Aritra Bose

IBM T.J. Watson Research Center
1101 Kitchawan Road
Yorktown Heights, NY 10598

Email: a.bose@ibm.com,
bosearitra08@gmail.com
Web: https://aritra90.github.io/
LinkedIn: aritrabose
GitHub: aritra90
Phone: +1(518) 522-7975

Research Interests

Bioinformatics, Computational Biology, Artificial Intelligence, Statistical Genetics, Biomedical Informatics, Quantum Computing, Data Mining, Topological Data Analysis, Randomized Numerical Linear Algebra.

Education

Purdue University — West Lafayette, IN, USA
Ph.D. in *Computer Science* — Aug 2016 - Aug 2019
Advisor: Prof. Petros Drineas
Thesis: Computational Methods for Population Genetics

Rensselaer Polytechnic Institute — Troy, NY, USA
M.S. in *Computer Science* — Aug 2014 - Jul 2016

West Bengal University of Technology — Kolkata, WB, India
B.Tech in *Information Technology* — Aug 2009 - Jun 2013

Experience

IBM T.J. Watson Research Center — Yorktown Heights, NY, USA
Staff Research Scientist — Feb 2021 - Present
Postdoctoral Researcher — Sep 2019 - Feb 2021
Research Intern — May 2018 - Jul 2018
Research Intern — May 2017 - Aug 2017
Research Intern — May 2016 - Aug 2016

IBM India Research Lab — Gurgaon, India

Broad Institute of MIT and Harvard — Cambridge, MA, USA
Visiting Scientist — Jun 2021 - Jan 2023

Purdue University — West Lafayette, IN, USA
Research Assistant — Aug 2016 - Aug 2019

Rensselaer Polytechnic Institute — Troy, NY, USA
Research Assistant — Aug 2015 - Aug 2016
Teaching Assistant — Aug 2014 - May 2015

Teradata Corporation — Hyderabad, India
Analyst — Oct 2013 - Apr 2014

Bose Institute — Kolkata, India
Research Trainee — Sep 2012 - Oct 2013

Indian Institute of Technology — Guwahati, India
Summer Intern — May 2012 - Jul 2012

Indian Statistical Institute Kolkata, India
Winter Intern Dec 2011 - Mar 2012

Publications

Journals

1. D.E. Platt, A. Guzmán-Sáenz, **A. Bose**, S. Saha, F. Utro, L. Parida, *Measuring Single Nucleotide Polymorphism Relevance by Significance and Predictivity in Alzheimers Disease for AI and Polygenic Risk Score Analyses*, iScience, 2024. **(IF = 5.8)**
2. **A. Bose**, M.C. Burch, A. Chowdhury, P. Paschou, P. Drineas, *Structure-informed clustering for population stratification in association studies.* BMC Bioinformatics 24, no. 1 (2023): 411. **(IF = 4.3)**
3. **A. Bose**, F. Utro, D.E. Platt, L. Parida, *Multiple Loci Selection with Multi-way Epistasis in Coalescence with Recombination*, Algorithms 14 (5), 136, 2021. **(IF = 2.27)**
4. **A. Bose**, D.E. Platt, L. Parida, P. Paschou, P. Drineas, *Integrating linguistics, social structure, and geography to model genetic diversity within India*, Molecular Biology and Evolution 38 (5), 1809-1819, 2021.**(IF = 16.24)**
5. **A. Bose**, V. Kalantzis, E. Kontopoulou, M. Elkady, P. Paschou, P. Drineas, *TeraPCA: a fast and scalable software package to study genetic variation in tera-scale genotypes*, Bioinformatics, Volume 35, Issue 19, 1 October 2019, Pages 36793683. **(IF = 6.93)**
6. G. Stamatoyannopoulos, **A. Bose**, A. Teodasiadis, F. Tsetsos, A. Plantiga, N. Psatha, N. Zogas, E. Yannaki, P. Zalloua, K.K. Kidd, B.L. Browning, J. Stamatoyannopoulos, P. Paschou, P. Drineas, *Genetics of the Peloponnesean populations and the theory of the extinction of the medieval Peloponnesean Greeks*, European Journal of Human Genetics, 25(5), pp. 637-645, 2017. **(IF = 5.35)**

Conferences

7. D. Gurnari, A. Guzmán-Sáenz, F. Utro, **A. Bose**, S. Basu, L. Parida, (2023). *Probing omics data via harmonic persistent homology.* Accepted, RECOMB-CCB, 2024.
8. D.E. Platt, **A. Bose**, C. Levovitz, K. Rhrissorrakrai, L. Parida, *Epidemiological topology data analysis links severe COVID-19 to RAAS and hyperlipidemia associated metabolic syndrome conditions*, In Intelligent Systems for Molecular Biology (ISMB) 2024.
9. M.C. Burch, **A. Bose**, L. Parida, G. Dexter, P. Drineas, *MaSk-LMM: A Matrix Sketching Framework for Linear Mixed Models in Association Studies*, Accepted, RECOMB 2024.
10. D.E. Platt, A. Guzmán-Sáenz, **A. Bose**, S. Saha, F. Utro, L. Parida, *Characterizing Single Nucleotide Polymorphism Relevance by Significance and Predictivity in Alzheimer's Disease using Machine Learning and Polygenic Risk Score Analysis*, RECOMB Genetics, 2023.
11. D. Machado Reyes[†], **A. Bose**[†], E. Karavani, L. Parida, *FairPRS: adjusting for admixed populations in polygenic risk scores using invariant risk minimization* , In PACIFIC SYMPOSIUM ON BIOCOMPUTING 2023: Kohala Coast, Hawaii, USA, 37 January 2023, pp. 198-208. ([†] Equal Contributors)

12. D.E. Platt, **A. Bose**, C. Levovitz, K. Rhrissorrakrai, L. Parida, *Epidemiological topology data analysis links severe COVID-19 to RAAS and hyperlipidemia associated metabolic syndrome conditions*, In AMIA Annual Symposium 2022. American Medical Informatics Association.

13. A. Chowdhury†, **A. Bose**†, S. Zhou, D. P. Woodruff, P. Drineas, *A Fast, Provably Accurate Approximation Algorithm for Sparse Principal Component Analysis Reveals Human Genetic Variation Across the World*, In Research in Computational Molecular Biology: 26th Annual International Conference, RECOMB 2022, San Diego, CA, USA, May 2225, 2022, Proceedings, pp. 86-106. Cham: Springer International Publishing, 2022. († Equal Contributors)

14. S. Dey†, **A. Bose**†, S. Saha, P. Chakraborty, M. Ghalwash, A.G. Sáenz, F. Utro, K. Ng, J. Hu, L. Parida, D. Sow, *Impact of Clinical and Genomic Factors on COVID-19 Severity*, In AMIA Annual Symposium Proceedings (Vol. 2021, p. 378). American Medical Informatics Association. († Equal Contributors)

15. **A. Bose**, M.C. Burch, A. Chowdhury, P. Paschou, P. Drineas, *CluStrat: a structure informed clustering strategy for population stratification*, In Research in Computational Molecular Biology: 24th Annual International Conference, RECOMB 2020, Padua, Italy, May 1013, 2020, Proceedings 24 (pp. 234-236). Springer International Publishing.

Preprints

16. **A. Bose**, C. Levovitz, D.E. Platt, S. Dey, U. Kartoun, K. Ng, L. Parida, *Clinical and genomic factors impacting Long COVID in UK Biobank*, In preparation, 2024.

17. D. Machado Reyes†, M.C. Burch, L. Parida, **A. Bose**, *COMICAL: A multimodal foundation model for contrastive multi-omics association learning in complex diseases*, Under Review, 2024.

18. D. Gurnari, A. Guzmán-Sáenz, F. Utro, **A. Bose**, S. Basu, L. Parida, (2023). *Probing omics data via harmonic persistent homology.* arXiv preprint arXiv:2311.06357.

19. S. Basu, J. Born, **A. Bose**, S. Capponi, D. Chalkia, T. A. Chan, H. Doga, et al., *Towards quantum-enabled cell-centric therapeutics.*, arXiv, 2023.

20. **A. Bose**†, D.E. Platt†, K. Ng, L. Parida, *Role of genetics in capturing racial disparities in cardiovascular disease*, medRxiv, 2023. († Equal Contributors)

21. D. Machado Reyes†, **A. Bose**†, E. Karavani, L. Parida, *FairPRS: a fairness framework for polygenic risk scores*, medRxiv, 2022. († Equal Contributors)

22. D.E. Platt, **A. Bose**, C. Levovitz, K. Rhrissorrakrai, L. Parida, *Epidemiological topology data analysis links severe COVID-19 to RAAS and hyperlipidemia associated metabolic syndrome conditions*, medRxiv, 2022.

23. **A. Bose**, D.E. Platt, N. Haiminen , L. Parida, *CuNA: Cumulant-based genotype-phenotype interaction networks in Parkinson's Disease*, medRxiv, 2021.

24. S. Dey, **A. Bose**, P. Chakraborty, M. Ghalwash, A.G. Saenz, F. Utro, K. Ng, J. Hu, L. Parida, D. Sow, *Impact of Clinical and Genomic Factors on SARS-CoV2 Disease Severity*, medRxiv, 2021.

25. S. Saha †, A.G Sáenz †, **A. Bose** †, F. Utro, D.E. Platt, L. Parida, *RubricOE: a learning framework for genetic epidemiology*, medRxiv, 2021. († Equal Contributors)

26. **A. Bose**, M.C. Burch, A. Chowdhury, P. Paschou, P. Drineas, *Structure informed clustering adjusts for population stratification in association studies*, BioRxiv, 2020.
27. **A. Bose**, D.E. Platt, L. Parida, P. Paschou, P. Drineas, *Dissecting Population Substructure in India via Correlation Optimization of Genetics and Geodemographics*, BioRxiv, 2017.
28. S. Hassan, P. Pal Choudhury and **A. Bose**, (2011), *A Quantitative model for Human Olfactory Receptors*, Nature Precedings, npre20126967-2, 2012.

Abstracts (peer reviewed only)

29. M. Burch, **A. Bose**, L. Parida, P. Drineas, *MaSk-LMM: a matrix sketching-based fast and scalable linear mixed model for association studies in large biobanks*, Annual meeting of the American Society for Human Genetics (ASHG), 2022.
30. D.E. Platt[†], **A. Bose**[†], K. Ng, L. Parida, *Race versus Genetics in clinical decision-making: a perspective from cardiovascular disease*, Intelligent Systems for Molecular Biology (ISMB), 2022. ([†] Equal Contributors)
31. M. Burch, P. Jain, Z. Yang, A. Topaloudi, P. Paschou, **A. Bose**, P. Drineas, *Predicting Complex Disorders by Combining Comorbidity Data and Polygenic Risk Scores*, ISMB, 2022.
32. A. Guzmán-Sáenz, D.E. Platt, F. Utro, **A. Bose**, S. Saha, L. Parida, *RubricOE: what Machine Learning can say about Alzheimers Disease*, ISMB, 2022.
33. **A. Bose**, M.C. Burch, A. Chowdhury, P. Paschou, P. Drineas, *Structure informed clustering for population stratification and genetic risk prediction*, ASHG, 2019.
34. **A. Bose**, F. Utro, D.E. Platt, L. Parida, *Algorithms to modulate ARG by Selection*, RECOMB-Genetics, 2018. **Selected for Platform presentation.**
35. **A. Bose**, V. Kalantzis, E. Kontopoulou, M. Elkady, P. Paschou, P. Drineas, *TeraPCA: a fast and scalable software package to study genetic variation in terascale genotypes*, ASHG, 2017.
36. **A. Bose**, D.E. Platt, L. Parida, P. Paschou, P. Drineas, *Correlation Optimization of Genetics and Geodemographics*, ASHG, 2016. **Selected for Platform presentation.**

Dissertation

37. **A.Bose**, *Computational Methods for Population Genetics*, `https://doi.org/10.25394/PGS.9752924.v1`, Purdue University, 2019.

Patents

1. Cumulant-enabled muti-omics neural network embeddings
   A. Bose, A. Guzmán-Sáenz, K. Rhrissorrakrai, L. Parida
   To be Filed, 2024.
2. Multi-omics Tensor Regression for Complex Diseases
   A. Bose, M.C. Burch, L. Parida
   To be Filed, 2024.
3. Complex disease marker discovery using cumulants and Ising Hamiltonians
   A. Guzmán-Sáenz, **A. Bose**, D. E. Platt, F. Utro, K. Rhrissorrakrai, L. Parida
   To be Filed, 2023.

4. Pharmacogenomics induced protein function of therapeutic targets
   A. Bose, F.Utro, L. Parida
   To be Filed, 2023.
5. Cross-disorder multi-omics feature ranking.
   A. Bose, F. Utro, M.C. Burch, L. Parida
   To be Filed, 2023.
6. Contrastive multi-omics association learning for complex diseases.
   A. Bose, D. Machado Reyes, M.C. Burch, L. Parida
   To be Filed, 2023.
7. A multi-modal Cumulant-based Risk Score for complex diseases.
   A. Bose, L. Parida
   Filed, 2023.
8. Interactive network for multi-modal biomarker discovery for complex diseases.
   A. Guzmán-Sáenz, **A. Bose**, D. E. Platt, L. Parida, N. Haiminen
   Filed, 2022.
9. Multivariate Gaussian GAN for generation of synthetic patient multi-view data for modal incompleteness.
   D.E. Platt, **A. Bose**, K. Rhrissorrakrai, A. Guzmán-Sáenz, N. Haiminen and L. Parida
   United States patent application US 17/930,477, 2024.
10. Discovering biomarkers via higher-order genotype-phenotype interactions in complex diseases.
    A. Bose, D.E. Platt, N. Haiminen and L. Parida
    United States patent application US 17/453,221, 2023.

Google Scholar

- Citations: 112
- h-index: 5
- i10-index: 2

Research Collaborations

- Biotherapeutics Discovery, Boehringer Ingelheim, Ridgefield, CT
- Lerner Research Institute, Cleveland Clinic, Cleveland, OH.
- Cardiovascular Disease Initiative at Broad Institute, Cambridge, MA.
- Biomedical Engineering Department , Rensselaer Polytechnic Institute, Troy, NY.
- Computer Science Department, Purdue University, West Lafayette, IN.

News Articles

1. *Severe COVID linked to RAAS and hyperlipidemia associated metabolic syndrome conditions*
   `https://www.news-medical.net/news/20220406/Severe-COVID-linked-to-RAAS-and-hyperlipidemia-associated-metabolic-syndrome-conditions.aspx`
2. *Combined clinical and genomic data better predicts COVID-19 severity*
   `https://www.news-medical.net/news/20210328/Combined-clinical-and-genomic-data-better-predicts-COVID-19-severity.aspx`
3. *Language (not geography) major force behind Indias gene flow.*
   `https://bigthink.com/culture-religion/indian-genetics`
4. *In India, People Who Speak the Same Language Have Similar DNA: Study*
   `https://theswaddle.com/in-india-people-who-speak-the-same-language-have-similar-dna-study/`

5. *New study ties Indias genetic diversity to language, not geography.* https://www.newsbug.info/lafayette_leader/news/local/new-study-ties-india-s-genetic-diversity-to-language-not-geography/article_52415487-9f63-5ce8-87d4-8edaba12aa0e.html
6. *New study ties Indias genetic diversity to language, not geography.* https://www.purdue.edu/newsroom/releases/2021/Q1/new-study-ties-indias-genetic-diversity-to-language,-not-geography.html
7. *In India, People Who Speak the Same Language Have Similar DNA.* https://theswaddle.com/in-india-people-who-speak-the-same-language-have-similar-dna-study/
8. *Genetic testing has a data problem. New software can help.* https://www.nsf.gov/discoveries/disc_summ.jsp?cntn_id=298521&org=NSF.
9. *Genetic testing has a data problem. New software can help.* https://www.purdue.edu/newsroom/releases/2019/Q2/genetic-testing-has-a-data-problem.-new-software-can-help..html.

Invited Presentations

- **Tutorial** organizer with 50 participants on *Quantum-enabled multi-omics analysis* in Intelligent Systems for Molecular Biology (ISMB) held in Montreal, Quebec, Canada in July 2024.
- **Platform** presentation on *Race versus Genetics in clinical decision-making: a perspective from cardiovascular disease* in Intelligent Systems for Molecular Biology (ISMB) held in Madison, WI in July 2022.
- **Platform** presentation on *A Fast, Provably Accurate Approximation Algorithm for Sparse Principal Component Analysis Reveals Human Genetic Variation Across the World* in Research in Computational Molecular Biology (RECOMB) held in San Diego, CA in May 2022.
- Impact of Clinical and Genomic Factors on COVID-19 Severity
  - IBM Got Science! 2021 Seminar series.
- *Machine Learning framework in Genetic Epidemiology*
  - Broad Institute of MIT and Harvard, Cambridge, MA, Jun 2021.
- *Computational methods in Population Genomics*
  - Regeneron Genetics Center, Tarrytown, NY, Dec 2020.
  - Inari Agriculture Inc., Cambridge, MA, Nov 2020.
  - Allen Institute of Brain Science, Seattle, WA, Nov 2020.
- *CluStrat: a structure informed clustering strategy for population stratification*
  - **Platform** presentation in Research in Computational Molecular Biology (RECOMB), held virtually in June 2020.
  - Poster presentation in American Society of Human Genetics (ASHG) meeting 2019, Houston, TX.
- **Platform** presentation on *Algorithms to modulate ARG by Selection* at the RECOMB-Genetics meeting, Paris, April, 2018. (This talk is given by Dr. Laxmi Parida)
- *TeraPCA: A fast and scalable method to study genetic variation in tera-scale genotypes*
  - Poster presentation in Conference of Scientific Computing and Approximation, Purdue University, West Lafayette, IN.

  - Poster presentation in ASHG 2017 meeting, Orlando, FL.

- *Integrating Linguistics, Social Structure and Geography to model genetic diversity within India.*
  - Poster presentation in Summer Intern Showcase 2017, IBM T.J Watson Research Center, NY.
  - Poster presentation in Biology of Genomes (BOG) 2017 meeting, Cold Spring Harbor Labs, NY.
  - **Platform** presentation in ASHG 2016 meeting, Vancouver, BC, Canada. (Abstract selected in top 8% of over 6000 submissions)
  - Poster presentation at Student Research Showcase in Computer Science Department, Purdue University, West Lafayette, IN.
  - Poster presentation in BOG 2016 meeting, Cold Spring Harbor Labs, NY.
  - Poster presentation in Student Research Symposium 2016 in Computer Science Department, Rensselaer Polytechnic Institute.
- **Summer school** on "Mathematics of Data", organized by **Park City Mathematics Institute (PCMI)** and the Institute for Advanced Study (IAS), held in, Midway, Utah, USA.
- **ASHG 2015**, Annual Meeting in Baltimore, MD, USA as a trainee researcher.
- **Gene Golub SIAM Summer School 2015**, held in, Delphi, Greece.

Mentoring

- PhD students
  - Diego Machado Reyes (4^{th} year, Rensselaer Polytechnic Institute)
  - Myson Burch (Graduated, 2023, Purdue University)
- High School students
  - Inaara Tuan, Mustafa Khan, Justin Gingrich, Romit Ghosh

Awards of Merit

- Plateau from IBM for inventors showcasing creativity and technical knowledge.
- IBM First Patent Application Invention Achievement Award.
- ISCB (International Society for Computational Biology) Travel Fellowship for RECOMB 2020 in Padua, Italy.
- NSF Travel Grants to the following conferences:
  - Biology of Genomes: 2016 and 2017.
  - American Society of Human Genetics (ASHG), 2015 - 2019.
  - International Conference for Distributed Computing and Internet Technologies (ICDCIT) 2017 meeting held at Bhubaneswar, Odisha, India.
- Received a 4 year fellowship from Ministry of Human Resource Development (M.H.R.D), Government of India for significant achievement in Higher Secondary Examination

Professional Activities

- Mentoring a group of five students from Purdue University in CS490, Spring 2024 on the project *Predictive power of lower-dimensional embeddings of single-cell RNA-seq data*.
- Served on the committee of the following PhD students:
  - Diego Machado Reyes, Rensselaer Polytechnic Institute
  - Myson Burch, Purdue University (Graduated 2023)
- Challenge lead in IBM Research projects and external partnerships in 2022 and 2023. Leading a team of over 10 people across global IBM Research labs.
- Reviewer for the following:

  Journals

  Nucleid Acids Research; Journal of the American Heart Association; IEEE Transactions on Computational Biology and Bioinformatics; Bioinformatics; BMC Bioinformatics; Bioinformatics Advances; Computational Biology and Chemistry; Scientific Reports; American Journal of Medical Genetics

  Conferences

  AMIA; Clinical Informatics; ISMB; NeurIPs; KDD; RECOMB; WABI
- Member: American Society of Human Genetics, International Society for Computational Biology.
- Peer Adviser to incoming graduate students in the Computer Science Department in Rensselaer Polytechnic Institute and in Purdue University.
- Co-Founder of the Robotics club of Meghnad Saha Institute of Technology which has over 400 students now.

Computer Skills

- ***Languages:*** Qiskit, PyTorch, Tensorflow, Python, R, C, C++, MATLAB, Java, postgreSQL, Scripting(AWK, bash,etc), Perl, HTML, LaTeX
- ***Operating Systems:*** GNU/Linux, Unix, Windows
- ***Computational Biology:*** SAIGE, REGENIE, BOLT-LMM, SKAT, PLINK, GATK, GCTA, Beagle, bcftools, Cytoscape and other computational biology and population genetic tools and workflows.
- ***Cloud Platforms:*** IBM, Google Cloud, AWS
- ***Databases:*** MySQL, TERADATA, Oracle, DB2

Graduate Coursework (selected)

Machine Learning, Computational Linear Algebra, Parallel Computing, Foundations of Data Science, Algorithms Design, Frontiers of Network Science, Distributed Systems, Randomized Algorithms, Theory of Computation

Independent Coursework

Coursera.org: Deep Learning Specialization; Python for Genomic Data Science; Algorithms: Design and Analysis; Bioinformatics I and II
IBM: Machine Learning Essentials, Qiskit Global Summer School 2023

References

Prof. Petros Drineas
Professor
Department of Computer Science
Purdue University
West Lafayette, IN, USA
drineas@gmail.com

Prof. Peristera Paschou
Associate Professor
Department of Biological Sciences
Purdue University
West Lafayette, IN, USA
ppaschou@gmail.com

Dr. Laxmi Parida
IBM Fellow & Manager,
Computational Genomics,
IBM T.J. Watson Research Center
Yorktown Heights, NY, USA
parida@us.ibm.com

Additional Information

- *Date of Birth*: August 8, 1990
- *Marital Status*: Married
- *Citizenship*: Indian