# Andrew Wiles

他证明了费马大定理，解决了著名数学家费马提出的悬而未决了几百年的重要问题

Modelling Framework for Radio Frequency Spatial Measurement（2006年）

The main crux of this thesis was to produce a model that was capable of simulating the theoretical performance of different configurations for a spatial measurement system using radio frequency technology. It has been important to study new modalities of spatial measurement since spatial measurement systems are an enabling technology that have allowed for the creation of better medical procedures and techniques, provided valuable data for motion capture in animation and biomechanics, and have improved the quality of manufacturing processes in many industries. However, there has been room for improvement in the functional design and accuracy of spatial measurement systems that will enhance current applications and further develop new applications in medicine, research and industry.

In this thesis, a modelling framework for the investigation of spatial measurement based on radio frequency signals was developed. The simulation framework was designed for the purpose of investigating different position determination algorithms and sensor geomatries. A finite element model using the FEMLAB partial differential equation modelling tool was created for a timedomain model of electromagnetic wave propagation in order to simulate the radio frequency signals travelling from a transmitting source antenna to a set of receiving antenna sensors. Electronic line signals were obtained using a simple receiving infinitesimal dipole model and input into a time difference of arrival localization algorithm. The finite element model results were validated against a set of analytical solutions for the free space case. The accuracy of the localization algorithm was measured against a set of possible applications for a potential radio frequency spatial measurement system design.

It was concluded that the simulation framework was successful should one significant deficiency be corrected in future research endeavours. A phase error was observed in the signals extracted at the receiving antenna locations. This phase error, which can be up to 40°, was attributed to the zeroth order finite elements implemented in the finite element model. This phase error can be corrected in the future if higher order vector elements are introduced into future versions of FEMLAB or via the development of custom finite element analysis software but were not implemented in this thesis due to time constraints. Other improvements were also suggested for future work.

避免使用第一人称

Navigation Accuracy of Image-Guided Interventional Systems（2009年）

Surgical procedures, when guided by preoperative images and tracked tools, can be compromised by the errors involved in registering the images and instruments to the patient. The process of minimizing these errors is analogous to optimizing quality on a manufacturing production line. For example, when building machinery or components to be assembled, the designer must specify a set of tolerances which describe the range in which a specific part can be built and still be properly assembled such that it will perform within the desired specifications. If the tools used cannot build the parts within the specified tolerances then the quality of the end product will suffer. Quality engineers regularly inspect the components being assembled to ensure that they meet the required specifications, so that costs are minimized and quality is maximized.

A similar framework can be applied to image-guided interventions (IGI), which comprise both surgical (IGS) and therapeutic (IGT) procedures, where the surgeon is generally trying either to (i) resect a section of tissue, (ii) implant a device, (hi) apply a dose of therapeutic energy, or (iv) apply some combination of the three. Before surgery, the surgeon uses the surgical workstation to develop a preoperative plan. On this plan, the path to the target is identified along with critical areas to avoid such as nerves, blood vessels, etc. A set of boundaries can be drawn to identify the critical areas along the path which essentially act as tolerances, where if the surgical tools move outside these boundaries additional complications may occur. Since IGI systems are based upon tracking technologies and image registration, the tool locations can be monitored to ensure that they fall within the tolerance boundaries.

However, due to the inherent uncertainties in the tracking technology and registration methods, the locations of the surgical targets, planned surgical path and tolerance boundaries are not known exactly, but rather take on a statistical distribution. At specific points of interest (targets), the uncertainties can be described by statistics of the target registration error (TRE), which is a function of the registration algorithms, the specific geometry and the uncertainty of identifying the landmarks, i.e., fiducial localizer error (FLE), used to infer the location of the point of interest. If the TRE statistics can be computed for a locus of points that define the targets, paths and boundaries, then the uncertainties of those features can be shown graphically to the surgeon via a comprehensive surgical plan, so that the surgeon can make strategic decisions and perioperatively to ensure success.

In this thesis, a complete framework is developed for estimating the overall error for IGI systems that use point-based registration, optical and/or magnetic tracking. In particular, TRE statistical models are developed for point-based registration where the FLE is either (i) anisotropic and homogeneous or (ii) anisotropic and nonhomogeneous. An algorithm that estimates the FLE and TRE in realtime is presented, which is useful for optical tracking systems that use point-based registration to determine the tool pose. Furthermore, statistical models are developed for tracking systems based on magnetic fields where there is no explicit registration but the FLE uncertainty models are available for sensors that report tracking information with five and six degrees of freedom. In this case, a new term, target tracking error (TTE), is introduced which is very similar to TRE but carries a distinct name to emphasize the lack of a registration taking place. Finally, the application of these models is examined in context of an intracardiac surgical guidance system where tracked 2D ultrasound is a feature.

The work presented here, in concert with related work in the. literature, provides a complete framework to estimate the uncertainty associated with the patient to image registration and tracking systems. The error estimates allow for the development of comprehensive surgical plans, improves patient registration algorithms, and provides the surgeon with a powerful tool for quality control during IGI procedures.

避免使用第一人称

# Terence Tao

数学家，在多个领域做出了突破性的工作，特别是在分析学和偏微分方程领域。

Nonlinear dispersive equations: local and global analysis（2006年）

Among nonlinear PDEs, dispersive and wave equations form an important class of equations. These include the nonlinear Schrodinger equation, the nonlinear wave equation, the Korteweg de Vries equation, and the wave maps equation. This book is an introduction to the methods and results used in the modern analysis (both locally and globally in time) of the Cauchy problem for such equations. Starting only with a basic knowledge of graduate real analysis and Fourier analysis, the textfirst presents basic nonlinear tools such as the bootstrap method and perturbation theory in the simpler context of nonlinear ODE, then introduces the harmonic analysis and geometric tools used to control linear dispersive PDE. These methods are then combined to study four model nonlinear dispersiveequations. Through extensive exercises, diagrams, and informal discussion, the book gives a rigorous theoretical treatment of the material, the real-world intuition and heuristics that underlie the subject, as well as mentioning connections with other areas of PDE, harmonic analysis, and dynamical systems. As the subject is vast, the book does not attempt to give a comprehensive survey of the field, but instead concentrates on a representative sample of results for a selected set of equations, ranging from the fundamental local and global existence theorems to very recent results, particularly focusing on the recent progress in understanding the evolution of energy-critical dispersive equations from large data. The book is suitable for a graduate course on nonlinear PDE.

避免使用第一人称

Multilinear weighted convolution of L2 functions, and applications to nonlinear dispersive equations（2001年）

The X s, b spaces, as used by Beals, Bourgain, Kenig-Ponce-Vega, Klainerman-Machedon and others, are fundamental tools to study the low-regularity behavior of nonlinear dispersive equations. It is of particular interest to obtain bilinear or multilinear estimates involving these spaces. By Plancherel's theorem and duality, these estimates reduce to estimating a weighted convolution integral in terms of the L 2 norms of the component functions. In this paper we systematically study weighted convolution estimates on L 2. As a consequence we obtain sharp bilinear estimates for the KdV, wave, and Schrödinger X s, b spaces.

避免使用第一人称

# Jon Kleinberg

算法和社交网络分析领域的权威人物，他的工作对于理解网络结构、信息传播和社交影响力等方面具有重要意义。

Authoritative sources in a hyperlinked environment（1999年）

The network structure of a hyperlinked environment can be a rich source of information about the content of the environment, provided we have effective means for understanding it. We develop a set of algorithmic tools for extracting information from the link structures of such environments, and report on experiments that demonstrate their effectiveness in a variety of context on the World Wide Web. The central issue we address within our framework is the distillation of broad search topics, through the discovery of “authorative” information sources on such topics. We propose and test an algorithmic formulation of the notion of authority, based on the relationship between a set of relevant authoritative pages and the set of “hub pages” that join them together in the link structure. Our formulation has connections to the eigenvectors of certain matrices associated with the link graph; these connections in turn motivate additional heuristrics for link-based analysis.

使用了we，但是只有一个作者

The small-world phenomenon: An algorithmic perspective（2000年）

Long a matter of folklore, the "small-world phenomenon" -- the principle that we are all linked by short chains of acquaintances -- was inaugurated as an area of experimental study in the social sciences through the pioneering work of Stanley Milgram in the 1960's. This work was among the first to make the phenomenon quantitative, allowing people to speak of the "six degrees of separation" between any two people in the United States. Since then, a number of network models have been pro- posed as frameworks in which to study the problem analytically. One of the most refined of these models was formulated in recent work of Watts and Strogatz; their framework provided compelling evidence that the small-world phenomenon is pervasive in a range of networks arising in nature and technology, and a fundamental ingredient in the evolution of the World Wide Web.

But existing models are insufficient to explain the striking algorithmic component of Milgram's original findings: that individuals using local information are collectively very effective at actually constructing short paths between two points in a social network. Although recently proposed network models are rich in short paths, we prove that no decentralized algorithm, operating with local information only, can construct short paths in these networks with non-negligible probability. We then de- fine an infinite family of network models that naturally generalizes the Watts-Strogatz model, and show that for one of these models, there is a decentralized algorithm capable of finding short paths with high probability. More generally, we provide a strong characterization of this family of network models, showing that there is in fact a unique model within the family for which decentralized algorithms are effective.

使用了we，但是只有一个作者

Bursty and hierarchical structure in streams（2002年）

A fundamental problem in text data mining is to extract meaningful structure from document streams that arrive continuously over time. E-mail and news articles are two natural examples of such streams, each characterized by topics that appear, grow in intensity for a period of time, and then fade away. The published literature in a particular research field can be seen to exhibit similar phenomena over a much longer time scale. Underlying much of the text mining work in this area is the following intuitive premise --- that the appearance of a topic in a document stream is signaled by a "burst of activity," with certain features rising sharply in frequency as the topic emerges. The goal of the present work is to develop a formal approach for modeling such "bursts," in such a way that they can be robustly and efficiently identified, and can provide an organizational framework for analyzing the underlying content. The approach is based on modeling the stream using an infinite-state automaton, in which bursts appear naturally as state transitions; in some ways, it can be viewed as drawing an analogy with models from queueing theory for bursty network traffic. The resulting algorithms are highly efficient, and yield a nested representation of the set of bursts that imposes a hierarchical structure on the overall stream. Experiments with e-mail and research paper archives suggest that the resulting structures have a natural meaning in terms of the content that gave rise to them.

避免使用第一人称。

# Haoyuan Li

数据编排公司Alluxio创始人，伯克利大学博士。

题目：Alluxio: A Virtual Distributed File System（2018年）

The world is entering the data revolution era. Along with the latest advancements of the Inter- net, Artificial Intelligence (AI), mobile devices, autonomous driving, and Internet of Things (IoT), the amount of data we are generating, collecting, storing, managing, and analyzing is growing ex- ponentially. To store and process these data has exposed tremendous challenges and opportunities.

Over the past two decades, we have seen significant innovation in the data stack. For exam- ple, in the computation layer, the ecosystem started from the MapReduce framework, and grew to many different general and specialized systems such as Apache Spark for general data processing, Apache Storm, Apache Samza for stream processing, Apache Mahout for machine learning, Ten- sorflow, Caffe for deep learning, Presto, Apache Drill for SQL workloads. There are more than a hundred popular frameworks for various workloads and the number is growing. Similarly, the storage layer of the ecosystem grew from the Apache Hadoop Distributed File System (HDFS) to a variety of choices as well, such as file systems, object stores, blob stores, key-value systems, and NoSQL databases to realize different tradeoffs in cost, speed and semantics.

This increasing complexity in the stack creates challenges in multi-fold. Data is siloed in various storage systems, making it difficult for users and applications to find and access the data efficiently. For example, for system developers, it requires more work to integrate a new compute or storage component as a building block to work with the existing ecosystem. For data application developers, understanding and managing the correct way to access different data stores becomes more complex. For end users, accessing data from various and often remote data stores often results in performance penalty and semantics mismatch. For system admins, adding, removing, or upgrading an existing compute or data store or migrating data from one store to another can be arduous if the physical storage has been deeply coupled with all applications.

To address these challenges, this dissertation proposes an architecture to have a Virtual Distributed File System (VDFS) as a new layer between the compute layer and the storage layer. Adding VDFS into the stack brings many benefits. Specifically, VDFS enables global data accessibility for different compute frameworks, efficient inmemory data sharing and management across applications and data stores, high I/O performance and efficient use of network bandwidth, and the flexible choice of compute and storage. Meanwhile, as the layer to access data and collect data metrics and usage patterns, it also provides users insight into their data and can also be used to optimize the data access based on workloads.

We achieve these goals through an implementation of VDFS called Alluxio (formerly Tachyon). Alluxio presents a set of disparate data stores as a single file system, greatly reducing the complexity of storage APIs, and semantics exposed to applications. Alluxio is designed with a memory centric architecture, enabling applications to leverage memory speed I/O by simply using Alluxio. Alluxio has been deployed at hundreds of leading companies in production, serving critical workloads. Its open source community has attracted more than 800 contributors worldwide from over 200 companies.

In this dissertation, we also investigate lineage as an important technique in the VDFS to im- prove the write performance, and also propose DFSPerf, a scalable distributed file system performance evaluation framework to help researchers and developers better design and implement systems in the Alluxio ecosystem.

使用了we，但是只有一个作者