Curriculum Vitae

Yongchao Liu, PhD
Research Scientist II
School of Computational Science & Engineering
Georgia Institute of Technology, Atlanta, GA 30332, USA
[Affiliation] [Google Scholar] [Scopus] [LinkedIn] [ResearchGate]

Personal Particulars		
	Gender:	Male
	Nationality:	China (U.S. EB1 green card petition approved)
	Languages:	Chinese (native), English(proficient)
	Filed of Specialty:	Parallel computing, Big data and Bioinformatics
	Homepage:	http://cc.gatech.edu/∼yliu
	Email:	yliu860@gatech.edu; yliu@cc.gatech.edu
	Mailing Address:	1341 Klaus Advanced Computing Building
		266 Ferst Drive, Atlanta, GA 30332, USA

Research Interests

Accelerators (e.g. FPGAs[2], GPUs and Xeon Phis), Big data, Bioinformatics, High performance computing, Parallel computing, and Parallelized machine learning

Brief Biography

Dr. Yongchao Liu (刘永超博士) is a Research Scientist II (a research faculty member working with Dr. Srinivas Aluru) in the School of Computational Science & Engineering, Georgia Institute of Technology (USA). Before that, he worked as a postdoctoral researcher (working with Dr. Bertil Schmidt) at the Institute of Computer Science, University of Mainz (Germany). He earned his **Ph.D. degree in computer engineering** from Nanyang Technological University (Singapore) in 2012 (supervised by Dr. Bertil Schmidt and Dr. Douglas Maskell). Prior to that, he earned the Master and Bachelor degrees in computer science and technology from Nankai University (China) in 2008 and 2005, respectively. His **research interests** focus on parallel and distributed algorithm design for bioinformatics, heterogeneous computing with accelerators (GPUs and Xeon Phi coprocessors), high performance computing on big data, and parallelized machine learning.

His **research philosophy** is to inspire technological innovation for health care and serve people around the globe. He is an active researcher in parallel computing and bioinformatics and his outstanding contributions to this interdisciplinary field are demonstrated in his novel parallel algorithms and software tools for large-scale biological data analysis. These parallel algorithms and tools streamline fundamental and computationally challenging biological issues with parallel computing, via indepth exploration of current high performance computing techniques and technologies such as hardware accelerators (e.g. GPUs and Intel Xeon Phis) and clusters (e.g. CPU/GPU/Xeon Phi clusters). On the other hand, these algorithms and tools investigate a set of related critical issues in bioinformatics and have actually established an innovative and unique analysis platform for large-scale biological datasets, especially next-generation sequencing read datasets. To address big data challenges, he proposed a new concept of computing, i.e. Compact Computing, which centers around data and enables full-stack computation by tightly coupling algorithms with systems.

He has published 26 journal papers, 17 conference papers and 5 book chapters, with > 1500 citations totally. Out of all journal papers, 23 are published in SCI jour-

---

[1]Last updated: May 16, 2017
[2]My Master's thesis researches on the design and implementation of FPGA-based system on chip (SoC) for portable e-readers.

nals (with > 76 impact factors in sum), 13 in SCI journals of > 3.0 impact factor each, and 7 in the top `Bioinformatics` journal (2016 impact factor 5.766). He has released a set of software tools for reproducible research and public use (already gained $> 20,000$ downloads). Among these algorithms, three CUDA-based open-source software tools, i.e. CUDASW++, mCUDA-MEME and CUSHAW, are `rated by NVIDIA Corporation as` `popular GPU-accelerated applications`, while DecGPU (the first parallel and distributed error correction algorithm for high-throughput short reads) was `reported by` `GenomeWeb`. Furthermore, his other open-source software algorithms are also leading in their respective areas, including MSAProbs (multiple protein sequence alignment), Musket (Illumina NGS read error correction), CUSHAW2 (NGS base-space read alignment), CUSHAW3 (NGS base-space and color-space read alignment), PASHA (NGS de novo genome assembly), SNVSniffer (NGS germline and somatic SNV calling), SWAPHI (Xeon-Phi-based protein database search), SWAPHI-LS (Xeon-Phi-based pairwise DNA sequence alignment), ParaBWT (parallel construction of Burrows-Wheeler transform and suffix array), LightSpMV (CUDA-based sparse matrix-vector multiplication) and LightPCC(parallel correlation computation on Xeon Phi clusters).

He won two Best Paper Awards from IEEE International Conference on Application-specific Systems, Architectures and Processors in 2009 and 2015, got one paper recommended for Best Paper Award from IEEE International Conference on Cluster Computing in 2014, and won the Program to Empower Partnerships with Industry Award from U.S. South Big Data Hub in 2016. Moreover, he is listed in the reputable Who's Who in America (Marquis research) in 2016. He co-chaired the first Workshop on Parallel Software Libraries for Sequence Analysis (pSALSA) in 2015 and founded the Workshop on Accelerator-Enabled Algorithms and Applications in Bioinformatics (WACEBI) in 2016. In addition, he serves as a reviewer for some leading journals, such as Nature Methods, Nature Communications, Bioinformatics, TPDS and TCBB, and as a program committee member for some leading conferences such as IPDPS, CCGrid and HiPC.

Education	**Nanyang Technological University**, Singapore

Ph.D, Computer Engineering $\hspace{6cm}$ $2008-2012$
Thesis: Design and implementation of parallel bioinformatics algorithms on heterogeneous computing architectures

Nankai University, China
Master, Computer Science and Technology $\hspace{4cm}$ $2005-2008$
Thesis: 嵌入式屏幕印刷格式多重解析及高速实现研究(*High-speed and multifaceted screen printing format parser on embedded systems) [*translated by myself].

Nankai University, China
Bachelor, Computer Science and Technology $\hspace{4cm}$ $2001-2005$

Proficient Computer Skills

C/C++, CUDA C/C++, SIMD vectorized assemblies, OpenMP, Pthreads, MPI, Linux

Work Experience

Research Scientist II $\hspace{6cm}$ $01/2015-\text{present}$
School of Computational Science & Engineering, Georgia Institute of Technology, USA

Postdoctoral researcher $\hspace{6cm}$ $11/2011-01/2015$
Institute of Computer Science, University of Mainz, Germany

Selected Awards and Honors

- Who's Who in America (Marquis research) 2016
- The Program to Empower Partnerships with Industry (PEPI) Award (U.S. South Big Data Regional Innovation Hub), 2016
- Best Paper Award (26th IEEE International Conference on Application-specific Systems, Architectures and Processors), 2015
- Best Paper Award Recommendation (2014 IEEE International Conference on Cluster Computing), 2014
- Best Paper Award (20th IEEE International Conference on Application-specific Systems, Architectures and Processors), 2009
- Second Prize of 100 Projects of Creative Research for Undergraduates of Nankai University, China, 2005
- Third-Class Award of Tianjin Academic Students Mathematics Contest, China, 2003-2004

Awarded Grants
- Yongchao Liu (2016) *Compressive computing for big data applications based on accelerators.* South Big Data Hub PEPI Grant. `Amount`: $11,000. **Role**: *principal investigator*
- Yongchao Liu (2016) *Donation to my [WACEBI 2016](#) workshop as the Best GPU Paper/Talk Award prize.* NVIDIA Academic Hardware Grant. One GeForce GTX Titan X graphics card (`Amount`: $1,000). **Role**: *principal investigator*
- Yongchao Liu (2015) *Parallel building blocks for large-scale sequence analysis.* NVIDIA Academic Hardware Grant. One GeForce GTX Titan X graphics card (`Amount`: $1,000). **Role**: *principal investigator*
- Yongchao Liu (2015) *Parallel assembler for metagenomic sequence data from next-generation sequencing.* NSF Extreme Science and Engineering Discovery Environment (**XSEDE**) Startup Grant. `Amount`: 50,000 SUs and 3,000 GB storage. **Role**: *principal investigator*
- Srinivas Aluru (2015) Intel Parallel Computing Center. `Amount`: $400,000. **Role**: *co-investigator*

Publications [3] [4] [5] [6]

Year 2017

J29 Yongchao Liu[‡], Tony Pan, Oded Green and Srinivas Aluru: "Parallelized Kendall's tau coefficient computation via SIMD vectorized sorting on many-integrated-core processors". **Journal of Parallel and Distributed Computing** (impact factor 1.320), 2017, under review [arXiv].

J28 Chao Liu, Tao Li, Yongchao Liu and Ying Zhang: "OrthoBBH: a parallel and distributed solution for orthology clustering". **BMC Bioinformatics** (impact factor 2.435), to be submitted.

---

[3][‡] indicates I am a corresponding author.
[4]B∗ refers to book, C∗ conference/workshop J∗ journal, P∗ poster and I∗ informal publications.
[5]Unless otherwise specified, *acceptance rate* refers to the rate of regular/long papers.
[6]As of April 04, 2017, I have got 1,506 citations (h-index 18 and i10-index 21) from Google Scholar.

J27 Tony C Pan, Patrick Flick, Chirag Jain, <u>Yongchao Liu</u> and Srinivas Aluru: "Kmerind: a flexible parallel library for k-mer indexing of biological sequences on distributed memory systems". **IEEE/ACM Transactions on Computational Biology and Bioinformatics** (impact factor 1.609), under review.

C18 Tony Pan, Rahul Nihalani, <u>Yongchao Liu</u> and Srinivas Aluru: "Fast de Bruijn graph compaction in distributed memory environments". 2017 International Conference for High Performance Computing, Networking, Storage and Analysis (**SC 2017**), under review.

C17 Yuandong Chan, Kai Xu, Haidong Lan, Weiguo Liu, <u>Yongchao Liu</u> and Bertil Schmidt: "PUNAS: a parallel ungapped-alignment-featured seed verification for next-generation sequencing read alignment". 31st IEEE International Parallel & Distributed Processing Symposium (**IPDPS 2017**), 2017, in press.

C16 Haidong Lan, Weiguo Liu, <u>Yongchao Liu</u> and Bertil Schmidt: "SWhybrid: a hybrid parallel framework for large-scale protein sequence database search". 31st IEEE International Parallel & Distributed Processing Symposium (**IPDPS 2017**), 2017, in press.

B5 <u>Yongchao Liu</u> and Bertil Schmidt: "CUSHAW Suite: parallel and efficient algorithms for next generation sequencing read alignment". **Algorithms for Next-Generations Sequencing Data: Techniques, Approaches and Applications**, edited by Mourad Elloumi, Springer, 2017, in press.

J26 <u>Yongchao Liu</u>, Fabian Ripp, Rene Koeppe, Hanno Schmidt, Lukas Hellmann, Mathias Weber, Christopher Felix Krombholz, Bertil Schmidt and Thomas Hankeln: "AFS: identification and quantification of species composition by metagenomic sequencing". **Bioinformatics** (impact factor 5.766), 2017, 33 (9): 1396-1398

J25 <u>Yongchao Liu</u>‡ and Bertil Schmidt: "LightSpMV: faster CUDA-compatible sparse matrix-vector multiplication using compressed sparse rows". **Journal of Signal Processing Systems** (impact factor 0.508), 2017, doi:10.1007/s11265-016-1216-4. (`citations:1`)

J24 Sharma V. Thankachan, Sriram P. Chockalingam, <u>Yongchao Liu</u>, Ambujam Krishnan and Srinivas Aluru: "A greedy alignment-free distance estimator for phylogenetic inference". **BMC Bioinformatics** (impact factor 2.58), 2017, in press.

Year 2016

J23 Jorge Gonzalez-Dominguez, <u>Yongchao Liu</u>, Juan Tourino and Bertil Schmidt: "MSAProbs-MPI: parallel multiple sequence aligner for distributed-memory systems". **Bioinformatics** (impact factor 5.766), 2016, 32(24): 3826-3828 (`citations:1`)

J22 <u>Yongchao Liu</u>‡, Martin Loewer, Srinivas Aluru and Bertil Schmidt: "SNVSniffer: an integrated caller for germline and somatic single-nucleotide and indel mutations". **BMC Systems Biology** (impact factor 2.435), 2016, 10(suppl 2): 47 (`citations:3`)

J21 Jorge González-Domínguez, <u>Yongchao Liu</u> and Bertil Schmidt:"Parallel and scalable short-read alignment on multi-core clusters using UPC++". **PLoS ONE** (impact factor 3.234), 2016, 11(1): e0145490 (`citations:4`)

J20 Tuan Tu Tran, <u>Yongchao Liu</u> and Bertil Schmidt: "Bit-parallel approximate pattern matching: Kepler GPU versus Xeon Phi". **Parallel Computing** (impact factor 1.511), 2016, 54: 128-138. (`citations:6`)

J19 Sharma V. Thankachan, Sriram P. Chockalingam, <u>Yongchao Liu</u>, Alberto Apostolico and Srinivas Aluru: "ALFRED: a practical method for alignment-free distance computation". **Journal of Computational Biology** (impact factor 1.737), 2016, 23(6):452-460. (`citations:5`)

J18 Yongchao Liu‡, Thomas Hankeln and Bertil Schmidt: "Parallel and space-efficient construction of Burrows-Wheeler transform and suffix array for big genome data". **IEEE Transactions on Computational Biology and Bioinformatics** (impact factor 1.536), 2016, 13(3): 592 -598. (`citations:5`)

C15 Yongchao Liu, Tony C Pan and Srinivas Aluru, "Parallel pairwise correlation computation on Intel Xeon Phi clusters". 28th International Symposium on Computer Architecture and High Performance Computing (**SBAC-PAD 2016**) (acceptance rate 35.1%), 2016, pp. 141-149. (`citations:3`)

C14 Tony C Pan, Patrick Flick, Chirag Jain, Yongchao Liu and Srinivas Aluru: "Kmerind: a flexible parallel library for k-mer indexing of biological sequences on distributed memory systems". 7th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (**ACM-BCB 2016**) (acceptance rate 27.7%), 2016, pp. 422-433. (`citations:3`)

 I4 Yongchao Liu: "Research statement for assistant professorship application". **ResearchGate**, 2016, doi: 10.13140/RG.2.2.28828.90243/1.

 I3 Yongchao Liu‡ and Srinivas Aluru: "LightScan: faster scan primitive on CUDA compatible manycore processors". arXiv:1604.04815 [cs.DC], 2016

Year 2015

J17 Yongchao Liu‡ and Bertil Schmidt: "GSWABE: faster GPU-accelerated sequence alignment with optimal alignment retrieval for short DNA sequences". **Concurrency and Computation: Practice and Experience** (impact factor 0.784), 2015, 27: 958-972 (`citations:10`)

 B4 Yongchao Liu and Bertil Schmidt: "Pairwise DNA sequence alignment optimization". **High Performance Parallelism Pearls Volume Two - Multicore and Manycore Programming Approaches**, edited by James Reinders and Jim Jeffers, Elsevier, 2015, pp. 43-54

C13 Yongchao Liu, Martin Loewer, Srinivas Aluru, Bertil Schmidt: "SNVSniffer: an integrated caller for germline and somatic SNVs based on Bayesian models". 2015 IEEE International Conference on Bioinformatics and Biomedicine (**BIBM 2015**) (acceptance rate 19.6%), 2015, pp. 83-90. (`citations:2`)

C12 Sharma V. Thankachan, Sriram P. Chockalingam, Yongchao Liu, Ambujam Krishnan, Srinivas Aluru: "A greedy alignment-free distance estimator for phylogenetic inference (extended abstract)". 5th IEEE International Conference on Computational Advances in Bio and Medical Sciences (**ICCABS 2015**) (acceptance rate 47.9%), 2015 (`citations:2`)

C11 Yongchao Liu and Bertil Schmidt: "LightSpMV: faster CSR-based sparse matrix-vector multiplication on CUDA-enabled GPUs". 26th IEEE International Conference on Application-specific Systems, Architectures and Processors (**ASAP 2015**) (acceptance rate 24.7%), 2015, pp. 82-89 (**Best Paper Award**) (`citations:13`)

 P3 Yongchao Liu, Jorge González-Domínguez, Bertil Schmidt: "Faster compressed sparse row (CSR)-based sparse matrix-vector multiplication using CUDA". GPU Technology Conference 2015 (**GTC 2015**), San Jose, USA, 2015

 I2 Yongchao Liu: "OpenGraphAssembly: abstract, modularize and parallelize fundamental building blocks for graph-based genome assembly". **ResearchGate**, 2015, doi: 10.13140/RG.2.2.11212.82568.

Year 2014

J16 Fabian Ripp, Christopher F Krombholz, <u>Yongchao Liu</u>, Mathias Weber, Anne Schäfer, Bertil Schmidt, Rene Köppel and Thomas Hankeln: "All-Food-Seq (AFS): a quantifiable screen for species in biological samples by deep DNA sequencing". **BMC Genomics** (impact factor 4.40), 2014, 15:639 [reported by Australian Food News] (citations:9)

J15 Adrianto Wirawan, Robert S Harris, <u>Yongchao Liu</u>, Bertil Schmidt and Jan Schröder: "HECTOR: A parallel multistage homopolymer spectrum based error corrector for 454 sequencing data." **BMC Bioinformatics** (impact factor 3.02), 2014, 15:131 (citations:20)

J14 <u>Yongchao Liu‡</u>, Bernt Popp and Bertil Schmidt: "CUSHAW3: sensitive and accurate base-space and color-space short-read alignment with hybrid seeding." **PLoS ONE** (impact factor 3.730), 2014, 9(1): e86869 (citations:25)

J13 <u>Yongchao Liu‡</u> and Bertil Schmidt: "CUSHAW2-GPU: empowering faster gapped short-read alignment using GPU computing". **IEEE Design & Test** (impact factor 1.623), 2014, 31(1): 31-39 (citations:31)

B3 <u>Yongchao Liu</u> and Bertil Schmidt: "Multiple protein sequence alignment with MSAProbs". **Methods in Molecular Biology**, Edited by David Russell, Springer, 2014, 1079: 211-218 (citations:6)

C10 Tuan Tu Tran, Simon Schindel, <u>Yongchao Liu</u> and Bertil Schmidt: "Bit-Parallel approximate pattern matching on the Xeon Phi coprocessor". 26th International Symposium on Computer Architecture and High Performance Computing (**SBAC-PAD 2014**) (acceptance rate 32.3%), 2014, pp. 81-88 (citations:2)

C9 <u>Yongchao Liu</u> and Bertil Schmidt: "SWAPHI: Smith-Waterman protein database search on Xeon Phi coprocessors". 25th IEEE International Conference on Application-specific Systems, Architectures and Processors (**ASAP 2014**) (2-page short paper acceptance rate 57.6%), 2014, pp. 184-185 (citations:27)

C8 <u>Yongchao Liu</u>, Tuan-Tu Tran, Felix Lauenroth and Bertil Schmidt: "SWAPHI-LS: Smith-Waterman algorithm on Xeon Phi coprocessors for long DNA sequences". 2014 IEEE International Conference on Cluster Computing (**Cluster 2014**) (acceptance rate 23.8%), 2014, pp. 257-265 (Best Paper Award Recommendation) (citations:17)

P2 <u>Yongchao Liu</u> and Bertil Schmidt: "CUSHAW Software Package: harnessing CUDA-enabled GPUs for next generation sequencing read alignment". GPU Technology Conference 2014 (**GTC 2014**), San Jose, USA, 2014

Year 2013

J12 <u>Yongchao Liu‡</u>, Adrianto Wirawan and Bertil Schmidt: "CUDASW++ 3.0: accelerating Smith-Waterman protein database search by coupling CPU and GPU SIMD instructions". **BMC Bioinformatics** (impact factor 3.02), 2013, 14:117 (citations:104)

J11 <u>Yongchao Liu‡</u>, Jan Schröder and Bertil Schmidt: "Musket: a multistage k-mer spectrum based error corrector for Illumina sequence data". **Bioinformatics** (impact factor 5.323), 2013, 29(3): 308-315 (citations:102)

C7 <u>Yongchao Liu</u> and Bertil Schmidt: "Faster GPU-accelerated Smith-Waterman algorithm with alignment backtracking for short DNA sequences". 10th International Conference on Parallel Processing and Applied Mathematics (**PPAM 2013**) (workshop acceptance rate 85.7%), appear in Lecture Notes in Computer Science 8385, pp. 247-257 (citations:4)

I1 <u>Yongchao Liu‡</u>, Bernt Popp, and Bertil Schmidt: "High-speed and accurate color-space short-read alignment with CUSHAW2". arXiv:1304.4766 [q-bio.GN], 2013

Year 2012

J10 Yongchao Liu‡ and Bertil Schmidt: "Long read alignment based on maximal exact match seeds". **Bioinformatics** (impact factor 5.468), 2012, 28(18): i318-i324 (also from ECCB 2012) (`citations:70`)

J9 Yongchao Liu‡, Bertil Schmidt, and Douglas L. Maskell: "CUSHAW: a CUDA compatible short read aligner to large genomes based on the Burrows-Wheeler transform". **Bioinformatics** (impact factor 5.468), 2012, 28(14): 1830-1837 (`citations:101`)

C6 Yongchao Liu and Bertil Schmidt: "Long read alignment based on maximal exact match seeds". 11th European Conference on Computational Biology (**ECCB 2012**) (acceptance rate 14.1%), Basel, Switzerland (also published in the *Bioinformatics* journal) (`citations:70`)

C5 Yongchao Liu and Bertil Schmidt: Evaluation of GPU-based seed generation for computational genomics using Burrows-Wheeler transform". 26th IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum (**IPDPSW 2012**) (workshop acceptance rate 50.0%), 684-690 (`citations:6`)

P1 Yongchao Liu, Bertil Schmidt, and Douglas L. Maskell: "A fast CUDA compatible short read aligner to large genomes". GPU Technology Conference 2012 (**GTC 2012**), San Jose, USA, 2012

Year 2011

J8 Yongchao Liu‡, Bertil Schmidt and Douglas L. Maskell: "Parallelized short read assembly of large genomes using de Bruijn graphs". **BMC Bioinformatics** (impact factor 3.43), 2011, 12:354 (`citations:46`)

J7 Yongchao Liu‡, Bertil Schmidt and Douglas L. Maskell: "DecGPU: distributed error correction on massively parallel graphics processing units using CUDA and MPI". **BMC Bioinformatics** (impact factor 3.43), 2011, 12:85 [**reported by GenomeWeb**] (`citations:44`)

J6 Lakshmi Kuttippurathu, Michael Hsing, Yongchao Liu, Bertil Schmidt, Douglas L.Maskell, Kyungjoon Lee, Aibin He, William T. Pu and Sek Won Kong: "CompleteMOTIFs: DNA motif discovery platform for transcription factor binding experiments". **Bioinformatics** (impact factor 4.926), 2011, 27(5): 715-717 (`citations:39`)

B2 Łukasz Ligowski, Witold Rudnicki, Yongchao Liu, and Bertil Schmidt: "Accurate scanning of sequence databases with the Smith-Waterman algorithm". **GPU Computing Gems**, Edited by Wen-mei W. Hwu, Elsevier 2011, Pages 155-172 (`citations:9`)

C4 Weiguo Liu, Bertil Schmidt, Yongchao Liu, and Wolfgang Müller-Wittig: "Mapping of the BLASTP algorithm onto GPU clusters". 17th IEEE International Conference on Parallel and Distributed Systems (**ICPADS 2011**) (acceptance rate 25.7%), 2011, 236-243 (`citations:7`)

C3 Yongchao Liu, Bertil Schmidt, and Douglas L. Maskell: "An ultrafast scalable many-core motif discovery algorithm for multiple GPUs". 25th IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum (**IPDPSW 2011**) (workshop acceptance rate 52.3%), 428-434 (`citations:26`)

Year 2010

J5 Yongchao Liu‡, Bertil Schmidt and Douglas L. Maskell: "MSAProbs: multiple sequence alignment based on pair hidden Markov models and partition function posterior probabilities". **Bioinformatics** (impact factor 4.926), 2010, 26(16): 1958 -1964 (`citations:115`)

J4 Yongchao Liu[‡], Bertil Schmidt and Douglas L. Maskell: "CUDASW++2.0: enhanced Smith-Waterman protein database search on CUDA-enabled GPUs based on SIMT and virtualized SIMD abstractions". **BMC Research Notes**, 2010, 3:93 (`citations:189`)

J3 Yongchao Liu[‡], Bertil Schmidt, Weiguo Liu and Douglas L. Maskell: "CUDA-MEME: accelerating motif discovery in biological sequences using CUDA-enabled graphics processing units". **Pattern Recognition Letters** (impact factor 1.559), 2010, 31(14): 2170 - 2177 (`citations:68`)

B1 Yongchao Liu, Bertil Schmidt , and Douglas L. Maskell: "Parallel bioinformatics algorithms for CUDA-enabled GPUs". **Bioinformatics: High Performance Parallel Computer Architectures**, Edited by Bertil Schmidt, CRC Press 2010, Pages 117–137 (`citations:1`)

Year 2009

J2 Yongchao Liu[‡], Douglas L. Maskell and Bertil Schmidt: "CUDASW++: optimizing Smith-Waterman sequence database searches for CUDA-enabled graphics processing units". **BMC Research Notes**, 2009, 2:73 (`citations:257`)

C2 Yongchao Liu, Bertil Schmidt, and Douglas L. Maskell: "MSA-CUDA: multiple sequence alignment on graphics processing units with CUDA". 20th IEEE International Conference on Application-specific Systems, Architectures and Processors (**ASAP 2009**) (acceptance rate 35.7%), 2009, 121-128 (**Best Paper Award**) (`citations:84`)

C1 Yongchao Liu, Bertil Schmidt, and Douglas L. Maskell: "Parallel reconstruction of neighbor-Joining trees for large multiple sequence alignments using CUDA". 23th IEEE International Symposium on Parallel and Distributed Processing (**IPDPS 2009**) (workshop acceptance rate 55.6%), 2009, 1-8. (`citations:33`)

Year 2007

J1 李庆诚(Qingcheng Li), 刘永超(Yongchao Liu)[‡] and 刘嘉欣(Jiaxin Liu), 平台无关的PDF嵌入式高性价比解析器设计与实现 (Design and implementation of a platform-independent, high-performance-cost-ratio, embedded PDF format parser[*]). 计算机应用 (**Computer Applications**), 2007, 27(z1). [*translated by myself] (`citations:6`)

Software Tools [7]

Big data

- ParaBWT[⋆]
  a leading parallel and space-efficient algorithm for Burrows-Wheeler transform construction on big genome data. (`downloads:141`)

Machine learning

- LightPCC[⋆]
  the first parallel and distributed pairwise correlation computation library on Intel Xeon Phi clusters for data science (e.g. co-expression network construction and feature selection). (`downloads:45`)

---

[7][⋆]indicates open-source; totally `20,999` downloads (available statistics only) as of March 04, 2017

Scientific computing

- LightSpMV⋆
  a faster compressed sparse row (CSR)-based sparse matrix-vector multiplication algorithm on CUDA-enabled GPUs. (`downloads:459`)

- LightScan⋆
  a faster parallel scan primitive for CUDA-enabled GPUs by investigating a hybrid model combining intra-block computation and inter-block communication. (`downloads:160`)

Motif finding

- CUDA-MEME⋆
  a fast parallel motif finding algorithm based on MEME (version 3.5.4) algorithm for a single GPU device using CUDA.

- mCUDA-MEME⋆
  a further extension of CUDA-MEME based on MEME (version 4.4.0) algorithm for multiple GPUs using a hybrid combination of CUDA, MPI and OpenMP.

- CompleteMOTIFs
  an integrated web tool developed by Harvard Medical School to facilitate systematic discovery of over-represented transcription factor binding motifs from high-throughput chromatin immunoprecipitation experiments. I contributed CUDA-MEME to accelerate motif discovery.

Next-generation sequencing

Short-read alignment

- CUSHAW⋆
  the first distribution of the CUSHAW software package for NGS read alignment. It is a CUDA compatible short read alignment algorithm for multiple GPUs sharing a single host. This aligner only provides support for ungapped alignment and has been incorporated to NVIDIA Tesla Bio Workbench. (`downloads:1,955`)

- CUSHAW2⋆
  the second distribution of the CUSHAW software package for NGS read alignment. It is a fast and parallel gapped read alignment to large genomes, such as the human genome. This aligner has been further accelerated using GPU computing and is implemented in CUSHAW2-GPU. (`downloads:2,260`)

- CUSHAW3⋆
  the third distribution of the CUSHAW software package for NGS read alignment. It is a parallel, sensitive and accurate short-read aligner for both base-space and color-space single-end/paired-end reads.This aligner has been further enhanced using cluster computing and is implemented in CUSHAW3-UPC.(`downloads:1,011`)

Short-read error correction

- DecGPU⋆
  the first parallel and distributed pre-assembly short read error correction algorithm using CUDA and MPI.(`downloads:637`)

- Musket⋆
  a parallel and scalable multistage k-mer spectrum based error corrector for Illumina sequence data. (`downloads:1,594`)

- Hector⋆
  a parallel multistage homopolymer spectrum based error corrector to handle homopolymer insertions or deletions in 454 sequencing data. (`downloads:240`)

Short-read assembly

- PASHA⋆
  a parallelized short read assembler for large genomes, such as the human genome, using de Bruijn graphs. (`downloads:773`)

SNV calling

- SNVSniffer
  an integrated caller for germline and somatic single nucleotide variants (SNVs) in diploid genomes. (`downloads:689`)

Metagenomics

- All-Food-Seq
  a software pipeline for quantitative measurement of species composition in foodstuff material. (`downloads:154`)

Sequence alignment

- CUDASW++⋆
  the fastest parallel Smith-Waterman protein database search algorithm for GPGPUs using CUDA. (`downloads:6,845`)
- MSAProbs⋆
  one of the most accurate multiple protein sequence alignment algorithms and also parallelized using multithreading for multi-CPU systems.(`downloads:3,278`)
- SWAPHI⋆
  the first parallel algorithm to accelerate the Smith-Waterman protein database search on Xeon Phi coprocessors. (`downloads:270`)
- SWAPHI-LS⋆
  the first parallel Smith-Waterman algorithm exploiting Xeon Phi clusters to accelerate the alignment of long DNA sequences. (`downloads:341`)
- XBitPar⋆
  a Bit-parallel approximate pattern matching algorithm that is based on the Wu-Manber algorithm and further accelerated by Xeon Phi coprocessors. (`downloads:147`)

Projects (2008-present)

Big Biological Data Analysis

The objective of this project is to research and develop parallel and distributed algorithms for bioinformatics, by employing a variety of tightly-coupled and loosely-coupled computing architectures, including heterogeneous computers with accelerators (e.g. Intel SSE, Intel AVX, Intel Xeon Phis, NVIDIA GPUs and AMD GPUs), cluster computing and cloud computing. The ultimate goal is to establish an analysis system for large-scale biological sequences to solve a series of pressing yet demanding problems in bioinformatics and computational biology, including but not limited to

next-generation sequencing, meta-genomics, motif discovery, sequence alignment, and phylogenetic inference.

Full-text Indexing and Pattern Search

The objective of this project is to design parallel and memory-efficient algorithms for full-text indexing and pattern search on heterogeneous computing environments comprising multi-core CPUs and accelerators (NVIDIA GPUs, AMD GPUs and Intel Xeon Phi coprocessors). In particular, this project will investigate at least four popular full-text indexing data structures: Burrows-Wheeler transform, FM-index, suffix array and enhanced suffix array. Based on these indexing data structures, it will further explore full-text pattern search, e.g. searching for maximal exact matches, super maximal exact matches, and approximate pattern matches, which is fundamental and critical to a myriad of applications in various fields such as bioinformatics and text/data mining.

Parallel Building Blocks With Heterogeneous Computing

The objective of this project is to build a parallel building block template library for heterogeneous computing. This library will contain popular fundamental building blocks for scientific computing and data science, e.g. sparse linear algebra, scan, reduction, sort, k-nearest neighbors and etc. As of today, some preliminary work has been done to accelerate sparse matrix-vector multiplication using NVIDIA GPUs for scientific computing, and pairwise correlation/dependence measures using Intel Xeon Phi clusters for data science and machine learning.

Professional Service

Membership

- IEEE Membership
- ACM SIGBio
- ACM SIGHPC

Conference/workshop organization

- Organizer and Co-chair, 2016 Workshop in Accelerator-Enabled Algorithms and Applications in Bioinformatics (WACEBI 2016), in conjunction with IEEE BIBM 2016 conference, 2016
- Co-chair, Workshop on Parallel Software Libraries for Sequence Analysis (pSALSA), in conjunction with ACM-BCB 2015 conference, 2015

Conference/workshop program committees

- 13th EAI International Conference on Collaborative Computing: Networking, Applications and Worksharing (**COLLABORATECOM 2017**)
- 8th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (**ACM-BCB 2017**) (Sequence analysis and genome assembly area)
- 11th International Conference on Frontier of Computer Science and Technology (**FCST-2017**)

- 8th International Workshop on Biological Knowledge Discovery and Data Mining (**BIOKDD 2017**) in conjunction with **DEXA 2017**
- 31st IEEE International Parallel and Distributed Processing Symposium (**IPDPS 2017**)
- 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid (**CCGrid 2017**) (Posters and research demos committee)
- 1st Workshop of Mission-Critical Big Data Analytics (**MCBDA 2016**)
- 16th International Conference on Algorithms and Architectures for Parallel Processing (**ICA3PP 2016**)
- 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (**CCGrid 2015**) (Posters committee)
- 2015 Workshop on Parallel Computational Biology (in conjunction with **PPAM 2015**)
- 20th IEEE International Conference on High Performance Computing (**HiPC 2015**)
- 15th International Conference on Algorithms and Architectures for Parallel Processing (**ICA3PP 2015**)
- 29th IEEE International Parallel & Distributed Processing Symposium (**IPDPS 2015**)
- 6th IEEE International Conference on Cloud Computing Technology and Science (**CloudCom 2014**)
- 19th IEEE International Conference on Parallel and Distributed Systems (**ICPADS 2013**)
- 2013 Workshop on Parallel Computational Biology (in conjunction with **PPAM 2013**)
- 18th IEEE International Conference on Parallel and Distributed Systems (**ICPADS 2012**)
- 4th Workshop on Using Emerging Parallel Architectures (in conjunction with **ICCS 2012**)

Invited journal reviewing

- Nature Methods
- Nature Communications
- ACM Computing Surveys
- Nucleic Acids Research
- Bioinformatics
- Briefings in Bioinformatics
- Genome Biology
- PLOS One
- BMC Bioinformatics
- BMC Genomics
- ACM Transactions on Parallel Computing
- ACM Transactions on Architecture and Code Optimization
- IEEE Transactions on Parallel and Distributed Systems
- IEEE/ACM Transactions on Computational Biology and Bioinformatics
- IEEE Transactions on Very Large Scale Integration Systems
- Parallel Computing
- Journal of Computational Science

- Journal of Parallel and Distributed Computing
- GigaScience
- The Journal of Signal Processing Systems
- BioMed Research International
- Computers in Biology and Medicine
- Microelectronics Journal
- EURASIP Journal on Wireless Communications and Networking
- International Journal of Data Mining and Bioinformatics
- International Journal of High Performance Computing and Networking
- International Journal of High Performance Computing Applications
- Knowledge and Information Systems
- Knowledge-Based Systems
- PeerJ
- BMC Research Notes

Invited conference/workshop reviewing

- CCGrid 2013, 2015, 2016, 2017
- CloudCom 2014
- Euro-Par 2017
- HiPC 2015
- ICA3PP 2015, 2016
- ICPADS 2012, 2013
- IPDPS 2015, 2016, 2017
- PPAM 2013, 2015
- SC 2015, 2016
- ACM PASC 2016

Invited Talks

- *High-performance bioinformatics algorithms: bridging big data and big compute for biological data analysis*, at **State University of New York at Binghamton**, Binghamton, USA, 03/2017

- *High-performance bioinformatics algorithms: bridging big data and big compute for biological data analysis*, at **Fairfield University**, Fairfield, USA, 01/2017

- *High-performance bioinformatics algorithms: bridging big data and big compute for biological data analysis*, at **Seven Bridges Genomics** (a biomedical data analysis company accelerating breakthroughs in genomics research for cancer, drug development and precision medicine), Boston, USA, 01/2017

- *An integrated solution to call germline and somatic single-nucleotide and indel mutations*, at 6th Workshop on Computational Advances for Next Generation Sequencing (**CANGS 2016**) in conjunction with 6th IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS), USA, 10/2016

- Invited speaker at South Big Data Hub Workshop on High Impact Applications of Data Science in Precision Medicine, Health Analytics, and Health Disparities, at **Georgia Institute of Technology**, USA, 09/2016

- *Gene prediction and genome annotation*, at **University of Mainz**, Germany, 10/2013

- *Parallel and accurate short-read alignment for next-generation sequencing*, at **University of Mainz**, Germany, 10/2013
- *PASHA: a parallelized short-read assembler for large genomes*, at **University of Cologne**, Germany, 06/2013
- *Parallel and accurate gapped alignment for next-generation sequencing reads*, at **Tianjin Polytechnic University**, China, 10/2012

Conference Presentations

- *Parallel pairwise correlation computation on Intel Xeon Phi clusters*, at 28th International Symposium on Computer Architecture and High Performance Computing (**SBAC-PAD 2016**), Los Angeles, USA, 10/2016
- *SNVSniffer: an integrated caller for germline and somatic SNVs based on Bayesian models*, at 2015 IEEE International Conference on Bioinformatics and Biomedicine (**BIBM 2015**), Washington DC, USA, 11/2015
- *SWAPHI: Smith-Waterman protein database search on Xeon Phi coprocessors*, at 25th IEEE International Conference on Application-specific Systems, Architectures and Processors (**ASAP 2014**), Switzerland, 06/2014
- *Long read alignment based on maximal exact match seeds*, at 11th European Conference on Computational Biology (**ECCB 2012**), Switzerland, 09/2012
- *Evaluation of GPU-based seed generation for computational genomics using Burrows-Wheeler transform*, at 26th IEEE International Parallel & Distributed Processing Symposium (**IPDPS 2012**), China, 05/2012

Conferences Without Talks

- 2016 International Plant & Animal Genome XXIV Conference (**PAG XXIV**), San Diego, CA, USA
- 2015 International Conference for High Performance Computing, Networking, Storage and Analysis (**SC15**), Austin, TX, USA
- 2015 ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (**ACM-BCB 2015**), Atlanta, GA, USA

Teaching Experiences

- Tutorials for Parallel Algorithms and Architectures in JGU, Germany (SS2014)
- Tutorials for High Performance Computing (WS 2013/2014) in JGU, Germany
- Tutorials for Parallel Algorithms and Architectures in JGU, Germany (SS2013)
- Tutorials for High Performance Computing in JGU, Germany (WS2012/2013)
- Tutorials for Parallel Algorithms and Architectures in JGU, Germany (SS2012)
- Tutorials for Data Mining in JGU, Germany (WS 2011/2012)
- Laboratory Supervisor for Data Structures & Object-Oriented Programming in NTU, Singapore (2009/2010 Semester 2)
- Laboratory Supervisor for Data Structures & Object-Oriented Programming in NTU, Singapore (2010/2011 Semester 1)

Academic Collaborators

- Dr. Adrianto Wirawan (University of Mainz, Germany)
- Dr. Alberto Apostolico (Georgia Institute of Technology, USA)
- Dr. Bertil Schmidt (University of Mainz, Germany)
- Bernt Popp (University of Erlangen-Nuremberg, Germany)
- Dr. Douglas L. Maskell (Nanyang Technological University, Singapore)
- Dr. Jan Schröder (University of Melbourne, Australia)
- Dr. John Castle (TRON Mainz, Germany)

- Dr. Jorge Gonzalez Dominguez (University of A Coruna, Spain)
- Dr. Martin Löwer (University of Mainz, Germany)
- Dr. Le Song (Georgia Institute of Technology, USA)
- Dr. Oded Green (Georgia Institute of Technology, USA)
- Dr. Sek Won Kong (Harvard Medical School, USA)
- Dr. Sharma V. Thankachan (University of Central Florida, USA)
- Dr. Srinivas Aluru (Georgia Institute of Technology, USA)
- Dr. Tran Tuan Tu (University of Mainz, Germany)
- Dr. Thomas Hankeln (University of Mainz, Germany)
- Dr. Tao Li (Nankai University, China)
- Dr. Weiguo Liu (Shandong University, China)
- Dr. Ying Zhang (The University of Rhode Island, USA)

Supervised Students

- Felix Kallenborn (Bachelor, University of Mainz, Germany, 2014)
  Thesis title: Parallel counting of k-mers in DNA sequences on Intel Xeon Phi
- Simon Schindel (Bachelor, University of Mainz, Germany, 2013-2014)
  Thesis title: Bit-vector algorithm for approximate string matching on the Xeon Phi
- Felix Lauenroth (Bachelor, University of Mainz, Germany, 2013-2014)
  Thesis title: Smith Waterman on Intel Xeon Phi coprocessors: optimization for long DNA sequences
- Alex Watkins (Bachelor, Georgia Institute of Technology, USA, 2017)
  Research proposal: Merge sorting using AVX-512
- Bhaskar Khaneja (Bachelor, Georgia Institute of Technology, USA, 2017)
  Research proposal: Execution of graph analytics on AMD GPUs

References   Available upon request.