

# Pratyush Das

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

## Education

### Purdue University

August, 2021 - Present

*PhD in Computer Science (Advisors - Milind Kulkarni, Ben Delaware)*

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 user-written Awkward Array kernels.

### 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  |  |   |
| <ul style="list-style-type: none"><li>• <b>P.Das</b>, A.Dias, A.Xhebraj, A.Pelenitsyn, K.Sundararajah, M.Kulkarni, “SABLE: Staging Blocked Evaluation of Sparse Matrix Computations”.</li></ul>   |  |   |
| Other Open Source Contributions   |  |   |
| Supervisor - Vassil Vassilev  |  | November, 2019 - May, 2021                                |
| <ul style="list-style-type: none"><li>• <b>ROOT</b>: An open-source data analysis framework storing over an exabyte of data<ul style="list-style-type: none"><li>- Improvements to interpreter (rootcling)</li></ul></li><li>• <b>Cling</b>: Interactive C++ interpreter built on top of Clang<ul style="list-style-type: none"><li>- Maintained cpt.py installer and packager</li></ul></li><li>• <b>Clang</b>: C language family frontend for LLVM<ul style="list-style-type: none"><li>- Several patches to print type information of C++ template arguments</li></ul></li></ul> |  |   |
| Supervisors - Jim Pivarski, Viktor Khristenko   |  | June, 2017 - August, 2017                                 |
| <ul style="list-style-type: none"><li>• <b>spark-root</b> - Apache Spark datasource for ROOT<ul style="list-style-type: none"><li>- Separated spark bindings from TTree reading code</li></ul></li><li>• <b>root4j</b> - Java implementation of ROOT file reader<ul style="list-style-type: none"><li>- Optimized codebase to facilitate interoperability</li></ul></li></ul>   |  |   |
| Teaching Experience (Teaching Assistant)  |  |   |
| CS 307: Software Engineering - Purdue University  |  | Spring 2025, Fall 2025                                    |
| 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  |  |   |
| <ul style="list-style-type: none"><li>• Specializing Data Access in a Distributed File System (GPCE)</li></ul>  |  | 2024  |
| <ul style="list-style-type: none"><li>• GSoC Experience - Enzyme (LLVM Developers’ Meeting)</li></ul>   |  | 2021  |
| <ul style="list-style-type: none"><li>• Python in High Energy Physics (SciPy India, PyCon USA)</li></ul>  |  | 2019, 2020  |
| <ul style="list-style-type: none"><li>• Writing files with uproot (PyHEP)</li></ul>   |  | 2019  |
| <ul style="list-style-type: none"><li>• Writing files with uproot (ROOT Users’ Workshop)</li></ul>  |  | 2018  |
| Invited talks at Research Groups  |  |   |
| <ul style="list-style-type: none"><li>• Language Transformations for the Awkward Array library (IRIS-HEP Fellow Presentations)</li></ul>  |  | 2020  |
| <ul style="list-style-type: none"><li>• CUDA backend for the Awkward Array project (Princeton University Liberty Research Group)</li></ul>  |  | 2020  |
| <ul style="list-style-type: none"><li>• Writing TTrees with uproot (IRIS-HEP Topical Meeting: Summer student project presentations)</li></ul>   |  | 2019  |
| <ul style="list-style-type: none"><li>• Writing files with uproot (DIANA Meeting: Updates on ROOT I/O)</li></ul>  |  | 2018  |
| <ul style="list-style-type: none"><li>• Separation of Concerns - Refactoring code between ROOT4J and Spark-Root (CMS Big Data Science, DIANA-HEP)</li></ul>   |  | 2017  |
| Posters   |  |   |
| <ul style="list-style-type: none"><li>• SABLE: Staging Blocked Evaluation of Sparse Matrix Computations (Midwest PL Summit)</li></ul>   |  | 2024  |
| Service   |  |   |
| <ul style="list-style-type: none"><li>• Purdue Programming Languages group (PurPL) co-ordinator</li></ul>   |  | 2024,2025   |
| <ul style="list-style-type: none"><li>• Artifact Evaluation Committee (SOSP)</li></ul>  |  | 2023  |