

Efficient Graph-Based Assembly of Short-Read Sequences on a Hybrid-Core Architecture

Tony Brewer
Chief Technology Officer
Convey Computer Corporation

Abstract:

Advanced architectures can deliver dramatically increased throughput for genomics and proteomics applications, reducing time-to-completion in some cases from days to minutes. One such architecture, hybrid-core computing, marries a traditional x86 environment with a reconfigurable coprocessor, based on field programmable gate array (FPGA) technology. In addition to higher throughput, increased performance leads to solving next-generation problems—problems that cannot practically be solved on today's commodity systems.

This presentation will discuss the approach used by Convey's de Bruijn graph constructor for short-read, *de novo* assembly. Bioinformatics applications that have random access patterns to large memory spaces, such as graph-based algorithms, experience memory performance limitations on cache-based x86 servers. Convey's highly parallel memory subsystem allows application-specific logic to simultaneously access up to 8192 individual words in memory, significantly increasing effective memory bandwidth over cache-based memory systems. Many algorithms, such as Velvet and other de Bruijn graph-based, short-read *de novo* assemblers, can greatly benefit from this type of memory architecture. Furthermore, the reconfigurable coprocessor permits instruction set definitions like small data type operations (four nucleotides can be represented in two bits). Such custom instructions sets execute faster and can make more efficient use of logic gates than the data types dictated by conventional programming models.