COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCEMENT/SOLICITATION NO./DUE DATE	☐ Special Exception to Deadline Date Policy	FOR NSF USE ONLY
PD 22-8069		NSF PROPOSAL NUMBER
FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) (Indicate the most specific unit known, i.e. program, division, etc.) DMS - CDS&E-MSS		2316011

DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	UEI (Unique Entity Identifier)	FILE LOCATION
01/20/2023	1	03040000 DMS	8069	E2NDENMDUEG8	

EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN)	SHOW PREVIOUS AWARD NO. IF THIS IS	IS THIS PROPOSAL BEING SUBMITTED TO ANOTHER FEDERAL AGENCY?
362170136	☐ A RENEWAL ☐ AN ACCOMPLISHMENT-BASED RENEWAL	YES ☐ NO ☒ IF YES, LIST ACRONYM(S)

NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE	ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE
ILLINOIS INSTITUTE OF TECHNOLOGY	10 W 35TH ST CHICAGO,IL 60616-3717 US
AWARDEE ORGANIZATION CODE (IF KNOWN) 0016915000	
NAME OF PRIMARY PLACE OF PERF Illinois Institute of Technology	ADDRESS OF PRIMARY PLACE OF PERF, INCLUDING 9 DIGIT ZIP CODE 10 W 35TH ST CHICAGO,IL 60616-3717 US

IS AWARDEE ORGANIZATION (Check All That Apply) ☐ SMALL BUSINESS ☐ FOR-PROFIT ORGANIZATION ☐ MINORITY BUSINESS ☐ WOMAN-OWNED BUSINESS

☐ IF THIS IS A PRELIMINARY PROPOSAL THEN CHECK HERE

TITLE OF PROPOSED PROJECT
Collaborative Research: Cost-Efficient and Confident Sampling for Modern Scientific Discovery

SHOW LETTER OF INTENT ID IF APPLICABLE

REQUESTED AMOUNT	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE
$ 487,349	36 months	07/01/2023	

THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW

- ☐ BEGINNING INVESTIGATOR
- ☐ DISCLOSURE OF LOBBYING ACTIVITIES
- ☐ PROPRIETARY & PRIVILEGED INFORMATION
- ☒ HISTORIC PLACES
- ☐ VERTEBRATE ANIMALS IACUC App. Date ______ PHS Animal Welfare Assurance Number ______
- ☒ TYPE OF PROPOSAL Research
- ☐ HUMAN SUBJECTS Human Subjects Assurance Number ______ Exemption Subsection ______ or IRB App. Date ______
- ☐ FUNDING OF INT'L BRANCH CAMPUS OF U.S IHE
- ☐ FUNDING OF FOREIGN ORGANIZATION OR FOREIGN INDIVIDUAL
- ☒ INTERNATIONAL ACTIVITIES: COUNTRY/COUNTRIES INVOLVED XX
- ☒ COLLABORATIVE STATUS A collaborative proposal from multiple organizations (PAPPG II.D.3.b)

PI/PD DEPARTMENT	PI/PD POSTAL ADDRESS
Applied Mathematics	10 West 32nd Street Chicago,IL 60616 US
PI/PD FAX NUMBER 312-567-3135	

NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Email Address
PI/PD NAME Fred J Hickernell	PhD	1981	312-567-8983	hickernell@iit.edu
CO-PI/PD Yuhan Ding	PhD	2015	312-567-8986	yding2@iit.edu
CO-PI/PD				
CO-PI/PD				
CO-PI/PD				

CERTIFICATION PAGE

Certification for Authorized Organizational Representative (or Equivalent)

By electronically signing and submitting this proposal, the Authorized Organizational Representative (AOR) is: (1) certifying that statements made herein are true and complete to the best of his/her knowledge; and (2) agreeing to accept the obligation to comply with NSF award terms and conditions if an award is made as a result of this application. Further, the applicant is hereby providing certifications regarding conflict of interest (when applicable), flood hazard insurance (when applicable), responsible conduct of research, and organizational support as set forth in the NSF Proposal & Award Policies & Procedures Guide (PAPPG). Willful provision of false information in this application and its supporting documents or in reports required under an ensuing award is a criminal offense (U. S. Code, Title 18, §1001).

Certification Regarding Conflict of Interest

The AOR is required to complete certifications stating that the organization has implemented and is enforcing a written policy on conflicts of interest (COI), consistent with the provisions of PAPPG Chapter IXA; and that, to the best of his/her knowledge, all financial disclosures required by the conflict of interest policy were made; and that conflicts of interest, if any, were, or prior to the organizations expenditure of any funds under the award, will be, satisfactorily managed, reduced or eliminated in accordance with the organizations conflict of interest policy. Conflicts that cannot be satisfactorily managed, reduced or eliminated and research that proceeds without the imposition of conditions or restrictions when a conflict of interest exists, must be disclosed to NSF via use of the Notifications and Requests Module in FastLane.

Certification Regarding Flood Hazard Insurance

Two sections of the National Flood Insurance Act of 1968 (42 USC §4012a and §4106) bar Federal agencies from giving financial assistance for acquisition or construction purposes in any area identified by the Federal Emergency Management Agency (FEMA) as having special flood hazards unless the:

(1) community in which that area is located participates in the national flood insurance program; and
(2) building (and any related equipment) is covered by adequate flood insurance.

By electronically signing the Certification Pages, the Authorized Organizational Representative (or equivalent) located in FEMA-designated special flood hazard areas is certifying that adequate flood insurance has been or will be obtained in the following situations:
(1) for NSF grants for the construction of a building or facility, regardless of the dollar amount of the grant; and
(2) for other NSF grants when more than $25,000 has been budgeted in the proposal for repair, alteration or improvement (construction) of a building or facility.

Certification Regarding Responsible Conduct of Research (RCR)

(This certification is not applicable to conference proposals.)
By electronically signing the Certification Pages, the Authorized Organizational Representative is certifying that, in accordance with the NSF Proposal & Award Policies & Procedures Guide, Chapter IX.B., the institution has a plan in place to provide appropriate training and oversight in the responsible and ethical conduct of research to undergraduates, graduate students and postdoctoral researchers who will be supported by NSF to conduct research.
The AOR shall require that the language of this certification be included in any award documents for all subawards at all tiers.

Certification Regarding Organizational Support

By electronically signing the Certification Pages, the Authorized Organizational Representative (or equivalent) is certifying that there is organizational support for the proposal as required by Section 526 of the America COMPETES Reauthorization Act of 2010. This support extends to the portion of the proposal developed to satisfy the Broader Impacts Review Criterion as well as the Intellectual Merit Review Criterion, and any additional review criteria specified in the solicitation. Organizational support will be made available, as described in the proposal, in order to address the broader impacts and intellectual merit activities to be undertaken.

Certification Regarding Dual Use Research of Concern

By electronically signing the certification pages, the Authorized Organizational Representative is certifying that the organization will be or is in compliance with all aspects of the United States Government Policy for Institutional Oversight of Life Sciences Dual Use Research of Concern.

Certification Regarding the Meeting Organizer’s Written Policy or Code-of-Conduct that Addresses Sexual Harassment, Other Forms of Harassment, and Sexual Assault

(This certification is only applicable to travel proposals)
By electronically signing the Cover Sheet, the AOR is certifying that prior to the proposer’s participation in the meeting, the proposer will assure that the meeting organizer has a written policy or code-of-conduct that addresses sexual harassment, other forms of harassment, and sexual assault, and that includes clear and accessible means of reporting violations of the policy or code-of-conduct. The policy or code-of-conduct must address the method for making a complaint as well as how any complaints received during the meeting will be resolved. The proposer is not required to submit the meeting organizer’s policy or code-of-conduct for review by NSF.

Certification Regarding Family Leave Status (or equivalent)

(This certification is only applicable to career-life balance supplemental funding requests)
By electronically signing the certification pages, the Authorized Organizational Representative hereby certifies that the request for a technician (or equivalent) is because the (PI/co-PI/senior personnel/ NSF Graduate Research Fellow/postdoctoral researcher/graduate student) is, or will be, on family leave status (or equivalent) from the organization in accordance with the organization’s policies. The Authorized Organizational Representative also affirms that the organization is able to fill the position for which funding is being requested, in an appropriate timeframe.

AUTHORIZED ORGANIZATIONAL REPRESENTATIVE	SIGNATURE	DATE
NAME **Robert Lapointe**	**Electronic Signature**	**Jan 20 2023 03:09 PM**

TELEPHONE NUMBER	EMAIL ADDRESS	FAX NUMBER
312-567-3035	**lapointe@iit.edu**	

COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCEMENT/SOLICITATION NO./DUE DATE	☐ Special Exception to Deadline Date Policy	FOR NSF USE ONLY
PD 22-8069		NSF PROPOSAL NUMBER
FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) (Indicate the most specific unit known, i.e. program, division, etc.) DMS - CDS&E-MSS		2316012

DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	UEI (Unique Entity Identifier)	FILE LOCATION
01/20/2023	1	03040000 DMS	8069	TP7EK8DZV6N5	

EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN)	SHOW PREVIOUS AWARD NO. IF THIS IS ☐ A RENEWAL ☐ AN ACCOMPLISHMENT-BASED RENEWAL	IS THIS PROPOSAL BEING SUBMITTED TO ANOTHER FEDERAL AGENCY? YES ☐ NO ☒ IF YES, LIST ACRONYM(S)
560532129		

NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE	ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE
DUKE UNIVERSITY	2200 W MAIN ST DURHAM,NC 27705-4640 US
AWARDEE ORGANIZATION CODE (IF KNOWN) 0029207000	
NAME OF PRIMARY PLACE OF PERF Duke University	ADDRESS OF PRIMARY PLACE OF PERF, INCLUDING 9 DIGIT ZIP CODE 2200 W MAIN ST STE 710 DURHAM,NC 27708-4677 US

IS AWARDEE ORGANIZATION (Check All That Apply) ☐ SMALL BUSINESS ☐ FOR-PROFIT ORGANIZATION ☐ MINORITY BUSINESS ☐ WOMAN-OWNED BUSINESS | ☐ IF THIS IS A PRELIMINARY PROPOSAL THEN CHECK HERE

TITLE OF PROPOSED PROJECT	SHOW LETTER OF INTENT ID IF APPLICABLE
Collaborative Research: Cost-Efficient and Confident Sampling for Modern Scientific Discovery	

REQUESTED AMOUNT	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE
$ 214,693	36 months	07/01/2023	

THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW

☐ BEGINNING INVESTIGATOR
☐ DISCLOSURE OF LOBBYING ACTIVITIES
☐ PROPRIETARY & PRIVILEGED INFORMATION
☐ HISTORIC PLACES
☐ VERTEBRATE ANIMALS IACUC App. Date ______
PHS Animal Welfare Assurance Number ______
☒ TYPE OF PROPOSAL Research

☐ HUMAN SUBJECTS Human Subjects Assurance Number ______
Exemption Subsection ______ or IRB App. Date ______
☐ FUNDING OF INT'L BRANCH CAMPUS OF U.S IHE
☐ FUNDING OF FOREIGN ORGANIZATION OR FOREIGN INDIVIDUAL
☐ INTERNATIONAL ACTIVITIES: COUNTRY/COUNTRIES INVOLVED
☒ COLLABORATIVE STATUS
A collaborative proposal from multiple organizations (PAPPG II.D.3.b)

PI/PD DEPARTMENT	PI/PD POSTAL ADDRESS
Statistical Science	415 Chapel Dr Durham,NC 27705 US
PI/PD FAX NUMBER	

NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Email Address
PI/PD NAME Simon Mak	PhD	2018	404-661-4431	sm769@duke.edu
CO-PI/PD				
CO-PI/PD				
CO-PI/PD				
CO-PI/PD				

CERTIFICATION PAGE

Certification for Authorized Organizational Representative (or Equivalent)

By electronically signing and submitting this proposal, the Authorized Organizational Representative (AOR) is: (1) certifying that statements made herein are true and complete to the best of his/her knowledge; and (2) agreeing to accept the obligation to comply with NSF award terms and conditions if an award is made as a result of this application. Further, the applicant is hereby providing certifications regarding conflict of interest (when applicable), flood hazard insurance (when applicable), responsible conduct of research, and organizational support as set forth in the NSF Proposal & Award Policies & Procedures Guide (PAPPG). Willful provision of false information in this application and its supporting documents or in reports required under an ensuing award is a criminal offense (U. S. Code, Title 18, §1001).

Certification Regarding Conflict of Interest

The AOR is required to complete certifications stating that the organization has implemented and is enforcing a written policy on conflicts of interest (COI), consistent with the provisions of PAPPG Chapter IXA; and that, to the best of his/her knowledge, all financial disclosures required by the conflict of interest policy were made; and that conflicts of interest, if any, were, or prior to the organizations expenditure of any funds under the award, will be, satisfactorily managed, reduced or eliminated in accordance with the organizations conflict of interest policy. Conflicts that cannot be satisfactorily managed, reduced or eliminated and research that proceeds without the imposition of conditions or restrictions when a conflict of interest exists, must be disclosed to NSF via use of the Notifications and Requests Module in FastLane.

Certification Regarding Flood Hazard Insurance

Two sections of the National Flood Insurance Act of 1968 (42 USC §4012a and §4106) bar Federal agencies from giving financial assistance for acquisition or construction purposes in any area identified by the Federal Emergency Management Agency (FEMA) as having special flood hazards unless the:

(1) community in which that area is located participates in the national flood insurance program; and
(2) building (and any related equipment) is covered by adequate flood insurance.

By electronically signing the Certification Pages, the Authorized Organizational Representative (or equivalent) located in FEMA-designated special flood hazard areas is certifying that adequate flood insurance has been or will be obtained in the following situations:
(1) for NSF grants for the construction of a building or facility, regardless of the dollar amount of the grant; and
(2) for other NSF grants when more than $25,000 has been budgeted in the proposal for repair, alteration or improvement (construction) of a building or facility.

Certification Regarding Responsible Conduct of Research (RCR)

(This certification is not applicable to conference proposals.)
By electronically signing the Certification Pages, the Authorized Organizational Representative is certifying that, in accordance with the NSF Proposal & Award Policies & Procedures Guide, Chapter IX.B., the institution has a plan in place to provide appropriate training and oversight in the responsible and ethical conduct of research to undergraduates, graduate students and postdoctoral researchers who will be supported by NSF to conduct research.
The AOR shall require that the language of this certification be included in any award documents for all subawards at all tiers.

Certification Regarding Organizational Support

By electronically signing the Certification Pages, the Authorized Organizational Representative (or equivalent) is certifying that there is organizational support for the proposal as required by Section 526 of the America COMPETES Reauthorization Act of 2010. This support extends to the portion of the proposal developed to satisfy the Broader Impacts Review Criterion as well as the Intellectual Merit Review Criterion, and any additional review criteria specified in the solicitation. Organizational support will be made available, as described in the proposal, in order to address the broader impacts and intellectual merit activities to be undertaken.

Certification Regarding Dual Use Research of Concern

By electronically signing the certification pages, the Authorized Organizational Representative is certifying that the organization will be or is in compliance with all aspects of the United States Government Policy for Institutional Oversight of Life Sciences Dual Use Research of Concern.

Certification Regarding the Meeting Organizer's Written Policy or Code-of-Conduct that Addresses Sexual Harassment, Other Forms of Harassment, and Sexual Assault

(This certification is only applicable to travel proposals)
By electronically signing the Cover Sheet, the AOR is certifying that prior to the proposer's participation in the meeting, the proposer will assure that the meeting organizer has a written policy or code-of-conduct that addresses sexual harassment, other forms of harassment, and sexual assault, and that includes clear and accessible means of reporting violations of the policy or code-of-conduct. The policy or code-of-conduct must address the method for making a complaint as well as how any complaints received during the meeting will be resolved. The proposer is not required to submit the meeting organizer's policy or code-of-conduct for review by NSF.

Certification Regarding Family Leave Status (or equivalent)

(This certification is only applicable to career-life balance supplemental funding requests)
By electronically signing the certification pages, the Authorized Organizational Representative hereby certifies that the request for a technician (or equivalent) is because the (PI/co-PI/senior personnel/ NSF Graduate Research Fellow/postdoctoral researcher/graduate student) is, or will be, on family leave status (or equivalent) from the organization in accordance with the organization's policies. The Authorized Organizational Representative also affirms that the organization is able to fill the position for which funding is being requested, in an appropriate timeframe.

AUTHORIZED ORGANIZATIONAL REPRESENTATIVE	SIGNATURE	DATE
NAME Lauren Faber	Electronic Signature	Jan 20 2023 02:59 PM

TELEPHONE NUMBER	EMAIL ADDRESS	FAX NUMBER
919-684-5693	lauren.faber@duke.edu	

Overview

With breakthroughs in experimental methods and computational technology, there are now novel sources of high-quality data for tackling a broad array of pressing problems in science and engineering. However, the generation of such high-fidelity data often requires costly experiments and/or simulations, which can significantly limit the amount of data available for scientific investigation. Given this cost bottleneck, it is of critical importance to develop *cost-efficient sampling methods* for data generation and model training. Furthermore, for scientific inference, such sampling methods need to be performed with *confidence*; they need to be coupled with theoretically sound and data-driven stopping rules, which guarantee the resulting statistical model achieves a desired error tolerance. This is paramount for *reliable* scientific discovery: it provides a quantification of uncertainty for scientific inference, thus protecting against spurious findings.

This project will develop a novel and timely suite of methods that jointly addresses this crucial need for cost-efficient and confident sampling for scientific discovery. Our framework features methodologies (with supporting theory and algorithms) that extend classical low discrepancy (i.e., highly stratified) sampling techniques for a broad range of challenging scenarios encountered in modern scientific problems, including cost-efficient Bayesian inference, efficient subsampling of massive data, multi-fidelity modeling, and density estimation. Major emphasis is placed on demonstrating the effectiveness of these methods for accelerating scientific discoveries, especially for the PIs' ongoing collaborations on the study of heavy-ion collisions and real-time engine control of unmanned aircraft vehicles, but also for new collaborations that will be developed over the project.

Intellectual Merit

Our project investigates four novel directions that extend low discrepancy sampling to complex settings in modern scientific and engineering problems. Each direction is necessitated by a motivating scientific problem from the PIs' multi-disciplinary collaborations and plays an integral role in our proposed suite of methods for accelerating scientific discovery. The first, called cost-efficient Stein points, extends low-discrepancy sampling for expensive Bayesian computation problems, where each posterior evaluation requires a forward run of a costly scientific simulator. The second direction explores adaptive algorithms and stopping rules for confident multifidelity sampling, which is widely used in the physical sciences. The third extends low discrepancy sampling for efficient and confident big data analysis to facilitate real-time decision-making. The last direction investigates low discrepancy sampling for distribution, density, and quantile estimation. Each direction will involve the development of novel methodology, theory, and algorithms, with a keen focus on addressing scientific needs in our aforementioned ongoing collaborations.

Broader Impacts

The proposed suite of methods paves the way for transformative scientific research, equipping practitioners with cost-efficient and confident sampling methods (with supporting theory and algorithms) for accelerating scientific and engineering discoveries. Our project will catalyze closer collaborations between the scientific and data science communities by broadening the application of low-discrepancy sampling for the complex settings featured in modern scientific and engineering problems. Such collaborations will be further strengthened via our open-source Python QMC library QMCPy, which provides an accessible and well-documented software connecting state-of-the-art low-discrepancy methods with the broader scientific community. This project will also train the next generation of science-based computational researchers, who can adeptly work in diverse and multi-disciplinary scientific teams pushing forward the frontiers of scientific knowledge.

TABLE OF CONTENTS

For font size and page formatting specifications, see PAPPG section II.B.2.

	Total No. of Pages	Page No.* (Optional)*
Cover Sheet for Proposal to the National Science Foundation		
Project Summary (not to exceed 1 page)	1	
Table of Contents	1	
Project Description (Including Results from Prior NSF Support) (not to exceed 15 pages) **(Exceed only if allowed by a specific program announcement/solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)**	15	
References Cited	9	
Biographical Sketches (Not to exceed 3 pages each)	6	
Budget (Plus up to 5 pages of budget justification. For proposals that contain subaward(s), each subaward must include a separate budget justification of no more than 5 pages)	6	
Current and Pending Support	8	
Facilities, Equipment and Other Resources	2	
Special Information/Supplementary Documents (Data Management Plan, Mentoring Plan and Other Supplementary Documents)	7	
Appendix (List below.) **(Include only if allowed by a specific program announcement/solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)**		

Appendix Items:

*Proposers may select any numbering mechanism for the proposal. The entire proposal however, must be paginated. Complete both columns only if the proposal is numbered consecutively.

TABLE OF CONTENTS

For font size and page formatting specifications, see PAPPG section II.B.2.

	Total No. of Pages	Page No.* (Optional)*
Cover Sheet for Proposal to the National Science Foundation		
Project Summary (not to exceed 1 page)		
Table of Contents	1	
Project Description (Including Results from Prior NSF Support) (not to exceed 15 pages) **(Exceed only if allowed by a specific program announcement/solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)**		
References Cited		
Biographical Sketches (Not to exceed 3 pages each)	3	
Budget (Plus up to 5 pages of budget justification. For proposals that contain subaward(s), each subaward must include a separate budget justification of no more than 5 pages)	5	
Current and Pending Support	7	
Facilities, Equipment and Other Resources	2	
Special Information/Supplementary Documents (Data Management Plan, Mentoring Plan and Other Supplementary Documents)	1	
Appendix (List below.) **(Include only if allowed by a specific program announcement/solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)**		

Appendix Items:

*Proposers may select any numbering mechanism for the proposal. The entire proposal however, must be paginated. Complete both columns only if the proposal is numbered consecutively.

Cost-Efficient and Confident Sampling for Modern Scientific Discovery

With breakthroughs in experimental methods and computational technology, there are now novel sources of high-quality data for tackling a broad array of pressing problems in science and engineering. However, the generation of such high-fidelity data often requires costly experiments and/or simulations, which can greatly limit the amount of data available for scientific investigation. Given this cost bottleneck, it is of critical importance to develop *cost-efficient sampling methods* for data generation and model training. Furthermore, for scientific inference, such sampling methods need to be performed with *confidence*; they need to be coupled with theoretically sound and data-driven stopping rules that guarantee the resulting statistical model achieves a desired error tolerance. This is paramount for *reliable* scientific discovery: it provides a quantification of uncertainty for scientific inference, thus protecting against spurious findings.

This project will develop a novel and timely suite of methods that jointly address this crucial need for cost-efficient and confident sampling for scientific discovery. Our framework features methodologies (with supporting theory and algorithms) that extend classical *low discrepancy* (LD[1]) (i.e., highly stratified) sampling techniques for a broad range of challenging scenarios encountered in modern scientific problems, including cost-efficient Bayesian inference, efficient subsampling of massive data, multi-fidelity modeling, and density estimation. Major emphasis is placed on demonstrating the effectiveness of these methods for tackling a wide array of complex scientific and engineering problems, especially for the PIs' ongoing collaborations on the study of heavy-ion collisions and real-time engine control of unmanned aircraft vehicles, but for also new collaborations that will be developed over the life of the project.

This project will be led by Fred Hickernell (FH, PI from Illinois Tech), Simon Mak (SM, PI from Duke U), Yuhan Ding (YD, co-PI from Illinois Tech), and Sou-Cheng Terrya Choi (SCTC, SAS & Illinois Tech, Senior Personnel). Our collaborators include Michael McCourt (MM, Intel), Chris Oates (CO, Newcastle U), Art Owen (AO, Stanford U), Jagadeeswaran Rathinavel (JR, Wi-Tronix), Pieterjan Robbe (PR, Sandia National Laboratories & KU Leuven), Illinois Tech PhD students Claude Hall, Jr. (CH) and Aleksei Sorokin (AS), Duke U PhD students Irene Ji (IJ), John Miller (JM), Kevin Li (KL), and Tao Tang (TT), and other students, alumni, and friends.

1. The Need for Cost-Efficient and Confident Sampling

1.1. Motivation. The nature of scientific discovery has undergone a radical paradigm shift over recent decades. With tremendous advances in experimental methodology and computational technology, scientists now have the capability to generate high-quality data for solving challenging problems which were once thought to be impossible or prohibitively expensive. In the physical sciences, complex phenomena such as universe expansions [77] and rocket propulsion [98, 146] can now be reliably studied via fine-scale virtual (i.e., computer) simulations on high performance computing systems. Similarly, in the biological sciences, fundamental developments in high throughput sequencing have led to key advances in genomics and epidemiology.

There are, however, two critical bottlenecks that greatly hinders the use of this high-quality data for scientific discovery. The first bottleneck is that, for complex problems, *such data can be very costly to generate*, requiring a large investment of computational and/or experimental resources. Take, e.g., the study of the Quark-Gluon Plasma (QGP), a deconfined phase of nuclear matter which filled the Universe shortly after the Big Bang (see recent papers by PI SM [37, 38, 67, 88] with the JETSCAPE collaboration, discussed later in Sect. 4.2). With advances in nuclear physics modeling, this plasma can be simulated by virtually colliding heavy ions together at near-light speeds. The ultimate goal is to perform an inverse problem, which uses such simulations with physical measurements from particle colliders to learn plausible properties of this plasma (call this $\boldsymbol{t}$) and hence shed light on the origins of matter. But such simulations are very costly: a single

[1]Acronyms and initials of personnel contain hyperlinks to their full names.

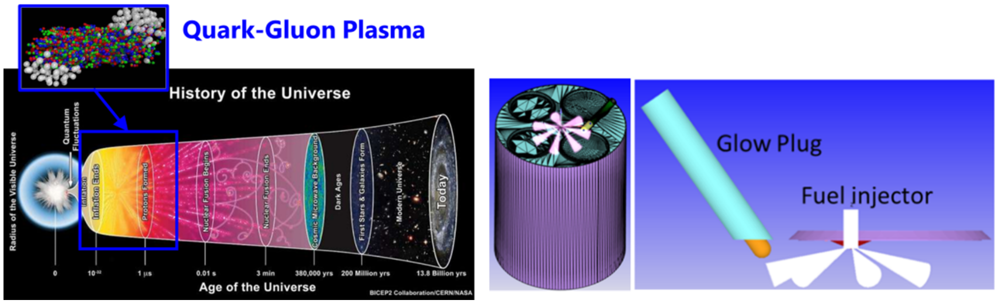

Figure 1. (Left) Visualizing the QGP shortly after the Big Bang [89], which became the building blocks of matter in the Universe. (Right) A schematic of the UAV metal engine: fuel is injected at seven nozzles, then ignited via a glow plug.

run at parameter $\boldsymbol{t}$, yielding simulation output $g(\boldsymbol{t})$, can take on the order of thousands of CPU hours [37]. This inverse problem may require thousands of simulation runs at different choices of $\boldsymbol{t}$, in order to find suitable parameters which best match cosmological measurements. Such a study can thus require *millions* of CPU hours, which places great strain on computational costs. Similar cost bottlenecks are faced in a broad variety of modern science and engineering problems.

The second bottleneck is that the increasing sophistication of scientific experiments results in *massive datasets which take on highly complex forms*. In our nuclear physics application, the simulated output $g(\boldsymbol{t})$ consists of fine-scale spatiotemporal flows for the hydrodynamic evolution of the plasma, which can take up to exabytes (10^{18} bytes) of storage. The efficient use of such massive data is thus critical for timely scientific inference and decision-making. Similar challenges arise in the PI SM's current collaboration with mechanical engineers at the University of Minnesota, on the engine control of unmanned aircraft vehicles (UAVs). Here, we wish to train a *real-time* control system for UAV engines with low cetane numbers, an important objective for U.S. Army's single fuel concept. Due to expensive prototyping costs, such systems are typically trained via complex computational fluid dynamics (CFD) simulations, which output a wide range of fine-scale engine characteristics. The key bottleneck is that simulation outputs are massive datasets, consisting of hundreds of response variables and millions of timesteps. The use of such big data for training a statistical model for *real-time* engine control is thus a critical need for practical implementation.

These challenges necessitate two crucial ingredients to facilitate scientific discovery: cost-efficiency and confidence. The first, *cost-efficiency*, refers to statistical methods which aim to maximize model performance given a limited cost (e.g., computational) budget. For the earlier nuclear physics problem, a cost-efficient method strives to provide an accurate solution to the inverse problem given a limited computational budget (and thus limited simulation runs). The second, *confidence*, refers to methods that provide a reliable and estimable measure (or quantification) of model error. Such confidence is essential for verifiable scientific discovery: it provides a quantification of uncertainty for findings, thus protecting against spurious findings. This further allows for theoretically sound stopping rules that guarantee model performance, which is particularly crucial in this cost-constrained setting. For our nuclear physics problem, a confident method yields a measure of uncertainty for the inverse problem, which can be used to guide the amount of simulation runs needed to ensure a desired error tolerance for physics discovery.

Both of the above bottlenecks can be alleviated via a careful integration of novel *sampling* algorithms within the statistical learning framework. Here, "sampling" refers to the drawing of representative samples $\boldsymbol{T}_1, \ldots, \boldsymbol{T}_n$ from a (potentially) complex probability distribution F, and the use of such samples for scientific inference and decision-making. For the first bottleneck of *costly* scientific data, a careful sampling design of input parameters for the expensive simulator can enable accurate and precise scientific inference in reasonable turnaround times. For the second bottleneck of *massive* data, a judicious sampling of large datasets can allow for timely scientific discovery and useful decision-making. There is, however, much to be done on *cost-efficient* and *confident* sampling algorithms, given the complexities present in modern scientific problems. We will thus propose a suite of methods, with supporting theory and algorithms, which address this important gap. We first present the prototypical problem of interest, then provide a survey of *low-discrepancy* sampling methods [107], which we extend for our suite of methods.

1.2. Background. We first provide a brief background on the prototypical problem we aim to address, then discuss existing literature and its limitations for our motivating applications.

1.2.1. Prototypical Problem. Consider the estimation of the *expectation of a random variable* Y whose distribution is some complicated function, g, of a random vector $\boldsymbol{T}$ with known distribution:

$$\mu := \mathbb{E}(Y) = \mathbb{E}[g(\boldsymbol{T})] = ? \tag{1a}$$

This arises in many practical problems, e.g., in our UAV application, $\boldsymbol{T}$ may represent uncertainties in engine operating conditions, $g(\boldsymbol{t})$ may denote the engine thrust at operating conditions $\boldsymbol{t}$, and we wish to learn the expected engine thrust $\mathbb{E}[g(\boldsymbol{T})]$ under uncertain conditions.

In turn, we define a transformation, $\boldsymbol{\Psi}$, of a *standard uniform* random vector, $\boldsymbol{X}$, into $\boldsymbol{T}$, so that we may also write Y as a function, f, of $\boldsymbol{X}$:

$$\boldsymbol{T} = \boldsymbol{\Psi}(\boldsymbol{X}), \qquad f(\boldsymbol{x}) = g(\boldsymbol{\Psi}(\boldsymbol{x}))\, |\partial \boldsymbol{\Psi}(\boldsymbol{x})/\partial \boldsymbol{x}|\,, \qquad \boldsymbol{X} \sim \mathcal{U}[0,1]^d, \tag{1b}$$

$$\mu = \mathbb{E}[f(\boldsymbol{X})] = \int_{[0,1]^d} f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = ?. \tag{1c}$$

This expectation can also be thought of as a *multivariate integration* problem. Here, we write $Y = f(\boldsymbol{X})$ because the LD sequences described later mimic uniform random vectors to approximate μ.

Besides the integral problem in (1), it is often useful to know the *distribution*, F, *density*, ϱ, and/or *quantile* function, Q, of Y, i.e.,

$$F(y) := \mathbb{P}(Y \le y) = \mathbb{P}[f(\boldsymbol{X}) \le y] = \int_{[0,1]^d} \mathbb{1}_{(-\infty,y]}(f(\boldsymbol{x}))\, \mathrm{d}\boldsymbol{x}, \quad \varrho(y) := F'(y), \quad Q(p) := F^{-1}(p). \tag{2a}$$

In the UAV problem, we may wish to learn the distribution of engine thrust Y under uncertain operating conditions; indeed, these are the primary quantities of interest when designing for robust engines. We will address these problems in Sect. 2.4 in the context of our motivating applications.

In practice, the population mean or multivariate integral, μ, often cannot be evaluated by analytic means, but it may be estimated by the sample mean, $\hat{\mu}_n$, i.e.,

$$\mu \approx \hat{\mu}_n := \frac{1}{n} \sum_{i=1}^{n} Y_i = \frac{1}{n} \sum_{i=1}^{n} f(\boldsymbol{X}_i), \qquad Y_i = f(\boldsymbol{X}_i). \tag{3}$$

Given an error tolerance, ε, we want to choose samples, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, to mimic $\mathcal{U}[0,1]^d$ and satisfy

$$|\mu - \hat{\mu}_n| \le \varepsilon \qquad \text{with high confidence.} \tag{4}$$

1.2.2. Low Discrepancy Sampling. Choosing $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$, for evaluating the sample mean in (3) to be *independent and identically distributed* (IID) yields a root mean square error of $\sqrt{\mathbb{E}[|\mu - \hat{\mu}_n|^2]} = \operatorname{std}(f(\boldsymbol{X}))n^{-1/2}$, which is independent of the dimension, d, but slowly vanishing as $n \to \infty$. The computational cost to satisfy the error tolerance (4) is $\mathcal{O}(\varepsilon^{-2})$. Tensor product generalizations of one-dimensional numerical integration rules using grid sampling give errors of $|\mu - \hat{\mu}_n| = \mathcal{O}(n^{-r/d})$ and a computational cost of $\mathcal{O}(\varepsilon^{-d/r})$, where r is limited by both the sophistication of the rule and the smoothness of the integrand, f. Such rules may be suited for small dimensions d, but they are inefficient for the larger d that occurs in our applications of interest.

A superior approach that combines the intentional structure of a grid with the essentially dimensionless error bound of IID sampling is LD sampling, $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \overset{\text{LD}}{\sim} \mathcal{U}[0,1]^d$, which has an error bound of [54, 107]

$$|\mu - \hat{\mu}_n| \le D(\{\boldsymbol{X}_i\}_{i=1}^n)\, \|f - \mu\|_{\mathcal{F}}\,. \tag{5}$$

The discrepancy, $D(\{\boldsymbol{X}_i\}_{i=1}^n)$, corresponds to the norm of the cubature error functional [53] for the function space $\mathcal{F}$. The discrepancy is also a measure of how close the empirical distribution (which assigns equal probability to each point) is to the uniform distribution. The discrepancy is typically $\mathcal{O}(n^{-1+\delta})$, where δ is arbitrarily small and positive. This faster convergence rate translates into a computational cost of $\mathcal{O}(\varepsilon^{-1-\delta})$ to satisfy error criterion (4), under mild smoothness conditions on f. The semi-norm $\|\cdot - \mu\|_{\mathcal{F}}$ is called the *variation* and is a measure of function roughness.

Popular LD sampling schemes include lattices [32, 107, 125] and digital nets [30, 107]. Fig. 2 displays $n = 64$ LD lattice points intended to mimic $\mathcal{U}[0,1]^2$. A two-dimensional grid of 64 points would only have 8 different values in each coordinate direction, whereas the LD sample covers 64 different values in each coordinate direction. Such schemes are available in many softwares, including BRODA [80], CUBA [47] MATLAB [135], NAG [136], PyTorch [115], SciPy [139], TensorFlow [134], and uncertainty quantification libraries such as Dakota [1], MUQ [114], UQLab [101], and UQTk [28, 29]. Efforts to identify better LD sequence generators include LatNet Builder [26] and the Magic Point Shop [109].

PI FH, co-PI YD, SCTC, MM, JR, AS and collaborators have developed more comprehensive QMC libraries, GAIL [21] for MATLAB and QMCPy [22] for Python. These libraries include the stopping criteria mentioned below, as well as flexible variable transformations of the form (1b). Our recent effort is focused on QMCPy, and includes an active repository [22], documentation [20], a tutorial [52], and a blog [19].

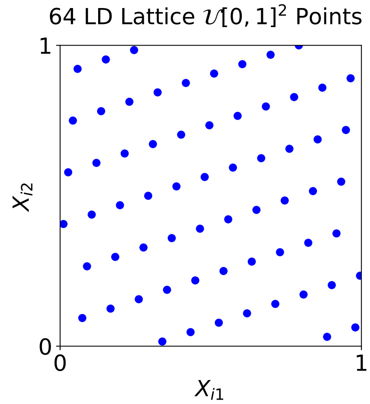

Figure 2. LD lattice points, which have fewer gaps and clusters of points than either the IID or grid points.

1.2.3. Error Bounds and Stopping Criteria. Algorithms based on efficient LD sampling are commonly called *quasi-Monte Carlo* (QMC) algorithms. To construct a QMC algorithm, one needs not only an LD sequence but also *reliable* and *practical* bounds on the error $|\mu - \hat{\mu}_n|$ which are based on the function data, $f(\boldsymbol{X}_1), f(\boldsymbol{X}_2), \ldots$. Such *data-based* error bounds inform the stopping criteria, which guide the determination of sample size n needed to satisfy the error tolerance (4). One approach is to use N different randomizations of a single LD sequence to compute N sample means, and then estimate the error of the grand sample mean, $\hat{\mu}_{nN} = (\hat{\mu}_n^{(1)} + \cdots \hat{\mu}_n^{(N)})/N$, via, e.g., bootstrapping. But this would be too costly for our motivating applications since it requires nN evaluations of the expensive function f (see first bottleneck in Sect. 1.1).

PI FH and his collaborators have developed two kinds of theoretically justified stopping criteria for LD sampling based on the Fourier coefficients of the data $f(\boldsymbol{X}_1), f(\boldsymbol{X}_2), \ldots$. The first kind determines a bound on $|\mu - \mu_n|$ by inferring the roughness of f from the decay of the discrete Fourier coefficients [57, 59, 70]. The second kind uses Bayesian credible intervals for $|\mu - \mu_n|$ assuming that f is an instance of a Gaussian process whose hyperparameters are tuned by the function data [56, 65, 66]. The cost of bounding the error $|\mu - \mu_n|$ is only $\mathcal{O}(n \log(n))$ for both kinds of stopping criteria. This can be achieved for the Bayesian approach by choosing covariance kernels that match the LD sampling schemes, and thus avoiding the typical $\mathcal{O}(n^3)$ cost.

1.3. Limitations of Existing Methodology and Software. There are, however, fundamental limitations which prevent the effective use of existing methods for tackling the complexities in our applications. These limitations, detailed below, are gaps which we will address in our proposal.

Limitation 1: *Expensive Bayesian Sampling.* Bayesian inference [42] is a popular statistical framework for tackling scientific problems. Here, parameters of interest are sampled from a "posterior" probability distribution ϱ, which captures both the the prior belief of the modeler and evidence from the collected data. In practical problems, the posterior distribution is often highly complex and available only in proportional form. This is further complicated by the *costly* nature of each posterior evaluation (see first bottleneck in Sect. 1.1). For example, *each* evaluation of the unnormalized posterior density in our heavy-ion physics inverse problem requires thousands of CPU hours [37]. Existing work on LD sampling, however, focuses largely on the sampling from uniform distributions (a literature review is provided later). We develop in Sect. 2.1 a novel cost-efficient LD posterior sampling method which addresses this need, with applications to our heavy-ion physics application and broader scientific problems.

Limitation 2: *Multifidelity Modeling.* For many scientific computing problems, the random variable of interest $Y_{\boldsymbol{\eta}} = f_{\boldsymbol{\eta}}(\boldsymbol{X}_{\boldsymbol{\eta}})$ is parameterized by a *fidelity* parameter $\boldsymbol{\eta}$, and the desired quantity of interest is the limiting mean $\mu_\infty = \lim_{\boldsymbol{\eta}\to\boldsymbol{\eta}_\infty} \mathbb{E}(Y_{\boldsymbol{\eta}})$. In our heavy-ion application, $f_{\boldsymbol{\eta}}(\boldsymbol{X}_{\boldsymbol{\eta}})$ may represent a observable simulated from a complex partial differential equation (PDE) system modeling the heavy-ion collision, with random coefficients $\boldsymbol{X}_{\boldsymbol{\eta}}$ representing uncertainties in simulation inputs, and mesh-size parameters $\boldsymbol{\eta}$ controlling simulator fidelity. As fidelity increases, the cost of sampling $Y_{\boldsymbol{\eta}}$ also increases (as each evaluation $f_{\boldsymbol{\eta}}$ becomes more expensive), and thus the evaluation of many high fidelity $Y_{\boldsymbol{\eta},i}$ to approximate μ_∞ can be prohibitively costly.

Multifidelity (also known as multi-level or multi-index) methods [45, 48, 49] choose a sequence of fidelity parameters, $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots$, and write the desired quantity as a telescoping sum:

$$\mu = \mu_\infty = (\mu_{\boldsymbol{\eta}_1} - \mu_{\boldsymbol{\eta}_0}) + (\mu_{\boldsymbol{\eta}_2} - \mu_{\boldsymbol{\eta}_1}) + \cdots + (\mu_{\boldsymbol{\eta}_L} - \mu_{\boldsymbol{\eta}_{L-1}}) + (\mu_\infty - \mu_{\boldsymbol{\eta}_L}), \qquad \mu_{\boldsymbol{\eta}_0} = 0.$$

The sequence of fidelity parameters is chosen so that the cost of evaluating $f_{\boldsymbol{\eta}_l}(\boldsymbol{X}_{\boldsymbol{\eta}_l})$ increases with l (greater fidelity), but the sample size required to estimate $\mu_{\boldsymbol{\eta}_l} - \mu_{\boldsymbol{\eta}_{l-1}}$ accurately decreases with l. The term $\mu_\infty - \mu_{\boldsymbol{\eta}_L}$ is approximated by zero. Substantial cost-efficiency is gained by using many cheap samples to approximate the low fidelity terms and relatively fewer expensive samples to compute the high fidelity terms. While there is a rich literature on multifidelity methods (including recent work by PIs SM and FH, which we discuss later), there has been little work on developing measures of confidence or stopping criteria (see Sect. 1.2.3) for such multifidelity methods. Such stopping criteria are important for our heavy-ion application, allowing for confident scientific inference with minimal experimental costs. We will addressed such limitations in Sect. 2.2.2.

Limitation 3: *Big Data Analysis.* Given sophisticated computing technology, scientific simulators typically output massive datasets with complex forms, and the efficient use of such data for scientific discovery is paramount. One solution is to take a small representative *subsample* of the big data, and use this for efficient model training. The careful selection of this subsample is critical for timely decisions. Machine learning algorithms often make use of stochastic gradient descent [6, 131], which takes a new *random* subsample of the big data at each gradient update; such random subsampling, however, may be practically and theoretically inefficient given a computational constraint (more on this later). There has been recent work on extending LD ideas for big data subsampling (discussed later), but such methods typically do not perform well or have sound theoretical guarantees for complex learning models, which are desired with massive training data. To address this, we propose in Sect. 2.3 a new LD subsampling method, which provides provably improved LD big data subsampling for a broad class of learning models.

Limitation 4: *Distribution, Density, and Quantile Estimation.* In many problems, practitioners wish to estimate not only the mean of $Y = g(\boldsymbol{T})$, but also its distribution, density or quantile (as defined in (2)). This is the case in our UAV problem: aerospace engineers wish to estimate not only the expected engine thrust $E(Y)$ under uncertain operating conditions $\boldsymbol{T}$, but also characterize its full distribution for designing robust engines. The error analysis for such problems, especially for LD sequences, is underdeveloped in the literature, and rigorous, data-based stopping criteria are non-existent. We address this limitation in Sect. 2.4.

Limitation 5: *Software Quality.* QMC software implemented by non-experts may be flawed. FH and MM found that randomized PyTorch Sobol' points fell on the boundaries of $[0, 1]^d$, when they never should [116] due to a lack of double precision. Lluís Antoni Jiménez Rugama (LlAJR), a

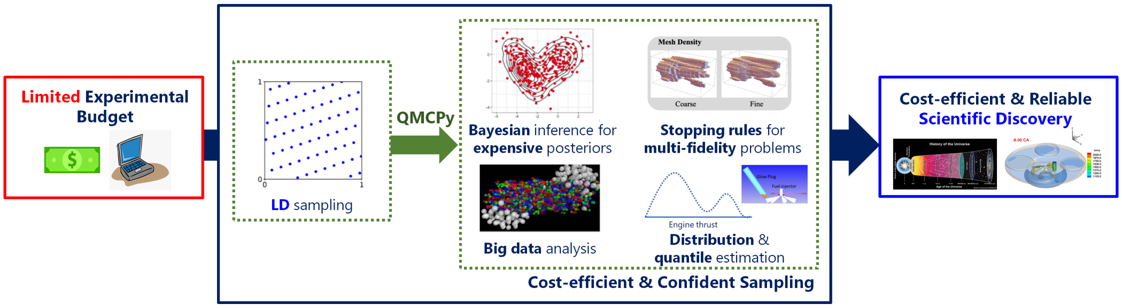

Figure 3. Project workflow: given a limited budget, the proposed tasks (Sect. 2) extend LD sampling to the complex settings motivated by our applications, providing a toolbox for cost-efficient and reliable scientific discovery. QMCPy (Sect. 3.2) serves as an open-source software package that disseminates our suite of methods to the scientific community.

former PhD student of FH, alerted that MATLAB's Sobol' sequence scrambling was incorrect; MATLAB later corrected this error. After a vigorous discussion on the PyTorch [116] and SciPy [124] issues sites, AO, FH, and other QMC researchers convinced the developers not to omit the first Sobol' point and to randomize by default. AO explained why this is crucial [112]. The need for a vigilant QMC software community is addressed in Sect. 3.2. Further, solving complex problems well and efficiently often requires multiple software libraries. For example, in uncertainty quantification, one integrand, $f(\boldsymbol{X}_i)$, may be the output from a PDE library. Not all libraries connect well, nor are there yet standards in the QMC software community on how to pass information from one library to the next. We address this issue in Sect. 3.2.

2. A Framework for Cost-Efficient and Confident Sampling

We now present a suite of novel methods (with supporting theory and algorithms) that extend LD sampling to the complex settings from our motivating applications. These methods provide a useful toolbox for cost-efficient and confident sampling to accelerate scientific discovery. Figure 3 shows the workflow for the four proposed tasks, which directly address the limitations in Sect. 1.3.

2.1. Bayesian Sampling for Expensive Posteriors [SM lead, FH, IJ, TT, JM]

2.1.1. Background and Preliminary Results. Bayesian methods [42] rely on MCMC sampling to explore the posterior distribution F, which captures information on model parameters (see PI SM's work [63, 97, 100]). However, F can often be *expensive* to evaluate in many scientific computing problems (see Limitation 1). This is compounded by the highly correlated nature of traditional MCMC samplers, which reduces the information provided by each sample [86]. Existing samplers for such problems thus be prohibitively *costly* [75], and an LD posterior sampling method can provide improved Bayesian learning given a computational budget.

One approach for LD posterior sampling is to minimize the *kernel discrepancy* $D_K(\{\boldsymbol{T}_i\}_{i=1}^n, F)$ [54], which measures the difference between the empirical distribution of $\{\boldsymbol{T}_i\}_{i=1}^n$ and the posterior F via a symmetric positive-definite kernel K. This approach, known as *kernel herding* [16], has a key limitation: the discrepancy requires an analytic form for the integral $\int K(\boldsymbol{t}, \cdot)\,\mathrm{d}F(\boldsymbol{t})$, which is unattainable for complex posteriors F. To address this, [15] proposed the "Steinized" kernel:

$$
\begin{aligned}
(6)\quad K_{\mathrm{ST}}(\boldsymbol{t},\boldsymbol{x}) = {} & \nabla_{\boldsymbol{t}} \cdot \nabla_{\boldsymbol{x}} K(\boldsymbol{t},\boldsymbol{x}) + \nabla_{\boldsymbol{t}} K(\boldsymbol{t},\boldsymbol{x}) \cdot \nabla_{\boldsymbol{x}} \log \mathrm{d}F(\boldsymbol{x}) \\
& + \nabla_{\boldsymbol{x}} K(\boldsymbol{t},\boldsymbol{x}) \cdot \nabla_{\boldsymbol{t}} \log \mathrm{d}F(\boldsymbol{t}) + K(\boldsymbol{t},\boldsymbol{x}) \nabla_{\boldsymbol{t}} \log \mathrm{d}F(\boldsymbol{t}) \cdot \nabla_{\boldsymbol{x}} \log \mathrm{d}F(\boldsymbol{x}),
\end{aligned}
$$

where ∇ and $\nabla\cdot$ are the gradient and divergence operators, respectively. With this, $\int K_{\mathrm{ST}}(\boldsymbol{t},\cdot)\,\mathrm{d}F(\boldsymbol{t})$ evaluates to 0, thus yielding a closed form expression for the *kernel Stein discrepancy* (KSD):

$$
(7)\qquad D_K(\{\boldsymbol{T}_i\}_{i=1}^n, F) := \sqrt{\frac{1}{n^2} \sum_{i,j=1}^{n} K_{\mathrm{ST}}(\boldsymbol{T}_i, \boldsymbol{T}_j)}.
$$

[15] then employs a sequential optimization of the KSD: the first sample $\boldsymbol{T}_1^*$ is taken at the mode of F, then subsequent samples $\boldsymbol{T}_2^*, \boldsymbol{T}_3^*, \cdots$ are obtained by sequentially optimizing the KSD, i.e.:

$$(\mathrm{ESP}_n) \qquad \boldsymbol{T}_n^* \leftarrow \arg\min_{\boldsymbol{t}} \mathrm{KSD}_n(\boldsymbol{t}) := \arg\min_{\boldsymbol{t}} D_K(\{\boldsymbol{T}_i^*\}_{i=1}^{n-1} \cup \boldsymbol{t}, F), \quad K = K_{\mathrm{ST}}, \quad n = 2, 3, \cdots .$$

These optimized samples, called *Stein points*, can yield improved representation of the posterior F over standard MCMC methods for a fixed sample size n.

However, Stein points have a key drawback: they are *not* cost-efficient. When the posterior F is expensive, the optimization of the Stein discrepancy for a *single* sample point will require *many* evaluations of F in the form of its score function $\nabla_{\boldsymbol{t}} \log \mathrm{d}F(\boldsymbol{t})$. For our heavy-ion application, given a fixed budget of 10^6 CPU hours, we can afford 10^3 evaluations of F, which translates to only $n \approx 50$ Stein points - this is inadequate for exploration of complex posterior distributions.

We thus propose the following cost-Efficient Stein Points (ESPs). The key idea is the construction of a sequence of carefully constructed Gaussian process (GP) surrogate models [123] on the expensive objective functions KSD_n. Consider first (ESP_n) for a given n. Suppose the posterior in the form of its score function $\nabla_{\boldsymbol{t}} \log \mathrm{d}F(\boldsymbol{t})$ has already been at evaluated at M_n candidate points $\{\boldsymbol{T}_j\}_{j=1}^{M_n}$, yielding objective evaluations $\mathcal{D}_n = \{\mathrm{KSD}_n(\boldsymbol{T}_j)\}_{j=1}^{M_n}$ via (7). Under a GP prior on $\mathrm{KSD}_n(\cdot)$, the posterior distribution of $\mathrm{KSD}_n(\boldsymbol{t})$ can be shown to be $\mathrm{KSD}_n(\boldsymbol{t})|\mathcal{D}_n \sim \mathcal{N}\{\mu_n(\boldsymbol{t}), \sigma_n^2(\boldsymbol{t})\}$; specific forms for $\mu_n(\boldsymbol{t})$ and $\sigma_n^2(\boldsymbol{t})$ can be found in [123]. Using this and following the literature on Bayesian optimization (e.g., [73] and work from PI SM [17]), the *expected improvement* in objective KSD_n from evaluating the posterior at a new point $\boldsymbol{t}$ takes the closed-form expression:

$$(8) \quad \mathrm{EIKSD}_n(\boldsymbol{t}) = (\mathrm{KSD}_{n,\min} - \mu_n(\boldsymbol{t}))\Phi\left(\frac{\mathrm{KSD}_{n,\min} - \mu_n(\boldsymbol{t})}{\sigma_n(\boldsymbol{t})}\right) + \sigma_n(\boldsymbol{t})\phi\left(\frac{\mathrm{KSD}_{n,\min} - \mu_n(\boldsymbol{t})}{\sigma_n(\boldsymbol{t})}\right),$$

where $\mathrm{KSD}_{n,\min}$ is the best observed KSD_n value. We thus wish to evaluate the expensive posterior at the point that maximizes $\mathrm{EIKSD}_n(\boldsymbol{t})$. This procedure, of refitting the GP surrogate and evaluating the posterior at the point of greatest expected improvement, is then iterated until a convergence criterion is met. The evaluated point with smallest KSD_n is then taken as the next sample $\boldsymbol{T}_n^*$ in (ESP_n).

Consider next the *sequence* of optimization problems in (ESP_n) for ESP sampling. A key observation is that the posterior evaluations used for solving previous problems $\mathrm{ESP}_2, \cdots, \mathrm{ESP}_{n-1}$ can be directly *reused* for the current problem ESP_n, since such evaluations translate directly to evaluations of KSD_n via (7). Thus, as sample size n increase, this allows for increasingly more data on the objective KSD_n to generate the n-th ESP $\boldsymbol{T}_n^*$. This recycling of posterior evaluations enables cost-efficient ESP sampling given a limited computational budget.

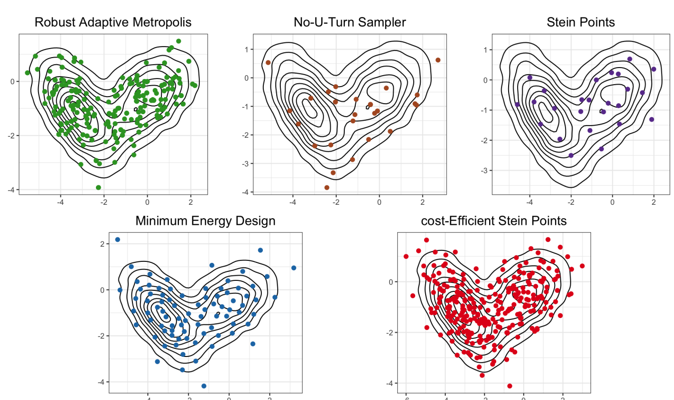

Figure 4. Visualizing the sampled points on a 2D two-mixture normal distribution, using four existing posterior samplers and the proposed ESPs. All samplers are limited to 500 posterior evaluations.

To demonstrate the cost-efficiency of ESPs, Fig. 4 compares several state-of-the-art samplers (the robust adaptive Metropolis sampler [138], the No-U-Turn sampler [62], Stein points [15], minimum energy designs [75]) on a 2D two-mixture normal distribution. All samplers are limited to $B = 500$ posterior evaluations. We see that, as expected, existing samplers that ignore the expensive nature of posterior evaluations provide a poor approximation of F: they either yield a small sample size ($n \approx 50$), or a highly correlated sample chain. ESPs, on the other hand, provide a noticeably larger sample size $n = 287$ with low sample correlation. This improved posterior representation is confirmed via a comparison of marginal statistics or distributional metrics.

2.1.2. Theory. [Years 1–3] Given these promising results, we will investigate key theoretical properties for ESP sampling. A key property to establish is, given an error tolerance ε for posterior approximation, what is the computational cost (i.e., the number of posterior evaluations B) required to achieve such a tolerance with ESPs. This provides a theoretical basis for comparison with existing samplers which ignore evaluation costs. Such rates will require a careful integration of cost complexity results for Monte Carlo [45] with Bayesian non-parametrics theory [61]. We will also explore the cost-efficiency of ESPs using a variety of probabilistic surrogate models, particularly models that can learn embedded low-dimensional structure for high-dimensional approximation, and models that can integrate prior scientific information (see papers by PI SM [13, 148, 151] and [67, 68, 88]).

2.1.3. Randomization and Central Limit Theorems. [Years 1–2] In QMC, the *randomization* of LD sequences is an important emerging topic, providing a basis for probabilistic inference (e.g., confidence intervals) on integral estimands [31]. Such randomization is particularly important in this expensive Bayesian setting, where a quantification of estimation uncertainty is desired given limited posterior evaluations. We will develop a randomized implementation of ESPs, where in addition to providing an efficient LD sampling of the posterior, each marginal sample $\boldsymbol{T}_n^*$ is *random* and follows the desired posterior distribution. We will further prove a Central Limit Theorem, which characterizes the asymptotic distribution of integral estimators as $n \to \infty$. With this, we will develop confidence intervals for the proposed randomized ESPs (following [120]), and use this to demonstrate the improved cost-efficiency of ESPs over existing MCMC samplers.

2.1.4. Implementation and Application. We will demonstrate the usefulness of ESPs in a wide range of modern scientific problems involving expensive Bayesian inference. This includes our motivating heavy-ion application (see Sect. 1.1), where the Bayesian inference of plasma properties from particle colliders requires costly forward runs (thousands of CPU hours) for each evaluation of the posterior. PI SM has worked extensively in this area (see [36–40, 82, 88]) as a member of the JETSCAPE collaboration (discussed later in Sect. 4.2). We will further explore broader applications of ESPs in Bayesian sensor imaging, rocket engine design and astrophysics, for which the PIs have close ongoing collaborations. The proposed algorithms will be implemented on our open-source package QMCPy (discussed later in Sect. 3.2).

2.2. Adaptive Multifidelity Algorithms [FH lead, SCTC, YD, MM, PR, CH, AO, AS]

2.2.1. Motivation. A common uncertainty quantification problem in the physical sciences involves the estimation of $\mu = \mathbb{E}(Y)$, where $Y = f(\boldsymbol{X})$ and f can only be approximated by $f_{\boldsymbol{\eta}}$, with $\boldsymbol{\eta}$ denoting the fidelity of the approximation. In geophysics, f may be the solution of a (partial) differential equation modeling fluid flow, whose boundary conditions are given by a random spatial field, and $\boldsymbol{\eta}$ may denote the mesh size of the numerical solver and the discretization of the random field. In our heavy-ion application, f may be the exact solution of a particle collision observable from a complex physics model with uncertain plasma properties $\boldsymbol{X}$ as inputs, and $\boldsymbol{\eta}$ may capture the spatial and temporal mesh size of the simulator (see work on this by PI SM [67, 68, 88]).

As mentioned in Sect. 1.3, approximating the true solution by a single high fidelity expectation can be prohibitively costly. Rather, multifidelity methods [45, 48, 49]—an active research area—consider a sequence of problems, $\{\mu_l = \mathbb{E}[f_{\boldsymbol{\eta}_l}(\boldsymbol{X}_{\boldsymbol{\eta}_l})]\}_{l=0}^{L}$, where the fidelity increases with l as does the computational cost of evaluating an instance $f_{\boldsymbol{\eta}_l}(\boldsymbol{X}_{\boldsymbol{\eta}_l,i})$. One can efficiently approximate μ by using more cheap samples to approximate $\mu_l - \mu_{l-1}$ for small l and fewer expensive samples to approximate $\mu_l - \mu_{l-1}$ for large l. There has been recent work (see [44, 45], including recent work by PI SM [132]) which show that, under mild conditions on the cost function of the simulator $f_{\boldsymbol{\eta}}$ and its convergence rates, multifidelity methods can provide noticeably more accurate and confident estimates over single-fidelity approaches, given a cost budget B.

Despite this, there has been little work on stopping criteria for multifidelity sampling methods (see Limitation 2). This is crucial for scientific discovery; in our heavy-ion application, such criteria provide physicists with a confident quantification of uncertainty given a computational budget, or equivalently, a confident estimate of budget required to achieve a desired precision on findings. We propose below two novel directions which address this in a cost-efficient manner.

2.2.2. Extending Stopping Rules to Multifidelity Problems. [Years 1-2] Adaptive algorithms for these multifidelity problems use ad hoc stopping criteria. We propose to develop rigorous stopping criteria like those described in Sect. 1.2.3. PI FH, AS, JR, and collaborators have developed stopping criteria for functions of several expectations, $C(\boldsymbol{\mu})$, [59, 130]. For example, Bayesian posterior means, [43], can be written as the ratio of two expectations, $C(\boldsymbol{\mu}) = \mu_2/\mu_1$. Sensitivity indices [121, 122, 128] also involve the computation of more than one expectation. However, in both these cases, the underlying random vector, $\boldsymbol{X}$, is the same for all expectations, $\mu_1, \mu_2, \ldots$, and only the functions f_l defining the expectations are different.

For multifidelity problems each μ_l depends on a different $\boldsymbol{X}_{\boldsymbol{\eta}_l} \sim \mathcal{U}[0,1]^{d_l}$ with a different d_l. Thus, adaptive algorithms need to manage LD sequences of different dimensions as well as different sample sizes. Adaptive decisions must be made on whether to devote more effort to sampling the low or high fidelity terms. The wealth of literature on multifidelity methods and our experience on developing rigorous stopping criteria for single fidelity problems gives us confidence in success.

2.2.3. Implementation and Application. We will demonstrate the effectiveness of the above developments for confident sampling in a broad spectrum of scientific problems involving multifidelity modeling. This includes our motivating heavy-ion application (Sect. 1.1), where the simulators for particle collisions have multiple fidelity parameters involving spatial meshes and time-steps at different stages (see PI SM's papers [67, 68]). We will further explore broader applications in geophysical applications, which the PIs have close ongoing collaborations (see [147]). The proposed algorithms will be implemented and fully documented in QMCPy (see Sect. 3.2)

2.3. Big Data Subsampling [SM lead, AO, IJ, KL]

2.3.1. Motivation and Preliminary Results. Big data is ubiquitous with advances in technology and computing. In our UAV application, the output of the numerical simulator can require terabytes of storage (see Sect. 1.2). A key challenge is that learning algorithms need to be *scalable* to extract useful information from such data for *real-time* decisions, e.g., real-time engine control for UAV flight. One strategy is to iteratively train the model on small batches of the data, typically sampled uniformly at random. This *subsampling* scales up state-of-the-art machine learning algorithms, such as stochastic gradient descent (SGD, [6]) and stochastic gradient boosting [41].

Consider SGD, which minimizes the loss $L(\theta; \mathcal{T}) = N^{-1} \sum_{m=1}^{N} l(\theta; \boldsymbol{T}_m)$ over model parameters $\theta \in \mathbb{R}^q$, where $\mathcal{T} = \{\boldsymbol{T}_m\}_{m=1}^N \subset \mathbb{R}^d$ is the large training data. Standard gradient descent [108] is impractical here, since they require evaluation of the full gradient $N^{-1} \sum_{m=1}^{N} \nabla_\theta l(\theta; \boldsymbol{T}_m)$, which is very expensive with N large. Mini-batch SGD [6] approximates this via a subsample $\mathcal{T}_s^{[l]} \subset \mathcal{T}$ of size $n \ll N$, taken IID and uniformly from $\mathcal{T}$. The descent steps are iterated until convergence:

$$\theta^{[l+1]} \leftarrow \theta^{[l]} - \zeta \left(\frac{1}{n} \sum_{\boldsymbol{T} \in \mathcal{T}_s^{[l]}} l(\theta; \boldsymbol{T}) \right), \quad l = 1, 2, \ldots, \tag{9}$$

where ζ is the gradient descent step size. Mini-batch SGD is widely used for scalable training of neural networks and deep learning models with big data [131].

Mini-batch SGD, however, has a key limitation. Since gradients are estimated by *random* subsampling, the solution sequence $(\theta^{[l]})_{l=1}^\infty$ converges to a *noise ball* of radius $\mathcal{O}(n^{-1})$ around the global optimum θ^*. For small subsample sizes n (as necessitated from our cost-constrained setting), SGD can thus return estimates *very far* from θ^*. Our solution is to choose an LD dataset that well-represents the big data $\mathcal{X}$. This is known as "data squashing" (termed by AO in [111]), and encompasses work on leverage-score subsampling [90], coresets [3, 9, 64], experimental design [142, 143], and work by PI SM [79, 92–94]. LD data squashing for SGD is a timely problem, but one largely unaddressed in the literature for complex non-linear models (see Limitation 3).

We propose a new data squashing method which makes use of LD subsampling of big data $\mathcal{T}$ for accelerating SGD. The preliminary result below guarantees the *existence* of such a subsample:

Theorem 1. *Let $\mathcal{T} = \{\boldsymbol{T}_m\}_{m=1}^N$ (the "big data") be any set of points on $[0,1]^d$, and suppose the feasible space Θ is convex. Further suppose $n \leq \sqrt{N}$ and the loss function l is convex with mild regularity conditions. Then there exists a subsample $\mathcal{T}_s \subseteq \mathcal{T}$ of size n which, when used within the SGD iterative updates (9), yields a solution sequence $(\theta^{[l]})_{l=1}^\infty$ converging to a noise ball of radius $\mathcal{O}\{(\log n)^{3d+1}/n^2\}$ around the global optimum θ^*.*

This theorem guarantees that, under mild assumptions on the loss function l (satisfied by a broad range of learning models), there exists an LD subsample of the big data which, when used within SGD, converges a noise ball of radius $\mathcal{O}\{(\log n)^{3d+1}/n^2\}$ around the desired solution θ^*. Thus, with a carefully chosen LD subsample, the proposed "LD-batch SGD" can yield *improved* optimization over standard mini-batch SGD, which converges to a *larger* noise ball of radius $\mathcal{O}(n^{-1})$. Put another way, this suggests that LD-batch SGD can yield comparable performance to mini-batch SGD with far fewer optimization iterations, thus allowing for *large computational savings for big data analysis*. We will tackle the following tasks to flesh out a comprehensive methodological framework for LD-batch SGD.

2.3.2. Optimization of LD Subsample. [Years 1–2] While the existence of an LD subsample is promising, one challenge is in finding such a subset efficiently. We will find this via the following optimization approach. Define the so-called "data kernel" using the big data $\mathcal{T} = \{\boldsymbol{T}_m\}_{m=1}^N$:

$$K_{\text{data}}(\boldsymbol{x},\boldsymbol{y}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^d\setminus\boldsymbol{0}} \lambda_{\boldsymbol{k}}\phi_{\boldsymbol{k}}(\boldsymbol{x})\overline{\phi_{\boldsymbol{k}}(\boldsymbol{y})}, \quad \phi_{\boldsymbol{k}}(\boldsymbol{x}) = e^{2\pi\mathrm{i}\boldsymbol{k}^T\boldsymbol{x}} - b_{\boldsymbol{k}}, \quad b_{\boldsymbol{k}} = \frac{1}{N}\sum_{m=1}^N e^{2\pi\mathrm{i}\boldsymbol{k}^T\boldsymbol{T}_m}. \tag{10}$$

where i is the imaginary number, and $\lambda_{\boldsymbol{k}} = \prod_{j=1}^d \max(1, 2\pi|k_j|)^{-2}$. Here, the coefficients $b_{\boldsymbol{k}}$ can be efficiently computed via non-linear fast Fourier transform [140]. The data kernel K_{data} has two nice properties. We can show that the subsample $\mathcal{X}_s$ minimizing the kernel discrepancy with K_{data} yields the improved rate in Theorem 1. We can also show that, with all coefficients computed, this discrepancy can be evaluated in $\mathcal{O}(n^2)$ work, *independent* of N (the big data size). We will develop *scalable algorithms to optimize this data discrepancy for LD subsampling*, leveraging recent developments in accelerated gradient descent [71] and randomized algorithms [91].

2.3.3. Implementation and Application. [Years 2–3] We will show the usefulness of the proposed LD subsampling on a broad range of scientific applications. In particular, we will showcase its effectiveness on our UAV application. The goal here is to train an *efficient* predictive model that can be used for *real-time* UAV flight, facilitating engine control decisions within a timeframe of several milliseconds. The challenge is that such a model has to be trained using massive simulation data from computational fluid dynamics models. We will show our LD subsampling approach can indeed facilitate this real-time control for UAV flight. We will also provide specific *implementations* of LD-batch SGD for a broad range of popular learning models (e.g., regression, neural networks, kernel methods), with full documentation on QMCPy (see Sect. 3.2).

2.4. Distribution, Density and Quantile Estimation [FH lead, AO, AS]

2.4.1. Motivation and Preliminary Results. Besides expectations, we may wish to know the distribution, density, and/or quantile function of $Y = f(\boldsymbol{X})$. As opposed to estimating a scalar, μ, we now need to approximate functions, F, ϱ, and/or Q, which require function approximation algorithms and error criteria involving function norms. This is critical in our UAV application: engineers are interested in estimating and quantifying uncertainty on the *distribution* of engine thrust under uncertain operating conditions, which enable robust engine design and control.

Although the distribution, F, is defined as an integral in (2), the integrand as it stands, $\mathbb{1}_{(-\infty,y]}(f(\cdot))$, is insufficiently smooth for QMC theory to apply directly, and thus there has been little developments on LD sampling for such a setting (see Limitation 4). Approximating the density (derivative of F) and quantile function (inverse of F) are harder problems. Despite such theoretical challenges, LD sequences show promise in approximating distributions, densities, and quantiles. Consider two different definitions of $Y = f(X)$, $X \sim \mathcal{U}[0,1]$ as shown in Fig. 5. For both choices the distribution function of Y is $F : y \mapsto \sin^{-1}(y)/\pi + 1/2$. Fig. 5 displays $N = 100$ replications of the empirical distribution, F_{emp} based on $n = 64$ IID and LD samples. The LD points provide better approximations of the distribution, and they provide better approximations for smoother f.

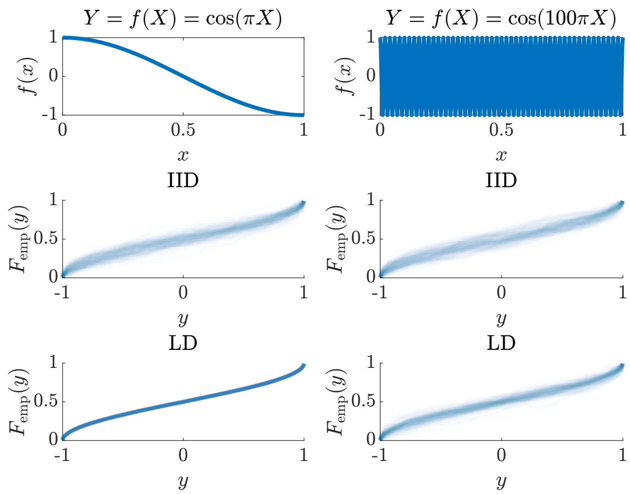

Figure 5. Comparison of the empirical distribution functions of $Y = f(X)$ using $n = 64$ IID and LD X_i and for two definitions of f. The empirical distributions generated by LD points are closer to the true distribution than IID sampling.

A recent study of LD sequences for kernel density estimation [4] has demonstrated that in practice they outperform IID sequences. The authors also derive theoretical error bounds on this estimator for LD sequences, but such bounds appear to be too loose. A tight quantification of uncertainty for density estimation is critical in our UAV application, enabling sound design and control decisions. There has also been little work on investigating (and mitigating) the curse-of-dimensionality for density estimation with LD sequences, a fundamental topic in QMC [31]. In QMC, effective dimension is defined roughly as the number of coordinates of the input vector $\boldsymbol{X}$ that make substantial contributions to $f(\boldsymbol{X})$. One expects that small effective dimension plays an important role in approximating ϱ with LD sequences, just as it is in approximating $\mu = \mathbb{E}[f(\boldsymbol{X})]$ with LD sequences [31]. This is critical in our UAV problem, where there are many design and control parameters for engine operation.

2.4.2. Exploring the Effectiveness of LD Sequences. [Year 1–3] The theory for using LD sequences to estimate distributions, densities, and quantiles is in its infancy, which tempers our ambitions for these problems. Our first step will be to empirically compare the efficiency of LD sequences to IID sequences by inserting LD sequences into standard algorithms that use IID sequences, most likely some form of kernel smoothing [141]. We will experimentally explore how to choose the band-width of kernel smoothing methods. We will explore how the nominal dimension, the effective dimension, and the smoothness of f influence the effectiveness of LD sequences. We will investigate its effectiveness in extensive experiments on the test function library [5].

Based on our computational experiments, we will derive data-driven algorithms that compute optimal bandwidth for smoothing and determine the sample size needed to satisfy the error criterion. We will pursue the two approaches used for stopping criteria for computing μ, namely i) analyzing the Fourier coefficients of f, and ii) assuming f comes from a Gaussian process. We will demonstrate the effectiveness of such methods on our UAV flight application.

3. Broader Impacts

3.1. Dissemination to the Broad Scientific Community. For the proposed suite of methods to become an integral tool for modern scientific discovery, scientists and engineers need to be convinced that our sampling framework can achieve confident and accurate estimates in a cost-efficient manner. This necessitates a *strategic outreach* and *targeted dissemination* of our methodologies to scientific communities who may not be familiar with the potential of LD sampling for accelerating scientific discovery. We will achieve this via a *multi-disciplinary* publication plan that targets a broad multi-disciplinary audience. We will publish in not only statistics, mathematics, and computer science journals, but also top subject-matter journals to make the proposed tools accessible to the scientific community. We will post all work as freely accessible papers on arXiv, with links to open source and well-documented code on GitHub to allow for easier dissemination of our methods to end-users. We will also give talks, tutorials and workshops at prominent statistics, machine learning, mathematics and scientific conferences, educating different communities on promising developments on novel LD methods for furthering scientific progress.

By broadening QMC methods for tackling complex scientific and engineering problems, our project will naturally *introduce new application areas* to the benefits of LD sampling. The accessibility of state-of-the-art QMC methods in our outreach will spur on novel advancements in many scientific disciplines. We will focus specifically on four areas:

3.1.1. Uncertainty Quantification (UQ). UQ is the science of quantifying and managing uncertainty in computational and physical systems [126, 127]. Since data are typically expensive for UQ, a key focus is the design of sampling points for experimentation. QMC can thus yield *great computational savings* for UQ (see [50] for a convincing application in fluid flows). PI SM has ongoing multidisciplinary collaborations on UQ in aerospace engineering [10, 11, 84, 85, 98, 146], nuclear physics [8, 36–38, 67, 68, 82, 88], astrophysics [96, 99, 151] and neuroscience [144, 145], and we will introduce the benefits of LD sampling for UQ in such disciplines.

3.1.2. Machine Learning (ML). Machine learning is a rapidly growing area with broad applications in science and engineering. Given the prevalent use of Monte Carlo in ML algorithms [6, 41, 117], QMC will undoubtedly have a significant impact in this area (see [78] for a stunning application of QMC in image rendering, and [12] for an application of QMC for learning PDEs). In addition to LD big data subsampling (Sect. 2.3), we will *show the advantages of QMC over IID sampling in cutting-edge ML problems*, an area where both PIs have prior publication record [74, 94, 99].

3.1.3. High Energy Physics. Physicists have begun to recognize the advantages of LD sampling for generating parton distributions [25]. We will continue our recent discussions with this group to explore how to best employ LD sampling for a cost-efficient simulation of such problems. PI SM has an extensive publication record in high energy physics (see Section 4.2 and [8, 36–38, 67, 68, 82, 88]).

3.1.4. Bayesian Methodology. In addition to expensive posterior sampling (Sect. 2.1), we will explore two Bayesian areas which can greatly benefit from QMC. The first is probabilistic numerics (PN), which presumes the solution of a mathematical problem follows a prior random process. PI FH with his former PhD student JR developed a fast Bayesian cubature method [65] using lattice LD sampling. Bayesian optimization (BO) aims to minimize a black-box objective function, where evaluations are optimized via an acquisition function in the form of a high-dimensional integral. MM illustrates the advantages of LD sampling for BO [19, qEI with QMCPy], and PI SM has also worked in this area [17, 95]. We will demonstrate the *efficacy of LD sampling* in BO and PN.

3.2. QMCPy as a Proving Ground. As mentioned in Sects. 1.2.2 and 1.2.3, FH, YD, SCTC, AS, MM, JR, PR, and collaborators have created the open source Python QMC library QMCPy [22]. QMCPy has a clear architecture and consistent user interface. It includes LD sequences, variable transformations, and stopping rules. QMCPy is hosted on a GitHub repository with documentation, a suite of doctests and unit tests, Jupyter notebook demos, and an issues board.

With this architecture, QMCPy will serve as an ideal vessel for bridging our suite of methods (and state-of-the-art QMC algorithms) to the scientific community, thus addressing Limitation 5.

QMCPy will continue to grow and *stand in the breach* between research code from individual groups and large-scale software packages. Research groups need to compare their new ideas with the best available. Developers for LD generators need to test them on a variety of use cases and as key components of QMC algorithms. Those with new QMC algorithms need to test them with the best generators. We will connect QMCPy to other libraries in Sect. 1.2.2, taking advantage of what they offer in QMCPy and pushing our more established developments into them. We recently collaborated with Uncertainty Quantification and Model Bridge (UM-Bridge) [27] to make QMCPy compatible with UM-Bridge. A recent gathering at an international QMC conference initiated by FH raised the need to standardize formats for LD generators to allow them to be shared across software libraries. We will pursue initiatives that promote same look-and-feel across QMC software.

3.3. Promoting Proper QMC Practice and Code. QMCPy—software, documentation, academic articles, and conference presentations—will *showcase the right way* to do LD sampling. As an example, the adoption of PyTorch into QMCPy and the tutorial given by FH at MCQMC 2020 [23, 51] prompted a vigorous discussion on the PyTorch issues site [116] that migrated to the SciPy issues site [124]. AO, FH, and other QMC researchers convinced the developers to not omit the first Sobol' point, but to randomize by default. Keeping the first point preserves the net property of the first 2^m Sobol' points and randomization can speed up convergence [112]. In these discussions, it was pointed out that UQLab [101], OpenTurns [110], and other packages routinely drop the first Sobol' point, a bad but understandable practice. The arguments we provided to PyTorch and SciPy developers addressed their concerns. We expect this project to produce fruitful discussions between QMC practitioners, which will promote better practice.

Having the eyes of the QMC community on QMCPy will more *quickly uncover and eradicate bugs*. FH found that randomized PyTorch Sobol' points fell on the boundaries of $[0, 1]^d$, when they never should [116] This was due to a lack of double precision, as discovered by MM. LlAJR found that the Sobol' scrambling in MATLAB was incorrect. This was rectified in R2017a. This highlights how having a larger community using a software library leads to higher quality code.

3.4. Training the Next Generation of *Science-Based* Computational Researchers. There is a critical need for individuals who have the computational and mathematical tools to adeptly work in modern *scientific* teams, geared towards pushing forward the frontier of scientific knowledge and engineering through improved science-based data science tools. Students involved in this project will be given a *multidisciplinary* background in science-based data science, and will learn relevant methods in statistics, mathematics and computer science. This will serve them well in transitioning to not just academic research teams but also government and industry positions.

Computational mathematics and statistics require the use of others' code, hopefully in the form of well-developed software packages. New scholars also need to be trained not only on how to use such packages, but how to contribute to them as well. Students supported by this project will learn to write clean, efficient code that fits package architecture, is documented, and passes doctests. Students will learn about repositories and software engineering tools, which will prepare them well for academic and/or industry opportunities.

3.5. Incorporating Diversity, Equity and Inclusion As with past projects, in seeking students, we will give preference to underrepresented minorities, women, and students from colleges where research experiences are rare. As noted in Sects. 4.1 and 4.2, we have had a strong track record of mentoring students from *diverse* backgrounds and experiences. Five of FH's fifteen students who earned PhDs are women, three of whom are academics. FH supervised an African-American student, CH, in the Summer Undergraduate Research Experience (SURE) program at Illinois Tech, which provides research opportunities particularly for underrepresented groups. CH is now a PhD student at Illinois Tech. Two of SM's seven current PhD students are women, and five of SM's ten undergraduate thesis advisees are women. Many of our undergraduates have enrolled in graduate programs. SM is on the leadership board of the IMS New Researchers Group, which aims to provide career development opportunities for young statisticians, particularly those from underrepresented backgrounds and institutions.

We will leverage these connections to further promote diversity and inclusion, providing students from underrepresented backgrounds and institutions with ample research training and opportunities. The senior personnel on this project include two women (SCTC and YD) and one early-career scholar (SM). Our senior personnel and collaborators include folks with diverse technical expertise, institutions and backgrounds. The students that we mentor in this project will also learn and benefit from thinking from these diverse perspectives.

4. Results from Prior NSF Support

4.1. NSF-DMS-1522687, *Stable, Efficient, Adaptive Algorithms for Approximation and Integration*, \$270,000, August 2015 – July 2018. Fred Hickernell (FH, PI) and Gregory E. Fasshauer (GEF, co-PI) led this project, and SCTC contributed as senior personnel. Other contributors were FH's research students YD (PhD 2015), LJ (PhD 2016), LlAJR (PhD 2016), Da Li (DL, MS 2016), Jiazhen Liu (JL, MS 2018), JR (PhD 2019), Xin Tong (XT, MS 2014, PhD 2020 at the University of Illinois at Chicago), Kan Zhang (KZ, PhD student), Yizhi Zhang (YZ, PhD 2018), and Xuan Zhou (KZ, PhD 2015). Articles, theses, software, and preprints supported in part by this grant include [2, 18, 21, 33, 34, 46, 55–60, 65, 69, 70, 72, 83, 87, 102–106, 118, 119, 149, 150, 152, 153].

4.1.1. Intellectual Merit from Prior NSF Support. FH, SCTC, YD, LlAJR, DL, JR, XT, YZ, and developed several adaptive algorithms for univariate integration, function approximation, and optimization [18, 24, 33, 137, 149], multivariate integration [57, 59, 70], and multivariate function approximation for Banach spaces, $\mathcal{F}$, defined by series representations [34, 35].

4.1.2. Broader Impacts from Prior NSF Support. Publications by GEF, FH, SCTC, students, and collaborators are listed above. We have spoken at many applied mathematics, statistics, and computational science conferences and given colloquium/seminar talks to mathematics and statistics departments. FH co-organized the 2016 Spring Research Conference, gave an invited tutorial at MCQMC 2016, was a program leader for the SAMSI 2017–18 Quasi-Monte Carlo (QMC) Program, and received the 2016 Joseph F. Traub Prize for Achievement in Information-Based Complexity. Our adaptive algorithms have been been implemented in GAIL [21], which has been used in the graduate Monte Carlo at Illinois Tech. The PIs mentored a number of research students; female students include YD, LJ, JL, XT, and Xiaoyang Zhao (MS 2017).

4.2. NSF CSSI Frameworks 2004571 (Subaward WSU20076). *X-Ion Collisions with a Statistically and Computationally Advanced Program Envelope (X-SCAPE)*, \$696,442, July 2020 – June 2024. High-energy colliders study the interaction between subatomic particles and environments produced in the collision of protons with protons and with nuclei. This requires an elaborate theoretical, statistical and computational framework. The X-SCAPE (JETSCAPE) collaboration is a multi-disciplinary team of physicists, computer scientists, and statisticians, who are engaged in the construction of such a framework. SM is a Duke co-PI in this ongoing project, and is responsible for leading statistical and ML developments.

4.2.1. Intellectual Merit from Prior NSF Support. The JETSCAPE collaboration has developed the first open-source simulation framework for the high energy sector of heavy-ion collisions and a Bayesian framework to rigorously compare event generators with experimental data. This has resulted in numerous publications in top physics journals and conferences [7, 8, 36–39, 76, 81, 82, 113, 129, 133] and top statistical / ML journals and conferences [13, 14, 63, 100, 145, 147, 148, 151].

4.2.2. Broader Impacts from Prior NSF Support. The primary broader impacts of the X-SCAPE collaboration have been in the training of its graduate students and postdocs, through regular meetings, collaboration gatherings, and joint projects. The collaboration also influences the training of the wider US nuclear physics workforce through its annual winter school and workshops. SM is supporting several students on this project, including two female PhD students and two female undergraduates. SM is co-organizing the 2023 Spring Research Conference, and received the 2022 Rosenbluth-Blackwell Award for exceptional career achievements by a junior Bayesian researcher.

5. Strengths of This Team and Collaboration Plan

Our team combines senior personnel with diverse backgrounds, career stages, and institutions. We will have regular meetings to share progress and collaborate on papers and software. These will be held both at our own institutions and via video conference between institutions.

5.1. Senior Personnel. FH has been the lead PI on the GAIL [21] MATLAB software project that contains many of the stopping criteria in QMCPy. His expertise is in the numerical analysis of QMC and other multivariate problems as well as theoretically justified adaptive numerical algorithms. As a former editorial board member for major computational mathematics journals, a Fellow of the Institute of Mathematical Statistics, and a co-leader of SAMSI's program on QMC in 2017-18, FH understands the interface between computational mathematics and statistics.

SM is an Assistant Professor at Duke, specializing in Bayesian computation, big data analytics, and computer experiments. As an Associate Editor for *Technometrics* (a top engineering statistics journal) and *Data Science in Science*, and the recipient of major awards from the American Statistical Association and the International Society of Bayesian Analysis, SM provides expertise on statistical / ML methods and applications. SM will lead the activities at Duke and oversee the proposed efforts on Bayesian sampling and big data subsampling.

YD is an associate teaching professor at Illinois Tech, who teaches Monte Carlo methods in finance and programming for data analytics. She is experienced in the theory and implementation of QMC and adaptive algorithms, and is one of the main developers of the GAIL project.

SCTC serves as a Principal Data Scientist at SAS in the financial risk group. She is a research associate professor at Illinois Tech. SCTC has co-led the GAIL and QMCPy projects and is an expert in best engineering practices for numerical software. She is a co-winner of the 2011 Society for Industrial and Applied Mathematics Activity Group on Linear Algebra best paper prize.

5.2. Students. AS began his PhD in applied mathematics at Illinois Tech in Fall 2021 after completing a dual BS/MS degree at Illinois Tech and doing a large majority of the coding of QMCPy to date. CH is an African American Illinois Tech PhD student who began in Fall 2022. IJ, TT, JM and KL are Statistical Science PhD students at Duke University, working on theory and methods for computer experiments, uncertainty quantification, Bayesian computation and big data analytics, with applications to physics and engineering. PhD students supported by this project will work with the senior personnel to tackle major theoretical and methodological challenges. Undergraduate students will focus on new features that can be implemented mostly over the course of a summer. They will be mentored by senior personnel and PhD students.

5.3. Collaborators. The responsibilities of these unpaid collaborators will be to discuss research ideas of common interest, involve their research groups as appropriate, and publish significant results. We have identifed above some of the projects that they may be involved in. MM convinced his company to fund the early development of QMCPy. He wanted to spread the advantages of LD sampling to the tech industry. MM will advise us on the continued development of QMCPy, and continue to help us spread the word among his network in the machine learning community. AO has engaged with the PIs in conversations about QMC for many years and is particularly an expert in randomized QMC. AO has taken a keen interest in QMCPy and used it in his own research. PR has wide experience in multi-level methods and uncertainty quantifciation applications of QMC. CO is an expert in PN. He will provide expertise on the use of Stein discrepancy points and also Bayesian numerics for stopping criteria for QMC methods.

References Cited

[1] Adams BM, Bohnhoff WJ, Dalbey KR, Ebeida MS, Eddy JP, Eldred MS, Hooper RW, Hough PD, Hu KT, Jakeman JD, Khalil M, Maupin KA, Monschke JA, Ridgway EM, Rushdi AA, Seidl DT, Stephens JA, Swiler LP, Tran A, Winokur JG (2022) Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.16 User's Manual. Sandia National Laboratories

[2] Ala G, Fasshauer GE, Francomano E, McCourt M, Ganci S (2017) An augmented MFS approach for brain activity reconstruction. Math Comput Simul 141:3–15, DOI 10.1016/j.matcom.2016.11.009

[3] Bachem O, Lucic M, Krause A (2017) Practical coreset constructions for machine learning. arXiv preprint arXiv:170306476

[4] Ben Abdellah A, L'Ecuyer P, Owen AB, Puchhammer F (2021) Density estimation by randomized quasi-Monte Carlo. SIAM/ASA J Uncertain Quantif 9(280-301), DOI 10.1137/19M1259213

[5] Bingham D (2017) The virtual library of simulation experiments: Test functions and data sets. URL https://www.sfu.ca/~ssurjano/index.html

[6] Bottou L (2010) Large-scale machine learning with stochastic gradient descent. In: Proceedings of COMPSTAT'2010, Springer, pp 177–186

[7] Cao S, Park C, Barbieri RA, Bass SA, Bazow D, Bernhard J, Coleman J, Fries R, Gale C, He Y, Heinz U, Jacak BV, Jacobs PM, Jeon S, Kordell M, Kumar A, Luo T, Majumder A, Nejahi Y, Pablos D, Pang LG, Putschke JH, Roland G, Rose S, Schenke B, Schwiebert L, Shen C, Sirimanna C, Soltz RA, Velicanu D, Vujanovic G, Wang XN, Wolpert RL (2017) Multistage Monte Carlo simulation of jet modification in a static medium. Phys Rev C 96:024,909, DOI 10.1103/PhysRevC.96.024909, URL https://link.aps.org/doi/10.1103/PhysRevC.96.024909

[8] Cao S, Chen Y, Coleman J, Mulligan J, Jacobs PM, Soltz RA, Angerami A, Arora R, Bass SA, Cunqueiro L, Dai T, Du L, Ehlers R, Elfner H, Everett D, Fan W, Fries RJ, Gale C, Garza F, He Y, Heffernan M, Heinz U, Jacak BV, Jeon S, Ke W, Kim B, Kordell M, Kumar A, Majumder A, Mak S, McNelis M, Nattrass C, Oliinychenko D, Park C, Paquet JF, Putschke JH, Roland G, Silva A, Schenke B, Schwiebert L, Shen C, Sirimanna C, Tachibana Y, Vujanovic G, Wang XN, Wolpert RL, Xu Y (2021) Determining the jet transport coefficient $\hat{q}$ from inclusive hadron suppression measurements using Bayesian parameter estimation. Phys Rev C 104:024,905, DOI 10.1103/PhysRevC.104.024905, URL https://link.aps.org/doi/10.1103/PhysRevC.104.024905

[9] Chan TM (2006) Faster core-set constructions and data-stream algorithms in fixed dimensions. Comput Geom 35(1-2):20–35

[10] Chang YH, Zhang L, Wang X, Yeh ST, Mak S, Sung CL, Wu CFJ, Yang V (2019) Kernel-smoothed proper orthogonal decomposition–based emulation for spatiotemporally evolving flow dynamics prediction. AIAA Journal 57(12):5269–5280

[11] Chang YH, Wang X, Zhang L, Li Y, Mak S, Wu CFJ, Yang V (2021) Reduced-order modeling for complex flow emulation by common kernel-smoothed proper orthogonal decomposition. AIAA Journal 59(9):3291–3303

[12] Chen J, Du R, Li P, Lyu L (2019) Quasi-Monte Carlo sampling for machine-learning partial differential equations. arXiv preprint arXiv:191101612

[13] Chen J, Mak S, Joseph VR, Zhang C (2020) Function-on-function kriging, with applications to three-dimensional printing of aortic tissues. Technometrics 63:384–395

[14] Chen J, Mak S, Joseph VR, Zhang C (2022) Adaptive design for gaussian process regression under censoring. The Annals of Applied Statistics 16(2):744–764

[15] Chen WY, Mackey L, Gorham J, Briol FX, Oates C (2018) Stein points. In: Dy J, Krause A (eds) Proceedings of the 35th International Conference on Machine Learning, PMLR, Proceedings of Machine Learning Research, vol 80, pp 844–853, URL https://proceedings.mlr.press/v80/chen18f.html

[16] Chen Y, Welling M, Smola A (2012) Super-samples from kernel herding. arXiv preprint arXiv:12033472

[17] Chen Z, Mak S, Wu CFJ (2019) A hierarchical expected improvement method for Bayesian optimization. arXiv preprint arXiv:191107285

[18] Choi SCT, Ding Y, Hickernell FJ, Tong X (2017) Local adaption for approximation and minimization of univariate functions. J Complexity 40:17–33, DOI 10.1016/j.jco.2016.11.005

[19] Choi SCT, Hickernell FJ, McCourt MJ, Sorokin A (2020) QMCPy blogs. URL https://qmcpy.wpcomstaging.com

[20] Choi SCT, Hickernell FJ, Jagadeeswaran R, McCourt MJ, Sorokin A (2020) QMCPy documentation. URL https://qmcpy.readthedocs.io/en/latest/

[21] Choi SCT, Ding Y, Hickernell FJ, Jiang L, Jiménez Rugama LlA, Li D, Jagadeeswaran R, Tong X, Zhang K, Zhang Y, Zhou X (2021) GAIL: Guaranteed Automatic Integration Library (versions 1.0–2.3.2). MATLAB software, http://gailgithub.github.io/GAIL_Dev/, DOI 10.5281/zenodo.4018189

[22] Choi SCT, Hickernell FJ, Jagadeeswaran R, McCourt M, Sorokin A (2022) QMCPy: A quasi-Monte Carlo Python library. DOI 10.5281/zenodo.3964489, URL https://qmcsoftware.github.io/QMCSoftware/

[23] Choi SCT, Hickernell FJ, Jagadeeswaran R, McCourt MJ, Sorokin AG (2022) Quasi-Monte Carlo software. In: Keller A (ed) Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Oxford, England, August 2020, Springer, Cham, Springer Proceedings in Mathematics and Statistics, pp 23–50, https://arxiv.org/abs/2102.07833

[24] Clancy N, Ding Y, Hamilton C, Hickernell FJ, Zhang Y (2014) The cost of deterministic, adaptive, automatic algorithms: Cones, not balls. J Complexity 30:21–45, DOI 10.1016/j.jco.2013.09.002

[25] Courtoy A, Huston J, Nadolsky P, Xie K, Yan M, Yuan CP (2022) Parton distributions need representative sampling. preprint, URL https://arxiv.org/pdf/2205.10444.pdf

[26] Darmon Y, Godin M, L'Ecuyer P, Jemel A, Marion P, Munger D (2018) LatNet builder. URL https://github.com/umontreal-simul/latnetbuilder

[27] Davis A, Parno M, Reinarz A, Seelinger L (2022) UQ and model bridge (UM-bridge). URL https://um-bridge-benchmarks.readthedocs.io/en/docs/

[28] Debusschere B (2022) UQ toolkit. URL https://www.sandia.gov/uqtoolkit/

[29] Debusschere B, Najm HN, Pébay PP, Knio OM, Ghanem RG, Maître OPL (2004) Numerical challenges in the use of polynomial chaos representations for stochastic processes. SIAM J Sci Comput pp 698–719, URL 10.1137/S1064827503427741

[30] Dick J, Pillichshammer F (2010) Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration. Cambridge University Press, Cambridge

[31] Dick J, Kuo FY, Sloan IH (2013) High-dimensional integration: the quasi-Monte Carlo way. Acta Numerica 22:133–288

[32] Dick J, Kritzer P, Pillichshammer F (2022) Lattice Rules: Numerical Integration, Approximation, and Discrepancy. Springer Series in Computational Mathematics, Springer Cham, DOI https://doi.org/10.1007/978-3-031-09951-9

[33] Ding Y (2015) Guaranteed adaptive univariate function approximation. PhD thesis, Illinois Institute of Technology

[34] Ding Y, Hickernell FJ, Jiménez Rugama LlA (2020) An adaptive algorithm employing continuous linear functionals. In: L'Ecuyer P, Tuffin B (eds) Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Rennes, France, July 2018, Springer, Cham, Springer Proceedings in

Mathematics and Statistics, vol 324, pp 161–181, DOI 10.1007/978-3-030-43465-6_8

[35] Ding Y, Hickernell FJ, Kritzer P, Mak S (2020) Adaptive approximation for multivariate linear problems with inputs lying in a cone. In: Hickernell FJ, Kritzer P (eds) Multivariate Algorithms and Information-Based Complexity, DeGruyter, Berlin/Boston, pp 109–145, DOI 10.1515/9783110635461-007

[36] Ehlers R, Angerami A, Arora R, Bass S, Cao S, Chen Y, Du L, Dai T, Elfner H, Fan W, Angerami A, Arora R, Cao S, Chen Y, Dai T, Du L, Ehlers R, Elfner H, Fries RJ, Gale C, He Y, Heffernan M, Heinz U, Jacak BV, Jacobs PM, Jeon S, Ji Y, Kauder K, Kasper L, Ke W, Kelsey M, Kordell M, Kumar A, Latessa J, Lee YJ, Liyanage D, Lopez A, Luzum M, Majumder A, Mak S, Mankolli A, Martin C, Mehryar H, Mengel T, Mulligan J, Nattrass C, Oliinychenko D, Paquet JF, Putschke JH, Roland G, Schenke B, Schwiebert L, Sengupta A, Shen C, Silva A, Sirimanna C, Soeder D, Soltz RA, Soudi I, Staudenmaier J, Strickland M, Tachibana Y, Velkovska J, Vujanovic G, Wang XN, Wolpert RL, Zhao W (2022) Bayesian analysis of QGP jet transport using multi-scale modeling applied to inclusive hadron and reconstructed jet data. arXiv preprint arXiv:220807950

[37] Everett D, Ke W, Paquet JF, Vujanovic G, Bass SA, Du L, Gale C, Heffernan M, Heinz U, Liyanage D, Luzum M, Majumder A, McNelis M, Shen C, Xu Y, Angerami A, Cao S, Chen Y, Coleman J, Cunqueiro L, Dai T, Ehlers R, Elfner H, Fan W, Fries RJ, Garza F, He Y, Jacak BV, Jacobs PM, Jeon S, Kim B, Kordell M, Kumar A, Mak S, Mulligan J, Nattrass C, Oliinychenko D, Park C, Putschke JH, Roland G, Schenke B, Schwiebert L, Silva A, Sirimanna C, Soltz RA, Tachibana Y, Wang XN, Wolpert RL (2021) Multisystem Bayesian constraints on the transport coefficients of QCD matter. Phys Rev C 103:054,904, DOI 10.1103/PhysRevC.103.054904, URL https://link.aps.org/doi/10.1103/PhysRevC.103.054904

[38] Everett D, Ke W, Paquet JF, Vujanovic G, Bass SA, Du L, Gale C, Heffernan M, Heinz U, Liyanage D, Luzum M, Majumder A, McNelis M, Shen C, Xu Y, Angerami A, Cao S, Chen Y, Coleman J, Cunqueiro L, Dai T, Ehlers R, Elfner H, Fan W, Fries RJ, Garza F, He Y, Jacak BV, Jacobs PM, Jeon S, Kim B, Kordell M, Kumar A, Mak S, Mulligan J, Nattrass C, Oliinychenko D, Park C, Putschke JH, Roland G, Schenke B, Schwiebert L, Silva A, Sirimanna C, Soltz RA, Tachibana Y, Wang XN, Wolpert RL (2021) Phenomenological constraints on the transport properties of QCD matter with data-driven model averaging. Phys Rev Lett 126:242,301, DOI 10.1103/PhysRevLett.126.242301, URL https://link.aps.org/doi/10.1103/PhysRevLett.126.242301

[39] Everett D, Oliinychenko D, Luzum M, Paquet JF, Vujanovic G, Bass SA, Du L, Gale C, Heffernan M, Heinz U, Kasper L, Ke W, Liyanage D, Majumder A, Mankolli A, Shen C, Soeder D, Velkovska J, Angerami A, Arora R, Cao S, Chen Y, Dai T, Ehlers R, Elfner H, Fan W, Fries RJ, Garza F, He Y, Jacak BV, Jacobs PM, Jeon S, Kelsey M, Kordell M, Kumar A, Latessa J, Lee YJ, Lopez A, Mak S, Martin C, Mehryar H, Mengel T, Mulligan J, Nattrass C, Putschke JH, Roland G, Schenke B, Schwiebert L, Silva A, Sirimanna C, Soltz RA, Staudenmaier J, Strickland M, Tachibana Y, Wang XN, Wolpert RL (2022) Role of bulk viscosity in deuteron production in ultrarelativistic nuclear collisions. Physical Review C 106(6):064,901

[40] Fan W, Vujanovic G, Bass S, Majumder A, Angerami A, Arora R, Cao S, Chen Y, Dai T, Du L, Ehlers R, Elfner H, Fries RJ, Gale C, He Y, Heffernan M, Heinz U, Jacak BV, Jacobs PM, Jeon S, Ji Y, Kauder K, Kasper L, Ke W, Kelsey M, Kordell M, Kumar A, Latessa J, Lee YJ, Liyanage D, Lopez A, Luzum M, Mak S, Mankolli A, Martin C, Mehryar H, Mengel T, Mulligan J, Nattrass C, Oliinychenko D, Paquet JF, Putschke JH, Roland G, Schenke B, Schwiebert L, Sengupta A, Shen C, Silva A, Sirimanna C, Soeder D, Soltz RA, Soudi I, Staudenmaier J, Strickland M, Tachibana Y, Velkovska J, Wang XN, Wolpert RL, Zhao W (2022) Multi-scale evolution of charmed particles in a nuclear medium. arXiv preprint arXiv:220800983

[41] Friedman JH (2002) Stochastic gradient boosting. Comput Statist & Data Anal 38(4):367–378
[42] Gelman A, Carlin JB, Stern HS, Rubin DB (1995) Bayesian Data Analysis. Chapman and Hall/CRC
[43] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian Data Analysis. CRC Texts in Statistical Science, Chapman & Hall
[44] Giles MB (2008) Multilevel Monte Carlo path simulation. Operations Research 56(3):607–617
[45] Giles MB (2015) Multilevel Monte Carlo methods. Acta Numer 24:259–328, DOI 10.1017/S096249291500001X
[46] Gilquin L, Jiménez Rugama LlA, Arnaud E, Hickernell FJ, Monod H, Prieur C (2017) Iterative construction of replicated designs based on Sobol' sequences. C R Math Acad Sci Paris 355:10–14, DOI 10.1016/j.crma.2016.11.013
[47] Hahn T (2005) CUBA—a library for multidimensional numerical integration. Comput Phys Commun 168:78–95, DOI 10.1016/j.cpc.2005.01.010, URL https://feynarts.de/cuba/
[48] Haji-Latif AL, Nobile F, Tempone R (2016) Multi-index Monte Carlo: when sparsity meets sampling. Numer Math 132:767–806, DOI 10.1007/s00211-015-0734-5
[49] Heinrich S (2001) Multilevel Monte Carlo methods. In: Margenov S, Wasniewski J, Yalamov PY (eds) Large-Scale Scientific Computing, Third International Conference, LSSC 2001, Springer-Verlag, Lecture Notes in Computer Science, vol 2179, pp 58–67
[50] Herrmann L, Schwab C (2020) Multilevel quasi-Monte Carlo uncertainty quantification for advection-diffusion-reaction. In: L'Ecuyer P, Tuffin B (eds) Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Rennes, France, July 2018, Springer, Cham, Springer Proceedings in Mathematics and Statistics, vol 324, pp 31–67, DOI 10.1007/978-3-030-43465-6_2
[51] Hickernell FJ (2020) Quasi-Monte Carlo software. URL https://mcqmc20.web.ox.ac.uk/programme
[52] Hickernell FJ, Sorokin A (2020) Quasi-Monte Carlo software tutorial. URL https://media.ed.ac.uk/playlist/dedicated/51612401/1_0z0wec2z/1_2k12mwiw
[53] Hickernell FJ (1998) A generalized discrepancy and quadrature error bound. Math Comp 67:299–322, DOI 10.1090/S0025-5718-98-00894-1
[54] Hickernell FJ (1999) Goodness-of-fit statistics, discrepancies and robust designs. Statist Probab Lett 44:73–78, DOI 10.1016/S0167-7152(98)00293-4
[55] Hickernell FJ (2018) The trio identity for quasi-Monte Carlo error analysis. In: Glynn P, Owen A (eds) Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Stanford, USA, August 2016, Springer-Verlag, Berlin, Springer Proceedings in Mathematics and Statistics, pp 3–27, DOI 10.1007/978-3-319-91436-7
[56] Hickernell FJ, Jagadeeswaran R (2019) Comment on "Probabilistic integration: A role in statistical computation?". Statist Sci 34:23–28, DOI 10.1214/18-STS660
[57] Hickernell FJ, Jiménez Rugama LlA (2016) Reliable adaptive cubature using digital sequences. In: Cools R, Nuyens D (eds) Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Leuven, Belgium, April 2014, Springer-Verlag, Berlin, Springer Proceedings in Mathematics and Statistics, vol 163, pp 367–383, arXiv:1410.8615 [math.NA]
[58] Hickernell FJ, Choi SCT, Jiang L, Jiménez Rugama LlA (2018) Monte Carlo simulation, automatic stopping criteria for. In: Davidian M, Everitt B, Kenett RS, Molenberghs G, Piegorsch W, Ruggeri F (eds) Wiley StatsRef-Statistics Reference Online, John Wiley & Sons Ltd., DOI 10.1002/9781118445112.stat08035
[59] Hickernell FJ, Jiménez Rugama LlA, Li D (2018) Adaptive quasi-Monte Carlo methods for cubature. In: Dick J, Kuo FY, Woźniakowski H (eds) Contemporary Computational Mathematics — a celebration of the 80th birthday of Ian Sloan, Springer-Verlag, pp 597–619, DOI 10.1007/978-3-319-72456-0

[60] Hickernell FJ, Kritzer P, Woźniakowski H (2020) Exponential tractability of linear tensor product problems. In: 2018 MATRIX Annals, MATRIX Book Series, Springer Nature Switzerland AG, pp 61–78

[61] Hjort NL, Holmes C, Müller P, Walker SG (2010) Bayesian Nonparametrics, vol 28. Cambridge University Press

[62] Hoffman MD, Gelman A (2014) The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res 15(1):1593–1623

[63] Huang C, Joseph VR, Mak S (2022) Population quasi-monte carlo. Journal of Computational and Graphical Statistics pp 1–14

[64] Huggins J, Campbell T, Broderick T (2016) Coresets for scalable Bayesian logistic regression. In: Advances in Neural Information Processing Systems, pp 4080–4088

[65] Jagadeeswaran R, Hickernell FJ (2019) Fast automatic Bayesian cubature using lattice sampling. Stat Comput 29:1215–1229, DOI 10.1007/s11222-019-09895-9

[66] Jagadeeswaran R, Hickernell FJ (2022) Fast automatic Bayesian cubature using Sobol' sampling. In: Botev Z, Keller A, Lemieux C, Tuffin B (eds) Advances in Modeling and Simulation: Festschrift in Honour of Pierre L'Ecuyer, Springer, Cham, pp 301–318, DOI 10.1007/978-3-031-10193-9_15

[67] Ji Y, Mak S, Soeder D, Paquet J (2021) A graphical Gaussian process model for multi-fidelity emulation of expensive computer codes. arXiv preprint arXiv:210800306

[68] Ji Y, Yuchi HS, Soeder D, Paquet JF, Bass SA, Joseph VR, Wu C, Mak S (2022) Multi-stage multi-fidelity gaussian process modeling, with application to heavy-ion collisions. arXiv preprint arXiv:220913748

[69] Jiménez Rugama LlA, Gilquin L (2018) Reliable error estimation for Sobol' indices. Statistics and Computing 28:725–738, DOI 10.1007/s11222-017-9759-1

[70] Jiménez Rugama LlA, Hickernell FJ (2016) Adaptive multidimensional integration based on rank-1 lattices. In: Cools R, Nuyens D (eds) Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Leuven, Belgium, April 2014, Springer-Verlag, Berlin, Springer Proceedings in Mathematics and Statistics, vol 163, pp 407–422, arXiv:1411.1966

[71] Jin C, Netrapalli P, Jordan MI (2018) Accelerated gradient descent escapes saddle points faster than gradient descent. In: Conference On Learning Theory, PMLR, pp 1042–1085

[72] Johnson T, Fasshauer GE, Hickernell FJ (2018+) Characterizing reproducing kernel Hilbert spaces. In preparation

[73] Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. Journal of Global Optimization 13(4):455–492

[74] Joseph VR, Mak S (2021) Supervised compression of big data. Statistical Analysis and Data Mining: The ASA Data Science Journal 14(3):217–229

[75] Joseph VR, Dasgupta T, Tuo R, Wu CJ (2015) Sequential exploration of complex surfaces using minimum energy designs. Technometrics 57(1):64–74

[76] Kauder K (2019) JETSCAPE v1.0 quickstart guide. Nuclear Phys A 982:615–618

[77] Kaufman CG, Bingham D, Habib S, Heitmann K, Frieman JA (2011) Efficient emulators of computer experiments using compactly supported correlation functions, with an application to cosmology. The Annals of Applied Statistics 5(4):2470–2492

[78] Keller A (2013) Quasi-Monte Carlo image synthesis in a nutshell. In: Dick J, Kuo FY, Peters GW, Sloan IH (eds) Monte Carlo and Quasi-Monte Carlo Methods 2012, Springer-Verlag, Berlin, Springer Proceedings in Mathematics and Statistics, vol 65, pp 213–249, DOI 10.1007/978-3-642-41095-6

[79] Krishna A, Mak S, Joseph R (2019) Distributional clustering: A distribution-preserving clustering method. arXiv preprint arXiv:191105940

[80] Kucherenko S (2022) BRODA. URL https://www.broda.co.uk/index.html

[81] Kumar A, Tachibana Y, Pablos D, Sirimanna C, Fries R, Angerami A, Bass S, Cao S, Coleman J, Cunqueiro L (2019) The JETSCAPE framework: p+p results. arXiv preprint arXiv:191005481
[82] Kumar A, Tachibana Y, Sirimanna C, Vujanovic G, Cao S, Majumder A, Chen Y, Du L, Ehlers R, Everett D, Elfner H, Fries RJ, Gale C, He Y, Heffernan M, Heinz U, Jacak BV, Jacobs PM, Jeon S, Ji Y, Kauder K, Kasper L, Ke W, Kelsey M, Kordell M, , Latessa J, Lee YJ, Liyanage D, Lopez A, Luzum M, Mak S, Mankolli A, Martin C, Mehryar H, Mengel T, Mulligan J, Nattrass C, Oliinychenko D, Paquet JF, Putschke JH, Roland G, Schenke B, Schwiebert L, Sengupta A, Shen C, Silva A, Soeder D, Soltz RA, Soudi I, Staudenmaier J, Strickland M, Velkovska J, Wang XN, Wolpert RL, Zhao W (2022) Inclusive jet and hadron suppression in a multi-stage approach. arXiv preprint arXiv:220401163
[83] Li D (2016) Reliable quasi-Monte Carlo with control variates. Master's thesis, Illinois Institute of Technology
[84] Li Y, Wang X, Mak S, Yeh ST, Lin LH, Wu CFJ, Yang V (2017) A two-stage transfer function identification methodology and its applications to bi-swirl injectors. In: 53rd AIAA/SAE/ASEE Joint Propulsion Conference, p 4933
[85] Li Y, Wang X, Mak S, Sung CL, Wu J, Yang V (2018) Uncertainty quantification of flame transfer function under a Bayesian framework. In: 2018 AIAA Aerospace Sciences Meeting, p 1187
[86] Link WA, Eaton MJ (2012) On thinning of chains in MCMC. Methods in Ecology and Evolution 3(1):112–115
[87] Liu J (2018) Adaptive quadrature with a general error criterion. Master's thesis, Illinois Institute of Technology
[88] Liyanage D, Ji Y, Everett D, Heffernan M, Heinz U, Mak S, Paquet JF (2022) Efficient emulation of relativistic heavy ion collisions with transfer learning. Physical Review C 105(3):034,910
[89] Lopes A (2020) Recreating Big Bang matter on Earth. https://phys.org/news/2020-12-recreating-big-earth.html, accessed: 2022-12-27
[90] Ma P, Mahoney MW, Yu B (2015) A statistical perspective on algorithmic leveraging. J Mach Learn Res 16(1):861–911
[91] Mahoney MW (2011) Randomized algorithms for matrices and data. arXiv preprint arXiv:11045557
[92] Mak S, Joseph VR (2017) Projected support points, with application to optimal MCMC reduction. arXiv preprint arXiv:170806897
[93] Mak S, Joseph VR (2018) Minimax and minimax projection designs using clustering. J Comput and Graph Statist 27(1):166–178
[94] Mak S, Joseph VR (2018) Support points. Ann Statist 46(6A):2562–2592
[95] Mak S, Wu CFJ (2019) Analysis-of-marginal-tail-means (ATM): A robust method for discrete black-box optimization. Technometrics 61(4):545–559
[96] Mak S, Xie Y (2018) Maximum entropy low-rank matrix recovery. IEEE J Sel Topics in Signal Process 12(5):886–901
[97] Mak S, Bingham D, Lu Y (2016) A regional compound Poisson process for hurricane and tropical storm damage. J R Stat Soc Ser C Appl Stat 65(5):677–703
[98] Mak S, Sung CL, Wang X, Yeh ST, Chang YH, Joseph VR, Yang V, Wu CFJ (2018) An efficient surrogate model for emulation and physics extraction of large eddy simulations. J Amer Statist Assoc 113(524):1443–1456
[99] Mak S, Yuchi HS, Xie Y (2021) Information-guided sampling for low-rank matrix completion. In: ICML Workshop on Information Theoretic Methods for Rigorous, Responsible, and Reliable Machine Learning

[100] Mak S, Zhou Y, Hoang L, Wu C (2022) Tsec: a framework for online experimentation under experimental constraints. Technometrics , to appear

[101] Marelli S, Sudret B (2014) UQLab: A framework for uncertainty quantification in MATLAB. In: The 2nd International Conference on Vulnerability and Risk Analysis and Management (ICVRAM 2014), ASCE Library, pp 2554–2563, URL https://www.uqlab.com

[102] Martin B, Elsherbeni A, Fasshauer GE, Hadi M (2018) Improved FDTD method around dielectric and pec interfaces using RBF-FD techniques. In: 2018 International Applied Computational Electromagnetics Society Symposium (ACES), DOI 10.23919/ROPACES.2018.8364209

[103] McCourt M, Fasshauer GE (2017) Stable likelihood computation for Gaussian random fields. In: Pesenson I, Le Gia Q, Mayeli A, Mhaskar H, Zhou DX (eds) Recent Applications of Harmonic Analysis to Function Spaces, Differential Equations, and Data Science: Novel Methods in Harmonic Analysis, vol 2, Basel, Birkäuser, pp 917–942

[104] McCourt MJ, Fasshauer GE, Kozak D (2018) A nonstationary designer space-time kernel. In: 32nd Conference on Neural Information Processing Systems (NIPS 2018)

[105] Mishra PK, Nath SK, Sen MK, Fasshauer GE (2018) Hybrid Gaussian-cubic radial basis function for scattered data interpolation. Comput Geosciences 22:1203–1218, DOI 10.1007/s10596-018-9747-3

[106] Mishra PK, Fasshauer GE, Sen MK, Ling L (2019) A stabilized radial basis-finite difference (RBF-FD) method with hybrid kernels. Comput Math Appl 77:2354–2368, DOI 10.1016/j.camwa.2018.12.027

[107] Niederreiter H (1992) Random Number Generation and Quasi-Monte Carlo Methods. CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia

[108] Nocedal J, Wright S (2006) Numerical Optimization. Springer Science & Business Media

[109] Nuyens D (2017) Magic point shop. URL https://people.cs.kuleuven.be/~dirk.nuyens/qmc-generators/

[110] OpenTURNS Developers (2020) An open source initiative for the Treatment of Uncertainties, Risks 'N Statistics. URL http://www.openturns.org

[111] Owen A (2003) Data squashing by empirical likelihood. Data Min Knowl Disc 7(1):101–113

[112] Owen AB (2020) On dropping the first Sobol' point. arXiv preprint arXiv:200808051

[113] Park C, Angerami A, Bass S, Cao S, Coleman J, Cunqueiro L, Dai T, Du L, Elfner H, Everett D (2019) Multi-stage jet evolution through QGP using the JETSCAPE framework: inclusive jets, correlations and leading hadrons. arXiv preprint arXiv:190205934

[114] Parno M, Davis A, Seelinger L, Marzouk Y (2014) MIT uncertainty quantification (muq) library. URL https://mituq.bitbucket.io/source/_site/index.html

[115] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2019) PyTorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems 32:8026–8037

[116] PyTorch Developers (2020) PyTorch discussion on Sobol' sequence implementation. URL https://github.com/pytorch/pytorch/issues/32047

[117] Quiroz M, Kohn R, Villani M, Tran MN (2018) Speeding up MCMC by efficient data subsampling. Journal of the American Statistical Association 114:831–843

[118] Rashidinia J, Fasshauer GE, Khasi M (2016) A stable method for the evaluation of Gaussian radial basis function solutions of interpolation and collocation problems. Comput Math Appl 72(1):178–193, DOI 10.1016/j.camwa.2016.04.048

[119] Rashidinia J, Khasi M, Fasshauer GE (2018) A stable Gaussian radial basis function method for solving nonlinear unsteady convection-diffusion-reaction equations. Comput Math Appl 75:1831–1850, DOI 10.1016/j.camwa.2017.12.007

[120] Rosenthal JS (2017) Simple confidence intervals for MCMC without CLTs. Electronic Journal of Statistics 11(1):211–214

[121] Saltelli A (2002) Making best use of model evaluations to compute sensitivity indices. Comput Phys Commun 145:280–297, DOI 10.1016/S0010-4655(02)00280-1

[122] Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008) Global Sensitivity Analysis: The Primer. John Wiley & Sons Ltd., Chichester, England

[123] Santner TJ, Williams BJ, Notz WI, Williams BJ (2003) The Design and Analysis of Computer Experiments. Springer

[124] SciPy Developers (2020) scipy discussion of Sobol' sequence implementation. URL https://github.com/scipy/scipy/pull/10844

[125] Sloan IH, Joe S (1994) Lattice Methods for Multiple Integration. Oxford University Press, Oxford

[126] SmartUQ Developers (2020) SmartUQ. URL https://www.smartuq.com/resources/uncertainty-quantification/

[127] Smith RC (2014) Uncertainty Quantification: theory, implementation, and applications. SIAM, Philadelphia

[128] Sobol' IM (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math Comput Simul 55(1-3):271–280

[129] Soltz R (2018) Bayesian extraction of q with multi-stage jet evolution approach. In: Proceedings, 9th International Conference on Hard and Electromagnetic Probes of High-Energy Nuclear Collisions: Hard Probes, pp 1–5

[130] Sorokin AG, Jagadeeswaran R (2022+) Monte Carlo for vector functions of integrals. In: Hinrichs A, Kritzer P, Pillichshammer F (eds) Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Linz, Austria, July 2022, Springer, Cham, Springer Proceedings in Mathematics and Statistics, in preparation for submission for publication

[131] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958

[132] Sung CL, Ji Y, Tang T, Mak S (2022) Stacking designs: designing multi-fidelity computer experiments with confidence. arXiv preprint arXiv:221100268

[133] Tachibana Y, Angerami A, Bass S, Cao S, Coleman J, Cunqueiro L, Dai T, Du L, Elfner H, Everett D (2018) Jet substructure modification in a QGP from a multi-scale description of jet evolution with JETSCAPE. arXiv preprint arXiv:181206366

[134] TF Quant Finance Contributors (2021) TF Quant Finance: TensorFlow based Quant Finance library. URL https://github.com/google/tf-quant-finance

[135] The MathWorks, Inc (2020) MATLAB R2020b. Natick, MA

[136] The Numerical Algorithms Group (2021) The NAG Library. Oxford, Mark 27 edn

[137] Tong X (2014) A guaranteed, adaptive, automatic algorithm for univariate function minimization. Master's thesis, Illinois Institute of Technology

[138] Vihola M (2012) Robust adaptive metropolis algorithm with coerced acceptance rate. Statistics and Computing 22(5):997–1008

[139] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat I, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, Vijaykumar A, Bardelli AP, Rothberg A, Hilboll A, Kloeckner A, Scopatz A, Lee A, Rokem A, Woods CN, Fulton C, Masson C, Häggström C, Fitzgerald C, Nicholson DA, Hagen DR, Pasechnik DV, Olivetti E, Martin E, Wieser E, Silva F, Lenders F, Wilhelm F, Young G, Price GA, Ingold GL, Allen GE, Lee GR, Audren H, Probst I, Dietrich JP, Silterra J, Webber JT, Slavič J, Nothman

J, Buchner J, Kulick J, Schönberger JL, de Miranda Cardoso J, Reimer J, Harrington J, Rodríguez JLC, Nunez-Iglesias J, Kuczynski J, Tritz K, Thoma M, Newville M, Kümmerer M, Bolingbroke M, Tartre M, Pak M, Smith NJ, Nowaczyk N, Shebanov N, Pavlyk O, Brodtkorb PA, Lee P, McGibbon RT, Feldbauer R, Lewis S, Tygier S, Sievert S, Vigna S, Peterson S, More S, Pudlik T, Oshima T, Pingel TJ, Robitaille TP, Spura T, Jones TR, Cera T, Leslie T, Zito T, Krauss T, Upadhyay U, Halchenko YO, Vázquez-Baeza Y, Contributors (2020) Scipy 1.0: fundamental algorithms for scientific computing in python. Nature Methods 17(3):261–272

[140] Wahls S, Poor HV (2015) Fast numerical nonlinear Fourier transforms. IEEE Transactions on Information Theory 61(12):6957–6974

[141] Wand MP, Jones MC (1995) Kernel Smoothing. Chapman & Hall, New York

[142] Wang H, Ma Y (2021) Optimal subsampling for quantile regression in big data. Biometrika 108(1):99–112

[143] Wang H, Zhu R, Ma P (2018) Optimal subsampling for large sample logistic regression. Journal of the American Statistical Association 113(522):829–844

[144] Wang H, Xie L, Cuozzo A, Mak S, Xie Y (2020) Uncertainty quantification for inferring Hawkes networks. In: Advances in Neural Information Processing Systems

[145] Wang H, Xie L, Xie Y, Cuozzo A, Mak S (2022) Sequential change-point detection for mutually exciting point processes over networks. Technometrics , to appear

[146] Yeh ST, Wang X, Sung CL, Mak S, Chang YH, Zhang L, Wu CJ, Yang V (2018) Common proper orthogonal decomposition-based spatiotemporal emulator for design exploration. AIAA Journal 56(6):2429–2442

[147] Yuchi HS, Mak S, Xie Y (2022) Bayesian uncertainty quantification for low-rank matrix completion. Bayesian Analysis , to appear

[148] Zhang R, Mak S, Dunson D (2022) Gaussian process subspace prediction for model reduction. SIAM Journal on Scientific Computing 44(3):A1428–A1449

[149] Zhang Y (2018) Guaranteed, adaptive, automatic algorithms for univariate integration: Methods, costs and implementation. PhD thesis, Illinois Institute of Technology

[150] Zhao X (2017) Simulating the Heston model via the QE method with a specified error tolerance. Master's thesis, Illinois Institute of Technology

[151] Zheng X, Mak S, Xie Y (2021) Online high-dimensional change-point detection using topological data analysis. Technometrics , to appear

[152] Zhou X (2015) Function approximation with kernel methods. PhD thesis, Illinois Institute of Technology

[153] Zhou X, Hickernell FJ (2016) Tractability of the radial function approximation problem with kernels of a product form. In: Cools R, Nuyens D (eds) Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Leuven, Belgium, April 2014, Springer-Verlag, Berlin, Springer Proceedings in Mathematics and Statistics, vol 163, pp 583–598, arXiv:1411.0790 [math.NA]

NSF BIOGRAPHICAL SKETCH

NAME: Hickernell, Fred J.

NSF ID: 000421071@nsf.gov
ORCID: 0000-0001-6677-1324

POSITION TITLE & INSTITUTION: Vice Provost for Research, Illinois Institute of Technology

(a) PROFESSIONAL PREPARATION -(see PAPPG Chapter II.C.2.f.(a))

INSTITUTION	LOCATION	MAJOR / AREA OF STUDY	DEGREE (if applicable)	YEAR YYYY
Pomona College	Claremont, CA	Mathematics and Physics	BA	1977
Massachusetts Institute of Technology	Cambridge, MA	Mathematics	PHD	1981

(b) APPOINTMENTS -(see PAPPG Chapter II.C.2.f.(b))

2018 - present Vice Provost for Research, Illinois Institute of Technology, Chicago, IL
2005 - 2022 Professor, Illinois Institute of Technology, Department of Applied Mathematics, Chicago, IL
2005 - 2017 Department Chair, Illinois Institute of Technology, Department of Applied Mathematics, Chicago, IL
1995 - 2005 Associate Professor, Professor, Hong Kong Baptist University, Department of Mathematics, Kowloon
1989 - 2002 Department Head, Hong Kong Baptist College/University, Department of Mathematics, Kowloon
1985 - 1995 Lecturer, Senior Lecturer, Hong Kong Baptist College, Department of Mathematics, Kowloon
1981 - 1985 Assistant Professor, University of Southern California, Mathematics, Los Angeles, CA

(c) PRODUCTS -(see PAPPG Chapter II.C.2.f.(c))

Products Most Closely Related to the Proposed Project

1. Tong X, Choi ST, Ding Y, Hickernell FJ, Jiang L, Jiménez Rugama L, Rathinavel J, Zhang K, Zhang Y, Zhou X. Guaranteed Automatic Integration Library (GAIL): An Open-Source MATLAB Library for Function Approximation, Optimization, and Integration. Journal of Open Software Research. 2022 July 29; 10(1):7. Available from: http://doi.org/10.5334/jors.381 DOI: 10.5334/jors.381
2. Choi ST, Hickernell FJ, Jagadeeswaran R, McCourt MJ, Sorokin AG. Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Oxford, England, August 2020. Keller A, editor. Cham: Springer; 2022. Quasi-Monte Carlo Software; p.23–50. DOI: 0.1007/978-3-030-98319-2_2
3. Jagadeeswaran R, Hickernell F. Fast automatic Bayesian cubature using lattice sampling. Statistics and Computing. 2019 September 10; 29(6):1215-1229. Available from: http://link.springer.com/10.1007/s11222-019-09895-9 DOI: 10.1007/s11222-019-09895-9
4. Jiménez Rugama L, Hickernell F. Adaptive Multidimensional Integration Based on Rank-1 Lattices. Springer Proceedings in Mathematics & Statistics [Internet] Cham: Springer

International Publishing; 2016. Chapter Chapter 20407-422p. Available from: http://link.springer.com/10.1007/978-3-319-33507-0_20 DOI: 10.1007/978-3-319-33507-0_20

5. Hickernell F, Woźniakowski H. Integration and approximation in arbitrary dimensions. Advances in Computational Mathematics. 2000; 12(1):25-58. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-0041638501&partnerID=MN8TOARS DOI: 10.1023/A:1018948631251

Other Significant Products, Whether or Not Related to the Proposed Project

1. Hickernell F. The Trio Identity for Quasi-Monte Carlo Error. Springer Proceedings in Mathematics & Statistics [Internet] Cham: Springer International Publishing; 2018. Chapter Chapter 13-27p. Available from: http://link.springer.com/10.1007/978-3-319-91436-7_1 DOI: 10.1007/978-3-319-91436-7_1
2. Hickernell F, Jiang L, Liu Y, Owen A. Guaranteed Conservative Fixed Width Confidence Intervals via Monte Carlo Sampling. Springer Proceedings in Mathematics & Statistics [Internet] Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. Chapter Chapter 5105-128p. Available from: http://link.springer.com/10.1007/978-3-642-41095-6_5 DOI: 10.1007/978-3-642-41095-6_5
3. Hickernell F, Müller-Gronbach T, Niu B, Ritter K. Multi-level Monte Carlo algorithms for infinite-dimensional integration on R^N. Journal of Complexity. 2010 June; 26(3):229-254. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0885064X10000191 DOI: 10.1016/j.jco.2010.02.002
4. Hickernell F. Uniform designs limit aliasing. Biometrika. 2002 December 01; 89(4):893-904. Available from: https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/89.4.893 DOI: 10.1093/biomet/89.4.893
5. Hickernell F. A generalized discrepancy and quadrature error bound. Mathematics of Computation of the American Mathematical Society. 1998; 67(221):299-322. Available from: http://www.ams.org/jourcgi/jour-getitem?pii=S0025-5718-98-00894-1 DOI: 10.1090/S0025-5718-98-00894-1

(d) SYNERGISTIC ACTIVITIES -(see PAPPG Chapter II.C.2.f.(d))

1. Fellow of the Institute of Mathematical Statistics (elected 2007)
2. Recipient of the 2016 Joseph F. Traub Prize for Achievement in Information-Based Complexity
3. Mentored dozens of high school, BS, MS, MPhil, and PhD students
4. Editorial board member for various academic journals
5. Steering Committee and Program Committee member for the International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing

NSF BIOGRAPHICAL SKETCH

NAME: Ding, Yuhan

NSF ID: 000814004@nsf.gov
ORCID: 0000-0002-7807-1752

POSITION TITLE & INSTITUTION: Associate Teaching Professor, Illinois Institute of Technology

(a) PROFESSIONAL PREPARATION -(see PAPPG Chapter II.C.2.f.(a))

INSTITUTION	LOCATION	MAJOR / AREA OF STUDY	DEGREE (if applicable)	YEAR YYYY
Shanghai University	Shanghai, Shanghai	Mathematics	BS	2006
Shanghai University	Shanghai, Shanghai	Mathematics	MS	2009
Illinois Institute of Technology	Chicago, IL	Applied Math	PHD	2015

(b) APPOINTMENTS -(see PAPPG Chapter II.C.2.f.(b))

2022 - present Associate Teaching Professor, Illinois Institute of Technology, Chicago, IL
2019 - 2022 Senior Lecturer, Illinois Institute of Technology, Chicago, IL
2017 - 2019 Assistant Professor, Misericordia University, Dallas, PA
2016 - 2017 Visiting Assistant Professor, Illinois Institute of Technology, Chicago, IL

(c) PRODUCTS -(see PAPPG Chapter II.C.2.f.(c))

Products Most Closely Related to the Proposed Project

1. Ding Y, Hickernell FJ, Jiménez Rugama L. An adaptive algorithm employing continuous linear functionals. Monte Carlo and Quasi-Monte Carlo Methods. MCQMC 2018. Springer Proceedings in Mathematics & Statistics. 2020 May 02; 324:161-181. Available from: https://doi.org/10.1007/978-3- 030-43465-6_8.
2. Ding Y, Hickernell FJ, Kritzer P, Mak S. Adaptive approximation for multivariate linear problems with inputs lying in a cone. In: Hickernnell FJ., Kritzer P, editors. Multivariate Algorithms and Information-Based Complexity [Internet] Berlin/Boston: DeGruyter; 2020. 109-148p. Available from: https://doi.org/10.1515/9783110635461-007
3. Choi ST., Ding Y, Hikcernell FJ., Tong X. Local Adaption for Approximation and Minimization of Univariate Functions. Journal of complexity. 2017; 40(Special Issue: Dedicated to the memory of Joseph F. Traub, Part II):17-33.
4. Clancy N, Ding Y, Hamilton C, Hickernell FJ., Zhang Y. The Cost of Deterministic, Adaptive, Automatic Algorithms: Cones, Not Balls. Journal of complexity. 2014; 30(1):21-45.
5. Choi ST., Ding Y, Hickernell FJ., Jiang L, Jiménez Rugama L, Tong X, Zhang Y, Zhou X, Rathinavel J, Zhang K. GAIL: Guaranteed Automatic Integration Library. [revised 2021 May]. [Internet]. Version 2.3.2. Chicago, IL: MATLAB; 2013. Available from: http://gailgithub.github.io/GAIL_Dev/

Other Significant Products, Whether or Not Related to the Proposed Project

1. Ding Y, Wang Q. The complex solution to a quaternion matrix equation with application.

Mathematical Sciences Research Journal. 2008; 12(9):215-224.

2. Wang Q, Ding Y. Least square solutions of generalized Hamiltonian quaternion matrices. Advances in Matrix Theory and its Applications. 2008; I:262-265.

(d) SYNERGISTIC ACTIVITIES -(see PAPPG Chapter II.C.2.f.(d))

1. Coordinator, SURE (Summer Undergraduate Research Experience) Program, Illinois Institute of Technology, Chicago, IL, Summer, 2021-2022
2. Program Director, Master of Data Science, Illinois Institute of Technology, Chicago, IL, 2021 Fall - now
3. Supervisor, Virtual Math Tutoring Center, Illinois Institute of Technology, 2020 August - now
4. Organizer of Minisymposium on Numerical Algorithms with Guaranteed Accuracy and Computational Cost Guaranteed Local Adaptive Interpolation, SIAM Annual Meeting, Portland, Oregon, July 9-13, 2018
5. Logistics Coordinator, Spring Research Conference 2016, Illinois Institute of Technology, Chicago, IL, May 25-27, 2016

NSF BIOGRAPHICAL SKETCH

NAME: Choi, Sou Cheng T.

ORCID: 0000-0002-6190-2986

POSITION TITLE & INSTITUTION: Research Associate Professor, Illinois Institute of Technology

(a) PROFESSIONAL PREPARATION -(see PAPPG Chapter II.C.2.f.(a))

INSTITUTION	LOCATION	MAJOR / AREA OF STUDY	DEGREE (if applicable)	YEAR YYYY
National University of Singapore	Singapore	Computational Science, Mathematics	BS	1997
National University of Singapore	Singapore	Statistics and Applied Probability	MS	2000
Stanford University	Stanford, CA	Computational and Mathematical Engineering	PHD	2007

(b) APPOINTMENTS -(see PAPPG Chapter II.C.2.f.(b))

2017 - present Research Associate Professor, Illinois Institute of Technology, Chicago, IL
2022 - present Principal Data Scientist, SAS Institute Inc., Chicago, IL
2020 - 2022 Chief Data Scientist, Kamakura Corporation, Chicago, IL
2018 - 2020 Lead Researcher, Allstate Corporation, Chicago, IL
2016 - 2017 Principal Data Scientist, Allstate Corporation, Chicago, IL
2014 - 2017 Research Assistant Professor, Illinois Institute of Technology, Chicago, IL
2014 - 2016 Senior Statistician, NORC at the University of Chicago, Chicago, IL
2010 - 2013 Research Scientist, University of Chicago/Argonne National Laboratory, Chicago, IL
2007 - 2013 University Affiliate, Stanford University, Stanfford, CA
2007 - 2010 Senior Member of Technical Staff, Oracle Inc., Redwood Shores, CA
1998 - 2000 Financial Software Engineer, Kamakura Corporation, Singapore
1997 - 1998 Systems Analyst, Union Bank of Switzerland, Singapore

(c) PRODUCTS -(see PAPPG Chapter II.C.2.f.(c))

Products Most Closely Related to the Proposed Project

1. Choi ST, Ding Y, Hickernell FJ, Jiang L, Jimenez Rugama L, Li D, Rathinavel J, Tong X, Zhang K, Zhang Y, Zhou X. GAIL: Guaranteed Automatic Integration Library (Versions 1.0--2.3.2). MATLAB Software. 2021 May. Available from: http://gailgithub.github.io/GAIL_Dev/
2. Choi ST, Hickernell FJ, McCourt M, Rathinavel J, Sorokin A. QMCPy: A quasi-Monte Carlo Python Library (Version 1.2). Python Software. 2022 June. Available from: https://qmcsoftware.github.io/QMCSoftware
3. Choi ST, Ding Y, Hickernell FJ, Tong X. Local adaption for approximation and minimization of univariate functions. Journal of Complexity. 2017; 40:17--33. Available from: https://www.sciencedirect.com/science/article/pii/S0885064X16301108
4. Hickernell FJ, Choi ST, Jiang L, Jimenez Rugama L. Monte Carlo simulation, automatic stopping criteria for. Wiley StatsRef: Statistics Reference. 2018. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat08035

5. Choi ST, Paige CC, Saunders MA. MINRES-QLP: A Krylov subspace method for indefinite or singular symmetric systems. SIAM J. Sci. Comput.. 2011; 33(4):1810--1836. Available from: https://epubs.siam.org/doi/10.1137/100787921

Other Significant Products, Whether or Not Related to the Proposed Project

1. Choi ST, Saunders MA. ALGORITHM 937: MINRES-QLP for Singular Symmetric and Hermitian Linear Equations and Least-Squares Problems. ACM TOMS. 2014; 40(2):16:1–16:12. Available from: https://dl.acm.org/doi/10.1145/2527267
2. Katz D, Choi ST, Lapp H, Maheshwari K, Löffler F, Turk M, Hanwell MD, Wilkins-Diehr N, Hetherington J, Howison J, Swenson S, Allen GD, Elster AC, Berriman B, Venters C. Summary of the First Workshop on Sustainable Software for Science: Practice and Experiences. Journal of Open Research Software. 2014; 2(1):e6. Available from: https://openresearchsoftware.metajnl.com/articles/10.5334/jors.an/
3. Wulfe B, Chintakindi S, Choi ST, Hartong-Redden R, Kodali A, Kochenderfer MJ. Real-Time Prediction of Intermediate-Horizon Automotive Collision Risk. 17th International Conference on Autonomous Agents and Multiagent Systems. 2018. Available from: https://dl.acm.org/doi/10.5555/3237383.3237858
4. Choi ST, Lin Y, Mulrow E. Comparison of Public-Domain Software and Services for Probabilistic Record Linkage and Address Standardization. Towards Integrative Machine Learning and Knowledge Extraction. 2015; :51-66. Available from: https://link.springer.com/chapter/10.1007/978-3-319-69775-8_3
5. Donoho DL, Flesia A, Huo X, Levi O, Choi ST, Shi D. BEAMLAB (Version 200). MATLAB Software. 2003. Available from: http://www-stat.stanford.edu/~beamlab/

(d) SYNERGISTIC ACTIVITIES -(see PAPPG Chapter II.C.2.f.(d))

1. Since 2013, served as a mentor to 12 graduate (8 doctoral and 4 master’s) students who are Applied Mathematics majors at IIT. Served as thesis committee members for two of the students. Seven of the PhD graduates are now working in either academia or industries. We continue to collaborate with our current students and alumni working in industries and academia.
2. Taught six research seminar courses at IIT and the University of Chicago (UC) between 2013 and 2020. Each course has between one to eight students at undergraduate or graduate levels from applied mathematics or computer science. We explored, for instance, modern machine learning methods for problems and big data sets stemming from computational finance or social sciences.
3. Gave over 80 scientific talks locally and internationally in the past ten years. Three of them are plenary talks at international conferences.
4. Co-organized multiple international conferences and (co-)hosted at least ten mini-symposiums in major conferences organized by the Society of Industrial and Applied Mathematics (SIAM), the American Mathematical Society (AMS), and the International Linear Algebra Society (ILAS).
5. Co-recipient of the SIAM (Society for Industrial and Applied Mathematics) Activity Group on Linear Algebra (SIAG/LA) Prize. International award for the best peer-reviewed journal paper from 2009 to 2011 with significant research contributions to the field of linear algebra, and with direct or potential applications

Effective 10/04/2021 **NSF BIOGRAPHICAL SKETCH** **OMB-3145-0058**

NAME: Simon Mak

POSITION TITLE & INSTITUTION: Assistant Professor, Duke University

A. PROFESSIONAL PREPARATION - (see PAPPG Chapter II.C.2.f.(i)(a))

INSTITUTION	LOCATION	MAJOR/AREA OF STUDY	DEGREE (if applicable)	YEAR (YYYY)
Simon Fraser University	Burnaby, BC, Canada	Statistics and Actuarial Science	BS	2013
Georgia Institute of Technology	Atlanta, GA	Statistics	MS	2018
Georgia Institute of Technology	Atlanta, GA	Industrial Engineering	PhD	2018
Georgia Institute of Technology	Atlanta, GA	Industrial Engineering	Postdoc	2018-2019

B. APPOINTMENTS - (see PAPPG Chapter II.C.2.f.(i)(b))

From - To	Position Title, Organization and Location
2019 - present	Assistant Professor, Department of Statistical Science, Duke University
2018 - 2019	Postdoctoral Fellow, School of Industrial and Systems Engineering, Georgia Institute of Technology

C. PRODUCTS - (see PAPPG Chapter II.C.2.f.(i)(c)) Products Most Closely Related to the Proposed Project

Zhang, R., Mak, S. and Dunson, D. (2022). Gaussian process subspace prediction for model reduction. SIAM Journal on Scientific Computing, 44(3), A1428-A1449.

Chen, J., Mak, S., Joseph, V. R. and Zhang, C. (2022). Adaptive design for Gaussian process regression under censoring. Annals of Applied Statistics, 16(2), 744-764

Huang, C., Joseph, V. R. and Mak, S. (2023+). Population Quasi-Monte Carlo. Journal of Computational and Graphical Statistics, to appear.

Mak, S. and Joseph V. R. (2018). Support points. Annals of Statistics, 46(64A):2562-2592.

Mak, S. Sung, C.-L., Wang, X. J., Yeh, S.-T., Chang, Y.-H., Joseph, V. R., Yang, V. and Wu, C. F. J. (2018). An efficient surrogate model for emulation and physics extraction of large eddy simulations. Journal of the American Statistical Association, 113(524):1443-1456.

Other Significant Products, Whether or Not Related to the Proposed Project

Liyanage, D., Ji, Y., Everett, D., Heffernan, M., Heinz, U., Mak, S., Paquet J.F. (2022). Efficient emulation of relativistic heavy ion collisions with transfer learning. Physical Review C 105 (3), 034910.

Yuchi, H.S., Mak, S., Xie, Y. (2022). Bayesian uncertainty quantification for low-rank matrix completion. Bayesian Analysis, to appear.

Zheng, X., Mak, S., Xie, L. and Xie, Y. (2022). PERCEPT: a new online change-point detection method using topological data analysis. Technometrics, to appear.

Mak, S., Zhou, Y., Hoang, L. and Wu, C.F.J. (2022). TSEC: a framework for online experimentation under experimental constraints. Technometrics, to appear.

Narayanan, S. R., Ji, Y.&, Sapra, H.D., Hessel, R., Yang, S., Mak, S., Sun, Z., Kokjohn, S., Kim, K., Kweon, C. B. (2023). Segmented Gaussian process learning for cost-efficient control system training of diesel engines for low cetane numbers. In Proceedings of the 2023 AIAA Science and Technology (SciTech) Forum and Exposition.

D. SYNERGISTIC ACTIVITIES - (see PAPPG Chapter II.C.2.f.(i)(d))

- Associate Editor, Technometrics (2019 - current)
- Associate Editor, Data Science in Science (2021 - current)
- Blackwell-Rosenbluth Award (2022), awarded annually by the International Society of Bayesian Analysis (ISBA), for exceptional career achievements and contributions to the community by junior Bayesian researchers.
- Statistics in Physical Engineering Sciences (SPES) award recipient, awarded annually by the American Statistical Association (ASA), for an innovative paper which uses statistics to solve a high-impact problem in the physical and engineering sciences (2019).
- Working Group Leader, Program on Quasi-Monte Carlo and High-Dimensional Sampling Methods for Applied Mathematics at the Statistical and Applied Mathematics Sciences Institute (SAMSI) (2017-2018)

Other Personnel Biographical Information

Data Not Available

Other Personnel Biographical Information

Data Not Available

SUMMARY PROPOSAL BUDGET

YEAR 1

ORGANIZATION Illinois Institute of Technology			FOR NSF USE ONLY PROPOSAL NO. 2316011	DURATION (months) Proposed	Granted
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Fred Hickernell			AWARD NO.		

A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)	NSF Funded Person-months CAL	ACAD	SUMR	Funds Requested By proposer	Funds granted by NSF (if different)
1. Fred Hickernell - Principal Inv	1.0			24,197	
2. Yuhan Ding	1.0			6,916	
3.					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.0			0	
7. (2) TOTAL SENIOR PERSONNEL (1 - 6)	2.0			31,113	
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)					
1. (0) POST DOCTORAL SCHOLARS	0.0			0	
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.0			0	
3. (1) GRADUATE STUDENTS				25,000	
4. (2) UNDERGRADUATE STUDENTS				12,000	
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0	
6. (0) OTHER				0	
TOTAL SALARIES AND WAGES (A + B)				68,113	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				2,645	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				70,758	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT				0	
E. TRAVEL 1. DOMESTIC (INCL. U.S. POSSESSIONS)				12,000	
2. INTERNATIONAL				6,000	
F. PARTICIPANT SUPPORT COSTS 1. STIPENDS $ 0 2. TRAVEL 0 3. SUBSISTENCE 0 4. OTHER 0					
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS				0	
G. OTHER DIRECT COSTS					
1. MATERIALS AND SUPPLIES				2,000	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				1,000	
3. CONSULTANT SERVICES				0	
4. COMPUTER SERVICES				0	
5. SUBAWARDS				0	
6. OTHER				14,814	
TOTAL OTHER DIRECT COSTS				17,814	
H. TOTAL DIRECT COSTS (A THROUGH G)				106,572	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 54.0, Base:91758.0)					
TOTAL INDIRECT COSTS (F&A)				49,549	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				156,121	
K. FEE				0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				156,121	
M. COST SHARING PROPOSED LEVEL $ 0	AGREED LEVEL IF DIFFERENT $				

PI/PD NAME Fred Hickernell	FOR NSF USE ONLY INDIRECT COST RATE VERIFICATION		
ORG. REP. NAME* Robert Lapointe	Date Checked	Date Of Rate Sheet	Initials - ORG

*ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET

YEAR 2

ORGANIZATION Illinois Institute of Technology	FOR NSF USE ONLY PROPOSAL NO. 2316011	DURATION (months) Proposed	Granted
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Fred Hickernell	AWARD NO.		

A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)	NSF Funded Person-months CAL	ACAD	SUMR	Funds Requested By proposer	Funds granted by NSF (if different)
1. Fred Hickernell - Principal Inv	1.0			25,165	
2. Yuhan Ding	1.0			7,192	
3.					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.0			0	
7. (2) TOTAL SENIOR PERSONNEL (1 - 6)	2.0			32,357	
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)					
1. (0) POST DOCTORAL SCHOLARS	0.0			0	
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.0			0	
3. (1) GRADUATE STUDENTS				26,000	
4. (2) UNDERGRADUATE STUDENTS				12,480	
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0	
6. (0) OTHER				0	
TOTAL SALARIES AND WAGES (A + B)				70,837	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				2,750	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				73,587	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT				0	
E. TRAVEL 1. DOMESTIC (INCL. U.S. POSSESSIONS)				12,480	
2. INTERNATIONAL				6,240	
F. PARTICIPANT SUPPORT COSTS 1. STIPENDS $ 0 2. TRAVEL 0 3. SUBSISTENCE 0 4. OTHER 0					
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS				0	
G. OTHER DIRECT COSTS					
1. MATERIALS AND SUPPLIES				2,080	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				1,040	
3. CONSULTANT SERVICES				0	
4. COMPUTER SERVICES				0	
5. SUBAWARDS				0	
6. OTHER				15,407	
TOTAL OTHER DIRECT COSTS				18,527	
H. TOTAL DIRECT COSTS (A THROUGH G)				110,834	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 54.0, Base:95427.0)					
TOTAL INDIRECT COSTS (F&A)				51,531	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				162,365	
K. FEE				0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				162,365	
M. COST SHARING PROPOSED LEVEL $ 0	AGREED LEVEL IF DIFFERENT $				

PI/PD NAME Fred Hickernell	FOR NSF USE ONLY INDIRECT COST RATE VERIFICATION		
ORG. REP. NAME* Robert Lapointe	Date Checked	Date Of Rate Sheet	Initials - ORG

***ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET**

SUMMARY PROPOSAL BUDGET

YEAR 3

ORGANIZATION Illinois Institute of Technology			FOR NSF USE ONLY PROPOSAL NO. 2316011	DURATION (months) Proposed	Granted
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Fred Hickernell			AWARD NO.		

A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)	NSF Funded Person-months CAL	ACAD	SUMR	Funds Requested By proposer	Funds granted by NSF (if different)
1. Fred Hickernell - Principal Inv	1.0			26,172	
2. Yuhan Ding	1.0			7,480	
3.					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.0			0	
7. (2) TOTAL SENIOR PERSONNEL (1 - 6)	2.0			33,652	
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)					
1. (0) POST DOCTORAL SCHOLARS	0.0			0	
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.0			0	
3. (1) GRADUATE STUDENTS				27,040	
4. (2) UNDERGRADUATE STUDENTS				12,980	
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0	
6. (0) OTHER				0	
TOTAL SALARIES AND WAGES (A + B)				73,672	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				2,861	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				76,533	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT				0	
E. TRAVEL 1. DOMESTIC (INCL. U.S. POSSESSIONS)				12,979	
2. INTERNATIONAL				6,490	
F. PARTICIPANT SUPPORT COSTS 1. STIPENDS $ 0 2. TRAVEL 0 3. SUBSISTENCE 0 4. OTHER 0					
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS				0	
G. OTHER DIRECT COSTS					
1. MATERIALS AND SUPPLIES				2,163	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				1,082	
3. CONSULTANT SERVICES				0	
4. COMPUTER SERVICES				0	
5. SUBAWARDS				0	
6. OTHER				16,023	
TOTAL OTHER DIRECT COSTS				19,268	
H. TOTAL DIRECT COSTS (A THROUGH G)				115,270	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 54.0, Base:99247.0)					
TOTAL INDIRECT COSTS (F&A)				53,593	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				168,863	
K. FEE				0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				168,863	
M. COST SHARING PROPOSED LEVEL $ 0	AGREED LEVEL IF DIFFERENT $				

PI/PD NAME Fred Hickernell	FOR NSF USE ONLY INDIRECT COST RATE VERIFICATION		
ORG. REP. NAME* Robert Lapointe	Date Checked	Date Of Rate Sheet	Initials - ORG

***ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET**

SUMMARY PROPOSAL BUDGET

Cumulative

ORGANIZATION	FOR NSF USE ONLY		
Illinois Institute of Technology	PROPOSAL NO. 2316011	DURATION (months) Proposed	Granted
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR: Fred Hickernell	AWARD NO.		

A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)	NSF Funded Person-months CAL	ACAD	SUMR	Funds Requested By proposer	Funds granted by NSF (if different)
1. Fred Hickernell - Principal Inv	3.0			75,534	
2. Yuhan Ding	3.0			21,588	
3.					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)					
7. (2) TOTAL SENIOR PERSONNEL (1 - 6)	6.0			97,122	
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)					
1. (0) POST DOCTORAL SCHOLARS	0.0			0	
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.0			0	
3. (3) GRADUATE STUDENTS				78,040	
4. (6) UNDERGRADUATE STUDENTS				37,460	
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0	
6. (0) OTHER				0	
TOTAL SALARIES AND WAGES (A + B)				212,622	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				8,256	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				220,878	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT				0	
E. TRAVEL 1. DOMESTIC (INCL. U.S. POSSESSIONS)				37,459	
2. INTERNATIONAL				18,730	
F. PARTICIPANT SUPPORT COSTS					
1. STIPENDS $ 0					
2. TRAVEL 0					
3. SUBSISTENCE 0					
4. OTHER 0					
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS				0	
G. OTHER DIRECT COSTS					
1. MATERIALS AND SUPPLIES				6,243	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				3,122	
3. CONSULTANT SERVICES				0	
4. COMPUTER SERVICES				0	
5. SUBAWARDS				0	
6. OTHER				46,244	
TOTAL OTHER DIRECT COSTS				55,609	
H. TOTAL DIRECT COSTS (A THROUGH G)				332,676	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)					
TOTAL INDIRECT COSTS (F&A)				154,673	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				487,349	
K. FEE				0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				487,349	

M. COST SHARING PROPOSED LEVEL $ 0 AGREED LEVEL IF DIFFERENT $

PI/PD NAME	FOR NSF USE ONLY		
Fred Hickernell	INDIRECT COST RATE VERIFICATION		
ORG. REP. NAME*	Date Checked	Date Of Rate Sheet	Initials - ORG
Robert Lapointe			

***ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET**

BUDGET JUSTIFICATION
Illinois Institute of Technology

Senior Personnel

Prof. Fred J. Hickernell, Professor of Applied Mathematics at Illinois Tech, will provide overall leadership for this project, and mentor the graduate and undergraduate student research assistants. He will contribute expertise in QMC methodology, especially error analysis and the theory underlying stopping criteria. The one-month summer salary compensates his time on the project.

Dr. Yuhan Ding, Senior Lecturer of Applied Mathematics at Illinois Tech, will co-lead this project, mentoring the students and carrying out the theoretical and methodological development. Dr. Ding has co-authored several articles on adaptive algorithms. The one-month summer salary compensates her time on the project.

Note: For purposes of NSF PAPPG section II.C.2.g(i)(a), the term "year" at Illinois Institute of Technology refers to IIT's fiscal year (June 1 – May 31).

Other Personnel

The graduate tuition scholarships and stipends will support PhD student(s) engaged in building out QMCPy as explained in the project. This includes ensuring that new contributions by themselves or others adhere to the QMCPy architecture, testing, and documentation requirements. The PhD students will also help develop some of the theoretical and methodological underpinnings of the new algorithms to be included in QMCPy. Aleksei Sorokin, a new, domestic PhD student at Illinois Tech and developer of QMCPy will be supported by this grant.

The summer undergraduate student stipends will fund smaller scale, but crucial components of QMCPy. These include, for example, novel use cases found in the literature and code essentially built by others but needing to be adapted to the QMCPy architecture.

Fringe Benefits
IIT's federally negotiated fringe benefit rates are: faculty academic salary, 20.9%; faculty summer salary, 8.5%; staff salary, 24.9%; and student stipends, 0.0%.

Travel
The senior personnel and research students will disseminate their results and introduce a broader audience to QMCPy through attendance at US and international conferences devoted to QMC and its applications.

Other Direct Costs

<u>Materials and Supplies</u>
Modest resources are needed for software license and website/blog maintenance fees.

<u>Publications</u>
Modest resources are required for making our publications open access.

<u>Tuition</u>
The PhD student(s) will be supported at 9 credits/yr. so that they may continue their studies while working on this grant.

Indirect Costs

IIT's current federally negotiated indirect cost rate (agreement date 02/18/2022) is 54% of modified total direct costs (MTDC). MTDC include all salaries and wages, fringe benefits, materials, supplies, services, travel and up to the first $25,000 of each subaward. MTDC excludes equipment, participant support, capital expenditures, student tuition, rental costs of off-site facilities, as well as the portion of each subaward in excess of $25,000.

	Y1	Y2	Y3	Total
Direct Costs	$106,572	$110,834	$115,270	$332,676
Indirect Costs	$49,549	$51,531	$53,593	$154,673
Total Costs	$156,121	$162,365	$168,863	$487,349
Modified Base	*$91,758*	*$95,427*	*$99,247*	*$286,432*

An inflationary rate of 4% is used for all categories for all years of the project.

SUMMARY PROPOSAL BUDGET

YEAR 1

ORGANIZATION Duke University		FOR NSF USE ONLY			
		PROPOSAL NO. 2316012	DURATION (months)		
			Proposed	Granted	
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Simon Mak		AWARD NO.			

A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)	NSF Funded Person-months CAL	ACAD	SUMR	Funds Requested By proposer	Funds granted by NSF (if different)
1. Simon Mak - Principal Inv	0.5			7,872	
2.					
3.					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.0			0	
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	0.5			7,872	
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)					
1. (0) POST DOCTORAL SCHOLARS	0.0			0	
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.0			0	
3. (1) GRADUATE STUDENTS				21,346	
4. (0) UNDERGRADUATE STUDENTS				0	
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0	
6. (0) OTHER				0	
TOTAL SALARIES AND WAGES (A + B)				29,218	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				4,592	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				33,810	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT				0	
E. TRAVEL 1. DOMESTIC (INCL. U.S. POSSESSIONS)				5,000	
2. INTERNATIONAL				0	
F. PARTICIPANT SUPPORT COSTS 1. STIPENDS $ 0 2. TRAVEL 0 3. SUBSISTENCE 0 4. OTHER 0					
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS				0	
G. OTHER DIRECT COSTS					
1. MATERIALS AND SUPPLIES				0	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				0	
3. CONSULTANT SERVICES				0	
4. COMPUTER SERVICES				0	
5. SUBAWARDS				0	
6. OTHER				7,983	
TOTAL OTHER DIRECT COSTS				7,983	
H. TOTAL DIRECT COSTS (A THROUGH G)				46,793	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 61.0, Base:38810.0)					
TOTAL INDIRECT COSTS (F&A)				23,674	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				70,467	
K. FEE				0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				70,467	
M. COST SHARING PROPOSED LEVEL $ 0	AGREED LEVEL IF DIFFERENT $				

PI/PD NAME Simon Mak	FOR NSF USE ONLY		
	INDIRECT COST RATE VERIFICATION		
ORG. REP. NAME* Lauren Faber	Date Checked	Date Of Rate Sheet	Initials - ORG

*ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET

YEAR 2

ORGANIZATION Duke University					FOR NSF USE ONLY
				PROPOSAL NO. 2316012	DURATION (months) Proposed / Granted
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Simon Mak				AWARD NO.	

A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)	NSF Funded Person-months CAL	ACAD	SUMR	Funds Requested By proposer	Funds granted by NSF (if different)
1. Simon Mak - Principal Inv	0.5			8,108	
2.					
3.					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.0			0	
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	0.5			8,108	
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)					
1. (0) POST DOCTORAL SCHOLARS	0.0			0	
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.0			0	
3. (1) GRADUATE STUDENTS				21,773	
4. (0) UNDERGRADUATE STUDENTS				0	
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0	
6. (0) OTHER				0	
TOTAL SALARIES AND WAGES (A + B)				29,881	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				4,820	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				34,701	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT				0	
E. TRAVEL 1. DOMESTIC (INCL. U.S. POSSESSIONS)				5,000	
2. INTERNATIONAL				0	
F. PARTICIPANT SUPPORT COSTS					
1. STIPENDS $ 0					
2. TRAVEL 0					
3. SUBSISTENCE 0					
4. OTHER 0					
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS				0	
G. OTHER DIRECT COSTS					
1. MATERIALS AND SUPPLIES				0	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				0	
3. CONSULTANT SERVICES				0	
4. COMPUTER SERVICES				0	
5. SUBAWARDS				0	
6. OTHER				7,533	
TOTAL OTHER DIRECT COSTS				7,533	
H. TOTAL DIRECT COSTS (A THROUGH G)				47,234	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 61.0, Base:39701.0)					
TOTAL INDIRECT COSTS (F&A)				24,218	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				71,452	
K. FEE				0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				71,452	
M. COST SHARING PROPOSED LEVEL $ 0	AGREED LEVEL IF DIFFERENT $				

PI/PD NAME Simon Mak	FOR NSF USE ONLY		
	INDIRECT COST RATE VERIFICATION		
ORG. REP. NAME* Lauren Faber	Date Checked	Date Of Rate Sheet	Initials - ORG

*ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET

YEAR 3

ORGANIZATION Duke University				FOR NSF USE ONLY PROPOSAL NO. 2316012	DURATION (months) Proposed	Granted
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Simon Mak				AWARD NO.		

A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)	NSF Funded Person-months CAL	ACAD	SUMR	Funds Requested By proposer	Funds granted by NSF (if different)
1. Simon Mak - Principal Inv	0.5			8,351	
2.					
3.					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.0			0	
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	0.5			8,351	
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)					
1. (0) POST DOCTORAL SCHOLARS	0.0			0	
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.0			0	
3. (1) GRADUATE STUDENTS				22,209	
4. (0) UNDERGRADUATE STUDENTS				0	
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0	
6. (0) OTHER				0	
TOTAL SALARIES AND WAGES (A + B)				30,560	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				4,935	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				35,495	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT				0	
E. TRAVEL 1. DOMESTIC (INCL. U.S. POSSESSIONS)				5,000	
2. INTERNATIONAL				0	
F. PARTICIPANT SUPPORT COSTS 1. STIPENDS $ 0 2. TRAVEL 0 3. SUBSISTENCE 0 4. OTHER 0					
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS				0	
G. OTHER DIRECT COSTS					
1. MATERIALS AND SUPPLIES				0	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				0	
3. CONSULTANT SERVICES				0	
4. COMPUTER SERVICES				0	
5. SUBAWARDS				0	
6. OTHER				7,577	
TOTAL OTHER DIRECT COSTS				7,577	
H. TOTAL DIRECT COSTS (A THROUGH G)				48,072	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 61.0, Base:40495.0)					
TOTAL INDIRECT COSTS (F&A)				24,702	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				72,774	
K. FEE				0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				72,774	
M. COST SHARING PROPOSED LEVEL $ 0	AGREED LEVEL IF DIFFERENT $				

PI/PD NAME Simon Mak	FOR NSF USE ONLY INDIRECT COST RATE VERIFICATION		
ORG. REP. NAME* Lauren Faber	Date Checked	Date Of Rate Sheet	Initials - ORG

*ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET

Cumulative

ORGANIZATION Duke University				FOR NSF USE ONLY PROPOSAL NO. 2316012	DURATION (months) Proposed	Granted
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR Simon Mak				AWARD NO.		
A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)	NSF Funded Person-months CAL	ACAD	SUMR	Funds Requested By proposer	Funds granted by NSF (if different)	
1. Simon Mak - Principal Inv	1.5			24,331		
2.						
3.						
4.						
5.						
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)						
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	1.5			24,331		
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)						
1. (0) POST DOCTORAL SCHOLARS	0.0			0		
2. (0) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.0			0		
3. (3) GRADUATE STUDENTS				65,328		
4. (0) UNDERGRADUATE STUDENTS				0		
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0		
6. (0) OTHER				0		
TOTAL SALARIES AND WAGES (A + B)				89,659		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				14,347		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				104,006		
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)						
TOTAL EQUIPMENT				0		
E. TRAVEL 1. DOMESTIC (INCL. U.S. POSSESSIONS)				15,000		
2. INTERNATIONAL				0		
F. PARTICIPANT SUPPORT COSTS 1. STIPENDS $ 0 2. TRAVEL 0 3. SUBSISTENCE 0 4. OTHER 0						
TOTAL NUMBER OF PARTICIPANTS (0) TOTAL PARTICIPANT COSTS				0		
G. OTHER DIRECT COSTS						
1. MATERIALS AND SUPPLIES				0		
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				0		
3. CONSULTANT SERVICES				0		
4. COMPUTER SERVICES				0		
5. SUBAWARDS				0		
6. OTHER				23,093		
TOTAL OTHER DIRECT COSTS				23,093		
H. TOTAL DIRECT COSTS (A THROUGH G)				142,099		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)						
TOTAL INDIRECT COSTS (F&A)				72,594		
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				214,693		
K. FEE				0		
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				214,693		
M. COST SHARING PROPOSED LEVEL $ 0	AGREED LEVEL IF DIFFERENT $					
PI/PD NAME Simon Mak				FOR NSF USE ONLY INDIRECT COST RATE VERIFICATION		
ORG. REP. NAME* Lauren Faber				Date Checked	Date Of Rate Sheet	Initials - ORG

*ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

BUDGET JUSTIFICATION

Key Personnel

Simon Mak, PhD, Principal Investigator, Duke. (.50 summer month each year). Dr. Mak will supervise all aspects of this project. This will include directing the work of the graduate student, holding regular meetings with the project team, interacting with the funding agency, and interfacing with the team from Illinois Institute of Technology.

Other Personnel

TBD, PhD student (4.50 academic months and 3 summer months – all years). The graduate student will be working under the direction of Professor Mak, while closely interacting with the project team. This work will consist of carefully checking quality of the available data, searching for new sources of relevant data, implementing the code in various context, comparing approaches with competitors, and modifying methods as appropriate to reflect additional information available from models.

Duke University Fringe Benefits

Fringe benefits are assessed at Duke University's projected rates for fiscal year 2024 (July 2023 – June 2024) and projected rates for subsequent years. Rates are pro-rated for budget years crossing different fiscal years.

		Faculty	Grad Students
FY 2024 (07/01/2023 – 06/30/2024)	approved	23.9%	12.7%
FY 2025 (07/01/2024 – 06/30/2025)	projected	24.0%	13.2%
FY 2026 (07/01/2025 – 06/30/2026)	projected	24.0%	13.2%

Tuition Remission

Duke University employs an Average Rate Basis method for tuition recovery from sponsored research. This rate is applied consistently to any PhD student salary charged directly to a sponsored project.

	Average Rate
Academic Year 2023-2024	32.9%
Academic Year 2024-2025	33.6%
Academic Year 2025-2026	34.3%

Domestic Travel

Funds for domestic travel are requested in the amount of $5,000 per year. These funds would be used to cover travel expenses for the PI and graduate student to attend relevant conference to present research and to travel to the collaborative university for research group meetings.

Facilities & Administrative (Indirect) Costs

The DHHS federally negotiated Facilities and Administrative (F&A) cost rate is used. Indirect costs for an on-campus research project are charged at Duke University's negotiated rate of 61% of modified total direct costs (MTDC), equal to total direct costs minus student tuition remission, and subaward costs above the first $25,000 of each individual subaward.

Effective 10/04/2021 NSF CURRENT AND PENDING SUPPORT OMB-3145-0058

PI/co-PI/Senior Personnel: Hickernell, Fred J.

PROJECT/PROPOSAL PENDING SUPPORT

1. Project/Proposal Title: Collaborative Research: Cost-Efficient Simulation via Quasi Monte Carlo for Scalable Scientific and Big Data Computing (This Proposal)

   Proposal/Award Number (if available):

   Source of Support: NSF - National Science Foundation

   Primary Place of Performance: Illinois Institute of Technology

   Project/Proposal Support Start Date (if available): 07/2023

   Project/Proposal Support End Date (if available): 06/2026

   Total Award Amount (including Indirect Costs): $487,349

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2024	1
2025	1
2026	1

   Overall Objectives: 1) To address frontier theoretical and implementation issues in quasi-Monte Carlo methods. 2) To educate the community in best practices in speedier and more accurate Monte Carlo simulation.

   Statement of Potential Overlap: The work done in this project will inform the projects in the proposed SURE program. Students supported by this award may work alongside SURE students.

2. Project/Proposal Title: REU Site: Summer Undergraduate Research Experience (SURE) at Illinois Tech

   Proposal/Award Number (if available):

   Source of Support: NSF - National Science Foundation

   Primary Place of Performance: Illinois Institute of Technology

   Project/Proposal Support Start Date (if available): 05/2023

Project/Proposal Support End Date (if available): 04/2026

Total Award Amount (including Indirect Costs): $404,893

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	0.01
2024	0.01
2025	0.01

Overall Objectives: We believe the proposed SURE program will fill a gap in training the research capabilities and enhance the positive impact of Illinois Tech on its neighboring community. We will attract more underrepresented undergraduate students and establish their interest in research in interdisciplinary areas of mathematics and data science. The research in the SURE program will result in publications in open-access, peer-reviewed journals in applied mathematics, statistics, and machine learning.

Statement of Potential Overlap: N/A

3. Project/Proposal Title: RTG: Research and Training in Complex Dynamical Systems

Proposal/Award Number (if available):

Source of Support: NSF - National Science Foundation

Primary Place of Performance: Illinois Institute of Technology

Project/Proposal Support Start Date (if available): 01/2023

Project/Proposal Support End Date (if available): 12/2027

Total Award Amount (including Indirect Costs): $2,391,869

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	0.25
2024	0.23
2025	0.23
2026	0.22
2027	0.25

Overall Objectives: The goal is to create a research and training group to nurture U.S. applied mathematicians in complex dynamical systems. This will be accomplished via an interdisciplinary platform that includes research theme teams, innovative courses, summer programs, group discussions, seminars, and workshops.

Statement of Potential Overlap: No overlap.

Effective 10/04/2021 NSF CURRENT AND PENDING SUPPORT OMB-3145-0058

PI/co-PI/Senior Personnel: Ding, Yuhan

PROJECT/PROPOSAL PENDING SUPPORT

1. Project/Proposal Title: REU Site: CAREM Academic Year REU Project Description

   Proposal/Award Number (if available): 76554

   Source of Support: NSF - National Science Foundation

   Primary Place of Performance: Illinois Institute of Technology

   Project/Proposal Support Start Date (if available): 09/2023

   Project/Proposal Support End Date (if available): 05/2026

   Total Award Amount (including Indirect Costs): $255,041

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	0.25
2024	0.5
2025	0.5
2026	0.25

   Overall Objectives: Provide an academic year research experience to students who have not had a research experience before.

   Statement of Potential Overlap: N/A

2. Project/Proposal Title: Collaborative Research: Cost-Efficient Simulation via Quasi Monte Carlo for Scalable Scientific and Big Data Computing (This Proposal)

   Proposal/Award Number (if available):

   Source of Support: NSF - National Science Foundation

   Primary Place of Performance: Illinois Institute of Technology

   Project/Proposal Support Start Date (if available): 07/2023

   Project/Proposal Support End Date (if available): 06/2026

   Total Award Amount (including Indirect Costs): $487,349

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2024	1
2025	1
2026	1

Overall Objectives: 1) To address frontier theoretical and implementation issues in quasi-Monte Carlo methods. 2) To educate the community in best practices in speedier and more accurate Monte Carlo simulation.

Statement of Potential Overlap: The work done in this project will inform the projects in the proposed SURE program. Students supported by this award may work alongside SURE students.

3. Project/Proposal Title: REU Site: Summer Undergraduate Research Experience (SURE) at Illinois Tech

Proposal/Award Number (if available):

Source of Support: NSF - National Science Foundation

Primary Place of Performance: Illinois Institute of Technology

Project/Proposal Support Start Date (if available): 05/2023

Project/Proposal Support End Date (if available): 04/2026

Total Award Amount (including Indirect Costs): $404,893

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	0.5
2024	0.5
2025	0.5

Overall Objectives: SURE is a ten-week program that engages ten undergraduate students---especially those from underrepresented groups and academic institutions where research opportunities are limited---in innovative computational mathematics and machine learning research. SURE students work in groups of three or fewer based on the students' preference, under the supervision of faculty mentors and with the assistance of graduate students. They will learn to work independently and creatively.

Statement of Potential Overlap: N/A

PI/co-PI/Senior Personnel: Choi, Sou-Cheng

PROJECT/PROPOSAL PENDING SUPPORT

1. Project/Proposal Title: Collaborative Research: Cost-Efficient Simulation via Quasi Monte Carlo for Scalable Scientific and Big Data Computing (This Proposal)

   Proposal/Award Number (if available):

   Source of Support: NSF - National Science Foundation

   Primary Place of Performance: Illinois Institute of Technology

   Project/Proposal Support Start Date (if available): 07/2023

   Project/Proposal Support End Date (if available): 06/2026

   Total Award Amount (including Indirect Costs): $487,349

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2024	0.1
2025	0.1
2026	0.1

   Overall Objectives: 1) To address frontier theoretical and implementation issues in quasi-Monte Carlo methods. 2) To educate the community in best practices in speedier and more accurate Monte Carlo simulation.

   Statement of Potential Overlap: The work done in this project will inform the projects in the proposed SURE program. Students supported by this award may work alongside SURE students.

2. Project/Proposal Title: REU Site: Summer Undergraduate Research Experience (SURE) at Illinois Tech

   Proposal/Award Number (if available):

   Source of Support: NSF - National Science Foundation

   Primary Place of Performance: Illinois Institute of Technology

   Project/Proposal Support Start Date (if available): 05/2023

Project/Proposal Support End Date (if available): 04/2026

Total Award Amount (including Indirect Costs): $404,893

Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	0.5
2024	0.5
2025	0.5

Overall Objectives: SURE is a ten-week program that engages ten undergraduate students---especially those from underrepresented groups and academic institutions where research opportunities are limited---in innovative computational mathematics and machine learning research. SURE students work in groups of three or fewer based on the students' preference, under the supervision of faculty mentors and with the assistance of graduate students. They will learn to work independently and creatively.

Statement of Potential Overlap: N/A

Effective 10/04/2021 **NSF CURRENT AND PENDING SUPPORT** **OMB-3145-0058**

*PI/co-PI/Senior Personnel Name: Mak, Simon

***Required fields**

Note: NSF has provided 15 project/proposal and 10 in-kind contribution entries for users to populate. Please leave any unused entries blank.

Project/Proposal Section:

Current and Pending Support includes all resources made available to an individual in support of and/or related to all of his/her research efforts, regardless of whether or not they have monetary value.[1] Information must be provided about all current and pending support, including this project, for ongoing projects, and for any proposals currently under consideration from whatever source, irrespective of whether such support is provided through the proposing organization or is provided directly to the individual. This includes, for example, Federal, State, local, foreign, public or private foundations, non-profit organizations, industrial or other commercial organizations, or internal funds allocated toward specific projects. Concurrent submission of a proposal to other organizations will not prejudice its review by NSF, if disclosed.[2]

[1] If the time commitment or dollar value is not readily ascertainable, reasonable estimates should be provided.
[2] The Biological Sciences Directorate exception to this policy is delineated in PAPPG Chapter II.D.2.

Projects/Proposals

1. *Project/Proposal Title : The X-SCAPE Collaboration: The X+ion Collision with a Statistically and Computationally Advanced Program Envelope Collaboration (co-PI) (PI: Steffen A. Bass)

*Status of Support : ◉ Current ○ Pending ○ Submission Planned ○ Transfer of Support

Proposal/Award Number (if available):

*Source of Support: National Science Foundation

*Primary Place of Performance : Duke University

Project/Proposal Start Date (MM/YYYY) (if available) : 07/2020

Project/Proposal End Date (MM/YYYY) (if available) : 06/2024

*Total Award Amount (including Indirect Costs): $ 696,442

*Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project

	*Year (YYYY)	*Person Months (##.##)
1.	2021	1.00
2.	2022	1.00
3.	2023	1.00

	Year (YYYY)	Person Months (##.##)
4.		
5.		

*Overall Objectives : High-energy colliders study the interaction between subatomic particles and environments produced in the collision of protons with protons, with nuclei, or between two nuclei. The study of such interactions requires an elaborate theoretical, statistical and computational framework. The X-SCAPE collaboration is a multi-disciplinary team of physicists, computer scientists, and statisticians, who are engaged in the construction of such an open-source framework.

*Statement of Potential Overlap : No overlap.

Projects/Proposals

2. *Project/Proposal Title : Meetings of New Researchers in Statistics and Probability (PI)

*Status of Support : (•) Current ( ) Pending ( ) Submission Planned ( ) Transfer of Support

Proposal/Award Number (if available): 2015380

*Source of Support: National Science Foundation

*Primary Place of Performance : Duke University

Project/Proposal Start Date (MM/YYYY) (if available) : 07/2020

Project/Proposal End Date (MM/YYYY) (if available) : 06/2023

*Total Award Amount (including Indirect Costs): $ 300,000

*Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project

	*Year (YYYY)	*Person Months (##.##)
1.	2020	0.00
2.	2021	0.00
3.	2022	0.00

	Year (YYYY)	Person Months (##.##)
4.	2023	0.00
5.		

*Overall Objectives : This project funds the New Researchers Conference (NRC) and the Preparing for Careers in Teaching Statistics and Data Science workshop. The NRC is the flagship meeting for junior researchers working in different areas of Statistics, Probability and Data Science. The primary objective of the conference is to provide a much needed platform for interaction among new researchers, as well as opportunities to seek mentorship from leading researchers in the field.

*Statement of Potential Overlap : No overlap.

Projects/Proposals

3. *Project/Proposal Title : Science-Integrated Predictive Modeling (SCINPL): a novel framework for scalable and interpretable predictive scientific computing

*Status of Support : (•) Current ( ) Pending ( ) Submission Planned ( ) Transfer of Support

Proposal/Award Number (if available):

*Source of Support: National Science Foundation

*Primary Place of Performance : Duke University

Project/Proposal Start Date (MM/YYYY) (if available) : 08/2022

Project/Proposal End Date (MM/YYYY) (if available) : 07/2025

*Total Award Amount (including Indirect Costs): $ 200,000

*Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project

	*Year (YYYY)	*Person Months (##.##)
1.	2023	0.30
2.	2024	0.30
3.	2025	0.30

	Year (YYYY)	Person Months (##.##)
4.		
5.		

*Overall Objectives : This project develops a Science-Integrated Predictive modeLing (SCINPL) framework, which incorporates scientific knowledge as prior belief within a probabilistic (Bayesian) predictive model. This integration of scientific principles within a probabilistic learning model enables timely and accurate emulation of expensive forward simulations, and provides an automated but principled approach for scientific discovery.

*Statement of Potential Overlap : No overlap.

Projects/Proposals

4.*Project/Proposal Title : Collaborative Research: ATD: a-DMIT: a novel Distributed, MultI-resolution, Topology-aware online monitoring framework of massive spatiotemporal data

*Status of Support : ○ Current ◉ Pending ○ Submission Planned ○ Transfer of Support

Proposal/Award Number (if available):

*Source of Support: National Science Foundation

*Primary Place of Performance : Duke University

Project/Proposal Start Date (MM/YYYY) (if available) : 07/2022

Project/Proposal End Date (MM/YYYY) (if available) : 06/2025

*Total Award Amount (including Indirect Costs): $ 123,363

*Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project

*Year (YYYY)	*Person Months (##.##)
1. 2023	0.15
2. 2024	0.15
3. 2025	0.15
4.	
5.	

*Overall Objectives : We propose a novel Distributed, MultI-source, Topology-aware (a-DMIT) online monitoring framework of massive spatiotemporal data. a-DMIT features three new monitoring methods: a multi-source Bayesian non-parametric monitoring approach via Gaussian processes, a topology-integrated change-point approach using persistence diagrams, and a distributed monitoring framework via conditional autoregressive modeling.

*Statement of Potential Overlap : No overlap.

Projects/Proposals

5.*Project/Proposal Title : Collaborative Research: Cost-Efficient and Confident Sampling for Modern Scientific Discovery (CURRENT)

*Status of Support : ○ Current ◉ Pending ○ Submission Planned ○ Transfer of Support

Proposal/Award Number (if available):

*Source of Support: National Science Foundation

*Primary Place of Performance : Duke University, Durham, NC

Project/Proposal Start Date (MM/YYYY) (if available) : 07/2023

Project/Proposal End Date (MM/YYYY) (if available) : 06/2026

*Total Award Amount (including Indirect Costs): $ 216,371

*Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project

*Year (YYYY)	*Person Months (##.##)
1. 2024	0.50
2. 2025	0.50
3. 2026	0.50

Year (YYYY)	Person Months (##.##)
4.	
5.	

*Overall Objectives : Quasi-Monte Carlo (QMC) methods replace independent and identically distributed points with low discrepancy points to improve computational efficiency. Given rapid developments in science and engineering, the spectrum of use cases for QMC needs to be broadened, and crucial methodological and theoretical issues must be addressed. This proposal aims to address these critical issues, providing a theoretically grounded and computationally efficient framework for scientific and big data applications.

*Statement of Potential Overlap : No overlap.

Projects/Proposals

6. *Project/Proposal Title :

*Status of Support : ○ Current ◉ Pending ○ Submission Planned ○ Transfer of Support

Proposal/Award Number (if available):

*Source of Support:

*Primary Place of Performance :

Project/Proposal Start Date (MM/YYYY) (if available) :

Project/Proposal End Date (MM/YYYY) (if available) :

*Total Award Amount (including Indirect Costs): $

*Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project

*Year (YYYY)	*Person Months (##.##)
1.	
2.	
3.	

Year (YYYY)	Person Months (##.##)
4.	
5.	

*Overall Objectives :

*Statement of Potential Overlap :

Facilities, Equipment and Other Resources

Facilities. All Illinois Tech faculty, PhD students, and visitors have offices provided at Illinois Tech. Summer MS, BS, and high school students have shared work areas. Faculty, student and visitor offices and conference rooms are provided by the Department of Applied Mathematics and the Office of Research.

The Department of Applied Mathematics has a research computer room that is available to all members of our department. The Center for Interdisciplinary Scientific Computation (CISC)—of which PI Hickernell is a member—has a 256-core cluster named von Neumann funded by Illinois Tech. Von Neumann is available available to all Illinois Tech research faculty and is centrally managed by Illinois Tech Office of Technology (OTS) Services. Illinois Tech is connected to the Open Science Grid through its own GridIIT.

Illinois Tech has site licenses for Mathematica, MATLAB, SAS, and JMP. Other open source software is also installed in our research and teaching laboratories.

Illinois Tech's university library provides access to journals, research monographs, and databases, either on-site, online or via inter-library loan.

Intellectual Resources.

Senior personnel. Dr. Sou-Cheng Choi, Research Asscoiate Professor at Illinois Tech and Principal Data Scientist at SAS, will provide in-kind, voluntary expertise in software engineering, documentation, and advising on important use cases. She has co-authored numerous articles in the field of computational mathematics. Dr. Choi is a co-author of the GAIL Matlab software library, a predecessor to QMCPy.

Unfunded collaborators. The PIs are part of a broad network of computational mathematicians and statisticians, including quasi-Monte Carlo (QMC) theorists and practitioners. In the supplemental attachments, we have included letters of collaboration from a key collaborators on this project, should it be funded.

Prof. Art B. Owen (Stanford University) has engaged with the PIs in conversations about QMC for many years. Prof. Owen is particularly expert in randomized QMC and the use of low discrepancy points for Markov chain Monte Carlo. He has taken a keen interest in QMCPy and will put forward new QMC use cases, advise on software features to be included, and possibly collaborate on joint publications with the PIs.

Dr. Michael J. McCourt (SigOpt) convinced his company to fund the early development of QMCPy. He was convinced of the advantages of low discrepancy sampling and wanted to spread these ideas to the tech industry. During the first two years of QMCPy's development, Mike advised us what we should prioritize for the benefit of tech practitioners. Although SigOpt is not in a position to fund QMCPy further, Mike will advise us on the continued development of SigOpt. He will also help us spread the word among his network in the machine learning community.

Prof. Chris Oates (Newcastle University) is an expert in probabilistic numerics and Stein discrepancies. He will collaborate with us on our development of new stopping criteria and on our development of cost-efficient Stein points for expensive Bayesian inference.

Dr. Jagadeeswaran Rathinavel (Wi-Tronix) has developed fast Bayesian stopping criteria for multivariate integration. He will help us extend those stopping criteria to multifidelity problems.

Dr. Pieterjan Robbe (Sandia National Laboratories and Katholieke Universiteit Leuven) is a postdoctoral researcher who has contributed multilevel QMC capabilities and use cases in QMCPy. He will extend those as well as contribute to the theory of multi-level and multifidelty model computation.

Short-term visitors. Some scholars whose expertise would be immensely helpful to the goals of this project will be hosted at Illinois Tech and/or Duke University.

Illinois Tech was listed on the National Federal Register of Historic Places in 2005. The proposed research activities will not make any physical changes to Illinois Tech's campus and buildings.

FACILITIES AND COMPUTING EQUIPMENT – Duke University

Office

The PI has an office in the Department of Statistical Science in the Old Chemistry building.

Group

Graduate students are provided shared office space in Statistical Science, including desk space and a computer with access to the OIT network as well as the networks available to the department.

Equipment

Duke University's Statistical Science Computing Resources

The Duke University Department of Statistical Science maintains a near state of the art network of approximately fifty single-and dual-processor x86 and x86-64-based Linux workstations, approximately a dozen Windows PCs in a Samba network environment, and a range of networked monochrome and color Postscript printers for its faculty, Ph.D. students, and staff. Rack-mounted servers offer file, e-mail, web, and authentication service. A RAID storage server facility offers something under one terabyte of disk capacity, backed up daily to an LTO tape changer. The software environment includes a wide array of scientific programming tools, including the GNU suite of libraries, compilers, and development tools, a range of scientific and statistical computing environments such as Matlab, Maple, Mathematica, S-Plus, R, OpenBUGS, etc. An MPI-based parallel computing environment is provided that is consistent with the HPC environment at Duke's Computational Science, Engineering & Medicine (CSEM) facility, to aid investigators in prototyping and debugging parallel computer code. The computing environment is maintained by a full-time systems manager and systems programmer.

University Facilities

The backbone and other university-level infrastructure needs of the University are maintained by a central IT organization, the Office of Information Technology (OIT). OIT is responsible for the operation, testing, support, and engineering of the campus-wide data, voice, and video communications infrastructure. This includes the design and subsequent implementation of structured wiring and switching systems, enterprise-level servers, including Domain Name Server (DNS) and Dynamic Host Configuration Protocol (DHCP) servers, routing systems, and wireless systems.

Duke University's high-speed backbone, DukeNet, provides researchers, staff, faculty and students with a robust, redundant conduit for data. The backbone consists of Cisco routers with redundant 10 gigabit ethernet links. Most buildings on campus are wired with Category 5 cabling and have 10M/100M Ethernet ports supplied to each desktop. Servers and high speed research workstations can be provided with gigabit or ten gigabit ethernet ports as needed. Building networks connect to the backbone via dual gigabit or 10 gigabit ethernet uplinks.

The Duke Shared Cluster Resource (DSCR) facility maintains a shared computational cluster facility of over 600 machines (1152 processors) to which we have access. The processors range from 2.8GHz to 3.6GHz. While the cluster must be shared by the entire community at Duke, it provides a useful resource for computational science.

The University also maintains a campus-wide AFS file system infrastructure with terabytes of storage; a campus-wide electronic mail infrastructure supporting over 35,000 mailboxes and handling in excess of a million messages a day; a server-based file service, authentication services; directory services; web service; and name service and other network services.

Data Management Plan

The research results will be shared with the academic community and general public through conference presentations, journal articles, the QMCPy.org blog and posts on the investigators web site. Software generated by the project will be freely available for educational, research and non-profit purposes on Github. When appropriate, preprints will be posted on https://arxiv.org/. Slide decks for talks will be posted on SpeakerDeck or similar platform. Publications, both articles and software, will be uploaded to NSF Public Access Repository. The emphasis on transparency and accountability of data management will be maintained throughout the project. We will maintain records, so that all mathematical calculations and simulations are reproducible.

Products of Research: The main contributions of this research project are the development and analysis of new simulation algorithms, including both theory and software. Details of the main research products will first appear in the investigators' notebooks and preprints. The significant research results will be published in peer-reviewed journal articles, book chapters, or conference proceedings. The primary journals for the published work will include those in computational mathematics, statistics, and the application areas where our algorithms will be used.

Data Format, Content, and Backup: All researchers will maintain research folders with the current state of the project. The electronic files will be backed up on a daily basis. Team members will use a shared Google/Dropbox folder and/or Github repository to provide access to computer files to all team members. In this way, all team members will have access to the ongoing files of the project. These files will contain computer code, text, plots, and images. Google/Dropbox will be used to share data and files among the project researchers and the files will be reasonably organized and coherent. The papers will be written in LaTeX or MS Word, the images will be in a standard format, such as jpg, tiff, or eps. The data files will be in format that is easily readable by a wide community of users, e.g. Excel or ascii text.

Data Access and Sharing: All participants in the project will publish the results of their work. The model data and other supporting materials created or gathered in the course of the work will be shared with other researchers upon reasonable request and within a reasonable time of the request, if the investigators have the authority to share the data. To ensure that data generated with this project is widely available and archived, the estimates for model parameters and references will be included in the research papers, and auxiliary material provided with their publication.

Software developed—in particular QMCPy algorithms and demos—will be available on a public repository. The software will be open to pull requests from those who wish to contribute to its development. This is consistent with our intention to make QMCPy a community-owned library.

Reuse and Redistribution: Public access to research products will be regulated in order to protect privacy and confidentiality concerns, as well to respect any proprietary or intellectual property rights. Legal offices will be consulted on a case-by-case basis to address any concerns, if necessary. Terms of use will include proper attribution to the PIs and authors along with disclaimers of liability in connection with any use or distribution of the research data.

Archiving and Preservation of Access: Research products will be made available immediately after publication. Journal publications will be available online from respective journal websites and linked to by the PIs' university websites. Again, publications, both articles and software, will be uploaded to NSF Public Access Repository. All computer data and files generated as a result of this project will backed up daily to protect from loss of data from hardware failures, fire, theft, etc.

KU LEUVEN

DEPARTMENT OF COMPUTER SCIENCE
NUMA SECTION
CELESTIJNENLAAN 200A BOX 2402
B-3001 LEUVEN

Dear NSF Proposal Review Committee

If the collaborative proposal submitted by Drs. Fred J. Hickernell and Simon Mak entitled *Collaborative Research: Cost-Efficient and Confident Sampling for Modern Scientific Discovery* is selected for funding by NSF, it is my intent to collaborate and/or commit resources as detailed in the Project Description or the Facilities, Equipment and Other Resources section of the proposal.

Sincerely,

Pieterjan Robbe, FWO postdoctoral fellow

December 13, 2022

School of Mathematics,
Statistics and Physics
Newcastle University
Herschel Building
Newcastle upon Tyne
NE1 7RU United Kingdom

13 December 2022

To whom it may concern,

If the collaborative proposal submitted by Drs. Fred J. Hickernell and Simon Mak entitled

Collaborative Research: Cost-Efficient and Confident Sampling for Modern Scientific Discovery

is selected for funding by NSF, it is my intent to collaborate and/or commit resources as detailed in the Project Description or the Facilities, Equipment and Other Resources section of the proposal.

Yours sincerely,

Prof. Chris Oates
Chair in Statistics, Newcastle University, UK
Group Leader and Fellow, Alan Turing Institute, UK

Tel :+44 (0) 191 208 3944
Fax :+44 (0) 191 208 8020

www.ncl.ac.uk/maths-physics/

SIGOPT

From the office of the general manager

If the collaborative proposal submitted by Drs. Fred J. Hickernell and Simon Mak entitled *Collaborative Research: Cost-Efficient and Confident Sampling for Modern Scientific Discovery* is selected for funding by NSF, it is my intent to collaborate and/or commit resources as detailed in the Project Description or the Facilities, Equipment and Other Resources section of the proposal.

Michael McCourt

Michael McCourt
General Manager
mccourt@sigopt.com

SigOpt
315 Montgomery St.
Floor 10, Suite 1016A
San Francisco, CA 94104

@SigOpt
/sigopt
contact@sigopt.com • sales@sigopt.com

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
STANFORD, CALIFORNIA 94305–4065

Art B. Owen
Max H. Stein Professor of Statistics
Phone: (650) 725-2232
Email: owen@stanford.edu

December 13, 2022

Program Directors
National Science Foundation
2415 Eisenhower Avenue
Alexandria, Virginia 22314

Dear Program Directors,

If the collaborative proposal submitted by Drs. Fred J. Hickernell and Simon Mak entitled Collaborative Research: Cost-Efficient and Confident Sampling for Modern Scientific Discovery is selected for funding by NSF, it is my intent to collaborate and/or commit resources as detailed in the Project Description or the Facilities, Equipment and Other Resources section of the proposal.

Sincerely,

Art B. Owen
Max H. Stein Professor of Statistics

Jagadeeswaran Rathinavel
631, E Boughton Rd, Suite 240,
Bolingbrook, IL 60440
(312) 493-2464
jrathin1@iit.edu

2nd January 2023

To whom it may concern

If the collaborative proposal submitted by Drs. Fred J. Hickernell and Simon Mak entitled *Collaborative Research: Cost-Efficient and Confident Sampling for Modern Scientific Discovery* is selected for funding by NSF, it is my intent to collaborate and/or commit resources as detailed in the Project Description or the Facilities, Equipment and Other Resources section of the proposal.

Sincerely yours,

Jagadeeswaran Rathinavel

Jagadeeswaran Rathinavel,

Principal Data Scientist, Wi-Tronix LLC,
Bolingbrook, IL, 60440

Sincerely,

Other Supplementary Documents

Data Not Available

Table 1

1	Your Name:	Your Organizational Affiliation(s), last 12 mo	Last Active Date
	Hickernell, Fred J.	Illinois Institute of Technology	

Table 2

2	Name:	Type of Relationship	Optional (email, Department)	Last Active Date
R	None			

Table 3

3	Advisor/Advisee Name:	Organizational Affiliation	Optional (email, Department)
G	Benney, David	Massachussetts Institute of Technology	deceased
G	Howard, Louis N.	Massachussetts Institute of Technology	deceased
G	Rosales, Ruben Rodolfo	Massachussetts Institute of Technology	Mathematics
T	Ding, Yuhan	Illinois Institute of Technology	
T	Hall, Jr., Claude	Illinois Institute of Technology	
T	Hong, Regina	Germany	
T	Huang, Fanglun	Anhui University	
T	Jiang, Lan	Compass	
T	Jiménez, Rugama, Ll. A.	UBS	
T	Li, Yiou	DePaul University	
T	Liu, Kwong-IP	Hong Kong Baptist University	
T	Niu, Ben	RMB Capital	
T	Rathinavel, Jagadeeswaran	Illinois Institute of Technology	
T	Sorokin, Aleksei	Illinois Institute of Technology	
T	Yue, Rongxian	Shanghai Normal University	
T	Zeng, Xiaoyan	Shanghai University	
T	Zhang, Kan	Illinois Institute of Technology	
T	Zhang, Yizhi	Jamran International	
T	Zhang, Yonglin		deceased
T	Zhou, Xuan	J. P. Morgan	

Table 4

4	Name:	Organizational Affiliation	Optional (email, Department)	Last Active Date
A	Ai, Mingyao	Peking University		12/31/20

A	Chen, Jianbin	Nankai University		12/31/20
A	Chen, Victoria	Yale School of Public Health		12/31/20
A	Choi, Sou-Cheng Terrya	Illinois Institute of Technology		
A	Cockayne, Jon	University of Warwick		08/31/20
A	Cui, Xiangzhao	Honghe University		12/31/20
A	Dewar, Jeremy	Tulane University		08/31/20
A	Ding, Yuhan	Illinois Institute of Technology		
A	Fan, Jianqing	Princeton University		12/31/20
A	Fang, Hong-Bin	Georgetown University Medical Center		12/31/20
A	Girolami, Mark	University of Cambridge		08/31/20
A	Hao, Chengcheng	Shanghai University of International Business and Economics		12/31/20
A	He, Ping	Bejing Normal University and Hong Kong Baptist University United International College		12/31/20
A	Hefter, Mario	Technische Universität Kaserslautern		08/31/20
A	Hesse, Kerstin	Paderborn University		08/31/20
A	Hinrichs, Aicke	Johannes Kepler Universität		08/31/20
A	Hu, Ting	Wuhan University		12/31/20
A	Huang, Yimin	Peking University		12/31/20
A	Jiang, Hongyan	Huaiyin Institute of Technology		12/31/20
A	Jiang, Lan	Compass		
A	Jiménez, Rugama, Ll. A.	UBS		
A	Keller, Alex	Nvidia		
A	Krieg, David	Johannes Kepler Universität		08/31/20
A	Li, Chun	Honghe University		12/31/20
A	Li, Da	Illinois Institute of Technology		12/31/18
A	Li, Feng	Central University of Finance and Economics		12/31/20
A	Li, Gang	University of California at Los Angeles		12/31/20
A	Li, Runze	Pennsylvania State University		12/31/20
A	Liang, Jiajun	University of New Haven		12/31/20
A	Lin, Dennis K. J.	Purdue University		12/31/20
A	Liu, Min-Qian	Nankai University		12/31/20
A	Liu, Wanjun	Pennsylvania State University		12/31/20
A	Liu, Xiaobing	University at Buffalo		12/31/20
A	Lu, Ying	Stanford University		12/31/20
A	Ma, Changxing	University at Buffalo		12/31/20

A	McCourt, Michael	SigOpt, an Intel Company		
A	Meyer, Lukas	Technische Universität Kaserslautern		08/31/20
A	Ning, Jianhui	Central China Normal University		12/31/20
A	Oates, Chris	University of Newcastle		08/31/20
A	Oftadeh, Elaheh	University of Kent		12/31/20
A	Pan, Jianxin	University of Manchester		12/31/20
A	Peng, Xiaoling	Bejing Normal University and Hong Kong Baptist University United International College		12/31/20
A	Prangle, Dennis	Newcastle University		08/31/20
A	Prochno, Joscha	Karl-Franzens Universität		08/31/20
A	Qin, Hong	Central China Normal University		12/31/20
A	Qiu, Si	Central China Normal University		12/31/20
A	Ritter, Klaus	Technische Universität Kaserslautern		08/31/20
A	Ruiz, Mirtha Pari	Universidad de Tarapacá		12/31/20
A	Sorokin, Aleksei	Illinois Institute of Technology		
A	Stehlík, Milan	University of Valparaíso		12/31/20
A	Stehlíkova, Silvia	Johannes Kepler University		12/31/20
A	Sullivan, Tim J.	University of Warwick		08/31/20
A	Sun, Xiaoying	York University		12/31/20
A	Tan, Ming T.	Georgetown University Medical Center		12/31/20
A	Tang, Man-Lai	Hang Seng Management College		12/31/20
A	Tang, Yu	Soochow University		12/31/20
A	Tian, Guoliang	Southern University of Science and Technology		12/31/20
A	Tian, Mu	Tulane University		08/31/20
A	Tong, Xing	University of Illinois, Chicago		
A	Ullrich, Mario	Johannes Kepler Universität		08/31/20
A	von Rosen, Dietrich	Swedish University of Agricultural Sciences		12/31/20
A	Wang, Baobin	South-Central University for Nationalities		12/31/20
A	Wasilkowski, Greg W.	University of Kentucky		08/31/20
A	Winker, Peter	Justus-Liebig-University Giessen		12/31/20
A	Wu, Yuehua	York University		12/31/20
A	Xie, Minyu	Central China Normal University		12/31/20
A	Xu, Jian-Lun	National Cancer Institute		12/31/20
A	Xu, Qinsong	Central South University		12/31/20
A	Yang, Jing	Tianjin Medical University		12/31/20
A	Yin, Hong	Renmin University of China		12/31/20
A	Yu, Jun	Beijing Institute of Technology		12/31/20

A	Yue, Rongxian	Shanghai Normal University		12/31/20
A	Zeng, Xiaoyan	Shanghai University		
A	Zhang, Aijun	The University of Hong Kong		12/31/20
A	Zhang, Defei	Honghe University		12/31/20
A	Zhang, Heping	Yale School of Public Health		12/31/20
A	Zhang, Jian	University of Kent		12/31/20
A	Zhang, Jin-Ting	National University of Singapore		12/31/20
A	Zhang, Mei	Sichuan University		12/31/20
A	Zhao, Xuejing	Lanzhou University		12/31/20
A	Zhou, Xuan	J. P. Morgan		
A	Zhou, Yongdao	Nankai University		12/31/20
A	Zhu, Tianming	National University of Singapore		12/31/20
C	Constantine, Paul	University of Colorado, Boulder		12/31/18
C	Ebert, Adrian	Radon Institue for Comp. & App. Math.		12/31/21
C	Fasshauer, Gregory E.	Colorado School of Mines		12/31/18
C	Giles, Michael B.	University of Oxford		
C	Hall, Jr., Claude	Illinois Institute of Technology		
C	Hyman, Mac	Tulane University		12/31/21
C	Kang, Lulu	Illinois Institute of Technology		
C	Kritzer, Peter	Radon Institue for Comp. & App. Math.		
C	Kuo, Frances	University of New South Wales		
C	L'Ecuyer, Pierre	University of Montreal		
C	Li, Yiou	DePaul University		
C	Mak, Simon	Duke		
C	McCourt, Michael	SigOpt, an Intel Company		
C	Novak, Erich	Friedrich-Schiller-Universität Jena		
C	Nuyens, Dirk	Katholieke Universiteit Leuven		
C	Osisiogu, Onyekachi	Radon Institue for Comp. & App. Math.		
C	Owen, Art	Stanford University		
C	Rathinavel, Jagadeeswaran	Illinois Institute of Technology		
C	Roshan, V	Georgia Tech		12/31/18
C	Wozniakowski, Henryk	University of Warsaw		12/31/21
C	Zhang, Kan	Illinois Institute of Technology		
C	Zhang, Yizhi	Jamran International		12/31/21

Table 5

5	Name:	Organizational Affiliation	Journal/Collection	Last Active Date
B	Dick, Josef	Univesity of New South Wales	Journal of Complexity	
B	Guo, Lei	Acad. of Math. \& System Sci., CAS	J. of Math. Research with Appl.	
B	Hinrichs, Aicke	Johann Kepler University Linz	Journal of Complexity	
B	Hsu, L. C.	Dalian University of Technology	J. of Math. Research with Appl.	
E	Kritzer, Peter	Radon Institue for Comp. & App. Math.	Journal of Complexity	
E	Kuo, Frances	University of New South Wales	Journal of Complexity	
B	Novak, Erich	Friedrich-Schiller-Universität Jena	Journal of Complexity	
B	Ritter, Klaus	Technische Universität Kaiserslautern	Journal of Complexity	
B	Sloan, Ian H.	University of New South Wales	Journal of Complexity	
B	Wang, Renhong	Dalian University of Technology	J. of Math. Research with Appl.	
B	Wasilkowski, Greg	Univesity of Kentucky	Journal of Complexity	
B	Wozniakowski, Henryk	University of Warsaw	Journal of Complexity	
B				
B				
B				
B				
B				
B				

Table 1

1	Your Name:	Your Organizational Affiliation(s), last 12 mo	Last Active Date
	Ding, Yuhan	Illinois Institute of Technology	
	Ding, Yuhan	Misericordia University	05/31/19

Table 2

2	Name:	Type of Relationship	Optional (email, Department)	Last Active Date
R	None			

Table 3

3	Advisor/Advisee Name:	Organizational Affiliation	Optional (email, Department)
G	Hickernell, Fred J.	Illinois Institute of Technology	Applied Mathematics

Table 4

4	Name:	Organizational Affiliation	Optional (email, Department)	Last Active Date
A	Hickernell, Fred J.	Illinois Institute of Technology		
A	Choi, Sou-Cheng Terrya	Illinois Institute of Technology		12/31/21
A	Jiang, Lan	Compass		12/31/21
A	Jimenez, Rugama, LI. A.	UBS		12/31/21
A	Tong, Xin	Florida State University		12/31/21
A	Rathinavel, Jagadeeswarar	Illinois Institute of Technology		12/31/21
A	Zhang, Kan	Illinois Institute of Technology		12/31/21
A	Zhang, Yizhi	Jamran International		12/31/21

Table 5

5	Name:	Organizational Affiliation	Journal/Collection	Last Active Date
B	None			
E	None			

Table 1

1	Your Name:	Your Organizational Affiliation(s), last 12 mo	Last Active Date
	Choi, Sou Cheng	Illinois Institute of Technology (adjunct)	
	Choi, Sou Cheng	SAS Institute Inc.	
	Choi, Sou Cheng	Kamakura Corporation, USA	06/30/22

Table 2

2	Name:	Type of Relationship	Optional (email, Department)	Last Active Date
R	Lim, Lek Heng	University of Chicago	lekheng@galton.uchicago.edu, Statistics	
R	Jarrow, Robert	Cornell University and SAS Institute Inc.	robert.jarrow@sas.com	
R	van Deventer, Donald	SAS Institute Inc.	donald.vandeventer@sas.com	

Table 3

3	Advisor/Advisee Name:	Organizational Affiliation	Optional (email, Department)
G	Saunders, Michael	Stanford University	saunders@stanford.edu, Management Science and Engineering
G	Golub, Gene (deceased)	Stanford University	Computer Science
G	Larsen, Rasmus Munk	Google Inc.	
G	Levy, Doron	University of Maryland, College Park	dlevy@math.umd.edu, Mathematics
G	Donoho, David L	Stanford University	donoho@stanford.edu, Statistics
T	Ding, Yuhan	Illinois Institute of Technology	yding2@hawk.iit.edu, Applied Mathematics

Table 4

4	Name:	Organizational Affiliation	Optional (email, Department)	Last Active Date
A	Hartong-Redden, Rory	Runtastic GmbH	roryhr@gmail.com	07/15/18
A	Kochenderfer, Mykel J.	Stanford University	mykel@stanford.edu, Aeronautics and Astronautics	07/15/18
A	Wulfe, Blake	Stanford University	wulfebw@stanford.edu, Computer Science	07/15/18
A	Huang, Jack	University of Chicago	jhuang11@uchicago.edu, Mathematics	05/03/19
A	Liu, Andy	Illinois Mathematics and Science Academy	aliu2@imsa.edu	05/03/19
A	Chintakindi, Sunil	Allstate Corporation	sunil.chintakindi@allstate.com, Automotive Innovation	01/31/20
A	Gibson, Tim	Allstate Corporation	tgibu@allstate.com, Automotive Innovation	01/31/20

A	Kodali, Anuradha	Allstate Corporation	akoda@allstate.com, Automotive Innovation	01/31/20
A	Li, Jinyang	Beijing Institute of Technology, China	jli221@hawk.iit.edu	05/26/21
A	Wang, Haoran	Beijing Institute of Technology, China	ishaoranwang@gmail.com	05/26/21
A	Zhang, Yiwei	Beijing Institute of Technology, China	yzhang176@hawk.iit.edu	05/26/21
A	Abelseth, Connie	Trinity International University		06/30/21
A	Dorrough, Storm	Trinity International University		06/30/21
A	Lok, Shu Yun	Trinity International University		06/30/21
A	Pao, Chrystal Ho	Trinity International University	chrystalhopao@gmail.com, Biology	06/30/21
A	Rentas, Angelo	Trinity International University		06/30/21
A	Shelton, Joyce	Trinity International University		06/30/21
A	Fung, Glenn	American Family Insurance, U.S.	GFUNG@amfam.com	12/01/21
A	Ma, Lawrence K. H.	Hong Kong Blockchain Society	lawrence.ma@hkbcs.org	12/01/21
A	Polania, Luisa	Target Corporation, U.S.	lfpolani@udel.edu	12/01/21
A	Wu, Victor	Ensembl AI, USA	wu.victor@gmail.com	12/01/21
A	Jiang, Lan	Compass Inc.	lanjiang61930@gmail.com	07/29/22
A	Jiménez Rugama, Lluís Antoni	Virtu Financial	lluisantoni@gmail.com	07/29/22
A	Tong, Xin	Florida State University	xtong5@hawk.iit.edu	07/29/22
A	Zhang, Kan	Illinois Institute of Technology	kzhang23@hawk.iit.edu, , Applied Mathematics	07/29/22
A	Zhang, Yizhi	Illinois Institute of Technology	yzhang97@hawk.iit.edu, Applied Mathematics	07/29/22
A	Zhou, Xuan	Morgan Stanley	xuanjzhou@gmail.com	07/29/22
A	Ding, Yuhan	Illinois Institute of Technology	yding2@hawk.iit.edu, Applied Mathematics	
A	Hickernell, Fred J.	Illinois Institute of Technology	hickernell@iit.edu, Applied Mathematics	
A	Jarrow, Robert A.	SAS Institute Inc.	robert.jarrow@sas.com	
A	McCourt, Michael	SigOpt, Inc.	mccourt@sigopt.com	
A	Mesler, Mark	SAS Institute Inc.	mark.mesler@sas.com	
A	Penanhoat, Eric	SAS Institute Inc.	eric.penanhoat@sas.com	
A	Rathinavel, Jagadeeswaran	Wi-Tronix, LLC	jrathin1@hawk.iit.edu, Applied Mathematics	
A	Sorokin, Aleksei	Illinois Institute of Technology	asorokin@hawk.iit.edu	
A	van Deventer, Don	SAS Institute Inc.	donald.vandeventer@sas.com	
A	Zorn, Martin	SAS Institute Inc.	martin.zorn@sas.com	
C	Ashbeck, Jonathan	University of Pennsylvania	jashbeck@sas.upenn.edu	08/31/20
C	Liu, Zongrui	University of Illinois Urbana-Champaign	zongrui2@illinois.edu	08/31/20
C	Negron, Alex	Illinois Institute of Technology	anegron1@hawk.iit.edu	08/31/20

C	Burwei, Rebecca	Mozilla Corporation	rebecca.burwei@gmail.com	
C	Ebert, Adrian	Austrian Academy of Sciences		
C	Giles, Michael	University of Oxford		
C	Herman, Joshua	Bank of America, USA	zitterbewegung@gmail.com	
C	Hofert, Marius	University of Waterloo		
C	Ji, Irene	Duke University		
C	Kang, Lulu	Illinois Institute of Technology	lkang2@iit.edu	
C	Kim, Jungtaek	Pohang University of Science and Technology		
C	Kucherenko, Sergei	Imperial College London		
C	L'Ecuyer, Pierre	University of Montreal		
C	Lemieux, Christiane	University of Waterloo		
C	Li, Shuwang	Illinois Institute of Technology	sli15@iit.edu	
C	Lim, Lek Heng	University of Chicago	lekheng@galton.uchicago.edu, Statistics	
C	Mak, Simon	Duke University	sm769@duke.edu	
C	Nuyens, Dirk	Katholieke Universiteit Leuven		
C	Osisiogu, Onyekachi	Austrian Academy of Sciences		
C	Owen, Art	Stanford University		
C	Robbe, Pieterjan	Sandia National Laboratories and Katholieke Universiteit Leuven		
C	Sullivan, Tim	Warwick University		
C	Tang, Tao	Duke University		
C	Zhang, Ruda	Duke University		

Table 5

5	Name:	Organizational Affiliation	Journal/Collection	Last Active Date
B	Chui, Charles K.	Stanford University; Hong Kong Baptist University	Mathematics of Computation and Data Science	12/01/21
E	Fung, Glenn	American Family Insurance, USA	GFUNG@amfam.com	12/01/21
E	Ma, Lawrence K. H.	Hong Kong Blockchain Society	lawrence.ma@hkbcs.org	12/01/21
E	Polania, Luisa	Target Corporation, USA	lfpolani@udel.edu	12/01/21
E	Wu, Victor	Ensembl, USA	wu.victor@gmail.com	12/01/21

Table 1

1	Your Name:	Your Organizational Affiliation(s), last 12 mo	Last Active Date
	Mak, Simon	Duke University	

Table 2

2	Name:	Type of Relationship	Optional (email, Department)	Last Active Date

Table 3

3	Advisor/Advisee Name:	Organizational Affiliation	Optional (email, Department)
G	Wu, C. F. Jeff	Georgia Institute of Technology	
G	Joseph, Roshan Vengazhiyil	Georgia Institute of Technology	

Table 4

4	Name:	Organizational Affiliation	Optional (email, Department)	Last Active Date
A	Joseph, Roshan Vengazhiyil	Georgia Institute of Technology		
A	Yeh, Shiang-Ting	Raytheon Co.		
A	Wang, Xingjian	Florida Institute of Technology		
A	Sung, Chih-li	Michigan State University		
A	Chang, Yu-Hung	Georgia Institute of Technology		
A	Yang, Vigor	Georgia Institute of Technology		
A	Xie, Yao	Georgia Institute of Technology		
A	Ding, Yuhan	Illinois Institute of Technology		
A	Hickernell, Fred J	Illinois Institute of Technology		
A	Kritzer, Peter	Austrian Academy of Sciences		
A	Zhang, Liwei	University of Texas, Arlington		
A	Li, Yixing	Cadence Design Systems		
A	Lin, Li-Hsiang	Georgia Institute of Technology		
A	Yuchi, Shaowu	Georgia Institute of Technology		
A	Chen, Jialei	Georgia Institute of Technology		
A	Zhang, Chuck	Georgia Institute of Technology		
A	Chen, Zhehui	Georgia Institute of Technology		
A	Ding, Liang	Texas A&M University		

A	Huling, Jared D	University of Minnesota		
A	Bass, Steffen	Duke University		
A	Dunson, David	Duke University		
A	Wolpert, Robert	Duke University		
A	Ruda Zhang	University of Houston		
A	Zhou, Yuanshou	Apple		
A	Rudin, Cynthia	Duke University		
A	Xie, Liyan	Chinese University of Hong Kong		

Table 5

5	Name:	Organizational Affiliation	Journal/Collection	Last Active Date
B	Joseph, Roshan Vengazhiyil	Georgia Institute of Technology	Technometrics	
B	Gramacy, Robert	Virginia Tech	Technometrics	
B	Matteson, David	Cornell University	Data Science in Science	

List of Suggested Reviewers

Data Not Available

List of Suggested Reviewers

Data Not Available

List of Reviewers Not to Include

Data Not Available

List of Reviewers Not to Include

Data Not Available