

## What each PI must do:

- ☒ ~~(Ed) Section 1.1.1: We need two paragraphs on your work in the first year~~
- ☐ **(Adrian)** Section 1.1.1: Can you add one more sentence of detail on graph query xcel?
- ☐ **(Adrian)** Section 1.1.1: Can you add one more sentence of detail on Caiman?
- ☒ ~~(Mike) Section 1.1.1: Can you add two more sentences of detail on panorama system?~~
- ☒ ~~(Pjotr) Section 1.1.1: We need more detail on biWFA and odgi~~
- ☒ ~~(Pjotr) Section 1.1.1: Please take pass through paragraph on human pangemoe~~
- ☒ ~~(Zhiru) Section 1.1.1: Should we have a paragraph on the genotype imputation work?~~
- ☐ **(All)** Section 1.3: Anything more to add about mentoring students?
- ☒ ~~(Pjotr) Section 1.5: Please add paragraph on goals for RT1 in year 2 (think broadly, not just your own work!)~~
- ☐ **(Adrian)** Section 1.5: Please add paragraph on goals for RT2 in year 2 (think broadly, not just your own work!)
- ☒ ~~(Ed) Section 1.5: Please add paragraph on goals for RT3 in year 2 (think broadly, not just your own work!)~~
- ☐ **(Erik)** Section 1.1.1: Sentence or two on differential privacy algorithm that you posted on the blog.

## 1. Accomplishments - What was done? What was learned?

“For NSF purposes, the PI should provide accomplishments in the context of the NSF merit review criteria of intellectual merit and broader impacts, and program specific review criteria specified in the solicitation. Please include any transformative outcomes or unanticipated discoveries as part of the Accomplishment section.”

“The PI is reminded that the grantee is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.”

“For purposes of NSF, please see Exhibit VII-1 of the Proposal & Award Policies & Procedures Guide for a complete listing of Grantee Notifications To and Requests For Approval From the National Science Foundation.”

### 1.1. What are the major goals of the project?

“List the major goals of the project as stated in the approved application or as approved by the agency. If the application lists milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.”

“Generally, the goals will not change from one reporting period to the next. However, if the awarding agency approved changes to the goals during the reporting period, list the revised goals and objectives. Also explain any significant changes in approach or methods from the agency approved application or plan.”

“For purposes of NSF, what are the major research goals and objectives of the project?”

Semiconductor technology scaling is slowing, and as a consequence, computer systems must increasingly rely on a heterogeneous mix of general-purpose and specialized computing engines. At the same time, computer users are tackling new problems with massive datasets that must be manipulated in irregular and rapidly changing ways while maintaining strict privacy guarantees. Efficiently supporting big, sparse, dynamic, and private data to solve large complex problems on heterogeneous systems is one of the grand challenges in software/hardware systems research. To address this grand challenge, the Panorama project is taking a truly cross-stack approach spanning applications, programming languages, compilers, architecture, security, and privacy including use of a one-of-a-kind Panorama prototype system to explore integrated rack-scale acceleration for computational pangenomics.

The project is pursuing five research goals organized into five tightly interconnected research thrusts. The first three thrusts are structured from top-down across the computing stack. Research Thrust 1 (RT1) investigates new computational pangenomics data structures and algorithms and will develop PanoBench, a new benchmark suite suitable for driving the remaining thrusts. Research Thrust 2 (RT2) investigates new programming language and compiler techniques including new DSLs, intermediate representations, compilers, and techniques targeting Panorama accelerators. Research Thrust 3 (RT3) investigates new computer architectures with support for a whole-rack manycore with 1M+ cores and application-specific accelerators for computational pangenomics. The final two thrusts cut across both software and hardware. Research Thrust 4 (RT4) investigates new security and privacy techniques including scalable secure computation and differential privacy and homomorphic encryption for pangenomics. Research Thrust 5 (RT5) involves holistically evaluating the research ideas in the other thrusts through the use of a one-of-a-kind Panorama prototype system.

This project is also pursuing two broader impact initiatives. Broader Impact Initiative (BI1) seeks to increase participation of under-represented minority students in computer science by developing a low-level computer systems module for a new four-week summer program targeting rising sophomores. Broader Impact Initiative (BI2) involves specific plans to grow the open-source software/hardware ecosystem in the computational biology and computer systems communities.

## **1.2. What was accomplished under these goals and objectives (you must provide information for at least one of the 4 categories below)?**

“As the project progresses, the emphasis in reporting in this section should shift from reporting activities to reporting accomplishments”

### **1.2.1. Major Activities**

The Panorama team has been meeting weekly since the start of the project with 15-20 researchers spanning 4+ different time zones regularly attending these hour-long meetings via zoom. These meetings vary in format. Each semester starts with PIs proposing a concrete list of independent and collaboration projects that they will be working on to help meet the overall project goals, and each semester ends with PIs presenting what they have accomplished on these projects. Some meetings involve short research project collaboration proposals or research project status updates. Other meetings focus on an introductory tutorial by one team member on topics such as typical computational pangenomic workflows, genotype imputation, hyperdimensional computing, garbled circuits, trusted execution environments, and GNU Guix package management. Other meetings involve a deep technical dive on a single research project. We have also hosted external speakers during our weekly meetings including Efraim Flashner

(open-source developer, spoke about his work on Guix packaging for RISC-V), Ekaitz Zarrag (open-source developer, spoke about bootstrapping Guix for RISC-V), Santiago Marco-Sola (postdoc at Universitat Politècnica de Catalunya in Spain, spoke about his work on GPU acceleration for wavefront algorithm for sequence alignment), and Andrea Guarracino (postdoc at Human Technopole in Italy, spoke about his work on the odgi tool). The team is making extensive use of zoom, slack, github, and google drive to facilitate collaboration across the three participating universities.

During the first year of the project, the Panorama team has made promising early progress on all five research thrusts including new algorithms and software tools for computational pangenomics; new domain-specific languages for pangenome query acceleration and heterogeneous hardware platforms; ... and steady progress on the Panorama prototype system.

#### *### Research Thrust 1 (RT1): Panorama Computational Pangenomics*

In RT1, the Panorama team has been developing new algorithms for computational pangenomics, implementing these algorithms in state-of-the-art computational pangenomic tools, and then using workload characterization to identify opportunities for pure-software performance optimization.

PI Garrison has developed the bidirectional Wavefront Alignment (BiWFA) algorithm in collaboration with Santiago Marco-Sola and Miquel Moreto at the Universitat Politècnica de Catalunya in Spain. The recent Wavefront Alignment (WFA) algorithm proposed by Santiago and Miquel can perform exact gap-affine alignment in  $O(ns)$  time and  $O(s^2)$  space, where  $s$  is the optimal score and  $n$  is the sequence length. The BiWFA algorithm is the first gap-affine algorithm capable of computing optimal alignments in  $O(s)$  memory while also retaining WFA's time complexity of  $O(ns)$ . In practice, our implementation never requires more than a few hundred MBs when aligning noisy Oxford Nanopore Technologies reads up to 1 Mbp long while maintaining competitive execution times. This work is currently under review and is published as a preprint.

PI Garrison has also developed the wfmash algorithm that builds up on BiWFA to create an aligner for pangenomes based on sparse homology mapping. wfmash uses a variant of MashMap to find large-scale sequence homologies. It then obtains base-level alignments using BiWFA, via the wfln hierarchical wavefront alignment algorithm. wfmash is designed to make whole genome alignment easy. On a modest compute node, whole genome alignments of gigabase-scale genomes should take minutes to hours, depending on sequence divergence. It can handle high sequence divergence, with average nucleotide identity between input sequences as low as 70%.

PI Garrison also designed the seqwish algorithm, which builds a variation graph from a set of sequences and alignments between them. It first transforms the alignment set into an implicit interval tree. To build up the variation graph, query this tree-based representation of the alignments to reduce transitive matches into single DNA segments in a sequence graph. By recording the mapping from input sequence to output graph, we can trace the original paths through this graph, yielding a pangenome variation graph. We present an implementation that operates in external memory, using disk-backed data structures and lock-free parallel methods to drive the core graph induction step. We demonstrate that our method scales to very large graph induction problems by applying it to build pangenome graphs for several species. This work as published this year in the Oxford Bioinformatics Journal.

PI Garrison helped lead the development of Optimized Dynamic Genome/Graph Implementation (ODGI), a novel toolsuite that implements scalable algorithms and has an efficient in-memory representation of DNA pangenome graphs in the form of variation graphs. ODGI supports pre-built graphs in the Graphical

Fragment Assembly format and includes tools for detecting complex regions, extracting pangenomic loci, removing artifacts, exploratory analysis, manipulation, validation and visualization. Its fast parallel execution facilitates routine pangenomic tasks, as well as pipelines that can quickly answer complex biological questions of gigabase-scale pangenome graphs. This work was published this year in the Oxford Bioinformatics Journal.

The BiWFA, wfmash, seqwisth, and ODGI tools are all open-source. PI Garrison and PI Prins, and others within the Human Pangenome Reference Consortium have been leveraging these tools to develop a new human pangenome reference genome. This community effort will create a more sophisticated and complete human reference genome with a graph-based, telomere-to-telomere representation of global genomic diversity. All of these algorithms, the corresponding tools, and representative datasets from the human pangenome reference genome are being used to create PanoBench, a computational pangenomic benchmark suite for driving the research in the rest of the project. To this end, PI Garrison has started working with PI Zhang and PI Sampson on a detailed workload characterization of the BiWFA, wfmash, seqish, and ODGI tools on realistic pangenome datasets.

Using ODGI from PanoBench, PI Zhang and his students conducted comprehensive profiling and performance characterization and identified pangenome layout algorithms as a critical bottleneck in current computational pangenomic workflows. These layout algorithms can provide effective visualization that reveals the detailed structure of regions of the human pangenome, which were completely invisible to genomics researchers before. However, generating the layout of a whole chromosome can take many hours even using massively parallel shared memory machines. PI Zhang and PI Garrison collaborated to develop an efficient implementation for a new pangenome layout algorithm that uses a lock-free method for parallel path-guided stochastic gradient descent. This new algorithm and high-performance implementation achieves 5-7x speedup over the state-of-the-art. The code has been contributed to the upstream odgi repository. This work illustrates the benefit of Panorama's interdisciplinary approach; the expertise in performance-oriented systems-level programming from PI Zhang's group combined with the algorithm and software work by PI Garrison produced meaningful performance improvements in real-world tools that are used every day by computational pangenomics researchers. The team is currently preparing a paper on this work to be submitted in early 2023.

#### *### Research Thrust 2 (RT2): Panorama PL and Compilers*

In RT2, the Panorama team has been focusing on exploring new domain-specific languages (DSLs) to help connect our ongoing work in RT1 on computational pangenomics (as captured in PanoBench/odgi) to the diverse range of computer architectures being explored in RT3.

Using ODGI from PanoBench, PI Sampson and his students have identified common algorithmic patterns around querying pangenomic data structures that are computationally intensive in the context of realistic pangenomes. PI Sampson's team is now working to add FPGA-based hardware acceleration to PanoBench/odgi in a flexible way, by automatically generating hardware implementations from high-level descriptions of the algorithms. PI Sampson's team is defining a DSL that describes the space of common queries and is starting to implement a compiler, leveraging the lab's existing open-source Calyx infrastructure, that translates this code into efficient hardware designs. ((Adrian: maybe one more sentence of detail?))

PI Sampson's lab has also been working on a new language and compiler approach for leveraging heterogeneous hardware platforms, which are pervasive in the Panorama system design. The new language, called Caiman, comprises facets that describe the abstract computation separately from its

partitioning and scheduling onto disparate hardware resources. The key objective is a theoretical guarantee that the program's semantics remain unchanged across different decisions about hardware mapping, i.e., that heterogeneity cannot introduce bugs. ((Adrian: maybe one more sentence of detail?)) The prototype compiler already demonstrates the basic principles working; future research stages will apply the design to hardware being explored elsewhere in the Panorama project (e.g., the compute-in-SRAM accelerator platform from GSI technologies, the Panorama prototype system being developed at UW).

### ### Research Thrust 3 (RT3): Panorama Computer Architecture

PI Batten has been collaborating with PI Taylor on software runtimes to help improve the programmability of the partitioned-global-address-space (PGAS) manycore architectures that are the key to the Panorama architecture's million-core design. These PGAS manycore architectures use scratchpad memories to improve the efficiency and scaling of the memory system by removing the need for a coherence protocol and associated network traffic. When used effectively, SPMs can yield critical performance and energy savings by reducing data movement, improving synchronization times, and eliminating overheads that can arise from false sharing. Unfortunately, manycore architectures that rely heavily on SPMs are notoriously challenging to program. To address this programmability challenge, we have developed a dynamic task parallel programming framework with a work-stealing runtime for such manycore architectures. Our preliminary results suggest that such a runtime is not just feasible but can also significantly improve the performance of irregular workloads when executing on these architectures. We are also exploring optimizations that allow the runtime to leverage unused SPM space for further performance benefit. Early results show 1.2-28.5x speedup on workloads that benefit from our techniques, and only minimal overhead for workloads that do not. A paper on this work is under review, and we are exploring mapping PanoBench/odgi workloads to this runtime.

In addition to more general-purpose manycore architectures, the Panorama project is also exploring specialized architectures for computational pangenomics. Early in the project, PI Batten gained access to a truly unique accelerator system commercially developed by GSI Technologies, a niche SRAM company. The accelerator is based on associative processing, a 50-year-old idea implemented via modern compute-in-SRAM techniques. PI Batten has done some work on associative processing evaluated through simulation, but now we have an opportunity to evaluate this idea on a real system based on a commercial-scale custom silicon implementation. Our hypothesis is that the massively parallel, low-precision, data-intensive nature of computational pangenomics makes it a good match for this compute-in-SRAM accelerator platform. Over the past year, we have been collaborating with Dan Ilan from GSI Technologies, and we have made significant progress developing a vector-based intermediate programming abstraction, learning best practices for low-level microcoding, microbenchmarking microcode fragments, and starting to evaluate alignment algorithms in PanoBench. We also hosted a virtual workshop for the entire Panorama team to give researchers hands-on experience using this compute-in-SRAM accelerator platform, and this workshop has sparked new collaborations with PI Zhang and PI Sampson. A paper on our microbenchmarking work is under review.

PI Batten has also started a collaboration with Santiago Marco-Sola and Miquel Moreto from the Universitat Politècnica de Catalunya in Spain. This collaboration was jump-started through an informal discussion and then engaging brainstorming session by PI Batten, PI Prins, Santiago, Miquel, and our graduate students at the 49th Int'l Symp. on Computer Architecture (ISCA) in New York City, NY. Santiago and Miquel have already been successfully collaborating with PI Garrison on accelerating the novel wavefront algorithm (WFA) for sequence alignment using both FPGAs and GPUs. WFA uses gap-affine-based sequence alignment, but edit-based sequence alignment has also shown promise in

many genomic workflows and edit-based sequence alignment algorithms can potentially offer even larger opportunities for hardware acceleration. We have begun exploring both dedicated hardware accelerators and instruction set specialization for the bitap algorithm and the Myer's bitparallel algorithms, and more importantly, we are exploring extensions to these algorithms to enable not just sequence-to-sequence but also sequence-to-graph alignment, which is critical for computational pangenomics and can eventually be integrated into PanoBench. A paper on our preliminary work in this direction is under review.

#### *### Research Thrust 4 (RT4): Panorama Security and Privacy*

In RT4, the Panorama team has been developing new security and privacy techniques for large-scale workloads and heterogeneous rack-scale systems. The effort in the first year focused on developing an effective combination of two secure computing techniques (i.e., trusted execution environments (TEEs) and secure multi-party computation (MPC)) and investigating algorithms to provide differential privacy guarantees in pangenomics.

PI Suh's team developed a new secure MPC protocol, named STAMP, which enables efficient and more easily deployable privacy-preserving machine learning inference by combining MPC with a lightweight TEE (LTEE). The traditional MPC protocols can be applied to commodity hardware, yet incur significant (>100x) performance overhead. Secure hardware such as Intel SGX, often referred to as TEEs, provides high performance yet requires custom hardware support for each computing platform (CPU, GPU, FPGA, etc.) and is vulnerable to side-channel attacks. STAMP leverages the intuition that non-linear operations that are expensive in MPC protocols are cheap in plaintext computation, and offloads non-linear operations to a small TEE, similar to today's discrete trusted platform modules (TPMs) or SoC security subsystems such as the Apple enclave processor, while leveraging MPC on commodity computing engines for linear operations. STAMP achieves significant performance improvements compared to state-of-the-art MPC protocols; STAMP is 4-63x faster and reduces the communication by 3.8-12x compared to the state-of-the-art MPC protocols - Falcon (PoPETs'21) and AriaNN (PoPETs'22). The work is under submission for a conference publication.

Extending STAMP, PI Suh and PI Zhang also collaborated in enabling secure computation for genomics workloads. In particular, the team observed that hyper-dimensional computing (HDC), an emerging paradigm in machine learning, mostly depends on linear operations, which can be performed efficiently in an MPC protocol. Leveraging this observation, the team developed a secure HDC-based genotype imputation scheme. The initial study shows a promising result; our secure genotype imputation using HDC+MPC is over 12 times faster than the previous work based on homomorphic encryption.

The team started an effort to develop a way to construct human pangenome graph with a differential privacy (DP) guarantee. The goal is to enable allowing human pangenome graph to be constructed and publicly released using genome data from a diverse set of genome data by ensuring strong privacy for individuals. As the first step, PI Garrison developed an algorithm to allow common haplotype information to be protected with DP by adding noise in the frequency counts of sub-haplotype variations. ... ((Erik: could you add a sentence or two on the DP algorithm that you posted on the blog?))

#### *### Research Thrust 5 (RT5): Panorama Prototype System*

PI Taylor's team is working on bringing up the Panorama prototype system which is a rack-scale system with custom rackmount chassis each with a custom motherboard, daughter card, FPGAs, and ASICs interconnected with hundreds of 100 GB/s links. This prototyping effort has been particularly challenging given the global supply chain slowdown; additional effort has been spent on rethinking design aspects to

mitigate these supply chain and other manufacturability issues. A new daughtercard design has been developed specifically for the needs of the Panorama project including support for significantly larger DRAM capacities. This new daughtercard design features 4 Zynq Ultrascale+ die serving as memory controllers to 4 DDR4 DIMMs and connecting to the primary ASIC socket using high-speed single-ended DDR links. The ASIC socket design has also been generalized as a generic platform so it can support a variety of PPOSS-specific ASICs. We have completed the placement of all components on the daughterboard, and are currently finalizing the pin connectivity, routing and decap selection. We have recently added the capability for the daughterboard will be able to function stand-alone or in the context of the Panorama Rack scale system.

### **1.2.2. Specific Objectives**

n/a

### **1.2.3. Significant Results**

n/a

### **1.2.4. Key Outcomes or Other Achievements**

n/a

## **1.3. What opportunities for training and professional development has the project provided?**

“Describe opportunities for training and professional development provided to anyone who worked on the project or anyone who was involved in the activities supported by the project. "Training" activities are those in which individuals with advanced professional skills and experience assist others in attaining greater proficiency.”

“Training activities may include, for example, courses or one-on-one work with a mentor. "Professional development" activities result in increased knowledge or skill in one's area of expertise and may include workshops, conferences, seminars, study groups, and individual study. Include participation in conferences, workshops, and seminars not listed under major activities.”

“If the research is not intended to provide training and professional development opportunities or there is nothing significant to report during this reporting period, please check "Nothing to Report" if applicable.”

“For NSF purposes, please summarize the contributions to the research and teaching skills and experience of those who have worked on the project, including undergraduate students, graduate students, post-docs, college faculty, and K-12 teachers. If your project supported postdoctoral researchers, then you must include a summary of the mentoring activities conducted.”

**ALL PIs:** Each PI should include a paragraph about how you have been mentoring the postdocs, graduate students, or undergraduate students that you listed in our nsf-pposs-panorama-people



spreadsheet and are working on the project. Talk about how they are gaining skills in research and what conferences you sent them to.

The Panorama weekly meetings have offered a supportive environment for both graduate and undergraduate students to practice their technical communication skills especially to other researchers outside their specific area of technical focus. PI Batten has made a concerted effort to reach out to student presenters across the project with constructive feedback after they present in the Panorama weekly meetings. Students working on the project have attended several conferences to help grow their professional networks (see next section for list). For many students these were the first in-person conferences they have attended since the start of the pandemic, and indeed some students attended their first in-person conference ever.

The Panorama project has also served as a context for mentoring junior researchers in some of their first experiences with independent work in computer science. In PI Sampson's lab, for example, three undergraduate students and one post-graduation research associate began their first projects on designing and implementing compilers: a skill that they all intend to bring to either PhD research or industrial applications after they graduate. Three PhD students, in addition to leading research projects of their own, served as co-mentors for one or more undergraduate collaborators, which required them to develop their skills as project managers and one-on-one instruction in technical skills. In PI Zhang's group, several undergraduates spent last summer exploring mapping machine learning algorithms to the compute-in-SRAM accelerator platform. In PI Batten's lab, an undergraduate student led the microbenchmarking work of the compute-in-SRAM accelerator platform and was first-author on a paper currently under review.

Finally, the project has also served as an environment for more senior researchers to hone their skills. For example, Dustin Richmond, a postdoc on the team who was engaged in the early phase of the project, has started as an tenure-track assistant professor in the UC Santa Cruz department of computer science and engineering.

#### **1.4. Have the results been disseminated to communities of interest? If so, please provide details.**

“Describe how the results have been disseminated to communities of interest. Include any outreach activities that have been undertaken to reach members of communities who are not usually aware of these research activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.”

**ALL PIs:** Each PI should include a paragraph that lists the papers they have published but no need to go into detail on those papers since we should have discussed those in detail in the major activities or key outcomes section. This is where you could also mention blog or twitter posts which we used to disseminate work done in this project too. This is also where you can discuss conferences and workshops you and/or your students attended whether the travel was funded by this grant or not. And you can discuss any invited talks, visits to companies, etc. Any place where you had a chance to talk about the work we are doing in this grant.

Several PhD students working on the project attended the 49th Int'l Symp. on Computer Architecture (ISCA) in New York City, NY. The conference featured two workshop papers by Panorama authors related to the open-source software ecosystem, and several papers on computational genomics including a paper specifically on hardware acceleration for computational pangenomics by UIUC, CMU, and ETH



Zurich. Members of the Panorama team enjoyed a vigorous intellectual debate with the SeGraM researchers, and we were encouraged to see computational pangenomics start to gain recognition as an important emerging workload within the computer architecture community.

PI Sampson and his students in the Panorama project attended the The 27th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS 2022) in Lausanne, Switzerland to promote the ongoing research in the project and to discuss its relevance to the adjacent fields of computer architecture, programming languages, and operating systems.

PI Zhang, and his graduate students involved in this project have attended multiple conferences in computer hardware and systems, in order to learn about recent work in the field, discuss their recent research results, and to generally engage industry and the research community. This past year, these conferences included DAC, FCCM, FPGA, and ISCA. PI Zhang delivered invited talks on hardware specialization at DATE'22 and DAC'22, as well as seminars hosted by Intel, Qualcomm, Rutgers, and the NSF A3D3 Institute.

PI Garrison gives talks on Pangenomics, including the HPRC, a course on advanced Bioinformatics organized by PI Prins (May 2022) and a summer school in Italy (July 2022). PIs Prins and Garrison organised a virtual conference on pangenomics (December 2021) attended by some 40 researchers, including Panorama PhDs. PI Prins gave a talk on GNU efforts at CARRV ISCA (June 2022) and helped organize Guix days and a devroom on 'Declarative and Minimalistic Computing' at FOSDEM (February 2022). In general, PIs Garrison and Prins are active in the free software communities for pangenomics and GNU Guix software distribution projects for RISC-V.

## **1.5. What do you plan to do during the next reporting period to accomplish the goals?**

["Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives."](#)

For RT1, in the second year of the project we will continue to develop and optimize the wfmask and odgi tools and use these tools to improve genotyping and all-2-all comparisons of pangenomes to build a 'tree of life'. We will also continue packaging these tools into PanoBench and begin porting these tools to RISC-V for use in other parts of the project. On the pure computational biology side, we will look at acrocentrics and other areas of variation and recombination in DNA of human, mouse, and rat to help motivate the specific computational pangenomic problems studied in this project.

For RT2, in the second year of the project we will ... ((Adrian: can you write this?))

For RT3, in the second year of the project we will continue our work on improving the performance and productivity of manycore architectures which form the heart of the million-core Panorama system. We will leverage our early work exploring a commercial-scale compute-in-SRAM accelerator platform to now study mapping various sequence-to-sequence and sequence-to-graph alignment algorithms to this platform. We will continue our collaboration with UPC to investigate both dedicated hardware accelerators and instruction set specialization for computational pangenomics, and we will begin exploring the potential for silicon photonic interconnection networks to enable manycore architectures to scale across an entire rack-scale system.

For RT4, in the second year of the project we will continue the two main research directions: secure genotype imputation and differentially private (DP) pangenome graph construction. The main goal for the secure genotype imputation is to improve the accuracy of the secure machine learning (ML) model. While HDC with MPC showed promising performance, its accuracy cannot match that of more complex ML models. We plan to further explore different ML model architectures and additional co-optimizations of an ML model and secure computing algorithms. For the DP pangenome graph construction, we plan to explore generating synthetic pangenome graphs using a Transformer model. Instead of directly constructing pangenome graphs from individual genome sequences, we plan to first train a Transformer model, which is shown to be effective in natural language processing, using DP, and use synthetic graphs from the trained Transformer for pangenome analyses.

For RT5, in the second year of the project we will complete the Panorama focused daughter card and bring up the baseline Panorama prototype system. We will also bring online the FPGA emulation portion of the Panorama prototype system to enable experimenting with novel accelerators for computational pangenomics at scale. Our goal is for the Panorama prototype system to be ready for other Panorama researchers to start experimenting with near the end of the second year of the project.

----

#### **Earlier Input from PIs:**

The team's subproject on hardware acceleration for querying pangenomic graph datasets will deliver its first suite of algorithm-specific hardware designs. The goal is to demonstrate their performance advantage over the odgi package's execution on a traditional multicore CPU. The work will also deliver an open-source implementation of a DSL that expresses these query algorithms at an abstraction level suitable for high-performance implementation. Similarly, the Caiman project on languages for orchestrating heterogeneous hardware will deliver an open-source implementation along with a report detailing its theoretical correctness guarantees.

PI Zhang's team will continue their work on accelerating important pangenome workloads, such as the layout algorithm for pangenome graph visualization. In addition to speeding up multithreaded implementation on multicore CPUs, the team will further explore GPU-based layout algorithms and aim to reduce the overall time to visualize a large multi-million node pangenome graph from hours to just minutes.

We plan to complete the V1 daughtercard design in RT5 by the end of Year 2 of the program; and to bring up the baseline Panorama Prototype; as well as enabling FPGA emulation of next-generation Panorama concepts.

The biology pangenomics team plans to continue developing new and enhanced methods for pangenomics and privacy. This includes improving pangenomics tools for the RISC-V platform in close collaboration with all Panorama initiatives.

## 2. Impact - What is the impact of the project? How has it contributed?

“Over the years, this base of knowledge, techniques, people, and infrastructure is drawn upon again and again for application to commercial technology and the economy, to health and safety, to cost-efficient environmental protection, to the solution of social problems, to numerous other aspects of the public welfare, and to other fields of endeavor.”

“The taxpaying public and its representatives deserve a periodic assessment to show them how the investments they make benefit the nation. Through this reporting format, and especially this section, recipients provide that assessment and make the case for Federal funding of research and education.”

“Agencies use this information to assess how their research programs: increase the body of knowledge and techniques; enlarge the pool of people trained to develop that knowledge and techniques or put it to use; and improve the physical, institutional, and information resources that enable those people to get their training and perform their functions.”

### 2.1. What is the impact on the development of the principal discipline(s) of the project?

“Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research and/or pedagogical methods in the principal disciplinary field(s) of the project.”

“For NSF purposes, the paragraph should read, How the fields or disciplines are defined is not as important as covering the impact the work has had on knowledge and technique.”

“For purposes of NSF, provide postdoctoral mentoring or other mentoring.”

This section is where we can talk about our Broader Impact Initiative (BI2) on open-source software/hardware advocacy. Eventually, I think this is where we can also talk about how our work is having an impact in our specific fields. I am open to suggestions. Citation counts? Number of groups using our tools? For our first year we can just say this is a new project and so the impact will come in later years?

We have made significant progress on Broader Impact Initiative 2 (BI2) which seeks to grow the open-source software/hardware ecosystem in the computational biology and computer systems communities.

PI Batten has continued his advocacy for free and open-source hardware within the computer architecture community in four key ways. (1) PI Batten was a co-organizer of the NSF Workshop on Integrated Circuit Research, Education, and Workforce Development held in San Jose, CA. This workshop brought together 40 participants from academia, government, and industry including representatives from fabless semiconductor companies, multi-project wafer companies, foundries, and EDA vendors. The workshop included working groups on technology nodes, packaging & testing, IP, EDA tool flows, and training & education. A report was published to help spur investment in this area. (2) PI Batten organized a DAC special/invited session titled "Breaking Down Physical Design Barriers with Open and Agile Flow Tools". The session included three presentations from UC Berkeley, Stanford, and Zero ASIC on novel flow tools that are attempting to address the physical design challenge from different perspectives including new

flow tool generators with a strong emphasis on modularity and reusability and a new distributed-systems approach to enable warehouse-scale execution of sophisticated flows. (3) A group of undergraduates on PI Batten's research team taped out BRGTC5: a 2x2.5mm chip in TSMC 180nm designed and implemented using the PyMTL3 framework. The chip includes a RISC-V TinyRV2 five-stage pipelined microcontroller with a 32KB instruction memory, 32KB data memory, 4–8 digital I/Os and a SPI master interface to enable attaching peripherals, low-power sleep mode which wakes up on a digital input, and an SPI minion interface to enable a host computer to test the chip and load programs. The chip used a sophisticated PyMTL3-based pre- and post-silicon testing strategy. (4) PI Batten helped lead a tapeout on SKY130 through the efabless Chiplgnite program this Spring as part of teaching ECE 5745 Complex Digital ASIC Design. 15 students (including sophomores, juniors, seniors, MEng, and PhD students) in four groups developed four different blocks: a CRC32 checksum implemented using high-level synthesis, a latency insensitive synthesizable memory using PyMTL3, a systolic array multiplier implemented using SystemVerilog, and a greatest common divisor unit. Each block also included a custom SPI interface for isolated testing.

PI Prins and PI Garrison have continued their advocacy for free and open-source software within the computational biology community. As part of the Panorama project, they have developed and released key open-source projects, including `odgi` and `vcflib`. PI Prins has also been a leader within the GNU Guix open-source community. Guix is a mature functional cross-platform package manager with hundreds of committers and over 20K packages. Guix has a strong emphasis on transparent and reproducible software development and enables bootstrapping entire Linux distributions from just a few hundred bytes of binary assembly code. PI Prins has been at the frontier of promoting RISC-V within the GNU Guix community. He is mentoring several open-source software developers working on porting packages to RISC-V (over 16K packages ported to date!) and bootstrapping a RISC-V Linux distribution using GNU Mes (a small C compiler which can be built from a just a small bit of RISC-V assembly). PI Prins is also working to port key computational pangenomics tools to GNU Guix to simplify installing, managing, and reproducing computational pangenomic workflows.

A highlight of our work on BI2 was a collaboration by PI Batten, PI Prins, and PI Garrison to make the case for using Guix in reproducible computer architecture research. Members of the Panorama team worked together to develop proof-of-concept Guix packages for the spike C++ RISC-V instruction set simulator, the gem5 C++ cycle-level RISC-V simulator, the Ariane Verilog RTL RISC-V model, and an example genomics workload which supported cross-compilation for RISC-V. We then used these packages to write two position papers presented at workshops held in conjunction with the 49th Int'l Symp. on Computer Architecture (ISCA) in New York City, NY. The first paper was titled "The Case for Using Guix to Solve the gem5 Packaging Problem" and was presented at the gem5 Users' Workshop. The second paper was titled "The Case for Using Guix to Enable Reproducible RISC-V Software & Hardware" and was presented at the 6th Workshop on Computer Architecture Research with RISC-V (CARRV). Both papers were well received and introduced many in the computer architecture community to the Guix packaging system for the first time. We are looking forward to continuing to using Guix as the centerpiece of our BI2 open-source software/hardware advocacy efforts.

## 2.2. What is the impact on other disciplines?

"Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines."

Not sure if we want to write anything here ... this would be for disciplines other than computational biology, computer architecture, compilers, PL, security, and privacy.

### 2.3. What is the impact on the development of human resources?

“For example, how has the project: provided opportunities for research and teaching in the relevant fields; improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions; developed and disseminated new educational materials or provided scholarships; or provided exposure to science and technology for practitioners, teachers, young people, or other members of the public?”

“For purposes of NSF, this should address all efforts to broaden participation in science and engineering.”

“For purposes of NSF, how has the project provided opportunities for research, teaching and mentoring in science and engineering areas.”

This section is where we can talk about our Broader Impact Initiative (BI1) on CS More each year.

We began our work on Broader Impact Initiative 1 (BI1) which seeks to increase participation of under-represented minority students in computer science through Cornell's CSMore pre-sophomore summer program. CSMore is a response to evidence that shows that disadvantaged groups in Cornell's CS curriculum encounter disproportionate difficulty when making the transition from first-year introductory courses to second-year advanced core topics. CSMore offers a funded, 4-week "preview" of sophomore-level courses that serve as a head start to make these challenging topics more accessible during the coming year. PI Sampson and PI Batten led the 2022 edition of CSMore's unit on digital logic, computer arithmetic, and simple processors. The PIs revamped the curriculum with all new lectures, in-class active learning exercises, and a newly designed series of hands-on lab assignments. By the end of the week, students had incrementally developed a "Femto-Processor" from just basic logic gates in the same logic design tools that students will encounter during Cornell's sophomore-level computer organization course, CS 3410. The Femto-Processor supported two instructions (i.e., write immediate and add) and was capable of executing up to 16-instruction programs to do simple arithmetic and generate the Fibonacci sequence.

PI Sampson and PI Batten ran the revamped unit for a class of 29 students from underrepresented backgrounds. In surveys after the program, students called the unit "very dynamic, interactive, and animated" and reported that "this class and it drastically raised my interest in and understanding of hardware" and that they "feel much more comfortable having seen a preview of the hardware material."

### 2.4. What was the impact on teaching and educational experiences?

“Describe how the project made an impact or is likely to make an impact on teaching and educational experiences. For example, has the project: developed and disseminated new educational materials; led to ideas for new approaches to course design or pedagogical methods; or developed online resources that will be useful for teachers and students and other school staff?”

**ALL PIs:** This is where you can mention if you have incorporated any aspects of the project into your teaching.

### 2.5. What is the impact on physical resources that form infrastructure?

“Describe ways, if any, in which the project made an impact, or is likely to make an impact, on physical resources that form infrastructure, including physical resources such as facilities, laboratories, or instruments.”

We have made progress on bringing up the Panorama prototype system at the University of Washington, which will be a new kind of rack-scale FPGA supercomputer with reconfigurable ASIC augmentation, that can be used as a resource for prototyping new kinds of scale-out systems and corresponding workloads. More details are in the major activities section.

## **2.6. What is the impact on institutional resources that form infrastructure?**

“Describe ways, if any, in which the project made an impact, or is likely to make an impact, on institutional resources that form infrastructure, including: institutional resources (such as establishment or sustenance of societies or organizations); including: information resources, electronic means for accessing such resources or for scientific communication, or the like.”

I don't think we usually ever write anything here.

## **2.7. What is the impact on information resources that form infrastructure?**

“Describe ways, if any, in which the project made an impact, or is likely to make an impact, on information resources that form infrastructure, including: information resources, electronic means for accessing such resources or for scientific communication, or the like.”

“For NSF purposes, the institutional resources parenthetical language should read: (such as policies, practices, programs, or establishment or sustenance of societies or organizations); and information resources should include data services and preservation.”

I don't think we usually ever write anything here.

## **2.8. What is the impact on technology transfer?**

“Describe ways in which the project made an impact, or is likely to make an impact, on commercial technology or public use. Including: transfer of results to entities in government or industry; instances where the research has led to the initiation of a start-up company; or adoption of new practices.”

I don't think we usually ever write anything here. I can just write some general stuff on open-source and refer to section 2.1

## **2.9. What is the impact on society beyond science and technology?**

“Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world. For example, in areas such as: improving public knowledge, attitudes, skills, and abilities; changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or improving social, economic, civic, or environmental conditions.”

n/a for now ...





### **3. Changes / Problems**

The PI is reminded that the grantee is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

#### **3.1 Actual or Anticipated problems or delays and actions or plans to resolve them**

Some GRA funding from the first year will be rolled over and used in the second year. Some participants were funded on first-year fellowships or had partial funding from synergistic grants. We have spent time during the first year recruiting new PhD students to work on the project and also transitioning mid-career PhD students to this project. The project is now fully staffed with a strong group of researchers including undergraduate students, graduate students, and postdocs.

PI Prins had an issue with his working visa because of backlogs in processing with the USA consulate from December 2021 to May 2022. More specifically, PI Prins' H1B visa expired and it took six months to schedule an appointment at the Netherlands consulate due to the backlog caused by pandemic travel restrictions. This required temporarily designating PI Garrison as the lead PI from UTHSC, although PI Prins continued his research duties on the project without interruption. We are happy to announce that Dr. Prins has resumed his role as the lead PI from UTHSC in June 2022 after resolving the H1B visa issue.

PI Prins and PI Garrison are in the process of submitting a budget adjustment request to NSF to change the funding for a GRA to a Postdoc. UTHSC is a medical school and it has proven difficult to find a GRA capable of the highly technical work required for the Panorama project. We did however find qualified postdocs interested to participating in this type of effort. We have supplemental NIH funding to make this a possibility.

PI Taylor reports some delays on the Panorama prototype system development due to the semiconductor supply chain crisis. His team has worked to mitigate these delays by redesigning the daughterboard and careful management of the corresponding component supplies and fabrication partners.