cell in the systolic array scoring chain delivers the final coprocessor score and its relative matrix location [X,Y] to coprocessor memory. Reference characters are shifted through the systolic cells at one character per clock cycle. The Convey Smith-Waterman personality contains 72 of these unique tiles on the HC-1, and 168 tiles on the HC-1^{ex}, each potentially executing independent searches.

A FAMILIAR LOOK AND FEEL: EASY TO USE, MANAGE, AND PROGRAM

Multi-core processors or asymmetric hardware-based solutions, such as GPGPUs and FPGAs, have been leveraged in the past to try to speed up bioinformatics applications. However, these solutions were not widely successful, because they were difficult to integrate and deploy.

The Convey systems solve the “ease of use” problem by marrying a reconfigurable coprocessor with a traditional x86 environment. Using field programmable gate array (FPGA) technology, application-specific instructions executed by the coprocessor appear as extensions to the x86 ISA. This solution combines the performance of application-specific hardware with the programmability and deployment ease of a commodity server. Applications can contain both x86 and coprocessor instructions in a single instruction stream—providing an ideal environment to harness the exponential growth of sequence data.

“Bioinformatics is making huge advances, but with the advent of modern technologies, such as the ultra high-throughput next-generation sequencers, the volume of generated data is out-stripping our ability to reduce that data to knowledge. The new computer capabilities that we will be bringing on-line will help to push the envelope of scientific discovery across a tremendous range of scientific disciplines.”

—Dr. Harold “Skip” Garner, executive director of VBI, professor in the departments of biological sciences and computer science at Virginia Tech, and professor in the department of basic science at the Virginia Tech Carilion School of Medicine

Convey systems are easy to program because they offer full support of ANSI standard C, C++ , and FORTRAN development environments. In addition, the Convey development environment includes a significant set of profiling tools and libraries that simplify programming.

The Convey coprocessor implements “personalities”—multiple, reloadable instruction sets that augment the host’s x86 instruction set. A personality can be applicable to a class of applications, such as bioinformatics, or a specific code, such as a search for an amino acid string in a protein molecule. Personalities make it easier for programmers to develop high-performance applications for the Convey systems, without the need to learn a special hardware programming language.

Convey’s FPGA-based architecture is an open platform, providing a blank slate for application innovation. Using the Convey Personality Development Kit (PDK), developers can choose to create their own application specific instruction set for either academic research or commercial purposes.

PROVEN RESULTS: DRAMATIC ACCELERATION

Working closely with numerous customers and partners, including instrument manufacturers, Convey is accelerating bioinformatics applications in a variety of areas. Applications include primary analysis, reference based and de novo assembly, sequence alignment, and downstream research such as phylogenetics.

Convey’s implementation of the Smith-Waterman algorithm, widely used to align DNA and protein sequences, is 172 and 401 times faster (on the HC-1 and HC-1^{ex}, respectively) than the best software implementation. Both performance numbers represent the fastest Smith Waterman to date.¹

The Virginia Bioinformatics Institute (VBI) is using Convey’s Smith-Waterman implementation as part of a project to analyze data from the 1000 Human Genomes Project.

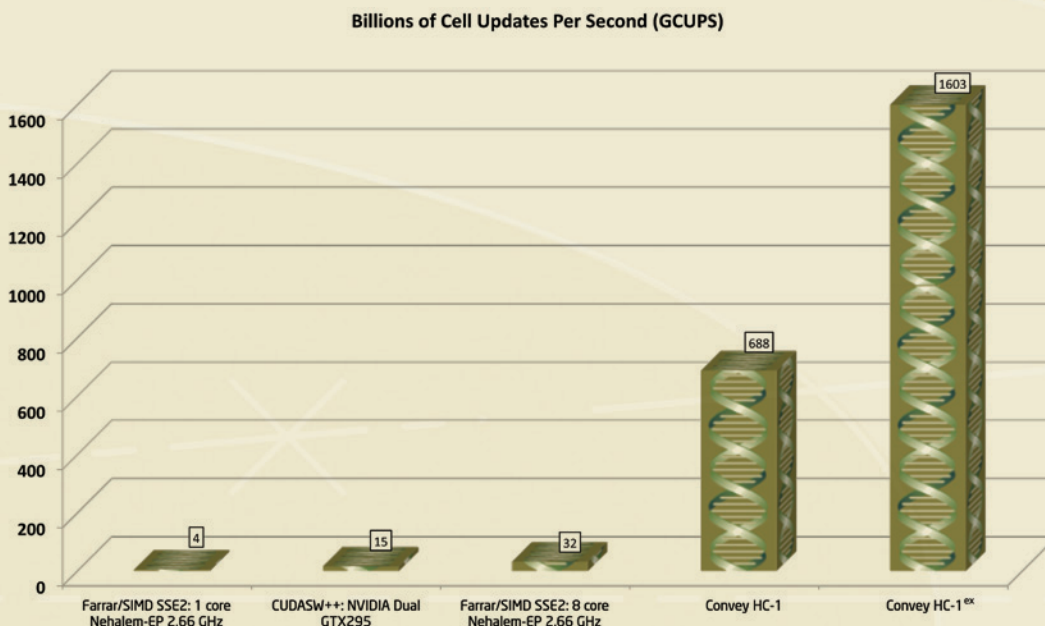
In addition to the Smith-Waterman algorithm, Convey is accelerating implementations of other common bioinformatics algorithms, such as BLAST, HMMER3 and Velvet. Additionally, Convey customers are using Convey systems to optimize a wide range of applications. Jason Bakos, assistant professor at the University of South Carolina’s department of computer science and engineering, is using a Convey HC-1 server to speed up the MrBayes phylogenetics algorithm. Researchers at the University of California, San Diego have optimized the InsPecT/MS-Alignment proteomics software package, and researchers at Lawrence Berkeley National Laboratory are optimizing a graph-based algorithm for short-read assembly.

¹ For a single system. According to Convey’s internal benchmarking, the Smith-Waterman algorithm runs 172 times faster on the HC-1, and 401 times faster on the HC-1^{ex} than a software implementation on an Intel Nehalem core.

Hybrid-Core Computing for High Throughput Bioinformatics



SMITH-WATERMAN KERNEL PERFORMANCE COMPARISON¹



AN ELEGANT SOLUTION FOR KEEPING UP WITH BIOINFORMATICS DATA

Convey’s groundbreaking hybrid-core technology is an elegant bioinformatics solution that dramatically increases throughput for genomics and proteomics applications. Many bioinformatics applications are uniquely suited to take advantage of the Convey architecture, achieving order of magnitude performance improvements.

Of course, performance improvements mean faster results. Yet more importantly, faster results mean that research organizations can now fundamentally improve the quality of their research. Scientists have the power to perform more accurate search methods, search larger databases, use more sequences in a query, and compare whole genomes rather than fragments. By speeding up some of the world’s most popular bioinformatics algorithms, the Convey HC-1 and Convey HC-1^{ex} are helping scientists and researchers better manage and analyze massive amounts of research data.

For more information, please see www.conveycomputer.com.

¹ According to Convey’s internal benchmarking, the Smith-Waterman algorithm runs 172 times faster on the HC-1, and 401 times faster on the HC-1^{ex} than a software implementation on an Intel® Nehalem core.