

**Research Experience for Undergraduates:  
Combinatorics, Algorithms, and AI applied to Real problems**  
PI: William Gasarch, co-PI: Laxman Dhulipala

## 1 Overview

This is a renewal request for the 2022–2024 REU grant titled *Combinatorics, Algorithms, and AI Applied to Real problems (CAAR)*. The theme of the grant since 2013 has been, and continues to be, *the interplay of theory and practice*.

### 1.1 Objectives and Intellectual Focus

Undergraduates in computer science take *theory* courses such as discrete math, algorithms, cryptography, complexity and formal languages; they also take *practical courses* on topics such as networking, databases, compiler design, and data science. Some courses can be taught either theoretically or practically such as machine learning. Sometimes students see connections between theory and practice (e.g., compiler design and formal languages, probability and machine learning) but often they do not. While there is some research combining theory and practice, undergraduates are rarely exposed to it. In REU-CAAR, students will do projects that either (1) apply theory to practice directly such as Cryptography-Security, Statistics-Machine Learning, or (2) apply theory to practice indirectly (or speculatively) such as quantum computing, graph theory, (3) apply practice to theory such as using programs to spot patterns that one can then prove. All of these types of projects will broaden their horizons, break down the artificial walls between disciplines, and expose them to problem formulation and modeling.

Our goal is to have a 10 week summer program that trains and mentors 10 to 20 undergraduates (10 by funded by NSF, 5-10 by other sources) and several (unfunded) high school students. Our primary objective is to recruit students who have a strong background in Discrete Mathematics, Programming, and Algorithms (some of the projects will require more specialized knowledge). At the same time, we are making every attempt to reach out to both underrepresented students and students at non-research schools (including 2-year colleges). As such, some of our projects have fewer prerequisites hence lowering the barrier-to-entry.

### 1.2 The Department's and the University's Commitment to the Project

We originally got this grant for 2013-2015, and it has been renewed three times. Over the last twelve years we have rapidly built up and significantly expanded the program to well beyond the size possible with just NSF funding.

With the help of **the Computer Science Department** we have found additional funds to support many more students, including students from underrepresented groups. In 2022–2024 we had 29 NSF-funded students and 25 non-NSF funded students. We also had 8 high school students participating informally. In addition the department has (1) helped us borrow equipment when needed (e.g., monitors and laptops), (2) had the computer staff help when needed and (3) supplied administrative personal.

**The University** has helped us with housing, food, and parking passes.

### 1.3 Student Recruitment

Our program has and will continue to focus on diversity and inclusion, in terms of cohort composition. As evidence of that latter statement, in 2022-2024 we had 61 students (that includes the HS students) of which 33 self-identified as Male, and 28 as Female; 3 were African-American, 4 were Hispanic, 1 was a Pacific-Islander, and the rest were White or Asian. We have also been keen on having students from non-research schools: Of the 61 students, 28 were from research schools, 25 were from non-research schools, and 8 were from high schools. We will describe Student Recruitment and admissions in Section 4.

### 1.4 Timeline for Admissions and Program

We will use the following timeline.

**Nov 1:** Post website and set up admissions. Begin making contacts and visiting schools, advertising, and recruiting. (This will be delayed if the decision to fund REU-CAAR has not been made yet.)

**March 1:** Deadline to apply.

**Mar 1-Mar 20:** Acceptances are emailed out in rounds to control how many come. The final list of who is coming will be known by April 1.

**Apr:** Mentors will be in contact with the students on their projects.

**Early June:** Students check into the dorms and have a welcome dinner with the PI and co-PI. The program runs for 10 weeks from early June to mid August.

**Early August:** Students will give talks on their projects.

**Early August:** Students will make arrangements with their mentors to keep in touch and continue work on their projects in the fall and perhaps beyond.

## 2 Research Environment

### 2.1 Student Teams

Students work in teams of 2-5 on a research project. When they apply, the students indicate which projects they are interested in, and we match students to projects. Each project has a mentor who meets the team frequently (3 times a week for the first month and once a week after that). Some of the projects will also have a co-mentor. A common combination is to have professor as mentor who will lay out the project and what needs to be done, and give general guidance, and meets with the students once a week, but also have and a grad student or postdoc co-mentor who meets more often and gives specific guidance.

### 2.2 Student Activities

Below we list some of our plans for student activities. Most of these have been done in Summer 2022, 2023, and 2024. These activities helped make for a better research environment and also impart valuable information to the students.

Some of the activities will be done jointly with the following programs: REU-BRIDGE (UMCP program: Bioinformatics Research in Data Science for Genomics), REU-EXERCISE (Salisbury University program: Explore Emerging Computing in Science and Engineering), (Since they are far away they only come to our campus once.) TRAILS (UMCP program: Trustworthy AI in Law and Society). We may add other programs if they arise and could benefit from the association. We note which activities are joint.

We plan to continue these associations which benefit all of the programs involved. The lunches and activities mentioned are funded privately by a donor, not by the NSF.

1. On the first day there is a talk that lays out what the program is about what is expected of them, and what they should expect of us. In the first week, there are talks by the mentors on each project. In the afternoons each group meets with their mentors.
2. Mondays the program provides a lunch, accompanied by an activity. Sample activities - (1) work on a fun math problem in groups, (2) discuss a TED talk we assigned them to watch on the prior day, (3) participate in a fun activity such as playing Telepictionary. Joint with some of the other programs depending on the activity.
3. Wednesdays at 4:00 (so it does not interfere with research) there is either (1) a research talk (e.g., a talk on *Understanding How People Share Passwords*), (2) a talk about research (e.g., a talk on *The Ethics of Science*), or (3) a talk on professional development. (e.g., (1) a former grad student who is in industry talks about their career path, (2) workshop on resume writing), Joint with all of the programs listed above (REU-EXERCISE attends by zoom).
4. Graduate school Lunch: There are talks by the CS Grad Director, and the Applied Math Grad Director, and a panel discussions about graduate school by graduate students. Topics covered are (1) is graduate school right for you?, (2) advice on how to get into graduate school, (3) what is graduate school really like, and (4) career options. Joint with all of the programs. For this one, REU-EXERCISE comes in person (see next point).
5. We try to have all of the programs attend each others talks. (1) The day of the Grad School Lunch is the only day that REU-EXERCISE comes to the campus. We take advantage of that and have them give presentations of their projects. (2) The 10th week of the program REU-CAAR, REU-BRIDGE, and TRAILS present their final talks and all go to each others groups talks. REU-EXERCISE will attend by zoom.
6. Social Activities: (a) Game Nights: For entertainment there are game nights where the students played games. We provided pizza for dinner (not funded by the NSF). Joint with all of the other programs except REU-EXERCISE since they are so far away. (b) A weekend Hike. (c) A trip to the Zoo. (d) Other activities that take advantage of being close to Washington DC.

### **2.3 Qualification of the Mentors**

We list the mentors, how many times they have mentored in an REU program, and a list of recent papers with undergraduates or high school students. They have all mentored many students on projects outside of the REU programs as well (except Aviva and Auguste who are graduate students).

All of the mentors have taken students to conferences, helped them network, and taught them not just the field of study, but also how to think about projects and research.

1. Victor Albert. Has mentored twice for REU-CAAR. Six papers with undergraduate students: [2], [3], [34], [38], [36], [41],
2. Auguste Gezalyan. Has mentored three REU-CAAR. Has three papers with students: [30], [31], [42]. (Dave Mount and Auguste Gezalyan are co-mentors.)

3. Jordan Boyd-Graber. Has mentored once for REU-CAAR. Ten papers with undergraduate students: [33], [45], [46], [47], [48], [49], [50], [53], [54], [55].
4. Andrew Childs. Has mentored six times for REU-CAAR. Four papers with undergraduates: [5], [10], [11], [12].
5. Dana Dachman-Soled. Has mentored five times for REU-CAAR. One paper with undergraduate students: [6].
6. Laxman Dhulipala. Has mentored twice for REU-CAAR. Five papers with undergraduates: [19], [20], [21], [52], [56].
7. William Gasarch. Has mentored eight times for REU-CAAR. five papers and one book with students: [4], [14], [27], [25], [26], [28] (book), [29].
8. Erin Molloy. Has mentored four times in REU programs, twice as a grad student and twice as a faculty for REU-BRIDGE. Three papers with students: [7], [13], [18], and one Emerging Research National (ERN) Conference poster [8].
9. Dave Mount. Has mentored three REU-CAAR. Has three papers with students: [30], [31]. [42], (Dave Mount and Auguste Gezalyan are co-mentors.)
10. Aviva Prins. Two REU-CAAR. One paper with students: [39].

Seven of the mentors are faculty in Computer Science, one is a professor in Electrical Engineering (Dana Dachman-Soled) and two are computer science graduate student (Auguste Gezalyan, Aviva Prins). Seven are male and three are female (Dana Dachman-Soled, Erin Molloy, and Aviva Prins).

### 3 Sample Projects

All of the projects require basic discrete math and some programming skill. If something additional is required then we state it as a prerequisite.

#### 3.0.1 Classical and Quantum Error Correction (Victor Albert)

*Prerequisite: Linear Algebra, At least one of the following: Quantum Computing, Quantum Mechanics.*

Error correction is what ensures that the audio in your phone calls remains sharp, your hard drives do not deteriorate too quickly, and signals can be reliably transmitted to remote satellites.

Over multiple decades, and with the explosion of the information age, an enormous variety of error-correction schemes were developed. Recently, a radically new type of error correction was introduced, one that can protect the quantum information that is stored in a quantum computer or that is communicated over a quantum network.

We created the Error Correction Zoo [1] to categorize and to organize known classical and quantum error-correction schemes. In this summer project, students will learn both classical and quantum error correction and contribute entries to the zoo's repository. The project can be scaled. On the easy end, there are tables and other readymade info to enter. On the hard end, the students can learn and write about advanced topics such as topological codes (see topological) and approximate quantum error correction (see approximate).

### 3.0.2 Generating Multimodal Examples for Adversarial Machine Learning (Jordan Boyd-Graber)

*Prerequisite: Programming, Linear Algebra, Building Interactive Web Applications.*

What makes a question difficult to computers? To humans? This can be measured through a metric called *item response theory* [44], which (in its extended forms) measures the properties of an example that are challenging to humans or computers [51]. Determining what is hard for computers and easy for humans (and vice versa) is useful for measuring the strengths and weaknesses of AI models.

Text-only models are very good at understanding questions; however, models that have to reason across modalities (e.g., answer a question by looking at a picture) are significantly worse than their text-only counterparts [32].

Students will build interactive interfaces where people can *author* new questions based on an image or sound file, see how a computer would answer the question, and then edit the question to make it more difficult for computers without making it more difficult for humans [55].

By creating interfaces for adversarial authoring, they will understand how to explain the predictions of AI models using nearest neighbors [37], influence functions [35], local approximations [43], and others.

This project can extend over multiple years: different modalities (audio, photographs, artwork, video), different forms of questions (common sense, cultural reasoning, mathematical reasoning), and different forms of answers (multiple choice, short form, long form).

### 3.0.3 Software Tools for Quantum Simulation (Andrew Childs)

*Prerequisites: Linear Algebra, Programming Experience, Knowledge of Quantum a plus but not Required*

Simulating quantum mechanics is arguably the most promising use case for quantum computers, with potential applications to chemistry, materials science, and condensed matter, nuclear, and high-energy physics. Some of the most effective quantum algorithms for simulating quantum dynamics on quantum computers are based on product formulas, which approximate the evolution according to a given Hamiltonian as a product of evolutions of its elementary terms. Recent work has provided good theoretical bounds on the quality of such approximations, but these bounds can be difficult to compute in practice, especially for high-order simulations of complicated quantum systems.

Building on past work at the University of Maryland—including a prior REU project—this project will develop computational methods for calculating product formula error bounds and implement them in software, handling a variety of physical models that include systems of spins, fermions, and/or bosons, and taking into account typical features of realistic systems such as translation invariance and particle number conservation. We will also aim to develop improved theoretical error bounds that are both reasonably tight and amenable to computation. These tools will improve our understanding of resource requirements for practical quantum simulation, and should also be useful for implementing large-scale simulations as quantum computing platforms mature.

### 3.0.4 Concrete Security Estimation for Post-Quantum Cryptosystems with Side Information. (Note: this is not a Quantum Project.) (Dana Dachman-Soled)

*Prerequisite: Python, Discrete Math, Algorithms, Linear Algebra. Knowledge of quantum computing is NOT needed for this project*

Most current cryptosystems depend on factoring (or other problems in Number Theory) being Difficult. Quantum computers pose a potential threat to the security of these systems since these problems are all

Easy for quantum computers. There is an ongoing effort by NIST to standardize a suite of so-called "post-quantum" cryptographic systems that will remain secure in the presence of a quantum computer.

One of the foremost avenues for efficient, post-quantum cryptosystems, is to construct cryptosystems from *lattice problems*, a type of mathematical problem that is believed to be hard even for quantum computers. It is important to research the best algorithms for breaking cryptosystems in order to set the parameters large enough.

What if the attacker has side information such as how much power the system is using? In prior work [16],[17],[15] we have developed publicly available toolkits that provide estimates for the concrete security of lattice-based cryptosystems when side information is incorporated.

In this project, students will experiment with the Toolkit to analyze the effects of side-channel attacks on lattice-based cryptosystems, using publicly available datasets. The students will gain experience in programming in Sage/Python and will learn about the state-of-the-art algorithms for solving lattice problems. Students may also participate in Dachman-Soled's group's current effort to extend and improve the Toolkit.

### 3.0.5 Space-Efficient Parallel Algorithms (Laxman Dhulipala)

*Prerequisites: Knowledge of elementary data structures and algorithms, programming skills in C++*

In parallel algorithm design we are interested in designing algorithms that obtain good speedup over the fastest sequential algorithms as we increase the number of cores used. The two primary measures of interest are the *work* of the algorithm, i.e., the total number of basic operations performed, and the *depth*, i.e., the longest chain of sequential dependencies in the algorithm. Due to the increasing size of modern data, designing algorithms that also minimize the amount of *space* used by the algorithm while having high parallelism are of interest.

This project will explore theoretical and practical parallel algorithms that have provably low space usage. Some directions we may consider include designing succinct or compact parallel algorithms, space-efficient parallel graph algorithms, and space-efficient parallel data structures.

### 3.0.6 The Ramsey Theory of Equations: An Empirical Approach (William Gasarch)

*Prerequisites: Programming skills and Combinatorics*

$(\mathbb{N} = \{1, 2, \dots\})$  and  $\mathbb{Z} = \{\dots - 3, -2, -1, -0, 1, 2, 3, \dots\}$ .

The following is known: For all finite colorings of  $\mathbb{N}$  there exists  $x, y, z$  that are the same color such that  $x + y - z = 0$ .

What about other equations? What about systems of equations? What if the number of colors is (say) 8? Rado's theorem answers many of these questions for homogenous systems of linear equations; however, even for a single quadratic equations, little is known.

The students will then consider equations  $E(x_1, \dots, x_n)$ , together with number-of-colors  $c$ , for which it is unknown if there is a  $c$ -coloring of  $\mathbb{N}$  without a monochromatic solution. For such pairs they will write a program to try to find a coloring of  $\mathbb{N}$  with no monochromatic solution to investigate the matter. The programming may require cleverness and special data structures.

The students will learn some Ramsey Theory and how to use programming to investigate open questions in mathematics.

### 3.0.7 Reconstructing Evolutionary Histories from Genomic Data (Erin Molloy)

*Prerequisites: Python, Probability, Discrete Math. Exposure to algorithms and C++ is preferred but not*

*required. No prior knowledge of biology is required.*

(This project is in REU-CAAR instead of REU-BRIDGE since this project involves combinatorial optimization which is a better fit with REU-CAAR.)

Reconstructing evolutionary histories (called phylogenies) from genomic data is a fundamental problem in computational biology. However, even with significant progress over the last decade, there are many open computational challenges. First, the current leading methods are heuristics for NP-hard optimization problems which sometimes take a long time and sometimes are wrong. Second, methods typically make simplifying model assumptions that are violated in practice. Third, methods typically assume that the input data are perfect, although genomic data can have high rates of missingness and/or error depending on the data collection techniques and processing pipelines.

REU students will join active projects in our lab that seek to address these computational challenges and develop fast and accurate computational methods for studying vertebrate evolution (with applications to human evolution and heritable diseases), agricultural pest evolution (with applications to resource management), virus evolution (with applications to epidemics like SARS-COV-2), or tumor evolution (with applications to cancer therapeutics).

REU students will gain experience designing and benchmarking algorithms on synthetic and real genomic data sets. Depending on the project, they will also have the opportunity to interact with our biological collaborators who provide novel data sets and inform the interpretation of biological results.

### **3.0.8 Computational Hilbert Geometry (Dave Mount and Auguste Gezalayan)**

*Prerequisites Knowledge of discrete structures, programming skills, very basic algorithms.*

The field of computational geometry looks at problems in Geometry computationally. For example, given  $n$  points in the plane, find which two are closest together. But there are other metrics [58] that have been useful in fields such as Genetics, Probability, and Physics. In this project we will be examining the Hilbert Metric [57], from the perspective of computational geometry. We will explore questions involving this metric. Examples of objects to study are (1) the probability simplex [40] and (2) areas [59]. The project will include contributing to a software package to help us visualize these structures. In this project you will learn computational geometry and combine computational geometry with programming.

## **4 Student Recruitment and Selection**

CAAR admissions will be run by Principal Investigator William Gasarch (henceforth *the PI* or *William*) and co-Principal Investigator Laxman Dhulipala. (henceforth *the co-PI* or *Laxman*). The recruiting effort is nationwide.

REU-CAAR has generated tremendous interest and many students have applied. (2024: 470 applicants of which about 200 were qualified, though we could only take 21).

1. We set up an REU-CAAR website which is on the NSF list of websites for REU programs.
2. We have, and will continue to, send representatives to the Grace Hopper Conference to talk to and encourage qualified women to apply.
3. Since beginning this REU program the PI has made contacts at University of Texas-Pan American (large Hispanic population), Haverford, Swarthmore, Bryn Mawr (non research schools), Morehouse, Spelman (both are HBCU's, and Spelman is a Women's College), and Howard (HBCU). He will use

these contacts to encourage qualified students from these schools to apply. He will also visit some of these schools.

4. Faculty mentors will contact top liberal arts colleges including Williams, Amherst, Pomona, Middlebury, Bowdoin, and Carleton.
5. Former students will act as ambassadors for us and tell qualified students about our program. This has already happened at several schools.
6. The high school students who were involved in the program informally may apply in later years when they are undergraduates. This has already happened several times.
7. Some of the projects have a lower prerequisite than others thus making them appropriate for people with less background, which is often people from underrepresented groups or non-research schools. (Of the sample projects Laxman's Parallelism project, and Auguste-Dave Hilbert Geometry project do not need much prerequisite.)

Our criteria to accept students are the following:

1. A high quality academic record including good grades in relevant courses and strong letters.
2. A good match of skills and interest for the available projects.
3. For some projects, specialized knowledge will be required.
4. Diversity.
5. Being from a non-research school (over half will be from such schools).
6. A compelling personal essay stating why they want to come.

We will be using the NSF Education and Training Application (ETAP) to manage student applications and to collect student demographic information. This information will be useful in tracking how well the program is doing for diversity and non-research schools.

## **5 Broader Impact**

The students in the program are given a chance to do research (this is especially important for those from non-research schools that might not have other opportunities) that they ordinarily would not have had (students in underrepresented groups are often in this category). The students in the program will learn how to do research as well as be exposed to modern topics from the fields of Theory and AI. This yields a long term benefit of (1) demystifying research, (2) building confidence in their problem-solving abilities, and (3) giving them information on career options (PhD and Industry). This will help them in whatever branch of Computer Science they go into. This yields a broad long term societal benefit.

The projects have short-term, medium-term, or potential long-term impact. Some of the projects are on-going so they have already had an impact.

1. The Comp Biology project has already lead to faster ways to build Phylogenic Trees, which has medical applications even in the short term.



2. The Crops project is being done in conjunction with Kheyti, a nonprofit in India, and the initial results have already been used help farmer in India have a bigger crop yield.
3. The Parallelism project may have an impact in the medium-term. With Moore's law about over, we have to rely on parallelism for fast computations. Note that the parallelism project is about real problems that people care about on real multicore machines.
4. The Quantum project may lead to long term society benefits but that is harder to say since its benefits are long term. Similar comments apply to the other projects.

## **6 Intellectual Merit**

The students in this REU program will work on projects that apply theory to practice. For undergraduates this is somewhat novel as theory and practice are often taught in different courses.

Roughly speaking the projects are in the following fields: algorithms (this permeates all of the projects), computational geometry, parallel algorithms for data science, cryptography as it applies to security, quantum computing, and other.

It will be an intellectual challenge to take questions from the real world (that is, not toy problems), make them rigorous, solve them, and then apply the results back to the real world. This will involve problem formulation and theoretical analysis. The challenge is to have questions that (1) are complex enough so that their answers are helpful in the real world, yet (2) simple enough so that they can be solved or partially solved within the timeframe of the REU.

## **7 Mentor Selection, Training, and Monitoring**

The PI emails the faculty asking who wants to be a mentor (a faculty can also suggest their graduate student as co-mentor). Once people respond the PI chooses based on the following criteria: (1) Diversity of Area. Note that in 2024 we had 3 projects on quantum computing, 2 on algorithms, and 2 on AI. This is a good balance. (2) Prior experience. As noted above, all of the mentors had mentored before. (More on first-time mentors later.) (3) Underrepresented groups. Note that the above list of mentors has six males and three females.

The mentors have several meetings before the program begins to review good mentoring practice and to talk to the first-time mentors. First time mentors (which may include graduate students will also be trained for mentoring undergraduates from underrepresented groups. For this we will work with the Iribe Initiative for Inclusion and Diversity to provide graduate student mentors with resources.

The PI will meet with the mentors once every 2 weeks, and (separately) the students every 2 weeks, to check that there are no problems. If the PI is also a mentor then the coPI will check in with the PI's group. If there are problems then the PI will talk to the mentor about it and find a solution. (True Example: the students thought the mentor spend to much time on review. I told the mentor this and he realized the students were correct and then spend much less time on review.)

## **8 Setting Expectations of Student Behavior**

The first day of the program the PI gives a talk about the program including what the mentors expect from the students, what the students should expect from the mentors, and what the PI expects from both of them.

This includes obvious things like showing up on time, and subtle ones like proper ways to credit prior work.

On the first day we also discuss respect for the other people in the program (and other people in general) what it means and how to deal with it if its being violated. Sexual harassment is treated as part of this; however, we also go over the rules for reporting it and where to go for help on it.

As part of the Wednesday afternoon talk series (see Section 2.2) there is a talk on the ethics of science. Many examples are given of unethical science such as deceptive graphs and p-hacking. There is a a discussion of the grey areas as well and how to spot if someone else, or even yourself, is wandering into unethical behavior. Plagiarism is also discussed.

## **9 Professional Development for Students**

The students are exposed to different career paths via (a) the grad student lunch, and (b) there will be talks at some of the other lunches from former grad students or professors who now work in industry.

There will be a talk on how to give a talk. For their final presentation they will first give a runthrough in front of a faculty member who is not in the program, and a student who is not in the area of the project. The students get a sharp critique of the talk and often need to redo their slides and re-present. The experience of seeing what they did wrong, and correcting it, will serve them well in their careers when they give future talks.

The same process happens, albeit less formally, with the students' writing. They will write up some of the work they did and their mentor will give them feedback on it.

They also learn (1) how to listen to a talk: after each talk during the summer we discuss what they learned and (2) how to read a paper: their mentors will give them reading assignments and they later discuss the papers with their mentors.

Learning how to give a talk, listen to a talk, read a paper, and write a paper will all serve them well whatever career path they take. And they learn this in a safe environment where, if they fail, the only consequence is to improve.

## **10 Project Evaluation and Reporting**

### **10.1 Evaluation of REU-CAAR**

REU-CAAR will work with the Computing Research Association's Center for Evaluating the Research Pipeline (CERP), which is an evaluation center that has been contracted by NSF CISE to provide evaluation for REU Sites and REU Supplements. CERP's evaluation work will focus on measuring the impact of our REU on students' self-perceptions (e.g., self-efficacy; scientific identity), academic development (e.g., research productivity; skills proficiency) and professional aspirations (e.g., intentions to pursue graduate school; career goals). At the end of data collection, CERP will provide a report that summarizes evaluation results alongside a comparison group of responses collected from similar REUs. Demographic data and other student characteristics will be provided in the report with an intersectional lens when possible, enabling the project team to understand the impact of REU-CAAR on different types of students. A designated liaison (likely the PI William Gasarch or the coPI Laxman Dhulipala) will be responsible for distributing the surveys and communicating with the CERP team, providing any information related to REU-CAAR that is necessary for data collection and reporting.

In addition we have in the past, and will in the future, ask the students to anonymously tell us how the program is going so that we can improve it.

## 10.2 Long Term Tracking

We will keep track of papers that came out of their work with REU-CAAR.

We will also, once a year, email the students asking them several questions to help us evaluate the long terms effect of the program. We will ask:

1. Where are you now? (e.g., Grad School? Industry?)
2. What effect has being in REU-CAAR had on you?

## 11 Qualifications of the PI and co-PI

### 11.1 PI William Gasarch

#### coPI and PI of REU-CAAR

From 2013-2016 William was the co-PI on the REU-CAAR grant. From 2016-2024 William was the PI on the REU-CAAR grant. In both time frames William did the following: recruited mentors, recruited students, did the admissions, set up activities and talks for the students, tracked the students after the left the program, maintained the REU-CAAR website, and coordinated activities with the other REU programs. He has also mentored projects in the program every year except 2018, 2021, 2023, 2024. By outreaching to various organizations he has also gotten more funding for the program, to pay for more students.

#### SPIRAL Summer Math Program

William was involved with SPIRAL (Summer Program in Research and Learning) from 2003-2007. This program was run by the Math Department to help math majors from HBCU's (Historically Black Colleges and Universities). Every summer, students applied and between 10 and 20 were admitted. William designed and taught a course on learning math through games, and designed many of the projects. William ran a similar program for CS in the Summer of 2005.

#### Undergrad Students advised by Dr. William Gasarch

Since 2014, William has mentored 61 undergraduates. Of these (1) 18 are either in computer science (or a related field) grad school or have completed such a grad school with a Masters or Phd, (2) 15 are still undergraduates, and (3) 28 are employed in computer science.

#### Other Undergraduate Advising Activities

Every year a set of students is selected to give out awards for which professors most influenced them. Sergey Brin (1993), Amy Castner (2004), and Katrina LaCurts (2007) all chose William. Formally he received, all three years, *Certificate of Teaching Excellence In Recognition of Significant Influence and Contributions to the Education of Outstanding Graduating Senior Sergey Brin (Amy Castner, Katrina LaCurts)*. Sergey Brin went on to found Google.

#### High School Mentoring

Since 2014 William has mentored over 200 High School students on projects. Of these, over 80 went to top places (Harvard, MIT, Stanford, Chicago, Princeton, Cal Tech, CMU) and the rest went to very good colleges or are still in High School.

Jacob Lurie, who William mentored in 1995 won the Westinghouse award, and later went on to be a full professor of Mathematics at Harvard. He won the Breakthrough award and the MacArthur Genius award. He mentioned William Gasarch in his acceptance speech and gave the REU program a one-time gift of \$25,000.

James Pinkerton and Rafael Setra, who William mentored in 2011, placed second in the Siemens competition.

### **Outreach Activities**

Complexityblog is a blog about computer science theory and related topics. William and co-blogger Lance Fortnow have run the blog since 2004. They have around 20,000 readers.

### **Relevant Prior NSF Grant Support**

1. 2005 Request by STAND (Science and Technology: Addressing the Need for Diversity) to Intel to Pilot CS SPIRAL (Summer Project in Research and Learning). Faculty Director (with Clyde Kruskal, Larry Davis, Joelle Carter). \$40,000.
2. 2013-2015 NSF REU Grant on Combinatorial Algorithms–Applied Research (CAAR). PI: Samir Khuller; coPI: William Gasarch. 3 years, \$300,000. CNS 1262805.
3. 2016-2018 NSF REU (REU) Grant on Combinatorial Algorithms–Applied Research (CAAR). PI: William Gasarch; coPI: Samir Khuller . 3 years, \$375,000. CNS 1560193.
4. 2019-2021 NSF REU Grant on Combinatorial Algorithms–Applied Research (CAAR). PI: William Gasarch; coPI: Samir Khuller. 3 years, \$375,000. CNS-1952352. (Samir Khuller left UMCP in 2020 and John Dickerson became co-PI.)
5. 2022-2024 NSF REU Grant on Combinatorics, Algorithms, and AI applied to Real problems (CAAR). PI: William Gasarch; coPI: John Dickerson. 3 years, \$422,092. CNS-2150382

## **11.2 Co-PI Laxman Dhulipala**

**Project Mentor in REU-CAAR** Laxman has mentored projects in for two years, from Summer’2023–Summer’2024. One student who worked with Laxman on parallel algorithms in 2023 will attend CMU to continue to work on parallel algorithms with Laxman’s PhD advisor at CMU.

**Undergrad Student Advising** Since July 2022, when Laxman started his position as Assistant Professor at the University of Maryland, he has mentored 5 undergraduate students, two of whom have gone on to top PhD programs in computer science (at CMU and Cornell) or to top firms and startups (such as Meta and Google and others).

Beyond the REU program, his undergraduate mentees have received top-level awards such as Goldwater Scholarships (George Li) and the Computing Research Association (CRA) Outstanding Undergraduate Researcher Award (George Li).

**Competitive Programming at UMD for Undergraduates and High Schoolers** Since July 2023, Laxman has been involved in increasing participation in competitive programming at the undergraduate and high school level. He is the faculty mentor for a new competitive programming student taught course (CMSC) called *Introduction to Competitive Programming (CMSC 398L)*, which provides students interested in competitive programming with a gentle introduction to programming contests, techniques, and problem solving approaches. Laxman also served as a problem setting lead for UMD’s High School Programming Contest in 2023 and 2024. Laxman will be the lead organizer in 2025.

### **Relevant Prior NSF Grant Support**

1. *Collaborative Research: SaTC: CORE: Medium: Graph Mining and Network Science with Differential Privacy: Efficient Algorithms and Fundamental Limits* (CNS-2317194, Jul 23, 2024–Jun 27,

2027 \$400,000). *Intellectual Merit*: This project aims to design scalable private graph algorithms with rigorous accuracy bounds. *Broader Impacts*: The project will result in a private graph processing system for network science cyber-infrastructures. *Preliminary Results*: The project has supported an undergraduate student at UMD with papers in submission on private-graph algorithms, and who recently won the CRA undergraduate research award.

2. *Collaborative Research: SHF: Medium: A Scalable Graph-Based Approach to Clustering*. (SHF-2403235, Oct 24, 2025–Sep 28 2027, \$364,000). *Intellectual Merit*: This project will design a graph-based approach to cluster large datasets. *Broader Impacts*: The project will result in highly-scalable pointset clustering algorithms.

## 12 Results From Prior NSF Support (REU-CAAR)

REU-CAAR was evaluated by Computing Research Association’s Center for Evaluating the Research Pipeline (CERP). The reports for 2022 and 2023 are in the supplementary documents. We also include information from 2021 and 2024 (not much in 2024 in terms of papers since the program ended so recently).

We comment on (1) how many students were NSF funded and how many are funded by other sources, (2) high school student involvement, (3) research produced, and (4) diversity statistics (some of the numbers in the report are off since not all students took the survey), (5) how many of the students were from Non-Research Schools.

We include REU-CAAR 2021 even though it was not on the prior grant to give a fuller picture of what we have accomplished.

### 12.1 Summary of REU-CAAR 2021,2022,2023,2024

We summarize the outcome of the grants for the years 2021, 2022, 2023, 2024. Even though 2021 was not the part of the prior grant, we include it since, for that year, we have tracking information that we (quite reasonably) do not have for 2024.

#### Summary of REU-CAAR 2021

In REU-CAAR 2021 there were 26 students. Of the 26, 19 were NSF funded (that was a COVID year so the program was virtual and we used funding saved on housing for stipends for more students), 5 were funded by other sources, and 2 were high school students. The Summer of 2021 was a COVID year so the program was virtual.

1. Number of applicants: 210. Number of students in the program: 26.
2. Ethnicity: 7 White, 13 Asian, 5 African-American, 1 Hispanic
3. Gender: 16 Males, 10 Females
4. Research Schools: 12 from research schools, 12 from non-research schools, 2 from high school.
5. Four papers that came out of this work: [24], [22], [23], [60].
6. **Where are they now?** 10 are in graduate school in CS or a related area. The rest have graduated (except for the HS students) and are employed in computer science, mostly as Software Engineers.

### **Testimonials by George Li (Mentored by Furong Huang)**

I really enjoyed spending my summer at REU CAAR, and I recommend it to anyone interested in computer science research. It's difficult to get research experience as a freshman, but REU-CAAR gave me the opportunity to do so. The work was engaging and just the right difficulty. Most importantly, my mentor treated me as just another researcher, spending hours sometimes discussing our problem and possible solutions.

### **Summary of REU-CAAR 2022**

In REU-CAAR 2022 there were 22 students. Of the 22, 10 were NSF funded 9 were funded by other sources, and 3 were high school students.

1. Number of applicants: 98. (The NSF wanted us to use ETAP for admissions; however, it was not available until mid-February.) Number of students in the program: 22.
2. Ethnicity: 8 White, 11 Asian, 1 African-American, 2 Hispanic.
3. Gender: 13 Males, 9 Females.
4. Research Schools: 11 from research schools, 8 from non-research Schools, 3 from high school.
5. Two papers that came out of the program so far: [9], [30].
6. **Where are they now?** 11 are in graduate school in Computer Science or a related area. Five are still in school. Six have graduated and have jobs in computer science.

### **Testimonials by Caesar Dai (Mentored by Dave Mount and Auguste Gezalyan)**

I feel incredibly fortunate for the opportunity to attend REU-CAAR2022! Coming from a small liberal arts college, the program was a broadening experience. I worked with peers and mentors from diverse academic and cultural backgrounds, learned the research processes, and experienced the graduate school lifestyle. Most importantly, I made meaningful connections, which helped me solidify my interest in pursuing a graduate degree.

Auguste and Dave, our mentors, were great. I genuinely felt that they want us to make the most out of this REU program. During the program, they showed us that research could be a fun and self-fulfilling process. Moreover, working alongside motivated peers and professors is an exhilarating experience. I learned about various fields of research in computer science from other mentors and students and also experienced the excitement they have for their field firsthand.

I want to shout out to our program coordinator, professor Gasarch! Whether a Monday morning or a Thursday night, he energizes the room with the tremendous energy he brings.

### **Summary of REU-CAAR 2023**

In REU-CAAR 2023 there were 19 students. Of those, 10 NSF funded, 7 were funded by other sources, and 2 were high school students.

1. Number of applicants: 300. Number of students in the program: 18.
2. Ethnicity: 6 White, 9 Asian, 1 African American, 2 Hispanics.
3. Gender: 9 Males, 9 Females.

4. Research Schools: 8 from research schools, 8 from non-research schools, 2 from high school.
5. Three papers came out of the program so far: [31],[38], [39].
6. **Where are they now?** 4 are in Grad School. 12 are still in school. 2 are out of school and working in CS field.

#### **Testimonials From Andrew Brady (mentored by Laxman Dhulipala)**

REU-CAAR was a great program! My project, mentored by Professor Laxman Dhulipala, was on "parallel algorithms for high dimensional clustering." It was cool to read lots of papers about many ways researchers have improved the k-means method; I had no idea there was so much research behind k-means until this project. I also got the chance to improve my C++, which will come in handy in the future. I learned some new board games (Olympus, Navigator, Carnegie) during the board game nights which was cool. The other students were fun to be around and enhanced my experience.

#### **Summary of REU-CAAR 2024**

Since this grant is being submitted around 2 weeks after the REU-CAAR-2024 program ended, we cannot include information about papers or *where are they now*.

In REU-CAAR 2024 there were 21 students. Of those, 9 NSF funded, 9 students funded by other sources, and 3 were high school students.

1. Number of applicants: 470. Number of students in the program: 21.
2. Ethnicity: 5 White, 14 Asian, 1 African-American, 1 Pacific Islander.
3. Gender: 11 Males, 10 Females.
4. Research Schools: 9 from research schools, 9 from non-research schools, 3 from high school.

#### **Testimonials From Anubha Mahajan (Mentored by Aviva Prins)**

I had an incredible experience working with my team, and we formed a close bond that made collaboration and learning truly enjoyable. I loved getting to know everyone else in the program and we are all still in contact with each other.

Bill will always have time for you and be willing to help you in every way he can. Auguste and Bill bring in so much energy to the program. Their enthusiasm made the program truly special.

My mentor, Aviva Prins, was an invaluable resource, especially when it came to exploring and applying for graduate schools. She provided not only guidance on our projects but also offered insightful advice that will benefit me long after the program.