COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCEMENT/SOLICITATION NO./DUE DATE PD 20-8069 09/15/2020	☐ Special Exception to Deadline Date Policy	FOR NSF USE ONLY NSF PROPOSAL NUMBER 2053714
FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) (Indicate the most specific unit known, i.e. program, division, etc.) DMS - CDS&E-MSS		

DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System)	FILE LOCATION
09/15/2020	1	03040000 DMS	8069	042084434	

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?
362-17-0136	☐ A RENEWAL ☐ AN ACCOMPLISHMENT-BASED RENEWAL	YES ☐ NO ☒ IF YES, LIST ACRONYM(S)

NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE Illinois Institute of Technology	ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE Illinois Institute of Technology 10 West 35th Street Suite 7D71 Chicago,IL 606163717 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 Illinois Institute of Technology 10 West 35th Street, Suite 7D71 Chicago,IL 606163717 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: Quasi-Monte Carlo Community Software

REQUESTED AMOUNT	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE
$ 395,338	36 months	06/01/2021	

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 ORG
☒ INTERNATIONAL ACTIVITIES: COUNTRY/COUNTRIES INVOLVED
XX
☒ COLLABORATIVE STATUS
A collaborative proposal from multiple organizations (PAPPG II.D.3.b)

PI/PD DEPARTMENT Applied Mathematics	PI/PD POSTAL ADDRESS 10 West 32nd Street Chicago, IL 60616 US
PI/PD FAX NUMBER 3125673135	

NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Email Address
PI/PD NAME Fred J Hickernell	PhD	1981	3125678983	hickernell@iit.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, Section §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 IX.A.; 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 organization's expenditure of any funds under the award, will be, satisfactorily managed, reduced or eliminated in accordance with the organization's 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 proposals for conferences, symposia, and workshops.)

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.

AUTHORIZED ORGANIZATIONAL REPRESENTATIVE	SIGNATURE	DATE
NAME Toni Allen	Electronic Signature	Sep 15 2020 02:28 PM

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

COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCEMENT/SOLICITATION NO./DUE DATE: PD 20-8069 — 09/15/2020

☐ Special Exception to Deadline Date Policy

FOR NSF USE ONLY — NSF PROPOSAL NUMBER: 2053715

FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) (Indicate the most specific unit known, i.e. program, division, etc.): DMS - CDS&E-MSS

DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System)	FILE LOCATION
09/15/2020	1	03040000 DMS	8069	044387793	

EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN): 560-53-2129

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)

NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE: Duke University

ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE:
Duke University
2200 W. Main St, Suite 710 Erwin Square
Durham,NC 277054010 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:
Duke University
2200 W. Main St, Suite 710, Erwin Square
Durham,NC 277054010 US

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

☐ IF THIS IS A PRELIMINARY PROPOSAL THEN CHECK HERE

TITLE OF PROPOSED PROJECT: Collaborative Research: Quasi-Monte Carlo Community Software

REQUESTED AMOUNT	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE
$ 207,985	36 months	06/01/2021	

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 ORG
☐ INTERNATIONAL ACTIVITIES: COUNTRY/COUNTRIES INVOLVED
☒ COLLABORATIVE STATUS
A collaborative proposal from multiple organizations (PAPPG II.D.3.b)

PI/PD DEPARTMENT: Statistical Science

PI/PD POSTAL ADDRESS:
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	4046614431	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, Section §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 IX.A.; 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 organization's expenditure of any funds under the award, will be, satisfactorily managed, reduced or eliminated in accordance with the organization's 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 proposals for conferences, symposia, and workshops.)

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.

AUTHORIZED ORGANIZATIONAL REPRESENTATIVE	SIGNATURE	DATE
NAME Lauren Faber	Electronic Signature	Sep 15 2020 01:51 PM

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

Overview

We will grow QMCPy, a Quasi-Monte Carlo (QMC) Python software library, from its nascent form to be a library embraced by the QMC community. QMC methods replace independent and identically distributed points by low discrepancy (highly stratified) points. This yields significant improvements in computational efficiency for a wide range of important problems, including Bayesian inference, financial risk, uncertainty quantification, and machine learning. Despite the demonstrated effectiveness of QMC methods, existing software comes primarily from *individual* research groups acting independently. There has yet to be a community owned effort which *integrates* these initiatives into a cutting edge package, which practitioners can apply to their problem of interest. QMCPy will take on this role, bridging the gap between research efforts of individual groups and QMC practitioners.

QMCPy will become a tightly-connected collection of best QMC software, including low discrepancy generators, QMC algorithms, and illuminating use cases. It will provide a user-friendly, consistent interface to the contributions of many scholars. QMCPy will serve as a proving ground for new QMC methods, allowing researchers to test their ideas with the state-of-the-art on a broad range of use cases. QMCPy will be owned and cultivated by the QMC community, and will provide an easy on-ramp for practitioners who are new to QMC.

Intellectual Merit

Currently, QMCPy has a basic suite of low discrepancy generators, QMC algorithms, and use cases. We will develop QMCPy in three crucial directions, guided by our motivation of bridging individual researchers to potential practitioners. First, we will implement in QMCPy a broader offering of low discrepancy sequence generators. This includes a richer scrambling of digital sequences, higher-order digital sequences, and low discrepancy sequences for Bayesian computation and experimental design. Second, we will implement a broader array of algorithms. This includes popular variance/variation reduction methods such as importance sampling and control variates, multilevel QMC algorithms, and density estimation algorithms. Lastly, we will develop and implement a broader spectrum of use cases for QMC, including big data analytics, Bayesian modeling, uncertainty quantification, computer vision, and probabilistic numerics. New algorithms will be underpinned by new theory. Performance will be enhanced by a computational framework including refactoring in C, parallel processing, and GPU computing.

Broader Impacts

By connecting individual QMC research groups with QMC practitioners, QMCPy will catalyze new application areas for low discrepancy sampling. QMCPy will serve as a proving ground for researchers to compare new ideas with the state-of-the-art on a broad range of realistic use cases. Large scale software packages, such as `scipy`, PyTorch, or MATLAB, do not offer the options that QMCPy does and will. QMCPy will offer practitioners a user-friendly platform for implementing the best QMC methods, which will bring about significant computational gains in a broad range of scientific disciplines. The success of QMC in new application areas will in turn raise new theoretical questions and computational challenges, which will advance the QMC community as a whole.

QMCPy will serve as an important vehicle for educating cross-disciplinary researchers and promoting proper QMC practice. Students supported by this project will learn to write clean, efficient code that fits the package architecture, is documented, and passes doctests. They will also become proficient in software engineering tools, which are essential skills for computational mathematicians and statisticians. The products of this project – software, academic articles, conference presentations, and documentation – will also showcase the *right* way to do low discrepancy sampling, both within the QMC community and to the broader scientific community.

TABLE OF CONTENTS

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

*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.

*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.

QMCPy: Quasi-Monte Carlo Community Software

Contents

Quasi-Monte Carlo (QMC) methods replace independent and identically distributed (IID) points by low discrepancy (LD) points to improve computational efficiency. We, Fred Hickernell (FH[1], PI from Illinois Tech), Simon Mak (SM, PI from Duke U), and Sou-Cheng Terrya Choi (SCTC, Senior Personnel) propose to grow the seedling QMC software library QMCPy [15] into a sapling, which will be embraced by the community of QMC researchers and benefit an expanding base of QMC practitioners. QMCPy will grow to become

1. A tightly-connected suite of the best QMC software from multiple sources:
   a. LD sequence generators,
   b. QMC algorithms, such as multilevel algorithms, automatic stopping criteria, variance/ variation reduction, density estimation, and
   c. Illuminating use cases;
2. A user-friendly, consistent interface to the contributions of many scholars;
3. A proving ground for new QMC ideas, which will migrate to other popular packages;
4. Owned and cultivated by our QMC community; and
5. An easy on-ramp for those who are new to QMC.

Our partners include collaborators Mike McCourt (MM, SigOpt), Art Owen (AO, Stanford University), and Tim Sullivan (TS, Warwick University), as well as students, alumni, and friends.

Here we provide an overview of QMC, a description of our early efforts, our plans to grow QMCPy, the benefits of QMCPy, and why our team is the right one for this project.

1. Overview of QMC and LD Sampling

1.1. **Applications of (Q)MC and of IID and LD Sampling.** Many situations are best understood by mathematical models that include *randomness*. Examples include Bayesian statistical inference [26, 30], financial risk [34, 71], particle transport [36, 139, 145], and uncertainty quantification [38, 137]. Simulations use random vectors to generate a myriad of possible outcomes. The statistical properties of these *sample* outcomes—such as their means (averages) and probability densities—can effectively estimate the *population* quantities. This is the (Q)MC process: IID sampling for simple MC and LD sampling for QMC.

1.2. **LD Versus IID Sampling.** The quantity of interest, Y (e.g., option payoff, pixel intensity, flow velocity at certain location), is commonly expressed as a function of a d-dimensional vector random variable $\boldsymbol{X}$, i.e., $Y = f(\boldsymbol{X})$. Here, $\boldsymbol{X}$ follows a cumulative distribution function $F : \Omega \subseteq \mathbb{R}^d \to [0, 1]$. When f has a complicated form[2], the population mean (multivariate integral), μ, cannot be computed analytically. But, it may be estimated by the sample mean, $\hat{\mu}_n$:

$$\mu := \mathbb{E}(Y) = \int_\Omega f(\boldsymbol{x}) \, \mathrm{d}F(\boldsymbol{x}) \approx \hat{\mu}_n := \frac{1}{n}\sum_{i=1}^n Y_i = \frac{1}{n}\sum_{i=1}^n f(\boldsymbol{X}_i). \tag{1}$$

One wants to choose $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ to make $|\mu - \hat{\mu}_n| \le \varepsilon$.

[1]Initials of personnel are hyperlinks to their full names.

[2]In our notation f denotes an integrand, F denotes a probability distribution, and ϱ denotes a probability density.

Fig. 1 displays $n = 64$ IID, LD, and grid points intended to mimic $\mathcal{U}[0,1]^2$, the standard uniform distribution in dimension $d = 2$. The random points on the left are independent and do not know about each other. The multivariate distribution of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{IID}}{\sim} \mathcal{U}[0,1]^2$ is $F_n(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = F(\boldsymbol{x}_1) \cdots F(\boldsymbol{x}_n) = x_{11}x_{12}x_{21} \cdots x_{n2}$, where $F(\boldsymbol{x}) = x_1 x_2$ is the bivariate uniform distribution.

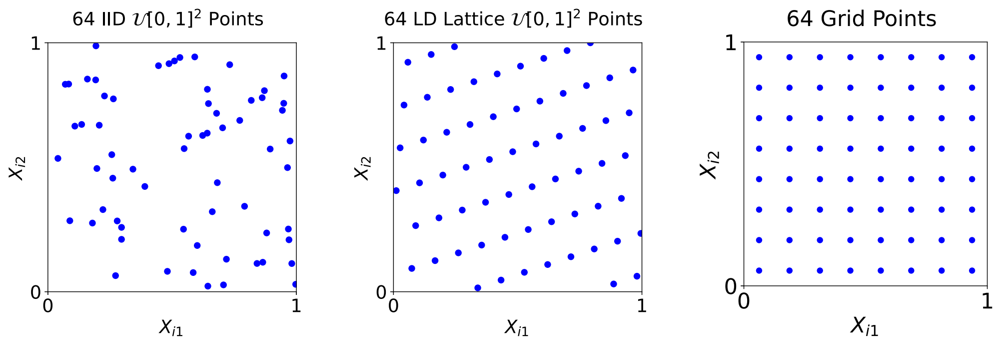

Figure 1. IID points (left), LD lattice points (center), and grid points (right). The LD points have fewer gaps and clusters of points than either the IID or grid points.

The LD points in the center of Fig. 1, $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \overset{\text{LD}}{\sim} \mathcal{U}[0,1]^2$, are carefully coordinated. They could be deterministic or random. They mimic the target distribution F (in this case bivariate uniform) by making the empirical distribution function of $\boldsymbol{X}_1, \ldots \boldsymbol{X}_n$—denoted $F_{\{\boldsymbol{X}_i\}_{i=1}^n}$—close to F. (The empirical distribution assigns equal probability to each point.) A *discrepancy*, $D(\{\boldsymbol{X}_i\}_{i=1}^n, F)$, measures the magnitude of $F - F_{\{\boldsymbol{X}_i\}_{i=1}^n}$. LD points make the discrepancy small.

The grid points on the right in Fig. 1 have only $\sqrt{n} = 8$ equally spaced values in each coordinate direction, whereas the LD points have $n = 64$ equally spaced values. Grids in large dimension d, have only $n^{1/d}$ values in each coordinate direction. They clump in lower dimensional projections.

1.3. **Efficiency Benefits from LD Sampling.** The root mean squared error of the sample mean, $\hat{\mu}_n$ when IID samples are used is

$$(2) \qquad \sqrt{\mathbb{E}\big[|\mu - \hat{\mu}_n|^2\big]} = \frac{\operatorname{std}(Y)}{\sqrt{n}}, \quad \operatorname{std}(Y) = \sqrt{\int_\Omega |f(\boldsymbol{x}) - \mu|^2 \, \mathrm{d}F(\boldsymbol{x})}, \qquad \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \overset{\text{IID}}{\sim} F.$$

The smoothness required of f is minimal, and the error bound has no curse of dimensionality (assuming that $\operatorname{std}(Y)$ does not explode with dimension), but the convergence rate is a modest $\mathcal{O}(n^{-1/2})$.

For LD sampling the absolute error has a deterministic upper bound of

$$(3) \qquad |\mu - \hat{\mu}_n| \le D(\{\boldsymbol{X}_i\}_{i=1}^n, F) \, \|f - \mu\|_{\mathcal{F}}, \qquad \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \overset{\text{LD}}{\sim} F.$$

The Banach space $\mathcal{F}$ requires somewhat more smoothness than the L^2 requirement for IID sampling, e.g., L^2 mixed partial derivatives of up to order one in each coordinate direction. The discrepancy, $D(\{\boldsymbol{X}_i\}_{i=1}^n, F)$, corresponds to the norm of the cubature error functional [41], and is typically $\mathcal{O}(n^{-1+\delta})$ for well-chosen LD sequences, where δ is arbitrarily small and positive.

Fig. 2 displays results from QMCPy for a computational physics example of Keister [63],

$$(4) \qquad \mu = \int_{\mathbb{R}^d} \cos(\|\boldsymbol{t}\|_2) \exp(-\|\boldsymbol{t}\|_2^2) \, \mathrm{d}\boldsymbol{t},$$

for the case $d = 5$, $\mu \approx 1.135$ under several absolute error tolerances, ε. QMCPy increases n until

$$(5) \qquad |\mu - \hat{\mu}_n| \le \varepsilon$$

is satisfied according to its stopping criteria. Both the number of function values and the computation time increase like $\mathcal{O}(\varepsilon^{-2})$ for IID sampling and $\mathcal{O}(\varepsilon^{-1-\delta})$ for LD sampling as ε decreases.

The advantage of LD sampling is illustrated by the sharp divergence of the function values and times required as ε decreases. For $\varepsilon = 0.01$ there is a hundred fold contrast between the n and time required for IID versus LD.

The orders of the computational cost dependence on ε for IID and LD sampling are *independent of* d. This is not the case for tensor product rules and grid sampling. If the derivatives of f up to total order rd exist, then tensor product rules may provide $|\mu - \hat{\mu}_n| \le \varepsilon$ at a computational cost of $\mathcal{O}(\varepsilon^{-d/r})$. Increased smoothness, r, helps but cannot overcome the exponential growth of the cost with d.

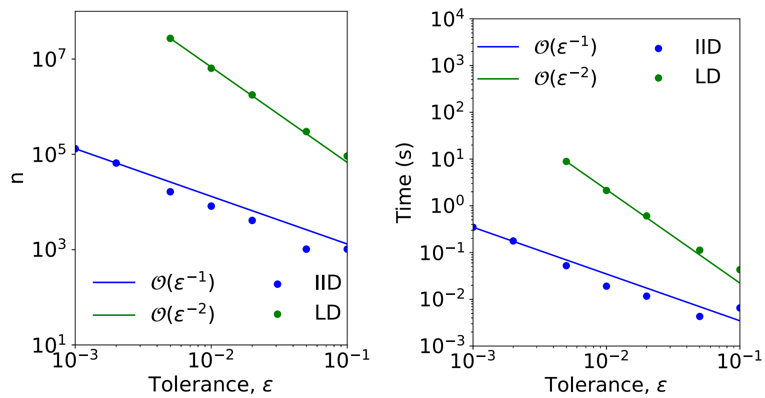

Figure 2. Number of function values (left) and run time (right) required to compute the Keister integral (4) using QMCPy. LD sampling is substantially more efficient than IID sampling, especially as the error tolerance decreases.

To harness the efficiency of LD sampling via QMC methods, practitioners need quality software that is easy to use. QMC researchers need a showcase for their latest work. We need QMCPy.

2. QMCPy Now

2.1. The Start of QMCPy. During the summer of 2018 FH began discussing with other QMC researchers, including AO, how we could combine our software efforts into a community owned library. The better existing QMC software efforts now and then include the following:

BRODA: Sobol' sequences in C, MATLAB, and Excel [67],
Burkhardt: various QMC software in C++, Fortran, MATLAB, & Python [5],
GAIL: Automatic stopping criteria in MATLAB by FH, SCTC and collaborators [13],
LatNet Builder: Generating vectors/matrices for lattices and digital nets [73],
MATLAB: Sobol' and Halton sequences, commercial [143],
ML(Q)MC: Multi-Level (Quasi-)Monte Carlo routines in C++, MATLAB, Python, and R [32],
MPS: Magic Point Shop, lattices and Sobol' sequences [109],
OpenTURNS: Open source initiative for the Treatment of Uncertainties, Risks 'N Statistics in Python [111],
Owen: Randomized Halton sequences in R by AO [118],
PyTorch: Scrambled Sobol' sequences [125],
QMC4PDE: QMC for elliptic PDEs with random diffusion coefficients [69],
qrng: Sobol' and Halton sequences in R [51],
Robbe: LD Sequences in Julia [129],
SSJ: Stochastic Simulation in Java [72], and
UQLab: Framework for Uncertainty Quantification in MATLAB [95].

Each software package above focuses on just certain aspects of QMC. Some packages focus on the generation of LD sequences, some packages focus on fundamental QMC algorithms, and other packages focus on particular applications of QMC. Most of the software packages listed above are the initiatives of individual research groups, not a shared community effort.

Fruitful discussions in 2018 led to further talks with MM in early 2019. MM's company, Silicon Valley startup SigOpt, offered to fund the early stage development of QMCPy, which would combine the efforts of several different research groups into a cutting edge package. Python 3 was chosen as the language because of its popularity among a broad spectrum of potential users, especially those in the high tech industry. Aleksei Sorokin (AS), a co-terminal BS applied mathematics and MS data science student at Illinois Tech, was hired to create the QMCPy code. QMCPy was released in the summer of 2020. QMCPy may be installed using `pip` and imported via `from qmcpy import *`.

2.2. **QMCPy Architecture.** The aforementioned discussions produced a skeleton for QMCPy consisting of four major abstract classes:

- `DiscreteDistribution` for generating LD and IID sequences,
- `TrueMeasure` to accommodate more general distributions or measures,
- `Integrand` to define the particular function, f, of interest, and
- `StoppingCriterion` to determine when to stop the simulation.

An auxiliary class, `AccumulateData`, keeps track of intermediate and final results.

2.2.1. `DiscreteDistribution` and `TrueMeasure`. LD sequences are constructed by creating the object and then generating points. The code below gives the points in the center panel of Fig. 1.

```
ld = qmcpy.Lattice(dimension = 2)  #define a discrete LD distribution
points = lattice.gen_samples(n = 64)  #construct points
```

All LD generators provide points designed to mimic the distribution $\mathcal{U}[0,1]^d$. QMCPy also has `Sobol` [21] and `Halton` [37] LD generators whose syntax are comparable. For comparison purposes, QMCPy has standard uniform and normal psuedo-random generators adopted from `numpy`. Our LD generators are extensible, meaning that one may generate additional points while reusing the original ones. For lattice and Sobol' points the preferred values of n are powers of 2. QMCPy LD generators are randomized by default to ensure that no points lie on the boundary of the unit cube and to improve the order of convergence of $\hat{\mu}_n$ to μ [115].

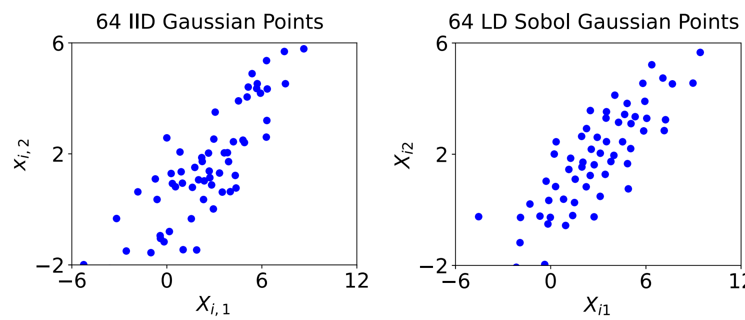

Figure 3. IID Gaussian (left) and Sobol' points transformed to mimic a Gaussian distribution (right). The LD points better represent the Gaussian distribution than the IID points.

The `TrueMeasure` class automates the transformation required to construct good points that mimic distributions other than standard uniform. The inverse distribution is often used.

```
gaussian_ld = qmcpy.Gaussian(qmcpy.Lattice(2), mean = [3,2], covariance =
↪  [[9,5],[5,4]])  #specify the distribution parameters
points = gaussian_ld.gen_samples(n = 64)  #construct points
```

Fig. 3 displays IID and Sobol' points transformed to mimic the Gaussian distribution with mean $\begin{pmatrix}3\\2\end{pmatrix}$ and covariance matrix $\begin{pmatrix}9&5\\5&4\end{pmatrix}$. Transformed low discrepancy points may or may not be low discrepancy with respect to the new distribution, depending on how one defines the discrepancy [78]. But, transformed low discrepancy points often outperform IID points in Monte Carlo calculations.

2.2.2. `Integrand` and `StoppingCriterion`. For computing expectations and multivariate integration as described in (1), one must specify the integrand. The original form of the problem may not be convenient for computation. For example, the Keister integral in (4) can be thought of as an integral of g with respect to the Gaussian distribution with zero mean and covariance $\mathsf{I}/2$:

$$\begin{aligned}(6)\quad \mu &= \int_{\mathbb{R}^d} \underbrace{\pi^{d/2}\cos(\|\boldsymbol{t}\|_2)}_{g(\boldsymbol{t})}\, \underbrace{\pi^{-d/2}\exp(-\|\boldsymbol{t}\|_2^2)\,\mathrm{d}\boldsymbol{t}}_{\mathcal{N}(\boldsymbol{0},\mathsf{I}/2)} = \int_{\mathbb{R}^d} \underbrace{\cos(\|\boldsymbol{t}\|_2)\exp(-\|\boldsymbol{t}\|_2^2)}_{h(\boldsymbol{t})}\, \underbrace{\mathrm{d}\boldsymbol{t}}_{\text{Lebesgue}} \\ &= \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \qquad \text{for an appropriate transformation } \boldsymbol{t} = \boldsymbol{T}(\boldsymbol{x}).\end{aligned}$$

Alternatively, μ can be thought of as an integral of h with respect to Lebesgue measure. QMCPy can compute the integral either way.

The first way begins by constructing the Gaussian transformed LD `TrueMeasure` instance `gs` as discussed in the previous section. Next, one defines the function g as in (6) above. `CustomFun` constructs an f for which the integral, μ, can be written in terms of the uniform distribution, which the LD sequence mimics. `CustomFun` automatically constructs the transformation $\boldsymbol{T}$ in (6).

```
gs = qmcpy.Gaussian(qmcpy.Sobol(5), covariance = 1/2)     #choose the Gaussian
↪  distribution
def g(t):  #your desired integrand, calculations must be vectorized
        d = t.shape[1]
        g_val = np.pi**(d/2) * np.cos(np.sqrt((t**2).sum(1)))
        return g_val  #size n vector
f = qmcpy.CustomFun(gs, custom_fun = g)
sc = qmcpy.CubQMCSobolG(f, abs_tol = 1e-3)   #stopping criterion must match
↪  points
solution = sc.integrate()
```

The final stage of the computation requires the construction of a `StoppingCriterion` instance, `sc`. The one here is due to PI FH and Lluís Antoni Jiménez Rugama (LlAJR) [45] and is based on Walsh transformations of the sampled integrand data. Invoking the `integrate` method of `sc` yields $\hat{\mu}_n$ satisfying (5).

Another way to compute the Keister integral, (4), is to think of it as an integral with respect to the Lebesgue measure. Again `CustomFun` makes the proper variable transformation. The answer returned by QMCPy is the same as the first way, but this second way takes about twice the time.

```
lb = qmcpy.Lebesgue(qmcpy.Sobol(5), lower_bound=-np.inf, upper_bound=np.inf)
↪  #choose the Lebesgue distribution
def h(t):  #your desired integrand, calculations must be vectorized
         norm = np.sqrt((t**2).sum(1))  #
         h_val = np.cos(norm) * np.exp(-norm**2)
         return h_val  #size n vector
 f = qmcpy.CustomFun(lb, custom_fun = h)
 sc = qmcpy.CubQMCSobolG(f, abs_tol = 1e-3)  #stopping criterion must match
 ↪  points
 solution = sc.integrate()
```

2.3. **QMCPy Options.** QMCPy has several options for users to choose:

- `backends` for `Sobol` from qrng [51], PyTorch [125] and MPS [109];

- `backends` for `Lattice` from GAIL [13] and MPS [109], and generators from LatNet Builder [73];
- `StoppingCriteria` based on the Central Limit Theorem (CLT) for IID sampling and random replications of LD sampling; a more rigorous criterion for IID sampling based on a Berry-Essen inequality [47]; criteria for LD sampling based on fast transforms of integrand data [45, 56, 59]; most of these have arisen from the NSF-supported research of FH, SCTC and their collaborators, AO, Lan Jiang (LJ), LlAJR, Jagadeeswaran Rathinavel (JR), and implemented in GAIL [13].
- Specification of relative error tolerances as well as absolute error tolerances;
- Multi-level Monte Carlo methods; and
- A few use cases, i.e., pre-programmed `Integrand` instances.

2.4. **QMCPy Support.** QMCPy is hosted on GitHub [15]. Bugs can be reported and features requested at the issues page. Pull requests can be made by those who wish to add features. All features are documented [16] using ReadtheDocs. Doctests ensure that features work as expected. Features are illustrated by Jupyter notebooks. FH gave a tutorial on QMC software—with a focus on QMCPy—at MCQMC 2020 [39]. A Google Colaboratory Notebook [40] is available so that those watching can try out QMCPy themselves in real time. FH, SCTC, AO, AS, and MM wrote a series of blogs [14] to introduce QMC to the broader community of QMC.

3. QMCPy in the Future

QMCPy has made a good start. Yet, much must be done to establish the critical mass of algorithms and collaborators that will make QMCPy self-sustaining. Here is our plan.

QMCPy will aggregate the best QMC software with a clear structure and a consistent user interface. When a potentially better or interesting LD generator, algorithm, or use case arises, it will be a simple matter to swap out the old for the new and measure the performance while all other pieces remain the same.

3.1. **Feature Rich LD Sequence Generators.** Although QMCPy already includes the most popular LD sequences, their implementation is rather rigid. Also, some newer sequences are missing.

At present, the digital shifting and scrambling recommended for Sobol' sequences are hard coded in the `Sobol backend`. For flexibility, the generating matrices for the Sobol' sequence should be specified independently of the random digital scrambling and shifting. *We will rectify this*, and thereby allow folks to try scrambled BRODA generators [67], which is not yet possible.

Sobol' sequences are a kind of digital sequence [21]. LatNet Builder [73] can construct the generating matrices for polynomial lattices, another kind of digital sequence. When our Sobol' sequence generator allows the generating matrices to be specified independently of the randomization, users will be able to explore *what advantages polynomial lattices may have* over Sobol' sequences.

The usual method for randomly scrambling and shifting digital sequences (such as the Sobol' sequence) is linear scrambling [53, 97]. A richer scrambling is the original nested uniform scrambling (NUS) proposed by AO [114]. NUS requires more complex code and a longer computation time, but it has a CLT [82]. We will *implement NUS* in QMCPy.

For integrands with higher order smoothness, higher order digital sequences [19, 20] provide faster convergence rates as $n \to \infty$, which translate into lower computational cost rates as $\varepsilon \to 0$. An example is computing probabilities over subsets of the sample space involving smooth densities. *Higher order digital sequences will be implemented* in QMCPy.

Discrepancies may be defined in terms of a symmetric, positive definite kernel, K. If we identify $K(\boldsymbol{t},\boldsymbol{x})$ as some inner product of point distributions, i.e., $K(\boldsymbol{t},\boldsymbol{x}) = \left\langle F_{\{\boldsymbol{t}\}}, F_{\{\boldsymbol{x}\}}\right\rangle_{\mathcal{M}}$ for all $\boldsymbol{t},\boldsymbol{x} \in \Omega$, then the discrepancy between two distributions is the norm of the distance between the two, as shown by PI FH in [42]:

$$D^2(F,G) := \|F-G\|_{\mathcal{M}}^2 = \int_{\Omega\times\Omega} K(\boldsymbol{t},\boldsymbol{x})\, \mathrm{d}(F(\boldsymbol{t})-G(\boldsymbol{t}))\mathrm{d}(F(\boldsymbol{x})-G(\boldsymbol{x})) = D^2(G,F). \tag{7}$$

When G is the empirical distribution for $\{\boldsymbol{X}_i\}_{i=1}^n$, we have an example of the discrepancy in (3):

$$D^2(\{\boldsymbol{X}_i\}_{i=1}^n, F) := D^2\big(F_{\{\boldsymbol{X}_i\}_{i=1}^n}, F\big) = \int_{\Omega\times\Omega} K(\boldsymbol{t},\boldsymbol{x})\,\mathrm{d}F(\boldsymbol{t})\mathrm{d}F(\boldsymbol{x}) - \frac{2}{n}\sum_{i=1}^n \int_\Omega K(\boldsymbol{t},\boldsymbol{X}_i)\,\mathrm{d}F(\boldsymbol{t}) + \frac{1}{n^2}\sum_{i,j=1}^n K(\boldsymbol{X}_i,\boldsymbol{X}_j). \tag{8}$$

In error bound (3) for LD cubature, K plays the role of a reproducing kernel for the Hilbert space, $\mathcal{F}$. A Bayesian version of (3) assumes random integrands with covariance kernel K [43]. QMCPy *will implement discrepancies* based on a variety of popular kernels, including those in [41].

One approach to constructing LD sequences, especially for small n, is to optimize numerically [78, 148]. This is non-trivial because the discrepancy as a function $\{\boldsymbol{X}_i\}_{i=1}^n$ is multi-modal. Changing the order of the vectors in the sequence $\{\boldsymbol{X}_i\}_{i=1}^n$ leaves the discrepancy unchanged.

We will *implement some numerical optimization algorithms* for constructing low discrepancy designs. Since evaluating $D(\{\boldsymbol{X}_i\}_{i=1}^n, F)$ generally requires $\mathcal{O}(n^2)$ operations, this numerical optimization effort will be cost effective when n is modest, and the kernel, K is smooth enough to allow $D(\{\boldsymbol{X}_i\}_{i=1}^n, F)$ to decay quickly to 0 with n. The corresponding integration problems would be characterized by a smooth f that is expensive to evaluate, hence n must be small. The discrepancies of typical LD sequences typically decay no faster than $\mathcal{O}(n^{-1})$. Hence we need numerical optimization to construct LD points for discrepancies defined with smooth kernels.

Markov Chain Monte Carlo (MCMC) differs from IID sampling in the sense that one often uses *one* infinite dimensional run, $\boldsymbol{X}_{\mathbb{N}} = (X_1, X_2, \dots)$, generated by a Markov chain to obtain X_i that approach the desired, very complicated, distribution as $i \to \infty$. Owen and Tribble [119] proposed a QMC variant of the Metropolis algorithm (a popular MCMC method) to speed up the convergence of the X_i to the target distribution. For QMC Metropolis, one needs completely uniformly distributed (CUD) sequences. A sequence $\{U_1, U_2, \dots\} \in [0,1]^{\mathbb{N}}$ is CUD, if for every positive integer d, the points $\boldsymbol{Z}_i = (U_i, \dots, U_{i+d-1}) \in [0,1]^d$ satisfy $\lim_{n\to\infty} D(\{\boldsymbol{Z}_i\}_{i=1}^n, F_{\mathrm{unif},d}) = 0$, where $F_{\mathrm{unif},d}$ is the standard uniform distribution in d dimensions.

CUD sequences must satisfy much stronger conditions than LD sequences. We *will implement CUD sequences* in QMCPy. This will allow further exploration into the situations where QMC-based Markov chain algorithms outperform ordinary MCMC algorithms.

3.2. **More Extensive Algorithms.**

3.2.1. *Variance/Variation Reduction Through Importance Sampling and Control Variates.* The QMC error bound in (3) contains the factor $\|f-\mu\|_{\mathcal{F}}$, which may be called the *variation* of the integrand. For randomized QMC, the mean squared error of $\hat{\mu}_n$ is its variance. Substantial error reduction can sometimes be obtained by rewriting the integral to reduce the variation or variance. Examples include rare event modeling [130] and Bayesian inference [131].

Importance sampling [113] is a variance/variation reduction tool whereby one samples from a new distribution (the *importance* distribution) to place greater weight where the integrand varies more. Since LD sampling at its heart is $\mathcal{U}[0,1]^d$, LD sampling to mimic a different distribution can be thought of as choosing a variable transformation. We describe it that way. The desired integral originally defined in terms of g is

$$\mu = \int_{\mathbb{R}^d} g(\boldsymbol{t})\,\varrho(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} = \int_{[0,1]^d} f_{\boldsymbol{T}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \qquad \text{for an appropriate transformation } \boldsymbol{t} = \boldsymbol{T}(\boldsymbol{x}). \tag{9}$$

Since $\boldsymbol{T}$ is not unique—as shown in the Keister example (6)—the challenge lies in finding a good choice for $\boldsymbol{T}$ that makes the variation, $\|f_{\boldsymbol{T}}-\mu\|_{\mathcal{F}}$, small. Key developments include the annealed importance sampler [104], the bridge sampler [29], and AO's safe importance sampler [113]. Recent work includes [102], which uses deep neural networks for importance sampling in image rendering, and [55], where PI SM extends importance sampling for covariate balancing in causal inference.

Control variates or de-trending [34] is another powerful variance/variation reduction technique. The idea is to obtain a new integrand, $\tilde{f}$, by subtracting from the original integrand a linear combination of functions (control variates) whose integrals are zero. If $\tilde{f}$ has smaller variation/variance than the original integrand, f, then substantial cost savings are possible. Control variates are widely used in financial pricing and stochastic simulations. Important works (among many) include [105], which investigates control variate remedies, and the QMC control variates papers [43, 46] by AO, PI FH, and collaborators. Recently, [103] introduced the neural control variates method, which uses a neural network to learn an optimal choice of control variate for variance reduction.

We will implement these variance/variation reduction methods in QMCPy and attempt to automate the choices of transformation and control variates. This will provide a comprehensive and accessible toolbox of variance reduction methods for a broad range of applications.

3.2.2. *Multilevel QMC (MLQMC).* For high or infinite dimensional integration problems, the computational cost of the integrand is often proportional to d, which makes the cost of the sample mean $\mathcal{O}(dn)$. Multilevel (Q)MC [31] decomposes the original integral into a sum of several integrals with different numbers of variables, $d_1 < \cdots < d_L$. The total cost is then $\mathcal{O}(d_1 n_1 + \cdots + d_L n_L)$. When done well, error criterion (5) can be met for a decreasing sequence $n_1 > \cdots > n_L$, which results in an overall large cost savings compared to a single level (Q)MC algorithm requiring $\mathcal{O}(d_L n_1)$ operations. *We will strengthen* QMCPy's rudimentary MLQMC, including extending the theory and implementation of the single level stopping criteria developed by PI FH, SCTC, and their collaborators [45, 47, 49, 56, 59] to the multilevel case.

3.2.3. *Density Estimation.* Much of the QMC literature focuses on estimating the population *mean*, $\mathbb{E}(Y)$. In many problems, however, one may be interested in estimating the underlying probability *density* of $Y = f(\boldsymbol{X})$. The state-of-the-art in the classical setting, is kernel density estimation [135], which uses IID samples $Y_1, \ldots, Y_n$ to approximate ϱ by:

$$\hat{\varrho}_n(y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} \mathfrak{K}\left(\frac{y - Y_i}{h}\right), \tag{10}$$

where $\mathfrak{K}$ is a symmetric kernel, and $h > 0$ is the bandwidth parameter.

Given the improvement that QMC gives for mean estimation, a natural question is whether QMC also provides gains for density estimation. A recent paper [1] by AO and coauthors explores this by replacing the IID samples $\{\boldsymbol{X}_i\}_{i=1}^n$ used to generate $Y_i = f(\boldsymbol{X}_i)$ with randomized LD samples [116]. They showed that this new QMC density estimator enjoys an improved rate of convergence over the standard estimator (10), both theoretically and in numerical experiments.

We will implement this QMC kernel density estimation method in QMCPy. This will provide users from a broad range of fields, from engineers to data scientists, a useful tool for accurate density estimation when sample data is limited or expensive to collect.

3.3. **A Broader Spectrum of Use Cases.**

3.3.1. *Sobol' Indices.* In sensitivity analysis, Sobol' indices, which are defined as multidimensional integrals, play an important role. We will implement the work done by LlAJR and others [33, 58] to adaptively compute Sobol' indices. AO will encourage his PhD student, Chris Hoyt, to help.

3.3.2. *Big Data Analytics.* With advances in experimental technology, mathematical modeling, and computing power, big data is ubiquitous. A key challenge is making use of this rich source of data; learning algorithms need to be highly *efficient* and *scalable* to extract information for real-time decision-making. The development of such algorithms is an important research direction in statistics and computer science.

One way to make these algorithms scalable is to iteratively train the model on small batches of the data, typically sampled uniformly at random. This is widely used to scale-up many powerful machine learning algorithms, e.g., stochastic gradient descent (SGD, [4]) and stochastic gradient boosting [28]. Consider SGD as an illustrative example. The objective is to minimize the loss function $L(\theta;\mathcal{X}) = N^{-1}\sum_{m=1}^{N} l(\theta;\boldsymbol{X}_m)$ over model parameters $\theta \in \mathbb{R}^q$, where $\mathcal{X} = \{\boldsymbol{X}_m\}_{m=1}^N \subset \mathbb{R}^d$ is the large training data. Standard gradient descent methods for optimization [106] are impractical here, since they require evaluation of the full gradient $N^{-1}\sum_{m=1}^{N} \nabla l(\theta;\boldsymbol{X}_m)$, which is very expensive for N large (i.e., for big data). Mini-batch SGD [4] approximates this gradient using $\mathcal{X}_s^{[l]} \subset \mathcal{X}$, which are subsamples of size $n \ll N$ taken IID and uniformly from the big data $\mathcal{X}$. The following descent steps are then iterated until convergence:

$$\theta^{[l+1]} \leftarrow \theta^{[l]} - \eta\left(\frac{1}{n}\sum_{\boldsymbol{X}\in\mathcal{X}_s^{[l]}} l(\theta;\boldsymbol{X})\right), \quad l = 1,2,\cdots, \tag{11}$$

where η 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 [140].

SGD has a notable weakness: since its 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 size n, SGD returns parameter estimates that can be very far from θ^*. One way to address this is by carefully choosing a small LD dataset that well-represents the big data $\mathcal{X}$. This is referred to as "data squashing" [112] by AO, and has garnered attention in recent years. Leverage-score subsampling was introduced in [83] to scale up linear regression for large datasets. A similar idea of *coresets* has been explored in computer science [7], with recent developments in [3] and [54] for cluster analysis and logistic regression. There is also recent work [17, 52] on applying QMC methods to generative adversarial networks [35], a popular machine learning model. PI SM has also done work in this area [65, 88–90], most notably on *support points*, which has been applied to mechanical engineering [27], computer simulations [122], and experimental design [66].

We will implement these existing methods in QMCPy, and explore their effectiveness in a suite of big data problems. There is little work on which data squashing procedure is most appropriate for SGD optimization. Our preliminary results show that a LD subsample exists for *any* big dataset $\mathcal{X} \subset \mathbb{R}^d$ that achieves a noise ball radius of $\mathcal{O}\{(\log n)^{3d+1}/n^2\}$. We will investigate this further and implement this in QMCPy.

3.3.3. *Posterior Sampling for Complex Bayesian Models.* Bayesian methods have become quite popular in recent decades, primarily due to their ability to fit complicated models via sampling [30]. Such methods have largely relied on MCMC sampling for exploration of the posterior distribution F, which captures information on model parameters. However, with the rise of big data and complex machine learning models, the posterior distribution F can be expensive to evaluate, which makes MCMC sampling very time-consuming [61]. This is compounded by the highly correlated nature of MCMC samples, which reduces the information provided by each sample [79]. For such problems, a LD posterior sampling method can greatly improve model learning given a time budget.

One approach for LD posterior sampling is to minimize the kernel discrepancy $D(\{\boldsymbol{X}_i\}_{i=1}^n, F)$ in (3) for samples $\{\boldsymbol{X}_i\}_{i=1}^n$. This is known as *kernel herding* [10]. Kernel herding has a key weakness for posterior sampling: the discrepancy, $D(\{\boldsymbol{X}_i\}_{i=1}^n, F)$, to be optimized requires an analytic expression for the integral $\int K(t,\boldsymbol{X}_i)\,\mathrm{d}F(t)$ (see (8)), which is unattainable for complex posteriors F. To address this, [9] proposes a new *Stein reproducing kernel* for discrepancy minimization:

$$\begin{aligned}K_{st}(\boldsymbol{t},\boldsymbol{x}) = \nabla_{\boldsymbol{t}}\cdot\nabla_{\boldsymbol{x}}K(\boldsymbol{t},\boldsymbol{x}) + \nabla_{\boldsymbol{t}}K(\boldsymbol{t},\boldsymbol{x})\cdot\nabla_{\boldsymbol{x}}\log\mathrm{d}F(\boldsymbol{x})\\ + \nabla_{\boldsymbol{x}}K(\boldsymbol{t},\boldsymbol{x})\nabla_{\boldsymbol{t}}\log\mathrm{d}F(\boldsymbol{t}) + K(\boldsymbol{t},\boldsymbol{x})\nabla_{\boldsymbol{t}}\log\mathrm{d}F(\boldsymbol{t})\cdot\nabla_{\boldsymbol{x}}\log\mathrm{d}F(\boldsymbol{x}),\end{aligned} \tag{12}$$

where K is a symmetric, positive definite kernel, and ∇ and $\nabla\cdot$ are the gradient and divergence operators. The key advantage is that the integral in question $\int K_{st}(t, \boldsymbol{X})\, \mathrm{d}F(t)$ evaluates to 0 for all $\boldsymbol{X}$, which allows one to generate LD posterior samples via minimization of discrepancy $D(\{\boldsymbol{X}_i\}_{i=1}^n, F)$. This is a promising method which speeds up learning of complex Bayesian models.

We will implement Stein points in QMCPy. PI SM has worked on a variety of complex Bayesian modeling problems, from climatology [93] to aerospace engineering [8, 94, 149], and we will explore the effectiveness of Stein points for such problems. We will also explore further extensions of Stein points for approximate Bayesian computation, a popular class of Bayesian methods for population genetics and epidemiology.

3.3.4. *Uncertainty Quantification.* Uncertainty quantification is the science of quantifying, characterizing, tracing, and managing uncertainty in computational and real worlds systems [136, 137]. Since data from such systems are typically expensive to simulate or collect, a key focus in this area is the sampling design for data collection, which can greatly benefit from QMC methods.

A recent application of QMC for uncertainty quantification is to compute the expectation of a functional of solution of a partial differential equation with random coefficients [38]. This has broad applications in engineering problems, e.g., fluid flow through a porous medium. QMCPy will implement a typical use case, which may show how to link QMCPy code with a PDE solver from another package. The PI SM has done extensive work on uncertainty quantification, for rocket engine design [8, 76, 77, 94, 149], neuroscience modeling [147], and active learning [92]. We will explore the performance of LD sampling for these applications.

3.3.5. *Ray Tracing.* Image rendering is a visually stunning application of QMC [64]. QMCPy will implement a simple example to showcase the advantage of LD over IID sampling.

3.3.6. *Probabilistic Numerics (PN).* PN assumes that the input function or solution of a mathematical problem is an instance of a random process. This allows one to apply a probabilistic approach to construct credible intervals for the solution. PI FH along with his former PhD student JR developed a fast Bayesian cubature method [56] using lattice LD sampling. Our collaborator TS is a PN expert.

Bayesian optimization applies this PN perspective to finding the minimum of an objective function. The next locations to sample the objective function are chosen as the optimum of an acquisition function, which in the Bayesian context takes the form of a d-variate integral, where d is the number of future points to be sampled times the dimension of those points. MM illustrates the advantages of LD sampling for computing an acquisition function in his blog [14, qEI with QMCPy]. The observed function values are noisy. As in the Keister example in Sect. 1.3, LD sampling shows a substantial advantage over IID sampling.

Bayesian optimization is a relatively unexplored QMC application area. PI SM has worked on black-box optimization [11, 91], most notably on a hierarchical Bayesian optimization method [11]. We will explore the *efficacy of LD sampling* in Bayesian optimization and other PN problems.

3.4. **Enhanced Performance.** Python's advantages are ease of coding and the rapidly growing code base. However, Python code does not execute particularly quickly. Python can be made to execute more quickly by re-writing it in C. For example, this is done for some of the PyTorch code. We will speed up critical QMCPy code by *re-writing it in C*.

There have been attempts at parallel implementations of LD sequences [75, 81, 110, 133, 146], but there is no generally accepted parallel LD software library. There are issues with load balancing. Does one give different processors different segments of the same LD sequence? Or does one combine the results of several randomized LD sequences, each from a different processor.

We will *implement LD sequences taking advantage of multiple cores* of the same CPU. We will also explore the possibility of GPU implementations. Python parallel computing resources [124], such as Numba [108] will prove invaluable.

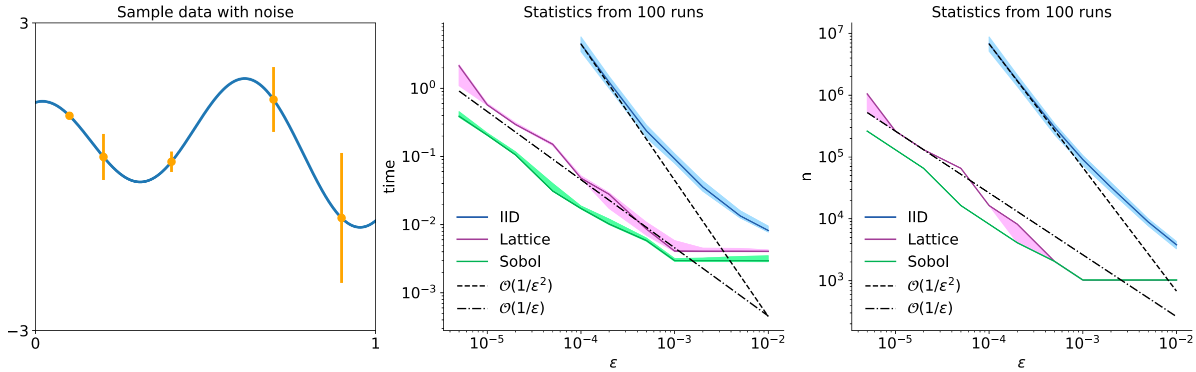

Figure 4. A one dimensional objective function (left) and the number of function values (center) and run time (right), required to compute the acquisition function in [14, qEI with QMCPy] to the desired tolerance, ε. LD sampling is much more efficient than IID sampling, especially as the error tolerance decreases.

4. Broader Impacts

4.1. **QMCPy as a Proving Ground.** QMCPy will *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. Those who develop LD generators need to test them on a variety of use cases and as key components of various QMC algorithms. Those with new QMC algorithms need to test them with the best generators. Those with juicy use cases want to try the best that QMC offers. Because large scale software packages like `scipy` or MATLAB, do not offer QMCPy's options, QMCPy *will attract a significant number of QMC researchers as contributors.*

Although large scale software packages cannot adopt every new QMCPy wrinkle, QMCPy algorithms attracting broad interest can be folded into these large scale packages. FH had this experience when MATLAB adopted his TOMS LD generators from [53].

By making QMC methods easily accessible, QMCPy will *introduce new application areas* to the benefits of LD sampling. This will lead to *LD sampling being incorporated* into the software packages used by practitioners in those disciplines. This includes, e.g., PyMC3 [132] and PyStan [141], which are popular Python packages for Bayesian statistical modeling and probabilistic machine learning.

As a library which strives to connect the best of QMC software to the scientific community, the success of QMCPy will be gauged by how well such a connection is made. We will thus measure project success via a variety of *engagement* metrics (e.g., number of package downloads, Github pull requests), as well as the degree of integration into large scale packages used by practitioners.

4.2. **New QMC Theory.** The history of QMC is marked by new applications leading to new theoretical insight. The success of QMC for a 360-dimensional financial risk application [121] spurred the theoretical study of QMC's effectiveness for problems with much higher dimensions than was previously thought feasible. This resulted in dozens of articles (see [22, 107] and the citations therein). These high dimensional applications gave impetus to the development of multi-level [31] and multivariate decomposition [70] methods. Looking forward, QMCPy's success in new application areas will *raise new theoretical questions* that we and others will address.

4.3. **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 [39] prompted a vigorous discussion on the PyTorch issues site [126] that migrated to the `scipy` issues site [134]. AO, FH, and other QMC researchers seem to have 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 [117].

In these discussions, it was pointed out that UQLab [95], OpenTurns [111], and other packages routinely drop the first Sobol' point, a bad, but understandable practice. The arguments we provided to the PyTorch and `scipy` developers answered 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 pointed out in [126] that randomized PyTorch Sobol' points fell on the boundaries of $[0,1]^d$, when they never should. MM later discovered that this was due to PyTorch not using double precision. LlAJR discovered that the Sobol' scrambling in MATLAB was incorrect. This was rectified in R2017a. These two examples highlight how having a larger community using a software library leads to higher quality code.

4.4. **Educating and Mentoring Cross-Disciplinary Computational Researchers.** Computational mathematics and statistics nearly always requires the use of others' code, hopefully in the form of well-developed software packages. New scholars needs to be trained not only how to use such packages, but how to contribute to them as well. QMCPy students supported by this project will learn to write clean, efficient code that fits the package architecture, is documented, and passes doctests. Some will learn how to combine code from different packages and even different languages. QMCPy students will learn about repositories and the software engineering tools that need to become second nature, just like Beamer became second nature to many a generation ago, and LaTeX two generations ago.

As with past NSF projects, in seeking summer undergraduate students and graduate students we will give preference to underrepresented minorities, women, and students from colleges where research experiences are rare. As noted in Sect. 5.1, we have had significant success in mentoring students. Many undergraduates have enrolled in graduate programs. Five out of FH's fifteen students earning PhDs are women, three of whom have entered academia.

The senior personnel on this project include one woman (SCTC) and one early-career scholar (SM). Also, TS is early-career. The backgrounds of our senior personnel and collaborators include folks with backgrounds in mathematics, statistics, and computer science. The students that we mentor will also learn to think from these different perspectives.

4.5. **Disseminating Our Work.** We will publish our theoretical and practical work in a variety of mathematics, statistics, and computer science journals and conference proceedings. We will present our work at conferences targeting theoreticians and practitioners. We will offer tutorials. Students will present at conferences and during group meetings, as part of their education.

5. Results from Prior NSF Support

5.1. **NSF-DMS-1522687, *Stable, Efficient, Adaptive Algorithms for Approximation and Integration*, $270,000, August 2015 – July 2018.** Gregory E. Fasshauer (GEF, co-PI) and FH (PI) led this project, and SCTC contributed as senior personnel. Other major contributors were FH's research students Yuhan Ding (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 @ 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, 12, 13, 23, 24, 33, 43–45, 48–50, 56, 58–60, 74, 80, 96, 98–101, 127, 128, 150–153].

5.1.1. *Intellectual Merit from Prior NSF Support.*
Adaptive Algorithms for Univariate Problems: FH, SCTC, YD, XT, YZ and collaborators developed several adaptive algorithms for univariate integration, function approximation, and optimization [12, 18, 23, 144, 150]. The function approximation and optimization algorithms constructed by

FH, SCTC, YD, and XT in [12] are *locally adaptive*—the nonuniform sampling density is influenced by the function data. For function approximation, the computational cost of $\mathcal{O}\left(\sqrt{\|f''\|_{1/2}/\varepsilon}\right)$, where ε is the error tolerance, and is essentially optimal. The 1/2-quasinorm $\|f''\|_{1/2}$ may be much smaller than $\|f''\|_{\infty}$ for peaky functions.

Globally Adaptive Cubature Based on LD Sequences: FH, LlAJR, DL, and JR developed globally adaptive algorithms for approximating $\int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ based on LD sequences [45, 49, 59]. These stopping criteria are now in QMCPy. Two common LD sequences are integration lattice nodes and digital sequences [22]. The error bounds underlying the adaptive cubatures developed by FH, LlAJR, DL track the Fourier coefficients of the sampled function values on these LD sequences. FH and JR base their automatic Bayesian cubature on credible intervals, where the hyper-parameters of the priors are treated by empirical Bayes (maximum likelihood estimation), full Bayes, and/or cross-validation. FH and JR chose covariance kernels that matched the LD sequences and reduced the computational cost to $\mathcal{O}(n\log(n))$, making Bayesian cubature practical.

Multivariate Function Approximation: FH, YD, and LlAJR and collaborators investigated function approximation problems for Banach spaces, $\mathcal{F}$, defined by series representations [24, 25]. For example, the bases can be general multivariate polynomials. Different definitions of a cone, $\mathcal{C}$, of functions were defined, all describing a reasonable behavior of the series coefficients. Adaptive function approximation algorithms constructed were shown to be essentially optimal.

5.1.2. *Broader Impacts from Prior NSF Support.*

Publications, Conference Participation, Conference Organization, and Leadership: 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, a long-running annual industrial statistics conference. FH gave an invited tutorial at MCQMC 2016 [43], a biennial conference for which he serves on the steering committee. FH was a program leader for the SAMSI 2017–18 Quasi-Monte Carlo (QMC) Program. FH received the 2016 Joseph F. Traub Prize for Achievement in Information-Based Complexity. In recognition of his research leadership, FH was appointed the director of Illinois Tech's new Center for Interdisciplinary Scientific Computation in 2017. In 2018, FH was appointed Vice Provost for Research.

GAIL Software: The results of this research have been implemented in GAIL, our open source MATLAB library hosted on Github. This software has input parsing, input validation, unit tests, inline documentation, and demonstrations. GAIL makes it easier for practitioners to try our new adaptive algorithms. SCTC has been key in this effort. GAIL has been used in the yearly graduate course in Monte Carlo methods taught by FH and YD.

Boosting the STEM Workforce: GEF, FH, and SCTC mentored a number of research students associated with this project. Female students include YD, LJ, JL, XT, and Xiaoyang Zhao (MS 2017). Mentees include undergraduate students involved more than a dozen Brazilian Science Mobility Program students in the summers of 2015 and 2016, plus eight other students (two female) from Illinois Tech, Biola U, U Minnesota, Macalester U, NUS, Colorado School of Mines. All but one of these eight have enrolled in graduate programs. All students have learned how to conduct theoretical and/or practical computational mathematics research.

5.2. **NSF CSSI Frameworks 2004571 (Subaward WSU20076).** ***X-Ion Collisions with a Statistically and Computationally Advanced Program Envelope (X-SCAPE),*** **\$696,442, July 2020 – June 2014.** 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 statistical and computational framework, which integrates volumes of data from diverse experiments with computer simulations from candidate theories. The X-SCAPE collaboration is a multi-disciplinary team of physicists, computer scientists and statisticians, is engaged in the construction of such an open-source framework. SM is a Duke co-PI in this ongoing collaboration (which started in the summer of 2020), and is responsible for leading the statistical developments on the project.

5.2.1. *Intellectual Merit from Prior NSF Support.* The X-SCAPE (previously JETSCAPE) collaboration has developed the first open-source, end-to-end, modular simulation framework for the high energy sector of heavy-ion and p-p collisions and a Bayesian statistical framework to rigorously compare any similar, complex event generator with extensive experimental data. This framework consists of several state-of-the-art modules implementing each successive stage of a heavy-ion collision: the initial state of two nuclei, the pre-equilibrium stage, the fluid-dynamical evolution, modules to generate high momentum partons, several modules that describe the shower of these partons in the viscous medium, the conversion of the quark-gluon plasma into an interacting gas of hadrons, the conversion of partonic jets into jets of hadrons. The development of the JETSCAPE framework and event generator led to several firsts: the calculation of the suppression of jets, high momentum hadrons from jets, and high momentum heavy hadrons from jets. It is only in the JETSCAPE analysis that the theory prediction is encapsulated by the data driven approach. The JETSCAPE manual has already appeared online [123] and been submitted to *Comp. Phys. Comm.*, and two papers [6, 68] and several conference papers [62, 120, 138, 142] have also appeared.

5.2.2. *Broader Impacts from Prior NSF Support.* The primary broader impacts of the JETSCAPE/X-SCAPE collaboration have been in the training of its graduate students and postdocs. Through regular weekly meetings, collaboration gatherings and joint projects, a multi-disciplinary, multi-institutional environment is fostered between the experimental and theoretical physicists, computer scientists and statisticians. This continuous cross-talk has imposed a far broader scope on the education and training of these individuals than would be possible in any monodisciplinary academic environment. Physicists learn about techniques in computer science and statistical analysis, while computer scientists encounter a unique physical problem requiring extensive software engineering, which in turn presents a difficult challenge in statistical model to data comparison. Beyond its students and postdocs, the collaboration strives to influence the training of the wider US nuclear physics workforce, through its winter school and workshops. Each of these annual gatherings consist of about 30 students and young postdocs, who are then trained in the essential components of Monte Carlo event generation, basics of heavy-ion collisions, and extensive hands on tutorials involving the JETSCAPE/X-SCAPE framework. Beyond this is the growing awareness in the nuclear physics community of the existence and utility of the JETSCAPE/X-SCAPE framework, as witnessed by the ever growing downloads and views of the JETSCAPE GitHub site [57].

6. Strengths of This Team and Collaboration Plan

We have a well-constructed team that combines senior personnel with diverse backgrounds, career stages, and institutions. Together with our students, collaborators, and friends, we will grow QMCPy into what it should become while providing new theory to underpin our new algorithms.

The senior personnel—together with our students—will have regular video conference meetings to share our progress and brainstorm next steps. These will be held both at our own institutions and via video conference among institutions.

6.1. **Senior Personnel.** FH has been the lead PI on the GAIL [13] MATLAB software project that contains many of the stopping criteria incorporated into QMCPy. His expertise is in the numerical analysis of QMC and other algorithms for multivariate problems. He has also developed several 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. FH will lead the activities at Illinois Tech and oversee QMCPy's architecture. He will also lead the development of higher order nets, net scrambling, and multilevel methods.

SM is a statistician who became quite familiar with QMC, having served as a Working Group Leader on the aforementioned SAMSI program led by FH. He is an Assistant Professor of Statistical Science at Duke, and his interests are in Bayesian modeling, big data analytics, and computer experiments. SM provides expertise on statistical methodology and applications: he is an Associate Editor for *Technometrics* (a top journal for engineering statistics), the recipient of the Statistics in Physical Engineering Science (SPES) Award and the Mary G. and Joseph Natrella Scholarship from the American Statistical Association, as well as multiple student paper awards. He also has experience in software library development, having authored four R packages [84–87] on the Comprehensive R Archive Network (CRAN). SM will lead the activities at Duke and oversee the big data learning and Bayesian inference efforts.

SCTC is a computational mathematician and engineer by training, currently serving as the Chief Data Scientist at Kamakura Corporation, a leading financial risk-software company. She has been a research assistant, and now associate, professor with the Department of Applied Mathematics at Illinois Tech since 2013. She has co-led the GAIL and QMCPy projects with FH since 2013 and was particularly responsible for educating the team in best software engineering practices for numerical software. SCTC will lead the effort to ensure that QMCPy is well-tested, well-documented, and meets the needs of the business world. She is a co-winner of the SIAM (Society for Industrial and Applied Mathematics) Activity Group on Linear Algebra (SIAG/LA) Prize with Michael A. Saunders and Chris C. Paige, an 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.

6.2. **Students.** The PhD students supported by this project will work with the senior personnel to address major algorithmic and theoretical issues. Promising QMCPy simulations may suggest open theoretical questions. Alternatively, new theoretical insights may be incorporated as new features in QMCPy. Our goal is to have theoretically justified, practically impactful algorithms. Undergraduate students will focus on new features or use cases that can be implemented mostly over the course of a summer. They will be mentored by the senior personnel and the PhD students. All students will learn good practices for contributing to a long term software project.

6.3. **Collaborators.** AO has engaged with the PIs in conversations about QMC for many years. He is particularly an expert in randomized QMC and the use of low discrepancy points for Markov chain Monte Carlo. AO 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. AO will encourage his student to help out with the implementation of Sobol' indices.

MM convinced his company to fund the early development of QMCPy. He wanted to spread the advantages of low discrepancy sampling to the tech industry. During the first year of QMCPy's development, MM advised us on what would benefit high tech. Although SigOpt cannot fund QMCPy further, MM will advise us on the continued development of QMCPy. He will also help us spread the word among his network in the machine learning community.

TS will provide expertise on two application domains for QMCPy. The first is PN, a Bayesian statistical approach to numerical tasks such as cubature. Here, we construct a statistical posterior distribution over the value of the solution to reflect the discretization uncertainty inherited from the finite computational budget. LD sampling is an attractive way to interrogate PN distributions at significantly lower cost than IID or MCMC sampling. The second is the use of QMC methods to train metamodels for heterogeneous (i.e. mixed atomistic-continuum) systems, which is of particular interest in Warwick's EPSRC Centre for Doctoral Training in Modelling of Heterogeneous Systems.

References Cited

[1] Abdellah AB, L'Ecuyer P, Owen AB, Puchhammer F (2018) Density estimation by randomized quasi-Monte Carlo. arXiv preprint arXiv:180706133

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

[5] Burkhardt J (2020) URL http://people.sc.fsu.edu/~jburkardt/

[6] Cao S, Park C, Barbieri RA, Bass SA, Bazow D, Bernhard J, Coleman J, Fries R, Gale C, He Y, et al (2017) Multistage Monte Carlo simulation of jet modification in a static medium. Phys Rev C 96(2):024,909

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

[8] 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

[9] Chen WY, Mackey L, Gorham J, Briol FX, Oates CJ (2018) Stein points. arXiv preprint arXiv:180310161

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

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

[12] 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

[13] Choi SCT, Ding Y, Hickernell FJ, Jiang L, Jiménez Rugama LlA, Li D, Jagadeeswaran R, Tong X, Zhang K, Zhang Y, Zhou X (2020) GAIL: Guaranteed Automatic Integration Library (versions 1.0–2.3.1). MATLAB software, http://gailgithub.github.io/GAIL_Dev/

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

[15] Choi SCT, Hickernell FJ, Jagadeeswaran R, McCourt M, Sorokin A (2020) QMCPy: A quasi-Monte Carlo Python library. https://qmcsoftware.github.io/QMCSoftware/, DOI 10.5281/zenodo.3964489.

[16] Choi SCT, Hickernell FJ, Jagadeeswaran R, McCourt MJ, Sorokin A (2020) URL https://qmcpy.readthedocs.io/en/latest/?badge=latest

[17] Chong MJ, Forsyth D (2020) Effectively unbiased FID and inception score and where to find them. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 6070–6079

[18] 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

[19] Dick J (2010) On quasi-Monte Carlo rules achieving higher order convergence. In: L'Ecuyer P, Owen A (eds) Monte Carlo and Quasi-Monte Carlo Methods 2008, Springer-Verlag, Berlin, pp 73–96

[20] Dick J (2011) Higher order scrambled digital nets achieve the optimal rate of the root mean square error for smooth integrands. Ann Statist 39:1372–1398

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

[22] Dick J, Kuo F, Sloan IH (2013) High dimensional integration — the Quasi-Monte Carlo way. Acta Numer 22:133–288, DOI 10.1017/S0962492913000044

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

[24] 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

[25] 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

[26] Efron B, Hastie T (2016) Computer Age Statistical Inference. Cambridge University Press, Cambridge

[27] Fernandez-Zelaia P, Joseph VR, Kalidindi SR, Melkote SN (2018) Estimating mechanical properties from spherical indentation using Bayesian approaches. Materials & Design 147:92–105

[28] Friedman JH (2002) Stochastic gradient boosting. Comput Statist & Data Anal 38(4):367–378

[29] Gelman A, Meng XL (1998) Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. Statist Sci pp 163–185

[30] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian Data Analysis. CRC Texts in Statistical Science, Chapman & Hall

[31] Giles M (2015) Multilevel Monte Carlo methods. Acta Numer 24:259–328, DOI 10.1017/S096249291500001X

[32] Giles M (2020) Multi-level (quasi-)Monte Carlo software. URL https://people.maths.ox.ac.uk/gilesm/mlmc/

[33] 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

[34] Glasserman P (2004) Monte Carlo Methods in Financial Engineering, Applications of Mathematics, vol 53. Springer-Verlag, New York

[35] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp 2672–2680

[36] Haghighat A (2014) Monte Carlo Methods for Particle Transport. CRC Press, Boca Raton, FL

[37] Halton JH (1960) On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numer Math 2:84–90

[38] 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

[39] Hickernell FJ (2020) Quasi-Monte Carlo software. URL https://mcqmc20.web.ox.ac.uk/programme

[40] Hickernell FJ, Sorokin A (2020) Quasi-Monte Carlo (QMC) software in QMCPy Google Colaboratory notebook. URL https://colab.research.google.com/drive/1Nh6OhZ5w6trTFjyauIf0YAXD5JReyKpM?usp=sharing

[41] Hickernell FJ (1998) A generalized discrepancy and quadrature error bound. Math Comp 67:299–322, DOI 10.1090/S0025-5718-98-00894-1
[42] Hickernell FJ (1999) Goodness-of-fit statistics, discrepancies and robust designs. Statist Probab Lett 44:73–78
[43] 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
[44] Hickernell FJ, Jagadeeswaran R (2019) Comment on "Probabilistic integration: A role in statistical computation?". Statist Sci 34:23–28, DOI 10.1214/18-STS660
[45] 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]
[46] Hickernell FJ, Lemieux C, Owen AB (2005) Control variates for quasi-Monte Carlo. Statist Sci 20:1–31, DOI 10.1214/088342304000000468
[47] Hickernell FJ, Jiang L, Liu Y, Owen AB (2013) Guaranteed conservative fixed width confidence intervals via Monte Carlo sampling. 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 105–128, DOI 10.1007/978-3-642-41095-6
[48] 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
[49] 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
[50] 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
[51] Hofert M, Lemieux C (2017) qrng R package. URL https://cran.r-project.org/web/packages/qrng/qrng.pdf
[52] Hofert M, Prasad A, Zhu M (2018) Quasi-random number generators for multivariate distributions based on generative neural networks. Stat 1050:1
[53] Hong HS, Hickernell FJ (2003) Algorithm 823: Implementing scrambled digital nets. ACM Trans Math Software 29:95–109, DOI 10.1145/779359.779360
[54] Huggins J, Campbell T, Broderick T (2016) Coresets for scalable Bayesian logistic regression. In: Advances in Neural Information Processing Systems, pp 4080–4088
[55] Huling JD, Mak S (2020) Energy balancing of covariate distributions. arXiv preprint arXiv:200413962
[56] Jagadeeswaran R, Hickernell FJ (2019) Fast automatic Bayesian cubature using lattice sampling. Stat Comput 29:1215–1229, DOI 10.1007/s11222-019-09895-9
[57] JETSCAPE Developers (2020) JETSCAPE. URL https://github.com/JETSCAPE
[58] 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
[59] 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

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

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

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

[63] Keister BD (1996) Multidimensional quadrature algorithms. Computers in Physics 10:119–122, DOI 10.1063/1.168565

[64] 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

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

[66] Krishna A, Joseph VR, Ba S, Brenneman WA, Myers WR (2020) Robust experimental designs for model calibration. arXiv preprint arXiv:200800547

[67] Kucherenko S (2020) BRODA. URL https://www.broda.co.uk/index.html

[68] 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

[69] Kuo FY, Nuyens D (2016) Application of quasi-Monte Carlo methods to elliptic PDEs with random diffusion coefficients – a survey of analysis and implementation. Found Comput Math 16:1631–1696, URL https://people.cs.kuleuven.be/~dirk.nuyens/qmc4pde/

[70] Kuo FY, Nuyens D, Plaskota L, Sloan IH, Wasilkowski GW (2017) Infinite-dimensional integration and the multivariate decomposition method. J Comput Appl Math 326:217–234, DOI 10.1016/j.cam.2017.05.031

[71] L'Ecuyer P (2009) Quasi-Monte Carlo methods with applications in finance. Finance Stoch 13:307–349

[72] L'Ecuyer P (2020) SSJ: Stochastic Simulation in Java. URL https://github.com/umontreal-simul/ssj

[73] L'Ecuyer P, Munger D (2018) LatNet builder. URL https://umontreal-simul.github.io/latnetbuilder/

[74] Li D (2016) Reliable quasi-Monte Carlo with control variates. Master's thesis, Illinois Institute of Technology

[75] Li JX, Mullen GL (2000) Parallel computing of a quasi-Monte Carlo algorithm for valuing derivatives. Parallel Comput 26:64–653

[76] 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

[77] 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

[78] Li Y, Kang L, Hickernell FJ (2020) Are transformed low discrepancy points also low discrepancy? In: Fan J, Pan J (eds) Contemporary Experimental Design, Multivariate Analysis and Data Mining (Festschrift in Honour of Professor Kai-Tai Fang), Springer Nature Switzerland AG, pp 69–92

[79] Link WA, Eaton MJ (2012) On thinning of chains in MCMC. Methods in Ecology and Evolution 3(1):112–115

[80] Liu J (2018) Adaptive quadrature with a general error criterion. Master's thesis, Illinois Institute of Technology

[81] Liu KI, Hickernell FJ (2004) A scalable low discrepancy point generator for parallel computing. In: Cao J, Yang LT, Guo M, Lau F (eds) Parallel and Distributed Processing and Applications, Springer-Verlag, Heidelberg, Lecture Notes in Computer Science, vol 3358, pp 257–262
[82] Loh WL (2003) On the asymptotic distribution of scrambled net quadrature. Ann Stat 31:1282–1324
[83] Ma P, Mahoney MW, Yu B (2015) A statistical perspective on algorithmic leveraging. J Mach Learn Res 16(1):861–911
[84] Mak S (2018) `minimaxdesign`: Minimax and Minimax Projection Designs. R package version 0.1.4
[85] Mak S (2018) `support`: Support Points. R package version 0.1.4
[86] Mak S (2019) `atmopt`: Analysis-of-Marginal-Tail-Means. R package version 0.1.0
[87] Mak S (2020) `cmenet`: Bi-Level Selection of Conditional Main Effects. R package version 0.1.1
[88] Mak S, Joseph VR (2017) Projected support points, with application to optimal MCMC reduction. arXiv preprint arXiv:170806897
[89] Mak S, Joseph VR (2018) Minimax and minimax projection designs using clustering. J Comput and Graph Statist 27(1):166–178
[90] Mak S, Joseph VR (2018) Support points. Ann Statist 46(6A):2562–2592
[91] Mak S, Wu CFJ (2019) Analysis-of-marginal-tail-means (ATM): A robust method for discrete black-box optimization. Technometrics 61(4):545–559
[92] Mak S, Xie Y (2018) Maximum entropy low-rank matrix recovery. IEEE J Sel Topics in Signal Process 12(5):886–901
[93] 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
[94] 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
[95] 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
[96] 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
[97] Matoušek J (1998) On the L_2-discrepancy for anchored boxes. J Complexity 14:527–556
[98] 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
[99] McCourt MJ, Fasshauer GE, Kozak D (2018) A nonstationary designer space-time kernel. In: 32nd Conference on Neural Information Processing Systems (NIPS 2018)
[100] 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
[101] 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
[102] Müller T, McWilliams B, Rousselle F, Gross M, Novák J (2019) Neural importance sampling. ACM Trans Graph 38(5):145:1–145:19, DOI 10.1145/3341156, URL http://doi.acm.org/10.1145/3341156
[103] Müller T, Rousselle F, Novák J, Keller A (2020) Neural control variates. arXiv:200601524

[104] Neal RM (2001) Annealed importance sampling. Stat Comput 11(2):125–139
[105] Nelson BL (1990) Control variate remedies. Oper Res 38(6):974–992
[106] Nocedal J, Wright S (2006) Numerical Optimization. Springer Science & Business Media
[107] Novak E, Woźniakowski H (2010) Tractability of Multivariate Problems Volume II: Standard Information for Functionals. No. 12 in EMS Tracts in Mathematics, European Mathematical Society, Zürich
[108] Numba Developers (2020) Numba. URL http://numba.pydata.org
[109] Nuyens D (2017) Magic point shop. URL https://people.cs.kuleuven.be/~dirk.nuyens/qmc-generators/
[110] Ökten G, Srinivasan A (2002) Parallel quasi-Monte Carlo methods on a heterogeneous cluster. In: Fang KT, Hickernell FJ, Niederreiter H (eds) Monte Carlo and Quasi-Monte Carlo Methods 2000, Springer-Verlag, Berlin, pp 406–421
[111] OpenTURNS Developers (2020) An open source initiative for the Treatment of Uncertainties, Risks 'N Statistics. URL http://www.openturns.org
[112] Owen A (2003) Data squashing by empirical likelihood. Data Min Knowl Disc 7(1):101–113
[113] Owen A, Zhou Y (2000) Safe and effective importance sampling. J Amer Statist Assoc 95(449):135–143
[114] Owen AB (1995) Randomly permuted (t, m, s)-nets and (t, s)-sequences. In: Niederreiter H, Shiue PJS (eds) Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Springer-Verlag, New York, Lecture Notes in Statistics, vol 106, pp 299–317
[115] Owen AB (1997) Scrambled net variance for integrals of smooth functions. Ann Stat 25:1541–1562
[116] Owen AB (2000) Monte Carlo, quasi-Monte Carlo, and randomized quasi-Monte Carlo. In: Monte-Carlo and Quasi-Monte Carlo Methods 1998, Springer, pp 86–97
[117] Owen AB (2020) On dropping the first Sobol' point. arXiv preprint arXiv:200808051
[118] Owen AB (2020) Randomized Halton sequences in R. URL http://statweb.stanford.edu/~owen/code/
[119] Owen AB, Tribble SD (2005) A quasi-Monte Carlo Metropolis algorithm. Proc Natl Acad Sci 102:8844–8849
[120] 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
[121] Paskov S, Traub J (1995) Faster valuation of financial derivatives. J Portfolio Management 22:113–120
[122] Pronzato L, Zhigljavsky A (2020) Bayesian quadrature, energy minimization, and space-filling design. SIAM/ASA J Uncertain Quantif 8(3):959–1011
[123] Putschke J, Kauder K, Khalaj E, Angerami A, Bass S, Cao S, Coleman J, Cunqueiro L, Dai T, Du L, et al (2019) The JETSCAPE framework. arXiv preprint arXiv:190307706
[124] Python Developers (2020) Parallel processing and multiprocessing in Python. URL https://wiki.python.org/moin/ParallelProcessing
[125] PyTorch Developers (2020) PyTorch. URL https://pytorch.org
[126] PyTorch Developers (2020) PyTorch discussion on Sobol' sequence implementation. URL https://github.com/pytorch/pytorch/issues/32047
[127] 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
[128] 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

[129] Robbe P (2020) Low discrepancy sequences in Julia. URL https://github.com/PieterjanRobbe/QMC.jl
[130] Rubino G, Tuffin B (2009) Rare event simulation using Monte Carlo methods. John Wiley & Sons
[131] Salmerón A, Cano A, Moral S (2000) Importance sampling in Bayesian networks using probability trees. Comput Statist Data Anal 34(4):387–413
[132] Salvatier J, Wiecki TV, Fonnesbeck C (2016) Probabilistic programming in Python using PyMC3. PeerJ Computer Science 2:e55
[133] Schmid WC, Uhl A (2001) Techniques for parallel quasi-Monte Carlo integration with digital sequences and associated problems. Math Comput Simul 55(1-3):249–257
[134] SciPy Developers (2020) scipy discussion of Sobol' sequence implementation. URL https://github.com/scipy/scipy/pull/10844
[135] Silverman BW (1986) Density estimation for statistics and data analysis, vol 26. CRC press
[136] SmartUQ Developers (2020) SmartUQ. URL https://www.smartuq.com/resources/uncertainty-quantification/
[137] Smith RC (2014) Uncertainty Quantification: theory, implementation, and applications. SIAM, Philadelphia
[138] 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
[139] Spanier J (1995) Quasi-Monte Carlo methods for particle transport problems. In: Niederreiter H, Shiue PJS (eds) Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Springer-Verlag, New York, Lecture Notes in Statistics, vol 106, pp 121–148
[140] 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
[141] Stan Development Team (2017) PyStan: the Python interface to Stan. Version 216 00
[142] 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
[143] The MathWorks, Inc (2020) MATLAB R2020a. Natick, MA
[144] Tong X (2014) A guaranteed, adaptive, automatic algorithm for univariate function minimization. Master's thesis, Illinois Institute of Technology
[145] Veach E (1997) Robust Monte Carlo methods for light transport simulation. PhD thesis, Stanford University
[146] Wan JL, Lai K, Kolkiewicz AW, Tan K (2006) A parallel quasi-Monte Carlo approach to pricing multi-dimensional American options. Int J of High Performance Computing and Networking 4:321–330
[147] Wang H, Xie L, Cuozzo A, Mak S, Xie Y (2020) Uncertainty quantification for inferring hawkes networks. arXiv preprint arXiv:200607506
[148] Winker P, Fang KT (1997) Application of threshold accepting to the evaluation of the discrepancy of a set of points. SIAM J Numer Anal 34:2028–2042
[149] 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
[150] Zhang Y (2018) Guaranteed, adaptive, automatic algorithms for univariate integration: Methods, costs and implementation. PhD thesis, Illinois Institute of Technology
[151] Zhao X (2017) Simulating the Heston model via the QE method with a specified error tolerance. Master's thesis, Illinois Institute of Technology
[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

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

(b) APPOINTMENTS

2018 - present	Vice Provost for Research, Illinois Institute of Technology, Chicago, IL
2005 - 2020	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
1999 - 2005	Professor, Hong Kong Baptist University, Department of Mathematics, Kowloon
1995 - 1999	Associate Professor, Hong Kong Baptist University, Department of Mathematics, Kowloon
1989 - 2002	Department Head, Hong Kong Baptist College/University, Department of Mathematics, Kowloon
1987 - 1995	Senior Lecturer, Hong Kong Baptist College, Department of Mathematics, Kowloon
1985 - 1987	Lecturer, Hong Kong Baptist College, Kowloon
1981 - 1985	Assistant Professor, University of Southern California, Mathematics, Los Angeles, CA

(c) PRODUCTS

Products Most Closely Related to the Proposed Project

1. Jiménez Rugama L, Hickernell F. Springer Proceedings in Mathematics & Statistics. Cham: Springer International Publishing; 2016. Chapter Chapter 20, Adaptive Multidimensional Integration Based on Rank-1 Lattices. 407-422p. Available from: http://link.springer.com/10.1007/978-3-319-33507-0_20 DOI: 10.1007/978-3-319-33507-0_20
2. Hickernell F, Jiang L, Liu Y, Owen A. Springer Proceedings in Mathematics & Statistics. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. Chapter Chapter 5, Guaranteed Conservative Fixed Width Confidence Intervals via Monte Carlo Sampling. 105-128p. 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. 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

4. Choi ST, Ding Y, Hickernell FJ, Jiang L, Jiménez Rugama L, Li D, Rathinavel J, Tong X, Zhang K, Zhang Y, Zhou X. GAIL: Guaranteed Automatic Integration Library. [Internet]. Version 2.3. Chicago, IL: Illinois Institute of Technology; 2020 May . Available from: http://gailgithub.github.io/GAIL_Dev/
5. Choi ST, Hickernell FJ, Rathinavel J, McCourt MJ, Sorokin A. QMCPy: a quasi-Monte Carlo Python Library. Chciago, IL: Illinois Institute of Technology; 2020 August . Available from: https://qmcsoftware.github.io/QMCSoftware/}

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

1. Hickernell F. Springer Proceedings in Mathematics & Statistics. Cham: Springer International Publishing; 2018. Chapter Chapter 1, The Trio Identity for Quasi-Monte Carlo Error. 3-27p. 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. Goodness-of-fit statistics, discrepancies and robust designs. Statistics & Probability Letters. 1999; 44(1):73-78. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0167715298002934 DOI: 10.1016/S0167-7152(98)00293-4
3. Gilquin L, Jiménez Rugama L, Arnaud É, Hickernell F, Monod H, Prieur C. Iterative construction of replicated designs based on Sobol' sequences. Comptes Rendus Mathematique. 2017 January; 355(1):10-14. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1631073X16302576 DOI: 10.1016/j.crma.2016.11.013
4. 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
5. 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

(d) SYNERGISTIC ACTIVITIES

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. Mentoring of 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: Choi, Sou Cheng T.

ORCID: 0000-0002-6190-2986

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

(a) PROFESSIONAL PREPARATION

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

2020 - present Chief Data Scientist, Kamakura Corporation, Chicago, IL
2017 - present Research Associate Professor, Illinois Institute of Technology, 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

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--2.3.1). MATLAB Software. 2020 June; 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 0.4). Python Software. 2020 August; 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 M. 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 M. 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 M. 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

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

Biographical Sketch

Simon Mak
Department of Statistical Science, Duke University
email: sm769@duke.edu

Professional Preparation

Simon Fraser University	Statistics & Actuarial Science	BS 2013
Georgia Institute of Technology	Statistics	MS 2018
Georgia Institute of Technology	Industrial Engineering	PhD 2018
Georgia Institute of Technology	Industrial Engineering	Postdoc 2018 – 2019

Appointments

Assistant Professor, Department of Statistical Science, Duke University, 2019 – current.

Postdoctoral Fellow, School of Industrial and Systems Engineering, Georgia Institute of Technology, 2018 – 2019.

Publications

Publications closely related to proposal

- Ding, Y., Hickernell, F. J., Kritzer, P. and Mak, S. (2019). Adaptive approximation for multivariate linear problems with inputs lying in a cone. In Hickernell F. J., Kritzer P. (eds.), *Multivariate Algorithms and Information-Based Complexity*, Berlin/Boston: De Gruyter.
- Mak, S. and Wu, C. F. J. (2019). cmenet: a new method for bi-level variable selection of conditional main effects. *Journal of the American Statistical Association*, 114(526):844-856.
- Mak, S. and Wu, C. F. J. (2019). Analysis-of-marginal-Tail-Means (ATM) - a new method for discrete black-box optimization. *Technometrics*, 61(4):545-559.
- Mak, S. and Joseph, V. R. (2018). Support points. *Annals of Statistics*, 46(6A):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.

Additional publications

- Chang, Y.-H., Zhang, L., Yeh, S.-T., Wang, X. J., Mak, S., Sung, C.-L., Wu, C. F. J. and Yang, V. (2019). Kernel-smoothed proper orthogonal decomposition (KSPOD)-based emulation for prediction of spatiotemporally evolving flow dynamics. *AIAA Journal*, 57(12):5269-5280.

- Mak, S. and Joseph, V. R. (2018). Minimax and minimax projection designs using clustering. *Journal of Computational and Graphical Statistics*, 27(1):166-178.
- Mak, S. and Xie, Y. (2018). Maximum entropy low-rank matrix recovery. *IEEE Journal of Selected Topics in Signal Processing*, 12(5):886-901.
- Yeh, S.-T., Wang, X. J., Sung, C.-L., Mak, S., Chang, Y.-H., Wu, C. F. J. and Yang, V. (2018). Data-driven analysis and common proper orthogonal decomposition (CPOD)-based spatio-temporal emulator for design exploration. *AIAA Journal*, 56(6):2429-2442.
- Mak, S., Bingham, D. and Lu, Y. (2016). A regional compound Poisson process for hurricane and tropical storm damage. *Journal of the Royal Statistical Society: Series C*, 65(5):677–703.

Synergistic Activities

- Associate Editor, Technometrics (2019 – current)
- Developed four open-source R packages (`support`, `cmenet`, `minimaxdesign`, `atmopt`) on the Comprehensive R Archive Network (CRAN)
- 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 Mathematical Sciences Institute (SAMSI) (2017 – 2018)
- Mary G. and Joseph Natrella Scholarship recipient, awarded annually to two PhD students by the ASA for engagement and experience in statistical applications, and service and leadership in the statistics community (2017)

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. 2053714	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			23,032	
2.					
3.					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.0			0	
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	1.0			23,032	
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)				60,032	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				1,773	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				61,805	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT				0	
E. TRAVEL 1. DOMESTIC (INCL. U.S. POSSESSIONS)				6,000	
2. INTERNATIONAL				3,000	
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				0	
3. CONSULTANT SERVICES				0	
4. COMPUTER SERVICES				0	
5. SUBAWARDS				0	
6. OTHER				14,526	
TOTAL OTHER DIRECT COSTS				16,526	
H. TOTAL DIRECT COSTS (A THROUGH G)				87,331	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 54.0, Base:72805)					
TOTAL INDIRECT COSTS (F&A)				39,315	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				126,646	
K. FEE				0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				126,646	
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* Toni Allen	Date Checked	Date Of Rate Sheet	Initials - ORG

1 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET

YEAR 2

ORGANIZATION Illinois Institute of Technology	FOR NSF USE ONLY PROPOSAL NO. 2053714	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			23,953	
2.					
3.					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.0			0	
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	1.0			23,953	
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)				62,433	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				1,844	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				64,277	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT				0	
E. TRAVEL 1. DOMESTIC (INCL. U.S. POSSESSIONS)				6,240	
2. INTERNATIONAL				3,120	
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				0	
3. CONSULTANT SERVICES				0	
4. COMPUTER SERVICES				0	
5. SUBAWARDS				0	
6. OTHER				15,107	
TOTAL OTHER DIRECT COSTS				17,187	
H. TOTAL DIRECT COSTS (A THROUGH G)				90,824	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 54.0, Base:75717)					
TOTAL INDIRECT COSTS (F&A)				40,887	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				131,711	
K. FEE				0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				131,711	
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* Toni Allen	Date Checked	Date Of Rate Sheet	Initials - ORG

2 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET

YEAR 3

ORGANIZATION Illinois Institute of Technology	FOR NSF USE ONLY			
	PROPOSAL NO. 2053714	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			24,911	
2.					
3.					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.0			0	
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	1.0			24,911	
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)				64,931	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				1,918	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				66,849	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT				0	
E. TRAVEL 1. DOMESTIC (INCL. U.S. POSSESSIONS)				6,490	
2. INTERNATIONAL				3,245	
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				0	
3. CONSULTANT SERVICES				0	
4. COMPUTER SERVICES				0	
5. SUBAWARDS				0	
6. OTHER				15,711	
TOTAL OTHER DIRECT COSTS				17,874	
H. TOTAL DIRECT COSTS (A THROUGH G)				94,458	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 54.0, Base:78747)					
TOTAL INDIRECT COSTS (F&A)				42,523	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				136,981	
K. FEE				0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				136,981	
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* Toni Allen	Date Checked	Date Of Rate Sheet	Initials - ORG

3 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET

Cummulative

ORGANIZATION Illinois Institute of Technology			FOR NSF USE ONLY PROPOSAL NO. 2053714	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			71,896	
2.					
3.					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)					
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	3.0			71,896	
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)				187,396	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				5,535	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				192,931	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT				0	
E. TRAVEL 1. DOMESTIC (INCL. U.S. POSSESSIONS)				18,730	
2. INTERNATIONAL				9,365	
F. PARTICIPANT SUPPORT COSTS					
1. STIPENDS $ 0					
2. TRAVEL 0					
3. SUBSISTENCE 0					
4. OTHER 0					
TOTAL NUMBER OF PARTICIPANTS (0.0) TOTAL PARTICIPANT COSTS				0	
G. OTHER DIRECT COSTS					
1. MATERIALS AND SUPPLIES				6,243	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				0	
3. CONSULTANT SERVICES				0	
4. COMPUTER SERVICES				0	
5. SUBAWARDS				0	
6. OTHER				45,344	
TOTAL OTHER DIRECT COSTS				51,587	
H. TOTAL DIRECT COSTS (A THROUGH G)				272,613	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)					
TOTAL INDIRECT COSTS (F&A)				122,725	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				395,338	
K. FEE					
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				395,338	
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* Toni Allen	Date Checked	Date Of Rate Sheet	Initials - ORG

C *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

BUDGET JUSTIFICATION

Senior Personnel

Prof. Fred J. Hickernell, Professor of Applied Mathematics at Illinois Tech, will lead 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 stopping criteria. The one-month summer salary compensates his time on the project.

Dr. Sou-Cheng Choi, Chief Data Scientist at Kamakura Corporation will provide in-kind, voluntary expertise in software engineering, documentation,

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 description. 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 underpinnings of the new algorithms to be included in QMCPy.

The summer undergraduate student stipends will fund smaller scale, but crucial components of QMCPy. These include, for example, novel use cases foun d 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, 24.5%; faculty summer salary, 7.7%; staff salary, 26.6%; and student stipends, 0.0%.

Equipment

Travel

The senior personnel and research students will disseminate their results and introduce a broader audience to QMCPy through attendance at US ($6,000 in Y1) and international ($3,000 in Y1) conferences devoted to QMC and its applications.

Participant Support

Other Direct Costs

<u>Materials and Supplies</u>

Modest resources are needed for software license and website/blog maintenance fees ($2,000 in Y1).

<u>Tuition</u>

The PhD student(s) will be supported at 9 credits/yr ($1,614 per credit hour in Y1) so that they may continue their studies while working on this grant.

Indirect Costs

IIT's current federally negotiated indirect cost rate (agreement date 02/24/2020) 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	87,331	90,824	94,458	272,613
Indirect Costs	39,315	40,887	42,523	122,725
Total Costs	126,646	131,711	136,981	395,338
Modified Base	*72,805*	*75,717*	*78,747*	*227,269*

SUMMARY PROPOSAL BUDGET

YEAR 1

ORGANIZATION Duke University				FOR NSF USE ONLY PROPOSAL NO. 2053715	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			7,222		
2.						
3.						
4.						
5.						
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.0			0		
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	0.5			7,222		
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				20,475		
4. (0) UNDERGRADUATE STUDENTS				0		
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0		
6. (0) OTHER				0		
TOTAL SALARIES AND WAGES (A + B)				27,697		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				3,756		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				31,453		
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				7,805		
TOTAL OTHER DIRECT COSTS				7,805		
H. TOTAL DIRECT COSTS (A THROUGH G)				44,258		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 61.0, Base:36453)						
TOTAL INDIRECT COSTS (F&A)				22,236		
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				66,494		
K. FEE				0		
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				66,494		
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

1 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET

YEAR 2

ORGANIZATION Duke University	FOR NSF USE ONLY PROPOSAL NO. 2053715	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			7,438	
2.					
3.					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.0			0	
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	0.5			7,438	
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				20,996	
4. (0) UNDERGRADUATE STUDENTS				0	
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0	
6. (0) OTHER				0	
TOTAL SALARIES AND WAGES (A + B)				28,434	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				4,819	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				33,253	
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				8,151	
TOTAL OTHER DIRECT COSTS				8,151	
H. TOTAL DIRECT COSTS (A THROUGH G)				46,404	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 61.0, Base:38253)					
TOTAL INDIRECT COSTS (F&A)				23,334	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				69,738	
K. FEE				0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				69,738	
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

2 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET

YEAR 3

ORGANIZATION Duke University	FOR NSF USE ONLY			
	PROPOSAL NO. 2053715	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			7,662	
2.					
3.					
4.					
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.0			0	
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	0.5			7,662	
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				21,627	
4. (0) UNDERGRADUATE STUDENTS				0	
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0	
6. (0) OTHER				0	
TOTAL SALARIES AND WAGES (A + B)				29,289	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				4,969	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				34,258	
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				8,548	
TOTAL OTHER DIRECT COSTS				8,548	
H. TOTAL DIRECT COSTS (A THROUGH G)				47,806	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) MTDC (Rate: 61.0, Base:39258)					
TOTAL INDIRECT COSTS (F&A)				23,947	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				71,753	
K. FEE				0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				71,753	

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

3 *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

SUMMARY PROPOSAL BUDGET

Cummulative

ORGANIZATION Duke University				FOR NSF USE ONLY PROPOSAL NO. 2053715	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			22,322		
2.						
3.						
4.						
5.						
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)						
7. (1) TOTAL SENIOR PERSONNEL (1 - 6)	1.5			22,322		
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				63,098		
4. (0) UNDERGRADUATE STUDENTS				0		
5. (0) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0		
6. (0) OTHER				0		
TOTAL SALARIES AND WAGES (A + B)				85,420		
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				13,544		
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				98,964		
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 (0.0) 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				24,504		
TOTAL OTHER DIRECT COSTS				24,504		
H. TOTAL DIRECT COSTS (A THROUGH G)				138,468		
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)						
TOTAL INDIRECT COSTS (F&A)				69,517		
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				207,985		
K. FEE						
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				207,985		
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			

C *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 (50% for 9 academic months and 100% for 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 2021 (July 2020 – June 2021) and projected rates for subsequent years. Rates are pro-rated for budget years crossing different fiscal years.

	Faculty	Grad Students
FY 2021 (07/01/2020 – 06/30/2021)	15.0%	13.4%
FY 2022 (07/01/2021 – 06/30/2022)	15.0%	13.0%
FY 2023 (07/01/2022 – 06/30/2023)	24.9%	14.3%
FY 2024 (07/01/2023 – 06/30/2024)	24.9%	14.3%

Tuition Remission

Duke University employs an Average Rate Basis method for tuition recovery from sponsored research. For the 2020-2021 academic year, the average rate is 37.7%, 38.4% in academic year 2021-2022 and 39.1% in academic year 2022-23. This rate is applied consistently to any PhD student salary charged directly to a sponsored project. For this project, funds are requested to support one (1) PhD student for 4.5 academic months and 3 summer months in all three years.

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.

NSF CURRENT AND PENDING SUPPORT

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

PROJECT/PROPOSAL CURRENT SUPPORT

1. Project/Proposal Title: Community Quasi-Monte Carlo Software Library

   Proposal/Award Number (if available): A20-0028

   Source of Support: SIGOPT

   Primary Place of Performance: IIT-Illinois Institute of Technology

   Project/Proposal Support Start Date (if available): 2019/08

   Project/Proposal Support End Date (if available): 2021/08

   Total Award Amount (including Indirect Costs): $20,000

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

Year	Person-months per year committed
2020	0.01

PROJECT/PROPOSAL PENDING SUPPORT

1. Project/Proposal Title: Collaborative Research: A Community Quasi-Monte Carlo Library for Uncertainty Quantification(this proposal)

   Proposal/Award Number (if available):

   Source of Support: NSF-National Science Foundation

   Primary Place of Performance: IIT-Illinois Institute of Technology

   Project/Proposal Support Start Date (if available): 2021/06

   Project/Proposal Support End Date (if available): 2024/05

   Total Award Amount (including Indirect Costs): $100,000

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

Year	Person-months per year committed
2021	3

2. Project/Proposal Title: Collaborative Research: Adaptive Multivariate Sampling to Accelerate Discovery

   Proposal/Award Number (if available): 2012834

   Source of Support: NSF-National Science Foundation

   Primary Place of Performance: IIT-Illinois Institute of Technology

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

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

   Total Award Amount (including Indirect Costs): $421,803

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

Year	Person-months per year committed
2020	3

NSF CURRENT AND PENDING SUPPORT

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

PROJECT/PROPOSAL PENDING SUPPORT

1. Project/Proposal Title: Collaborative Research: A Community Quasi-Monte Carlo Library for Uncertainty Quantification(This Proposal)

   Proposal/Award Number (if available):

   Source of Support: NSF-National Science Foundation

   Primary Place of Performance: IIT-Illinois Institute of Technology

   Project/Proposal Support Start Date (if available): 2021/06

   Project/Proposal Support End Date (if available): 2024/05

   Total Award Amount (including Indirect Costs): $100,000

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

Year	Person-months per year committed
2021	3

Current and Pending

Mak, Simon TSZ Fung

CURRENT

Project Title: **The X-SCAPE Collaboration: The X+ion Collision with a Statistically and Computationally Advanced Program Envelope Collaboration** (co-PI)
Source of Support: National Science Foundation
Award Total Amount: $696,442 Total Award Period Covered: 07/01/2020 – 6/30/2024
Location of Project: Duke University
Person Months Per Year Committed to the Project: 1 summer month, Years 1-3

Project Title: **Meetings of New Researchers in Statistics and Probability** (PI)
Source of Support: National Science Foundation
Award Total Amount: $300,000 Total Award Period Covered: 07/01/2020 – 6/30/2023
Location of Project: Duke University
Person Months Per Year Committed to the Project: 0.0 summer

PENDING

Project / Proposal Title: **Physics-Integrated Surrogate Modeling (PhISM): a New Machine Learning Framework for Predictive Scientific Computing** (PI)
Source of Support: Department of Energy
Award Total Amount: $300,000 Total Award Period Covered: 07/01/20 – 06/30/22
Location of Project: Duke University
Person Months per Year Committed to the Project: 1.25 summer (Year 1), 1.75 summer (Year 2)

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 the College of Science. 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.

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 supple- mental attachments, we have included letters of collaboration from a few key collaborators on this project, should the project 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. Prof. Owen will also encourage his student to help out with the implementation of Sobol' indices.

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 year 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 SigOpt 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.

Dr. Tim Sullivan (University of Warwick, UK) provides expertise in two application domains for the QMC software developed in this project. The first area is probabilistic numerics, a Bayesian statistical approach to numerical tasks such as cubature and the solution of differential equations, in which the solution object is a statistical posterior distribution and reflects the discretization un- certainty inherited from the finite computational budget. The second is the use of QMC methods in the training of metamodels for heterogeneous (i.e. mixed atomistic-continuum) systems, which is an area of particular interest in Warwick's EPSRC Centre for Doctoral Training in Modelling of Heterogeneous Systems (HetSys).

Shot-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. Access to 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/. 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 quasi-Monte Carlo 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.

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. 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.

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
STANFORD, CALIFORNIA 94305–4065

Art B. Owen
Professor of Statistics, and Chair
Phone: (650) 725-2232
Email: owen@stanford.edu

August 31, 2020

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

Dear Program Directors,

If the proposal submitted by Dr. Fred J. Hickernell entitled Collaborative Research: Quasi-Monte Carlo Community Software 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
Professor of Statistics, and Chair

SigOpt
100 Bush St. Ste 1100
San Francisco, CA 94104

August 31, 2020

To whom it may concern:

If the proposal submitted by Dr. Fred J. Hickernell entitled **Collaborative Research: Quasi-Monte Carlo Community Software** 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,

Michael McCourt
________________________________ August 31, 2020__

Michael McCourt date
Research Engineer
(216)409-4644
mccourt@sigopt.com

Dr T. J. Sullivan PhD MMath
Assistant Professor in Predictive Modelling
Mathematics Institute and School of Engineering
The University of Warwick
Coventry CV4 7AL
United Kingdom

t.j.sullivan@warwick.ac.uk
https://warwick.ac.uk/tjsullivan

National Science Foundation
2415 Eisenhower Avenue
Alexandria
VA 22314
United States of America

September 2, 2020

Collaborative Research: Quasi-Monte Carlo Community Software

To the National Science Foundation of the United States of America,

If the proposal submitted by Dr. Fred J. Hickernell entitled "Collaborative Research: Quasi-Monte Carlo Community Software" is selected for funding by NSF, then it is my intent to collaborate with this project and/or commit resources as detailed in the Project Description or the Facilities, Equipment and Other Resources section of the proposal.

Yours faithfully,

T. J. Sullivan

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	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	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	Élise, Arnaud	Universite Grenoble Alpes		12/31/17
A	Fan, Jianqing	Princeton University		12/31/20
A	Fang, Hong-Bin	Georgetown University Medical Center		12/31/20
A	Gilquin, Laurent	Inria Grenoble Rho^ne-Alpes		12/31/17
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	Hervé, Monod	MaIAGE, INRA, Universite Paris-Saclay		12/31/17
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	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	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	Prieur, Clémentine	Universite Grenoble Alpes		12/31/17
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	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	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	Ye, Qi	South China Normal University		12/31/15
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		12/31/16
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	Fasshauer, Gregory E.	Colorado School of Mines		12/31/18
C	Giles, Michael B.	University of Oxford		
C	Hyman, Mac	Tulane University		
C	Ipsen, Ilse	North Carolina State University		07/31/18
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	Georgia Tech		
C	McCourt, Michael	SigOpt		
C	Novak, Erich	Friedrich-Schiller-Universität Jena		
C	Owen, Art	Stanford University		
C	Rathinavel, Jagadeeswaran	Illinois Institute of Technology		
C	Roshan, V	Georgia Tech		12/31/18
C	Tong, Xing	University of Illinois, Chicago		
C	Wozniakowski, Henryk	University of Warsaw		
C	Zhang, Kan	Illinois Institute of Technology		
C	Zhang, Yizhi	Jamran International		

Table 5

5	Name:	Organizational Affiliation	Journal/Collection	Last Active Date
B	Babuska, Ivo	Univesity of Texas, Austin	Intl. J. Numerical & Applied Math.	03/31/18

B	Bochev, Pavel	Sandia National Laboratories	SIAM J. on Numerical Analysis	12/31/17
B	Cui, Jun-zhi	Inst. Comp. Math., Chinese Acad. Sci.	Intl. J. Numerical & Applied Math.	03/31/18
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	
B	Kunoth, Angela	Universität zu Köln	SIAM J. on Numerical Analysis	12/31/18
E	Kuo, Frances	University of New South Wales	Journal of Complexity	
B	Lan, Yanping	University of Alberta	Intl. J. Numerical & Applied Math.	03/31/18
B	Novak, Erich	Friedrich-Schiller-Universität Jena	Journal of Complexity	
B	Ritter, Klaus	Technische Universität Kaiserslautern	Journal of Complexity	
B	Shi, Zhongci	Inst. Comp. Math., Chinese Acad. Sci.	Intl. J. Numerical & Applied Math.	03/31/18
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	

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	Kamakura Corporation, USA	
	Choi, Sou Cheng	Allstate Insurance Company, USA	01/31/20

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	

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	Saunders, Michael	Stanford University	saunders@stanford.edu, Management Science and Engineering	02/28/14
A	Munson, Todd S.	Argonne National Laboratory	tmunson@mcs.anl.gov, Mathematics and Computer Science Division	03/31/14
A	Berriman, Bruce	California Institute of Technology	gbb@ipac.caltech.edu, Infrared Processing and Analysis Center	07/09/14
A	Elster, Anne C.	Norwegian University of Science and Technology	elster@idi.ntnu.no, Computer Science	07/09/14
A	Hanwell, Marcus D.	Kitware Inc.	marcus.hanwell@kitware.com	07/09/14
A	Lapp, Hilmar	Duke University	hlapp@duke.edu, Center for Genomic and Computational Biology	07/09/14
A	Maheshwari, Ketan	Oak Ridge National Laboratory	ketan@mcs.anl.gov	07/09/14

A	Swenson, Shel	Emory University	Mathematics and Computer Science	07/09/14
A	Turk, Matthew J.	University of Illinois Urbana-Champaign	mjturk@illinois.edu, Astronomy	07/09/14
A	Clune, Thomas L	National Aeronautics and Space Administration	thomas.l.clune@nasa.gov, Goddard Space Flight Center	02/22/16
A	Cranston, Karen A.	Duke University	karen.cranston@gmail.com, Biology	02/22/16
A	Hong, Neil Chue	University of Edinburgh	N.ChueHong@software.ac.uk, Software Sustainability Institute	02/22/16
A	Howison, James	University of Texas at Austin	jhowison@ischool.utexas.edu, School of Information	02/22/16
A	Jones, Matthew	University of California Santa Barbara	jones@nceas.ucsb.edu, National Center for Ecological Analysis and Synthesis	02/22/16
A	Littauer, Richard	University of Saarland	richard.littauer@gmail.com, Computational Linguistics	02/22/16
A	Seinstra, Frank	Netherlands eScience Centre	F.Seinstra@esciencecenter.nl, Computer Science	02/22/16
A	Wilkins-Diehr, Nancy	San Diego Supercomputer Center	wilkinsn@sdsc.edu	02/22/16
A	Accomazzi, Alberto	Harvard University	HarvardSmithsonian Center for Astrophysics	09/19/16
A	Allen, Alice	Astrophysics Source Code Library		09/19/16
A	Altman, Micah	Massachusetts Institute of Technology		09/19/16
A	Billings, Jay Jay	Oak Ridge National Laboratory		09/19/16
A	Boettiger, Carl	University of California, Berkeley		09/19/16
A	Brown, Jed	University of Colorado Boulder		09/19/16
A	Crick, Tom	Cardiff Metropolitan University		09/19/16
A	Crosas, Mercè	Harvard University	Institute for Quantitative Social Science	09/19/16
A	Edmunds, Scott	BGI Hong Kong	GigaScience	09/19/16
A	Erdmann, Christopher	Harvard University	HarvardSmithsonian Center for Astrophysics	09/19/16
A	Fenner, Martin	DataCite		09/19/16
A	Finkbeiner, Darel	The Office of Scientific and Technical Information		09/19/16
A	Gent, Ian	University of St Andrews	School of Computer Science	09/19/16
A	Goble, Carole	The University of Manchester	Software Sustainability Institute	09/19/16
A	Groth, Paul	Elsevier Inc.	Elsevier Labs	09/19/16
A	Haendel, Melissa	Oregon Health and Science University		09/19/16
A	Hagstrom, Stephanie			09/19/16

A	Hanisch, Robert	National Institute of Standards and Technology		09/19/16
A	Henneken, Edwin	Harvard University	HarvardSmithsonian Center for Astrophysics	09/19/16
A	Herman, Ivan	World Wide Web Consortium (W3C)		09/19/16
A	Ingraham, Thomas	F1000Research		09/19/16
A	Jones, Catherine	Science and Technology Facilities Council		09/19/16
A	Konovalov, Alexander	University of St Andrews		09/19/16
A	Kratz, John	California Digital Library		09/19/16
A	Lin, Jennifer	Public Library of Science		09/19/16
A	Matthews, Brian	Science and Technology Facilities Council		09/19/16
A	Mayes, Abigail Cabunoc	Mozilla Science Lab		09/19/16
A	Mietchen, Daniel	National Institutes of Health		09/19/16
A	Mills, Bill	TRIUMF		09/19/16
A	Misshula, Evan	The City University of New York	Graduate Center	09/19/16
A	Muench, August	American Astronomical Society		09/19/16
A	Murphy, Fiona	Independent Researcher		09/19/16
A	Nielsen, Lars Holm	CERN		09/19/16
A	Ram, Karthik	University of California, Berkeley		09/19/16
A	Rios, Fernando	Johns Hopkins University		09/19/16
A	Sands, Ashley	University of California, Los Angeles		09/19/16
A	Scott, Soren	Independent Researcher		09/19/16
A	Smith, Arfon M	Space Telescope Science Institute	Data Science Mission Office	09/19/16
A	Thaney, Kaitlin	Mozilla Science Lab		09/19/16
A	Van Hauwermeiren, Daan	Ghent University, Belgium		09/19/16
A	Van Hoey, Stijn	Ghent University, Belgium		09/19/16
A	Weaver, Belinda	The University of Queensland		09/19/16
A	Allen, Gabrielle	University of Illinois Urbana-Champaign	dallen@illinois.edu, Astronomy	10/21/16
A	Brandt, Steven R.	Louisiana State University	sbrandt@cct.lsu.edu, Computer Science and Engineering	10/21/16

A	Davis, Ethan	University Corporation for Atmospheric Research	edavis@ucar.edu	10/21/16
A	de Val-Borro, Miguel	Princeton University	Astrophysical Sciences	10/21/16
A	Gesing, Sandra	University of Notre Dame	sandra.gesing@nd.edu, Computer Science and Engineering	10/21/16
A	Gunter, Dan	Lawrence Berkeley National Laboratory	dkgunter@lbl.gov, Computational Research Division	10/21/16
A	Hetherington, James	University College London	j.hetherington@ucl.ac.uk, Research Software Development Group	10/21/16
A	Hwang, Lorraine	University of California, Davis	lorraine@geodynamics.org, Earth and Planetary Sciences	10/21/16
A	Idaszak, Ray	University of North Carolina at Chapel Hill	rayi@renci.org, Renaissance Computing Institute	10/21/16
A	Jones, Nick	The University of Auckland	n.jones@auckland.ac.nz, New Zealand eScience Infrastructure	10/21/16
A	Katz, Daniel S.	University of Illinois Urbana-Champaign	dskatz@illinois.edu, National Center for Supercomputing Applications (NCSA)	10/21/16
A	Löffler, Frank	Louisiana State University	knarf@cct.lsu.edu, Center for Computation & Technology	10/21/16
A	Marru, Suresh	Indiana University	smarru@iu.edu, Research Technologies	10/21/16
A	Miller, Mark A.	San Diego Supercomputer Center	mmiller@sdsc.edu	10/21/16
A	Niemeyer, Kyle	Oregon State University	Kyle.Niemeyer@oregonstate.edu, Mechanical Engineering	10/21/16
A	Patra, Abani	University at Buffalo	abani@buffalo.edu, Mechanical and Aerospace Engineering	10/21/16
A	Penzenstadler, Birgit	California State University Long Beach	birgit.penzenstadler@csulb.edu, Computer Science	10/21/16
A	Todorov, Ilian T.	Science and Technology Facilities Council	ilian.todorov@stfc.ac.uk, Computational Chemistry Groups	10/21/16
A	Venters, Colin C.	University of Huddersfield	C.Venters@hud.ac.uk, Software Systems Engineering	10/21/16
A	Weber, Nic	University of Washington	nmweber@uw.edu, Information School	10/21/16
A	Li, Da			08/28/17
A	Zhang, Yizhi	Illinois Institute of Technology	yzhang97@hawk.iit.edu, Applied Mathematics	08/28/17
A	Zhou, Xuan	JPMorgan Chase and Co.	xuanjzhou@gmail.com	08/28/17
A	Lin, Yongheng	Apple Inc.	yongheng.lin@gmail.com	10/29/17
A	Mulrow, Edward	NORC at the University of Chicago	Mulrow-Edward@norc.org, Statistics and Methodology	10/29/17
A	Hartong-Redden, Rory	Runtastic GmbH	roryhr@gmail.com	07/15/18
A	Kochenderfer, Mykel	Stanford University	mykel@stanford.edu, Aeronautics and Astronautics	07/15/18

	J.			
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	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	
C	Jarrow, Robert A.	Kamakura Corporation, USA	RJarrow@kamakuraco.com	
A	Jiang, Lan	Compass	lanjiang61930@gmail.com	
A	Jiménez Rugama, Lluís Antoni	UBS AG	lluisantoni@gmail.com	
C	Ashbeck, Jonathan	University of Pennsylvania	jashbeck@sas.upenn.edu	
A	Li, Jinyang	Beijing Institute of Technology, China	jli221@hawk.iit.edu	
C	Lim, Lek Heng	University of Chicago	lekheng@galton.uchicago.edu, Statistics	
C	Liu, Zongrui	University of Illinois Urbana-Champaign	zongrui2@illinois.edu	
A	McCourt, Michael	SigOpt, Inc.	mccourt@sigopt.com	
C	Mesler, Mark	Kamakura Corporation, USA	MMesler@kamakuraco.com	
C	Negron, Alex	Illinois Institute of Technology	anegron1@hawk.iit.edu	
A	Pao, Chrystal Ho	Trinity International University	chrystalhopao@gmail.com, Biology	
C	Penanhoat, Eric	Kamakura Corporation, USA	epenanhoat@kamakuraco.com	
A	Rathinavel, Jagadeeswaran	Illinois Institute of Technology	jrathin1@hawk.iit.edu, Applied Mathematics	
A	Sorokin, Aleksei	Illinois Institute of Technology	asorokin@hawk.iit.edu	
A	Tong, Xin	University of Illinois at Chicago	xtong20@uic.edu, Mathematics, Statistics, and Computer Science	
C	van Deventer, Don	Kamakura Corporation, USA	DvanDeventer@kamakuraco.com	
A	Wang, Haoran	Beijing Institute of Technology, China	ishaoranwang@gmail.com	
A	Zhang, Kan	Illinois Institute of Technology	kzhang23@hawk.iit.edu, , Applied Mathematics	
A	Zhang, Yiwei	Beijing Institute of Technology, China	yzhang176@hawk.iit.edu	
C	Zorn, Martin	Kamakura Corporation, USA	mzorn@kamakuraco.com	

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	07/12/18
E	Fung, Glenn	American Family Insurance, U.S.	GFUNG@amfam.com	
E	Ma, Lawrence K. H.	Hong Kong Blockchain Society	lawrence.ma@hkbcs.org	
E	Polania, Luisa	Target Corporation, U.S.	lfpolani@udel.edu	
E	Wu, Victor	Tilt Dev Computer Software	wu.victor@gmail.com	

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		

Table 5

5	Name:	Organizational Affiliation	Journal/Collection	Last Active Date
B	Joseph, Roshan Vengazhiyil	Georgia Institute of Technology	Technometrics	

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