Léopold Cambier

+1(650)665-4481 • leopoldcambier@gmail.com • https://www.leopoldcambier.xyz

Education

- **Stanford University:** Ph.D. in Computational & Mathematical Engineering. 2015—2021. GPA 4.193.
  - Thesis: Fast and scalable hierarchical linear solvers. Advisor: Prof. Eric Darve.
  - Awards: Total Innovation Fellowship, ICME Teaching Fellow.
- **UCLouvain, Belgium:** Bachelor's & Master's in Mathematical Engineering. 2010—2015. *Summa Cum Laude* (18/20).
  - Specializations: optimization, numerical linear algebra, numerical PDEs.
  - Master's thesis: "Robust Low-Rank Matrix Completion" with Prof. P.-A. Absil. Resulted in one publication.

Selected Publications

- L. Cambier and E. Darve. A Task-Based Distributed Parallel Sparsified Nested Dissection Algorithm. To appear in the Proceedings of the Platform for Advanced Scientific Computing (**PASC21**).
- L. Cambier, Y. Qian and E. Darve. TaskTorrent: a Lightweight Distributed Task-Based Runtime System in C++. 2020 IEEE/ACM 3rd Annual Parallel Applications Workshop: Alternatives To MPI+ X (**PAW-ATM**).
- L. Cambier, A. Bhiwandiwalla, T. Gong, M. Nekuii, O. Elibol and H. Tang. Shifted and Squeezed 8-bit Floating Point Format for Low-Precision Training of Deep Neural Networks. **ICLR 2019**.
- L. Cambier, C. Chen, E. Boman, S. Rajamanickam, R. Tuminaro and E. Darve. An Algebraic Sparsified Nested Dissection Algorithm using Low-Rank Approximations. **SIAM Journal on Matrix Analysis and Applications**, 41(2) (2020), 715-746.
- L. Cambier and E. Darve. Fast Low-Rank Kernel Matrix Factorization through Skeletonized Interpolation. **SIAM Journal on Scientific Computing**, 41(3) (2019), A1652–A1680.
- L. Cambier and P.-A. Absil. Robust Low-Rank Matrix Completion by Riemannian Optimization. **SIAM Journal on Scientific Computing**, 38(5) (2016), S440-S460.

Selected Work Experience

- **Software Engineer, NVIDIA.** 2021—Present
  - I work in the cuFFT team, helping accelerate fast Fourier transforms on NVIDIA GPUs.
- **Graduate Intern Technical, Intel.** Summer 2019
  - Project: study of low-precision training of large neural networks. Resulted in one ICLR paper.
    - Implemented a framework based on TensorFlow to simulate low-precision arithmetic and developed the S2FP8 format, a new 8 bits number format designed for training.
    - Simulated training of ResNet on ImageNet, Transformers, GNMT and NCF. S2FP8 can train large networks, reaching the FP32 baseline, with up to 2x improved bandwidth and 4x improved flops.
- **Software Intern, Nvidia.** Summer 2016 and 2017
  - Project: added functionalities to cuDNN (dilated and low-precision convolutions) and cuBLAS (low precision GEMMS) and improved by more than 10x the speed of the test framework, leading to improved heuristics training for convolutions. All work done in C++.
- **Teaching Fellow, Stanford University.** 2017—2019
  - Instructor for Introduction to Python (ICME summer workshop, 2 times). Attendance of more than 100 students, faculty and industry partners.
  - TA for Numerical Linear Algebra (3 times). Office hours and lectures for more than 70 graduate students.

Skills

- Programming languages: C++ (MPI, OpenMP, CUDA and threads), Python, MATLAB, Julia.
- Tools and frameworks: TensorFlow, Pandas (Python), Git, Kubernetes, Slurm, LaTeX, Unix tools.