

The background is a solid teal color with a pattern of white polka dots of varying sizes scattered across the entire surface. In the center, there is a white rectangular border containing the text.

WPCCS  
2018

The postgraduate population in Computer Science continues to grow and there are now more than 100 registered PhD students in the Department. At the same time, the number of academics in Computer Science is also growing and we are pleased to be adding expertise in areas including database systems, multi-agent systems, graph algorithms, multimedia computing amongst others. Research students and staff are enjoying the benefits of the Department's association with the Center for Urban Science in New York and the Alan Turing Institute in London, as well as with the many companies and organisations that work with us.

In September we expect to expand into our new 6-storeys high Mathematical Sciences Building, a project that has won £2 million in funding from the Wolfson Foundation. We often have the pleasure of meeting alumni, who are now business leaders, wealthy entrepreneurs or leading academics. They remind us that Warwick continues to be an excellent institution in which to conduct research, with some of the strongest academic departments in the country, and that Computer Science has an enviable reputation for world-class research which is at the forefront of change in our society.

Best wishes with WPCCS 2018 and I hope you have an exciting and instructive day.

- Ranko Lazic

It is my pleasure to welcome you to the 2018 Warwick Postgraduate Colloquium in Computer Science (WPCCS). Following a great success last year of the reloaded WPCCS, the organizers worked very hard to match the expectations for another great event. It is now your turn to contribute by sharing your research experiences, ideas, and visions with your peers.

What brings you all together is your passion for innovation and an inherent desire to achieve something outstanding in the field of Computer Science or Urban Science. Unlike most of your former colleagues you have chosen the unknown path, yet you are confident that your perseverance to face the murky facets of research and your ingenuity to solve open problems will guide you through. And they will.

You have been fortunate to explore your talent in Warwick's Department of Computer Science, which is one of the most fertile breeding grounds for creativity like yours. Our department is equipped with a fine team of researchers who once undertook unique challenges like yours. Not only did they succeed, but they are now here because they want to further succeed through you. Learn from them, leverage their experience and intuition, but do remain assertive by trusting your instincts. Working closely together will largely narrow the gap to your joint success.

WPCCS is an excellent opportunity to widely expose where you currently stand. Speak up and do enjoy the challenge!

- Florin Ciucu

Welcome to  
**WPCCS 2018**

---

The Oculus  
Fri 29th June

# Guest Talks

## Design Patterns in 2018

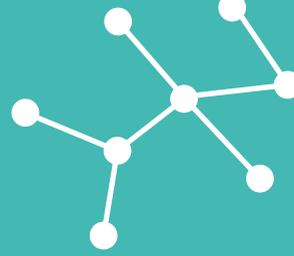
**Michael Gale · Presentation at 9:15am in OC0.02**

Most of us are familiar with design patterns in object-oriented programming languages, which attempt to capture reusable ideas with respect to the design of software. However, these design patterns are often viewed as overly convoluted and their implementations are rarely reusable. In this talk, we will take a look at design patterns in functional programming languages. Unlike their object-oriented relatives, functional design patterns are grounded in abstract mathematics and their implementations are highly reusable. To illustrate this, we will see how many useful functions can be defined in terms of a single, overloaded function. This allows us to write highly reusable code which works on almost arbitrary data structures.

## FPGAs: From Glue Logic to High Performance Compute Acceleration

**Suhaib Fahmy · Presentation at 10:55am in OC0.02**

Field Programmable Gate Arrays (FPGAs) emerged over three decades ago, offering a way to build small compute functions to bind discrete components in electronic systems. Since then, their capabilities have increased tremendously, resulting in their use today in wide-ranging application domains, from high frequency trading, through bioinformatics, networking, high energy physics, to machine learning. They offer one way to overcome the stalled performance scaling of traditional Von Neumann processor architecture, while also addressing the energy challenges posed by massive scale computation. This talk will discuss the evolution of FPGAs, how they differ from other architectures, some of the exciting ways they are being used today, and exciting research on the horizon.



## WebEXR: An Online High Dynamic Range Image Viewer

**Jonathan Hatchett · Presentation at 1:30pm in OC0.02**

We present a tool for sharing full range high dynamic range (HDR) images online. Using Emscripten together with existing C++ decoding libraries, HDR image content is made available to JavaScript in a web browser environment. Using WebGL, the results are presented to the user with real-time, interactive tone-mapping. We compare the decoding performance of Emscripten and native code, showing less than an order of magnitude difference between the two.

Web content has traditionally been limited to 8-bit, low dynamic range (LDR) images. With HDR imaging, all of the light in the scene can be captured, processed and displayed. The existing pipeline for sharing HDR images online is to convert the data to an 8-bit representation, known as a tone-mapping, before uploading the resulting image to the internet. The quantisation required is one-way and the inverse cannot be calculated exactly. Therefore, the final experience of the image is reduced as information is lost.

By providing full HDR content to the browser, tone-mapping is left up to the user, who can choose to view the information they are most interested. This view can be shared with other users simply with a reference to the image and the metadata.

# When & Where

## SCHEDULE

---

9:00 am Welcome from Prof. Ranko Lazic in OC0.02

9:15 am Guest talk from Michael Gale in OC0.02

---

### 9:45 am Tracks 1, 2 & 3

10:35 am Break with refreshments in foyer

10:55 am Guest talk from Suhaib Fahmy in OC0.02

### 11:25 am Tracks 1, 3 & 4

12:30 pm Lunch and posters in ground floor foyer

1:30 pm Guest talk from Jonathan Hatchett in OC0.02

---

### 2:00 pm Tracks 3, 5 & 6

3:05 pm Break with refreshments in foyer

### 3:25 pm Tracks 5, 7 & 8

### 5:05 pm Prize giving in OC0.02

---

## POSTERS

---

Presented in the ground floor foyer at lunch

### Theory, Foundations & D.M.

Tejas Kulkarni

### Education & Reliable Systems

Nouf Almujaali  
Huda Alrashidi  
Ebtehal Quqandi

### Machine Learning & A.I.

Elena Kochkina  
Nicole Peinelt  
Liam Steadman

### Image, Vision & Comp. Bio.

Najah Alsubaie  
Simon Graham  
Talha Kaiser  
Ian Tu  
Zhuo Wan

### Urban Science

Zakiyya Adam  
Alex Caton  
Henry Crosby  
Katherine Harris  
Vikki Houlden  
Melissa Kenny  
Maria Liatsikou  
Vangelis Pitidis  
Bradley Sheridan  
Timothy Sit  
Isabella Slattery

### Networks

Al Alharbi  
Jack Kirton

### High Performance Computing

Mohammed Alghamdi

**TRACK 1 · OC1.01**  
Theory, Foundations &  
Discrete Mathematics

Eleanor Davies  
Alex Dixon  
David Purser

**TRACK 2 · OC1.08**  
Education &  
Reliable Systems

Noor Hasimah Ibrahim Teo  
Dimah Al-Fraihat  
Denys Flores Armas

**TRACK 3 · OC1.09**  
Machine Learning &  
Artificial Intelligence

Qingzhi Ma  
Muhammad Shaban  
Junyu Li

**BREAK**

**TRACK 1 continued**  
Theory, Foundations &  
Discrete Mathematics

Yani Pehova  
Jacques Dark  
Charlie Dickens  
Christopher Hickey

**TRACK 4 · OC1.08**  
Networks

Matthew Bradbury  
Jasmine Grosso  
David Richardson  
Betty Agbons

**TRACK 3 continued**  
Machine Learning &  
Artificial Intelligence

Zhenyu Li  
Man Luo  
Edward Chuah  
Ania Lisowska

**LUNCH**

**TRACK 5 · OC1.01**  
Urban Science

Neha Gupta  
Konstantin Klemmer  
Shanaka Perera  
Elisabeth Titis

**TRACK 6 · OC1.08**  
Security

Sara Alhajaili  
Bowen Du  
Mansour Aldawood  
Farrukh Qazi

**TRACK 3 continued**  
Machine Learning &  
Artificial Intelligence

James Van Hinsbergh  
Helen McKay  
Caroline Player  
James Archbold

**BREAK**

**TRACK 5 continued**  
Urban Science

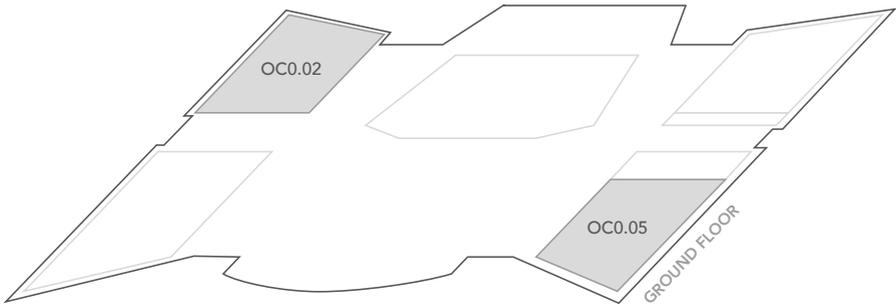
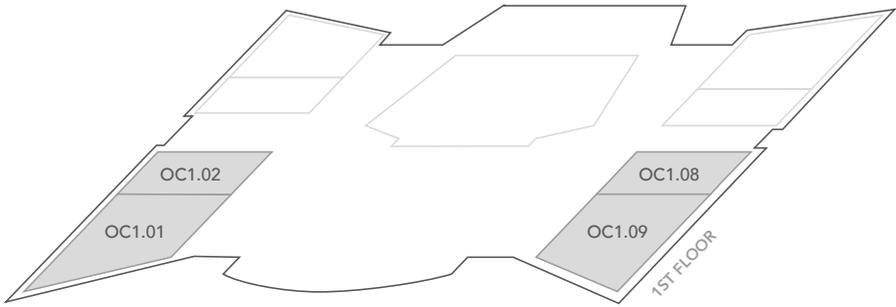
Aseel Alturki  
Corinne Muir  
Oba Waiyaki  
Philipp Ulbrich  
Ciaran Devlin  
Amit Chaudhary

**TRACK 7 · OC1.08**  
Image, Vision &  
Computational Biology

Greg Watson  
Abdullah Algamdi  
Ruqayya Awan  
Navid Alemi Koohbanani  
Jamie Bayne  
Mary Shapcott

**TRACK 8 · OC1.09**  
High Performance  
Computing

Dom Brown  
Andrew Owenson  
Dean Chester  
Richard Kirk  
Mohammed Maray  
David Truby



## **50th ANNIVERSARY**

OC0.05

To celebrate the 50th anniversary of Computer Science at Warwick, we have assembled a collection of photos and archives of the past 50 years in OC0.05.

## **PRIZES**

Awarded in  
OC0.02 at  
5:05pm

WPCCS aims to engage attendees with their colleagues' research and with research in the wider community. For this purpose, the Department of Computer Science and the Research Student Skills Programme have kindly agreed to sponsor the event.

For today's posters and presentations, 12 prizes will be awarded by the Programme Committee. An Amazon voucher will be awarded per session for the presentation that best informs and educates with engaging deliverance. Two further prizes will also be awarded, 'Best Poster in Colloquium' and 'Best Presentation in Colloquium'.

# Theory, Foundations & Discrete Mathematics

Track 1

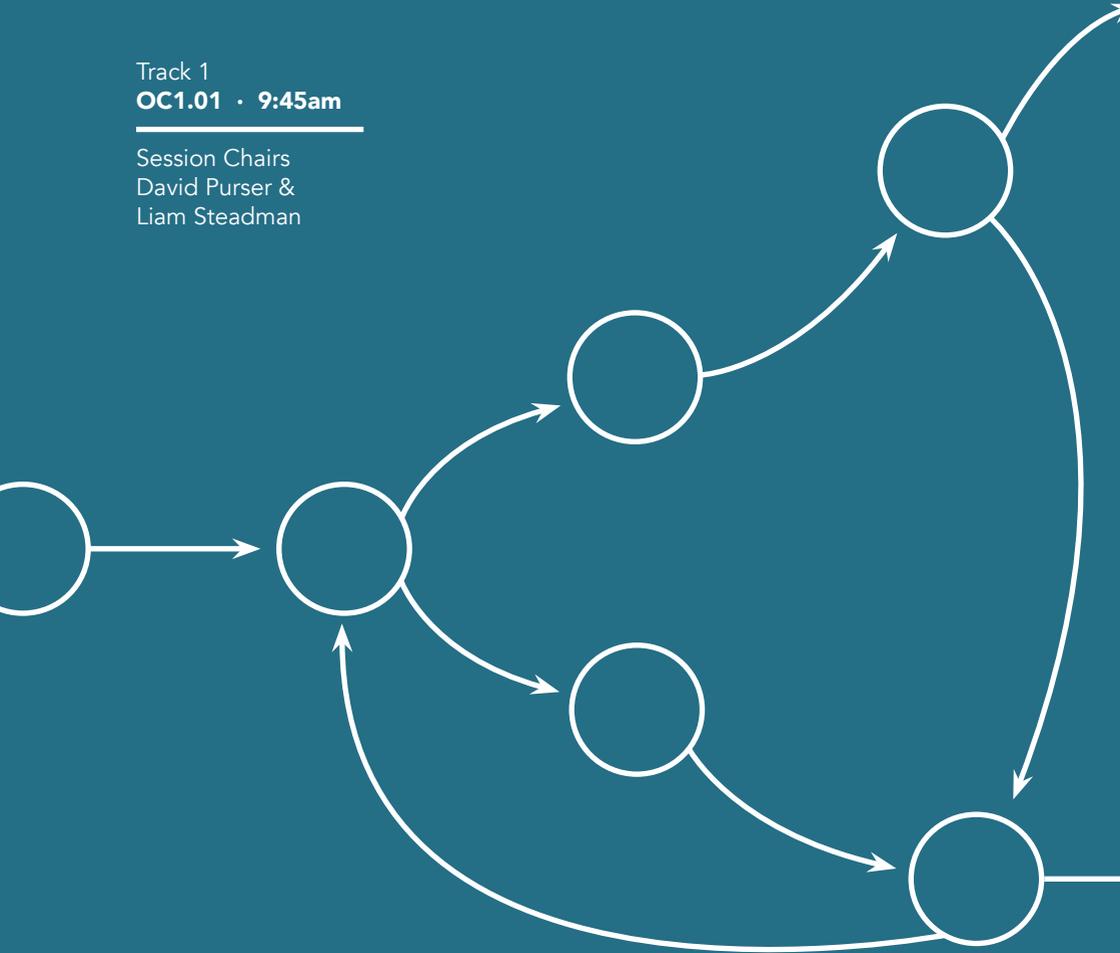
**OC1.01** · 9:45am

---

Session Chairs

David Purser &

Liam Steadman



# On the Soundness of Miniphases

**Eleanor Davies · Presentation at 9:50am**

The Miniphase approach is used to fuse compiler phases, allowing them to be specified separately but run in combination. This lets developers implement a large number of phases which each perform a single task, making the compiler easier to maintain and debug, whilst still achieving good performance. In its implementation in Dotty, a project developing the next major version of Scala, compilation using these fused phases has been shown to be 35% quicker than executing them consecutively. We look at how to model and formalise Miniphases in order to prove the soundness of this approach, ensuring that fusing phases does not change the result of compilation

---

# Efficiently Solving Decision Problems on Petri Nets

**Alex Dixon · Presentation at 10:05am**

Petri nets are networks augmented with counters in each place. The formalism is gaining traction in the domains of both formal methods (for the verification of programs) and in abstracting real-world systems such as transportation and traffic flow modelling. Many properties of these models can be characterised by the coverability problem - finding whether there is a limit to the counts in some set of places. Efficiently solving the coverability problem on Petri nets is essential for the practical application of these formal methods. Naive solutions to the coverability problem may operate in no better than 2-EXP time. Through careful pruning this bound can be reduced significantly. We aim to implement recent academic findings and bring them to an effective, practical and extensible framework.

---

# Bisimilarity Distances for $\epsilon, \delta$ Differential Privacy

**David Purser · Presentation at 10:20am**

Differential privacy often relies upon manually crafted mathematical proofs, we look to use automata-based modelling to automatically verify differential privacy properties. The Kantorovich metric has been used to analyse bisimilarity of Markov chains (Desharnais et al., 2004) and probabilistic automata (van Breugel, Worrell, 2014) and was later generalised to analyse  $\epsilon$ -differential privacy (Xu, Chatzikokolakis, Lin, 2014). We present a work-in-progress to further generalise the Kantorovich metric to analyse  $\epsilon, \delta$ -differential privacy, using a distance introduced by Barthe et al. (2012). We consider its properties, how it can be calculated and problems it can be applied to.

---

# Ordered Ramsey Theory

**Yani Pehova · Presentation at 11:30am**

The Ramsey number  $R(H)$  of a graph  $H$  is defined as the smallest number  $n$  such that every colouring of the edges of the complete graph on  $n$  vertices in two colours contains a monochromatic copy of  $H$ . This notion was first introduced by Ramsey in 1930 in a paper called “On a problem of formal logic”, where he proved that  $R(H)$  exists for all graphs  $H$ . We consider a variant of this problem in which  $H$  is an ordered graph and we seek a monochromatic ordered copy of  $H$ . In this setting different orderings of the same graph can have vastly different Ramsey numbers. We will survey some existing results on this topic and present new findings, as well as some open questions. This is joint work with J. Geneson, A. Holmes, B. Lidicky, X. Liu, D. Neidinger and I. Wass.

---

# Maximum Independent Set in Streams of Vertices

**Jacques Dark · Presentation at 11:45am**

Maximum independent set (MIS) is a fundamental problem on graphs. Unfortunately, MIS is known to be hard to approximate in both offline and edge-arrival streaming models. We show that MIS remains hard for general graphs in the vertex-arrival streaming model.

However, there are interesting classes of graphs which do admit efficient approximation algorithms for MIS. We show algorithms and lower bounds for some types of geometric intersection graphs in two variants of the vertex-arrival model. We also consider the weighted version of MIS.

---

# Data Summarisation Techniques to Exploit Sparsity in Regression Problems

**Charlie Dickens · Presentation at 12:00pm**

Regression problems are ubiquitous in machine learning but solving them exactly can often be costly. Various randomised algorithms have been proposed to solve forms of regression more efficiently; we show that a data summarisation technique, ‘CountSketch’, from the streaming literature has a wider reaching application and can often slot in to previously proposed methods for solving different types of regression. The advantage of CountSketch is that it can be applied to data as it is seen, and in time proportional to the number of nonzeros in the dataset - this can yield a significant speedup in comparison to other random transforms.

---

# Practically Outsourcing Verification for Linear Algebra

**Christopher Hickey · Presentation at 12:15pm**

Our work is looking to improve protocols for the practical verification of outsourced linear algebra. We consider two parties, a client with low resources, and a untrusted helper, who claims to be able to perform the computation. This is achieved by having the helper giving more information alongside the solution, allowing the client to use a small amount of computing power to correctly verify the response with high probability.

We explore non-interactive and multi-round protocols for matrix multiplication and vector-matrix-vector multiplication and analyse the real-world practicality of these approaches. We investigate the costs for the client, as well as trying to keep the overheads for the helper at a minimum, and consider what is causing bottlenecks. We discuss the potential uses of these protocols, specifically examining the prospective application to distributed machine learning.

---

## Constrained Private Mechanisms for Count Data

**Tejas Kulkarni · Poster presented in foyer**

Differential privacy has emerged as an accepted model to release sensitive information while giving a statistical guarantee for privacy. We focus on the core problem of count queries, and seek to design mechanisms to release data associated with a group of  $n$  individuals. Prior work has focused on designing mechanisms by raw optimization of a loss function, without regard to the consequences on the results. This can lead to mechanisms with undesirable properties, such as never reporting some outputs, and over-reporting others. We tame these pathological behaviors by introducing a set of desirable properties that mechanisms can obey. Any combination of these can be satisfied by solving a linear program (LP) which minimizes a cost function, with constraints enforcing the properties. We focus on a particular cost function, and provide explicit constructions that are optimal for certain combinations of properties, and show a closed form for their cost.

# Education & Reliable Systems

Track 2

**OC1.08** · 9:45am

---

Session Chair

Dean Chester



# Ontological Approach to Automatic Question Generation

**Noor Hasimah Ibrahim Teo · Presentation at 9:50am**

Machine generated questions research has been widely studied. Developing automatic methods for Question Generation (QG) helps to reduce instructors' effort in preparing assessment questions. Different strategies have been study to transform semantic information from online resources such as ontology, Wikipedia and learning material into questions. In this work, we explore strategies to construct formative assessment questions using ontology. We have identified 3 challenges in generating an automatic question from an ontology for educational purposes. Therefore, we divided each of this challenge into different stages. The result of the experiments will be discussed.

---

# Evaluating E-Learning Systems Success - The Case of Moodle at Warwick University

**Dimah Al-Fraihat · Presentation at 10:05am**

Although much research on e-learning has focused on the pedagogies and the technologies used, assessment of the success of e-learning tools has not received a holistic or unified approach. The current research investigates the measurement factors that influence the success of e-learning systems and accordingly proposes a multi-dimensional model for assessing the success of e-learning. To achieve this purpose, a model was developed based on intensive literature review and analysis of four approaches for measuring the success of e-learning as a theoretical basis: DeLone and McLean IS success model (D&M model), Technology Acceptance Model (TAM); User Satisfaction Models; E-learning Quality Models. This new model is believed to be holistic because different perspectives have been considered in relation to quality, usefulness, user satisfaction, attitude, social factors and benefits of using the e-learning systems, and these encompass the main components of the existing four approaches.

## Know-Your-Adversary: Towards a Forensically-Aware Architecture for Analysing Malicious Insider Actions against Transactional Databases

**Denys Flores Armas · Presentation at 10:20am**

Database audit records can be used as forensic evidence to investigate suspicious actions against transactional databases. Currently, these actions are being perpetrated by trusted users (insiders) who interact with databases using data manipulation (DML) operations. Although the investigation of such actions rely on audit records, these must be produced during the normal operation of a forensically-aware database architecture. I.e. following an unbroken accountability trail known as Chain-of-Custody (CoC) to ensure their proper production. Hence, our research aims to define the functional requirements of such an architecture, assuming role segregation, provenance, event timelining and causality as CoC requirements, and introducing forensic controllers for the generation, collection and preservation of audit records. These controllers use vector clocks in centralised settings, and hybrid logical clocks in distributed environments to produce a timeline that can be used for analysing provenance and causality of DML operations derived from insider activity.

---

## Toward a New Model for the Adoption of Web 2.0-based Knowledge Sharing Platform Amongst Academics in Saudi Universities

**Nouf Almujally · Poster presented in foyer**

Knowledge sharing is a significant source of success in knowledge management, however, in Saudi Arabian universities, knowledge management is often insufficient when it comes to web-based knowledge sharing amongst academics. To ensure that a knowledge sharing platform can be well implemented and used when communicating internally in an academic context, there is a need to know why academics accept or reject the use of web-based knowledge sharing platform. Therefore, the aim of this poster is to investigate the important factors that would influence academics' behaviour towards adopting the use of web 2.0- based knowledge sharing platform in Saudi Arabian universities. A conceptual model is proposed based on the Unified Theory of Acceptance and Use of Technology (UTAUT) model, as well as other factors which are explored in knowledge sharing literature reviews to enrich the proposed model. Then, the model will be evaluated using mixed method approach.

---

## Automatic Reflective Writing Analysis Based on Semantic Concepts

**Huda Alrashidi · Poster presented in foyer**

This study proposes a semantic-based approach for reflective writing analysis to overcome the limitations of existing shallow processing approaches. The semantic analysis examines the meaning of linguistic input that depends mainly on transferring words, phrases and sentences into concepts in knowledge sources, such as WordNet-Affect and analysing the relationships of mapped concepts in the underlying knowledge sources. The proposed reflective writing analysis approach focuses on the efficiency of using affectional and emotional concepts identified in WordNet-Affect to classify text into reflective or non-reflective.

---

## Mobile Augmented Reality in Educational Environments

**Ebtehal Quqandi · Poster presented in foyer**

The possibility of using Augmented Reality (AR) in learning and training has become more straightforward than before, as a result of the extensive use of ICT in the computer and mobile industries. Even though AR is used in education, and a general acknowledgement it has a positive impact on learning outcomes, the value of integrating AR applications into learning environments has not yet been fully investigated (Diegmann et al., 2015). This study considers the Integration of AR technology into nursing clinical lab training, introduces new ways of interacting with the manikins and allows students to view patient scenarios instead of relying on teacher explanations. AR it allows students to visualize hidden objects such as internal organs, which makes simulations more realistic and immersive. The study aim to investigate the potential of this technology in term of improving nursing students' self—learning.

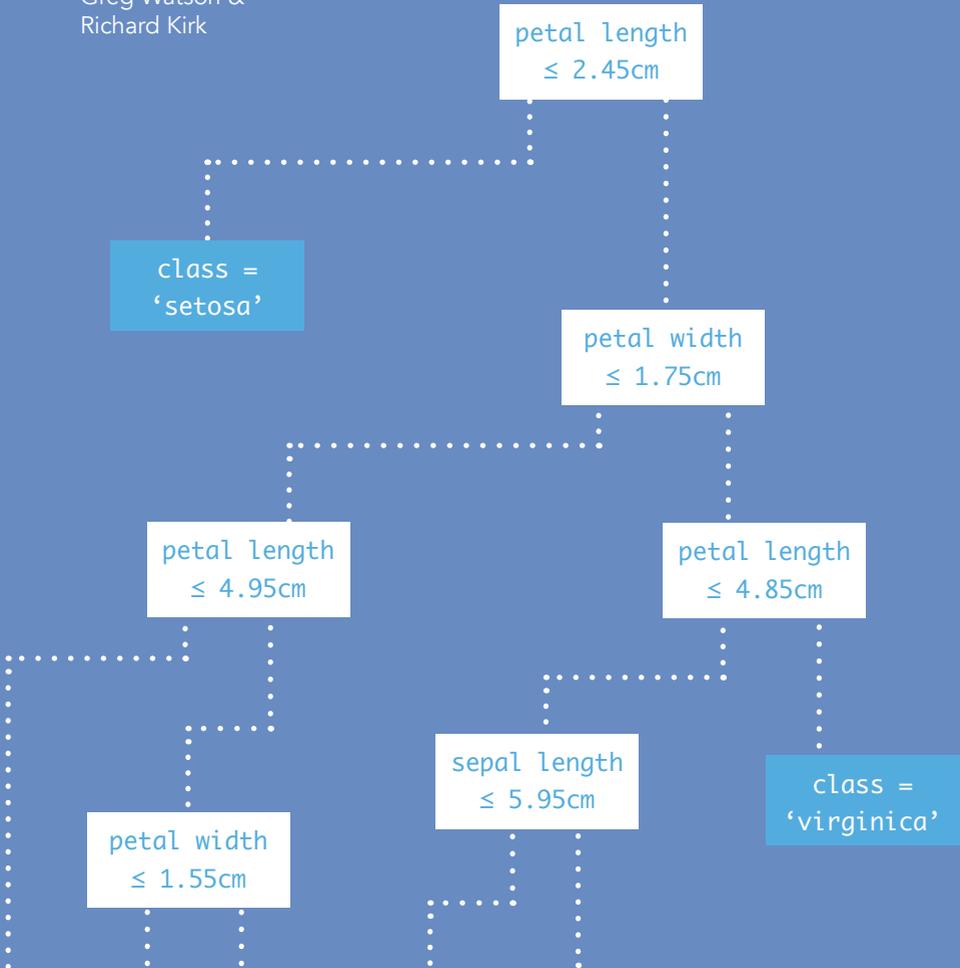
# Machine Learning & Artificial Intelligence

Track 3

OC1.09 · 9:45am

---

Session Chairs  
Greg Watson &  
Richard Kirk



## Query-Centric Regression for In-DBMS Analytics

**Qingzhi Ma · Presentation at 9:50am**

Research in data analytics is receiving increasingly greater attention from the data management community and efforts are underway to enrich DB functionality with Machine Learning models for data analytics. In this realm, regression models, being a principal means for predictive analytics, are of particular interest. This study addresses the problem of empowering DB systems with regression models. In data management it is widely recognized that query distributions do not necessarily follow data distributions. Hence, when queries target data subspaces on which a model is weak, the prediction accuracy suffers. This study puts forth a new ensemble model QReg that addresses this problem, while also offering better overall expected accuracy. The evaluation will utilize both query-centric metrics as well as overall-workload-centric metrics. Furthermore, QReg is currently be applied to approximate query processing to serve aggregate queries like AVG, COUNT, etc in HIVE.

---

## Automated Detection of Tumour Infiltrating Lymphocytes and Invasive Front in OSCC Histology Images

**Muhammad Shaban · Presentation at 10:05am**

Oral cancer is among the top ten cancers in the world and tumour infiltrating lymphocytes (TILs) and invasive front (IF) are two key prognostic indicators. An automated method for TILs and IF detection can facilitate better stratification and prognostic prediction of oral cancer patients. We propose an end-to-end solution for detection of TILs and IF in cancer histology images, the first of its kind in the literature. We use a novel deep learning based method to segment the histology image into regions of four clinically significant classes. We employed statistical colocalization measure (the Morisita-Horn index) for TILs detection and for invasive front detector we utilise the predicted region labels to extract global and local features. We evaluate our TILs detection method using ROC measure and the proposed method achieved 0.77 AUC on test dataset. Our invasive front detection also shows promising results with 95% detection accuracy.

# Data Fine-Tuning: A Simple Way to Accelerate Neural Network Training

**Junyu Li · Presentation at 10:20am**

Training neural network is the most time-consuming process before deploying to applications. With scaling up layers and parameters of the neural network, it has been shown to lead to increased performance in benchmark tasks and to enable discovery of sophisticated high-level features; on the other hand, it also presents enormous challenges where the core one is the training efficiency of the deep neural network. We present an approach that fine-tunes input data during training process by evaluating the performance of the data in real time. Predicting performance on test data shows that our approach achieves a quite similar accuracy comparing to the original networks but approximate 10% of time reduction.

---

# MapRDD: Finer Grained Resilient Distributed Dataset for Machine Learning

**Zhenyu Li · Presentation at 11:30am**

The Resilient Distributed Dataset (RDD) is the core memory abstraction behind the popular data-analytic framework Apache Spark. We present an extension to the Resilient Distributed Dataset for map transformations, that we call MapRDD, which takes advantage of the underlying relations between records in the parent and child datasets, in order to achieve random-access of individual records in a partition. The design is complemented by a new MemoryStore, which manages data sampling and data transfers asynchronously. We use the ImageNet dataset to demonstrate that: (I) The initial data loading phase is redundant and can be completely avoided; (II) Sampling on the CPU can be entirely overlapped with training on the GPU to achieve near full occupancy; (III) CPU processing cycles and memory usage can be reduced by more than 90%, allowing other applications to be run simultaneously; (IV) Constant training step time can be achieved.

## On Quantifying the Role of Tumour Infiltrating Lymphocytes in Head and Neck Squamous Cell Carcinoma

**Ania Lisowska · Presentation at 11:45am**

Head and neck squamous cell carcinoma (HNSCC) develops in the mucous membranes of the nose, mouth and throat and is among the most prevalent cancers worldwide. The purpose of this study is to investigate how the tumour microenvironment is affected by genetic mutations that have been previously associated with patient survival. Since host immune response to tumour was previously shown to play a major role in patient outcome, we specifically focus this analysis on tumour infiltrating lymphocytes (TIL) and their interaction with cancer cells. To do this, we first develop a deep learning algorithm to identify TIL-rich areas in Haematoxylin and Eosin stained Whole Slide Images of HNSCC histopathology sections. Then, we quantify the degree of tumour infiltration by TILs and finally, investigate the relationship between TILs infiltration, genetic mutations and patient survival. This study should improve our ability to make more precise outcome predictions based on patients genetic profile.

---

## Efficient On-Device Learning for Embedded and Mobile Platform

**Man Luo · Presentation at 12:00pm**

Deep Neural Networks have become ubiquitous in many application domains, ranging from computer vision to natural language processing. Though these neural networks are powerful, the large number of weights consume considerable storage and memory, making them difficult to deploy on embedded system with limited hardware resources and power budgets. To address this limitation, much work has been done to reduce the size of networks. This presentation will talk about the existing work to overcome this issue, mainly focus directly on fewer parameters but equivalent accuracy, making it possible to fit them on mobile device, and finally, the challenges and future work related to this problem will also be discussed.

## Assessing Inter-Component Error Propagation on HPC Resources from Cluster Log-Data

**Edward Chuah · Presentation at 12:15am**

Recent work have used both resource use data and failure logs to detect failure-inducing errors and to diagnose dependability issues such as error propagation in large cluster systems. A system failure typically occurs as a result of error propagation and the unsuccessful execution of error recovery protocols. In this presentation, a new framework which integrates feature analysis and correlation analysis will be presented. Application of the framework on the resource use data and message logs from the Ranger supercomputer have produced interesting results. The framework has: (i) identified significant resource use and significant messages, (ii) identified network packet drops and Lustre I/O errors, (iii) identified virtual memory allocation and harddisk I/O errors, and (iv) showed that times of different granularities are required for identifying the error propagation patterns.

---

## Vehicle Visit Detection Using In-Car Data

**James Van Hinsbergh · Presentation at 2:05pm**

GPS trajectories can be obtained using inexpensive hardware. Using the Gradient-based Visit Extractor (GVE), these trajectories can be used to calculate meaningful locations, or visits, of users. When applied in the automotive domain falsely identified visits are found, such as in heavy traffic where the vehicle is stationary for periods of time. To isolate these visits, we build a classifier using signals from the vehicle and determine the types of stops. In this paper, we (i) demonstrate that GVE using an existing error metric, is useful for location extraction in vehicles and (ii) build a classification model the used signals from the vehicle to eliminate noise generated by GVE. In particular, we use a support vector machine (SVM) with features selected using minimal redundancy maximal relevancy (mRMR), that achieves high accuracy.

# Online Transfer Learning for Concept Drifting Data Streams

**Helen McKay · Presentation at 2:20pm**

Transfer learning uses knowledge learnt in a source domain to aid predictions in a target domain. When both source and target domains are online they are both susceptible to concept drift, which alters the mapping of knowledge between source and target. This makes additional information available and necessitates its transfer both from the source to the target and vice versa. To address this we introduce the Bi-directional Online Transfer Learning framework (BOTL), which uses knowledge learnt in an online source domain to aid predictions in an online target domain. We present an empirical evaluation of the BOTL framework using two data stream generators: the drifting hyperplane emulator, and the smart home heating simulator. The results obtained show that BOTL outperforms the existing state of the art techniques that use transfer learning for online data streams.

---

# Addressing Concept Drift in Reputation Assessment

**Caroline Player · Presentation at 2:35pm**

Evaluating trust and reputation in environments where agent behaviours are highly dynamic is challenging. In this paper, we address the limitations of existing methods when agent behaviours can change at varying speeds and times across a system. Modelling such environments as a multi-agent system, we propose a method that expands on existing trust, reputation and stereotype models and uses concept drift detection to identify and exclude unrepresentative past experiences. Our method improves the selection of relevant data for evaluating trust and reputation, without excluding relevant historical data (as is the case with sliding windows and forgetting factors). We show that our approach enables agents to achieve higher utility than existing methods, by more accurately evaluating trust and reputation robustly with respect to dynamic agent behaviours.

## Indirect Influence Manipulation with Partial Observability

**James Archbold · Presentation at 2:50pm**

The propagation of concepts through a population of agents can be modelled as a cascade of influence spread from a set of individuals. In real-world environments there may be many concepts spreading and interacting, and we may not be able to control the target concept we wish to manipulate, requiring indirect manipulation through a controllable concept. Previous work on influence spread assumes that we have full knowledge of a network, which may not be the case. We investigate indirect influence manipulation when we can only observe a sample of the network. We propose a heuristic, known as Target Degree, for selecting seed nodes for a secondary controllable concept that uses the limited information in a partially observable environment to indirectly manipulate the target concept. Target degree is shown to be effective in small-world networks and in real-world networks when the controllable concept is introduced after the target concept.

---

## All-in-one: Multi-task Learning for Rumour Verification

**Elena Kochkina · Poster presented in foyer**

Automatic resolution of rumours is a challenging task that can be broken down into smaller components that make up a pipeline, including rumour detection, rumour tracking and stance classification, leading to the final outcome of determining the veracity of a rumour. In previous work, these steps in the process of rumour verification have been developed as separate components where the output of one feeds into the next. We propose a multi-task learning approach that allows joint training of the main and auxiliary tasks, improving the performance of rumour verification. We examine the connection between the dataset properties and the outcomes of the multi-task learning models used.

## A Context-based Classifier Prediction System for Chinese Language Learners

**Nicole Peinelt · Poster presented in foyer**

Classifiers are function words that are used to express quantities in Chinese and are especially difficult for non-native language learners. In contrast to previous studies, we argue that the choice of Chinese classifiers is highly contextual and present a novel publicly available dataset for sentence-based classifier prediction. Our context-aware machine learning models trained on this resource outperform previous word-based baselines. We further present use cases for our database and models in an interactive demo system.

---

## A Reduction Technique for Spatiotemporal Datasets

**Liam Steadman · Poster presented in foyer**

Often, processing and finding trends in large datasets can be too computationally intensive or require too much memory to be practical. We therefore need to reduce the quantity of data needed to be processed whilst answering queries without losing any information or affecting the query results. In these scenarios compression is not applicable and data sketching techniques have limited capabilities. We introduce a novel method for reducing spatiotemporal datasets and examine its use in a range of scenarios. We present the effectiveness of the method on a real world dataset collected from traffic counting sensors embedded on motorways in England.

# Networks

Track 4

**OC1.08 · 11:25am**

---

Session Chair

Dean Chester



# A Source Location Privacy-Aware Duty Cycle for Wireless Sensor Networks

**Matthew Bradbury · Presentation at 11:30am**

Source Location Privacy (SLP) is an important property for monitoring privacy-critical assets using wireless sensor networks. Many SLP-aware routing techniques exist with most striking a trade-off between various key metrics and SLP level. An example of such a metric is energy, since there is limited battery power available. To save energy the radio and CPU are duty cycled by putting them to sleep when no messages are expected to arrive. Since the radio is asleep, the duty cycle protocol can cause vital control messages to be lost, leading to reduced SLP levels. Any duty cycling protocol will need to ensure that (i) the duty cycle is as low as possible, and (ii) control messages are still transmitted in a timely manner. In this presentation, the first duty cycling protocol for an SLP-aware routing protocol will be presented. Simulation results show that a duty cycle of 10% can be achieved.

---

# Using Ant Colony Optimisation for Routing in Many-to-Many Wireless Sensor Networks

**Jasmine Grosso · Presentation at 11:45am**

Wireless sensor networks (WSNs) have been widely studied in the context of many-to-one communication, however, there is little research in the area of many-to-many communications. WSNs of this form consist of multiple source nodes broadcasting data towards multiple sink nodes. Many-to-many communication in WSNs is a growing application area, with examples including detection of fire in both natural and urban areas. There is a need for scalable many-to-many routing protocols in the context of WSNs that is adaptive to topology changes. The goal of this research is to create a distributed routing protocol for many-to-many WSNs that makes use of Ant Colony Optimisation (ACO), a heuristic inspired by the movement of ants in nature to find short paths to food sources. This protocol for WSNs is adaptive to changes in topology such as node failures, as well as being efficient with the use of energy resources.

## Broadcasting in Transiently Powered Sensor Networks

**David Richardson · Presentation at 12:00pm**

Over the past 15 years sensor node technology has improved leaps and bounds in computational power and energy efficiency. Modern sensor nodes can be powered by rechargeable battery cells or harvest energy from their environment, allowing for extended periods of node uptime. However, these power sources are not totally reliable and thus some downtime will occur. Nodes that crash mid-task need to recover to a previous state to re-attempt computation. Many techniques toward making this recovery process, such as checkpointing make use of stable storage on the node itself. We discuss using redundancy in the network to store and retrieve a sensor's state upon recovery, looking specifically at the problems posed by broadcasting in a wireless sensor network.

---

## TyfeSensor: An eHealth Sensor for Monitoring Relapse of Salmonella Typhi in Outpatient

**Betty Agbons · Presentation at 12:15pm**

Typhoid fever remains a major health problem in developing countries. Reliable mobile monitoring systems for observing patient physiological signals is important when designing hardware and algorithms for Healthcare Sensor Networks (HSNs). In this article, we propose mobile preventive monitoring methodology, which does not rely on bacterial culture or stool culture for monitoring Enteric fever. The methodology has its basis in biometrics and mobile sensor design. We present a generic design for a monitoring apparatus, which can be used in the monitoring methodology. We also present a specific design, which we use, in implementing an actual monitoring apparatus. We test the efficiency of our monitoring apparatus and to demonstrate the usefulness of the methodology, perform experiments akin to those required in projects involving biometric HSNs. Results from the said monitoring and experiments demonstrate that the methodology is an invaluable tool for software programmers and algorithm designers working with biometric HSNs.

## Monitors Selection for Predicate Detection in Wireless Sensor Networks

**Al Alharbi · Poster presented in foyer**

Wireless Sensor Network (WSN) is a new way to gather and monitor information by small and cheap hardware devices. A WSN consists of spatially distributed autonomous devices using sensors to monitor physical or environmental conditions, such as temperature, humidity, sound, vibration, pressure, motion or pollutants, at different locations. In WSNs not all the sensors are in the communication range of each other, so they cannot communicate directly with each other and with the base station. That's why communication or relaying data typically occurs via wireless multi-hop routing. The majority of WSN architecture contains source and sink, source nodes, which generate data; in our proposed system we will use sensors to measure temperature and humidity. Sink nodes collect all data gathered by source nodes. When WSN size increases it becomes difficult to control, manage and monitor all network nodes (source and sink).

---

## Towards Optimal Source Location Privacy-Aware TDMA Schedules in Wireless Sensor Networks

**Jack Kirton · Poster presented in foyer**

Source Location Privacy (SLP) is becoming important for wireless sensor networks where the source of messages is kept hidden from an attacker. In this work, we conjecture that similar traffic perturbation to altering the routing protocol can be achieved at the link layer through assignment of time slots to nodes. This paper presents a multi-objective optimisation problem where SLP, schedule latency and final attacker distance are criteria. We employ genetic algorithms to generate Pareto-optimal schedules using two fitness criteria, examining the Pareto efficiency of selecting either and confirming the efficiency by performing simulations which show a near optimal capture ratio.

# Urban Science

Track 5

**OC1.01 · 2:00pm**

---

Session Chair

Melissa Kenny



## Exploring Twitter and UK Land Registry Data to Understand Working Hours and Sentiment in Industry

**Neha Gupta · Presentation at 2:05pm**

Economic research on working hours has shown that the potential adverse effects of variable and longer working hours on happiness, is offset by better income and job satisfaction. Such studies to date have been conducted using data available from traditional sources such as surveys. Whilst surveys and self reported methods provide precise longitudinal data, they lack scalability and real-time monitoring. Therefore, we propose and test a new method for incorporating secondary sources of geo-spatial data, both real-time and static, to evaluate the working hours and sentiment of employees across several industrial sectors. We test our methodology on data from London UK and whilst our results reveal no general trends across industrial sectors, our scalable approach shows that for the financial and insurance sector, the duration of working hours does impact on employee sentiment. Our new data-driven, scalable methodology has application beyond the analysis of mood prevalence in various economic sectors.

---

## Community Structures, Interactions and Dynamics in London's Bicycle Sharing Network

**Konstantin Klemmer · Presentation at 2:20pm**

Bikesharing schemes are transportation systems that not only provide an efficient mode of transportation in congested urban areas, but also improve last-mile connectivity with public transportation and local accessibility. Bikesharing schemes around the globe generate detailed trip data sets with spatial and temporal dimensions, which, with proper mining and analysis, reveal valuable information on urban mobility patterns. In this paper, we study the London bicycle sharing dataset to explore community structures. Using a novel clustering technique, we derive distinctive behavioural patterns and assess community interactions and spatio-temporal dynamics. The analyses reveal self-contained, interconnected and hybrid clusters that mimic London's physical structure. Exploring changes over time, we find geographically isolated and specialized communities to be relatively consistent, while the remaining system exhibits volatility, especially during and around peak commuting times. By increasing our understanding of the collective behaviour of the bikesharing users, this analysis supports policy appraisal, operational decision-making and motivates improvements in infrastructure design and management.

---

# Spatiotemporal Clustering in Commercial Properties

**Shanaka Perera · Presentation at 2:35pm**

Spatial clustering is a popular topic to academics and industry alike. Real estate professionals are looking to maximise their profit on the use of land, whereas academics may be interested in the long term sustainability of a city, environmental phenomena's or other geographies. The choice of industrial location affects the national spatial distribution of resources and where the population agglomerates. As such, I will explore to what extent the choice of location is “accidental” or a function of space and time. Furthermore, this research attempts to apply unsupervised machine learning techniques to understand the spatial and temporal structures of multiple industries across England. I will achieve this by utilising a more comprehensive property database, which consists of spatial and temporal data for all the non-domestic properties, in partnership with a specialist in the Property Technology industry.

---

## Estimating Interdependences Across Space, Time and Spatial Disparities in Diet and Health in Agent-Based Model for Policy Interventions: The Case of Food Deserts in Greater London/UK

**Elisabeth Titis · Presentation at 2:50pm**

In examining spatial inequalities, there has been little examination of the interdependencies across space and time that contribute to the creation of food deserts, FDs (1). Simulation models allow to simulate a theoretically realistic environment and generate synthetic data for evaluation of hypothetical policies in real time, provide for the feedback between incorporated variables. The research will investigate issues of FDs in the context of low-income and social exclusion, poor accessibility in terms of retail geography, and nutrition-related ill-health, utilising agent-based modeling to explore the emergent of FD environment and to build data model useful to policymakers when proposing policy interventions, such as introducing a junk food tax or placing new healthful food stores in the FDs.

## Increasing Complex Networks' Resilience by Taking Inspiration from Food Webs

**Aseel Alturki · Presentation at 3:30pm**

Human built systems are increasingly under the strain of both natural and man-made stressors. Many of the critical infrastructure systems have network dimensions. Despite their national importance, the complexity of interdependent and multi-scale networks means we do not fully understand how to invest and adapt them to different risks and uncertainties. This feasibility study on the UK's rail network explores whether we can learn from natural complex systems that have evolved under constant predation and environmental stress. In particular, we draw inspiration from food webs and their common adaptation mechanism when faced with perturbations. The study will use transport data, and explore the coupling between small-scale failures (e.g., accident leading to road closure) and large-scale effects (e.g., cascade effects lead to increased vulnerability to further failures). The scientific innovation is that we draw lessons from resilient ecosystems, to inform the design of real-time routing mechanisms and long term investment strategies.

---

## Effect of Airbnb on the Local Housing Rental Market

**Amit Chaudhary · Presentation at 3:45pm**

The Peer to Peer market i.e. Airbnb has disrupted the hotel Industry and benefited travelers. However, the effect of Airbnb on the local housing rental market remains a field of inquiry. In this paper, I will model the supply constraint of local housing rental market due to Airbnb's entry. For empirical evidence, Individual home level panel data of rental market supply and, Airbnb's listings are collected from the City of London over 7 years (2010-2017). The local market housing supply elasticity is then estimated using the regulatory cost shock on the Airbnb market. The increase in Airbnb reduces the supply of housing in the rental market thus, creating the upward pressure on home rental prices. The reduced form results obtained after assuming different wards as local market supplements the findings.

# Neo-Urbanisation Towards Sustainable Policy for Cities in Sub-Saharan Africa

**Oba Waiyaki · Presentation at 4:00pm**

Humanity's quest to create the most ideal urban spaces has been in constant development since the world's first cities emerged 8000 years ago, that sort of ingenuity in order to thrive and survive. Africa in the 21st century is no different. Urbanisation agendas are driven towards job creation with industrialisation at its core and thus, the main resolve towards development. Given the environmental destruction and inequality wrought by Western models of urbanisation, does the continent need to re-shape its urban development priorities and models in order to advance more humane, environmentally sensitive and equitable cities? And how easy will this be to achieve? Many cities in Africa are at a crossroads in history – between an established 'Western' model wrought with post-colonial path dependencies that constrain adaptation and new and innovative neo-urbanisation approaches that focus more on combining the need of the citizens and nature, and the requirements of commerce.

---

## Localising the Urban SDGs: Participatory Geospatial Data Practices and Resilience Building in Slums

**Philipp Ulbrich · Presentation at 4:15pm**

The indicator framework to measure progress towards the “urban” Sustainable Development Goals (SDGs) is currently informed by conventional, top-down data practices, such as censuses and household surveys. While this may be useful for producing country-level indicators, these methods tend to overlook intra-urban variations regarding access to services and vulnerability, which is problematic for slum areas, whose residents tend to be exposed to different hazards than their neighbours living in less deprived urban areas. There are proposals to address this gap with granular data from big data-driven means - however data scarcity, especially in slums, still limits their effectiveness.

In this context, this research provides a critical assessment regarding the potential of participatory geospatial practices to localise the urban SDG indicators. In addition to enhanced local relevance, participatory geospatial data initiatives can build resilience. Notably, this can result in increased levels of trust between previously less connected local stakeholders.

# Planning in the Smart City

**Ciaran Devlin · Presentation at 4:30pm**

Urban planning has traditionally been the method used to assess and manage how cities develop and incorporate new infrastructure and technological advances. The smart city has emerged as a concept which can profoundly enhance urban data collection and systems, providing an opportunity to improve urban planning and governance. This study aims to investigate these opportunities and implications by answering the following research questions; How can urban planning utilise the emerging technologies? How will the use of technology affect urban governance? What are the implications for the planning profession?

The relationship between urban planning and smart city technology will be reciprocal, affecting the decisions and plans made in relation to urban development and smart technology. Therefore it is vital to understand the interactions to ensure the most effective strategies for future urban planning and smart city implementation are adopted.

---

## Are Changing Management Practices in Urban Parks and Greenspaces Increasing Biodiversity, and Are There Trade-Offs with Amenity?

**Corinne Muir · Presentation at 4:45pm**

Parks are an easy target for budget cuts as there are no statutory obligations for their provision or maintenance. However, they have many important roles for urban populations including recreation, healthy exercise, air pollution reduction and flood mitigation. Various techniques are being tried in the UK's cities to reduce management costs in parks, for example new types of floral display. Could some cost saving exercises improve parks for people and wildlife, or does amenity suffer?

Methods:

- Investigating changing management practices in parks and their effects on ecosystem services. Local fieldwork results from the West Midlands will be compared to national data.
- Comparing parks mapping (from Ordnance Survey) to scenic-ness ratings (from [scenicornot.datasciencelab.co.uk](http://scenicornot.datasciencelab.co.uk)).
- Comparing changing management/maintenance data to habitat data in Coventry.
- Investigating public opinion of different types of habitat/management type using on-site surveys in the West Midlands and a national online survey.

## Commuting and Experienced Well-Being Throughout the Day

**Zakiyya Adam** · Poster presented in foyer

Using the UK Time Use Survey (UKTUS) dataset, the experienced well-being effects of commuting, in terms of enjoyment, were evaluated relative to other daily activities. Those that commuted using passive modes of transport found commuting to be the least enjoyable activity carried out in the day. Commuting was also amongst the least enjoyable activities for those using active forms of commuting, although their enjoyment was significantly higher than that of passive commuters. This study also assessed differences in the experienced well-being of other daily activities between those days on which people did, and did not, commute. Using a series of multilevel analyses, those who worked from home were found to enjoy personal care related activities more than those that commuted used passive modes. Those that solely used active modes were found to experience enjoyment of all daily activities just as much as those who did not commute to/ from work.

---

## A Systematic Literature Review in Examining the Use of Spatial Methods in Epidemiology

**Alex Caton** · Poster presented in foyer

This study aims to perform a systematic literature review of the current research on the spatial methods and applications for mitigating the spread of infectious diseases. Furthermore, using both semantic and spatiotemporal information from location-based social networks such as Twitter can provide new and innovative ways to study epidemics. From the literature review performed in this study, it is apparent that few existing literature reviews tackle the study of spatial methods and disease tracking, and that future research should be dedicated to rectifying this. This study investigates the existing literature on how diseases are mitigated, finding correlations between geographic based data and outbreaks. Working with existing research, this project finds current trends in the methodologies of diseases being researched, informing researchers of gaps and potential avenues for future study.

---

## Embedding Road Networks and Travel Time into Distance Metrics for Urban Modelling

**Henry Crosby · Poster presented in foyer**

Urban environments are restricted by various physical, regulatory and customary barriers such as buildings, one-way systems and pedestrian crossings. These constraints create challenges for modelling processes over an urban space because the typical distance metrics used, e.g. Euclidean, can no longer capture and describe proximity in these settings. In this work we propose and develop a low-dimensional Euclidean distance metric approximating restricted road distance and time travel using an Isomap approach. This new approach is compared using Kriging to predict house prices on 3,669 properties in Coventry, UK.

---

## How Can Geo-Resources Build Sustainable and Resilient Cities?

**Katherine Harris · Poster presented in foyer**

The general trend in cities around the world is that urbanisation is on the rise. As a result of this, cities are becoming increasingly reliant on geo-resources to support their everyday services and expansion. Geo-resources are naturally occurring assets of the earth that can be harnessed to create something functional for our consumption, including; geo-materials, sub-surface space, groundwater and geothermal energy. Above ground space, ground properties (permeability, stability, etc) and topography can also be considered valuable geo-resources that have sometimes been overlooked in city development. This study is investigating how the use of geo-resources can enhance the resilience and sustainability of urban design in cities. Current perceptions and practises of utilising geo-resources are being reviewed, as well as an examination of current resilience and sustainability assessment methods proposed in different cities today. Ultimately, a geo-resource site specific indexing tool is envisioned to propose a way forward to resilient urban design.

## A Spatial Analysis of Proximity to Greenspace and Mental Wellbeing in London

**Vikki Houlden** · Poster presented in foyer

Exposure to greenspace has been shown to be associated with improved health, productivity and societal prosperity. While the government recommends that greenspace should be provided within 300m of individuals, to ensure that urban residents can “access the natural environment”, this is the first study to examine the potential mental wellbeing benefits of this guideline. Using survey data on mental wellbeing and socio-demographic, alongside GIS greenspace maps, Geographically Weighted Regression is used to investigate how the association between proximate greenspace and 3 mental wellbeing outcomes varies across London. Greenspace within 300m is found to be positively and significantly associated with life satisfaction, happiness and sense of worth, with a weaker association for greenspace within 500m, becoming negligible by 1000m distance. The strongest associations are observed towards the north and west of London; this variation may be due, in part, to the types and quality of these greenspaces.

---

## Planning for Urban Development and Industrial Growth in the Circumpolar North

**Melissa Kenny** · Poster presented in foyer

Development and urbanisation the Arctic is likely to become a central feature of the region. As the natural resource extraction industry is set to grow, alongside the impacts of climate change, the landscape and the livelihoods of those who live there may be subject to extreme changes in the near future. Urban areas will become the focal points for development in the region. Increased levels of accessibility due to climate change, as well as new opportunities for resource extraction, trade routes, rural-urban migration and strategic interest mean that urban development the Arctic is becoming increasingly significant, economically, politically and socially. These opportunities come with threats and consequences; remote regions are particularly environmentally vulnerable to human influence. Furthermore, cost, human adaptability and resilience all present challenges. These complex challenges need to be anticipated and consequently well planned for to ensure that the urbanisation of the Arctic is sustainable, manageable and inclusive.

---

# Resilience of Critical Infrastructure Systems

**Maria Liatsikou · Poster presented in foyer**

Critical National Infrastructure is of high importance for a country and is characterized by high complexity: multi-dimensional systems, consisting of subsystems interacting with each other, forming complex networks and governed by non-linear dynamics, subject to failures and environmental threats, are highly important, since they determine their safety and efficiency. A system's resilience is its ability to adjust and retain its functionality when failures and environmental changes and disasters happen. The scope of this research is to model, and predict the dynamics of multi-dimensional systems and to apply them in the urban infrastructure domain, such as in water distribution networks.

---

## Understanding Geohazards for Enhancing Urban Resilience: A Study on Thessaloniki, Greece

**Vangelis Pitidis · Poster presented in foyer**

Urban Resilience has recently emerged as both a conceptual approach and directive agenda in the attempt to enhance the capabilities of cities to withstand and manage their environmental, economic or social pressures more effectively. The malleable definition of resilience has rendered its operationalisation an intriguing task for contemporary cities trying to address their organisational problems and confront uncertainty in a holistic manner. Concurrently the role of civic society in this process has moved to the centre of attention, while local geomorphological context along with urban planning and design are often playing a secondary role. Our goal is to highlight links between urban landscape and physical planning for confronting earthquakes and surface flooding in Thessaloniki, Greece. Building on the Resilient Thessaloniki Strategy, we demonstrate how considerations of local topography and geomorphology, surface and subsurface geological conditions and condition of the building stock can strengthen the proposed strategy.

# Understanding the Engineering Characteristics of Embankments

**Bradley Sheridan · Poster presented in foyer**

Approximately 52,160km of various embankments have been constructed across England and Wales with an annual budget exceeding £450 million. It is important to understand what influences the behaviour and stability behind these geo-structures. By understanding these influences it will allow for future behaviour forecasting of geo-structures and will allow for a greater development of maintenance and management that will result increased resilience of these infrastructures. This poster describes an insight into what the research project will be focusing on, which is to understand the engineering characteristics of embankments, and to show the importance of these geo-structures.

---

# Using the Weather Research and Forecasting Model to Validate and Improve High Resolution Weather Forecasts

**Timothy Sit · Poster presented in foyer**

Simulations of city-sized mesoscale meteorology at the sub-kilometre scale rarely succeeds due to the incompatibility of the numerics used at global scales with the requirements of doing simulations with local resolution below 1 kilometre. Numerical schemes used in the Weather Research and Forecasting (WRF) model are capable of simulating conditions as fine as 100 metres and successful in generating sufficient resolved scale turbulence, something which current Met Office Unified Model (MetUM) calculations fail at when resolving at these sub-kilometre scales.

My plans will be to demonstrate that WRF can indeed predict high rainfall events at these finer resolutions, then compare with MetUM simulations of the same data. This will eventually lead into providing a set of reliable high-resolution reference simulations over a range of urban environments so that city councils can better prepare for hazardous weather events such as flooding.

---

# Exploring city Placemaking capacity and perceived Quality of Life through local policies, initiatives and social media engagement'

**Isabella Slattery · Poster presented in foyer**

City Placemaking and Quality of Life (QoL) are not terms that are often used together in established urban studies aimed at improving life in the city. Separately however, they have long been key drivers for governments, city administrations and private industry. What is less understood is what City Placemaking can do to impact perceived QoL and how. Assessing this impact (and in particular utilising social media capacity, searches and data to aid in this) is a relatively new field of research and the focus of this work.

# Security

Track 6  
**OC1.08 · 2:00pm**

---

Session Chair  
Liam Steadman



# Internet of Things Forensics

**Sara Alhajali · Presentation at 2:05pm**

Researching internet of things (smart homes) forensics, on determining a method to generate a sound digital evidence, by logging the IoT activities in a smart home network using a modified version of Chandy & Lamport algorithm to record a snapshot of the network that is sound and correct to satisfy the condition to stand in court, while also defining the concept of chain of custody in IoT forensics investigation and how to apply it.

---

# Sniffing Passwords and Verifying Users on Smartwatch via Deep Sequence Learning

**Bowen Du · Presentation at 2:20pm**

Demand for smartwatches has taken off in recent years with new models which can run independently from smartphones and provide more useful features, becoming first-class mobile platforms. One can access online banking or even make payments on a smartwatch without a paired phone. This makes smartwatches more attractive and vulnerable to malicious attacks, which to date have been largely overlooked. This talk will demonstrate Snoopy, a password extraction and inference system which is able to accurately infer passwords entered on Android/Apple watches just by eavesdropping on motion sensors and a novel authentication approach for smartwatches, which adds another layer of security on top of the traditional passwords by considering the unique motion signatures when different users input passwords on their watches.

## Security-Aware Virtual Machine Placement Model in Cloud Computing

**Mansour Fahad Abdulaziz Aldawood · Presentation at 2:35pm**

Cloud computing technology offers many advantages for cloud users such as the sharing of the resources among physical components and the ability for virtual machines (VM) to migrate between physical nodes in order to increase the availability and efficiency of them. However, this feature opens the door for serious security threats that affect the confidentiality of the data processed in these systems. Recent studies showed that the attacker able to extract private information by utilizing the misconfiguration of the physical resources through a side channel, and the weakness of placement algorithm of the VMs. In this study, we focused on tackling this issue by modelling a security-aware virtual machine placement system that aimed to mitigate the risk of side-channel attack and reduce the probability of a VM placed with an attacker VM in cloud computing systems.

---

## Automating Cloud SLA Enforcement using Fair Exchange Protocol

**Farrukh Qazi · Presentation at 2:50pm**

In the modern e-commerce realm, a “business fairness” is extremely paramount. Evaluating the extent of fairness is hard for non-tangible eg. quality of service (QoS). The Cloud Service Providers (CSP) offer an on-demand and elastic services to Cloud Service Subscribers (CSS). Service Level Agreement (SLA), a legal instrument, yields the service definitions against some monetary commitments. A dishonest party such as the CSP could cause series of SLA violations. The entire burden of proof gets on the CSS, if the QoS is compromised. To cover this gap, we suggest that Fair Exchange Protocol (FEP) is an excellent way to manage a dispute resolution scenario. It also addresses to mitigate exchange unfairness using automated service anomaly detection. Our architecture based upon FEP, which constitutes the least involvement of a Trusted Third Party (TTP), who intelligently takes controls to manage business transactions in a controlled fashioned with minimal monitoring overheads.

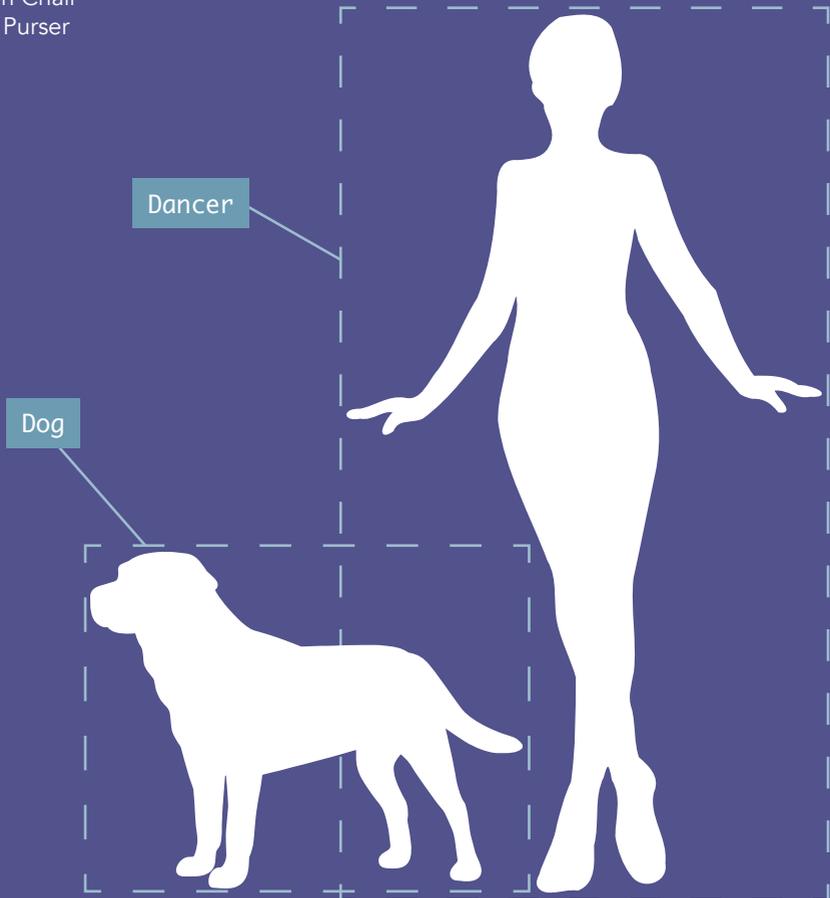
# Image, Vision & Computational Biology

Track 7

**OC1.08** · 3:25pm

---

Session Chair  
David Purser



# Person Re-ID using Deep Foreground Appearance Modelling

**Greg Watson · Presentation at 3:30pm**

Person Re-Identification is the process of matching people from images taken of them at different times and under different circumstances. Most methods extract features from the entire image, giving no consideration to the spatial context present within it. We propose using a convolutional neural network based on ResNet-50 to predict the foreground of an image. We can then use the LOMO and Salient Colour Name feature descriptors to extract features primarily from these areas. We also use the XQDA distance metric learning technique to calculate optimally weighted distances between the features. We evaluate on the VIPeR, QMUL GRID and CUHK03 data sets, and compare our results against a linear foreground estimation approach, and show competitive or better matching performance.

---

# Two-Stream Capsule Network for Action Recognition

**Abdullah Algamdi · Presentation at 3:45pm**

Convolutional Neural Network (CNN) is considered the state-of-the-art in action recognition. Recent CNN models have shown substantial results on both images and video datasets. However, CNN still needs large data to generalize. Furthermore, CNN loses spatial relation due to extensive use of pooling layers. In 2017, Sabour et al. proposed a new architecture called 'Capsule Network'. Their network made improvement in spatial correlation and perform better in affine transformation. CapsNet was able to reach the state-of-the-art in MNIST dataset. Action recognition is one of the challenging machine learning tasks in computer vision. Many methodologies have been introduced to classify video. One of which is the two-stream CNN network architecture which shows competitive results on different datasets such as UCF-101 and HMDB51. Our work focus on using the Capsule Network to learn both spatial and temporal dimensions from a video. We used the two-stream architecture to classify actions in the video.

# Deep Autoencoder Based Registration of Histology Images

**Ruqayya Awan · Presentation at 4:00pm**

Currently registration of cross-slide histology images is performed manually by pathologists. This is time-consuming due to a large number of sections per tissue block. There is a need to automate the registration by developing methods that do not require corresponding points as ground truth information. Keeping this in mind, we employed autoencoders (AEs) to learn the latent representation features that could be used for registration of histology images, followed by a feature matching method. Our proposed method estimates the transformation which maximises the mutual information (MI) between the learnt AE features of the two images. For comparison purpose, we experimented with image intensities using the same MI method. Our results demonstrate that the use of AE features reduces the error rate and computational time by a significant margin.

---

# Cell Detection using Mixture Density Network

**Navid Alemi Koohbanani · Presentation at 4:15pm**

Nuclei detection is an important task in histology domain as it is the main step toward further analysis such as cell counting, cell segmentation, study of cell connections, etc. This is a challenging task due to complex texture of histology image, variation in shape, touching cells. To tackle these hurdles, many approaches have been proposed in the literature where deep learning methods stand on top in terms of performance. Hence, in this paper, we propose a novel framework for nuclei detection based on Mixture Density Network and Deep features. Candidate images and their corresponding features are estimated through a deep network and then Mixture Density Network is responsible for generating locations of the nucleus through a probabilistic framework. The experimental results show the state-of-the-art performance on complex colorectal denocarcinoma dataset.

## Optimising Vision Algorithms with Visual Saliency

**Jamie Bayne · Presentation at 4:30pm**

Over the past few years, the computer vision requirements of emerging autonomous vehicle technology have been the subject of intense academic and industrial research. At the same time, breakthroughs machine learning have propelled many fields forward. One of these is the problem of visual saliency, which assigns importance to visual features according to expected human focus. Recent work has provided the DR(EYE)VE dataset for visual saliency in a driving context, and this work aims to use this as a heuristic for multi-resolution optimisation of several algorithms fundamental to autonomous vehicle systems, such as stereo disparity, optical flow, and object recognition. This presentation will give an overview of visual saliency, multi-resolution methods, and how they can be combined.

---

## The Cancer Genome Atlas - Tumour Profiles from Whole-Slide Images

**Mary Shapcott · Presentation at 4:45pm**

We describe how convolutional neural networks (CNNs) can be trained to detect and classify the cells within whole-slide images of colorectal carcinomas. Using the CNNs we have processed 145 colorectal adenocarcinoma pathology images from The Cancer Genome Atlas (TCGA), a publicly available repository of cancer data. The resulting cell location patterns are summarised as morphological features, such as abundances of epithelial, inflammatory and necrotic cells, and fibroblasts. The TCGA colorectal cancer data includes clinical information concerning age, gender, diagnosis, cancer grade and stage, as well as gene expression data. We have examined relationships between morphological profiles and demographic, clinical and molecular features, and present the results of our analysis.

## AI based Identification and Quantification of Growth Patterns in Lung Adenocarcinoma

**Najah Alsubaie · Poster presented in foyer**

According to the latest 2015 WHO of lung tumour, lung adenocarcinoma has five growth patterns: Acinar, papillary, micro-papillary, solid and lepidic. Adenocarcinoma tumour could contain one or more of these patterns in the same biopsy. Studies show that the growth patterns are correlated with patient survival, with solid and micro-papillary having the worst prognosis, while cases with predominant lepidic pattern have good survival.

Automatic identification of growth patterns is, therefore, critical and could affect patients' lives. It will not only reduce the pathologists effort and variability, it could also provide a second opinion to support pathologists decision.

We proposed a deep learning-based framework that mimics pathologists examination. Our method identifies all possible patterns by examining the tissue at several resolutions. We then fuse classification results delivered from these resolutions to construct the final classification over the whole slide image. We measure the pattern classification

---

## MILD-Net: Minimal Information Loss Dilated Network for Gland Instance Segmentation in Colon Histology Images

**Simon Graham · Poster presented in foyer**

The analysis of glandular morphology within colon histopathology images is crucial in determining the stage of colon cancer. Despite the importance of this task, manual segmentation is laborious and can suffer from subjectivity among pathologists. The rise of computational pathology has led to the development of automated methods for gland segmentation that aim to overcome the challenges of manual segmentation. However, this task is non-trivial due to the large variability in glandular appearance and the difficulty in differentiating between certain histological structures. We propose a fully convolutional neural network that counters the loss of information caused by max-pooling by re-introducing the original image at multiple points within the network. We also use atrous spatial pyramid pooling with varying dilation rates for resolution maintenance and multi-level aggregation. The proposed network achieves state-of-the-art performance on the GlaS challenge dataset, as part of MICCAI 2015, and on a second colorectal adenocarcinoma dataset.

## Detection of $\beta$ Cells in Immunohistochemical (IHC) Stained Mouse Pancreatic Tissue

**Talha Qaiser · Poster presented in foyer**

Beta ( $\beta$ ) cells in the pancreas are primarily responsible for releasing and storing of insulin, that eventually affect the concentration of blood glucose. Accurate detection and density estimation of  $\beta$  cells in mouse pregnancy may help in our understanding of the formation and growth of beta cells in pancreatic islets. We approach this problem by proposing a fast deep convolutional regression model that predicts spatial coordinates of a bounding box containing nuclei. The proposed framework is capable of encoding the spatial context, as opposed to sliding windowing approaches, which are computationally complex and their analysis is limited to a small subsection of an input image. Our aim is to achieve almost real-time detection of beta cells. Our method is closely related to YOLO with residual-blocks instead of only convolutional layers to alleviate the vanishing gradient problem. The training dataset contains more than 9,000 hand-marked nuclei from pancreatic islets.

---

## Dual Viewpoint Passenger State Classification Using 3D CNNs

**Ian Tu · Poster presented in foyer**

The rise of intelligent vehicles systems will lead to more human machine interactions so there is a need to create a bridge between the system and the actions and behaviours of the people inside the vehicle. We propose an in-vehicle camera setup to monitor the actions and behaviour of vehicle passengers and a deep learning architecture which can utilise the video data to classify their actions. The method incorporates the two different views as input to a 3D CNN and uses transfer learning from other action recognition data. The performance of this method is evaluated using videos of people performing a range of common in-vehicle actions. We show that the combination of transfer learning and using dual viewpoints in a 3D CNN offers an increase in classification accuracy of actions with distinct poses whilst it does not apply as well for classifying actions with small movements.

# Individualized Prediction of Sensation Seeking Personality based on Functional Connectivity

**Zhuo Wan · Poster presented in foyer**

Sensation seeking is a personality trait which is described as high impulsiveness to search for various, intense and complex experiences and prepared to take risks for the sake of these experiences. Current understanding of the neural mechanism underlying the components of personality is lacking, and whether and how neuroanatomical features can be used to predict this sensation seeking trait remain largely unexplored. In the presented study, I analyzed a large sample from the enhanced Nathan Kline Institute-Rockland Sample (NKI-RS) dataset and successfully built multivariate predictive models for the sensation-seeking using the functional connectivity (FC) features between brain regions. The results showed that the model effectively captured individual differences in this personality and was able to significantly predict for unseen individuals. The k-fold cross-validation was applied to estimate the performance of the prediction model.

# High Performance Computing

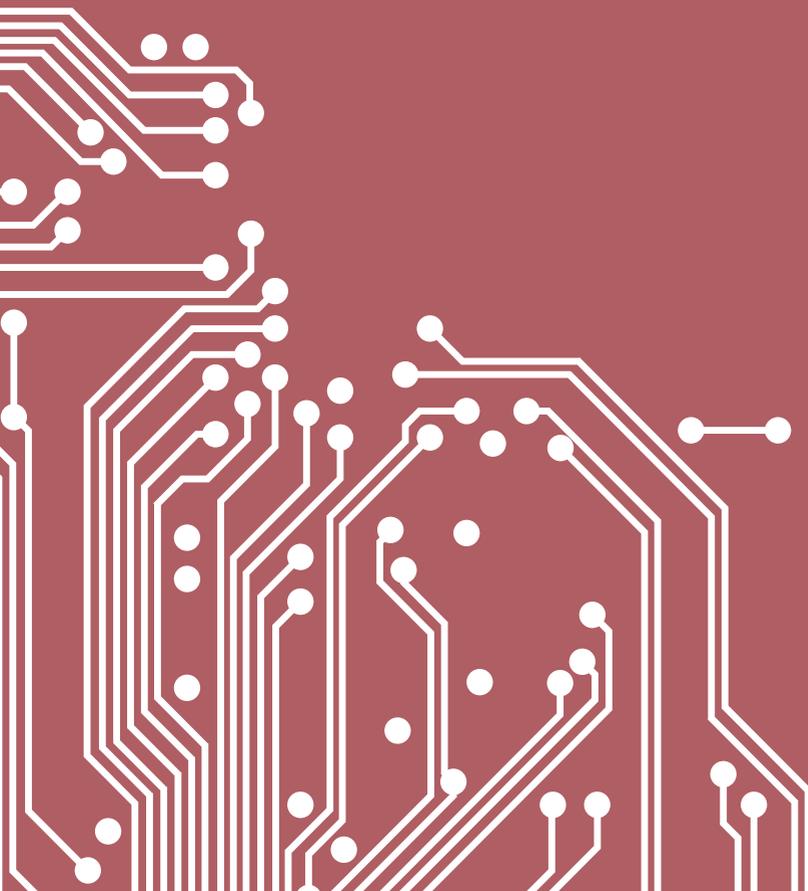
Track 8

**OC1.09 · 3:25pm**

---

Session Chair

Matthew Bradbury



# Performance of a Second Order Electrostatic Particle-in-Cell Algorithm on Modern Many-Core Architectures

**Dom Brown · Presentation at 3:30pm**

In this paper we present the outline of a novel electrostatic, second order Particle-in-Cell (PIC) algorithm, that makes use of ‘ghost particles’ located around true particle positions in order to represent a charge distribution. We implement our algorithm within EMPIRE-PIC, a PIC code developed at Sandia National Laboratories. We test the performance of our algorithm on a variety of many-core architectures including NVIDIA GPUs, CPUs, and Intel’s Knights Landing. Our preliminary results show the viability of second order methods for PIC applications on these architectures when compared to previous generations of many-core hardware. Although these initial results show a large increase in runtime over first order methods, we hope to be able to show improved scaling behaviour and increased simulation accuracy in the future.

---

# Developing a Geometric Multigrid, Unstructured Grid Mini-Application

**Andrew Owenson · Presentation at 3:45pm**

Achieving high-performance of large scientific codes is a difficult task. This has led to the development of numerous mini-applications that are more tractable to analyse, while retaining performance characteristics of their full-sized counterparts. These “mini-apps” also enable faster hardware evaluation, and for sensitive codes allow evaluation of systems outside of access approval processes.

In this paper we develop a mini-application of a geometric multigrid, unstructured grid Computational Fluid Dynamics (CFD) code, designed to exhibit similar performance characteristics without sharing code. We detail our experiences developing this application, using guidelines detailed in existing research, and contribute further additions to these to aid future mini-application developers. Our application is validated against the inviscid flux routine of HYDRA, a CFD code developed by Rolls-Royce, which confirms that the parent kernel and mini-application share fundamental causes of parallel inefficiency.

## Understanding Communication Patterns in HPCG

**Dean Chester · Presentation at 4:00pm**

Conjugate Gradient (CG) algorithms form a large part of many HPC applications, examples include bioinformatics and weather applications. These algorithms allow numerical solutions to complex linear systems. Understanding how distributed implementations of these algorithms use a network interconnect will allow system designers to gain a deeper insight into their exacting requirements for existing and future applications.

This talk documents my initial investigation into the communication patterns present in the High Performance Conjugate Gradient (HPCG) benchmark.

---

## Data Structures and Parallelisation for High Performance Computing

**Richard Kirk · Presentation at 4:15pm**

In order to gain the performance out of large production codes, having the correct underlying data structure is crucial. It ensures temporal and spacial locality of the data can be exploited, thus gaining performance. However, the optimal data structure that should be used differs for each algorithm. As such, in a production code which contains a large array of algorithms, it is very difficult to produce a data structure that works optimally for all algorithm. The hardware which the program is run on will also affect what the optimal data structure should be. This is due to how data parallelism is achieved. In this presentation, an API will be presented which allows for different data structures can be tested quickly, with the aim of seeing which data structures are optimal.

# Computation Offloading in Dynamic Environment of Mobile Cloud Computing

**Mohammed Maray · Presentation at 4:30pm**

In our study, we are trying to find the beneficial solution in the worst case in the dynamic environment of computation offloading mechanism.

We used an internet edge “Cloudlets” to be deployed in the region of users movements and all cloudlets will be connected to the access points in different places. So we consider the user mobility patterns as an input Trajectory of cloudlets in the way of user direction. We represent tasks assignment scheduling as Directed Acyclic Graph with release time and finish time for each task.

Our objective is to minimize the execution time of dependency tasks of ecosystems during mobility among several cloudlets and make the computation offloading from ecosystems to cloudlets more efficient.

---

# Implicit Mapping of Pointers Inside C++ Lambda Closure Objects in OpenMP Target Offload Regions

**David Truby · Presentation at 4:45pm**

With the diversification of HPC architectures beyond traditional CPU-based clusters, a number of new frameworks for performance portability across architectures have arisen. One way of implementing such frameworks is to use C++ templates and lambda expressions to design loop-like functions. However, lower level programming APIs that these implementations must use are often designed with C in mind and do not specify how they interact with C++ features such as lambda expressions.

This talk proposes a change to the behavior of the OpenMP specification with respect to lambda expressions such that when functions generated by lambda expressions are called inside GPU regions, any pointers used in the lambda expression correctly refer to device pointers. This change has been implemented in a branch of the Clang C++ compiler and demonstrated with two representative codes.

# Three Communication patterns for Parallelizing All-Pairs Shortest-Path Algorithm

**Mohammed Alghamdi · Poster presented in foyer**

The All-Pairs Shortest Paths (APSP) is a fundamental graph problem aiming to find the shortest path between any two nodes in a graph. In this paper, we present a new method for solving the APSP problem for big graphs by partitioning and processing it in parallel. In this work Breath First Search Algorithm is used to prepare the graph for partition. After we partition the graph into a smaller sub-graph, we use the Floyd-Warshall algorithm or Dijkstra's algorithm to find the APSP in each sub-graph. Then we use one of the two ways we implemented to combine the output of each part of the partitioned graph and find the APSP for the whole graph. The method works in a high level of parallelism. We presented three different communication patterns can be used when we parallelize the method.

“Science means constantly walking a tightrope between blind faith and curiosity; between expertise and creativity; between bias and openness; between experience and epiphany; between ambition and passion; and between arrogance and conviction - in short, between an old today and a new tomorrow.”

**Heinrich Rohrer**

---

“The hardest thing is to go to sleep at night, when there are so many urgent things needing to be done. A huge gap exists between what we know is possible... and what we have so far been able to finish.”

**Donald Knuth**

Warwick Postgraduate Colloquium in Computer Science 2018

Sponsored by the Department of Computer Science and the Research Student Skills Programme

[warwick.ac.uk/wpccs18](http://warwick.ac.uk/wpccs18)