Isaac Newton Institute for Mathematical Sciences

Annual Report
2007–2008

UNIVERSITY OF CAMBRIDGE

Contents

APPENDICES

Please note that the following statistical information may be obtained from the Institute on request, or from **www.newton.ac.uk/reports/0708/appendices.html**

Director's Foreword

David Wallace,
Director of the Institute

Welcome to the Annual Report of the Newton Institute, restructured this year under the headings: Science at the Institute; Serving the UK Community; and Management and Statistical Reports. Each of these sections is introduced by the relevant Chair: John Ball for the Scientific Steering Committee (SSC), Elmer Rees for the National Advisory Board (NAB), and Ekhard Salje for the Management Committee (MC). I am really grateful for the vital contribution which they and committee colleagues make to the continuing success of the Institute.

Scientific activities continue to flourish, with five programmes this year: Strong Fields, Integrability and Strings; Bayesian Non-Parametric Regression; Phylogenetics; Statistical Theory and Methods for High-Dimensional Data; and Combinatorics and Statistical Mechanics. In total 22 workshops were held, including five satellite meetings and two meetings oriented to interests in business and industry. More than 1060 scientists participated (64% from overseas), and 500 others attended specific events. I record deep gratitude to organisers.

Significant changes are underway in oversight of how the Institute serves the UK community. One condition of the new Engineering and Physical Sciences Research Council (EPSRC) grant was clarification of the role of the NAB, which was established in 1999 following an earlier EPSRC panel visit. The issue was discussed at all Institute committees; and Heads of UK Mathematics Departments, the 80 Institute Correspondents, and learned societies were consulted. The outcome was agreement that the NAB had fulfilled its role, not least through establishing Correspondents, that it should be disbanded, and that focus should be given to building up the role of Correspondents.

It is a great strength of the Institute that we can call on a network of scientists in relevant academic departments and research establishments to help us in the two-way communication within the UK community. Correspondents have a non-trivial task, given the breadth of the programmes, spanning pure and applied mathematics and all their applications in science, engineering and technology. The identification of a Chair of Correspondents, who will be co-opted as an additional member of the MC, will give them more influence on the direction and operation of the Institute.

Statistical information on the performance of the Institute this year is given in pages 23 to 25. The Institute's continuing existence is dependent on the financial support we receive, and it is a pleasure to thank the supporters listed on page 28. In particular, the EPSRC award of £9.6 million over six years from March 2008 is a tremendous foundation, most obviously financially, and also in the very welcome freedom that EPSRC gives the SSC; it underlines the Institute's huge responsibility to the UK community. I thank Ekhard Salje, who steps down as Chair of the MC, and am delighted to welcome his successor, Howard Covington, Chief Executive of New Star Asset Management.

Institute staff have again performed superbly in supporting the participants, with questionnaire returns of 91% 'excellent' and 9% 'good'. I thank them for their commitment and achievements, and congratulate particularly Steve Greenham, our Audio-Visual Technician, on his first class honours degree in Audio Music Technology.

Finally, I congratulate Robert Hunt on his appointment as Senior Tutor at Christ's College, and thank him for the exceptional contribution that he has made in his eight years as Deputy Director. We welcome Ben Mestel from the Open University, who succeeds him from 1 November.

JHN Chin

Dr Robert Hunt
Deputy Director
to August 2008

SGJ Penton

Dr Ben Mestel
Deputy Director
from November 2008

Science at the Institute


SJ Wilkinson

John Ball, Chair of the Scientific Steering Committee

Over the year, the Scientific Steering Committee (SSC) received ten new proposals, plus two resubmissions from the previous year. The pipeline of future programmes was significantly extended - see the following page. We are extremely grateful for the serious attention which referees give to our requests for their help - there were 90 reports in total this year. We thank the Fields Institute, which compiles a list of future programmes at other Institutes worldwide, which helps to avoid clashes and spot potential synergies.

A strategy discussion was held jointly with the National Advisory Board (NAB), reviewing the potential of new programme areas, the duration and nature of programmes, fundraising, and how the Institute might encourage increased participation from business and industry. We also discussed two governance issues: we endorsed the move to disband the NAB, given the planned strengthening of the role of Correspondents and increased external membership on the Management Committee; and established a Nominating Committee, which gives more detailed attention to the crucial issue of future membership.

I thank the members of the SSC, particularly John McWhirter (University College London) and Marie-France Vigneras (Paris 7, Institut de Mathématiques de Jussieu) who stepped down at the end of their term, and welcome David Fearn and Étienne Ghys. We look forward to John Toland (Bath) joining us; as Research Director of ICMS, he will be *ex-officio* in attendance, as part of the new arrangements *post* NAB. Finally, we give our best wishes to SSC secretary Kate Gilbert on her move to the Scott Polar Research Institute.

Scientific Steering Committee

Membership of the Scientific Steering Committee at 31 July 2008 was as follows:

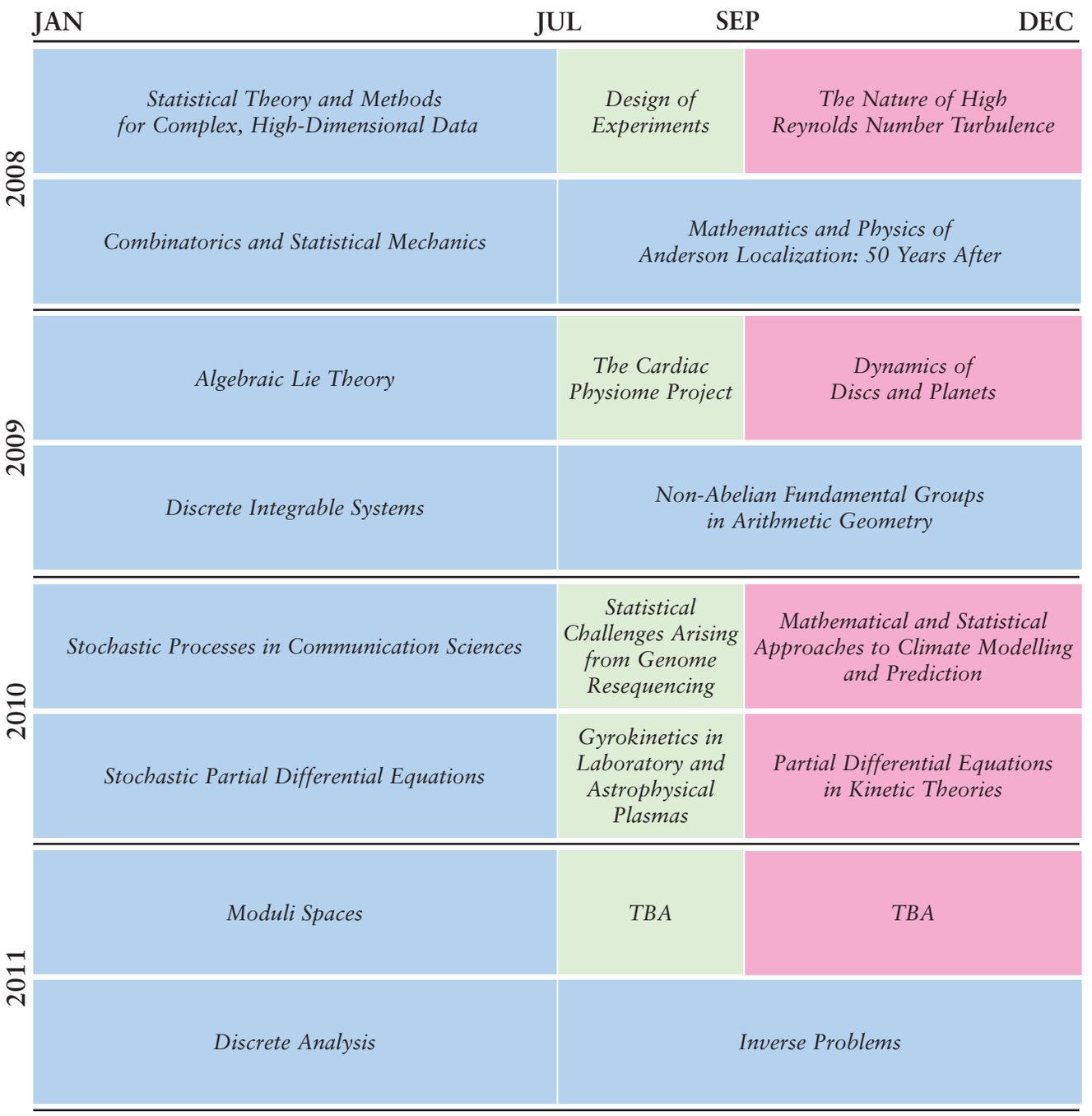
Sir John Ball FRS FRSE (Chair)	University of Oxford
Professor KG Binmore CBE FBA	University College London
Professor M Broué	Institut Henri Poincaré
Professor EK Burke	University of Nottingham
Professor PV Coveney FInstP FRSC	University College London
Professor DR Fearn FRSE	University of Glasgow
Professor CS Frenk FRS	University of Durham
Professor É Ghys	École Normale Supérieure de Lyon
Professor PJ Green FRS	University of Bristol
Professor SR Reid FREng	University of Aberdeen
Professor CM Series	University of Warwick
Professor LN Trefethen FRS	University of Oxford
Professor M Vingron	Max Planck Institute for Molecular Genetics
Sir David Wallace CBE FRS FREng (Secretary)	Director, Newton Institute

The Scientific Steering Committee (SSC) meets twice each year to consider proposals for programmes (of 4-week, 4-month or 6-month duration) to run two or three years later. Successful proposals are usually developed in a discussion between the proposers and the SSC conducted through the Director, and may well be considered at more than one SSC meeting before selection is recommended. Proposers may wish to submit a shorter 'preliminary' proposal in the first instance with a view to obtaining feedback from the SSC prior to the submission of a 'full' proposal. Complete details of the Institute's regular call for proposals, including guidelines for submission, can be found on the Institute's website at **www.newton.ac.uk/callprop.html**

Future Programmes

The schematic below shows the forthcoming programmes selected by the Scientific Steering Committee. Long-stay participation in a programme is at the invitation of the programme organisers; anyone interested should contact them directly. One- or two-week workshops are advertised, and applications invited with specific deadlines. Visits of one or two days are always welcome; we ask only that **reception@newton.ac.uk** is emailed in advance to assist us with planning.

Further details of each of these programmes, including the names and contact details of the organisers, can be found on the Newton Institute website at **www.newton.ac.uk/programmes/** Further information on how to participate in programmes can also be found on the website at **www.newton.ac.uk/participation.html**

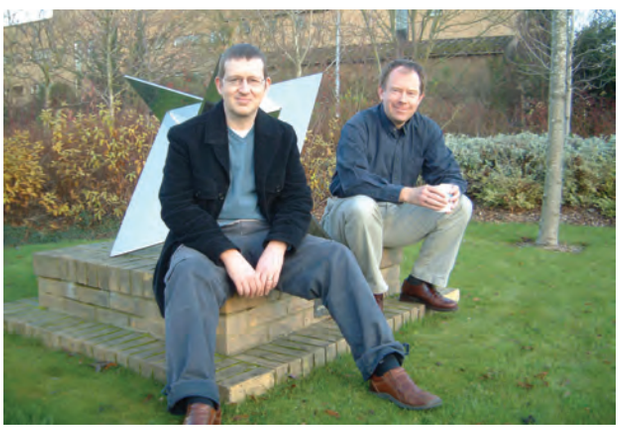


Strong Fields, Integrability and Strings

23 July to 21 December 2007

Report from the Organisers:
N Dorey (Cambridge), S Hands (Swansea) and N MacKay (York)

SJ Wilkinson



N Dorey and N MacKay

Theme of the Programme

For over 25 years, Quantum Chromodynamics (QCD), the gauge theory of quarks and gluons based on a gauge group SU(3), has been the accepted theory underlying strong interactions. However, many fundamental questions, which impact on the early universe as well as particle physics, still await satisfactory explanation: the masses of the observed particles; the nature of colour confinement of quarks through chromoelectric flux tubes and its relationship to strings; and the properties of the deconfined 'quark-gluon plasma' (QGP), the object of experiments at RHIC and the LHC.

To date the most systematic way of studying QCD non-perturbatively is via Lattice QCD, with fields defined on a Euclidean space-time lattice. Using Monte-Carlo importance sampling, much progress has been made in reproducing the hadron spectrum at the 5–10% level, in calculating important matrix elements, in identifying topologically interesting field configurations, and in bulk thermodynamics. Currently much attention is focussed on light fermions with the correct global chiral symmetries, which are required both for prediction of physics beyond the standard model, and for the formulation of supersymmetric (SUSY) models.

A productive avenue explored in analytic approaches is the observation that some dramatic simplifications occur in QCD-like theories with a large number N_c of colour charges. Perhaps the most surprising recent result is Maldacena's 1997 conjecture that gauge theory and string theory are equivalent in a particular model (strings on $AdS_5 \times S^5$ and N = 4 SUSY gauge theory) with the result that strongly-interacting gauge theory can be explored in the weak curvature limit of string theory, and vice versa, that gauge theory can shed light on information loss from black holes and provide a possible non-perturbative definition of string theory in certain space-times. This 'holographic principle' realises 't Hooft's vision that the large-N_c limit would converge into a string description of QCD.

The gauge theory/string theory correspondence has led to many powerful and unexpected results in strongly interacting theories, e.g., analytic predictions for the glueball spectrum; studies of chiral symmetry breaking and pion dynamics; and the calculation of the viscosity of N = 4 SUSY gauge theory, an effective model of the QGP.

In parallel, work on integrable models has provided a rich variety of insights into the analytic and algebraic structure of quantum systems and various beautiful connections between different branches of mathematical physics and mathematics. In recent years a recurring theme in the gauge/string correspondence has been the presence of integrability. For example, recent progress uses the techniques of integrability (exact S-matrices and the Bethe ansatz) to posit exact results for the energies of certain string states and thus (on the gauge side) the scaling dimensions of the corresponding operators. It seems that integrability plays a ubiquitous and powerful role in fundamental physics.

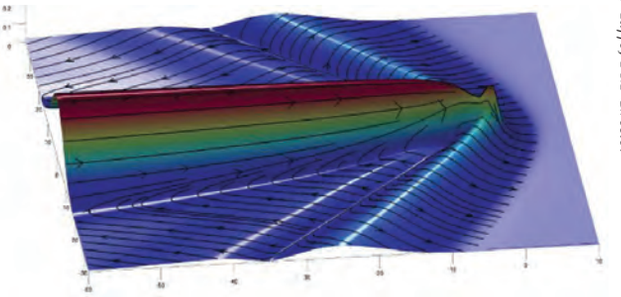
Structure of the Programme

The programme included two workshops and a 10-day school. We planned a rough schedule of themes: strongly-interacting thermodynamics and vacuum structure; the relation between gauge and string theory; and integrable models. Flexibility and serendipity brought together workers from different backgrounds – a major success of the programme. Rapid progress in the year leading up to the programme underlined the value of participation by experts in gauge/string duality.

As well as the typically daily seminar, some participants led extended introductory lectures and subsequent discussions: for instance, Aharony covered holographic models of deconfinement and chiral symmetry restoration in QCD, and Hanany covered dimers, quivers, counting problems and Hilbert schemes.

The first workshop, *Exploring QCD: Deconfinement, Extreme Environments and Holography* (organisers Evans, Hands, Teper), attracted 75 participants and 46 talks were given. It was especially topical given the conjunction of: heavy-ion collision experiments at RHIC; the imminent start-up of LHC; the exploitation of multi-teraflop computing resources; and recent dramatic advances in understanding gauge/gravity duality. Several talks focused on applications of the latter to study transport and the passage of heavy quarks through a strongly-interacting non-abelian plasma. Others reported progress in extending the holographic principle to non-supersymmetric theories. The workshop succeeded in its aim of initiating and sustaining dialogue between phenomenologists, lattice experts and gauge/string theorists.

The second workshop, *Integrability and the Gauge/String Correspondence* (organisers Dorey, MacKay, Tseytlin, Zarembo), attracted 83 participants and 40 talks were given. Talks covered the spin-chain of operator-traces in N = 4 SUSY gauge theory, the hidden integrability of the AdS string, as well as the techniques of classical and quantum integrability. The mix of expertise from gauge fields of strings (especially in AdS/Conformal Field Theory) and the integrable models community worked extremely well, with extended and lively discussions.



LG Yaffe, PM Chesler

Energy flux for $v = 0.75$

Energy flux in the wake of a heavy quark moving through a maximally supersymmetric Yang-Mills plasma at three-quarters of the speed of light. The flux magnitude is both colour-coded and indicated by the height of the surface; flow lines show the direction of the flux. Clearly visible in the flux is the Mach cone, as well as the diffusive tail behind the quark.

The school on *Gauge Fields and Strings* organised by D Tong provided both introductory and advanced lectures on topics related to the recent explosion of activity at the interface between gauge and string theory. Ten invited lecturers, including Microsoft Research Fellow Larry Yaffe, each gave 3–4 talks to an audience which included 47 registered students and younger researchers, many from the 29 programme participants present at the time, and 25 students from DAMTP at Cambridge. The online videos are a valuable resource. In total, 42 talks were given. Popular features included several 'gong shows' where students were given a glass of wine and invited to present their own research in the most exciting fashion possible!

Finally, with the help of the London Mathematical Society, we organised a 'Spitalfields Day' of talks aimed at undergraduate and postgraduate students, given by Evans, Shifman and Gorsky. The highlight was a talk entitled *The Coming Revolutions in Fundamental Physics* by 2004 Nobel Laureate and Rothschild Visiting Professor David Gross, a fitting occasion to mark the Institute's 15th anniversary, and so popular that it had to be relocated to a larger lecture theatre.

Outcomes and Achievements

The programme attracted a total of 254 researchers, including 124 workshop participants, with 51 papers or preprints listed in exit questionnaires. Participants gave 35 invited talks at UK universities, including Brunel, Cambridge, Cardiff, Durham,

Heriot Watt, Imperial College, King's College London, Liverpool, Oxford, Plymouth, Queen Mary, Swansea and York.

Given the nature of the programme and its success, it is difficult and somewhat invidious to pick out just a few outstanding topics for special mention. Some highlights include:

- A clarification of the relation between gravity backgrounds including black holes and the hydrodynamic regime of strongly interacting gauge theories, enabling the calculation of transport coefficients and sound propagation in the deconfined phase of gauge theories, which underpins the idea of the strongly interacting quark-gluon plasma (Policastro, Starinets, Minwalla, Stephanov, Reall, Hubeny, Gibbons, Rangamani).

- A study of the propagation of heavy quarks using the techniques of gauge/gravity duality, in particular showing the development of a Mach cone in the particle's wake, which may conceivably help in the interpretation of heavy ion collision data from RHIC (Yaffe, Vuorinen).

- The first steps towards a numerical calculation of transport coefficients have been made by relating Euclidean Green functions, calculable by numerical simulations of lattice QCD, to real-time response functions via the Kubo relation (Meyer, Aarts).

- The inclusion of fundamental, as well as adjoint, degrees of freedom in AdS/CFT models has expanded the scope of the predictions accessible to gauge/gravity duality, for instance, permitting the thermodynamic phase diagram of QCD-like theories to be probed (Aharony, Erdmenger, Evans, Peeters, Zamaklar). This can then be compared with lattice gauge theory simulations with non-zero baryon chemical potential (de Forcrand, Philipsen, Hands).

- Developments in simulating lattice models with exact supersymmetry, aided by theoretical ideas related to twisted or topological field theories and orbifolding, together with recent algorithmic advances in the simulation of light quarks. There has also been progress in understanding the relation between the various different methods currently being explored (Akemann, Catterall, Damgaard, Matsuura).

- Increased activity in simulating QCD-like theories with varying numbers of colours and with quark fields in varying representations of the gauge group. This has been inspired by the need to challenge recent analytic progress in understanding QCD in the large-N_c limit, but also has a phenomenological application in aiding truly non-perturbative study of walking technicolor scenarios for electroweak symmetry breaking, which may be tested shortly at LHC (Del Debbio, Bringoltz, Teper, Lucini).

- Enhanced understanding of the phase structure of Super Yang–Mills theory: specifically, exploring the relation between Polyakov–Maldacena loops in the supergravity duals and ordinary Wilson loops in the gauge theory. Of particular significance is the eigenvalue distribtion of the P-M loops, which shed light on how the phase structure of SYM theory changes between weak and strong coupling (Aharony, Hartnoll, Gursoy, Kumar). A related issue in SYM theory is the connection between Wilson–Maldacena loops with cusps and multi-gluon scattering amplitudes (Branhuber, Alday, Tseytlin, Travaglini, Roiban).

- Identification of a new supersymmetric gauge theory in three dimensions starting from a field theory model for M2-branes (Bagger, Lambert).

- Progress towards treating strings on $AdS_5 \times S^5$ in a manner which retains 2D Lorentz covariance, by using Pohlmeyer's reduction technique for integrable sigma models to reformulate the string sigma model using only the physical degrees of freedom (Grigoriev, Tseytlin).

- Use of the integrability of the sine-Gordon model, and its connection with $AdS_5 \times S^5$ strings, to construct two-magnon string states and thereby improve our understanding of magnon interactions and multi-magnon states (Klose, McLoughlin, Mikhailov, Schafer-Nameki).

- Development of a generalised scaling function for planar N = 4 SUSY Yang–Mills which yields predictions for a new one-parameter family of observables computable in the dual string theory (Staudacher and Freyhult).

A full report on this programme can be found at

www.newton.ac.uk/reports/0708/sis.pdf

Bayesian Nonparametric Regression

30 July to 24 August 2007

Report from the Organisers:
CC Holmes (Oxford), N Hjort (Oslo), P Müller (Texas) and SG Walker (Kent)

S Greenham

CC Holmes, SG Walker, P Müller and N Hjort

Scientific Background

Bayesian nonparametric inference is a relatively young research area, with the earliest papers appearing in the mid-1970s and greatly increased activity during the last 10 years. The group of researchers working in this field is still moderately small and there was a need for a focussed opportunity to survey the field, exchange ideas, identify gaps in the literature and coordinate research efforts. There is a lot of current activity, particularly in the area of regression, with many possible avenues of investigation.

Thus the purposes of the programme were to review the current state of nonparametric Bayesian research, to foster collaborations between researchers with different focus areas, and to identify important open problems. One prominent example was (and is) the need to combine ideas in the construction of random probability measures with research into random mean functions in a regression problem.

There is a need to build and strengthen connections between emerging areas of nonparametric Bayesian inference. Besides the already mentioned connection of random distributions and regression functions, examples include the relevance of asymptotic results to model choice and inference, generalisations of predictive probability functions, and random clustering models. We were particularly pleased that a number of people from the Machine Learning community were able to attend the programme, and that they highlighted practical problems to be solved.

Structure of the Programme

The programme was four weeks long, with an intensive workshop held during the second week. The remaining three weeks were deliberately left with maximum flexibility and opportunity for spontaneous interaction. We organised two informal talks each day, some being descriptions of current research ideas, some describing work in progress, and some reporting on recently completed research. The format was chosen to encourage interaction and lively discussions. There were 24 participants at the Institute for some or all of this informal part of the programme.

Bayesian Nonparametric Regression

6–10 August 2007

The first day of the workshop (which was named after the programme itself) was devoted to four tutorials given by Ghosal, Lijoi, Teh and Dunson. The remaining four days were taken up with general talks. There were 71 invited participants at the workshop, including over 40 speakers.

The tutorials provided an excellent survey of methods in nonparametric Bayesian inference, and they were extremely successful. Unusually for introductory material, more than 80% of participants stayed for all four tutorials.

The general talks included some on innovative applications of nonparametric Bayesian inference (Herring, House, Laud, Mena, Mukherjee, Popova, Ruggiero, Williams), novel methods for survival analysis (De Iorio, Guglielmi, Johnson, Yin), issues of prior choice (Cox, MacEachern), new nonparametric models (Arjas, Griffin, Prünster, Quintana), constructions of families of dependent random probability measures (Basu, Dunson, Petrone, Spano, Steel), and asymptotic results (Choi, De Blasi, Lee).

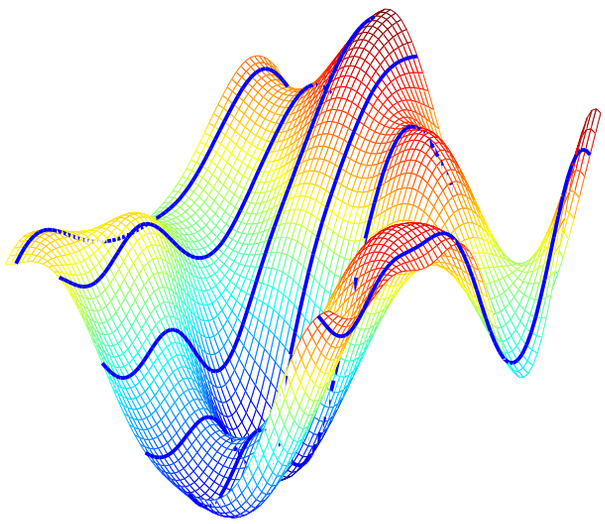
Outcome and Achievements

The programme successfully surveyed the field and pointed towards specific directions for future applications. It generated interest in alternative models that generalise the most traditional ones, which currently dominate practical nonparametric Bayesian methods. At the same time, several talks highlighted and discussed distinguishing features of these traditional models.

Some new models under discussion aimed to introduce dependence into Pólya trees, one of the main methods of creating nonparametric Bayesian models. One formulation of the problem is to drive the dependence with a collection of stochastic processes introduced as a piece of the model. A second formulation is to drive the dependence through shared portions of the trees.

Other ideas included the exploration of random effects models that preserve selected features of a set of distributions as covariate values change. The feature to preserve in this instance is the expected probability of an event. With the asymmetry of the mapping from the probability scale to the logit scale, this requires a random effects distribution that changes with the covariate in a very particular fashion.

Dependent Dirichlet processes (and, a bit more generally, dependent nonparametric processes) provide a framework which encompasses many of the nonparametric Bayesian models that are currently being proposed. The relationship between the models becomes clearer when set in the framework of dependent Dirichlet processes. This context also indicates possible ways in which to refine the models, as there are generally effective strategies that allow one to capture specific types of behaviour.



C Williams

An example of multi-task modelling with Gaussian processes from machine learning.

Many other ideas were discussed, such as modelling dependence via partial exchangeability; sampling methods and models for tree- and time-structured dependent Dirichlet processes; and hierarchical mixtures and general consistency issues.

Most of the talks and presentations in the informal part of the programme addressed the question that initially motivated the programme, namely the combination of random probability measures and random regression mean functions to construct real nonparametric regression models. Several talks focussed on the construction of families of random probability measures indexed with a covariate. This formally provides a constructive definition of the desired combinations.

Feedback from participants makes it quite clear that many collaborations are under way and many new projects have been initiated as a consequence of the intensity of the programme. The organisers would like to express their deep gratitude to the Institute for their outstanding hospitality.

A volume entitled *Bayesian Nonparametrics in Practice*, containing material from the four tutorials given during the workshop plus one chapter provided by each of the four organisers, is currently in preparation and will be published by Cambridge University Press.

A full report on this programme can be found at

www.newton.ac.uk/reports/0708/bnr.pdf

Phylogenetics

3 September to 21 December 2007

Report from the Organisers:
V Moulton (East Anglia), MA Steel (Canterbury) and DH Huson (Tübingen)

SJ Wilkinson

MA Steel, V Moulton and DH Huson

Scientific Background

Phylogenetics is the reconstruction and analysis of trees and networks to describe and understand the evolution of species, populations and individuals. It is widely used in molecular biology and other areas of classification (such as linguistics), and has both led to, and benefited from, the development of new mathematical, statistical and computational techniques. Although the foundations of phylogenetics were laid down many decades ago, it is currently experiencing an exciting renaissance due to the wealth and types of biological data that are becoming available.

This Newton Institute programme attracted 159 researchers in phylogenetics and related areas from around the world, including 57 Visiting Fellows who stayed at the Institute for extended periods. The programme aimed to develop our knowledge on the following main themes, which provide a rich source of problems in diverse areas such as combinatorics, algorithmic complexity, graph theory, probability theory, topology and algebraic geometry.

New data types in phylogenetics: Until quite recently most modern methods for constructing phylogenetic trees have been designed with sequence data in mind, usually constructing evolutionary trees from genes as an approximation to species phylogenies. However, the abundance of new types of molecular data is creating interesting new challenges. Not only do we have to reconsider previous estimates of phylogeny in view of the new data, but new methods need to be established that allow the incorporation of subtle phylogenetic signals in the data. Moreover, the incorporation of phylogenetic information into bioinformatics methods for tackling problems such as motif discovery in genomes/biochemical networks, also known as phylogenetic footprinting, can significantly improve sensitivity, although often at the price of introducing hard mathematical and computational variants of well-studied problems.

Reticulate evolution: How can we best model reticulate evolution by networks generalising trees? For example, from genomic data can we determine how much gene transfer occurred early in the Tree-of-Life by comparing the genomes of extant species? Various techniques for building networks have been proposed. For example, a rich mathematical theory is emerging for representing phylogenetic relationships using so-called split networks (see figure opposite). These networks provide a snapshot of data which can indicate the presence of incompatibilities that are often the consequence of non-tree-like evolutionary processes. Even so, there is great interest in new theories and constructions for phylogenetic networks that provide a more concrete representation of reticulate evolution.

Constructing large trees: Biologists wish to build large trees across thousands of species leading to substantial combinatorial and statistical problems. These trees not only deepen our understanding of the Tree-of-Life, but also provide useful information to understand global biodiversity. However, popular methods for tree reconstruction (such as maximum parsimony and maximum likelihood) can sometimes

be far too computationally expensive for deriving large trees. Moreover, biologists commonly wish to combine several trees from overlapping data sets to obtain overall estimates of phylogeny.

Mathematical modelling of evolution: Stochastic models have long played an important role in phylogenetics. Indeed, Yule in 1924 showed how simple branching-type processes could model the distribution of species numbers by genera. More recently, statisticians have begun to study how the 'shape' of phylogenetic trees could be predicted from simple speciation models. Further investigations by probability theorists and biologists have allowed for features of published trees to be studied, with the goal of learning more about the process of speciation and testing specific hypotheses. Other processes in phylogenetics where models are of interest include the study of character evolution – for example, how does DNA evolve, and how can we use these models to refine methods for tree reconstruction? Another is the use of species-level phylogenetic techniques to study population-level processes through the coalescent process.

Programme Structure

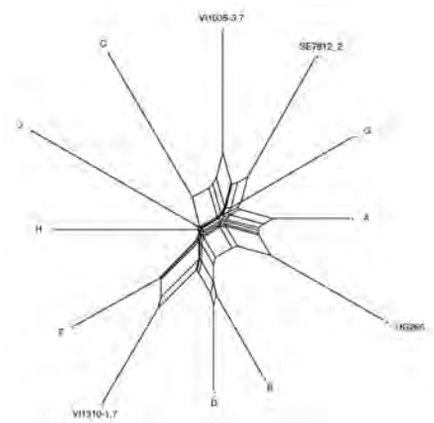
The programme included three workshops together with a half-day meeting aimed at newcomers to phylogenetics:

- *Current Challenges and Problems in Phylogenetics (EMBO)*, 3–7 September: 85 participants attended, 11 invited talks and 28 contributed talks were given, and there were 15 contributed posters.
- *Phyloinformatics Workshop*, e-Science Institute, Edinburgh, 22–24 October: 24 participants, 12 invited talks.
- *Yggdrasil: Reconstructing the Tree of Life (Spitalfields Day)*, 6 December: 35 participants.
- *Future Directions in Phylogenetic Methods and Models*, 17–21 December: 71 participants, 41 invited talks.

The events were a great success, including between them many world-class invited talks, contributed talks, lectures and discussion sessions.

Outcome and Achievements

New data types in phylogenetics: There are now thousands of whole genome sequences available, allowing us to dig even even more deeply into the evolutionary history of present day organisms. For example, phylogenetic trees are now being built for HIV viruses based on whole genome sequences, and programme participants (e.g., Lemey, Pybus, Rambaut) worked on developing new tools to understand virus evolution. In related work, other participants (e.g., Gascuel, Spencer, Székely, Vision) grappled with problems in bacterial genome evolution such as how to deal with subsets of genes having different behaviours, and how to compare multiple genomes. It also became clear that there is still debate on how tree reconstruction methods originally designed to deal with single genes should be extended to whole genomes: several of the participants (e.g., Allman, Kim, Matsen, Rhodes, Steel) worked on developing a more unified theory for mixture models. These models have been proposed as a way for biologists to analyse data in which certain DNA sequence sites evolve quite differently to other sites, due to structural or functional constraints. Such models can seriously mislead existing phylogenetic approaches, and further work is needed. A further insight into the problem of tree reconstruction from non-homogeneous data was a theoretical result concerning the complexity of computing a most parsimonious tree for two genes (Grünewald, Moulton). An exciting new direction was also presented by the new generation of sequencing technologies, which present the possibilities of sequencing short genomes in hours or of gathering large numbers of markers from larger genomes. One topical application is metagenomics, the study of genetic material recovered from envir-



A split network computed from HIV sequences detailed in 'The Phylogenetic Handbook' (CUP, 2003). Letters A–J denote HIV-1 subtypes, and the remaining labels denote recombinant viruses.

onmental samples (e.g., of soil). In this context, participants (e.g., Huson, Rodrigo, Spencer) developed new methods to deal with issues such as *How to separate mixtures of genomes?* and *How to decide statistically the abundance of sequences coming from each genome in the sample?*

Reticulate evolution: There was much interest in the further development of the theory of phylogenetic networks based on acyclic digraphs. These networks can provide an intuitive representation of the evolutionary relationships between species, although surprisingly little is known concerning the combinatorial properties of such networks and general methods for their construction. The programme generated new insights: for example, methods were developed for constructing networks from combinatorial data such as triplets and clusters, as well as new software for their computation (e.g., Dress, Huber, Huson, Kelk, Rupp, Stougie, Willson). Applications of such networks to recombination and reconstruction of whole genome phylogenies were also pushed forward (e.g., Holland, Huson, Gusfield, Lockhart, Willson). New results were developed concerning tanglegrams (Gusfield, St John) and the reticulate evolution of languages (Warnow).

Constructing large trees: Some approaches by graph theorists are already being applied by biologists in the construction of large trees in the form of 'supertrees' (combining trees that classify overlapping sets of species). Algorithmic approaches were developed to handle constraints such as edge-lengths, divergence dates and ancestral taxa (e.g., Semple, Willson), and to build such trees (and networks) from dense data (e.g., Kelk, Huber, Willson). Alternative methods for efficiently constructing large trees based on distance measures and likelihood scores were also developed, together with a theoretical analysis of the issues when ancestral data is involved (e.g., Holland, Roch, Warnow, Whelan). Applications of large trees to the understanding of diversity generated much interest. For example, a conjecture concerning phylogenetic diversity for two trees was solved and new methods were developed to improve the applicability of phylogenetic diversity (e.g., Bordewich, Hartmann, Klaere, Rodrigo, Semple, Spillner, von Haeseler).

Mathematical modelling of evolution: One of the main tools in understanding how DNA evolves is the study of Markov models of sequence evolution (on a tree or network), and it is the basis of widely-used likelihood-based and Bayesian approaches to phylogenetics (as well as 'corrected distance' approaches). Participants worked on improving the accuracy of such models through, for example, estimating empirical substitution matrices from huge alignment databases (e.g., Gascuel, Goldman, Holder). Research was also done on the consequences of model mis-specification (e.g., Howe, Lockhart, Naylor, Steel) and on methods for accelerating Bayesian MCMC inference (Nicholls, Rodrigo). Many stochastic models lead to interesting mathematical problems which are of interest in their own right. For example, Markov models for sequence evolution give rise to polynomial ideals ('phylogenetic invariants') that have a rich algebraic and geometric structure; they were intensively studied by several participants (e.g., Allman, Kim, Matsen, Rhodes), leading to new results concerning model identifiability, and geometry of phylogenetic models.

Related probabilistic questions were also studied (e.g., Mossel, Roch, Steel, Székely) yielding solutions to two problems: *How can we efficiently reconstruct species trees from gene trees that conflict due to lineage sorting?* and *Is the amount of data required to 'test' whether or not a given phylogenetic tree is 'true' fundamentally less than the amount of data required to reconstruct the phylogenetic tree from scratch?* Participants (e.g., Dress, Grünewald, Huber, Koolen, Moulton, Spillner, Steel) also pushed forward the new mathematical theory of phylogenetic combinatorics, which is concerned with the combinatorial problems involved in modelling evolution and constructing trees. New insights were also gained concerning SPR/TBR combinatorial tree moves and their relation to tree-space (e.g., Bordewich, Erdös, Gascuel, Huber, Steel, Székely).

One of the main outcomes will be a special issue of the journal IEEE/ACM Transactions in Computational Biology and Bioinformatics. In addition, three books are in preparation and 50 papers were written.

A full report on this programme can be found at

www.newton.ac.uk/reports/0708/plg.pdf

Statistical Theory and Methods for Complex, High-Dimensional Data

7 January to 27 June 2008

Report from the Organisers:
D Banks (Duke), P Bickel (California, Berkeley), IM Johnstone (Stanford) and DM Titterington (Glasgow)

SJ Wilkinson

DM Titterington and D Banks

Scientific Background

Most twentieth-century statistical theory was restricted to problems in which the number of 'unknowns' (such as parameters) is much smaller than the number of experimental units. However, the number of 'unknowns' in the modern massive datasets which arise in areas such as image analysis, genomics, astronomy and climatology can hugely exceed the number of experimental units. Innovative core statistical theory is required, along with new methodology and novel approaches to graphical display, that cope with these important practical scenarios and exploit state-of-the-art computing capability. Key advances engage both mainstream statisticians and the machine-learning community.

The general aim of the programme was to promote research in the following areas: strategies for dimension-reduction, including latent-structure modelling and the exploitation of sparsity; classification methods for large-scale problems; asymptotics for increasing dimension; and visualisation methods for complex datasets. Applications of breakthroughs in theory and methods covered by the programme are essential for the proper analysis of data in a wide range of contexts in modern-day life.

The organisers benefited from the advice, as Scientific Advisors, of C Bishop (Microsoft), P Hall (Melbourne), J Shawe-Taylor (University College London) and S van de Geer (Zurich).

Programme Structure

The programme attracted 88 visiting fellows and 16 programme participants, with about 75% of the total from overseas. Apart from the workshops detailed below, most day-to-day activity was left to spontaneity but there were two or three more formal seminars per week, and a dedicated short series on metabolomics, organised by Banks and Dianne Cook. Butucea, with Clarke, initiated a working group on performance bounds for inference.

Contemporary Frontiers in High-Dimensional Statistical Data Analysis

7–11 January 2008

Organisers: D Banks, M Titterington, S van de Geer

The opening workshop, with 111 participants, laid out the main themes for the programme. Talks by Donoho, Wainwright, Niyogi and Candes addressed the information-theoretic limits that arise in problems where the number of observations is smaller than the number of potentially explanatory variables. Van de Geer, Young, Lee, Yu, Dennis Cook, Samworth, Nadler, Bickel and Jordan described ways to approach those limits, using methods such as the Lasso, high-dimensional bootstrapping, treelets, regularisation, dimension-reduction and kernel-based contrast functions. Visualisation was addressed by Dianne Cook, Stuetzle and Wegman, and applications in finance and bioinformatics were covered by West, Mammen and Fan. These presenters gave perspectives from statistics, mathematics and computer science, and

that cross-disciplinary interaction has been the engine for recent progress in this field. There were 27 invited talks and 20 posters.

High-Dimensional Statistics in Biology

31 March – 4 April 2008

Organisers: P Bickel, E Birney, R Durbin, W Huber

This workshop took advantage of local strength in molecular biology in Cambridge University, the European Bioinformatics Institute and the Sanger Institute. There were 133 participants, 23 talks and 20 posters. The mix of speakers achieved the major aim of exposing mathematical scientists to the great variety and complexity of genomic data and to the underlying biological goals. Topics included multiple comparisons (Benjamini), the statistical complexity of the genome (Bickel), biological database issues (Huang) and classification (McLachlan, West). The amount of data, already huge, is clearly on the threshold of another level of exponential growth, due largely to new, high-throughput, relatively cheap, sequencing technologies: Durbin described the 1000 Genomes project, which compares the genomes of 1000 humans at extremely fine resolution. Comparative genomics, inferring structure from high conservation between species, and phylogenetics, will have to incorporate new dimension-reduction ideas, as discussed by McVean, Margulies and Birney.

Bayesian Analysis of High-Dimensional Data

14–16 April 2008

Organisers: D Banks, J Griffin, F Rigat, M Steel

This satellite meeting at the University of Warwick provided the programme with a distinctive Bayesian element. There were 63 participants, who enjoyed 20 invited talks, 3 contributed talks and 17 contributed posters. The workshop highlighted recent methodological and applied advances in the Bayesian analysis of complex data. Keynote and themed talks focussed on selected topics in biostatistics, computational systems biology and mathematical statistics. Methodological contributions were provided by Hjort, Liang, Li, Dobra, Guindani, Kass, Teh and Rasmussen, and context-related modelling featured strongly in many talks. Those by Schmid, Ghosh, Morris, Wilkinson and Purutcuoglu all had a biological flavour, and climatological applications were described by Sanso, Nychka and Haslett.

Inference and Estimation in Probabilistic Time-Series Models

18–20 June 2008

Organisers: D Barber, T Cemgil, S Chiappa

This workshop, partially funded by the European Network of Excellence PASCAL2, encouraged cross-fertilisation of ideas among the machine-learning, engineering and statistics communities. For example, parameter estimation in state-space models often uses variational procedures in machine learning, Markov chain Monte Carlo (MCMC) methods in statistics and subspace methods based on Hankel matrices in engineering. Opper's application of deterministic approximate inference to continuous Markov processes contrasted very well with Papaspiliopoulos's talk on advanced MCMC-based techniques. The work of Sykulski and Olhede on approximating integrated volatility in stochastic differential equations was also directly related. Godsill raised doubts about the ultimate feasibility of inference for complex models of an environment, given that parameter estimation is compounded by strong temporal posterior correlations; this point was emphasised in Turner's talk on variational inference for time-series models. Altogether there were 19 talks, 4 posters and 80 participants.

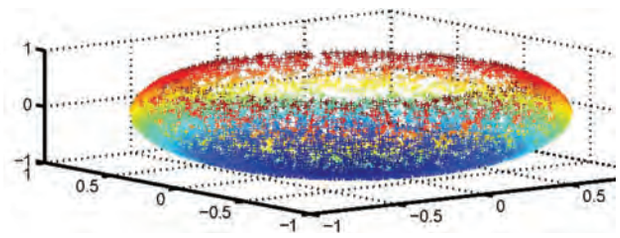
Future Directions in High-Dimensional Data Analysis

23–27 June 2008

Organisers: D Barber, I Johnstone, R Samworth, M Titterington

The closing workshop attracted 110 participants, and had 20 invited talks, 9 contributed talks and 13 posters. Several talks were based on advances made during the programme, and many looked to future directions. Talks by Koltchinskii, Meinshausen, Tibshirani, Tsybakov and Yuan concerned algorithms and properties for large linear regression models when sparsity in the coefficients is expected and exploited. Machine-learning perspectives on sparse regression and models for large graphs or networks were also discussed. The workshop's subtitle was *New Methodologies, New Data Types and New Applications*, with the last including challenges raised by large-scale astronomical surveys (Rice), and genetics and systems biology (Li, in particular).

PJ Wolfe



An important goal of the programme was to develop an efficient approach to the analysis of massive datasets such as the 10^5 random point "fishbowl" example shown here. From Belebbas M-A & Wolfe PJ (2009) Spectral methods in machine learning and new strategies for very large data sets. *Proc. Natl. Acad. Sci. 106 (2) (copyright 2009, National Academy of Sciences, USA).*

The longer workshops were complemented by a variety of shorter events. A 'Financial Data Day', organised by C Rogers, attracted an audience of about 70, mostly from the finance industry itself. Two sessions were given by members of Cambridge groups: the Machine Learning Group in Engineering, and the Statistical Laboratory. One afternoon brought together participants from the parallel programme on *Combinatorics and Statistical Mechanics*, fruitfully generating substantive collaborations. An 'Open for Business' day, organised in collaboration with R Leese of the Knowledge Transfer Network for Industrial Mathematics, alerted senior people in business and industry to the expertise and applicable research associated with the programme. The format of expository talks, a discussion panel and the opportunity to network and to engage with participants in the informal ambience of the Institute was extremely successful.

The most important individual presentation of the programme was the lecture entitled *More unknowns than equations? Not a problem! Use sparsity!*, by the Rothschild Visiting Professor, David Donoho. In addition, D Banks provided a well-received talk as the Institute's contribution to the Cambridge Science Festival.

Outcome and achievements

The main aims of the programme were to accelerate ideas in a crucial area of core statistical research and to bring together world leaders in both statistics and machine learning.

Topics on which progress was made include the following: issues concerning the approximation and estimation of large matrices (Cai, Hero, Claeskens, Wolfe, Müller, Yuan, Donoho, Johnstone, Zhou); nonparametric Bayesian analysis (Bickel, Kleijn, Gamst); issues of sparsity (many participants, including Koltchinskii, Levina, Nadler, Rohde, Tsybakov); oracle inequalities and properties (Bunea, Nan, Tsybakov, Wegkamp); and visualisation (Dianne Cook, Hero, Marron). The Rothschild Professor, David Donoho, interacted with many people, especially Tanner on random polytopes, Jin on optimal feature selection, and Kutyniok on the geometric separation problem. Wellner worked on Nemirovski's inequality with Dümbgen and van de Geer, and on empirical process theory with van der Vaart. Microsoft Fellow Dennis Cook developed a general framework for sparse dimension reduction and, with Kent, looked into ways of unifying the understanding of regularisation methods. With Dryden, Olhede developed a new method for estimating diffusion tensors in medical image analysis; and with Kovac he worked on bivariate curve analysis. With Wagner of the *Combinatorics and Statistical Mechanics* programme, Banks worked on agent-based modelling. Rajaratnam and Young initiated a collaboration on bootstrap ideas.

The dialogue between statistics and machine learning was strongly apparent. The former involved Rajaratnam, Kent, McLachlan, Robert, Samworth, Shi, Yuan and Titterington, and the latter, amongst others, Lafferty, Lawrence, Seeger, Pontil, Barber, Teh, Shawe-Taylor, Cristianini, Roweis and Murray. The hope is that seeds have been sown for productive interactions and further constructive blurring of this interface.

As well as individual journal publications, two compilations are envisaged: one is with Cambridge University Press for a book of contributed chapters based on the Time-Series workshop. Secondly, the Royal Society has accepted a proposal from the programme organisers for a theme issue in the Society's *Philosophical Transactions, Series A*.

Feedback indicated universal praise for the scientific and social excellence of the Institute's environment and for the efficiency and friendliness of its staff.

A full report on this programme can be found at

www.newton.ac.uk/reports/0708/sch.pdf

Combinatorics and Statistical Mechanics

14 January to 4 July 2008

Report from the Organisers:
PJ Cameron (Queen Mary, London), B Jackson (Queen Mary, London), A Scott (Oxford), A Sokal (NYU/UCL) and DG Wagner (Waterloo)

SGJ Penton

A Sokal, D Wagner, P Cameron, A Scott and B Jackson

Scientific Background

The aim of the programme was to stimulate the growing interactions among combinatorialists, probabilists, computer scientists and theoretical physicists concerned broadly with probability theory or statistical mechanics on graphs and other structures such as matroids, set partitions and constraint satisfaction problems. In particular, increasing computer power has greatly widened the class of examples which can be examined, and methods from physics are increasingly used in combinatorics.

Programme Structure

The programme included six workshops, as follows

- *Introductory Cross-Disciplinary Symposium,* 14–18 January: 48 participants attended, and 19 invited talks were given.
- *Zeros of Graph Polynomials,* 21–25 January: 60 participants, 12 invited talks, 16 contributed talks and 3 contributed posters.
- *Markov-chain Monte Carlo Methods,* 25–28 March: 74 participants, 29 invited talks.
- *Combinatorial Identities and their Application in Statistical Mechanics*, 7–11 April: 86 participants, 28 invited talks and 3 contributed posters.
- *Statistical-Mechanics and Quantum-Field Theory Methods in Combinatorial Enumeration,* 21–25 April: 57 participants and 23 invited talks.
- *Combinatorial and Probabilistic Inequalities*, 23–27 June: 47 participants and 22 invited talks.

There were 100 longer-term participants, including Professor C Thomassen (Rothschild Visiting Professor) and Professor S Janson (Microsoft Research Fellow). Regular seminars were held during weeks when no workshops ran, and the participants interacted in the excellent surroundings of the Newton Institute.

A cross-programme seminar was held with participants of the *Statistical Theory and Methods for Complex and High-Dimensional Data* (SCH) programme. This was very successful, and revealed a number of points of contact between the two programmes, for example, convergence rates for Markov chains, and use of Markov chain methods for exploring high-dimensional datasets.

The workshop on combinatorial identities was dedicated to Pierre Leroux, who helped to organise it and died shortly before. As a participant commented, it was “a moving celebration of Pierre Leroux and his work”.

Outcomes and Achievements

A number of areas saw significant advances as a result of the programme. Given space constraints, a few selected examples follow.

- A working seminar on the algebraic number theoretical properties of chromatic roots explored some general questions: which algebraic numbers arise as chromatic roots, and what is the typical behaviour of Galois groups of chromatic polynomials? We formulated some conjectures, made progress towards their solution, and found a number of remarkable examples. This involved the

Director and a number of long-term participants (especially Cameron, Dong, Farr, Jackson, Morgan, Sellers, Sokal, Wagner), and also V Dokchitser, a number theorist from the University of Cambridge.

- Jackson, Noble and Wagner made significant progress on a 10-year old conjecture of Merino and Welsh, answering a related question of Jerrum in the process. The conjecture concerns values of the Tutte polynomial at specific points; for graphs, these count spanning trees, acyclic orientations and totally cyclic orientations. Cameron and Thomassen attacked the Merino-Welsh conjecture from another point of view and obtained new estimates for the numbers involved.

- There was significant work on negative correlation, negative association and log-concavity by Borcea, Brandan, Kahn, Neiman, van den Berg, Wagner and Chaiken, which has continued since the end of the programme. In particular, conjectures of Markström and Welsh were disproved.

- Fernandez, Procacci, Salas, Kotecky, Severini, Sokal, Shrock, Faris and Jackson worked on cluster expansions and produced several new approaches, including a general set-up for cluster expansions for systems with two-body attractive interactions; cluster expansions for not necessarily positive interactions; and the meaning of cluster expansions in the theory of species.

- Royle and Sokal found a bound on the chromatic roots of series-parallel graphs in terms of maxmaxflow (a bound that is also surprisingly close to sharp). Also, Royle found examples of graphs with real flow roots above 4, disproving a conjecture of Welsh, and he is working with Jacobsen and Salas to try to show that there can be roots approaching 5 from below.

- Among many other successful research collaborations, we note work on spin models; computing Ottaviani's invariant; using improved bounds on the Mayer expansion to prove existence of mean field phase transition near the Kac limit (using insights from the Identities workshop); multipermutation solutions of the set-theoretic Yang–Baxter equation; counting defective parking functions; matroids defined by root systems; computational complexity of partition functions and counting hypergraph homomorphisms; the connection between Ashkin–Teller model partition functions and generalised Tutte–Whitney polynomials; the subgraph enumerating polynomial; the asymptotic upper matching conjecture; extending the 'entropy method' for counting graph homomorphisms to non-bipartite graphs; zero-free regions for the Potts model; complex-temperature phase diagrams and associated partition function zeros; chromatic zeros of three-dimensional lattices; Cayley-type identities (new identities and new proofs of traditional identities by methods of fermionic integration, hopefully giving meaning to a 'fractional' Cayley-type identity); Capelli identities (a proof technique which extends the results beyond Weyl algebras); and invariant theory for OSP(1|2) and OSP(1|2m) and the relationship with counting spanning forests.

- As a result of the cross-programme session with SCH, Penman worked with Marron on a problem arising from DNA, which boils down to 'coding with an involution'; Severini worked with M Pontil on ground states of Hamiltonians; Banks, Said, Wagner, and Wegman worked on relations between agent-based models in statistics and interacting particle systems in statistical mechanics.

Some 73 reports were produced during the course of the programme. A survey paper on graph and matroid polynomials containing information important for statistical mechanics is being prepared by Ellis-Monaghan.

Participants' comments were extremely positive, with the interdisciplinary nature of the programme being much appreciated. For example, comments included:

"The high quality of the environment and facilities enabled me to perform research particularly effectively."

"A fantastic opportunity to meet great mathematicians and physicists and learn new techniques and results."

"A terrific workshop. The support provided by the Institute will have a lasting effect on my research."

A full report on this programme can be found at

www.newton.ac.uk/reports/0708/csm.pdf

Serving the UK Community

Elmer Rees, Chair of the National Advisory Board

The major issue this year has been the EPSRC requirement to clarify the role of the National Advisory Board (NAB). Two aspects were important in arriving at the conclusion that the NAB had served its main purpose and that its remaining duties could be carried out as effectively by other mechanisms it had put in place.

Firstly, the NAB has established the network of Correspondents, introduced systematic data collection and dissemination, and expanded the number of satellite workshops at other institutions. It has also introduced training events, developed the Junior Membership scheme, and encouraged the complementary relationship with the ICMS in Edinburgh. In addition, the new Director was making some 30 visits to Departments, and was giving presentations to Heads of Mathematics Departments and to the British Mathematical Colloquium. These successful changes instigated and encouraged by the NAB have enabled the Institute to better serve UK national interests.

Second, changes are being set in train to strengthen the role of the Correspondents. The NAB felt that a Chair of Correspondents, selected independently by the learned societies from nominations by Correspondents, would be a strong voice for UK community interests, particularly after secondment to the Management Committee. The NAB supported the identification of Howard Covington as an independent Chair of the Management Committee. The additional requirements of the EPSRC award provide further assurance, and I am pleased to see the progress towards these on page 21.

The process to identify the first Chair of Correspondents has resulted in Professor Caroline Series from the University of Warwick being appointed for a two-year term. I am confident that the new arrangements will be successful and offer those involved my very best wishes.

Professor Caroline Series
Chair of Correspondents

National Advisory Board

Membership of the National Advisory Board during its final year was as follows:

Sir John Ball FRS FRSE	University of Oxford
Professor JW Bruce	University of Hull
Professor P Grindrod CBE	Lawson Software, Oxford
Dr RE Hunt	Deputy Director, Newton Institute
Sir Peter Knight FRS	Imperial College London
Professor JG McWhirter FRS FREng	University of Cardiff
Professor EG Rees FRSE (Chair)	University of Edinburgh
Professor GO Roberts	University of Warwick
Dr FA Rogers	King's College London
Professor E Salje FRS	University of Cambridge
Professor SM Schaefer	London Business School
Professor JF Toland FRS FRSE	Director, ICMS
Sir David Wallace CBE FRS FREng	Director, Newton Institute

Institute Activities

Short Visits to the Institute

Any researcher associated with a UK University, academic institution or R&D group in industry or commerce may visit the Newton Institute for up to two days without an invitation, in order to attend seminars or to work with colleagues. We ask only that **reception@newton.ac.uk** is emailed in advance to assist us with planning.

Unfortunately we are unable to guarantee office space, accommodation or meals during such short visits, but visitors are welcome to use the common areas of the building and our library.

More information is available at

www.newton.ac.uk/shortvisits.html

UK Correspondents

The Newton Institute has for several years maintained a list of Correspondents in UK Universities to act as a channel of communication between the Institute and the mathematical sciences community in the UK. This list has in recent years been extended to include relevant learned societies, commercial organisations and institutions not attached to Universities. All Correspondents are regularly informed about activities of the Institute, and it is their responsibility to ensure that the information is disseminated to relevant individuals within their institution, whether in mathematics departments or in other scientific groups appropriate to each event. Correspondents also provide invaluable feedback, particularly at the Annual Meeting of Correspondents held at the Institute. The names of all Correspondents as at 31 July 2008 can be found on page 20.

Seminars on the Web

To increase the benefit of Institute programmes to the UK mathematical community, seminars delivered during workshops or at special events at the Institute are recorded and made available on the web at

www.newton.ac.uk/webseminars/

For most past seminars, audio files together with accompanying transparencies or PowerPoint files, etc., can be downloaded. Starting from September 2007, however, full **video** of every seminar has been available for either streaming or download. In time the library of online seminars should build into a substantial national resource. At current rates over 400 seminars are being added *per annum*.

Seminars in the UK

Visiting Fellows on Newton Institute programmes are strongly encouraged to visit other institutions within the UK during their stay at the Institute, and many did so during 2007/08. To promote this activity, the Institute covers on request the travel costs within the UK for any overseas Fellow.

Lists of future participants, with dates of their visits to the Institute, can be found on the individual programme web pages. In addition, the Institute has set up a register, with titles of topics, of those Fellows who are willing to travel to other UK institutions to give seminars. Correspondents are urged to ensure that organisers of local seminar series know about and consult this register when planning their schedule of speakers. Potential speakers may be contacted directly using the details listed in the register, which can be found at

www.newton.ac.uk/programmes/speakers.html

Alternatively, advice on suitable speakers may be obtained from the organisers of any Institute programme via the Institute.

Junior Membership

The Institute recognises that early career researchers have much to contribute to, and gain from, Institute programmes and events. In order to maximise the information available to them, and to facilitate their involvement in Institute activities by offering additional funding opportunities, there is a special scheme for Junior Membership of the Newton Institute. To be eligible you must be either a Research Student or within 5 years of having received your PhD (with appropriate allowance for career breaks), and you must work or study in a UK University or a related research institution.

Those wishing to join the scheme should consult the Institute's web site at

www.newton.ac.uk/junior.html

Members will receive regular advance information regarding programmes, workshops, conferences and other Institute events. The Institute also makes available some of its general funds specifically to support early career researchers taking part in Institute activities. Members may apply for grants from these funds. Types of involvement supported include attendance at workshops, conferences, etc., and visits of up to two weeks to work or study with longer-stay participants in the Institute's research programmes. The Institute registered 186 new Junior Members in 2007/08; the current total is 604 as at the end of July 2008.

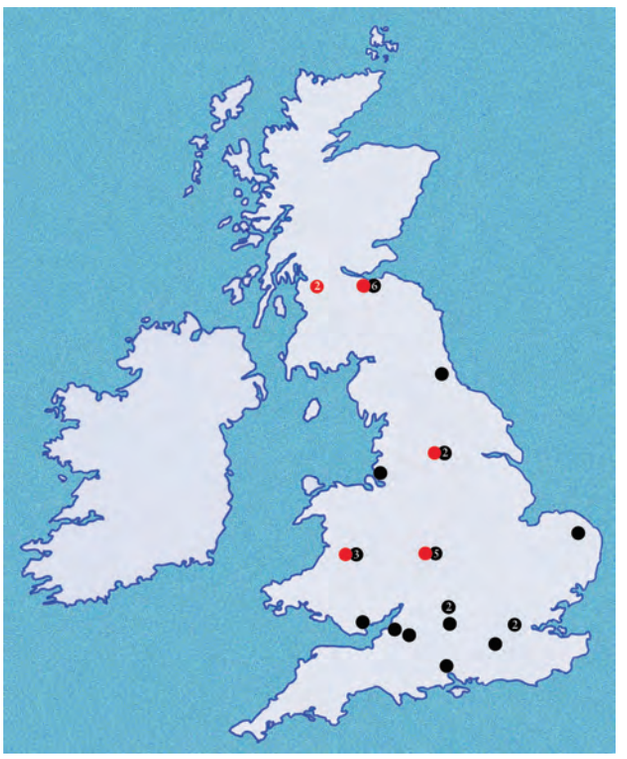
Satellite Workshops

The Institute encourages organisers of 4- or 6-month programmes to cooperate with local organisers in holding 'satellite' workshops at UK Universities and institutions outside Cambridge. Satellite workshops are on themes related to an Institute programme, and involve a significant number of longer-stay overseas participants from the Institute. They also, crucially, draw in and involve UK mathematicians and scientists who might not otherwise be able to participate substantially in the Institute programme; and they enable the expertise of the Institute's overseas participants to be shared more widely.

Satellite workshops have been held at Bath, Bristol, Brunel, Cardiff, Durham, East Anglia, Edinburgh, ICMS, Leeds, Liverpool, London, the National e-Science Centre in Edinburgh, Oxford, Reading, Southampton, Surrey, Wales (Gregynog Hall) and Warwick. Future workshops are being planned at Edinburgh Royal Observatory, Glasgow, Leeds and Warwick.

Costs for satellite workshops are typically approximately £15,000 (excluding the overseas travel costs of Institute participants, which are covered separately) and are shared between the Institute and the host institution. From mid-2008 the host institution's share has been provided by EPSRC.

The Institute is keen to continue to expand the geographical range of satellite workshop locations. Institutions interested in holding a workshop should contact either the organisers of the relevant programme or the Deputy Director, Dr Ben Mestel (B.Mestel@newton.ac.uk).



Distribution of Satellite Workshops

● *Planned* ● *Held*

Follow-up Events

As stated in the Institute's Scientific Policy Statement it is intended that each Institute programme will have long-term impact well beyond the programme itself in terms of breakthroughs, new research directions and collaborations. The Institute has therefore become proactive in arranging short follow-up events some years after programmes finish, whenever the original organisers are enthusiastic.

The Logic and Algorithms programme (January to July 2006) held a follow-up event as a one-week satellite workshop at the International Centre for Mathematical Sciences in Edinburgh, 21–25 July 2008. There were a total of 90 participants, of which 12 were funded by the Institute. Futher details can be found at

www.icms.org.uk/workshops/logicalg

Newton Institute Correspondents (as at 31 July 2008)

Newton Institute Correspondents act as a channel of communication between the Institute and the UK mathematical sciences community (see page 18). Further nominations are encouraged.

Aberdeen	A Sevastyanov	Manchester (Mathematics)	M Prest
Bath	JF Toland	Manchester (Physics)	AJ McKane
Birmingham	IV Lerner	Napier	T Muneer
Brighton	SW Ellacott	Newcastle	J Stoyanov
Bristol	F Mezzadri	Nottingham	Y Mao
Brunel	J Kaplunov	Open	B Mestel
Cambridge	N Dorey	Oxford	P Chruściel
City	O Kerr	Plymouth	C Christopher
Dundee	G Hornig	Portsmouth	AD Burbanks
Durham	WJ Zakrzewski	Queen Mary, London	PJ Cameron
East Anglia	G Everest	Queen's	AW Wickstead
Edinburgh	A Olde Daalhuis	Reading	EA Hanert
Essex	DB Penman	Royal Holloway, London	CS Elsholtz
Exeter	D Stephenson	St Andrews	D Dritschel
Glasgow	C Athorne	Salford	RD Baker
Greenwich	T Mann	Sheffield	K Mackenzie
Heriot-Watt	S Foss	Southampton	CJ Howls
Hertfordshire	S Kane	Staffordshire	BL Burrows
Huddersfield	A Crampton	Stirling	R Norman
Hull	JW Elliott	Strathclyde	M Ainsworth
Imperial College London	M Plenio	Surrey	PE Hydon
Keele	JJ Healey	Sussex	M Hintermüller
Kent	PA Clarkson	Ulster	M McCartney
King's College London	A Recknagel	University College London	ER Johnson
Lancaster	S Power	West of England	K Henderson
Leeds	A Pillay	Wales (Aberystwyth)	R Douglas
Leeds Metropolitan	E Guest	Wales (Cardiff)	KM Schmidt
Leicester	F Neumann	Wales (Swansea)	T Brzeziński
Liverpool	V Goryunov	Wales (WIMCS)	T Lyons
Liverpool John Moores	PJG Lisboa	Warwick	M Reid
Loughborough	AP Veselov	York	N MacKay
LSE	M Luczak		

ATM	P Andrews	Met Office	MJP Cullen
British Computer Society	M Rodd	Microsoft Research Group	CM Bishop
Edinburgh Mathematical Society	TH Lenagan	OR Society	R Hibbs
		Proudman Oceanographic Laboratory	PJM Huthnance
EPSRC	K Bowes		
ICMS	JF Toland	RAL	I Duff
IMA (Academic)	D Abrahams	Rothamsted Research	M Semenov
IMA (Organisational)	D Youdon	Royal Academy of Engineering	J McWhirter
Institute of Actuaries	M Lyons	Royal Society	M Taylor
Institute of Physics	G Watts	Royal Statistical Society	P Gentry
LMS	P Cooper	Schlumberger	J Sherwood
Mathematical Association	RH Barbour	Smith Institute	T Armour

Meeting EPSRC Challenges

Following referees' reports and a panel visit, EPSRC awarded £9.6 million over six years from March 2008. Given the exceptional nature of the award, EPSRC identified several targets, in addition to normal terms and conditions, and a rigorous review of the Institute will be held in 2011. These are welcome challenges, which should help the Institute to meet the interests of all its stakeholders. Below we give a brief progress report on each target.

(1) *Report on progress towards achieving the objectives of the grant.* Management Committee (MC) receives regular updates, covering the objectives in the grant proposal.

(2) *A long-term business plan, with inclusive financial projections.* The first task has been to develop robust five-year financial projections which incorporate targets for fundraising from all sources, and which will be presented regularly to MC and to the School of Physical Sciences.

(3) *A ten-year plan, including a ten-year vision.* This will build on point 2, and involve discussion of scenarios with stakeholders, taking into account the widening opportunities provided by upgrades to our audio-visual and communication capacity.

(4) *A full review of staffing requirements and efficiency.* The Institute keeps its organisational efficiency under review; MC is regularly updated: see comments in the MC Chair's introduction.

(5) *Clarification of the role of the National Advisory Board.* Following extensive discussion and consultation, the NAB is disbanded, with focus now on Correspondents and external representation on the MC.

(6) *Details on how the under-representation of key groups such as women and industrial participants in the Institute's activities has been addressed.* Good dialogue has been established with the LMS Women in Mathematics Committee, with the aim of putting in place a practical set of actions supported by additional funding to be sought from trusts and individuals. See below for details of industrial participation.

(7) *Details on how the Institute has extended its international benchmarking and sharing of best practice with similar organisations overseas.* Detailed comparison is difficult, since activities, local academic and administrative support and operating policies vary enormously. A pattern of comparability is being developed through visits by the Director to ICMS (Edinburgh), IHES (Bures), IHP (Paris), Fields Institute (Toronto), KITP (Santa Barbara) and MSRI (Berkeley).

Open for Business

From its inception, the Institute has engaged with business and industry, and indeed for ten years a most generous arrangement with Hewlett-Packard Laboratories exemplified this. Sustaining this kind of arrangement was always going to be difficult, however, given the Institute's commitment to long-term research carried out in programmes lasting from one to six months, the reality of our programme mix, and increasing business expectations on shareholder value. Engagement has been given renewed attention over the past year, with a focus on three areas:

(1) **Open for Business.** The Institute has initiated a series of half-day meetings in collaboration with the Knowledge Transfer Network for Industrial Mathematics. The programme involves one or two high-level talks, a panel discussion, and a reception and dinner. Participants, who have come from the aerospace, finance and pharmaceuticals sectors, have been enthusiastic about the format, greatly appreciating the opportunity to benefit from the informal and stimulating ambience of the Institute. We hope to hold at least two per year, based around both current and past programmes.

(2) **Opening up to the City.** The incoming Chair of the MC, Howard Covington, has initiated a series of informal dinners which give senior people in the City the opportunity to engage with leading researchers. These have generated great interest, and are creating an influential network of highly placed individuals, who value fundamental science, many of whom will become important advocates for it.

(3) **Visits of up to two days.** In a new initiative, the Institute encourages visits of up to two days from any researcher attached to an academic institution or R&D group in industry or commerce. No prior approval is required. Further details are found at

www.newton.ac.uk/shortvisits.html

Management and Statistical Reports


L Salje

Ekhard Salje, Chair of the Management Committee

The Management Committee oversees all operational and financial aspects of the Institute, and its membership reflects the relevant stakeholders: EPSRC, as a major funder; the University of Cambridge, as employer and underwriter of any financial loss; and the national community, as beneficiary. The new arrangements will greatly strengthen the position of the latter on the committee.

This section provides various statistical data on participation. On the previous page is an update on how the Institute is addressing the additional requirements of the EPSRC grant, particularly steps to increase the engagement of end-users from business and industry.

Institute staffing has been an active area of discussion, recognising the need to balance efficiency with excellent service to participants. On the resignation of the Deputy Director, the committee reviewed this role, and agreed that it was essential. Staff numbers have been reduced by two (part-time) to twelve, as well as Director, Deputy Director (both part-time) and Administrator. The Director's visits to other institutes provide useful benchmarks on staffing and operational issues.

The surplus for 2006/07 was the result of a one-off additional grant of £185k from the University of Cambridge. The full economic costings of the EPSRC award should be a great help for the future. On behalf of all who benefit, I thank all sponsors, and particularly those who have donated personally.

It has been gratifying to watch the Institute's progress over the two years in which I have been Chair. I congratulate all the staff on this success, and warmly welcome my successor, Howard Covington.

Howard Covington
Incoming Chair of the Management Committee

Management Committee

Membership of the Management Committee at 31 July 2008 was as follows:

Sir John Ball FRS FRSE	Chair of the Scientific Steering Committee
Professor JW Bruce	London Mathematical Society
Professor N Dorey	Faculty of Mathematics
Professor WJ Fitzgerald	Council of the School of Technology
Professor WT Gowers FRS	Trinity College
Mr D Harman	EPSRC
Professor PH Haynes	Head of Department, DAMTP
Dr RE Hunt (Secretary)	Deputy Director, Newton Institute
Professor JME Hyland	Head of Department, DPMMS
Professor PT Johnstone	St John's College
Professor PB Littlewood FRS	Council of the School of the Physical Sciences
Professor EKH Salje FRS (Chair)	General Board
Sir David Wallace CBE FRS FREng	Director, Newton Institute

The Management Committee is responsible for overall control of the budget of the Institute and for its financial planning. The Director is responsible to the Management Committee, which provides essential advice and support in relation to fund-raising activities, employment of the staff of the Institute, appointment of the organisers of programmes and general oversight of Institute activities. Its aim is to facilitate to the fullest possible extent the smooth and effective running of the Institute's programmes and all related activities.

Programme Participation

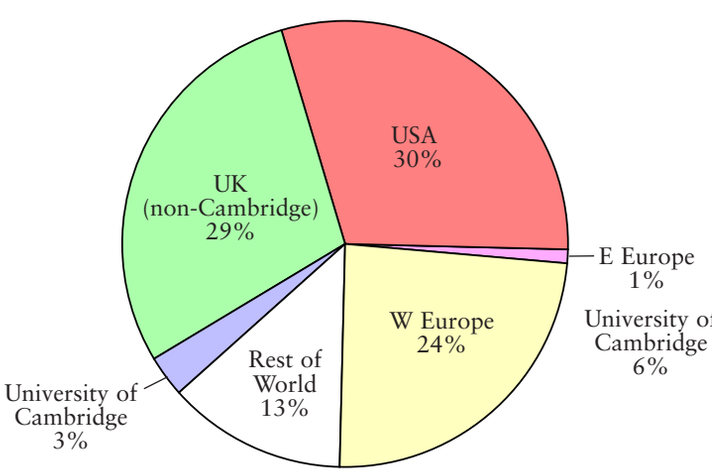
A total of 1566 visitors was recorded for 2007/08.

This includes 339 Visiting Fellows and 83 Programme Participants. These new categories for visitors replace the 'long-term' and 'short-term' categories of previous years which were determined simply by the duration of the visit; a Visiting Fellowship instead reflects *total* contribution to the programme in terms of international reputation and importance in the field as well as the length of stay. Within the five programmes during the year there were 22 workshops (periods of intense activity on specialised topics) which attracted a further 638 visitors (i.e., those not already attending the programme). In addition to workshops, which serve to widen UK participation in programmes, the Institute from time to time arranges less formal special academic meetings as well as talks for the general public, so further opening up the activities of the Institute. More than 500 visitors attended such events and took part informally in Institute activities, including satellite events and follow-up meetings.

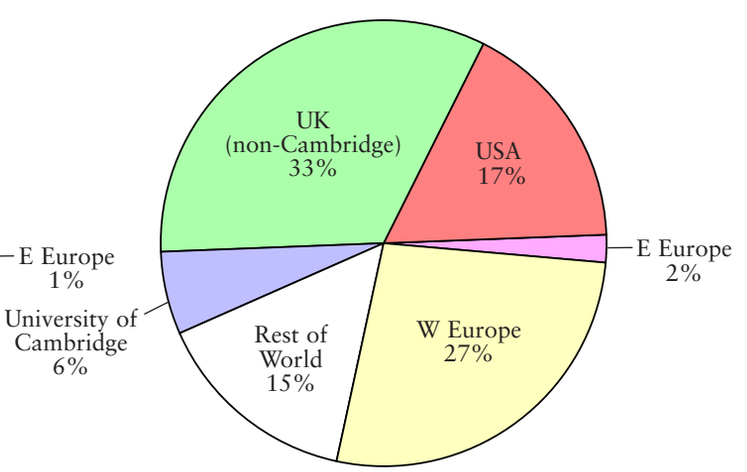
Within all the programmes, workshops and other activities, 707 seminars were given in total at the Institute during the year. The Institute also funds visits by overseas programme participants to other UK institutions to give seminars (see page 18), and 175 such seminars took place last year.

Programme	Visiting Fellows	Mean stay (days)	Programme Participants	Mean stay (days)	Workshop Participants
Strong Fields, Integrability and Strings	110	25	21	19	121
Bayesian Nonparametric Regression	24	19	1	13	62
Phylogenetics	57	35	9	7	87
Statistical Theory and Methods for Complex, High-Dimensional Data	88	43	14	48	238
Combinatorics and Statistical Mechanics	60	56	38	31	118
Follow-up Events	–	–	–	–	12
Totals	**339**	**36**	**83**	**28**	**638**

The pie charts below show the percentages of Visiting Fellows, Programme Participants and Workshop Participants broken down by country of residence:

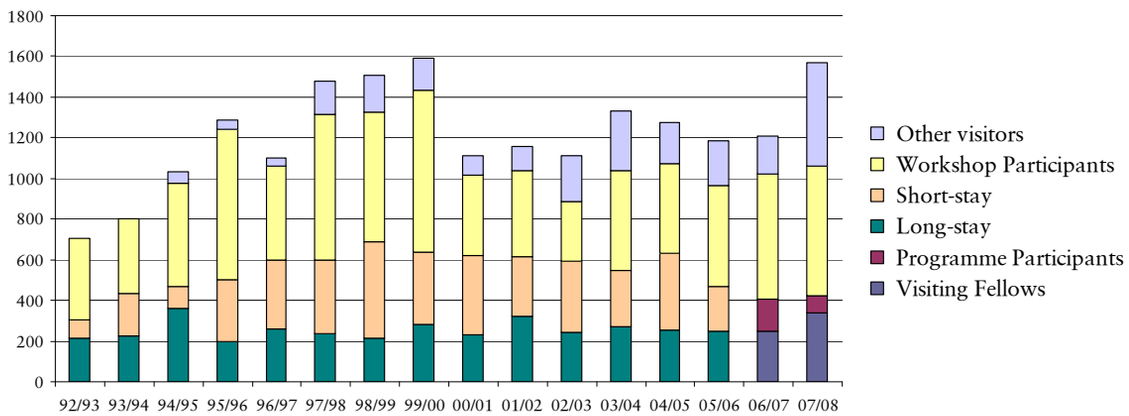


Visiting Fellows and Programme Participants

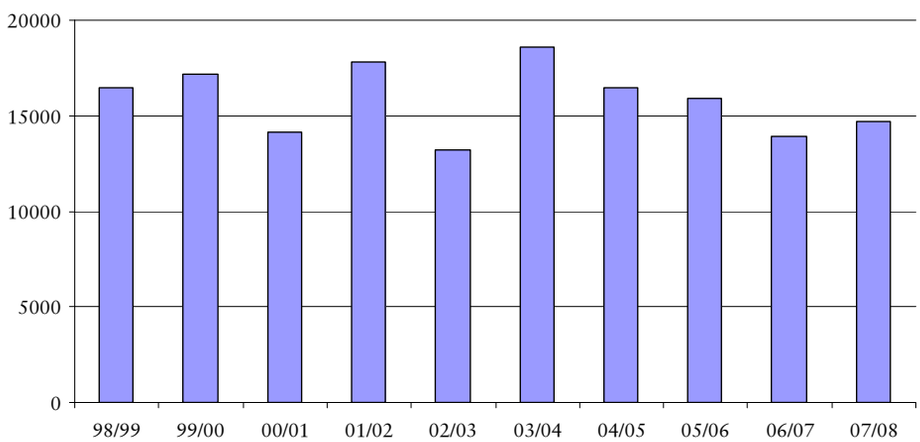


Workshop Participants

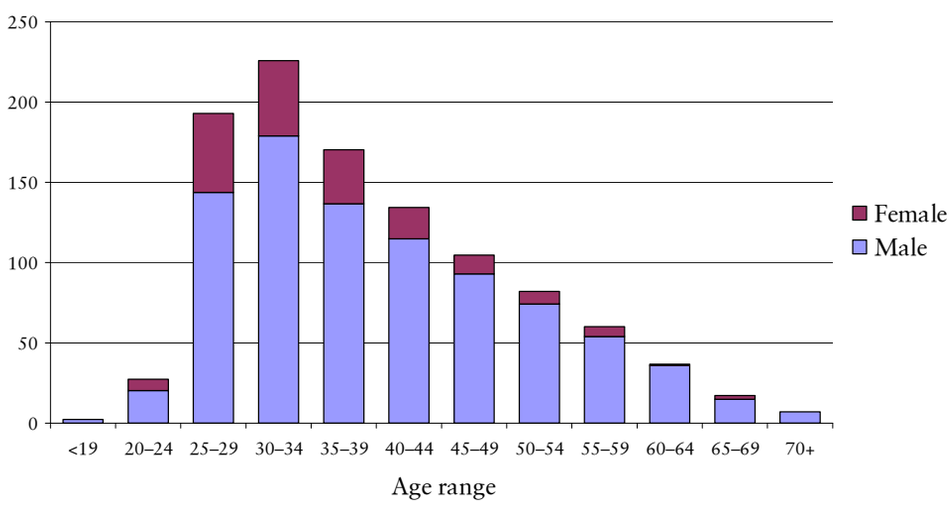
The following chart summarises the total participation figures since the Institute began:



The chart below summarises the total number of person-days for Visiting Fellows and Programme Participants combined (or long-stay and short-stay combined prior to 2006/07; see page 23), including follow-up events but *excluding* Workshop Participants.

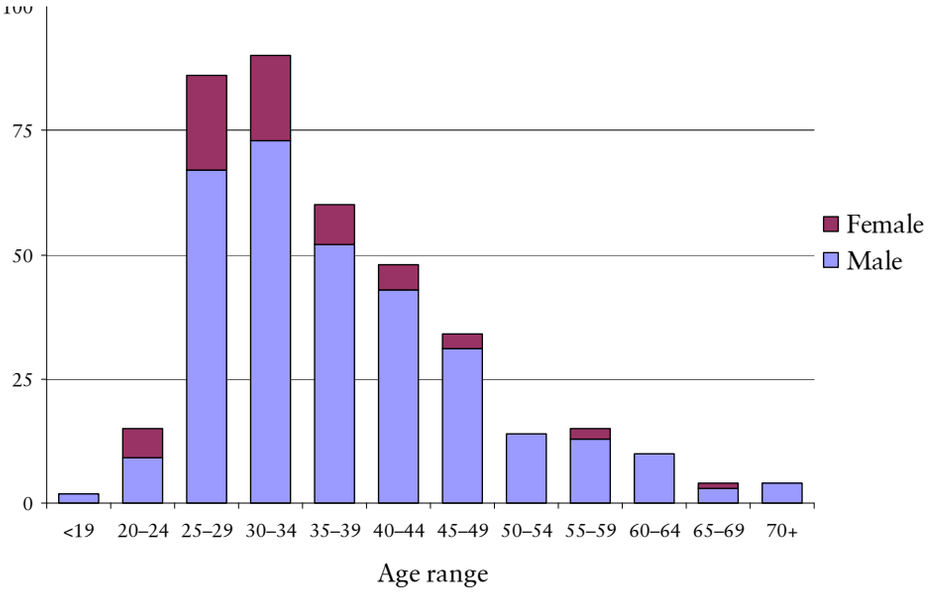


The numbers of all Visiting Fellows, Programme Participants *and* Workshop Participants combined in 2007/08 are shown below, by age and gender:

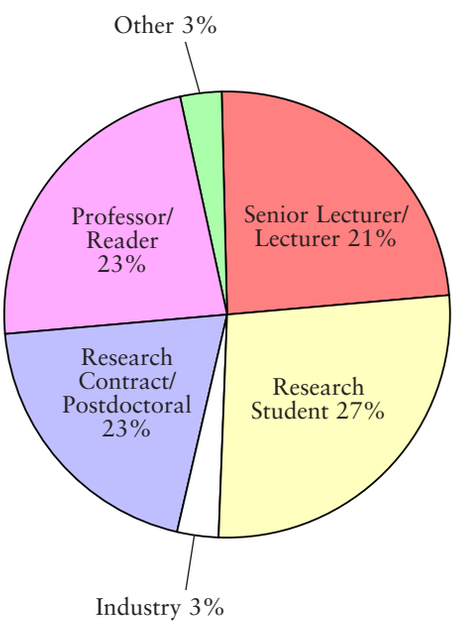


The statistics presented on this page relate only to visitors whose home institutions are in the UK: overseas visitors have been excluded.

The age range and gender balance of all Visiting Fellows, Programme Participants and Workshop Participants from UK institutions in 2007/08 are illustrated below:



The following diagrams indicate the academic status and geographical distribution of all Visiting Fellows, Programme Participants and Workshop Participants from UK institutions during 2007/08:



More detailed statistics, including visit dates, home institutions, seminars given and papers written are shown in the Appendices, available at

www.newton.ac.uk/reports/0708/appendices.html

Finances

Accounts for August 2007 to July 2008 (Institute Year 16)

	2006/07 Year 15 £'000	2007/08 Year 16 £'000
Income		
Grant Income – Revenue[1]	914	1,113
Grant Income – Workshop[2]	214	135
Grant from the University of Cambridge	233	107
NM Rothschild and Sons Trust Funds[3]	109	117
Investment Income	130	141
Donations, Reimbursements and Other Income[4]	47	10
Total Income	**1,648**	**1,623**
Expenditure		
Scientific Salaries[5]	353	413
Scientific Travel and Subsistence	362	432
Scientific Workshop Expenditure	202	216
Other Scientific Costs[6]	18	18
Staff Costs	388	426
Net Housing Costs[7]	79	17
Computing Costs	38	80
Library Costs	12	9
Building – Repair and Maintenance	10	10
Estates and Indirect Costs[8]	97	56
Consumables	20	18
Equipment – Capital	2	4
Equipment – Repair and Maintenance	8	3
Publicity	3	2
Recruitment Costs	2	2
Total Expenditure	**1,594**	**1,706**
Surplus / (Deficit)[9]	**55**	**(83)**

Notes to the Accounts

1. Grant Income – Revenue. The income breaks down as follows:

	2006/07 Year 15 £'000	2007/08 Year 16 £'000
EPSRC/PPARC Salaries	466	482
EPSRC/PPARC Travel and Subsistence	360	449
EPSRC Workshop income	0	60
EPSRC other costs	0	11
PPARC/STFC	0	11
Microsoft Research Cambridge	0	33
Leverhulme Trust	0	21
PF Charitable Trust	37	20
London Mathematical Society	25	25
Centre National de la Recherche Scientifique	24	0
Cambridge Philosophical Society	2	2
Total	**914**	**1,113**

2. Grant Income – Workshop. Included in the EPSRC grant is a specific award towards workshops (see note 1 above). The figure given here is in addition to EPSRC workshop income.

3. NM Rothschild and Sons Trust Funds. The amounts received break down as follows:

Rothschild Visiting Professorships (drawdown)	25	27
Rothschild Mathematical Sciences (income)	84	90
Total	**109**	**117**

The income from the Rothschild Mathematical Sciences Fund supports the Professorship held by the Director of the Institute.

4. Donations, Reimbursements and Other Income. The figure for 2007/08 includes additional programme sponsorship received from the European Molecular Biology Organisation, AWE, Microsoft Research Cambridge and the PASCAL EU network. Additional money from the National Science Foundation (USA) was paid directly to participants in two Institute programmes and is not shown here.

5. Scientific Salaries. This includes stipends paid to Microsoft Research and EPSRC Fellows, Rothschild Visiting Professors, the Director and the Deputy Director.

6. Other Scientific Costs. This includes costs relating to meetings of the Institute's committees, Institute Correspondents' expenses, programme organisers' expenses and entertainment, as well as the travel expenses of overseas participants for their visits to other UK institutions to give seminars (see page 18).

7. Net Housing Costs. These figures include the salary of the Housing Officer, and break down as follows:

Income	294	358
Expenditure	372	375
Total	**(79)**	**(17)**

The costs for 2006/07 include £15k spent on redecorating flats in Benians Court.

8. Estates and Indirect Costs. The figure for 2007/08 results from a new calculation carried out under the full Economic Costs (FEC) regime, which involves different elements than in previous years. The figure is therefore not comparable to previous years.

9. Surplus/(Deficit). The deficit recorded for 2007/08 must be added to accumulated past deficits, amounting at the end of 2006/07 to £361k.

Grants and Donations August 2007 to date

In addition to its substantial funding from the Engineering and Physical Sciences Research Council, the Institute is indebted for continuing funding to the Cambridge Philosophical Society, the Centre National de la Recherche Scientifique, the London Mathematical Society, PF Charitable Trust, NM Rothschild and Sons, and the University of Cambridge. We are very grateful to the following organisations for their specific support: the Alan Wilson Centre of Molecular Ecology and Evolution (New Zealand), the Boston Foundation, the PASCAL European Network, the European Molecular Biology Organisation, the Institute of Physics, Microsoft Research Cambridge, Microsoft Research (USA), the National Institute of General Medical Sciences (USA), the National Science Foundation (USA), the Royal Commission for the Exhibition of 1851, and the Science and Technology Facilities Council (previously the Particle Physics and Astronomy Research Council). Individuals also gave generously in support of our activities: Iain Bratchie, Howard and Veronika Covington, Professors David and Miriam Donoho, Professor Roy and Mrs Ann Garstang, Dr Jonathan Hodgson, David Malcolm and David Wallace.

Cumulative Grants and Donations above £25,000

SERC/EPSRC/PPARC	£20,190k	over 22 years (to 2014)
Trinity College (Isaac Newton Trust)	£2,610k	over 14 years
NM Rothschild and Sons	£2,083k	over 10 years
University of Cambridge	£1,502k	over 16 years
European Union	£1,415k	over 16 years
Leverhulme Trust	£1,109k	over 19 years (to 2011)
Anonymous Donation	£1065k	
Hewlett-Packard	£1065k	over 10 years
Dill Faulkes Foundation	£1000k	
St John's College	£750k	over 5 years
NATO	£728k	over 10 years
Centre National de la Recherche Scientifique	£459k	over 15 years
London Mathematical Society	£358k	over 15 years
Rosenbaum Foundation	£330k	over 7 years
PF Charitable Trust	£240k	over 3 years
Clay Mathematics Institute	£160k	
Gonville and Caius College	£100k	
Prudential Corporation plc	£100k	over 4 years
Microsoft Corporation/ Microsoft Research	£91k	over 3 years
National Science Foundation	£74k	
Institute of Physics	£69k	over 14 years
Wellcome Trust	£65k	
Meteorological Office	£64k	
Nuffield Foundation	£57k	
Howard and Veronika Covington	£51k	
TSUNAMI	£40k	
John Templeton Foundation	£37k	
Daiwa Anglo-Japanese Foundation	£36k	over 4 years
BNP Paribas	£35k	
Cambridge Philosophical Society	£35k	over 20 years (to 2013)
American Friends (Hamish Maxwell)	£32k	
American Friends (Anonymous Donation)	£32k	
Office of Naval Research	£31k	
Emmanuel College	£30k	
Jesus College	£30k	over 6 years

The Whitworth Art Gallery University of Manchester

In the library of the Newton Institute there is a maquette sculpted by Sir Eduardo Paolozzi (1924–2005), which depicts Newton in the pose made famous in the etching by William Blake. The maquette is one of several created by Paolozzi as preliminary studies for the large bronze statue which stands in the courtyard at the British Library in St Pancras, London.

In early 2008, our maquette was exhibited in the Whitworth Art Gallery at Manchester University, and it is now appearing in the Museum of Art at Seoul National University, South Korea (below).

H Birchall

Isaac Newton Institute for Mathematical Sciences

Director: Sir David Wallace CBE FRS FREng

20 Clarkson Road, Cambridge, CB3 0EH, UK

Tel: +44 (0) 1223 335999
Fax: +44 (0) 1223 330508

info@newton.ac.uk

http://www.newton.ac.uk/

Front cover: Planktonic foraminifera (marine microplankton whose shells are preserved in deep-sea sediments). The continuous and complete fossil record of marine microplankton over the last 70 million years allows direct reconstruction of phylogenetic relationships by the method of stratophenetic tracing. This was one of the areas studied during the *Phylogenetics* programme. Image courtesy of M Kucera.

Editors: Dr Robert Hunt, Dr Ben Mestel
Design and production: Stephen Penton, Sara Wilkinson