HPC-BLAST
Scaling the Life Sciences for the Intel® **M**any **I**ntegrated **C**ore future

R. Glenn Brook, Ph.D., Project Lead
Shane Sawyer, Research Associate
Co-Director, U.T. Intel Parallel Computing Center

Joint Institute For Computational Sciences,
University of Tennessee, Knoxville

HPC-BLAST Development

- **HPC-BLAST** is a high-throughput, highly thread count version of the **NCBI B**asic **L**ocal **A**lignment **S**earch **T**ool. Designed from the outset for use with the Intel® Many Integrated Core architecture.
- The NCBI BLAST algorithm is used extensively in the field of Bioinformatics.
- **HPC-BLAST** is a scalable code employing a **symmetric MPI** execution mode.
- The development of **HPC-BLAST** made extensive use of Rogue Wave Totalview, Allinea DDT, Intel VTune and Intel Inspector in addition to the 2013 and 2015 Intel C++ compilers.

The NCBI-BLAST algorithm: Overview

[1] Concatenate a batch of queries together

word	offset
TCG	13
TAC	14
CCA	15
…	…

[2] Build a lookup table for the concatenated query

..ATCGAGGTCATGCTACGGA..

word	offset
TCG	13
TAC	14
CCA	15
…	…

[3] For each entry in the comparison database mark relevant matches (seeds) in the lookup table

[4] Alignment proceeds in two stages around relevant matches. An ungapped alignment occurs first and is followed by a gapped alignment. The result is a sub-string and associated score.

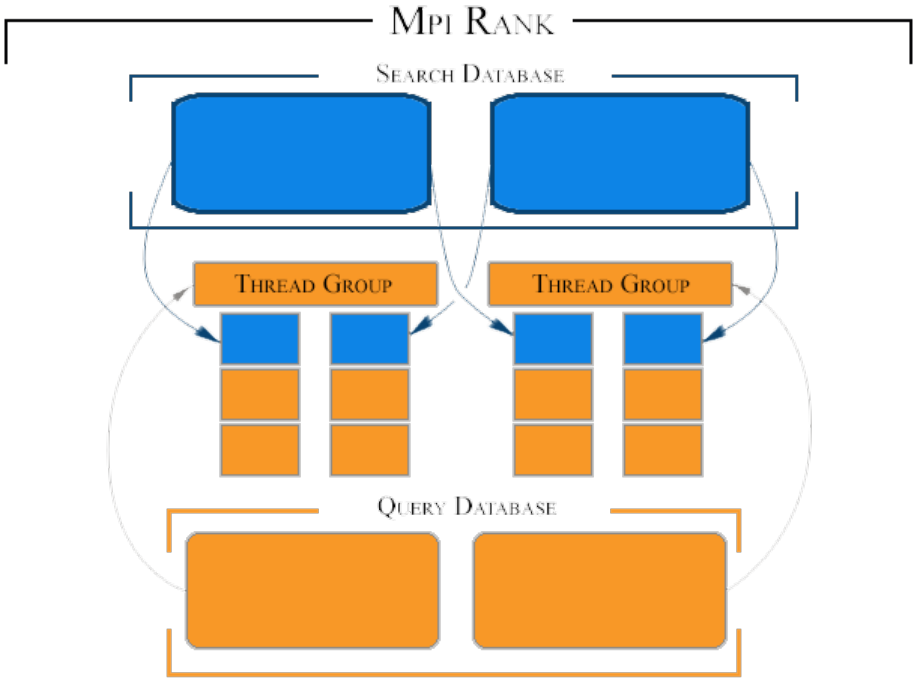
..ATCGAGGTCATGCTACGGA..

..GGGTCATGATACGA..

GATACG → USER

[5] Report the highest scoring sub-string matches to the user

HPC-BLAST Approach



Problem:
Large volumes of un-sequenced data are being generated at increasingly exponential rates. Scientists and industry, using traditional non-HPC approaches, are struggling to handle the deluge of data.

Solution:
A **M**any **I**ntegrated **C**ore approach

High thread counts can enable scalability, performance and significant throughput of results

With **HPC-BLAST** our team has designed a new, highly threaded variant of the NCBI-BLAST approach

Completed Capability

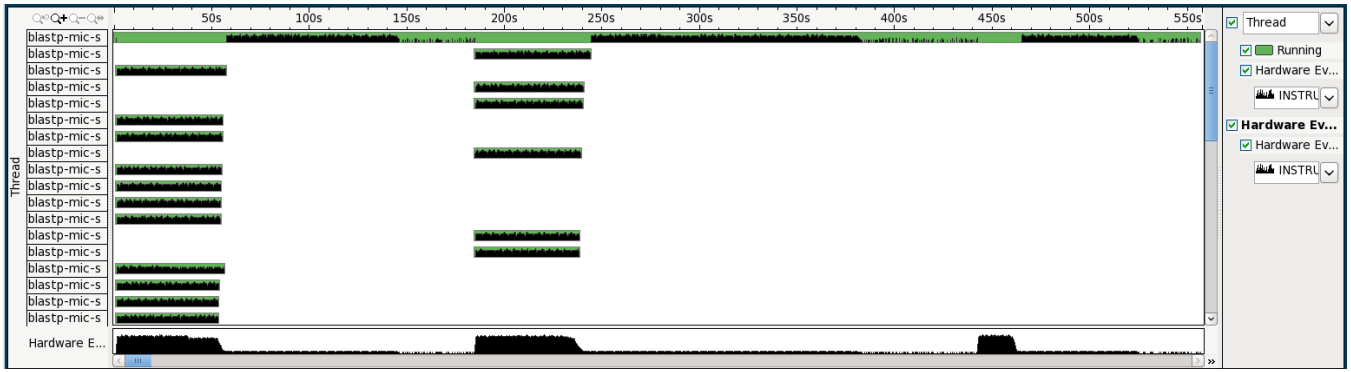
- ✓ A coarse grain, MPI layer implementation for search **database distribution**.
- ✓ A highly threaded approach to **query distribution** and **parallel searching** for mid-grain parallelism.
- ✓ A bin-packing load balancer.

Enhancements in progress

- Distributed, dynamic load balancing based on local work loads.
- Large scale, parallel I/O for more efficient output of results.
- Exploring SIMD opportunities.

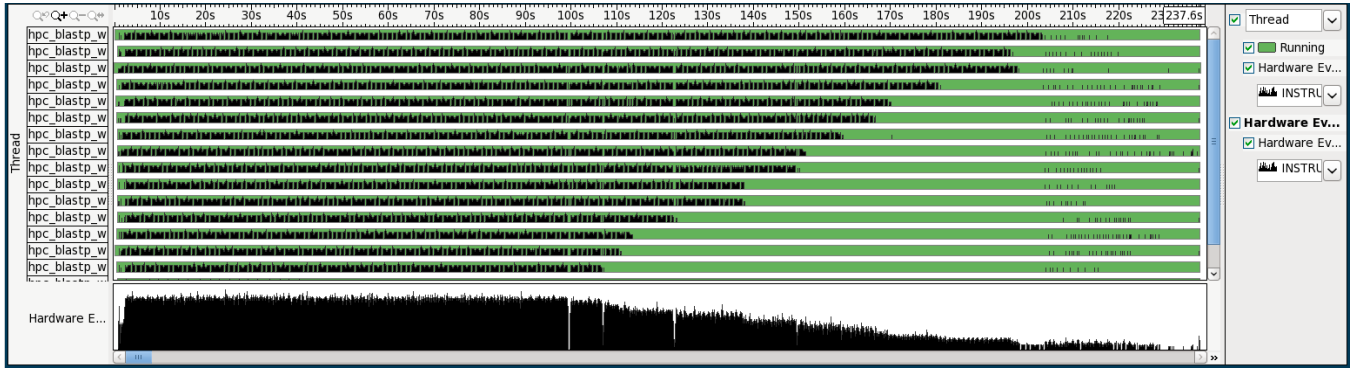
Preliminary Performance: Concurrency

Intel® VTune™ 2013 concurrency analysis: 16 threads. NCBI-BLAST v.s. HPC-BLAST



NCBI-BLAST

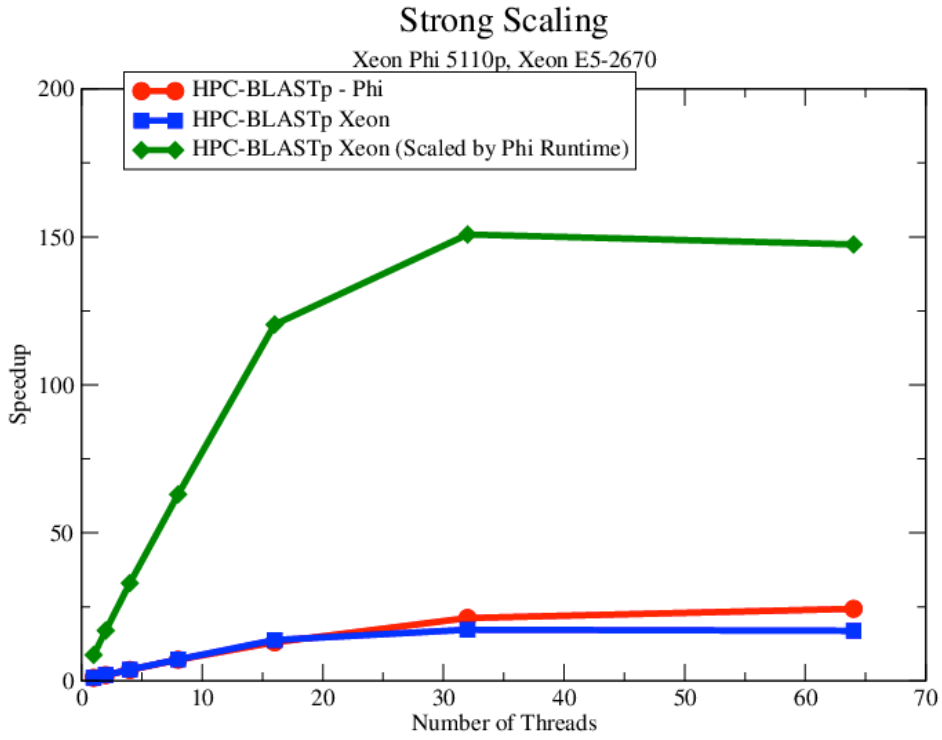
- Green bars show running threads
- Clear serialization halting performance

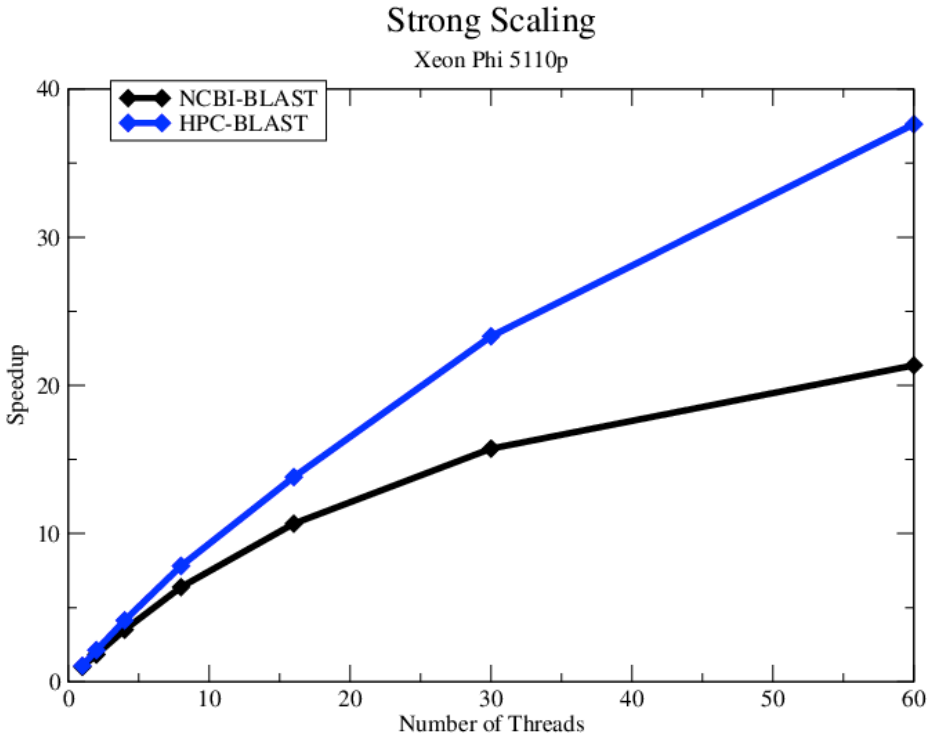


HPC-BLAST

- Concurrency is dramatically improved
- First pass at threading model implementation. Improvements continue with optimization*!*

Preliminary Performance: Speedup





- Speedups scaled by 1 thread NCBI BLAST on Xeon Phi.
- Used nr.01 partition as subject database.
- 128 Queries taken from subject database.

- Speedups scaled by 1 thread NCBI BLAST on Xeon Phi.
- HPC-BLAST used two ranks each with same subjects.
- Used nr.01 partition as subject database.
- 534 Queries taken from subject database.

Insights

- **Leveraging full performance requires thread concurrency and vectorization.**
- **Flexibility in expressing parallelism at all levels to tune performance crucial for code not readily mapped to architecture.**
- **VTune is very helpful in identifying hotspots; Inspector is useful in finding race conditions.**
- **One of the key remaining challenges is to refactor hotspots to utilize SIMD operations while matching output produced by standard implementation.**

THANK YOU

Shane Sawyer
Research Associate, Joint Inst. for Comp. Sciences
Shane-Sawyer@Tennessee.edu

R. Glenn Brook, Ph.D.
Co-Director, U.T. Intel Parallel Computing Center
Director, Application Acceleration Center of Excellence
Chief Technology Officer, Joint Inst. for Comp. Sciences
Project Lead: HPC-BLAST

Glenn-Brook@Tennessee.edu

Bhanu Rekepalli, Ph.D.
Computational Scientist, Joint Inst. for Comp. Sciences
Adjunct Asst. Professor, Univ. of Tennessee
Project Co-lead: HPC-BLAST

Brekapal@Tennessee.edu