

Vipin Sachdeva

CONTACT INFORMATION	IBM Future Technology Design Center Indianapolis, IN 46204	<i>Email:</i> vipin.sachdeva@gmail.com
RESEARCH INTERESTS	Multi-Core Programming, Healthcare and Life Sciences workloads, Workload-driven architecture design, Performance analysis, Heterogeneous architectures, Computational Science.	
EDUCATION	University of New Mexico , Albuquerque, New Mexico USA Master of Sciences with Distinction Overall GPA: 3.88/4 Computer Engineering, September 2003 - December 2005 <ul style="list-style-type: none">• Concentration: High Performance Computing• Advisor: Dr. David A. Bader, Georgia Institute of Technology• Thesis Topic: High-performance computing for Computational Biology and Graph theory• Outstanding Graduate Student in Computer Engineering, 2005 Indian Institute of Technology , Guwahati, Assam India Bachelor of Technology, Mechanical Engineering, July, 1998 - May, 2002 <ul style="list-style-type: none">• Advisor: Dr. M. K. Das.• Thesis Topic: Design and Implementation of Parallel Delaunay Triangulation Algorithm.	
RESEARCH EXPERIENCE	<i>HPC Specialist (IBM Future Technology Solution Design Center)</i> Jan 2008 - Present <ul style="list-style-type: none">• Evaluation of Multi-Core Processors and Programming Platforms with Jaccard coefficient: This project ports and evaluates the performance of the Jaccard coefficient used in chemical database fingerprinting and numerous other areas to available multi-core architectures such as Intel Woodcrest (scalar and SSE3), Cell Broadband Engine and the Nvidia 8800 GPUs. The implementations are completed in dynamic code-generation platforms such as Rapidmind as well as vendor-released programming platforms with specialized hardware such as IBM SDK and CUDA released by Nvidia, and SSE3 from Intel. For the workload being evaluated, the goal is to find out the best platform for metrics such as performance (absolute performance as well as performance per watt and dollar) and programmer productivity. <i>Long-Term Supplemental (Novel Systems Architecture Group)</i> July 2006 - Jan. 2008 <ul style="list-style-type: none">• Exploring the Cell architecture for computational biology workloads: This project investigates the suitability of the Cell Broadband Engine Architecture (CBEA) for common computational biology kernels such as dynamic programming, database searches and other kernels commonly found in the popular bioinformatics applications. So far, we have focused on Smith-Waterman from FASTA, Clustalw and the P7viterbi kernel from HMMER for this analysis. We have a fully functional Clustalw running on the Cell processor, with Smith-Waterman and P7Viterbi running for limited sizes, with excellent speedups in all cases. The Cell processor is the first truly multi-core heterogeneous chip, optimized for compute-intensive workloads with nine cores on one chip, eight of which are specially-designed vectorized cores connected with a high-speed memory controller and high-bandwidth bus interface, all integrated on-chip. Significant changes had to be made in each of the applications to account for parallelization of tasks on the Cell broadband engine, managing explicit data movement and vectorization of time-consuming	

kernels. This work was covered by the **Genome Technology** magazine in their **October 2007** issue as part of an article on the Cell processor

- **FFTW for the Cell processor:** This project ported FFTW to the Cell processor, with Matteo Frigo, who is one of the developers of FFTW. Issues involved including parallelizing Fourier transforms across the SPUs, adapting existing vector codelets to the Cell SPU APIs and managing the local store memory of the SPUs for transpose and data-movements. FFTW is now released with Cell support, which allows software developers on Cell to use FFTW with the same portability as other platforms.
- **NAS FT benchmark:** This project changed the MPI NAS FT code to use the LAPI communication interface by IBM instead of MPI calls. The Low-level Application Programming Interface (LAPI) is a low level communication interface for the IBM Scalable Powerparallel (SP) super-computer Switch and provides one-sided communication performance between tasks on IBM SP system. LAPI offers better message passing performance than MPI on small and medium size messages. The LAPI implementation was subsequently compared to a UPC/MPI implementation for performance. This work is part of a submission, that won the **most productive research implementation in the HPC Challenge Class 2 awards at Supercomputing 2007**.

Technical Co-op (Novel Systems Architecture Group)

January 2006 - July 2006

- **Optimization of next-generation processors for bioinformatics workloads:** The project aimed at uncovering limitations in present architectures and finding architectural trade-offs that can benefit computational biology workloads without degrading performance of traditional supercomputing workloads. As a first step, the bottlenecks of representative bioinformatics applications was analyzed from the *BioPerf* suite (read below) at both the instruction and memory level, on current Power processors. This data was used to design future systems and processors (simulated through the IBM proprietary simulator *SystemSim*) for better computational biology applications performance. Adding conditional assignments to the PowerPC ISA, removing the two-cycle branch penalty currently in the Power5 architecture, as well as the impact of changing the number of function units were some of the changes simulated. The new instructions were inserted, both through hand-inserted assembly and compiler support, and the combined effect of these changes was 64% on an average, for sequence homology applications.

Affiliate, IBM PERCS

Advisor: Dr. David A. Bader

September 2005 - December 2005

Georgia Institute of Technology, Atlanta, GA, USA

Research Assistant, IBM PERCS

Advisor: Dr. David A. Bader

Jan. 2004 - August 2005

University of New Mexico, Albuquerque, NM, USA

- **Creation of *BioPerf* benchmark suite to evaluate high-performance computer architectures:** This project created and maintains a benchmark suite of representative bioinformatics applications *BioPerf* (**available from www.bioperf.org**) to facilitate the design and evaluation of high performance computer architectures for emerging computational biology workloads. The applications cover the broad biological problems of sequence alignment, phylogeny and protein structure prediction, spanning different algorithms and performance statistics. The benchmark suite includes source codes of fifteen executables, pre compiled Alpha, x86 and PowerPC binaries, and scripts for running, installing, compiling the source codes for other architec-

tures and for saving the outputs to appropriate directories.

- **Workload characterization and performance analysis of Bioinformatics workload *BioPerf*:** This project analyzed the workload *BioPerf* thoroughly at the instruction and memory level on the PowerPC G5 (PowerPC 970) processor to give an assessment of the limitations of present-day microprocessors for computational biology applications. In addition to reporting the aggregate performance characteristics at the end of each run, we also analyze the “live-graph” performance data which show the variation of the performance exhibited during the execution of the application. (*Collaborations: Ram Rajamony, Manager, Novel System Architectures Group, IBM Research Labs, Austin*)
- **Maximum flow problem on shared-memory multiprocessors:** This project implemented the push-relabel network flow algorithm on present-day symmetric multiprocessors (SMPs) with large shared memories. A major contribution of this project was the new cache-aware implementation of Goldberg’s parallel algorithm for modern shared-memory parallel computers, and a parallel implementation and analysis of the gap relabeling heuristic. The implementation was performed on Sun E4500 SMP with 14 processors, with programming in POSIX threads, an in house library called SIMPLE and C language with notable speedups.

Undergraduate Thesis Research

Advisor: Dr. M. K. Das

August 2001 - April 2002

Indian Institute of Technology , Guwahati, Assam, India

- **Parallelization of Delaunay Triangulation and Linear Algebra Problems:** This work designed and implemented a parallel algorithm for Delaunay triangulation that works on general distributions. The method of reducing a two-dimensional delaunay triangulation to find the three-dimensional convex hull of points on a paraboloid was used. The algorithm was implemented using C and MPI-based toolkit, speedups of four to five times was achieved compared to the best sequential algorithm. Experiments were performed on PARAM 10000 parallel computer and later on the SGI ORIGIN 2000.

Project Assistant

Advisor: Dr. Ravi S. Nanjundiah

July 2002 - July 2003

Indian Institute of Science, Bangalore, Karnataka, India

- **Second generation wavelet based solution of shallow water equations :** A second generation wavelet solution for solution to the shallow water equations was designed. This approach could generate automatic grid adaption in regions of sudden turbulence. (*Collaborations: Dr. A. S. Vasudevamurthy, Mathematics Deptt., TIFR Bangalore*).

Summer Trainee

Advisor: Dr. Ravi S. Nanjundiah

May 2001 - July 2001

Indian Institute of Science, Bangalore, India

- **Validation and Parallelization of Numerical Solution of Shallow Water Equations:** Shallow Water equations contain both slow moving Rossby waves and high-frequency gravity waves. Thus explicit schemes to solve the equations require a very small time step to progress the solution. A semi-implicit scheme is used where the nonlinear parts are treated explicitly. This results in solving a Helmholtz equation. The code was parallelized using the OpenMP implementation on IBM SP3 (shared memory architecture) and ORIGIN 2000.

Winter Trainee

Advisor: Dr. Ravi S. Nanjundiah

Nov. 2001 - Dec. 2001

Indian Institute of Science , Bangalore, India

- **Implementation and parallelization of a wavelet based solution of Burgers and Helmholtz equations:** A wavelet galerkin approach for a solution to the one dimensional Burgers equations and Helmholtz equations was designed. Wavelets due to their localized nature are able to capture discontinuities better than other series. Such an approach also speeds up the solution considerably compared to a conventional finite difference approach. As a second step, the wavelet based solution was parallelized in C++ using MPICH.

PROFESSIONAL
ACTIVITIES

Invited Speaker: V. Sachdeva, “Opportunities and Challenges for Cell Broadband Engine in Life Sciences”, Minisymposia on *Unleashing the Power of the Cell Broadband Engine Processor for HPC*, SIAM Conference on Parallel Processing for Scientific Computing, Atlanta, GA, Mar. 12-14, 2008.

Invited Speaker: V.Sachdeva, “Evaluating the Cell from a Bioinformatics Perspective”, Birds of Feather on *Unleashing the Power of the Cell Broadband Engine Processor for HPC*, 2007 Supercomputing Conference (SC2007), Reno, NV, Nov. 10-16, 2007.

Conference Tutorial: P. Bohrer, M. Kistler, D. Murrell, V. Sachdeva, “Using Systemsim to Guide Application Transformation and Tuning for the Cell Broadband Engine”, The 12th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS 2006), San Jose, CA, Oct. 21-25, 2006.

PUBLICATIONS

V. Sachdeva, M. Kistler, E. Speight, Tzy Hwa K. Tzeng, “Exploring the viability of the Cell Broadband Engine for bioinformatics applications”, *Parallel Computing (Journal)*.

V. Sachdeva, E. Speight, M. Stephenson, L. Chen, “Characterizing and Improving the Performance of Bioinformatics Workloads on the POWER5 Architecture”, 2007 International Symposium on Workload Characterization (IISWC 2007), Boston, MA, Sept. 27-29, 2007.

V. Sachdeva, M. Kistler, E. Speight, Tzy Hwa K. Tzeng, “Exploring the viability of the Cell Broadband Engine for bioinformatics applications”(Invited Paper), The 6th IEEE International Workshop on High Performance Computational Biology (HiCOMB 2007), Long Beach, CA, Mar. 26, 2007.

V. Sachdeva, M. Kistler, Evan Speight, “Pairwise Alignments on the Cell Processor”(Poster Session), 2006 IEEE and ACM Supercomputing Conference (SC2006), Tampa, FL, Nov. 11-17, 2006.

D. A. Bader, V. Sachdeva, “An Open Benchmark Suite for Evaluating Computer Architecture on Bioinformatics and Life Science Applications”, 2006 SPEC Benchmark Workshop, Jan. 23, 2006.

D. A. Bader, Tao Li, Yue Li, V. Sachdeva, “BioPerf: A Benchmark Suite to Evaluate High-Performance Computer Architecture on Bioinformatics Applications”, 2005 International Symposium on Workload Characterization (IISWC 2005), Austin, TX, Oct. 6-8, 2005.

D.A. Bader, V. Sachdeva, “A Portable Shared-Memory Implementation of the Push-Relabel Network Flow Algorithm with Global and Gap Relabeling”, The 18th ISCA International Conference on Parallel and Distributed Computing Systems (PDCS 2005), Las Vegas, NV, September 12-14, 2005.

D.A. Bader, V. Sachdeva, A. Trehan, V. Agarwal, G. Gupta, and A.N. Singh, “BioSPLASH: A sample workload from bioinformatics and computational biology for optimizing next-generation high-performance computer systems”(Poster Session), 13th Annual International Conference on In-

telligent Systems for Molecular Biology (ISMB 2005), Detroit, MI, June 25-29, 2005.

D.A. Bader, V. Sachdeva, "Incorporating life sciences applications in the architectural optimizations of next-generation petaflop-system"(Poster Session), The 4th IEEE Computational Systems Bioinformatics Conference (CSB 2005), Stanford University, CA, August 8-11, 2005.

D. A. Bader, V. Sachdeva, V. Agarwal, G. Gupta and A. N. Singh, "Life Science Applications", IBM/DARPA High Productivity Computing Systems (HPCS) Project Meeting, Austin, TX, February 8, 2005.

D.A. Bader, V. Sachdeva, V. Agarwal, G. Goel, A.N. Singh, "BioSPLASH: A sample workload for bioinformatics and computational biology for optimizing next-generation performance computer systems", Technical Report, University of New Mexico, May 2005.

REVIEWER

- 2007 IEEE and ACM Supercomputing Conference (SC2008), Austin, TX, Nov. 15-21, 2008.
- BioMed Central Bioinformatics Journal.
- The 2008 International Conference on Parallel Processing (ICPP 2008), Portland, OR, Sept. 8-12, 2008.
- Special Issue of the Journal of Parallel and Distributed Computing: General-Purpose Parallel Processing Using GPUs
- 2007 IEEE and ACM Supercomputing Conference (SC2007), Reno, NV, Nov. 10-16, 2007.
- The 5th International Symposium on Image and Signal Processing and Analysis (ISPA 2007), Istanbul, Turkey, Sept. 27-29 2007.
- The 6th IEEE/ACM International Workshop on Grid Computing (GRID 2005), Seattle, WA, Nov. 13-14, 2005.
- The 2006 International Conference on Parallel Processing (ICPP 2006), Columbus, OH, Aug. 14-18, 2006.

HONORS AND MEMBERSHIPS

- Selected for the 10th Annual San Diego Summer Institute organized by San Diego Supercomputing Center on data-intensive computing.(Aug 23rd - Aug 27th) 2004.
- Member, International Society for Computational Biology (ISCB) 2005-2006.
- ISCB Student Council Leader 2005-2006 (reviewer of posters for student symposium at ECCB 2005).
- Member, IEEE Computer Society 2005-2006.
- Ranked in the top 1% among the students writing the IIT entrance test 1998.

COMPUTER SKILLS

- Languages: C, C++, Perl, POSIX threads, OpenMP, MPI, Unix shell scripts, MATLAB.
- Performance Analysis Packages and Simulators: Mambo (IBM Austin), CHUD (Apple G5), Hpmcount (IBM), Sun One / Forte / Workshop (Sun).
- Operating Systems: Unix/Linux, Solaris, AIX, Windows.