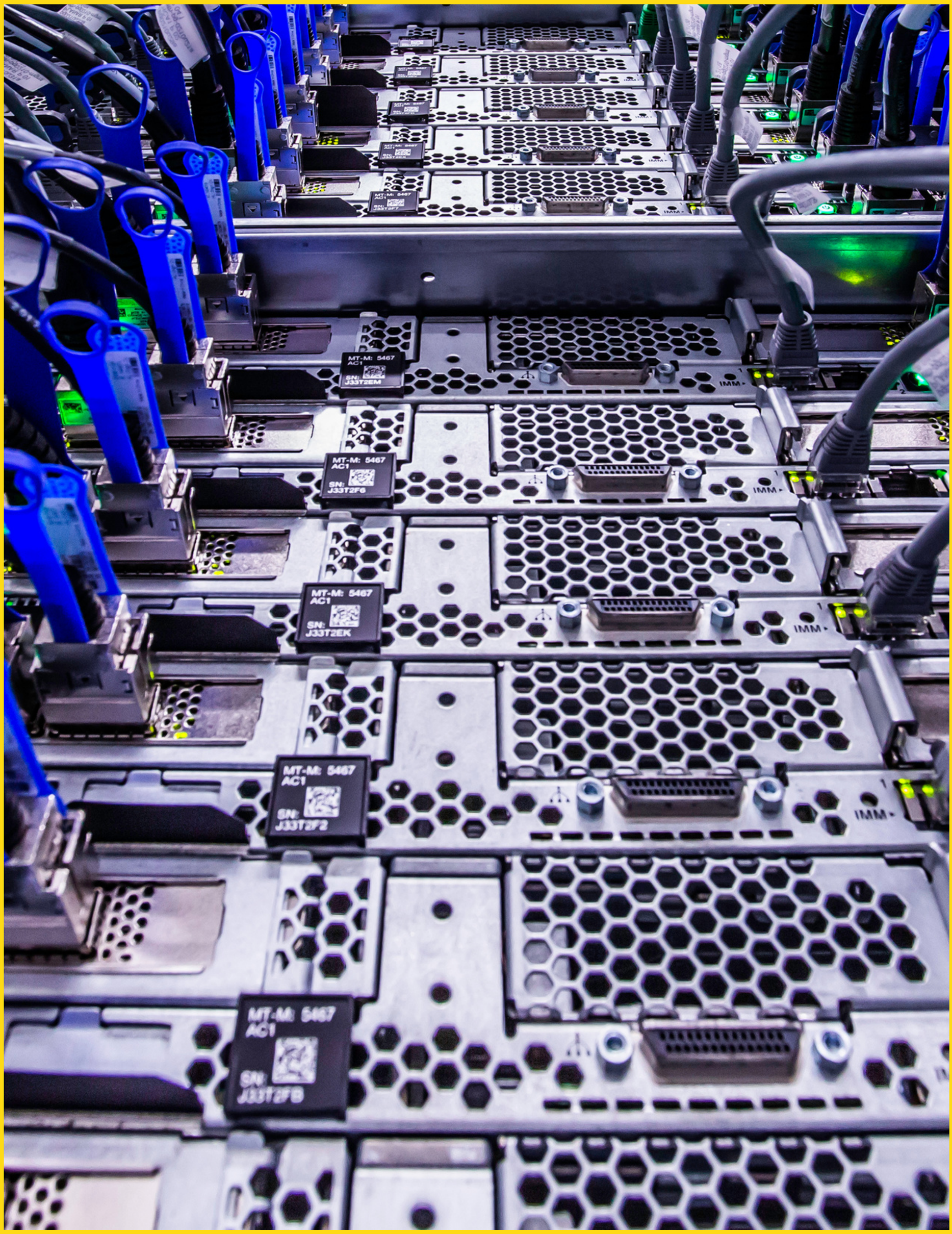
BEAR

BIRMINGHAM ENVIRONMENT
FOR ACADEMIC RESEARCH



19-20TH APRIL
2023

BEAR Conf'23

COPYRIGHT & ACKNOWLEDGEMENTS

Published in the United Kingdom by The University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.

Aslam Ghumra,
The University of Birmingham,
IT Services, Building G5
Edgbaston,
Birmingham, B15 2TT, UK.

Acknowledgements

We would like to thank all the people who contributed to this publication.

Cite

JJenkins, Ben and Sykes, Jack and Thompson, Jacqueline and Softley, Tim and Allsopp, James and Nicusan, Leonard and Gilbert, Sebastian and Chen, Zihan and McCready, Kirsty and Henes, Jacqueline and Zhang, Xiaoqi and Price, Matt and Owen, Chris (2023) BEAR PGR Conference 2023 - Conference proceedings. In: BEAR PGR Conference 2023, 19 - 20 June, University of Birmingham.

DOI : https://doi.org/10.25500/epapers.bham.00004279

FOREWORD

DR DIETMAR HEINKE

CHAIR OF THE BEAR USER GROUP

The BEAR PGR Conference served as a gathering for BEAR users to convene and learn about the captivating research conducted by Postgraduate Researchers (PGRs) and academic staff utilizing the University's supercomputer, BlueBEAR. The BEAR PGR Conference takes place annually and offers PGRs a valuable opportunity to gain conference management experience, enhancing their curriculum vitae.

This year's conference spanned two days and featured poster presentations and talks encompassing a wide range of disciplines, including mathematics, chemical engineering, physics & astronomy, applied health research, social policy, sociology and criminology, computer science, psychology, and the BEAR Team of Research Software Engineers. We were privileged to host two distinguished keynote speakers, namely Professor Stephen Jarvis (Provost and Vice-Principal) and Professor Ole Jenson (Co-Director of the Centre for Human Brain Health).

We trust that these abstracts will prove highly stimulating and motivate you to delve deeper into the presenters' research and explore how BlueBEAR can contribute to your own research endeavours.

The conference culminated in the announcement of the best poster and talk winners. Given the outstanding quality of presentations, selecting the winners was no easy task. However, we are delighted to acknowledge Kirsty McCready as the recipient of the best poster award, commending her engaging and clear presentation. Additionally, Leonard Nicusan earned the best talk award for his visually impactful delivery. We extend our gratitude to Lenovo for generously providing tablets as prizes for the winners.

These proceedings and the conference as a whole would not have been possible without the remarkable efforts of numerous dedicated individuals working tirelessly behind the scenes. We extend a heartfelt thank you to the Organizing Committee, comprised of Antony Lee, Jordan Dorrell, Ulviyya Khalilova, Samiratu Wahab, and Gunay Ismayilova, as well as the advisory Committee members: Aslam Ghumra, Dr. Stephanie Thompson, and Debbie Carter.

Dietmar Heinke

ORGANISING COMMITTEE MEMBERS

Advisory Committee
Dr. Dietmar Heinke
Aslam Ghumra
Dr. Stephanie Thompson
Debbie Carter

CHAIR

Anthony Lee
Computer Science

PUBLICITY & PUBLIC RELATIONS

Gunay Ismayilova
School of Law

DIGITAL ENGAGEMENT

Jordan Dorrell
Metallurgy and Materials

PUBLICITY & PUBLIC RELATIONS

Samiratu Wahab
Economics

PUBLICITY & PUBLIC RELATIONS

Ulviyya Khalilova
Social Policy

KEYNOTE SPEAKERS

PROFESSOR STEPHEN JARVIS BSC, MSC, PHD, FBCS

Provost and Vice-Principal
University of Birmingham

PROFESSOR OLE JENSEN

Professor in Translational Neuroscience
Co-Director of the Centre for Human Brain Health

CONTRIBUTORS

	PAGE
Ben Jenkins (Talk)	7
Jack Sykes (Talk)	8
Jacqueline Thompson (Talk)	9
James Allsopp (Talk)	10
Leonard Nicusan (Talk)	11
Sebastian Gilbert (Talk)	12
Zihan Chen (Talk)	13
Jacqueline Thompson (Poster)	14
Kirsty McCready (Poster)	15
Jacqueline Henes (Poster)	16
Leonard Nicusan (Poster)	17

THE NEXT-GENERATION OF POWDER AND PARTICLE CHARACTERISATION TOOLS

Ben Jenkins, Andrei Nicuşan, Aurélien Neveu, Geoffroy Lumay, Filip Francqui, Jonathan Seville, Kit Windows-Yule

email: bdj746@student.bham.ac.uk

Discrete element method (DEM) is a way of simulating powders on a computer where each particle is modelled as an individual element. DEM can be used to simulate powders in industrial equipment to design better equipment. However, to get good results the microscopic properties of the powder - friction, cohesion - need to be calibrated to ensure the simulation provides accurate results. Current calibration are time-intensive and often inaccurate. Therefore, a better calibration method is needed. Powder characterisation tools are lab scale pieces of equipment that take a sample of powder or granular material and measure a bulk property. One simple example is pouring a powder on a table to form a heap. The angle this heap makes with the surface is a common measurement used for powders called the angle of repose.

These bulk measurements are inherently linked to the microscopic properties. If the particles that make up a powder are more frictional, then the heap will be steeper, and the angle of repose measured will be higher. Therefore, if we knew the relationship between each bulk measurement of a powder and the microscopic properties of the particles, the bulk measurements could be used to calculate what the microscopic properties could be. Unfortunately, this relationship is very complex and currently impossible to find. Instead, a data driven approach is being used. Digital twins of powder characterisation tools are being used to generate data of bulk measurements at different microscopic properties on BlueBear. This data can then be used with a data-driven method to calibrate DEM simulations in any system

ACCES: AUTONOMOUS CHARACTERISATION AND CALIBRATION USING EVOLUTIONARY SIMULATION

Jack Sykes, Andrei Leonard Nicusan, Dominik Werner, Kit Windows-Yule, Tzany Kokalova Wheldon

email: jas653@student.bham.ac.uk

The discrete element method (DEM) is a powerful simulation technique that is capable of numerically modelling the behaviour of complex granular media, being used to better understand and optimise the internal dynamics of a large number of systems and multiphase flow processes in both academic fields and industrial sectors, from fundamental research into contact mechanics to improving plant-scale reactors [Rosato and Windows-Yule, 2020]. DEM can offer exceptional accuracy through its lack of approximations over meshes and, if correctly calibrated, simulations can provide results with quantitative precision. It is this “if”, however, that also represents DEM’s biggest drawback: without choosing appropriate contact models and carefully calibrating multiple DEM parameters, the simulation outputs simply cannot be trusted. This calibration is a time-consuming process, typically involving the measurement of diverse particle properties including size, density, restitution and friction coefficients and, for purely “virtual” parameters such as the cohesive energy density, a great deal of experimentation [Luding, 2008].

To automate DEM calibration against experimental measurements, we have developed ACCES - Autonomous Characterisation and Calibration using Evolutionary Simulation. ACCES enables a researcher to calibrate virtually any DEM parameters against a user-defined cost function, quantifying and subsequently minimising the disparity between the simulated system and the experimental reality using state-of-the-art evolutionary strategies – in essence, autonomously ‘learning’ the physical properties of the particles within the system, without the need for human input. This cost function is completely general, allowing ACCES to calibrate DEM against measurements as simple as photographed occupancy plots, or complex system properties captured through e.g. Lagrangian particle tracking in multiphase flow processes. The algorithm itself is completely DEM engine-agnostic; it was implemented in an open-source Python library, providing an interface that is easy to use, but powerful enough to automatically parallelise arbitrary user scripts through code inspection and metaprogramming.

METHODS FOR ESTIMATING COVARIATE-ADJUSTED RELATIVE RISKS: A SIMULATION EVALUATION

Jacqueline Thompson, Karla Hemming, Lee Middleton, Sam Watson

email: j.y.thompson@bham.ac.uk

Binary outcomes are widespread in clinical trials. The odds ratio is a common and established approach for estimating covariate-adjusted binary treatment effects when comparing a treatment and control group. Its popularity is primarily because of its stability and robustness to model mis-specification. However, the situation is different for the relative risk; there is no equivalent, widely acceptable approach to estimate an adjusted relative risk (aRR) when conducting clinical trials. This lack of consistency in practice is partly due to the need for a comprehensive evaluation of available candidate methods to identify optimal approaches for estimating aRRs.

A literature review was performed, and a simulation study was designed to evaluate the performance of several candidate methods for estimating aRR that represent parametric and non-parametric estimation approaches. We consider the log-binomial, generalised linear models with iteratively re-weighted least-squares (IRWLS) and model-based standard errors (SE); log-binomial with convex optimisation and Hessian SEs; modified-Poisson IRWLS and robust SEs; log-binomial and Poisson generalised estimation equations with robust SEs; marginal standardisation with delta method SEs and permutation test SEs.

Independent and identically distributed datasets are simulated from a randomised controlled trial to evaluate these candidate methods. Performance measures (bias, empirical and mean square errors, relative efficiency and convergence rates) are considered across scenarios. Simulations are replicated 10 000 times for each scenario across all combinations of sample sizes (100, 200, 500, 1000, and 5000), outcomes (5%, 20%, 50% and 80%), and covariates (ranging from -0.05 to -0.02 on the log scale) with main and interaction effects. The treatment effect of 0 (on the log scale) under the null (H0) hypothesis is used to evaluate coverage and power. Subsequently, candidate methods are assessed using datasets with correlated covariates (correlation coefficients ranging from 0.0 to 0.25 on the log scale) and mis-specified models to illustrate the behaviour of candidate methods under these settings.

COULOMB CRYSTALS - TOWARDS A GREATER UNDERSTANDING OF CHEMICAL BONDING

Tim Softley, **James Allsopp**

email: j.allsopp@bham.ac.uk

Since Paul Traps were developed in the late eighties, they have provided a way of studying chemical species and their reactions in a very dense and extremely cold environment. When trapped, ionised species form an ordered state called a Coulomb Crystal, which can be made to fluoresce with an appropriately tuned laser, and the light is then captured by a microscope.

This allows chemists to work with a cleaner signal when studying how reactions progress, as the species should be almost all in the ground state, and collisions should be frequent enough to be experimentally practical.

Due to its lack of an activation energy, one of the simplest reactions is between an ionised species and a neutral species, the theory of which has remained essentially unchanged since 1905. Interpreting the fluorescence images of the crystals is demanding and it can take up to an hour of trained experimenter's time to find the number of species. To streamline this process, BlueBear is being used to simulate the data and train a machine learning model that can analyse an image in seconds.

EVOLUTIONARY INDUSTRIAL DESIGN OPTIMISATION USING HARPPP

Leonard Nicusan, Kit Windows-Yule

email: aln705@student.bham.ac.uk

Currently, industrial geometric design - such as shapes of impellers, pellets, beds, vessels, baffles or placement of any process assembly such as heat recovery or reactant injection points in larger plants - is virtually always done empirically, by human experts with vast domain experience. Improvements are done in very small steps based on well-known previous designs, and one design iteration may take years to go from CAD drawings to prototype manufacture and finally experimental campaigns; a single such design iteration may form the basis of an entire grant.

To solve this, we have developed HARPPP, a highly autonomous computational solution to optimise geometric designs in conjunction with physically-based simulations: given only a CAD model of the system and details of the material it processes, HARPPP uses AI to design and re-design the system of interest, iteratively improving its function toward a goal (e.g. “minimise power draw” or “maximise throughput”) defined by the user.

The optimisation backend was implemented in an open-source Python library, providing an interface that is easy to use, but powerful enough to automatically parallelise arbitrary user scripts through code inspection and metaprogramming. It was used successfully from laptop-scale shared-memory machines to multi-node supercomputing clusters on BlueBEAR.

TOPOLOGICAL BIOMARKERS FOR COLORECTAL CANCER PROGNOSIS

Sebastian Gilbert

email: sgg897@student.bham.ac.uk

Evidence shows that, throughout their life, a person will develop tumours which are identified and eradicated by normal immunological mechanisms, without notice or consequence by the individual. However, in many cases when ‘normal’ immunological responses fail, we observe the detrimental results as cancer.

Multiplex immuno-fluorescence imaging identifies the proximity of many different immune cell types within the tumour microenvironment. Designing mathematical and computational methods to interpret the spatially-resoled ex-vivo dataset (specifically in colorectal and kidney cancer) can lead to crucial ways in understanding and predicting disease progression.

We introduce topological techniques, such as persistent homology and winding numbers, to interpret the communities within these multi-layered networks. Combining these techniques with prior knowledge of immunological mechanisms, we present intuitive and predictive biomarkers which go beyond previous prognostic tools such as grading, multi-satellite stability/instability, and Immunoscore®

STUDY ON STRUCTURAL DILEMMA AND COMPLEX LOGIC OF URBAN MIGRANTS' INTEGRATION IN CHINA- BASED ON CHINA MIGRANTS DYNAMIC SURVEY

Zihan Chen

email: zxc203@student.bham.ac.uk

Urban migrants in China face various integration challenges due to the excessive speed urbanization and unique hukou system, which is a special population registration and an identity document that distinguishes individuals as “urban hukou” or “rural hukou” based on their family’s work and birthplace. As urban migrants encounter a Structural Dilemma during the integration process, leading to higher social risk, which makes Complex Logic more obvious, this project aims to use Machine Learning Methods, Decision Methods and Statistical Methods to achieve the following goals: 1) Find structural dilemmas that urban migrants face in terms of integration; 2) The complex logic of urban migrants’ integration; 3) The factors affecting urban migrants' integration and the relationship between the structural dilemma and the complex logic of integration; 4) The ways that structural dilemmas and the complex logic of integration can be improved.

To fill the literature gaps, this project will construct a complete theoretical framework. For instance, conceptualize and operationalize the structural dilemma from a policy angle, establish a policy evaluation system and construct the complex logic research system from risk society theory. Meanwhile, the empirical evaluation of integration as multidimensional is in development, so the work in this project like using machine learning, coupling coordination degree model and so on would contribute to the further development of work in this area. Furthermore, it will make an important contribution to evaluate existing policies and to find new solutions about urban migrants’ integration as well as superdiversity in China.

Methods for estimating dichotomous treatment effects: a comparative simulation study

Jacqueline .Y Thompson, Sam Watson, Lee Middleton, Karla Hemming

BESTATE Research Group, Institute of Applied Health Research,
The University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

SUMMARY OF FINDINGS

Clinical trial practitioner impact:

- This project addresses a gap in the relative performance of selected statistical methods for clinical trials, which show negligible bias in a neutral simulation study using a homogenous sample of data.
- Methods show different efficiency (i.e., how information is used) across scenarios and potential sample size gains with covariate adjustment. **Future work will explore which methods are most powerful**.

Patient and public impact:

- Trialists can now use suitable methods for estimating risks (as relative risks) in clinical trials while accounting for factors that influence the size of benefit observed; maximising information gained from data.
- This will also facilitate communication of scientific findings with the lay public -making it more accessible; improving understanding of the impact of health effectiveness interventions.

BACKGROUND

- Covariate-adjusted treatment effects, commonly reported as an adjusted odds ratio, estimated using logistic regression -due to its stability and robustness to model mis-specification-can improve precision.
- However, adjusted relative risks (aRR) are easier to understand, and adjusted risk differences (aRD) are better suited for equivalence or non-inferiority trials.
- CONSORT and the FDA recommend reporting aRRand aRDas complementary treatment effects, but it is unclear how approaches for estimating aRRperform under different scenarios in clinical trials.

OBJECTIVE

- ***To evaluate the performance of selected candidate methods for estimating adjusted relative risks in a head-to-head comparison.***

CANDIDATE METHODS FOR ESTIMATING RELATIVE RISK

Candidate methods	Model	Optimisation technique	Standard errors (SE)
Generalised linear model: Direct methods			
Log-binomial, IRWLS	Log-binomial distribution	IRWLS	Default (model-based) SEs
Log-binomial, GEE	Log-binomial	GEE	Robust SEs
Log-binomial, convex optimisation	distribution Log-binomial distribution	Convex optimisation	Default (Hessian) SEs
			Design-based SEs (permutation test)
Modified Poisson	Poisson distribution	IRWLS	Robust SEs (HC1)
Generalised linear model: Indirect methods			
Marginal standardisation, delta SE	Logistic distribution	IRWLS	Delta method SEs of marginal probabilities
Marginal standardisation, permutation test	Logistic distribution	IRWLS	Design-based SEs (permutation tests) of predictive probabilities

IRWLS -Iteratively reweighted least-squares; GEE -Generalised Estimating Equations; SE –Standard errors; HC –heteroskedasticity consistent; MS –Marginal standardisation

DATA GENERATING MECHANISM

- 10000 iterations from log-binomial model*(iid)*

For each individual i with dichotomous outcome $y_i \in [0,1]$

$$y_i \sim Bernoulli\ h(X_i\beta)$$

where $h(.)$ is the link function; $h(x) = \log(x)$

- *Allocation ratio: 1:1*
- *Sample sizes: 100, 200, 1000 & 5000*
- *Continuous covariates (unadjusted and adjusted):*
- *Number: maximum 4*
- *Strength: -0.05 to -0.02*
- *Frequency of outcomes:*
- *Infrequent (.20); Common (.50); Prevalent (very common) (.80)*
- *Null (H0) hypothesis: (0 treatment effect on the log-scale)*

Scenarios were evaluated for:

- *Convergence rates and Coverage*
- *Empirical standard errors and inverse of variance (1/(Variance))*
- *Relative efficiency (Adjusted mean square error (MSE) / Unadjusted MSE)*
- *Proportion reduction in sample size with covariate adjustment (1 -relative efficiency)*

BENCHMARK SCENARIO

- Base scenario -coverage was "only" optimal across candidate methods at a sample size of ≥5000.
- Convergence was worse for log-binomial IRWLS and GEE at high event rates and small sample sizes.

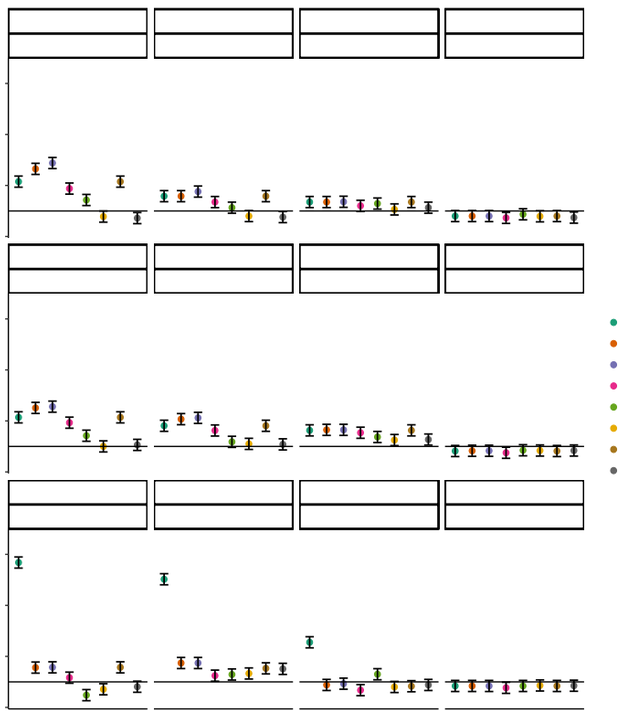


Figure 1: Benchmark scenarios with no factors considered, showing base performance and non-convergence

PERFORMANCE OF CANDIDATE METHODS

- Log-binomial IWRLS optimised using convex optimisation with Hessian SE or permutation test showed improved performance over IWRLS optimisation alone.
- Modified-Poisson and log-binomial GEE have similar performance and straightforward syntax.
- Marginal standardisation with delta method or permutation test showed optimal performance across all scenarios but is the most ***computationally intensive.***
- Marginal standardisation with ***delta SE has a simpler syntax.***

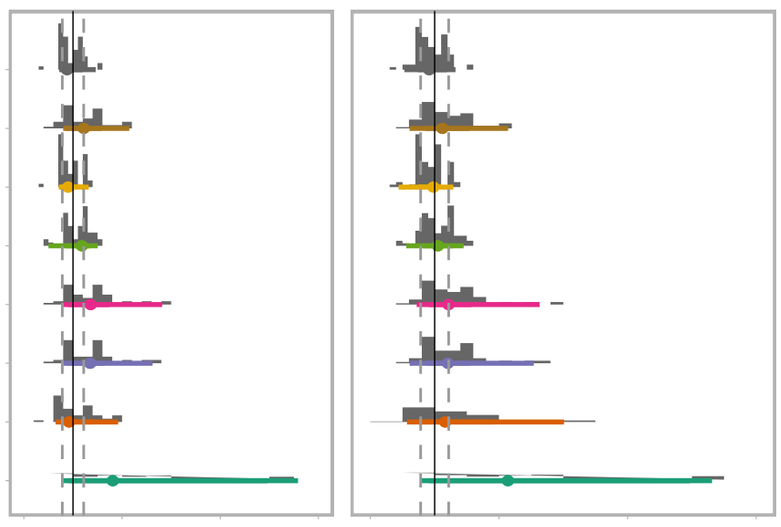


Figure 2: Coverage of candidate methods (probability that empirical confidence interval contains true value), showing improving precision with covariate adjustment across scenarios, set up to reflect real-world scenarios.

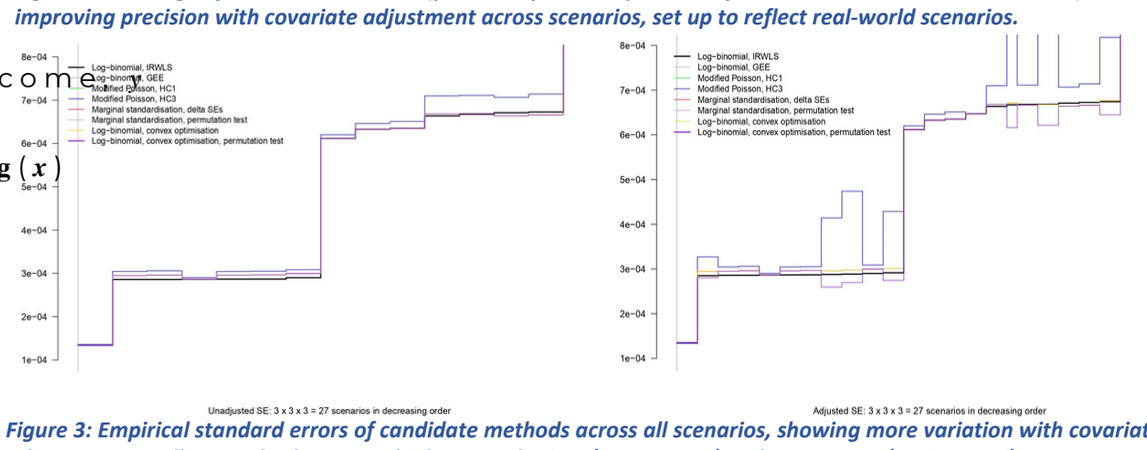


Figure 3: Empirical standard errors of candidate methods across all scenarios, showing more variation with covariate adjustment, smaller standard errors at higher sample sizes (5000 to 100) and event rates (80% to 20%).

Efficiency of candidate methods	Adjusted models relative to Unadjusted models		
	Precision (%)	Relative efficiency (%)	Sample size reduction (%)
Log-binomial, IRWLS	94.47%	78.59%	**21.41%**
Modified Poisson, HC1	93.76%	78.66%	21.34%
Modified Poisson, HC3	93.76%	78.66%	21.34%
Log-binomial, GEE	83.87%	**80.66%**	19.34%
Marginal standardisation, delta SE	93.76%	78.63%	**21.37%**
Marginal standardisation, permutation test	93.76%	78.63%	**21.37%**
Log-binomial, convex optimisation	93.76%	78.50%	**21.50%**
Log-binomial, convex optimisation, permutation test	**98.46%**	78.31%	**21.69%**

IRWLS -Iteratively reweighted least-squares; GEE -Generalised Estimating Equations; SE –Standard errors; HC –heteroskedasticity consistent; MS –Marginal standardisation

CONCLUSION

- √ Bias across candidate methods for unadjusted and adjusted treatment effects are small (<0.01).
- √ Candidate methods had a similar distribution of small empirical SEs, which improved with covariate adjustment.
- √ **Overall, Log-binomial IWRLS and GEE showed poor performance.**
- √ **We recommend that Log-binomial convex optimisation with or without a permutation test and Modified-Poisson is used when non-convergence is anticipated.**
- √ **Marginal standardisation with delta standard error or permutation test out-perform other candidate models; suitable when marginal estimates are of interest.**
- √ Precision improved with "complete" covariate adjustment, but lower gains can be expected in real-world scenarios with missing outcomes and weaker covariate strengths.

The search for missing genes in the *Ceratopteris richardii* genome

Kirsty McCready, Dr. Juliet Coates, Dr. Andrew Plackett

MIBTP Midlands Integrative Biosciences Training Partnership

UNIVERSITY OF BIRMINGHAM

BEAR BIRMINGHAM ENVIRONMENT FOR ACADEMIC RESEARCH

Significant biological data is routinely lost during RNA-seq processing

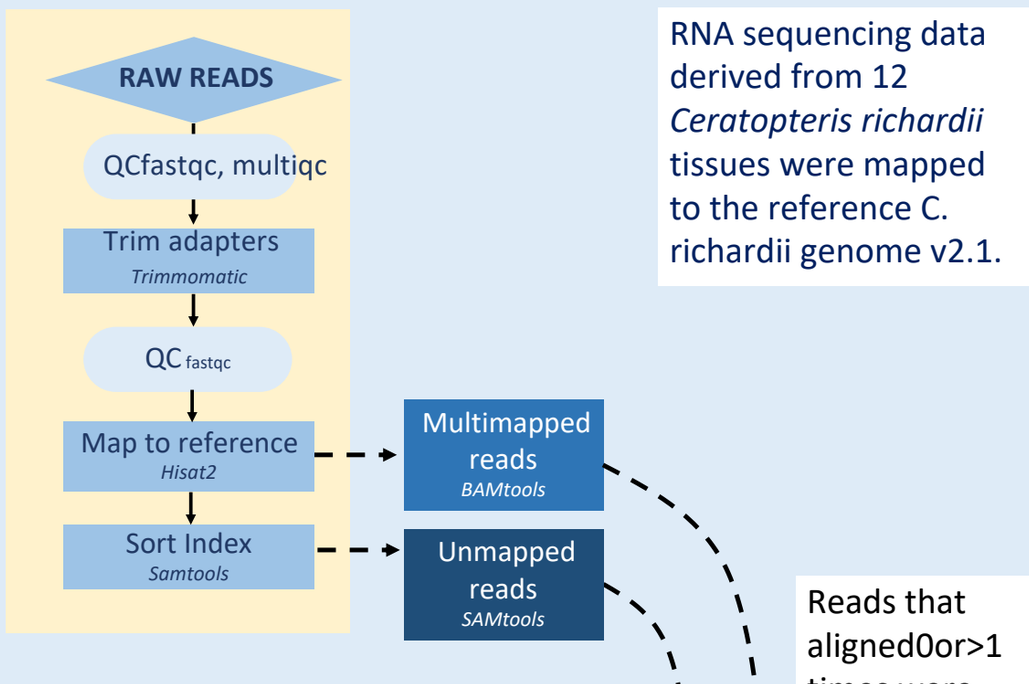
Alignment of reads to a reference genome is a common first step in the analysis of Next Generation Sequencing data. However, despite advancements in accuracy and efficiency of mapping, a significant proportion of reads remain unmapped/multimapped after alignment. These reads are subsequently discarded, leading to the loss of potentially important biological information.

The *Ceratopteris richardii* (*Cr*) genome was recently published, however its completeness and accuracy remains largely untested. Here, unmapped and multimapped RNA-seq reads were extracted after mapping to the *C. richardii* assembly and *de novo* assembled to find potentially missing genes.

The Pipeline

Reference-led mapping

RNA sequencing data derived from 12 *Ceratopteris richardii* tissues were mapped to the reference C. richardii genome v2.1.



Reads that aligned0or>1 times were extracted and combined.

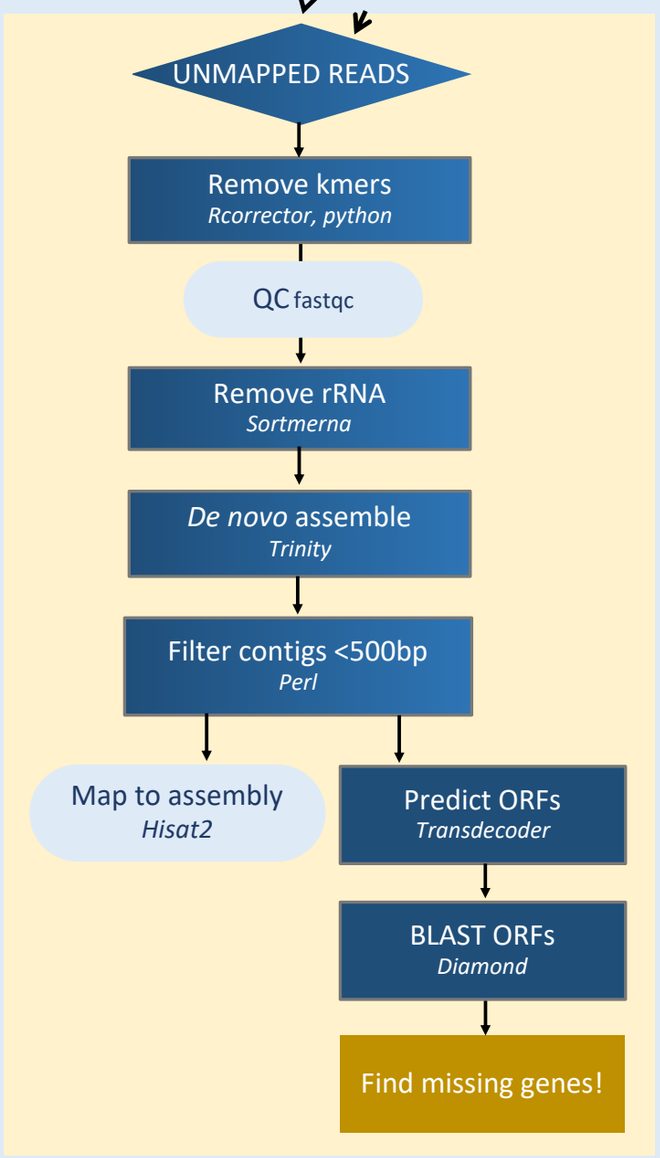
De novo assembly

Unmapped reads were cleaned and assembled following an adapted pipeline by Harvard FAS informatics.

ORFs were predicted from the assembled contigs and blasted against the NCBI nr database.

ORFs that matched to plants other than *C. richardii* were retained for further analysis.

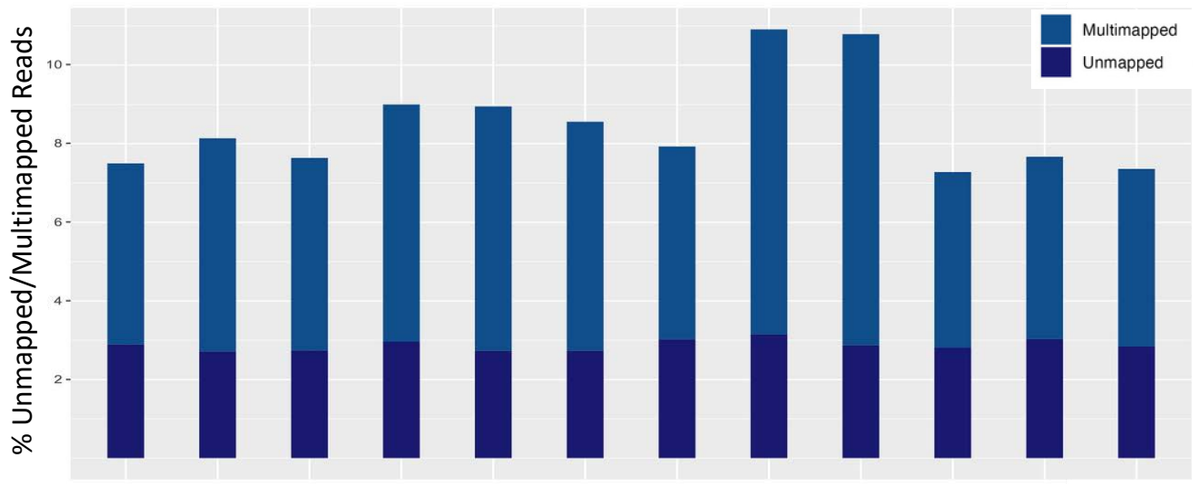
Non-*Cr* plantORFswere aligned back to the *Cr* genome. Unmapped reads were deemed “missing genes”.



Scan for References

Finding missing genes

>115 million reads didnotmaptothe genome



Total Reads	3773015	0392435	0537284	8953822	8782444	7069336	4014603	5061663	5426680	4065900	4356945	0336625	0123626	3242
Unmapped	1087241	1065318	1016071	1142855	1233924	998295	1073310	1110899	1178111	1000460	1116818	1036921		
%Unmapped	2.89	2.71	2.74	2.96	2.73	2.73	3.02	3.15	2.87	2.81	3.03	2.84		
Multimapped	1757342	2132936	1808981	2324952	2752536	2119393	1724291	2749130	3242827	1602064	1702121	1651940		
%Multimapped	4.61	5.43	4.9	6.04	6.22	5.83	4.91	7.76	7.92	4.47	4.64	4.52		

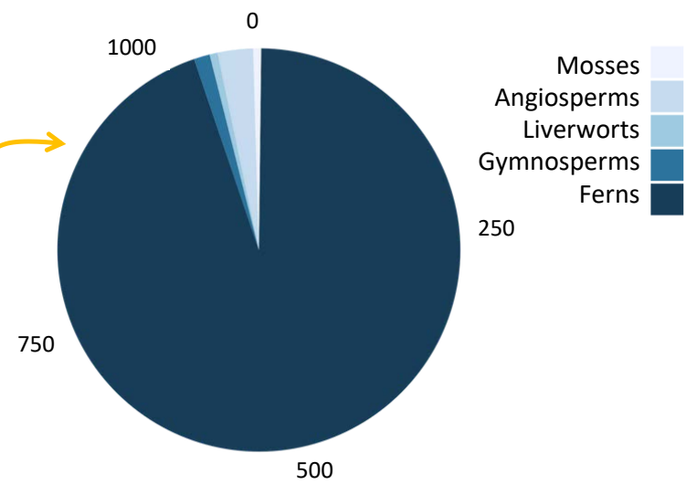
Graphical andtabularsummary of the number and percentage of unmapped/multimapped reads per tissue type after alignment to the *C. richardii* genome.

The assembly

115,886,213 unmapped or multimapped reads were de novo assembled into 61,943 trinity contigs (>500bp), with an N50 of 1,447 bp and a median contig length of 994 bp. From these contigs, 55,130 open reading frames (ORFs) were identified.

50,524 ORFs aligned to the nr database with an e value lower than 1e-10:

Group		# Alignments
Ceratopteris richardii	Other	43547
plants		1092
Bacteria		73
Fungi		11
Virus		20
Molluscs		2
Arthropods		2



Number of times ORFs aligned to each taxonomic group. *C. richardii* is separated from all other plant groups.

A closer look at the number of alignments within “Other plants”, separated by major plant lineage.Thesecontigsare candidate missing genes from the *Cr* annotation.

Of the non-Cr plant contigs, 248 had no match when mapped back to the *Cr* genome. Therefore **248 new genes** that do not exist in the *C. richardii* annotation have potentially been identified!

What does this mean?

248 new genes have been identified from data that would have otherwise been discarded.

Unmapped reads hold important biological information.

With new genomes being published frequently, it is important that we question their completeness, or risk losing vital information.

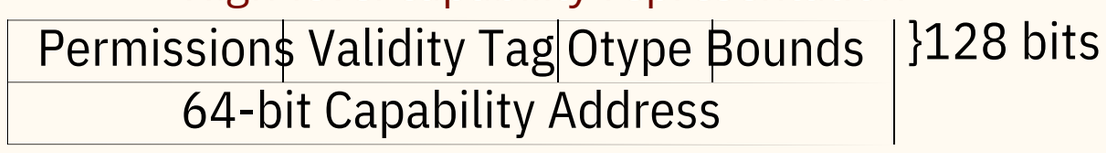
Evaluating the Microarchitectural Safety of CHERI

What is CHERI?

Capability Hardware Enhanced Risc Instructions is a joint research project of SRI International and the University of Cambridge [1].
The focus of this development has been on the CHERI Instruction-Set Architecture (ISA), which introduces an architecture-neutral capability-based protection model.
CHERI extends traditional pointers into what we call a capability:

- Capabilitiesarepointerswithadditionalmetadatathat strictly controls what they can have access to
- Providesstrongspatialandreferentialintegrity
- Securesmemory-unsafelanguageslikeC/C++with hardware-backed data protection

High-level capability representation



State of Affairs

The CHERI protection model has been applied into various base ISAs as extensions:

- CHERI-MIPS,CHERI-RISC-Vbeingexamplesofmature applications
- ArmMorello,aexperimentalextensionofArmv8-A
  - Morello boards in use for wider testing in academia and industry

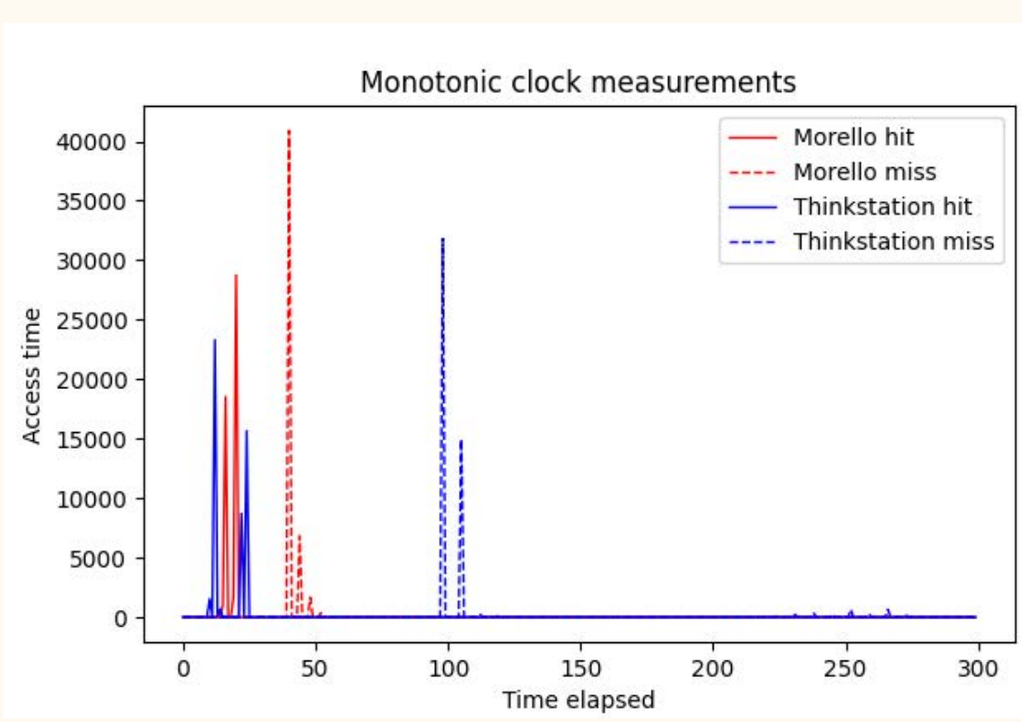


Figure 1: Measurements using FLUSH+RELOAD to measure cache access time on Arm Morello vs an x86 system with libflush

Microarchitectural Attacks

In recent years attacks have been found that take advantage of microarchitectural behaviour. One type of these attacks are transient-execution attacks, which combine:

- Speculativeexecution(inducederroneously)
- Microarchitecturalcovertchannelse.g.cacheside-channels

After mistraining:

```
if (x < array1_size)
y = array2[array1[x]]
```

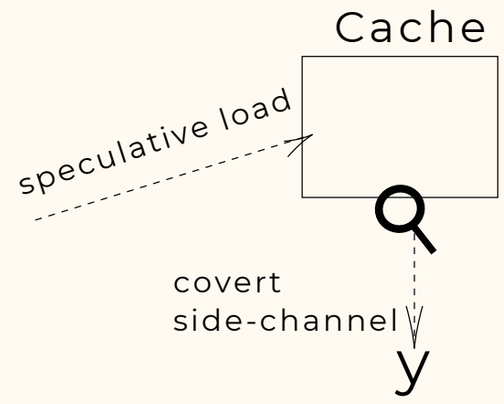


Figure 2: An example of a typical variant of a transient-execution attack [2], with x as attacker-controlled input. The branch pre- dictor is mistrained so that line 2 is speculatively executed with an out-of-bounds x, which puts y into the cache. The value of array1[x] can then be recovered.

Protections against certain variants exist but often introduce large performance penalties.

High Performance Computing

Data Confidentiality

- CertainHPCenvironmentsneedguaranteesthatdatais confidential [3]
  - e.g.Medical,defense,orintelligence
- Transient-executionattacksthreatenthisconfidentiality

Compartmentalisation

HPC systems are naturally quite compartmentalised already!

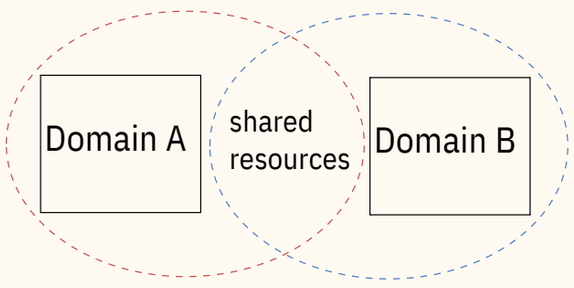


Figure 3: High-level abstraction of software compartmentalisation. In a security context, isolating a system can mitigate the impact of attacks.

CHERI not only inherently provides software compartmentalisation but improves scalability

Current Work: Transient-execution attacks on Morello

Focus is on testing Arm Morello's susceptibility to current transient-execution attacks

- Background:Currentexistingworkproposes CHERI-basedprotectionsagainstmicroarchitecturalattacks

Cache Side-channel on Morello

- WorkingonportingthelibflushlibrarytocheriBSD[4]
- Preliminarytestsshowexpectedresultswith FLUSH+RELOAD(seeFigure1)
- Refiningporttousehardwareperformanceregistersforcyclemeasurements
- Nextsteps:Testcurrenttransient-executionattackvariantsi.e.knownvariantsofSpectreandMeltdown

Research Questions

- HowdodifferentCHERIapplicationscompareintermsofmicroarchitecturalsafety?
- HowcantheCHERImodelprotectagainstmicroarchitecturalattacks?
- Does CHERI make a system more prone to physical attacks, and to what extent?

[1] R. N. M. Watson, P. G. Neumann, J. Woodruff, et al., "Capability Hardware Enhanced RISC Instructions: CHERI Instruction-Set Architecture (Version 8),", Oct. 2020.
[2] P. Kocher, J. Horn, A. Fogh, et al., "Spectre Attacks: Exploiting Speculative Execution,", Jan. 2018.
[3] S. Peisert, "Security in high-performance computing environments," Communications of the ACM, vol. 60, no. 9, pp. 72–80, Aug. 23, 2017.
[4] M. Lipp, D. Gruss, R. Spreitzer, C. Maurice, and S. Mangard, ARMageddon: Cache Attacks on Mobile Devices, Nov. 2015.

Jacqueline K. Henes, David Oswald
jkh703@student.bham.ac.uk
University of Birmingham

This research is funded by the Digital Security by Design (DSbD) Programme delivered by UKRI.

CAP TEE

UNIVERSITY OF BIRMINGHAM

Andrei Leonard Nicusan
aln705@student.bham.ac.uk

Dominik Werner
Jack Sykes
Dr. Kit Windows-Yule
Prof. Jonathan Seville

Data-Driven Engineering Open-Source Ecosystem

Learning Simulation Parameters from Experiments

Problem: what microscopic parameters do I need to match a macroscopic measurement?

Manual calibration is tedious and error-prone
Grid-based calibration is expensive and limited

ACCES

Autonomous Calibration and Characterisation using Evolutionary Software

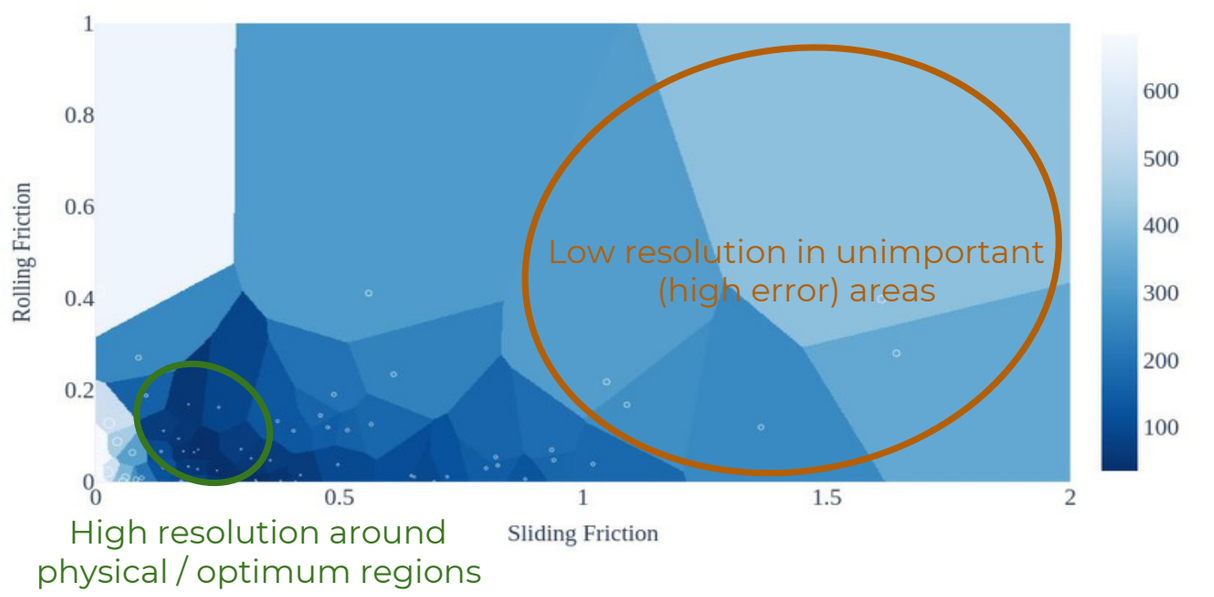
Vary < **Evolutionary Optimisation** > **Minimise**

Free parameters → Simulation → "Difference to Reality"

e.g. Friction, Restitution, Cohesive Energy Density, etc.

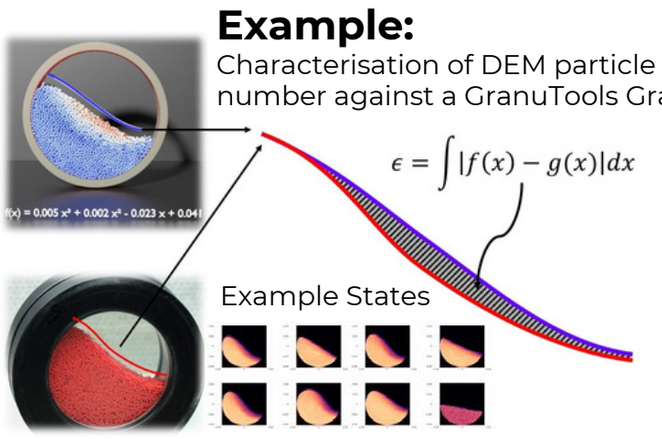
e.g. DEM, Geometry Design, DNS, Monte Carlo, etc.

e.g. Velocity Distribution, Free Surface Shape, Residence Time, etc.

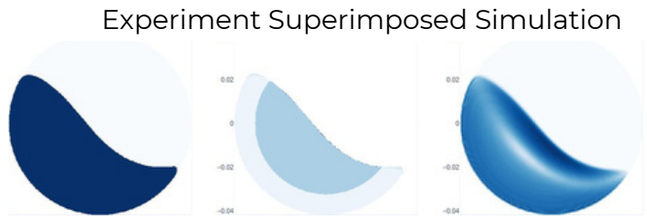


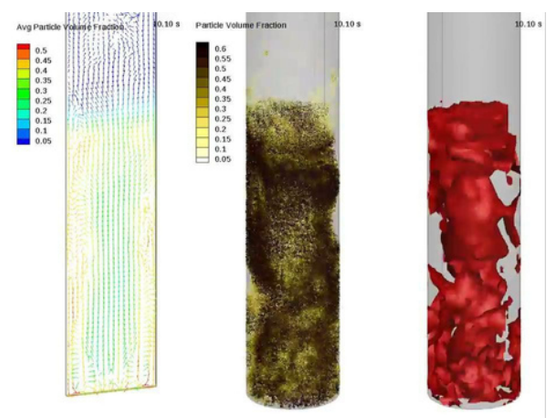
Example:

Characterisation of DEM particle friction, restitution and particle number against a GranuTools GranuDrum-imaged free surface shape.



Experiment Superimposed Simulation





Calibrating & Optimising CFD-DEM Fluidised Bed:

1.**Calibrate DEM particles** using a macroscopic measurement, e.g. reproduce PEPT velocity distribution

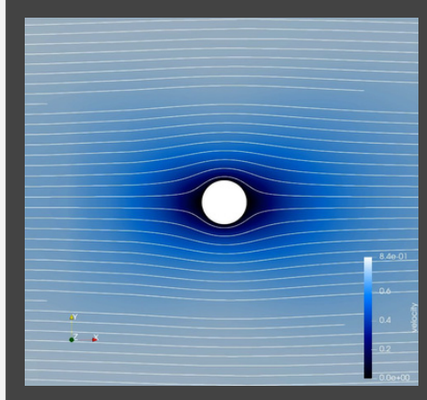
2.**Optimise bed design** - distributor plate design, bed geometry

Both using ACCES

M²E³D: Multiphase Materials Exploration via Evolutionary Equation Discovery

Autonomously discover the equations required to accurately reproduce a system's full, 3D dynamics.

1. Re-"discovering" Stokes' Law



- Used direct numerical simulations (DNS) of flow around a static 3D sphere
- Ran 32 combinations of the particle radius, fluid dynamic viscosity and velocity (0.02 < Re < 0.7)
- With **no a priori information**, we let M²E³D discover the equation fitting the resulting drag force

M²E³D found the equation on the left!

$$\frac{(-0.249 - 1.039)uR}{\frac{-0.069}{\mu}} = 18.67\mu Ru \approx 6\pi\mu Ru$$

Discovering Equations / Laws from Data

Classical approach: find a predetermined model's / equation's coefficients that fit experimental data.

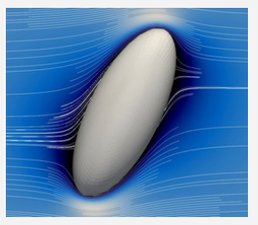
ML / AI:find a blackbox, uninterpretable surrogate model.

M²E³D: **Find the equation itself!**

2. Novel Correlation for Ellipsoids

- Apply **M²E³D** to open problems: **drag on Ellipsoids for Re > 133** As a function of aspect ratio (A), orientation (ϕ) and Re
- Tested up to **Re < 1500 with < 3% error**!

$$C_D(A, \phi, Re) = \frac{5.68 + \phi A + \frac{Re+19}{A}}{Re^{0.96}}$$



Experimental or Simulated Measurements → Discover Multiphase Drag Model → Implement in CFD-DEM → Calibrate Model against Experiment → Optimise Entire System → Discover Correlations

Any Data | M²E³D | **Any** Engine / Workflow | **ACCES** | **ACCES** | M²E³D

Interactive Example

OBJECTIVE TRACKING OF TROPICAL CYCLONES AND THEIR IMPACT ON RELEVANT WIND FIELDS IN THE WEST-PACIFIC FOR CONSTRUCTION OF PHYSICALLY CONSISTENT EVENT SETS

https://doi.org/10.5194/egusphere-egu23-3390

Xiaoqi Zhang

email: xxz252@student.bham.ac.uk

TOPOLOGICAL BIOMARKERS FOR COLORECTAL CANCER PROGNOSIS

Sebastian Gilbert

email: sgg897@student.bham.ac.uk

ELUCIDATION OF THE MECHANISM OF POLY(ETHYLENE) TEREPHTHALATE DUAL-CATALYSED ALCOHOLYSIS

Matt Price

email: mjp930@student.bham.ac.uk

MODELLING AMORPHOUS METAL-OXIDE FILMS USING AB-INITIO MOLECULAR DYNAMICS

Chris Owen

email: cbo738@student.bham.ac.uk