



A /ORNL PARTNERSHIP
NATIONAL INSTITUTE FOR COMPUTATIONAL SCIENCES

NICS

Systems Biological Workbench Development on NICS Resources

Bhanu Rekepalli

Mark Fahey, Aaron Vose, Chuan He, and Igor Jouline

University of Tennessee
NICS



NATIONAL INSTITUTE FOR COMPUTATIONAL SCIENCES

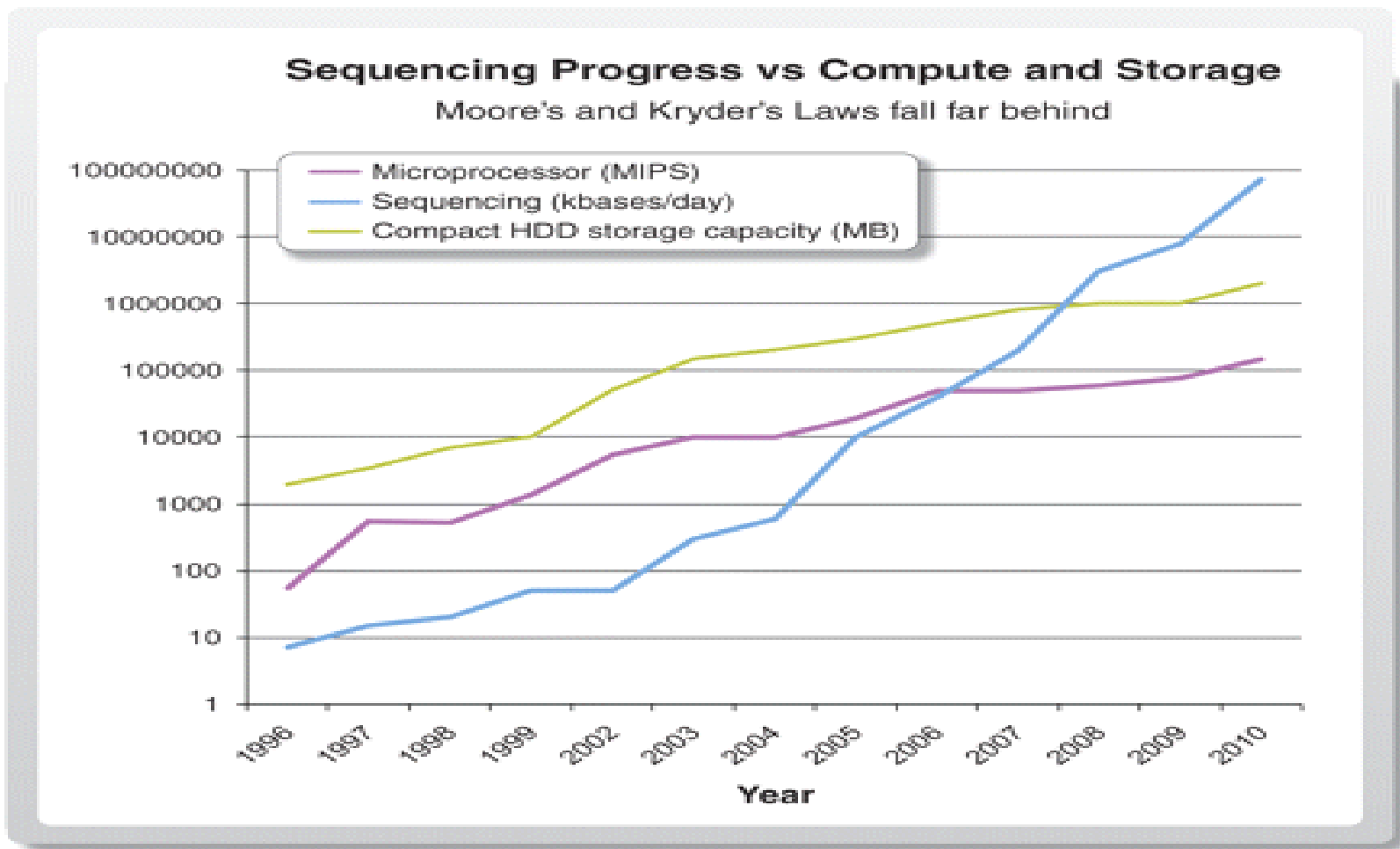


Systems Biological Workbench

- **Improve tools for analysis**
 - Existing bioinformatics tools run on desktops, workstations, and small clusters
- **Improving performance and scalability of tools is critical to transform exponentially growing data into biological knowledge**
 - Data from computational biology and biomedical informatics applications is growing rapidly
 - This data contains invaluable information directly related to human health.
- **Develop novel software applications that include optimization techniques to improve the scalability of the most widely used bioinformatics tools on advanced parallel architectures**
 - Improvements allow a near-linear scaling to tens of thousands of processing cores along with full machine capability runs on current petaflop supercomputers such as Kraken
- **New tools, still under development, increase data analysis by four to five orders of magnitude**
- **Tool development fosters research collaborations between six universities from Tennessee and South Carolina**



Problem: Exponential Genomic Data Growth



Solution: Next Generation Architectures

- **Large Scale Analysis**
 - Supercomputing
 - Hybrid Computing
- **Medium Scale Analysis**
 - Grid computing
 - Cloud computing
- **Develop Highly Scalable Informatics Toolboxes**

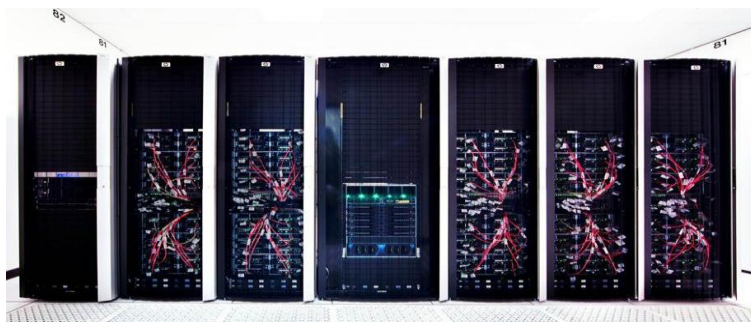


Applications Under Development

- **Bioinformatics**
 - BLAST, HMMER
 - ClustalW, MUSCLE
 - PhyML, GARLI, RAxML, MrBayes
- **Cheminformatics**
 - DOCK6, AutoDock, AMBER, NAMD
- **Large Scale Data Handling on Advanced Architectures and Data Mining**
 - Job Scheduling Algorithms
 - Very Large Scale Data Analysis and Databases
- **Web Interfaces and Data Visualization Tools**



XSEDE Resources Used for Development



Keenland

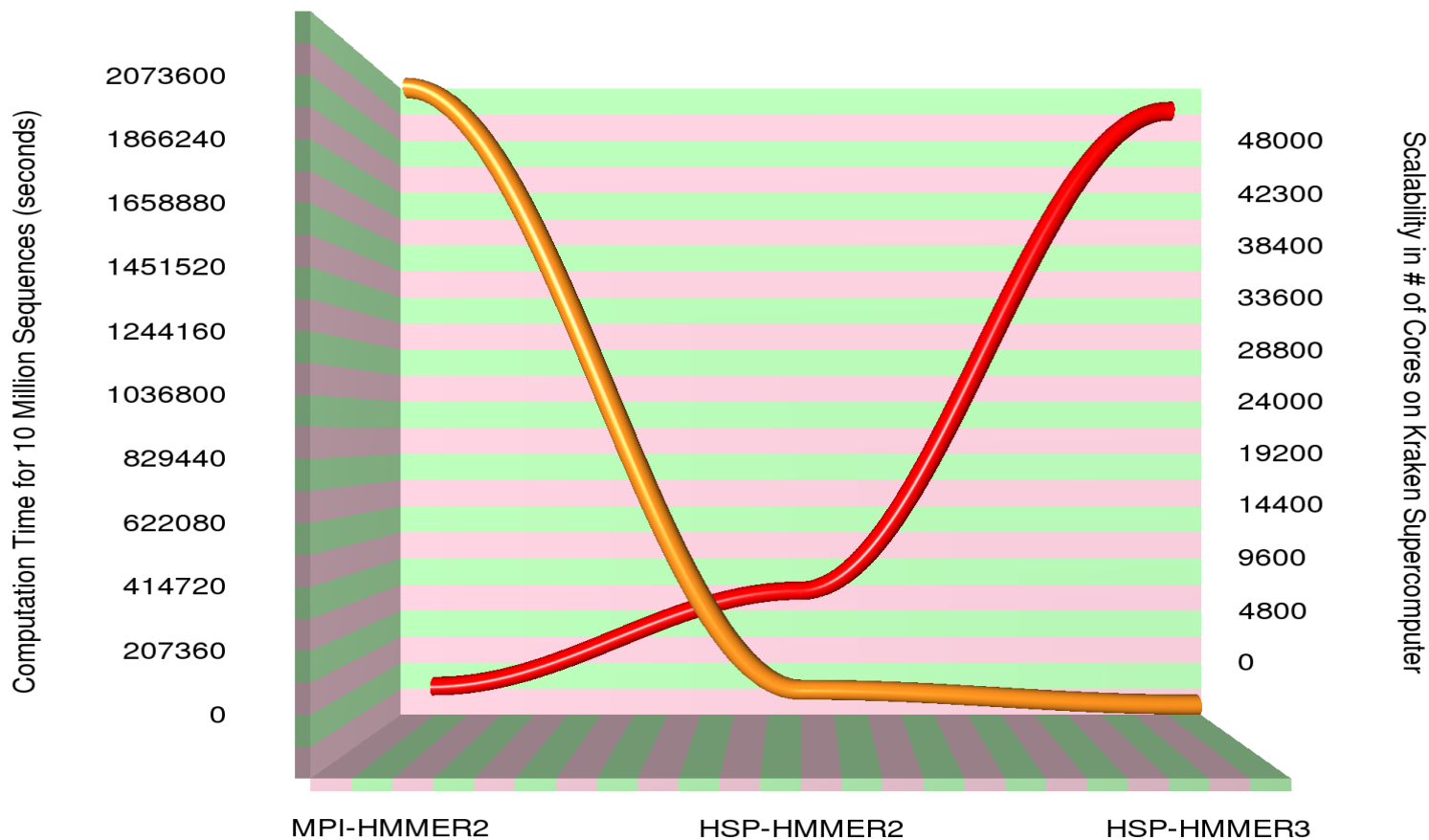


Nautilus

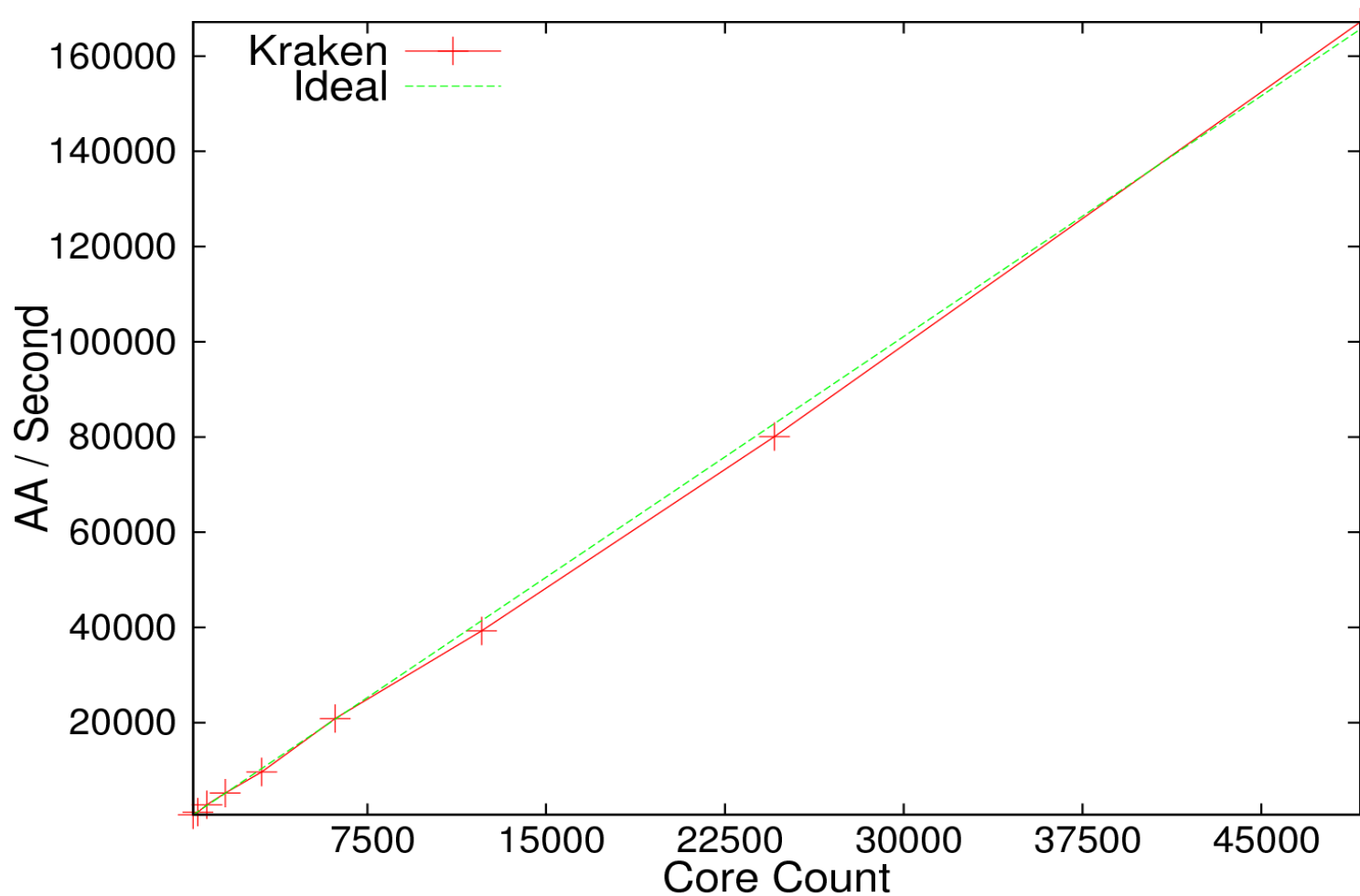


HSP-HMMER

Performance Improvements of Scaled HMMER

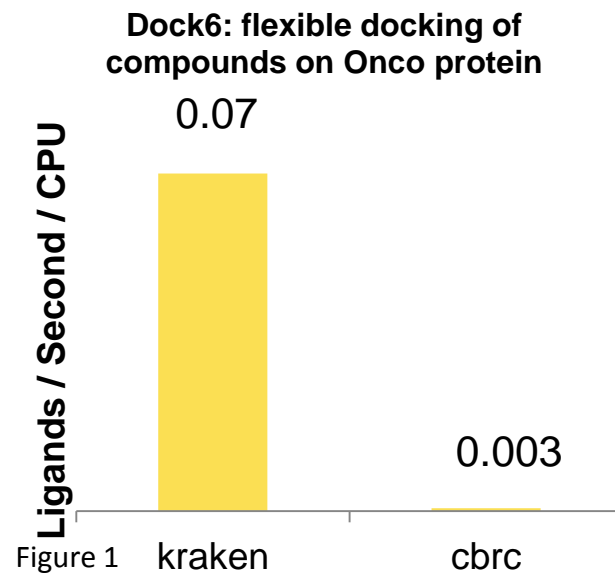


HSP-BLAST



Highly Scalable Parallel Docking

- **Improve speed and scaling of molecular docking tools**
 - Tools used to search compounds for novel drug discovery
 - Libraries such as ZINC and PubChem contain millions of vendor and academic compounds and are growing rapidly
- **Algorithmic and I/O improvements of the MPI version of Dock6**
 - Used profiling tools to tune the code
 - Modified compiling options and linked in faster math libraries
 - Developed wrapper code in which the master process dynamically load-balances the computation
 - The wrapper code can use other docking tools
 - Scaled code to 8,000 cores on Kraken
 - Next step is to scale docking codes to full machine runs
 - Achieved a 20X speedup per core on Kraken
 - Runs took 2 months on the CBRC cluster (136 cores) versus 40 minutes on NICS Kraken (8,000 cores)
 - It is now possible to dock ~55 million oncology related receptors in one day versus 2 years
- **Speedups achieved reduce computational time for drug discovery from years to months or days, significantly reducing the time to market of new drug**



Speedups achieved on Kraken vs CBRC

With Algorithmic, I/O improvements and code optimizations we achieved a 20X speedup per core on NICS Kraken supercomputer compared to MUSC Computational Biology Resource Center (CBRC) cluster .



Collaborations: NSF EPSCoR

- **Clemson University**
 - Jill Gemmill, Amy Lawton-Rauh, and Galen Collier
- **Medical University of South Carolina**
 - Yuri Peterson and Brendan Duggan
- **Vanderbilt University**
 - Peter Cummings, and Junhwan Jeon
- **Clafin University**
 - William L. Mondy, Nick Panasik
- **Tennessee State**
 - Ali Sekmen
- **University of South Carolina**
 - Jerry Ebalunode





Contact

Bhanu Rekepalli

University of Tennessee, NICS
bhanu@utk.edu



NATIONAL INSTITUTE FOR COMPUTATIONAL SCIENCES

