

# Curriculum vitae of Fahad Saeed, Ph.D.

Associate Professor of Computational Sciences,  
School of Computing and Information Sciences (SCIS) at Florida International University (FIU)

Mail: 11200 SW 8th St, Room ECS 261, Miami, FL 33199 USA

Email: [fsaeed@fiu.edu](mailto:fsaeed@fiu.edu)

Phone: (305) 348-2744

Web: <http://www.saeedfahad.com>

Document updated on August 3, 2018

## Contents

<b>1</b>	<b><u>Current Fields of Interests and Research Activities</u></b>	<b>3</b>
<b>2</b>	<b><u>Education and Professional Preparation</u></b>	<b>4</b>
<b>3</b>	<b><u>Experience</u></b>	<b>4</b>
3.1	Research Experience . . . . .	4
3.2	Teaching Experience . . . . .	6
<b>4</b>	<b><u>Honors and Awards</u></b>	<b>6</b>
<b>5</b>	<b><u>Research Funding</u></b>	<b>7</b>
5.1	Funded External Grants . . . . .	7
5.2	Funded Intramural Grants . . . . .	8
5.3	Pending . . . . .	8
<b>6</b>	<b><u>Publications and Creative Scholarship</u></b>	<b>8</b>
6.1	Edited Conference Proceedings . . . . .	8
6.2	Edited Special Issue of Journals . . . . .	9
6.3	Peer-Reviewed Journal Publications . . . . .	9
6.4	Peer-Reviewed Conference Proceedings . . . . .	11
6.5	Peer-Reviewed Abstracts and Posters . . . . .	13
6.6	Talks and Presentations . . . . .	14
6.7	Thesis . . . . .	16
6.8	Technical Reports . . . . .	16
6.9	Software (Released under GNU GPL or provided as a web-service) . . . . .	17
<b>7</b>	<b><u>Student Research Supervision</u></b>	<b>19</b>
7.1	PhD Dissertations (as Chair) . . . . .	19
7.2	PhD Student Supervision . . . . .	19
7.3	M.S. Student Supervision . . . . .	20
7.4	Undergraduate Researchers . . . . .	21
7.5	High School Research Students . . . . .	22
<b>8</b>	<b><u>Professional and Academic Service</u></b>	<b>22</b>
8.1	Editorial Boards . . . . .	22
8.2	Guest Editor Special Issues . . . . .	23
8.3	Conference/Workshop Chair . . . . .	23
8.4	Selected Conference/Workshop Program Committee Membership . . . . .	24

8.5	Peer Reviewer . . . . .	25
8.6	Proposal Panels and Reviews . . . . .	26
8.7	University and Department service . . . . .	26
8.7.1	Florida International University (FIU) . . . . .	26
8.7.2	Western Michigan University (WMU) . . . . .	27
8.8	Outreach and Community Service (recent) . . . . .	27
8.9	Miscellaneous . . . . .	28
<b>9</b>	<b><u>Consultation Experience</u></b>	<b>28</b>
<b>10</b>	<b><u>Professional Membership</u></b>	<b>28</b>

## 1 Current Fields of Interests and Research Activities

Fahad Saeed is a tenured Associate Professor in the School of Computing and Information Sciences at Florida International University (FIU), Miami FL and is the director of Saeed Lab which is a parallel computing and data science group: (<https://saeedlab.cis.fiu.edu/>). Prior to joining FIU, Prof. Saeed was an Assistant Professor (2014-2018) in the Department of Electrical & Computer Engineering and Department of Computer Science at Western Michigan University (WMU), Kalamazoo Michigan. He was tenured and promoted to the rank of Associate Professor at WMU in July 2018. He also serves as visiting scientist at National Institutes of Health (NIH) Bethesda, Maryland. Dr. Saeed was a Research Fellow in the National Institutes of Health (NIH) from 2011 to 2014 and a postdoctoral fellow from 2010-2011. He received his Ph.D. in the Department of Electrical and Computer Engineering, University of Illinois at Chicago (UIC) in 2010. He has served as visiting scientist in world-renowned prestigious institutions such as ETH Zurich, Swiss Institute of Bioinformatics (SIB) and National Institutes of Health (NIH). He has authored 42 research articles in peer-reviewed journals and conference proceedings.

Dr. Saeed's research interests are at the intersection of high performance computing and real-world applications, especially in computational biology. His research focus is to design and develop algorithmic and high performance computing (HPC) foundations for practical sublinear and parallel algorithms for big proteogenomics data. The proposed tools will reveal tremendous biological insight such as novel genes and proteins and is a crucial step towards understanding the genomic, proteomic and evolutionary aspects of species in the tree of life. Such analysis can lead to prevention, diagnosis and treatment of diseases with genetic predisposition such as cancer, obesity, diabetes, heart disease and mental illness. More info about his research activities can be found at <http://www.saeedfahad.com>

Dr. Saeed's research is supported by highly competitive grants mainly from National Science Foundation (NSF) and National Institutes of Health (NIH). He has secured over **US\$ 1.17 million** in external research funds as principal investigator and about **US\$1.33 million** overall. He was awarded the **NSF Research Initiation Initiative (CRII) Award** bestowed to young and promising scientists in the first two years of their tenure-track position. Most recently he was awarded the **NSF Faculty Early Career Development (CAREER) Award** which is NSF's most prestigious award in support of early-career faculty who have the potential to serve as academic role models in research and education. During his brief time in WMU since Jan 2014 his research has been supported by WMU, NVIDIA, Intel/Altera, National Science Foundation (NSF) and National Institutes of Health (NIH).

He has served as the program co-chair of the Bioinformatics and Computational Biology (BI-CoB) Conference and IEEE International Conference on Bioinformatics and Biomedicine (IEEE BIBM). He is also a founding chair of IEEE Workshop on HPC solutions to Big Data Computational Biology (IEEE HPC-BCB). He also serves on the editorial board of Springer Journal of Network Modeling Analysis in Health Informatics and Bioinformatics since 2014. He has served on numerous IEEE/ACM program committees and is peer-reviewer for more than a dozen journals. Dr. Saeed is the senior member of ACM and also the Senior Member of IEEE.

His honors include ThinkSwiss Fellowship (2007, 2008), NIH Postdoctoral Fellowship Award (2010), Fellows Award for Research Excellence (FARE) at NIH (2012), NSF CRII Award (2015), Outstanding New Researcher Award at WMU (2016), WMU Distinguished Research and Creative Scholarship Award (2018) and NSF CAREER Award (2017).

## 2 Education and Professional Preparation

**National Institutes of Health (NIH)** Bethesda, Maryland  
*Research Fellow* June 2011-January 2014  
Research Focus: Parallel algorithms for high-throughput proteomics & genomics

**National Institutes of Health (NIH)** Bethesda, Maryland  
*Postdoctoral Fellow* August 2010-June 2011  
*Mentor:* Mark A. Knepper  
Research Focus: High performance algorithms for high-throughput proteomics

**University of Illinois at Chicago (UIC)** Chicago, IL  
*Ph.D. Electrical & Computer Engineering* Fall 2006-Summer 2010  
*Advisor:* Ashfaq A. Khokhar  
Research Focus: High performance algorithms for high-throughput(NGS) genomics  
Dissertation: High performance computational biology algorithms

**University of Engineering & Technology (UET)** Lahore, Pakistan  
*B.Sc. Electrical Engineering* Jan 2002-Jan 2006  
*Advisor:* Shahid H. Bokhari

## 3 Experience

### 3.1 Research Experience

- **Florida International University (FIU)** Miami, FL, USA  
School of Computing and Information Sciences (SCIS)  
*Associate Professor (Tenured)* August 2018 to date  
Independent and collaborative research with colleagues, students, visiting scholars, grant writing, graduate and undergraduate student supervision, teaching graduate and undergraduate classes.
- **Western Michigan University (WMU)** Kalamazoo, MI, USA  
Department of Electrical & Computer Engineering (50%) and  
Department of Computer Science (50%)  
*Associate Professor (Tenured)* July 2018 to date (on leave)  
*Assistant Professor (Tenure-Track)* January 2014 to July 2018  
Independent and collaborative research with colleagues, students, visiting scholars, grant writing, graduate and undergraduate student supervision, teaching graduate and undergraduate classes.
- **Center of HPC and Big Data at WMU** Kalamazoo, MI, USA  
*Founding Co-Director* October 2014 to July 2018  
A 1400 sq. ft. area was acquired from the university by a competitive proposal to set up a Hadoop cluster, high performance server, GPU's and, FPGA's. Independent and collaborative research with colleagues, students, visiting scholars, grant writing, graduate and undergraduate student supervision are some of the activities accomplished using the center.

- **National Institutes of Health (NIH)** Bethesda, MD, USA  
*Visiting Scientist* January 2014 to date  
 Involved in various collaborative projects in systems biology, large-scale proteomics and genomics, grant writing and consultation for high-performance computing algorithms and infrastructure.
- **National Institutes of Health (NIH)** Bethesda, MD, USA  
*Research Fellow* June 2011 to January 2014  
 Independent and collaborative research with colleagues, students, visiting scholars, grant writing, graduate student supervision.
- **National Institutes of Health (NIH)** Bethesda, MD, USA  
*Postdoctoral Fellow* August 2010 to June 2011  
 Independent and collaborative research with colleagues, students, visiting scholars, graduate student supervision.
- **University of Illinois at Chicago (UIC)** Chicago, IL , USA  
*Research Assistant* August 2006 to August 2010  
 Included Ph.D. research, Ph.D. and Masters level coursework, aid in grant writing, reviewing, contributing numerous conference and journal papers.
- **University of Illinois at Chicago (UIC)** Chicago, IL , USA  
*Research Graduate Assistant* August 2009 to August 2010  
 Graduate assistant in Academic Computing and Communication Center at UIC. Responsibilities included maintaining websites for high traffic application, load-balancing webserver traffic and interactive scripts for academic and research applications
- **Swiss Institute of Bioinformatics (SIB)** Basel, Switzerland  
*Visiting Scientist in Mihaela Zavolan Group* Summers 2009  
 Proposed, designed and implemented a graph-theoretic framework for efficient computation of HMM based models for motif finding.
- **Swiss Federal Institute of Technology Zurich (ETH Zurich)** Basel, Switzerland  
*Visiting Scientist in Computational Biology Group (Niko Beerenwinkel)* Summers 2008  
 Designed a high performance multiple alignment algorithm for pyrosequencing reads of very large number. The algorithm was then used for viral population system of HIV haplotypes and resulted in more than 5 journal papers.
- **Space and Upper Atmosphere Research Commission** Lahore, Pakistan  
*Assistant Manager (Research)* Dec 2005 to August 2006  
 Analyzed the security vulnerability of organization wide computer networks. Designed and implemented a model for wireless communication incorporating different atmospheric conditions on memory-distributed clusters.
- **University of Engineering & Technology (UET)** Lahore, Pakistan  
*Undergraduate Research Assistant* Jan 2005 to Dec 2005  
 Introduced, designed and implemented an ad-hoc based wireless network in the computer communication laboratory which allowed the students to work in the same way as in the case of conventional (client/server) setup while eliminating single points of failure.

## 3.2 Teaching Experience

- **Florida International University (FIU)** Miami, FL, USA  
CDA 3103, Fundamentals of Computer Systems Fall 2018
- **Western Michigan University** Kalamazoo, MI, USA  
CS 4310, Design & Analysis of Algorithms Spring 2018  
ECE 2510, Introduction to Microprocessors Fall 2017  
ECE 2510, Introduction to Microprocessors Spring 2017  
CS6030/ECE6950, High Performance Architectures & Algorithms for Big Data Spring 2017  
ECE 2510, Introduction to Microprocessors Fall 2016  
CS 4310, Design & Analysis of Algorithms Spring 2016  
ECE 2510, Introduction to Microprocessors Fall 2015  
CS6030/ECE6950, High Performance Architectures & Algorithms for Big Data Spring 2015  
ECE 2510, Introduction to Microprocessors Fall 2014  
ECE/CS 7250, Doctoral Research Seminar Spring 2014
- **National Institutes of Health (NIH)** Bethesda, MD, USA  
Instructor for Community College Summer Enrichment Program (CCSEP) Summer 2012
- **University of Illinois at Chicago (UIC)** Chicago, IL, USA  
Instructor for Digital Communications Course (ECE 432) Fall 2007  
Teaching Assistant for Electric Circuit Analysis Lab (ECE 225) Spring 2007

## 4 Honors and Awards

1. Distinguished Research and Creative Scholarship Award, Office of Vice President of Research WMU, Feb 2018
2. Nominated for Blavatnik Award of The New York Academy of Sciences, Nov 2017
3. **NSF CAREER Award**, 2017-2022
4. **ACM Senior Member**, May 2017
5. Nominated for ACM Future of Computing Academy, Nov 2016
6. **Outstanding New Researcher Award**, College of Engineering and Applied Science (CEAS), Western Michigan University, Jan 2016 (1 faculty member gets the award in a single year for the entire college consisting of 7 academic departments)
7. **IEEE Senior Member**, March 2015
8. **NSF CISE Research Initiation Initiative (CRII) Award**, Feb 2015 - Feb 2018
9. Fellows Award for Research Excellence (FARE), National Institutes of Health (NIH), June 2012 (Official award ceremony and US\$1000 travel grant)
10. Nominated for NHLBI DIR Orloff Science Awards NIH, Jan 2012.
11. Postdoctoral Fellowship Award, National Institutes of Health (NIH), 2010-2011.

12. Nominated for the Best Dissertation Award of the year 2010 by ECE Department at UIC (One dissertation is nominated per year).
13. Tuition & Fee Waiver Award from Electrical and Computer Engineering Department, University of Illinois at Chicago, Summers 2010.
14. Recipient of full tuition fellowship and research assistantship by the Department of Electrical & Computer Engineering, University of Illinois at Chicago during doctoral studies.
15. Travel award from Swiss Institute of Bioinformatics (SIB), Summers 2009.
16. UIC Graduate Student Council (GSC) travel award for the BiCoB April 2009.
17. Recipient of Think Swiss Scholarship, by the Government of Switzerland for two years (2007 and 2008).
18. Travel award from D-BSSE ETH Zurich, Summers 2008.
19. Higher Education Commission Scholarship by Government of Pakistan in 2007 ( $\frac{13}{800} = 1.63\%$  acceptance rate in engineering discipline).
20. Satellite Research and Development Center Space and Upper Atmosphere Research Commission Pakistan (SUPARCO) in 2006 for PhD studies.

## **5 Research Funding**

The total dollar amount of external grants as a PI or Co-PI is **US\$1.33 million**. Since Jan 2014 share from these grants of over **US\$1.17 million to WMU**, of which my share has amounted to over **US\$1.17 million**.

### **5.1 Funded External Grants**

1. **National Science Foundation (NSF) ACI-1651724 [US\$ 500,000]**, “CAREER: Towards Fast and Scalable Algorithms for Big Proteogenomics Data Analytics”, **Fahad Saeed (PI)**, April 2017-April 2022
2. **National Institutes of Health (NIH) R15GM120820 [US\$ 418,533]**, “Parallel Algorithms for Big Data from Mass Spectrometry based Proteomics”, **Fahad Saeed (PI)**, April 2017 - April 2020
3. **National Science Foundation XSEDE renewal grant TG-CCR150017 [30,000 Service Units (SU)/6TB SDSC Disk Storage: US\$ 6564]**, “A Distributed-Shared Memory Strategy to Speedup the Compression of Big Next-Generation Sequencing Datasets”, **Fahad Saeed (PI)**, (June 2016 - June 2018)
4. **National Science Foundation (NSF) REU Supplement [US\$ 16,000]**, “CRII: SHF: HPC Solutions to Big NGS Data Compression”, **Fahad Saeed (PI)**, Feb 2016 - Feb 2018
5. **National Science Foundation XSEDE startup grant TG-CCR150017 [30,000 Service Units (SU)]**, “Scalability study of compression algorithms for peta scale NGS data”, **Fahad Saeed (PI)**, (June 2015 - June 2016)

6. **National Science Foundation (NSF) CCF-1464268 [US\$ 171,341]**, “CRII: SHF: HPC Solutions to Big NGS Data Compression”, **Fahad Saeed (PI)**, (Feb 2015 - Feb 2018)
7. **Altera [US\$ 16,000]**, “Short Reads mapping to the genome using reconfigurable hardware”, **Fahad Saeed(PI)** (Equipment Grant, 2 DE5-NET-450 FPGA’s), April 2014
8. **NVIDIA [US\$ 5499]**, “High Performance Algorithms for Genome Alignments”, **Fahad Saeed (PI)** (Equipment Grant for Tesla K40 GPU), Feb 2014
9. **National Science Foundation (NSF) CNS-1250264 [US\$ 200,000]**, “EAGER: High Performance Algorithms and Implementations for Biological Sequence Analysis and Genome Alignment”, Ashfaq Khokhar, **Fahad Saeed (Co-PI)** (Sept 2012 - Aug 2015)

## 5.2 Funded Intramural Grants

1. **Office of Vice President of Research, Western Michigan University (WMU) [US\$ 129,570]**, “Scalable Algorithms for Big Proteogenomics Data Analytics”, **Fahad Saeed (PI)**, April 2017 - April 2020
2. **College and Engineering and Applied Science (CEAS), Western Michigan University (WMU) [US\$ 41,650]**, “Developing HPC solutions to big fMRI data”, **Fahad Saeed (PI)**, April 2017 - June 2018

## 5.3 Pending

1. **National Institutes of Health (NIH) [US\$ 405,972]**, “Machine learning and high-performance computing algorithms for big fMRI data for ADHD classifications”, **Fahad Saeed (PI)** with Alvis Fong (Co-PI), June 2018
2. **Biological Technologies Office - DARPA [US\$ 350,823]**, “Near Real-Time Proteogenomics Data Analysis using FPGA’s”, **Fahad Saeed (PI)**, April 2018
3. **National Institutes of Health (NIH) [US\$ 687,639]**, “Fast and HPC Algorithms for big data proteomics and proteogenomics”, **Fahad Saeed (PI)**, March 2018

## 6 Publications and Creative Scholarship

† Indicates Corresponding and/or Senior Author.

### 6.1 Edited Conference Proceedings

1. Proceedings of 7th International Conference on Bioinformatics and Computational Biology (BICoB), with Hisham Al-Mubaid and Nurit Haspel, (ISBN: 9781510800137) March 2015
2. Proceedings of 6th International Conference on Bioinformatics and Computational Biology (BICoB), with Bhaskar Dasgupta, Hisham Al-Mubaid and Nurit Haspel (ISBN: 9781632665140) March 2014
3. Proceedings of 5th International Conference on Bioinformatics and Computational Biology (BICoB), with Bhaskar Dasgupta, Hisham Al-Mubaid and Reda Al-Hajj (ISBN: 978-1-880843-89-5), March 2013

4. Proceedings of 4th International Conference on Bioinformatics and Computational Biology (BICoB), with Hisham Al-Mubaid and Ashfaq Khokhar (ISBN: 978-1-880843-85-7), March 2012

## 6.2 Edited Special Issue of Journals

1. Special issue on selected papers from the 7th international conference on bioinformatics and computational biology (BICoB 2015) with Nurit Haspel and Hisham Al-Mubaid in the *Journal of Bioinformatics and Computational Biology (JBCB)* Vol. 14, No. 3. March 2016
2. Special issue on selected papers from the 6th international conference on bioinformatics and computational biology (BICoB 2014) with Bhaskar Dasgupta, Nurit Haspel and Hisham Al-Mubaid in the *Journal of Bioinformatics and Computational Biology (JBCB)* Volume 12, Issue 05, October 2014
3. Special issue on selected papers from the 5th international conference on bioinformatics and computational biology (BICoB 2013) with Bhaskar Dasgupta and Hisham Al-Mubaid, *Journal of Bioinformatics and Computational Biology (JBCB)* Volume 11, Issue 05, October 2013

## 6.3 Peer-Reviewed Journal Publications

1. **Fahad Saeed**<sup>†</sup>, “Towards quantifying psychiatric diagnosis using machine learning algorithms and big fMRI data”, *BMC Big Data Analytics*, Vol. 3, No. 1, pp. 1-7, May 2018
2. Taban Eslami, and **Fahad Saeed**<sup>†</sup>, “Fast-GPU-PCC: A GPU-Based Technique to Compute Pairwise Pearsons Correlation Coefficients for Time Series Data - An fMRI Study”, *MDPI High-Throughput*, Vol. 7, No. 2, article 11, April 2018
3. Mohammed Aledhari, Marianne E Di Pierro, Mohamed Hefeida, and **Fahad Saeed**<sup>†</sup>, “A Deep Learning-Based Data Minimization Algorithm for Fast and Secure Transfer of Big Genomic Datasets”, *IEEE Transactions on Big Data*, Feb 2018
4. Sandino N. Vargas-Pérez and **Fahad Saeed**<sup>†</sup>, “A Hybrid MPI-OpenMP Strategy to Speedup the Compression of Big Next-Generation Sequencing Datasets”, *IEEE Transactions on Parallel and Distributed Systems*<sup>1</sup>, vol. 28, no. 10, pp. 2760-2769, Oct. 1 2017 (**Impact Factor: 4.2**)
5. Pablo C. Sandoval, JNeka S. Claxton, Jae Wook Lee, **Fahad Saeed**, Jason D. Hoffert and Mark A. Knepper, “Systems-level analysis reveals selective regulation of Aqp2 gene expression by vasopressin”, *Nature Scientific Reports*, Vol. 6, article number 34863, October 2016 (**Impact Factor: 5.6**)
6. Muaaz Gul Awan and **Fahad Saeed**<sup>†</sup>, “MS-REDUCE: An ultrafast technique for reduction of Big Mass Spectrometry Data for high-throughput processing”, *Oxford Bioinformatics*<sup>2</sup>, Vol. 32, No. 10, pages 1518-1526, Jan 2016 (**Impact Factor: 7.3**)

---

<sup>1</sup>IEEE Transactions on Parallel and Distributed Systems is considered *the top* (highest ranked) peer-reviewed journal in its respective sub-category [https://scholar.google.com/citations?view\\_op=top\\_venues&hl=en&vq=eng\\_computingsystems](https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng_computingsystems)

<sup>2</sup>Oxford Bioinformatics is considered *the top* (highest ranked) peer-reviewed journal in its respective sub-category [https://scholar.google.com/citations?view\\_op=top\\_venues&hl=en&vq=eng\\_bioinformatics](https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng_bioinformatics)

7. Sookkasem Khositseth, Panapat Uawithya, Poorichaya Somparn, Komgrid Charngkaew, Nattakan Thippamom, Jason D. Hoffert, **Fahad Saeed**, D. Michael Payne, Shu Hui Chen, Robert A. Fenton and Trairak Pisitkun, "Autophagic degradation of aquaporin-2 is an early event in hypokalemia-induced nephrogenic diabetes insipidus", *Nature Scientific Reports*, Vol. 5, article number 18311 Dec 2015 (**Impact Factor: 5.6**)
8. Akshay Sanghi, Matthew Zaringhalam, Callan Corcoran, **Fahad Saeed**, Jason Hoffert, Pablo Sandoval, Trairak Pisitkun, and Mark Knepper, "A Knowledge Base of Vasopressin Actions in Kidney", *American Journal of Physiology (AJP)*, Vol. 307, No. 6, pages F747–F755, July 2014 (**Impact Factor: 7.6**)
9. **Fahad Saeed**<sup>†</sup>, Jason Hoffert, Trairak Pisitkun and Mark Knepper, "Exploiting Thread-Level and Instruction-Level Parallelism to Cluster Mass Spectrometry Data using Multicore Architectures", *Springer Journal of Network Modeling Analysis in Health Informatics and Bioinformatics*, Volume 3, No. 1, pages 1-19, April 2014 (**Impact Factor: N/A**)
10. **Fahad Saeed**<sup>†</sup>, Jason Hoffert and Mark Knepper, "CAMS-RS: Clustering Algorithm for Large-Scale Mass Spectrometry Data using Restricted Search Space and Intelligent Sampling", *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol.11, no.1, pp.128-141, Feb. 2014 (**Impact Factor: 1.4**)
11. Jason Hoffert, Trairak Pisitkun, **Fahad Saeed**, Justin Wilson, and Mark Knepper, "Global Analysis of the Effects of the V2 Receptor Antagonist Satavaptan on Protein Phosphorylation in Collecting Duct", *American Journal of Physiology (AJP)*, Vol. 306, No. 410-421 February 2014 (**Impact Factor: 7.6**)
12. **Fahad Saeed**<sup>†</sup>, Trairak Pisitkun, Jason D. Hoffert, Sara Rashidian, Guanghui Wang, Marjan Gucek, and Mark A. Knepper, "PhosSA: Fast and Accurate Phosphorylation Site Assignment Algorithm for Mass Spectrometry Data", *Proteome Science* Volume 11, Supplement 1, November 2013 (**Impact Factor: 2.4**)
13. Pablo C. Sandoval, Dane H. Slentz, Trairak Pisitkun, **Fahad Saeed**, Jason D. Hoffert and Mark A. Knepper, "Proteome-wide measurement of protein half-lives and translation rates in vasopressin-sensitive collecting duct cells", *Journal of the American Society of Nephrology (JASN)*, Volume 24, Issue 11, pages 1793-1805, March 2013 (**Impact Factor: 9.3**)
14. Boyang Zhao, Trairak Pisitkun, Jason D. Hoffert, Mark A. Knepper, and **Fahad Saeed**<sup>†</sup>, "CPhos: a program to calculate and visualize evolutionary conserved functional phosphorylation sites", *PROTEOMICS*, Vol. 12, No. 22, pages 3299-3303, October 2012 (**Impact Factor: 3.8**)
15. Steven Bolger, Patricia Gonzales Hurtado, Jason Hoffert, **Fahad Saeed**, Trairak Pisitkun, and Mark Knepper, "Quantitative Phosphoproteomics in Nuclei of Vasopressin-Sensitive Renal Collecting Duct Cells", *American Journal of Physiology (AJP)*, Vol. 303, No. 10, pages C1006-C1020, Sept. 2012 (**Impact Factor: 7.6**)
16. Jacqueline Douglass, Ruwan Gunaratne, Davis Bradford, **Fahad Saeed**, Jason D. Hoffert, Peter J. Steinbach, Mark A. Knepper, and Trairak Pisitkun, "Identifying Protein Kinase Target Preferences Using Mass Spectrometry", *American Journal of Physiology (AJP)*, Vol. 303, No. 7, pages C715-C727, June 2012 (**Impact Factor: 7.6**)

17. **Fahad Saeed**, Alan Perez-Rathke, Jaroslaw Gwarnicki, Tanya Berger-Wolf, Ashfaq Khokhar, “A high performance multiple sequence alignment system for pyrosequencing reads from multiple reference genomes”, *Journal of Parallel and Distributed Computing (JPDC)*, Volume 72, Issue 1, pages 83-93, January 2012 (**Impact Factor: 1.2**)
18. Jason D. Hoffert, Trairak Pisitkun, **Fahad Saeed**, Jae Song and Mark A. Knepper, “Dynamics of the vasopressin V2 receptor signalling network revealed by quantitative phosphoproteomics” *Molecular & Cellular Proteomics (MCP)*<sup>3</sup>, Vol. 11, Issue No. 2, Feb 2012 (**Impact Factor: 6.9**)
19. Trairak Pisitkun, Jason D. Hoffert, **Fahad Saeed** and Mark Knepper, “NHLBI-AbDesigner: An online tool for design of peptide-directed antibodies”, *American Journal of Physiology (AJP)*, Vol. 302, pages C154-C164, Jan 2012 (**Impact Factor: 7.6**)
20. **Fahad Saeed**, Ashfaq Khokhar, “A domain decomposition strategy for alignment of multiple biological sequences on multiprocessor platforms”, *Journal of Parallel and Distributed Computing (JPDC)* Vol 69, Issue 7, pages 666-677, July 2009 (**Impact Factor: 1.2**)
21. **Fahad Saeed**, Ashfaq Khokhar, Osvaldo Zagordi and Niko Beerenwinkel, “Multiple Sequence Alignment System for Pyrosequencing Reads” *Lecture Notes in Computer Science (LNCS)*, Volume 5462/2009, pp 362-375, 2009. (**Impact Factor: 0.5**)

#### 6.4 Peer-Reviewed Conference Proceedings

22. Mohammed Aledhari, Marianne Di Pierro, and **Fahad Saeed**<sup>†</sup>, “A Fourier-Based Data Minimization Algorithm for Fast and Secure Transfer of Big Genomic Datasets”, Proceedings of *IEEE Big Data Congress*, San Francisco CA USA, July 2-7, 2018 (19.8% acceptance rate)
23. Taban Eslami, and **Fahad Saeed**<sup>†</sup>, “Similarity based classification of ADHD using Singular Value Decomposition”, Proceedings of *ACM International Conference on Computing Frontiers (ACM-CF)*, Ischia, Italy, May 2018 (33% acceptance rate)
24. Usman Tariq, Umer Cheema and **Fahad Saeed**<sup>†</sup>, “Power-Efficient and Highly Scalable Parallel Graph Sampling using FPGAs”, Proceedings of *International Conference on Reconfigurable Computing and FPGAs (ReConFig 2017)*, Cancun, Mexico, December 4-6, 2017
25. Sandino N. Vargas-Pérez and **Fahad Saeed**<sup>†</sup>, “Scalable Data Structure to Compress Next-Generation Sequencing Files and its Application to Compressive Genomics”, Proceedings of *IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2017)* Kansas City, MO, USA, Nov 13-16, 2017
26. Muaaz Gul Awan and **Fahad Saeed**<sup>†</sup>, “An Out-of-Core GPU based dimensionality reduction algorithm for Big Mass Spectrometry Data and its application in bottom-up Proteomics”, Proceedings of *ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB)*, pp. 550–555, Boston MA, August 20-23, 2017 (29% acceptance rate)

---

<sup>3</sup>Molecular & Cellular Proteomics (MCP) is considered *the top* (highest ranked) peer-reviewed journal in its respective sub-category [https://scholar.google.com/citations?view\\_op=top\\_venues&hl=en&vq=bio\\_proteomicspeptides](https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=bio_proteomicspeptides)

27. Taban Eslami, Muaaz Gul Awan and **Fahad Saeed**<sup>†</sup>, “GPU-PCC: A GPU based technique to compute pairwise Pearson’s Correlation Coefficients for big fMRI data”, *Proceedings of ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB)*, pp. 723–728, Boston MA, August 20-23, 2017 (29% acceptance rate)
28. Mohammed Aledhari, Ali Marhoon, Ali Al-Qaabi, and **Fahad Saeed**<sup>†</sup>, “A New Cryptography Algorithm to Protect Cloud-based Healthcare Services”, *Proceedings of IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (IEEE/ACM CHASE-SEARCH)*, pp. 37–43, Philadelphia PA, July 2017
29. Muaaz Gul Awan and **Fahad Saeed**<sup>†</sup>, “GPU-ArraySort: A parallel, in-place algorithm for sorting large number of arrays”, *Workshop on High Performance Computing for Big Data, Proceedings of International Conference on Parallel Processing (ICPP-2016)*, pp. 78-87, Philadelphia PA, August 16-19, 2016
30. Majdi Maabreh, Ajay Gupta and **Fahad Saeed**<sup>†</sup>, “A Parallel Peptide Indexer and Decoy Generator for Crux Tide using OpenMP”, *Workshop on High Performance Computing Systems for Biomedical, Bioinformatics and Life Sciences, Proceedings of International Conference on High Performance Computing & Simulation (HPCS 2016)*, Innsbruck, Austria, July 2016 (**Nominated for Best Paper Award**)
31. Mohamed S Hefeida and **Fahad Saeed**<sup>†</sup>, “Data Aware Communication for Energy Harvesting Sensor Networks”, *Proceedings of International Conference on Wired & Wireless Internet Communications (WWIC 2016)*, Thessaloniki, Greece May 2016 (27 papers accepted out of 55 papers submitted: 49% acceptance rate)
32. Mohammed Aledhari, Mohamed S Hefeida and **Fahad Saeed**<sup>†</sup>, “A Variable-Length Network Encoding Protocol for Big Genomic Data”, *Proceedings of International Conference on Wired & Wireless Internet Communications (WWIC 2016)*, Thessaloniki, Greece May 2016 (27 papers accepted out of 55 papers submitted: 49% acceptance rate)
33. **Fahad Saeed**<sup>†</sup>, “Big Data Proteogenomics and High Performance Computing: Challenges and Opportunities”, *Proceedings of Symposium on Signal and Information Processing for Software-Defined Ecosystems, and Green Computing, IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Orlando Florida, Dec 2015
34. Sandino N. Vargas-Pérez and **Fahad Saeed**<sup>†</sup>, “A Parallel Algorithm for Compression of Big Next-Generation Sequencing (NGS) Datasets”, *Proceedings of Parallel and Distributed Processing with Applications (IEEE ISPA-15) Vol.3*. pp. 196-201, Helsinki Finland, August 2015
35. Mohammed Aledhari and **Fahad Saeed**<sup>†</sup>, “Design and Implementation Network Transfer Protocol for Big Genomic Data”, *Proceedings of IEEE Big Data Congress*, pp.281-288, New York City, USA, June 2015 (18% acceptance rate)
36. Muaaz Gul Awan and **Fahad Saeed**<sup>†</sup>, “On the sampling of Big Mass Spectrometry Data”, *Proceedings of ISCA International Conference on Bioinformatics and Computational Biology (BICoB)*, Honolulu, Hawaii, USA, March 2015
37. **Fahad Saeed**<sup>†</sup>, Jason Hoffert, and Mark Knepper, “A High Performance Algorithm for Clustering of Large-Scale Protein Mass Spectrometry Data using Multi-Core Architectures”, *Proceedings of IEEE/ACM International Symposium on Network Enabled Health Informatics, Biomedicine and Bioinformatics (HI-BI-BI)*, August 2013 (25% acceptance rate for full papers)

38. **Fahad Saeed**<sup>†</sup>, Trairak Pisitkun, Jason Hoffert, Guanghui Wang, Marjan Gucek, and Mark Knepper, "An Efficient Dynamic Programming Algorithm for Phosphorylation Site Assignment of Large-Scale Mass Spectrometry Data", Proceedings of *International Workshop on Computational Proteomics, Proceedings of IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Philadelphia USA, Oct 2012 (20% acceptance rate)
39. **Fahad Saeed**<sup>†</sup>, Trairak Pisitkun, Jason Hoffert, and Mark A. Knepper, "High Performance Phosphorylation Site Assignment Algorithm for Mass Spectrometry Data using Multicore Systems", proceedings of *ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB)*, 667-672, Orlando Florida USA, Oct 2012. (33 papers accepted out of 159 papers submitted: 21% acceptance rate)
40. **Fahad Saeed**<sup>†</sup>, Trairak Pisitkun, Mark A. Knepper, and Jason D. Hoffert, "An Efficient Algorithm for Clustering of Large-Scale Mass Spectrometry Data", Proceedings of *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1-4, Philadelphia USA, Oct 2012. (62 short paper accepted out of 299 papers submitted: 20.7 % acceptance rate)
41. **Fahad Saeed**<sup>†</sup> and Ashfaq Khokhar, "Parallel algorithm for center-star sequence alignments with applications to short reads" Proceedings of *ISCA International Conference on Bioinformatics and Computational Biology (BICoB)*, Las Vegas Nevada, USA, March 2012
42. **Fahad Saeed**<sup>†</sup>, Trairak Pisitkun, Mark A. Knepper, and Jason D. Hoffert, "Mining Temporal Patterns from iTRAQ Mass Spectrometry (LC-MS/MS) Data" Proceedings of *ISCA Bioinformatics and Computational Biology Conference (BICoB)* pp 152-159, New Orleans USA, March 2011 .
43. **Fahad Saeed** and Ashfaq Khokhar, "Sample-Align-D: A High Performance Multiple Sequence Alignment System using Phylogenetic Sampling and Domain Decomposition", Proceedings of *IEEE International Workshop on High Performance Computational Biology, IPDPS*, Monday, April 2008.

## 6.5 Peer-Reviewed Abstracts and Posters

1. Jae Wook Lee, Chung-Lin Chou, **Fahad Saeed**, and Mark A. Knepper, "RNA-seq of Microdissected Renal Tubules Identifies Segment-Specific Transcription Factors", poster presentation at *American Society of Nephrology (ASN)*, August 2013.
2. Jae Wook Lee, Chung-Lin Chou, **Fahad Saeed**, and Mark A. Knepper, "RNA-seq Identification of Transcriptome of Native DCT1 Cells", poster presentation at *American Society of Nephrology (ASN)*, August 2013.
3. Jae W. Lee, **Fahad Saeed**, Chung Lin Chou and Mark A. Knepper, "RNA-Seq Mapping Of G Protein-Coupled Receptor Expression along the Nephron and Collecting Duct", poster presentation at *American Society of Nephrology (ASN)*, August 2013.
4. Steven J Bolger, Patricia A Gonzales, Jason D Hoffert, **Fahad Saeed**, Trairak Pisitkun and Mark A Knepper, "Quantitative phosphoproteomics implicates clusters of proteins involved in cell-cell adhesion and transcriptional regulation in the vasopressin signaling network", meeting abstracts *Experimental Biology (EB)*, July 2013

5. Boyang Zhao, Trairak Pisitkun, Jason D. Hoffert, Mark A. Knepper, and **Fahad Saeed**, "An Information Theory-Based Approach to Assess the Functional Significance of Phosphorylation Sites in Proteomes of Renal Tubule Epithelia", poster at *International Society of Nephrology (ISN) Symposium*, Ann Arbor Michigan, USA, June 2012
6. **Fahad Saeed**, J. Hoffert, P. Pisitkun, M. Knepper, "Mapping-based temporal pattern mining algorithm identifies unique clusters of phosphopeptides regulated by vasopressin in collecting duct", meeting abstracts *Experimental Biology (EB)*, Washington DC, USA, April 2011.
7. J. Hoffert, T. Pisitkun, **Fahad Saeed**, J. Song, M. Knepper, "Large-scale iTRAQ-based quantification of phosphorylation changes during vasopressin signaling", Featured Topic and abstract *Experimental Biology (EB)*, Washington DC USA, April 2011.
8. **Fahad Saeed** and Ashfaq Khokhar, "Parallel Algorithm for Center Star Sequence and Alignments with applications to short reads", International Conference On *Bioinformatics and Computational Biology (ACM-BCB)* in August 2010.

## 6.6 Talks and Presentations

1. "Data-Reduction Algorithms for Big Data Proteomics: Challenges, Opportunities and Progress", at the Department of Mathematics and Statistics , *Western Michigan University (WMU)*, Kalamazoo, MI, April 2018 (**Invited**)
2. "Big Data Proteogenomics and High Performance Computing: Challenges, Progress and Opportunities", at the School of Computing and Information Sciences, *Florida International University (FIU)*, Miami, Florida, Jan 2018
3. "Reductive and HPC Analytics for Big Mass Spectrometry Data and its applications to Proteomics", *ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB)*, Boston MA, August 2017
4. "Big Data Proteogenomics and High Performance Computing: Challenges and Opportunities", *Symposium on Signal and Information Processing for Software-Defined Ecosystems, and Green Computing, (IEEE GlobalSIP)*, Orlando Florida, Dec 2015
5. "High Performance Computing and Big Data Computational Biology", introductory remarks at the inaugural *IEEE Workshop of High Performance Computing and Big Computational Biology (HPC-BCB)*, *IEEE BIBM*, Washington DC, Nov 2015
6. "Big Data Proteomics: High Performance Computing Methods & Applications", at School of Science & Engineering, *Lahore University of Management Science (LUMS)*, Lahore Pakistan, January 2015 (**Invited Seminar**)
7. "Solving Big Data Problems in Computational Biology using High Performance Architectures and Algorithms", at Department of Electrical Engineering and Computer Systems, *University of Cincinnati*, Cincinnati Ohio, November 2013
8. "Big Data Problems in high-throughput Genomics and Proteomics", at *Cincinnati Childrens Hospital Medical Center*, Cincinnati Ohio, November 2013

9. "Solving Big Data Problems in Computational Biology using High Performance Architectures and Algorithms", at Department of Computer Science, *Western Michigan University*, Kalamazoo MI , October 2013
10. "Solving Big Data Problems in Computational Biology using High Performance Architectures and Algorithms", at Department of Electrical and Computer Engineering, *Western Michigan University*, Kalamazoo MI , August 2013
11. "Solving Big Data Problems in Computational Biology using High Performance Architectures and Algorithms", at Innovation Center for Biomedical Informatics, *Georgetown University Medical Center*, Washington DC, August 2013
12. "Clustering and Consensus of Large-Scale Mass Spectrometry Data", Work in Progress Series, Systems Biology Interest Group, National Heart, Lung and Blood Institute (NHLBI), *National Institutes of Health (NIH)*, May 2013.
13. "High Performance Algorithm Engineering for Large-Scale Next Generation Genomics (NGS) Data", at Computer Science Department, *University of Massachusetts Boston*, Boston MA, March 2013.
14. "High Performance Algorithm Engineering for Large-Scale Next Generation Genomics (NGS) Data", at Electrical, Computer and Systems Engineering Department, *Rensselaer Polytechnic Institute (RPI)*, Albany NY, March 2013.
15. "High Performance Algorithm Engineering for Large-Scale Next Generation Genomics (NGS) Data", at Electrical and Computer Engineering Department, *George Washington University*, Washington DC, Feb 2013.
16. "High Performance Algorithm Engineering for Large-Scale Next Generation Sequencing (NGS) Data", at Center for Genome Research and Biocomputing (CGBC), *Oregon State University*, Corvallis Oregon, Oct 2012.
17. "An Efficient Algorithm for Clustering of Large-Scale Mass Spectrometry Data", *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Philadelphia Pennsylvania, Oct 2012
18. "An Efficient Dynamic Programming Algorithm for Phosphorylation Site Assignment of Large-Scale Mass Spectrometry Data", International workshop on computational proteomics, *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Philadelphia Pennsylvania, Oct 2012
19. "HPC approaches to computational problems in high-throughput genomics and proteomics", at Electrical and Computer Engineering Department, *State University of New York (SUNY) at Binghamton*, Feb 2012
20. "A graph theoretic approach to spectral clustering of mass spectrometry data", Work in Progress Series, Systems Biology Interest Group, National Heart, Lung and Blood Institute (NHLBI), *National Institutes of Health (NIH)*, Jan 2012.
21. "Computational problems in large-scale protein phosphorylation studies using iTRAQ labeled LC MS/MS data", *Bioinformatics and Computational Biology Conference (BICoB)*, New Orleans USA, March 2011.

22. "Dynamic Programming Algorithm for Phosphorylation Site Assignment of Mass Spectrometry Data", Work in Progress Series, Systems Biology Interest Group, National Heart, Lung and Blood Institute (NHLBI), *National Institutes of Health (NIH)*, Feb 2011.
23. "Challenges in multiple alignment of huge number of pyrosequencing reads on multiprocessor platforms", at *Institute of Genetic Medicine, Keck School of medicine, University of Southern California (USC)*, March 2010.
24. "Challenges in multiple alignment of huge number of pyrosequencing reads on multiprocessor platforms", at *school of medicine, University of Colorado at Denver(UCD)*, Feb 2010.
25. "How to multiple align huge number of short reads", at *Critical Assessment of Massive Data Analysis (CAMDA)* Oct 2009. **(Invited)**
26. "An Efficient Multiple Alignment system for Pyrosequencing Reads", at *Bioinformatics and Computational Biology Conference (BICoB)* April 09.
27. "Online scheduling of equal length Jobs on Parallel Machines", *Seminar UIC CS501*, Dec 08.
28. "A Domain Decomposition strategy for Multiple Alignment of Biological Reads", at *Laboratory of Computational Population Biology, UIC*, Nov 08. **(Invited)**
29. "A Sampling based High Performance Multiple Alignment systems", at *Proc. IEEE International Workshop on High Performance Computational Biology, IPDPS* April 08.
30. "An overview of Multiple Alignment systems, parallel approaches and their limitations", *Seminar UIC CS502*, May 07.

## 6.7 Thesis

1. **Fahad Saeed**, "High performance computational biology algorithms", (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (Accession Order No. AAT 3431281, ISBN: 9781124308067), 2010.
2. **Fahad Saeed**, Fahad A. Arshad and Salman Javed, "Design and Implementation of Linux Based Wireless Ad-Hoc Network based Compute Laboratory", (Undergraduate Final Report), Submitted as partial requirements for BSc. degree in Electrical Engineering University of Engineering and Technology Lahore, 2005.

## 6.8 Technical Reports

1. **Fahad Saeed**, Lukas Burger, Ashfaq Khokhar, and Mihaela Zavolan, "A graph-theoretic framework for efficient computation of HMM based motif finder", Technical Report, *University of Illinois at Chicago*, Jan 2010.
2. **Fahad Saeed**, "High Performance Graph Theoretic model for finding Regulatory Elements and motifs", Technical Report, Zavolan Group, *Swiss Institute of Bioinformatics (SIB), University of Basel Switzerland*, August 2009.
3. **Fahad Saeed**, "High Performance Multiple Alignment Systems for Pyrosequencing Reads of very large number", Technical Report, Beerenwinkel Group Computational Biology, *Department of Biosystems Science and Engineering, Eth Zurich Switzerland*, August 2008.

4. **Fahad Saeed** and Ashfaq Khokhar, “An overview of parallel programs for multiple sequence alignments and their limitations”, technical report, *Parallel Algorithms and Multimedia System Laboratory, University of Illinois at Chicago*, May 2007.
5. **Fahad Saeed**, “Securing Suparco’s Computer Networks and Linux Operating Systems”, Internal research Report on *Suparco’s Computer Networks Security X11/04/2006-I*, made official on April 2006.
6. **Fahad Saeed**, Fahad Ali Arshad and Salman Javed, “Design and Implementation of Linux based wireless ad-hoc network”, *Linux Focus*, article 390, Dec 2005.

## 6.9 Software (Released under GNU GPL or provided as a web-service)

### 1. J-Eros

J-Eros is a machine learning algorithm for computing similarity between two multivariate time series along with k-Nearest-Neighbor classifier, to classify healthy vs ADHD children using just fMRI data without using any other data or demographics. We applied this technique to the public data provided by ADHD-200 Consortium competition and our results show that J-Eros is capable of discriminating healthy from ADHD children such that we outperformed other state of the art techniques. This machine learning algorithm is a major step towards diagnosing ADHD using quantitative methods and will be an essential part for diagnosing mental illnesses. The algorithm is developed in python and can be accessed at <https://github.com/pcdslab/J-Eros>

### 2. PHYNGSC

phyNGSC is a hybrid strategy between MPI and OpenMP to accelerate the compression of big FASTQ datasets by combining the best features of distributed and shared memory architectures to balance the load of work among processes, alleviate memory latency by exploiting locality and accelerate I/O by reducing excessive read/write operations and inter-node message exchange. The algorithm introduces a novel timestamp-based approach which allows concurrent writing of compressed data in a non-deterministic order and thereby allows us to exploit a high amount of parallelism. As a proof-of-concept, we implemented some methods developed for DSRC v1 to underline the compression portion of our hybrid parallel strategy, since it exhibits superior performance for sequential solutions. The parallel algorithm is developed using C/C++, MPI and OPENMP constructs and is available at <https://github.com/pcdslab/PHYNGSC>.

### 3. GPU-ArraySort

GPU-ArraySort is a highly scalable parallel algorithm for sorting large number of arrays using a GPU. Existing techniques focus on sorting a single large array and cannot be used for sorting large number of smaller arrays in an efficient manner. Such small number of large arrays are common in many big data applications in fields such as proteomics, genomics, connectomics, and astronomy. Our algorithm performs in-place operations and makes minimum use of any temporary run-time memory. Our results indicate that we can sort up to 2 million arrays having 1000 elements each, within few seconds. We compare our results with the unorthodox tagged array sorting technique based on NVIDIA’s Thrust library. GPU-ArraySort out-performs the tagged array sorting technique by sorting three times more data in a much smaller time. The developed tool and strategy is made available at <https://github.com/pcdslab/GPU-ArraySort-2.0>.

#### 4. **MS-Reduce**

MS-Reduce is a linear-time tool that allows massive reduction in amount of mass spectrometry data without significantly reducing the quality of the peptide deduction. Our novel data-reductive strategy for analysis of Big MS data is called MS-REDUCE and is capable of eliminating noisy peaks as well as peaks that do not contribute to peptide deduction before any peptide deduction is attempted. Our experiments have shown up to 100x speed up over existing state of the art noise elimination algorithms while maintaining comparable high quality matches. Using our approach we were able to process a million spectra in just under an hour on a moderate server which will be especially useful for processing in high-throughput environments. The algorithm has been implemented in Java and code/associated data sets are available on GitHub <https://github.com/pcdslab/MSREDUCE>.

#### 5. **ParaDSRC**

ParaDSRC is a high-performance tool for compressing next generation sequencing data using memory-distributed clusters. It uses domain decomposition and message passing interface (MPI) to distributed data on memory-distributed compute nodes. Our implementation gives near-linear speedups for most of the data sets with some evidence of super-linear speedups for some data sets. We report experimental results for up to 1 tera byte (TB). The algorithm has been implemented using C/C++ and MPI and the code is available on GitHub here <https://github.com/PCDS/paraDSRC>

#### 6. **PhosSA**

PhosSA is a program for phosphorylation site assignment of LC-MS/MS data. It uses a linear-time and linear space dynamic programming strategy for phosphorylation site assignment. The algorithm optimizes the objective function defined as the summation of intensity peaks that are associated with theoretical peptide fragmentation ions. A classifier introduced in the algorithm exploits the specific characteristics of mass spectrometry data to distinguish between the correctly and incorrectly assigned site(s). The algorithm has been implemented in JAVA. An executable and instruction to use the software can be downloaded from <http://helixweb.nih.gov/ESBL/PhosSA/>.

#### 7. **CPhos**

Cphos is a program to calculate and visualize evolutionary conserved Phosphorylation sites. CPhos utilizes an information theory-based algorithm to assess the conservation of phosphorylation sites among species. A conservation established from this approach can be used to potentially assess the functional significance of a particular phosphorylation site. A web-service and executable are available from <http://helixweb.nih.gov/CPhos/>

#### 8. **P-Pyro-Align**

P-Pyro-Align is an open source parallel algorithm for multiple alignment of pyrosequencing reads from multiple genomes. The proposed alignment algorithm accurately aligns the erroneous reads and the accuracy of the alignment is confirmed from the consensus obtained from the multiple alignments. The algorithm uses domain decomposition for parallel computations of the local multiple alignments and a novel merging technique for global alignment of the reads. The proposed algorithm shows super-linear speedups for large number of reads. Note that the algorithm is for multiple alignment of reads coming from different strains of genomes which cannot be handled using mapping of the reads to a reference genome. The code has been implemented using C/C++ and MPI library.

## 7 Student Research Supervision

### 7.1 PhD Dissertations (as Chair)

1. Sandino N. Vargas-Pérez, Summer 2014 - Fall 2017 (Defended Summer 2017)  
Dissertation Title: A Hybrid Parallel Approach to High-Performance Compression of Big Genomic Files and In-Compresso Data Processing  
Honors: Winner of Regional 3MT competition  
First Position: Lecturer in CS department at WMU  
Last Known Position: Assistant Professor (Tenure-Track), Kalamazoo College, Kalamazoo MI USA
2. Mohammed Aledhari, Fall 2014 - Fall 2017 (Defended Summer 2017)  
Dissertation Title: A Deep Learning-based Data Minimization Algorithm for Big Genomics Data in Support of IoT and Secure Smart Health Services  
Honors: Recipient of 2016 Graduate Student Research Grant at WMU for US\$1000, Recipient of Gwen Frostic Doctoral Fellowship from Western Michigan University for US\$4000  
Last Known Position: N/A
3. Mohammad Abu Shattal (jointly supervised with Prof. Ala Al-Fuqaha), Summer 2014 - Fall 2017 (Defended Fall 2017)  
Dissertation Title: Bio-Socially Inspired Strategies in Support of Dynamic Spectrum Access: An Evolutionary Game Theory Perspective  
Last Known Position: Post-doctoral Fellow ECE department, Ohio State University, Columbus, OH

### 7.2 PhD Student Supervision

#### *Current PhD students (post prelim)*

1. Muaaz Gul Awan, Fall 2014 - Fall 2019 (expected)  
PhD student in the CS Department, WMU  
Recipient of 2015 Department-Level Graduate Research and Creative Scholars Award at WMU, Research Internship at Pacific Northwest National Laboratory (PNNL), Richmond WA
2. Taban Eslami, Fall 2015 - Fall 2020 (expected)  
PhD student in the CS department, WMU

#### *Current PhD students (post qualifier)*

1. N/A

#### *Current PhD students (pre-qualifier)*

1. Muhammad Haseeb, Fall 2017 - Fall 2022 (expected)  
PhD student in the SCIS, FIU
2. Usman Tariq, Fall 2018 - Fall 2023 (expected)  
PhD student in the SCIS, FIU

#### *Former PhD Students supervision*

1. Sarah Rashidian, May 2012 to Aug 2012  
PhD Student at National Institutes of Health (NIH)  
Project Title: Phosphorylation site assignment for large-scale mass spectrometry data  
Last Known Position: PhD student at Towson University, MD USA

#### *Member of PhD Dissertation Committee*

1. Semsi Coskun, Civil Engineering Department, 2017 - (Advisor: Houssam A Toutanji)
2. Hasnaa Al Shaikhli, CS department, 2017 - (Advisor: Elise DeDoncker)
3. Samah Rahamneh, ECE department, 2017 - (Advisor: Ikhlas Abdel-Qader)
4. Ting-Yu Mu, CS department, 2017 - (Advisor: Ala Al-Fuqaha)
5. Omar Abed Darwish, CS department, 2017 - (Advisor: Ala Al-Fuqaha)
6. Chung-Ling Lin, CS Department, 2014 - (Advisor: Wuwei Shen)

### **7.3 M.S. Student Supervision**

#### *Current MS Thesis Students*

1. N/A

#### *Current Non-Thesis Research Students*

1. N/A

#### *Former MS Thesis Students*

1. Usman Tariq (ECE, WMU), Fall 2016 - Spring 2018  
Thesis Title: Power-Efficient and Highly Scalable Parallel Graph Sampling using FPGAs  
Last Known Position: PhD student WMU
2. Ansab Ali (ECE, IIT Chicago co-advised with Ashfaq Khokhar), Fall 2014 - Spring 2016  
Thesis Title: Clustering Algorithm for Mass Spectrometry data using general-purpose computing on graphics processing units (GPU's)  
Last Known Position: Engineer at Intel Corp., Portland OR
3. Dana Emad Abdul Qader (ECE, WMU), Fall 2014 - Fall 2015  
Thesis Title: A High Performance Architecture for an Exact Match Short-Read Aligner Using Burrow-Wheeler Transform on FPGAs  
Last Known Position: QA Engineer at Stryker Corporation, San Jose CA

#### *Former Non-Thesis MS Research Students*

1. Blake Wrege, Spring 2017 - Summer 2017  
MS student in CS department at WMU  
Project Title: Graphical user interfaces for networking solutions for big data  
Last position: Site Reliability Engineer (SRE) at Apple

2. Srikanth Aravamuthan, Fall 2015 - Summers 2016  
MS student in Statistics department at WMU  
Project Title: statistical models for mass spectrometry based proteomics data  
Last position: N/A
3. Alan Perez-Rathke, Jan 2009 to Jan 2010  
MS student at University of Illinois at Chicago  
Project Title: Multiple sequence alignment algorithms using HPC systems for pyrosequencing reads  
Last Known Position: MD student at University of Illinois at Chicago
4. Jarek Gwarnicki, Jan 2009 to Jan 2010  
MS student at University of Illinois at Chicago  
Project Title: Quality assessment of MSA of pyrosequencing reads using domain decompositions.  
Last Known Position: Software Engineer at a Chicago based gaming company

## 7.4 Undergraduate Researchers

### *Current Undergraduate Research Students*

1. N/A

### *Former Undergraduate Research Students*

1. Alexander S. Cadigan, May 2017 - Aug 2017  
B.Sc. student in CS Department, Kalamazoo College  
Project Title: Scalability study of Open-Search Algorithms for MS based Proteomics data  
Last Known Position: Computer Science undergrad student at Kalamazoo College, Kalamazoo MI
2. James A Novorita, August 2017 - Dec 2017  
B.Sc student in the ECE Department, WMU  
Project Title: k-mer based alignment of big genomic data using FPGA's  
Last Known Position: Engineer at Stryker Corporation, Kalamazoo MI
3. Blake Wrege, Summer 2016 - Fall 2016  
B.Sc student in the CS Department, WMU Project Title: Graphical User interfaces for big data transmission and sharing  
Last Known Position: MS student at WMU
4. Melissa Basileo, Summer 2015 - Summer 2015  
B.Sc student in the CS Department, WMU
5. Akshay Sanghi, May 2013 to Aug 2013  
B.Sc. student at National Institutes of Health (NIH)  
Project Title: A Knowledge Base of Vasopressin Actions in Kidney  
Last Known Position: MD student at Johns Hopkins University
6. Boyang Zhao, Dec 2011 to Aug 2012  
B.Sc. student at National Institutes of Health (NIH)  
Project Title: Assessing the Functional Relevance of Phosphorylation Sites using Information

Theory and Naive Bayes Classifications

Last Known Position: PhD student at Massachusetts Institute of Technology (MIT)

7. Jacqueline Douglass, Sept 2010 to Aug 2011  
B.Sc. student at National Institutes of Health (NIH)  
Project Title: Protein Kinase Target-Sequence Profiling Using LC-MSMS  
Last Known Position: MD/PhD student at Johns Hopkins University

## 7.5 High School Research Students

### *Current High School Research Students*

1. N/A

### *Former High School Research Students*

1. Zoe Srackangast, Gwen Park, and Isabel Holton, Sept 2017 - March 2018  
High School student at Kalamazoo Area Mathematics & Science Center  
Project Title: Development of Algorithm for Classification of Proteins found in Environmental Samples (Recipient of Out Standing Achievement for Ability and Creativity Award @ SW-Michigan International Science & Engineering Fair (ISEF) 2018)  
Last Known Position: Isabel Holton joined Undergraduate program in Electrical Engineering at University of Michigan Ann Arbor
2. Lily Kitagawa, Oct 2015 - March 2016  
High School student at Kalamazoo Area Mathematics & Science Center  
Project Title: Counting kmers in big genomic data (Recipient of Intel Excellence in Computer Science Award 2016)  
Last Known Position: Undergraduate student (Computer Science) at California Institute of Technology (Caltech)
3. Binh Le, August 2014 - March 2015  
High School student at Kalamazoo Area Mathematics & Science Center  
Project Title: GUI development of computational biology tools for domain scientists  
Last Known Position: Undergraduate student at Massachusetts Institute of Technology (MIT)
4. Adam Loles, August 2014 - March 2015  
High School student at Kalamazoo Area Mathematics & Science Center  
Project Title: GUI development of computational biology tools for domain scientists  
Last Known Position: Undergraduate student at Colgate University.

## 8 Professional and Academic Service

### 8.1 Editorial Boards

- **Associate Editor** for Springer Journal of Network Modeling Analysis in Health Informatics and Bioinformatics, Jan 2014 to date.

## 8.2 Guest Editor Special Issues

- Guest Editor for *Journal of Bioinformatics and Computational Biology (JBCB)* special issue containing selected papers from the 7th International Conference on Bioinformatics and Computational Biology, Feb 2016 (with Nurit Haspel and Hisham Al-Mubaid).
- Guest Editor for *Journal of Bioinformatics and Computational Biology (JBCB)* special issue containing selected papers from the 6th International Conference on Bioinformatics and Computational Biology, Oct 2014 (with Hisham Al-Mubaid and Bhaskar Dasgupta).
- Guest Editor for *Journal of Bioinformatics and Computational Biology (JBCB)* special issue containing selected papers from the 5th International Conference on Bioinformatics and Computational Biology, Oct 2013 (with Hisham Al-Mubaid and Bhaskar Dasgupta).

## 8.3 Conference/Workshop Chair

- Track Chair for Mathematical Modeling and Scientific Computing, IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur India, 2016
- Session Chair, IEEE Workshop on High Performance Computing for Big Data Computational Biology (IEEE HPC-BCB), in conjunction with IEEE BIBM, Washington DC, Nov 2015
- **Founding Workshop Chair**, IEEE Workshop on High Performance Computing for Big Data Computational Biology (IEEE HPC-BCB), in conjunction with IEEE BIBM, Nov 2015
- Tutorials Co-Chair for IEEE International Conference on Collaboration Technologies and Systems (CTS) Atlanta, Georgia, USA, May 2015
- **Program Co-Chair** for 7th Bioinformatics and Computational Biology (BICoB) Conference, March 2015
- Tutorials Co-Chair for IEEE International Conference on Collaboration Technologies and Systems (CTS) Minneapolis, Minnesota, USA, May 2014
- **Program Co-Chair** for International Workshop on Algorithms for Computational Biology (ACB), 2014
- **Program Co-Chair** for 6th Bioinformatics and Computational Biology (BICoB) Conference, Las Vegas, Nevada, March 24-26, 2014
- Session Chair, International workshop on computational proteomics, IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Oct 2012
- Session Chair, IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Oct 2012
- **Program Co-Chair** for 5th Bioinformatics and Computational Biology (BICoB) Conference, Honolulu, Hawaii, March 4-6, 2013
- **Program Co-Chair** for Bioinformatics and Computational Biology (BICoB) Conference, Las Vegas, Nevada, USA, March 12 - 14, 2012

#### 8.4 Selected Conference/Workshop Program Committee Membership

- Program committee member, Symposium on Bioinformatics and Bioforensics (SBB18), ICACCI, Bangalore India, Sept 2018
- Program committee member, Workshop on Large Scale Computational Physics (LSCP), International Conference on Computational Science and Applications (ICCSA), June 2018
- Program committee member, First International Symposium on Signal and Image Processing (ISSIP-2017), Kolkata, India Nov 2017
- Program committee member, IEEE International Conference on Advanced Computational and Communication Paradigms (ICACCP-2017), Sikkim, India 8-10th September 2017
- Program committee member, 10th International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage (SpaCCS 2017), Guangzhou, China, December 12-15, 2017.
- Program committee member, International Symposium on Bioinformatics and Bioforensics (SBB17), International Conference on Advances in Computing, Communications and Informatics (ICACCI), Manipal, India September 13-16, 2017
- Program Committee Member, International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India 2017
- Program committee member, International Symposium on Bioinformatics and Bioforensics (SBB16), International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India September 21-24, 2016
- Program committee member, International Symposium on Network Enabled Health Informatics, Biomedicine and Bioinformatics (HI-BI-BI), Aug 2016
- Program committee member, Workshop on Large Scale Computational Physics (LSCP), International Conference on Computational Science and Applications (ICCSA), June 2016
- Program committee member, The 15th IEEE International Conference on Data Mining (ICDM), 2015
- Program committee member, IEEE Symposium on Signal and Information Processing for Software-Defined Ecosystems and Green Computing (IEEE GlobalSIP), Orlando Florida Dec 2015
- Program committee member, Advancement in Petroleum and Chemical Engineering Technology and Applications International Conference (APCETA), Dec 2015
- Program committee member, IEEE Information Reuse and Integration in Health Informatics (IRI-HI), San Francisco CA August 2015
- Program committee member, IEEE/ACM International Symposium on Network Enabled Health Informatics, Biomedicine and Bioinformatics (HI-BI-BI), Paris, France, Aug 26-27, 2015
- Program committee member, International Workshop on Algorithms for Computational Biology (ACB-2014), Delhi, India, September 2014

- Program committee member, IEEE/ACM International Symposium on Network Enabled Health Informatics, Biomedicine and Bioinformatics (HI-BI-BI), Beijing, China, August 18-19, 2014
- Member Steering Committee, 2nd Computational Science Conference, Pakistan Society of Computational Science and Biology (PSCS/PSCB) for High School Students, Islamabad Pakistan, Oct 2013
- Program committee member for IEEE International Workshop on High Performance Computational Biology (HiCOMB), International Parallel & Distributed Processing Symposium (IPDPS), May 27, 2013
- Program committee member for ACM conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB) conference, Orlando Florida, Oct 7-10, 2012
- Program committee member for 11th International Workshop on Data Mining in Bioinformatics (BIOKDD '12)
- Program committee member for 8th International Conference on Bioinformatics and Genome Regulation and System Biology (BGRS/SB) 2012

## 8.5 Peer Reviewer

- Peer Reviewer for *Springer Signal, Image and Video Processing Journal*, July 2018
- Peer Reviewer for *IEEE Transactions on Parallel and Distributed Systems*, Feb 2018
- Peer Reviewer for *Wiley Proteomics*, Nov 2017
- Peer Reviewer for *Wiley Proteomics*, Oct 2017
- Peer Reviewer for *BMC Bioinformatics*, Sept 2017
- Peer Reviewer for *Wiley Proteomics*, Aug 2017
- Peer Reviewer for *IEEE Transactions on Parallel and Distributed Systems*, Aug 2017
- Peer Reviewer for *Software Quality Journal*, Jan 2017
- Peer Reviewer for *Springer Journal Network Modeling Analysis in Health Informatics and Bioinformatics (NHIB)*, Jan 2017
- Peer Reviewer for *Oxford Bioinformatics*, Nov 2016
- Peer Reviewer for *Nature Protocols*, Oct 2016
- Peer Reviewer for *Oxford Bioinformatics*, Sept 2016
- Peer Reviewer for *Springer Journal Network Modeling Analysis in Health Informatics and Bioinformatics (NHIB)*, Sept 2015
- Peer Reviewer for *Journal of Bioinformatics and Computational Biology (JBCB)*, July 2015
- Peer Reviewer for *Journal of Computer Science and Technology (JCST)*, May 2015

- Peer Reviewer for *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, March 2015
- Peer Reviewer for *Springer Journal Network Modeling Analysis in Health Informatics and Bioinformatics (NHIB)*, Feb 2015
- Peer Reviewer for *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Feb 2015
- Reviewer for *Oxford Bioinformatics*, Jan 2015
- Reviewer for *International Journal of Information and Communication Technology (IJICT)*, Jan 2015
- Reviewer for *International Journal of Computers and Their Applications (IJCA)*, September 2014
- Peer Reviewer for *Springer Journal Network Modeling Analysis in Health Informatics and Bioinformatics (NHIB)*, March 2014
- Peer Reviewer for *International Journal of Computer Systems Science and Engineering (IJCSSE)*, 2013
- Peer Reviewer for *BMC Proteome Science*, March 2013
- Peer Reviewer for *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2012
- Peer Reviewer for *International Journal of Computer Systems Science and Engineering (IJCSSE)* 2009 to 2011.

## 8.6 Proposal Panels and Reviews

- Reviewer for Croatian Science Foundation (CSF), March 2018
- NSF Panel, Office of Advanced Cyberinfrastructure (OAC), Oct 2017
- ORAU NSF Big Data Regional Innovation Hub Panel, 2017
- Early Career Reviewer (ECR), Center for Scientific Review (CSR), National Institutes of Health (NIH), March 2017 to date
- Reviewer for Croatian Science Foundation (CSF), March 2014

## 8.7 University and Department service

### 8.7.1 Florida International University (FIU)

- Member of SCIS Graduate Committee, Aug 2018 - to date
- Member of SCIS Tenure Track Faculty Hiring Committee, Aug 2018 - to date
- Member of SCIS ACM Student Chapter Faculty Advisor, Aug 2018 - to date

### **8.7.2 Western Michigan University (WMU)**

- Panelist for NSF CAREER Workshop for Faculty, Office of Vice President of Research (OVPR), WMU, March 2018
- Member of CEAS Faculty Creative Initiative Review Committee, WMU, Feb 2018
- Member of Undergraduate Computer Engineering Curriculum Revision Committee, ECE Department, Sept 2017 - to date
- Panelist for Discover Discovery Workshop for New Faculty, Office of Vice President of Research (OVPR), WMU, Sept 2017
- Member of committee on designing and building Biomedical Engineering Program for CEAS WMU, July 2017 - to date
- Member of CEAS outstanding awards nomination, Jan 2017
- Member of Graduate Committee, CS Department, Sept 2016 - to date
- Member of Undergraduate Computer Engineering Curriculum Revision Committee, ECE Department, Sept 2016 - Dec 2016
- Member sub-committee for research Strategic Plan for College of Engineering and Applied Science (CEAS) WMU, March 2016
- Member of Strategic Plan Committee for College of Engineering and Applied Science (CEAS) WMU, March 2016
- Panelist for Discover Discovery Education and Training for New Faculty, Office of Vice President of Research (OVPR), WMU, Sept 2015
- Member of Undergraduate Computer Engineering Curriculum Revision Committee, ECE Department, Oct 2014 - Aug 2015
- Member of Seminar Committee, ECE Department, Oct 2014 - to date
- Member of Computing Hardware Committee, Computer Science Department, Sept 2014 - to date
- Proctoring service for ECE Qualifying Exam, March 2014
- ECE representative for the College of Engineering and Applied Sciences (CEAS) Scholarship Committee, Feb 2014 - to date

### **8.8 Outreach and Community Service (recent)**

- Judge for Regional Intel Science and Engineering Fair, for high school students (K-12) Kalamazoo Area Math and Science Center (KAMSC), March 2017
- Judge for 16th Annual Southwest Michigan Regional (Intel Science and Engineering) ISEF Science Fair for high school students (K-12) Kalamazoo Area Math and Science Center (KAMSC), March 2016

- Judge for Southwest Michigan Science and Engineering Fair for high school students (K-12) Kalamazoo Area Math and Science Center (KAMSC), March 2015
- Judge for Regional Science and Engineering Fair for high school students (K-12) Kalamazoo Area Math and Science Center (KAMSC), March 2014

## 8.9 Miscellaneous

- Judge in genomics study section for Fellows Award for Research Excellence (FARE), National Institutes of Health (NIH) 2013-2014
- Student representative for Electrical and Computer Engineering Department, Graduate Student Council, University of Illinois at Chicago (Jan 2008 to Summers 2010).

## 9 Consultation Experience

- **Chulalongkorn University** Bangkok, Thailand  
*Systems Biology (CUSB) Center* Oct 2013 to Nov 2013  
 This work involved design of setting up a compute infrastructure for large-scale system biology computational needs. A memory-distributed compute cluster, a high-performance workstation, an intelligent memory system and GPU's were part of the setup.
- **University of Illinois at Chicago (UIC)** Chicago, IL USA  
*Parallel algorithms and multimedia systems laboratory* May 2010 to August 2010  
 This work involved design of setting up a high-performance compute clusters for computational biology algorithms. A 16-node system was setup as a pilot project.

## 10 Professional Membership

- Institute of Electrical and Electronics Engineers (IEEE), since 2005
- Association for Computing Machinery (ACM), since 2010