

# Pratyush Das

• *Email* das160@purdue.edu  
 • *GitHub* <https://github.com/reikdas>

## Education

### Purdue University

August, 2021 - Present

*PhD in Computer Science (Advisor - Milind Kulkarni)*

Research interests - Compilers, Automatic parallelization, Sparse tensors, High performance computing

### Institute of Engineering & Management, Kolkata (MAKAUT)

August, 2017 - May, 2021

*Bachelor of Technology in Computer Science and Engineering*

Awarded the Director's Award for Best Scientific Mind

## Experience

### Swift Platform Experience - Compiler Intern

May, 2023 - August, 2023

Apple

Manager - Richard Wei

- Swift compiler
  - Designed a new Intermediate Representation used internally across multiple teams at Apple.
  - Extended Swift's code generation and runtime to work with this new Intermediate Representation.

### Google Summer of Code - Student

June, 2021 - August, 2021

The LLVM Compiler Infrastructure Organization

Supervisors - William Moses, Johannes Doerfert

- Enzyme: LLVM Pass to perform automatic differentiation of statically analyzable LLVM IR
  - Integrated custom derivatives of several BLAS (Basic Linear Algebra Subprograms) functions into Enzyme.
  - Wrote an LLVM pass to inline function definitions from bitcode files into LLVM IR.

### IRIS-HEP - Fellow

June, 2020 - September, 2020

Princeton University

Supervisor - Jim Pivarski

- Awkward Array: Library for nested, variable-sized data using NumPy-like idioms
  - Implemented a source to source compiler to generate equivalent Python for a subset of C++.
  - Designed a property based testing framework.
  - Created a source to source compiler to generate equivalent parallel CUDA from specification (Python and type info).
  - Enabled automatically parallization of ...

### IRIS-HEP and DIANA-HEP - Fellow

June, 2018 - September, 2018; June, 2019 - September, 2019

Fermi National Accelerator Laboratory and Princeton University

Supervisor - Jim Pivarski

- Uproot: Python implementation of ROOT I/O, an open source file format storing over an exabyte of HEP data
  - Enabled writing fundamental HEP data structures like TTrees and histograms to ROOT files.
  - Uproot has become one of the most widely used HEP libraries.

## Publications

- **P.Das**, A.Xhebraj, T.Rompf, "Specializing Data Access in a Distributed File System (Generative Pearl)", 23rd ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences (GPCE, 2024), DOI: <https://doi.org/10.1145/3689484.3690736>.
- T.Mustafa, K.Kallas, **P.Das**, N.Vasilakis, "DiSh: Dynamic Shell-Script Distribution", 20th USENIX Symposium on Networked Systems Design and Implementation (NSDI, 2023).
- J.Pivarski, I.Osborne, **P.Das**, D.Lange, P.Elmer, "AwkwardForth: accelerating Uproot with an internal DSL", 25th International Conference on Computing in High-Energy and Nuclear Physics (vCHEP, 2021), DOI: 10.1051/epj-conf/202125103002.
- J.Pivarski, I.Osborne, **P.Das**, A.Biswas, P.Elmer, "Awkward Array: JSON-like data, NumPy-like idioms", Proceedings of the 19th Python in Science Conference (SciPy USA, 2020), Pages 68-74, DOI: 10.25080/Majora-342d178e-00b.
- E.Rodrigues, et al., "The Scikit HEP Project - overview and prospects", Proceedings of the 24th International Conference on Computing in High Energy and Nuclear Physics (CHEP, 2019), DOI: 10.1051/epjconf/202024506028.
- N.Saha, **P.Das**, H.N.Saha, "Authorship Attribution of Short Texts using a Multi Layer Perceptron", International Journal of Applied Pattern Recognition, 2018 Vol. 5 No. 3, Pages 251-259, DOI: 10.1504/IJAPR.2018.10016100.

---

## Preprints under submission

---

- **P.Das**, A.Dias, A.Xhebraj, A.Pelenitsyn, K.Sundararajah, M.Kulkarni, “SABLE: Staging Blocked Evaluation of Sparse Matrix Computations”.

---

## Other Open Source Contributions

---

*Supervisor* - Vassil Vassilev

November, 2019 - May, 2021

- **ROOT**: An open-source data analysis framework storing over an exabyte of data
  - Improvements to interpreter (rootcling)
- **Cling**: Interactive C++ interpreter built on top of Clang
  - Maintained cpt.py installer and packager
- **Clang**: C language family frontend for LLVM
  - Several patches to print type information of C++ template arguments

*Supervisors* - Jim Pivarski, Viktor Khristenko

June, 2017 - August, 2017

- **spark-root** - Apache Spark datasource for ROOT
  - Separated spark bindings from TTree reading code
- **root4j** - Java implementation of ROOT file reader
  - Optimized codebase to facilitate interoperability

---

## Teaching Experience

---

**CS 307: Software Engineering** - Purdue University

Spring 2024

**CS 354: Operating Systems** - Purdue University

Fall 2022, Spring 2023, Fall 2023, Spring 2024, Fall 2024

**CS 240: Programming in C** - Purdue University

Fall 2021, Spring 2022

---

## Programming Languages and Tools

---

**Experienced:** Python, CUDA

**Familiar:** C, C++, Java, Scala, Coq, Bash, L<sup>A</sup>T<sub>E</sub>X, Swift, WebAssembly, Triton

---

## Summer Schools

---

**Oregon Programming Languages Summer School** - University of Oregon

2021

**Computational and Data Science for High Energy Physics** - Princeton University

2019

---

## Invited talks at Conferences

---

- Specializing Data Access in a Distributed File System (GPCE) 2024
- GSoC Experience - Enzyme (LLVM Developers' Meeting) 2021
- Python in High Energy Physics (SciPy India, PyCon USA) 2019, 2020
- Writing files with uproot (PyHEP) 2019
- Writing files with uproot (ROOT Users' Workshop) 2018

---

## Invited talks at Research Groups

---

- Language Transformations for the Awkward Array library (IRIS-HEP Fellow Presentations) 2020
- CUDA backend for the Awkward Array project (Princeton University Liberty Research Group) 2020
- Writing TTrees with uproot (IRIS-HEP Topical Meeting: Summer student project presentations) 2019
- Writing files with uproot (DIANA Meeting: Updates on ROOT I/O) 2018
- Separation of Concerns - Refactoring code between ROOT4J and Spark-Root (CMS Big Data Science, DIANA-HEP) 2017

---

## Posters

---

- SABLE: Staging Blocked Evaluation of Sparse Matrix Computations (Midwest PL Summit) 2024

---

## Service

---

- Purdue Programming Languages group (PurPL) co-ordinator 2024 - Present
- Artifact Evaluation Committee (SOSP) 2023