Producing Software for Science with Class

Damian Rouson
Computer Languages and Systems Software (CLaSS) Group (http://go.lbl.gov/class)

SIAM CSE 2023, 1 March 2023
Software Developed with CLaSS

— LLVM Flang, Caffeine, Matcha, Inference-Engine
— GASNet-EX, UPC, UPC Runtime, UPC++
— MetaHipMer, SIMCoV, symPACK
— Berkeley Quantum Synthesis Toolkit (BQSKit)

Thoughts on Sustainability

— Socially sustainable development
— Technologically sustainable development
— At the intersection of the social and the technological

Conclusions
The Fortran front-end in the LLVM Compiler Infrastructure Project.

Using agile techniques employed across many CLaSS projects, we aim to accelerate Flang’s support for Fortran’s parallel features.

Our agile practices include test-driven development (TDD), continuous integration, pair programming, and git workflows.

TDD

— Compile-time semantics tests for parallel Fortran 2018 features drive our contributions to the LLVM Flang frontend*.
— Parallel runtime tests drive the development of Caffeine.

---

* This research was supported in part by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.

---

For more on Caffeine, see: Rouson & Bonachea (2022) "Caffeine: CoArray System Runtime & Memory Technologies for a Language Framework of Efficient Interfaces to Application GasNet-EX, GASNet-EX", doi:10.25344/S4459B.
Caffeine supports the parallel features of Fortran 2018 for compilers.

Caffeine leverages GASNet-EX, a high-performance networking middleware that undergirds a broad ecosystem of languages, libraries, frameworks, and applications.
Motility analysis of T-cell histories in activation (Matcha)

A parallel virtual T-cell model that captures the speed and turning angle distribution of T-cell motions in tissue.

Matcha is the first target application for Caffeine.
Inference-Engine is a library for researching the efficient runtime inference in high-performance computing (HPC) applications using deep neural networks exported from Python by the companion package nexport.

The implementation language, Fortran 2018, makes it suitable for integration into high-performance computing (HPC) applications. First target: the Intermediate Complexity Atmospheric Research (ICAR) model.

A pure, elemental inference procedure facilitates optimized, including GPU-accelerated, large-batch inference via array statements or do concurrent loops.

https://go.lbl.gov/inference-engine

https://go.lbl.gov/nexport
UPC++ & GASNet-EX

UPC++ is a C++ template library supporting Partitioned Global Address Space (PGAS) parallel programming and interoperability with other common HPC frameworks, including MPI, OpenMP, C++/POSIX threads, CUDA, ROCm/HIP. [https://go.lbl.gov/upcxx](https://go.lbl.gov/upcxx)

UPC++ leverages the GASNet-EX networking middleware to deliver low-overhead, fine-grained communication, including Remote Memory Access (RMA) and Remote Procedure Call (RPC). [https://gasnet.lbl.gov/](https://gasnet.lbl.gov/)
Application case studies

UPC++ has been used successfully in many applications to improve programmer productivity and runtime performance.

We discuss several applications written in UPC++:

- symPack, a sparse symmetric matrix solver
- SIMCoV, agent-based simulation of lungs with COVID
- MetaHipMer, a genome assembler
BQSKit

Berkeley Quantum Synthesis Toolkit

- **QSearch**: Optimal depth synthesis up to four qubits
- **LEAP**: Best quality of solution synthesis up to eight qubits
- **QFAST**: Scales good solution quality synthesis up to eight qubits
- **QGO**: Optimizing compiler combining partitioning and synthesis
- **QUEST**: Scalable circuit approximations
- **QFACTOR**: Fastest quantum optimizer

An optimizing quantum compiler framework.

Quantum synthesis converts a quantum program’s mathematical description, given as a unitary matrix, to an executable quantum circuit.
Software Developed with CLaSS
— LLVM Flang, Caffeine, Matcha, Inference-Engine
— GASNet-EX, UPC, UPC Runtime, UPC++
— MetaHipMer, SIMCoV, symPACK
— Berkeley Quantum Synthesis Toolkit (BQSKit)

Thoughts on Sustainability
— Socially sustainable development
— Technologically sustainable development
— At the intersection of the social and the technological

Conclusions
Socially Sustainable Development

In a diversifying workforce, any sufficiently large project must diversify to be sustainable.
Sustainable Research Pathways

2015 - 2022

Sustainable Research Pathways (SRP)

- Build relationships centered on research collaborations
- Recruit
  - Faculty working with underrepresented students
  - Students from underrepresented backgrounds
- Provide opportunities for staff scientists
- Research collaborations
- Learn/contribute to diversity and inclusion efforts
- Supplement existing D&I Laboratory programs

HPCwire Workforce Diversity and Inclusion Award 2021

2022-

Sustainable Research Pathways for HPC (SRP-HPC)

- Expands SRP into a multi-lab program throughout the ECP community.
- Integrates participants into the broader ECP community through research presentations at the ECP Annual Meeting starting in 2023.
- Normalizes inclusion partly through Guided Affinity Groups and engaging DEI exercises open to all at the ECP Annual Meeting, other planned activities
- Blends the benefits of SRP with the Broader Engagement (BE) program led by Sustainable Horizons Institute (SHI) at SIAM CSE.

Leung ASCAC presentation, July 2021
LLVM Flang & Caffeine Team

Broadening participation includes engaging a diverse ensemble of educational and professional backgrounds.

Hugh Kadhem
M.S. Pure Mathematics
Ph.D. Candidate, Math

Katherine Rasmussen
B.S. History/M.S. Linguistics

Brad Richardson
B.S./M.S., Nuclear Engineering

Damian Rouson
B.S./M.S./Ph.D., Mechanical Engineering
Engaging Users in Development Broadens Participation

The Caffeine Proposition:

- A subset of the Fortran 2018 non-parallel features suffice for writing a runtime library, mostly in Fortran, to support the Fortran 2018 parallel features.

- Writing a parallel runtime library in the language of the users improves sustainability by lowering a barrier to community maintenance.

- In Caffeine, writing in Fortran also improves sustainability by reducing complexity and maintenance costs.
Engaging Users in Development Broadens Participation

The Caffeine Proposition:

A subset of the Fortran 2018 non-parallel features suffice for writing a runtime library, mostly in Fortran, to support the Fortran 2018 parallel features.

Writing a parallel runtime library in the language of the users improves sustainability by lowering a barrier to community maintenance.

In Caffeine, writing in Fortran also improves sustainability by reducing complexity and maintenance costs.
Develop Collaboratively and Train

The BQSKit Team

Costin Iancu
Senior Staff Scientist

Ed Younis
Computer Systems Engineer

Wim Lavrijsen
Computer Systems Engineer

Marc Davis
Graduate Student

Mathias Weiden
Graduate Student

Lindsay Bassman
Postdoctoral Fellow

Justin Kalloor
Graduate Student

Roel Van Beeumen
Research Scientist

Alon Kukliansky
Graduate Student

Ji Liu
Postdoctoral Appointee

Roel is a computational scientist.
Technologically Sustainable Development

Backwards compatibility ensures continuity of user experience, protects users’ investments in code and thus improves sustainability.

- GASNet-EX (2002)
- GASNet-1

- Fortran 2018

- Fortran 66 (1957)
Reducing maintenance costs improves sustainability.

- Application source code never directly references GASNet.
- Applications benefit from new GASNet feature releases without revising their applications.

- Parallel programming languages require specialized compilers.
- Using UPC++ requires only a standard C++ compiler.
- This compiler-free approach greatly reduces the size of the code base that the developers of UPC++ must maintain.
At the Intersection of Social & Technical Workflow

- Run ICAR & Save Training Data
- Import training data into PyTorch and train neural network
- Run nexport to export network to JSON
- Import network into ICAR via Inference-Engine and validate
The Telephone Game

Solution Time = Development Time + Runtime
Runtime Training in ICAR with embedded Inference-Engine

Sustainable Workflow

Rinse, Repeat...
Software Developed with CLaSS
— LLVM Flang, Caffeine, Matcha, Inference-Engine
— GASNet-EX, UPC, UPC Runtime, UPC++
— MetaHipMer, SIMCoV, symPACK
— Berkeley Quantum Synthesis Toolkit (BQSKit)

Thoughts on Sustainability
— Socially sustainable development
— Technologically sustainable development
— At the intersection of the social and the technological

Conclusions
Conclusions

The CLaSS Group co-develops open-source software:

1. Parallel programming compilers, runtime libraries, networking middleware,
2. An optimizing quantum synthesis toolkit,
3. HPC deep learning tools, and
4. Some targeted applications that use the above.

Our socially sustainable development practices include

1. Broadening participation through workforce development programs, involving contributors with varied educational backgrounds, and lowering barriers to community maintenance
2. Developing openly and collaboratively
3. Training new entrants to the field.

Our technologically sustainable practices include

1. High levels of backwards compatibility.
2. Minimizing maintenance costs for users and contributors.

Exciting avenues for exploration lie at the intersection between the social and the technological.