ISS2011

LIGHT & MATTER

36TH PROFESSOR HARRY MESSEL INTERNATIONAL SCIENCE SCHOOL

THE UNIVERSITY OF SYDNEY

THE LECTURE SERIES OF THE

36TH PROFESSOR HARRY MESSEL INTERNATIONAL SCIENCE SCHOOL

LIGHT AND MATTER

3–16 JULY 2011

THE SCIENCE FOUNDATION FOR PHYSICS
WITHIN THE UNIVERSITY OF SYDNEY

"IN THE PURSUIT OF EXCELLENCE"

THE UNIVERSITY OF SYDNEY

ISS2011

LIGHT & MATTER

EDITOR
Dr Chris Stewart, Manager, International Science School
School of Physics
The University of Sydney, Australia

A course of lectures given at the
36th Professor Harry Messel International Science School
for High School Students
3–16 July 2011

The Science Foundation for Physics
School of Physics A28
The University of Sydney NSW 2006
Australia
www.physics.usyd.edu.au/foundation

Design and print production: Peter Thorn
Printed using Forestry Stewardship Council approved paper

Light and Matter
ISBN: 978-1-74210-236-8

THE UNIVERSITY OF SYDNEY

CONTENTS

THE MESSEL ENDOWMENT

To ensure the continuation of the Professor Harry Messel International Science School the Science Foundation for Physics established The Messel Endowment in 1999.

While the Endowment holds over $3.69million in funds the Foundation seeks to raise a further $3.31million to reach a total goal of $7million to ensure the ISS can be continue in perpetuity allowing for rising costs over the years. The Foundation seeks to do this via donations, grants and bequests.

The Messel Endowment is open to accept donations at any time and currently has over 200 supporters — all of whom the Foundation is highly appreciative and values greatly. Donations of $2.00 and over are tax-deductible. Pledged gifts i.e. donations spread over a three to five year period are also accepted and are tax deductible.

The ISS now has over 4,000 alumni with many going onto outstanding achievements in their chosen fields including science, medicine, engineering and technology. The ISS honours excellence in our high-achieving youth and encourages them to reach their full potential and pursue careers in science and its related areas.

A donation to the Messel Endowment is really an investment in the future of these young scientists. A donation form can be found at the back of this book and at www.physics.usyd.edu.au/foundation

the MESSEL ENDOWMENT
Science Foundation for Physics.

MESSEL ENDOWMENT DONATION HONOUR BOARD

The Science Foundation for Physics established the Messel Endowment in 1999 to honour Emeritus Professor Harry Messel AC CBE and to ensure the Professor Harry Messel International Science School, established in 1962 continues in perpetuity. The Endowment is managed to preserve the real value of all donations received. To date approximately $3.69 million has been raised. A list of donors who have donated $10,000 or more appears below. All donors to the Endowment are listed on the Messel Endowment Honour Banner in the Slade Lecture Theatre, School of Physics and can be viewed online at: www.physics.usyd.edu.au/foundation/donations/MEhonour.shtml

The Science Foundation for Physics appreciates and thanks all our Messel Endowment donors.

EXTRA GALACTIC DONORS

$1 MILLION AND OVER

Australian Government
Department of Innovation, Industry, Science and Research

Mulpha Australia Limited

Australian Government through the Department of Innovation, Industry, Science and Research (DIISR)
Mulpha Australia
Mr Ming Tee Lee through Mulpha Australia

GALACTIC DONORS

$100,000 TO $999,999

Dr John Graham
Hermon Slade Foundation
Nell & Hermon Slade Trust
Science Foundation for Physics
Mr Albert YL Wong & Mrs Sophie Wong

STELLAR DONORS

$10,000 TO $99,999

Anonymous
ANZ Banking Group
Mr Terrey Arcus AM
Mr Robert Arnott
Australian Business Limited
Emeritus Professor Maxwell H Brennan, AO & Mrs Ionie M Brennan
Dr Gregory Clark
Cochlear Limited
Emeritus Professor Richard Collins & Mrs Marilyn Collins
Mr Trevor Danos
Cecil & Ida Green Foundation
James Hardie Industries Pty Ltd
Associate Professor Robert Hewitt & Mrs Helen Hewitt
Mr John Hooke CBE
IBM Australia Limited
James Hardie Industries Pty Ltd
Dr Peter Jones AM
Macquarie Group Foundation
Emeritus Professor Harry Messel AC CBE
Mr Michael Messel
Mr Jim O'Connor
OneSteel Limited
Queensland Cyber Infrastructure Foundation Limited
Mr John Slade
Mr Paul Slade
The James N Kirby Foundation Pty Limited
USA Foundation Donors
Vonwiller Foundation
Westpac Banking Corporation
Mr Thomas Yim

SUPPORTERS OF ISS2011

THE SCIENCE FOUNDATION FOR PHYSICS WARMLY THANKS ALL DONORS, SUPPORTERS AND VOLUNTEERS OF THE 36TH PROFESSOR HARRY MESSEL INTERNATIONAL SCIENCE SCHOOL: LIGHT AND MATTER

The Messel Endowment
Department of Education and Training, NSW Government (DET)
Department of Innovation, Industry, Science and Research, Australian Government (DIISR)
Adolph Basser Trust
The James N Kirby Foundation
Faculty of Science, The University of Sydney
Mr Ron Enestrom
Associate Professor Robert Hewitt and Mrs Helen Hewitt
Dr Peter Jones AM
And all donors to the Foundation's Sydney Development Fund ISS2011 Appeal

The ISS2011 Volunteer Staffies:
Ms Charlotte Fletcher, ISS2009
Mr Matthew Georgiades, ISS2009
Mr Matthew Hourigan, ISS2007
Ms Pamela Karas, ISS2009
Mr Victor Lu, ISS2009
Mr Gabriel Nguyen, ISS2007
Ms Sukey Pan, ISS2007
Ms Clare Paynter, ISS2009
Mr Andrew Perryman, ISS2009
Ms Rebekah Raymond, ISS2009
Mr Jun Tong, ISS2009
Ms Jessica Whiting, UK Scholar - ISS2009

Young Scientists of Australia (YSA)

Australian students were selected in their states and territories with the support of the Science Teachers Associations in:
Australian Capital Territory
Northern Territory
Queensland
South Australia
Tasmania
Victoria
Western Australia
and in New South Wales through the NSW Department of Education and Training

The following institutions assisted in the selection and travel of the overseas students:
The Affiliated High School of Peking University, China
Raman Research Institute, India
Ministry of Education, Culture, Sports, Science and Technology in Japan (MEXT)
Ministry of Education, Malaysia
The Royal Society of New Zealand
Ministry of Education, Singapore (MOE)
Ministry of Education, Thailand
The Association for Science Education, United Kingdom
The Royal Institution of Great Britain
The Ogden Trust, Great Britain
The Comino Foundation, United Kingdom
Department of Energy, United States of America

PREFACE

THE SCHOOL OF PHYSICS AND THE SCIENCE FOUNDATION FOR PHYSICS WITHIN THE UNIVERSITY OF SYDNEY ARE DELIGHTED TO PRESENT ISS2011, THE 36TH PROFESSOR HARRY MESSEL INTERNATIONAL SCIENCE SCHOOL, FROM 3-16 JULY 2011.

The theme for ISS2011, Light and Matter, marks the recent 50th anniversary of the laser, as well as acknowledging the immense contribution made to science by pioneer researchers in the 21st Century.

The primary aim of the ISS is to acknowledge the excellence of the scholars who have been selected on the basis of their academic abilities, passion for science and leadership qualities. The presence of 145 gifted Year 11 and 12 student from all parts of Australia, China, India, Japan, Malaysia, New Zealand, Singapore, Thailand, the UK and the USA creates an environment in which each scholar can experience the values of many cultures and to learn new ways of doing things. The Science Foundation for Physics stands for the Pursuit of Excellence, and is pleased to have an opportunity to honour and reward this spirit in these young people.

The International Science School can only be held because of the generous financial contributions of our supporters to the Messel Endowment and to this year's ISS, and because of the time and energy donated by the lecturers, the many university staff and students who run activities and workshops, and the many volunteers who give their time and energy to the program. Like the Foundation itself, ISS supporters, donors and lecturers are committed to promoting science education at the very highest level of excellence. On behalf of the Foundation and the School of Physics, I express our grateful thanks to all these benefactors.

To our scholars this year, I wish you a most enriching fortnight here at the University of Sydney, and trust that you, like those before you, will enjoy a remarkable and memorable experience, make many life-long friends, and feel empowered to continue to pursue your passion for science.

With very warm wishes,

Professor Clive Baldock
Head, School of Physics
The University of Sydney
June 2011

HISTORY OF THE INTERNATIONAL SCIENCE SCHOOLS

The Professor Harry Messel International Science School (ISS) has a long and distinguished history. The 145 students attending ISS2011: *Light and Matter* are the 36th group to gather in the School of Physics at the University of Sydney for this outstanding free science education program. In all, well over 4,000 scholars have attended an ISS with many remembering fondly their time at the school.

Initially the Schools were annual events, and the first four Schools, held between 1958 and 1961, were for teachers. In 1962 Professor Harry Messel, the founder of the ISS, changed the focus to honour excellence in senior high school students and to encourage them to consider careers in science.

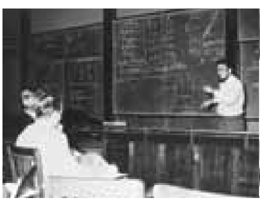
A TRULY INTERNATIONAL SCIENCE SCHOOL

One student from New Zealand attended the very first Science School in 1962, and overseas students have been a feature of the ISS ever since. In 1967, ten students traveled from the USA to attend the School; the following year they were joined by five from the United Kingdom and five from Japan.

South-East Asia joined the ISS in 1985 when students attended from Singapore, Malaysia, Thailand and the Philippines — sadly, that was the only time the Philippines has participated. China has sent five students to every ISS since 1999, except for 2003 when the SARS epidemic restricted travel in the region and they reluctantly withdrew. In 2007 we were unfortunate not to be joined by Malaysia but we did welcome India for the first time.

In 2005 the Indigenous Science Scholars program was introduced, offering five places to indigenous students from across Australia. Since then, the Foundation has worked with the University of Sydney's Koori Centre and other centres and programs in Indigenous education, and we are thankful for their support, encouragement and assistance in ensuring the success of this important program. In 2011, the program has been expanded to include eight Indigenous science scholars.

PROF. HARRY MESSEL



DR GEORGE PORTER

PROF. MARGARET BURBIDGE

PROF. CARL SAGAN

The ISS2011 students attend from ten countries in total: China, India, Japan, Malaysia, New Zealand, Singapore, Thailand, the UK and the USA — and of course, every state and territory of Australia.

GREAT LECTURERS OF THE ISS

One of the features of the International Schools is the lecture series. Past ISS lecturers include James Watson, who won a Nobel Prize for discovering the structure of DNA, and Jerome Friedman, also a Nobel laureate his for work on particle physics. Sir Hermann Bondi (physicist and astronomer at Cambridge University), Margaret Burbidge (astronomer with the Hubble Space Telescope), Carl Sagan (famous astronomer and science broadcaster) and Lord Robert May (President of the Royal Society in the UK) have all given talks at the ISS.

And of course, who could forget the brilliant science demonstrations of Professor Julius "Why is it so?" Sumner Miller, which were such a popular feature of the ISS that they spawned a television show! These days, Dr Karl Kruszelnicki, the University's Julius Sumner Miller Fellow, entertains and enthuses the ISS Scholars with his famous Great Moments in Science.

Between 1960 and 1979 the ISS lectures were shown on television, in fact, many people recall waking up early on Sundays to make sure they didn't miss the telecast! One member of the School of Physics here at the University of Sydney is adamant that the lectures shown on TV were a key part of her decision to become an astronomer.

In 2003 part of the lecture series was broadcast on the Internet as a trial run, and in 2007 the entire series was made available as both video webcast and audio podcast. In 2011, we will once again put all ISS lectures onto the web for the world to share. In addition, this book of the ISS lecture series is available on-line, to encourage a truly international science readership.


PROF. HERMANN BONDI


PROF. LORD WINSTON


2009 SCIENCE CHALLENGE


ISS SCHOLARS 2009

SCIENCE SCHOOL FOR HIGH SCHOOL TEACHERS

Year	Teachers	Theme
1958	123	Selected Lectures in Modern Physics and the Astronomer's Universe
1959	123	Lecture notes on an introductory course in modern physics and nuclear power and radioisotopes
1960	123	From Nucleus to Universe
1961	123	Space and the Atom
TOTAL	**492**	

INTERNATIONAL SCIENCE SCHOOLS FOR HIGH SCHOOL STUDENTS

Year	Boys	Girls	Total	Theme
1962	108	45	153	A Journey through Space and the Atom
1963	104	51	155	The universe of Time and Space
1964	106	53	159	Light and Life in the Universe
1965	114	42	156	Time (and Relativity)
1966	104	52	156	Atoms to Andromeda
1967	101	57	158	Apollo and the Universe
1968	109	20	129	Man in Inner and Outer Space
1969	118	21	139	Nuclear Energy Today and Tomorrow
1970	99	33	132	Pioneering in Outer space
1971	87	35	122	Molecules to Man
1972	95	28	123	Brain Mechanisms and the Control of Behaviour
1973	93	29	122	Focus on the Stars
1974	90	33	123	Solar Energy
1975	76	43	119	Our Earth
1977	54	50	104	Australian Animals and their Environment
1979	63	52	115	Energy for Survival
1981	50	65	115	Biological Manipulation of Life
1983	67	51	118	Science Update 1983
1985	71	59	130	The Study of Populations
1987	70	56	126	Highlights in Science
1989	69	58	127	Today's Science Tomorrow's Technology
1991	61	70	131	Living with the Environment
1993	60	72	132	Carbon: Element of Energy and Life
1995	55	80	135	Breakthrough! Creativity & Progress in Science
1997	72	65	137	Light
1999	73	66	139	Millennium Science
2001	70	71	141	Impact Science
2003	54	85	139	From Zero to Infinity
2005	73	66	139	Waves of the Future
2007	68	65	133	EcoScience
2009	67	71	138	Genes to Galaxies
TOTALS	**2434**	**1573**	**4007**	

ISS2011 LECTURERS

A/PROF STEPHEN BARTLETT

SCHOOL OF PHYSICS
UNIVERSITY OF SYDNEY

A/Professor Stephen Bartlett completed his Ph.D. in mathematical physics at the University of Toronto in 2000. Moving to Australia, he directed his research to the theory of quantum computing, first as a Macquarie University Research Fellow and then as an ARC Postdoctoral Research Fellow at the University of Queensland. Since 2005, he has lead the Quantum Physics research group at University of Sydney, pursuing research in quantum computing, quantum measurement and control, and the foundations of quantum theory.

PROFESSOR CHRISTINE CHARLES

SPACE PLASMA, POWER AND
PRODUCTION LABORATORY, ANU

Professor Christine Charles is Head of the Space Plasma, Power and Propulsion Laboratory at the Australian National University in Canberra. She has a French Engineering degree in applied physics, a Ph D in plasma physics, a French Habilitation thesis in materials science and a Bachelor of Music degree in Jazz. She won a tenure research position with the French National Centre for Scientific Research (CNRS) in Nantes in 1990 and moved to the Australian National University late 1992. Since the early 90s she has been working on experimental expanding plasmas and their applications to electric propulsion, microelectronics, optoelectronics, astrophysical plasmas, and hydrogen fuel cells.

She is the inventor of the Helicon Double Layer Thruster, a new electrode-less magneto-plasma space engine, which applications include satellite station keeping or interplanetary space travel. She has demonstrated the thruster concept both at NASA and ESA (the European Space Agency) and more recently at EADS-ASTRIUM, the company which runs the Airbus 380 and the Ariane Rocket program. She has published over 120 refereed papers and was elected by the Australian Institute of Physics as International Women In Physics Lecturer of the year in 2009. Since 2004, she has been awarded five Visiting Professorships at the University of Orleans (France).

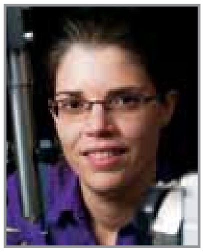
PROFESSOR ALLAN G. CLARK

DIRECTOR, DEPARTMENT OF NUCLEAR AND PARTICLE PHYSICS, UNIVERSITY OF GENEVA

Allan Clark had his first course in particle physics in the 1960's, as an undergraduate at the University of Tasmania. At the same time, particle physicists first suggested that nucleons were not indivisible, but were made up of quarks (and we know now of many other things). This was the start of a new period in the understanding of fundamental particles and their interactions. This sparked his interest and shaped his subsequent research activities.

His research activites have concentrated on the collision of hadrons (mostly protons or anti-protons) at the highest possible energies, using particle accelerators. He has been a member of several experimental teams that have made major contributions to our understanding of particle interactions, for example the UA2 experiment that co-identified the W and Z bosons at the CERN proton-antiproton collider, the CDF experiment that identified the top quark at the Fermilab Tevatron proton-antiproton collider near Chicago, and the ATLAS experiment that is now collecting data at the CERN Large Hadron Collider in Geneva.

In 1989, he was appointed Professor of Physics at the University of Geneva and since 1998 he has served as Director of the Department of Nuclear and Particle Physics. He has over the years served on many advisory panels, including the European Committee for Future Accelerators and the Swiss Institute of Particle Physics where he was the founding Chair.



DR DEANNA D'ALESSANDRO

SCHOOL OF CHEMISTRY
UNIVERSITY OF SYDNEY

Dr Deanna D'Alessandro was awarded a 2010 L'Oreal Australia for Women in Science Fellowship, a 2010 University of Sydney Postdoctoral Research Fellowship, and a 2011-2015 Australian Research Council QEII Fellowship for her work into new materials for capturing carbon emissions.

Deanna's research spans the areas of inorganic chemistry, physical chemistry and materials science and focuses on the development of functional inorganic materials that exhibit novel electronic, optical and magnetic phenomena. Potential applications range from the capture of greenhouse gases by using molecular sponges that can mop up carbon dioxide and release it on cue, to sensors, optoelectronics devices and photocatalysts.

Deanna received a PhD (Chemistry) in 2006 and a Bachelor of Science with Honours (Chemistry) in 2001 from James Cook University. She conducted postdoctoral research at the University of California, Berkeley from 2007-2009.

Photo Credit: Sam D'Agostino, SDP Photos

PROFESSOR MARTIN GREEN

UNIVERSITY OF NEW SOUTH WALES, SYDNEY AUSTRALIA

Martin Green is currently a Federation Fellow and Scientia Professor at the University of New South Wales and Executive Research Director of the ARC Photovoltaic Centre of Excellence. He is also a Director of CSG Solar, a company formed specifically to commercialise the University's thin-film, polycrystalline-silicon-on-glass solar cell. His group's contributions to photovoltaics are well known including the development of the world's highest efficiency silicon solar cells and the successes of several spin-off companies. He is the author of six books on solar cells and numerous papers in the area of semiconductors, microelectronics, optoelectronics and, of course, solar cells. His work has resulted in several major awards including the 1999 Australia Prize, the 2002 Right Livelihood Award (also known as the Alternative Nobel Prize), the 2004 World Technology Award for Energy, the 2007 SolarWorld Einstein Award, the 2009 Zayed Future Energy Prize (one of two finalists) and the 2009 ENI Award for Renewable and Non-conventional Energy.

PROFESSOR BEN EGGLETON

FEDERATION FELLOW, CUDOS, SCHOOL OF PHYSICS, UNIVERSITY OF SYDNEY

Professor Benjamin Eggleton is currently an ARC Federation Fellow and Professor of Physics at the University of Sydney. He is Research Director of the Centre for Ultrahigh-bandwidth Devices for Optical Systems (CUDOS), an ARC Centre of Excellence, and Director of the Institute of Photonics and Optical Science (IPOS) at the University of Sydney.

He studied at the University of Sydney, obtaining his BSc (Hons 1) in 1992 and his PhD in Physics in 1996. After graduation, he went to the United States to join the world's leading research institute in his field, Bell Laboratories, as a Postdoctoral Fellow in the Optical Physics Department. He was subsequently promoted to Research Director of the Specialty Fiber Business Division of Bell Lab's parent company, Lucent Technologies, where he drove Lucent's research program in optical fibre devices.

He has co-authored more than 250 journal papers, presented more than 60 invited and plenary presentations at international conferences, and has filed 35 patents. He has received several significant awards, most notably the NSW Scientist of the year for Physics and Astronomy, the Pawsey Medal from the Australian Academy of Sciences, the Prime Minister's Malcolm McIntosh Science Prize for Physical Scientist of the Year, the ICO Prize (International Commission for Optics), and the Adolph Lomb Medal from the Optical Society of America.

PROFESSOR THOMAS MASCHMEYER

ARC FUTURE FELLOW,
SCHOOL OF CHEMISTRY
UNIVERSITY OF SYDNEY

In 1994, as Australian Bicentennial Fellow, Thomas Maschmeyer went to the Royal Institution, London. Subsequently, he was made Assistant Director of their Davy Faraday Laboratories and also Associate Lecturer and Affiliated Fellow (Peterhouse) at the University of Cambridge. In 1998, at the age of 31, he was appointed Professor and Head of the Department of Industrial Organic Chemistry at the Delft University of Technology, making him the youngest Professor of Chemistry in Europe at that time. Additionally, in 2000, Thomas became Vice-Chairman of the Delft Institute of Chemical Technology. During this time he was also advisor to the Dutch Federal Ministry of Economic Affairs. He returned to his alma mater in late 2003, when he commenced his new positions as Federation Fellow and Professor of Chemistry, and most recently, ARC Future Fellow.

PROFESSOR SIR JOHN PENDRY

IMPERIAL COLLEGE, LONDON

John Pendry has been working at the Blackett Laboratory, Imperial College London since 1981. He began his career in the Cavendish Laboratory at Cambridge University, followed by six years at the Daresbury Laboratory of the Science Research Council, where he headed the theory group. In collaboration with the Marconi Company, he designed a series of completely novel artificial materials, or 'metamaterials', with properties not found in nature. Successively metamaterials with negative electrical permittivity, then with negative magnetic permeability were designed and constructed. This project culminated in the proposal for a 'perfect lens' whose resolution is unlimited by wavelength. He is popularly known for creation of the first practical 'Invisibility Cloak'. At Imperial College he was at various times: head of the physics department, Dean of the Royal College of Science, and Principal of the Faculty of Physical Sciences. The long list of awards he has received includes, among others, his fellowship of the Royal Society (1984), honorary fellowship of Downing College at Cambridge University, the Dirac prize (1996), the EU Descartes prize (2005), the Royal Medal of the Royal Society (2006), the UNESCO-Niels Bohr gold medal (2009), as well as being knighted for his services to science (2004).

PROFESSOR STEPHEN SIMPSON

LAUREATE FELLOW, SCHOOL OF BIOLOGICAL SCIENCE, UNIVERSITY OF SYDNEY

Professor Stephen Simpson is an Australian Research Council Laureate Fellow Within Sydney University's School of Biological Sciences. He returned to Australia in 2005 after 22 years at Oxford, where he was Professor of Entomology and Curator of the University Museum of Natural History.

Stephen has pioneered developments at the interface of nutritional physiology, ecology, and behaviour. With colleagues he developed state-space models for nutrition, devised and tested using insects but applied since to a wide range of animals and problems, from aquaculture and conservation biology to human obesity. His research has led to an understanding of locust swarming that links chemical events in the brains of individual insects to landscape-scale mass migration.

In 2007 he was elected a Fellow of the Australian Academy of Science, in 2008 he was awarded the Eureka Prize for Scientific Research, in 2009 he was named NSW Scientist of the Year, and in 2010 he was named as the Wigglesworth Medallist by the Royal Entomological Society of London.

PROFESSOR FRED WATSON

ASTRONOMER-IN-CHARGE, AUSTRALIAN ASTRONOMICAL OBSERVATORY

Professor Fred Watson says he has spent so many years working in large telescope domes that he has started to look like one. His main scientific interest at the AAO in Coonabarabran is gathering information on very large numbers of stars and galaxies. He is also an adjunct professor at the Queensland University of Technology and the University of Southern Queensland. Fred is the author of "Stargazer — the life and times of the telescope", and is a regular broadcaster on ABC radio. In 2003, he received the David Allen Prize for communicating astronomy to the public, and in 2006 was the winner of the Australian Government Eureka Prize for Promoting Understanding of Science.

DR JOANNE WHITTAKER

SCHOOL OF CHEMISTRY,
UNIVERSITY OF SYDNEY

Joanne Whittaker's research interests are predominantly in plate tectonics, marine geophysics and geodynamics. Jo is currently a postdoctoral fellow in the EarthByte Group, School of Geosciences at the University of Sydney, Australia, working on investigating Indian Ocean plate reconstructions and continental break-up in the context of Gondwanan extension, breakup and continental margin evolution. She received her PhD, on the tectonic consequences of mid-ocean ridge formation, evolution and subduction, from the University of Sydney in 2008. Following completion of her PhD, she worked as a Geodynamicist for Getech in Leeds, UK, before returning to Australia.

ROBYN WILLIAMS AM

SCIENCE BROADCASTER, ABC

Science journalist and broadcaster, Robyn Williams, presents ABC Radio National's *Science Show*, *Ockham's Razor* and *In Conversation*.

Although he graduated with a Bachelor of Science (Honours) in England, Robyn admits to spending as much time acting as studying. Early in his career he made guest appearances in *The Goodies*, *Monty Python's Flying Circus* and *Dr Who*, and stood in for Tom Jones for four months in his TV series.

He has conducted countless interviews with scientists on ABC TV on programs such as *Quantum* and *Catalyst*, narrated the *Nature of Australia* series and appeared in *World Safari* with David Attenborough.

Outside the ABC, Robyn has served in various capacities, including President of the Australian Museum Trust, Chairman of the Commission for the Future, and President of the Australian Science Communicators. In 1987, he was proclaimed a National Living Treasure.

In 1993, Robyn was the first journalist elected as a Fellow Member of the Australian Academy of Science. He was appointed AM in the 1988 Australian Bicentenary honours list and in the same year received Honorary Doctorates in Science from the University of Sydney and Macquarie and Deakin Universities. The ANU awarded him a Doctorate of Law, and he is a Visiting Professor at the University of NSW and an Adjunct Professor at the University of Queensland.

THE LECTURES

A/PROFESSOR STEPHEN BARTLETT

PROFESSOR CHRISTINE CHARLES

PROFESSOR ALLAN G. CLARK

DR KARL KRUSZELNICKI

DR DEANNA D'ALESSANDRO

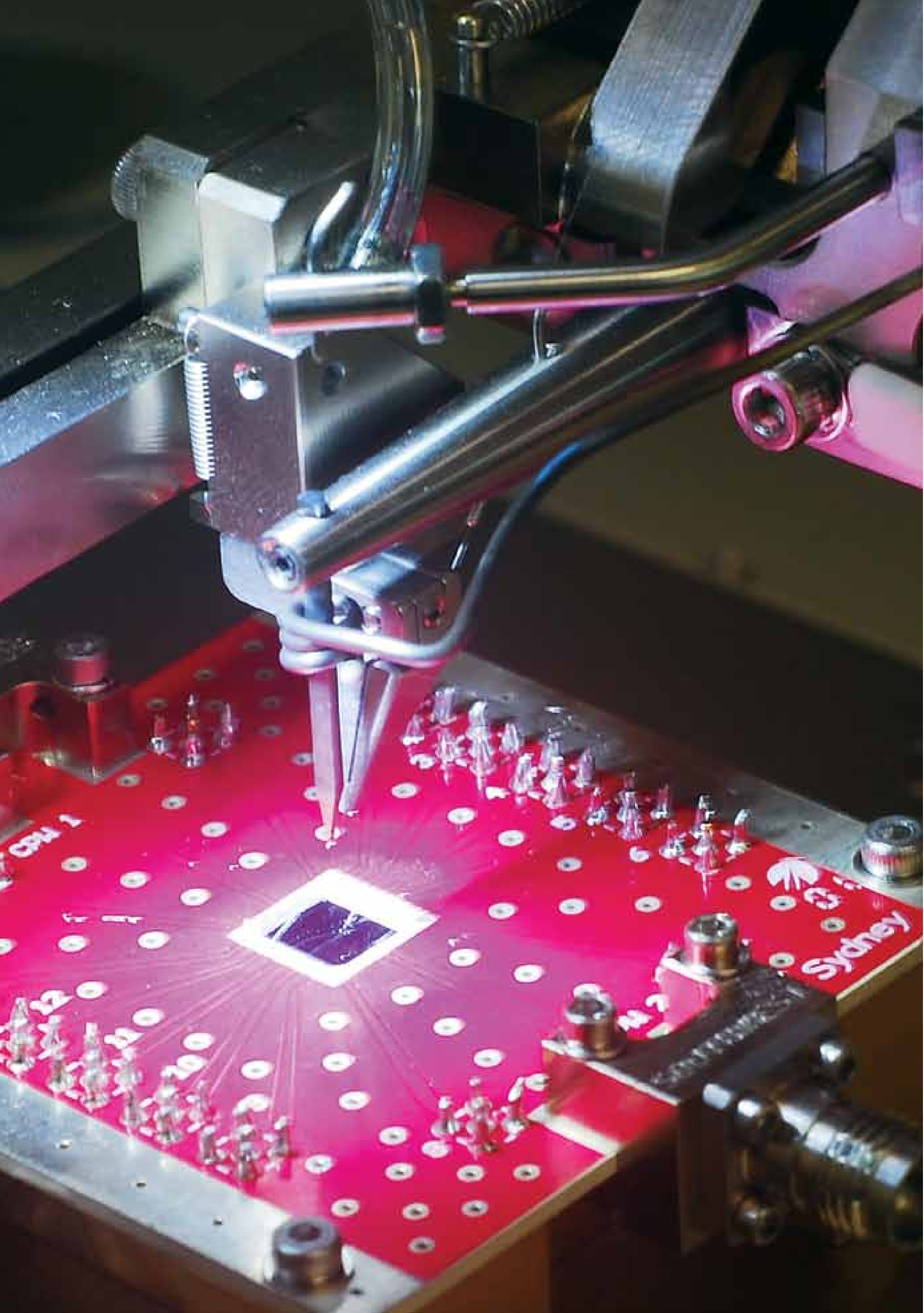
PROFESSOR MARTIN GREEN

PROFESSOR SIR JOHN PENDRY

PROFESSOR STEPHEN SIMPSON

PROFESSOR FRED WATSON

DR JOANNE WHITTAKER


Sydney

SMALLER, FASTER, BETTER, UNIMAGINABLE-ER? THE QUANTUM REVOLUTION IS COMING.

A/PROFESSOR STEPHEN BARTLETT,
WITH MICHAEL J. BIERCUK, ANDREW DOHERTY AND DAVID J. REILLY
QUANTUM PHYSICS GROUP, SCHOOL OF PHYSICS, THE UNIVERSITY OF SYDNEY

INTRODUCTION

In the 20th century a dramatic upheaval began as classical physics was augmented by the theory of quantum mechanics, operational on the scale of individual photons, atoms, and molecules. The results were astounding; with this new framework long-standing problems in physics were solvable, a new understanding of matter at the atomic scale emerged, and as we all know a new class of technologies founded on quantum mechanics arose.

It is not an exaggeration to say that the modern information economy exists primarily because of the discovery of quantum mechanics. Quantized (discrete) energy levels give us the underlying mechanism enabling the laser – used for high-bandwidth optical communications. Quantum statistics give us the semiconductor band structure used in producing transistors – the fundamental hardware elements in a computer. The phenomenon of Giant Magnetoresistance gives us a way to easily read-out magnetic information – enabling high-capacity, small form-factor hard-disk storage. When one examines nearly any component or functionality in a microprocessor, server, or communications network, quantum mechanics is at the heart of that technology.

As amazing as that is, it's only the first chapter of our story. These technologies, and many others, exploit only the most basic principles of quantum mechanics: discrete energy levels, the Pauli exclusion principle, and the like. But inherent in the formalism of quantum mechanics was a host of "quantum coherent" effects that were largely swept under the rug because they were just too strange; for instance, the principle of quantum superposition, in which a particle can exist in two or more different states at once — until, upon measurement, the particle winds up in just one of those allowable states; or quantum entanglement, in which multiple particles can be intricately linked, even when they're separated by huge distances, and any action on one will still affect the other. This is what Einstein famously and pejoratively characterized as "Spooky action at a distance".

These phenomena have long been considered mathematical oddities, "quantum weirdness". But a new field is emerging looking to explicitly leverage quantum coherence for the development of more powerful and fundamentally new technologies.

Thanks to decades of work in atomic, optical, and solid-state physics we are now able to access the full spectrum of quantum

Figure 1. A semiconductor spin-qubit chip is connected to a high-frequency circuit for quantum measurement and control. Quantum systems based on single spins are a key research thrust of the Mesoscopic Physics Laboratory at the University of Sydney.

phenomena in a variety of systems — from individual trapped atoms to optical devices, superconducting circuits, and single electrons in semiconducting materials. The strangeness of quantum mechanics is not only being observed in these systems, but it is being controlled, yielding a new discipline — quantum science, founded on the ability to coherently manipulate quantum systems.

Quantum science has the promise to truly revolutionise modern technology. In the near term we can expect new sensors capable of using quantum entanglement to provide dramatic performance enhancements over classical technology, novel biomedical imaging techniques giving sub-cellular resolution as well as rapid readout, and new precision time standards for improved global positioning systems — devices that are smaller and faster, while providing better performance.

But it's not the near term that's really exciting; what the future will bring is completely open. When the transistor was invented — the basic logic element in a modern computer — it was believed that it could only be useful for … hearing aids!

At the time, the modern digital microprocessor was unimaginable. The same holds true for quantum-enabled technologies. Rather than simply take what we find in Nature, can we engineer the quantum world to our purposes? Can the strangeness of quantum mechanics be used as a new and untapped resource? We just don't know what's down the road, but we can't wait to get there.

FROM QUANTUM MECHANICS TO QUANTUM SCIENCE

Quantum mechanics gradually developed as a mathematical framework for the description of matter on the atomic scale. Although it was cobbled together from the work of a number of independent scientists, essentially all phenomena at the microscopic scale — including the orbitals of electrons around an atom, atom-photon interactions, and the statistics of electrons in solids — have been elegantly described using the quantum formalism. To date, quantum mechanics represents our most accurate and widely applicable scientific theory, predicting technologically relevant phenomena such as semiconductor band structure, lasing, and superconductivity. All told, quantum mechanics has already enabled a quantum revolution, supporting the development of an entire generation of technology that has given us the modern information economy.

However, the quantum mechanics that we encounter daily, in devices like semiconductor electronics or lasers, incorporates only the most basic phenomena. These systems behave "semi-classically", possessing quantum characteristics such as discrete energy levels and quantum statistics, but lacking any signatures of the more interesting — and "spooky" — quantum coherence or phase information. The most exotic behaviours, such as quantum superposition and entanglement, that have so intrigued and confounded physicists are rapidly lost through coupling to a noisy environment, spontaneous emission, and the like. Most importantly, these devices make use of bulk properties, where quantum coherent effects are masked by averaging over large numbers of incoherent quantum systems — that is to say, quantum effects like entanglement between small numbers of coherent atoms, for example, are totally washed out when you look at the very, very large numbers of atoms operating in an electronic device or laser.

The past two decades have seen an explosion in the number of systems that have been able to provide a window from our classical world to the world of quantum coherence. This research has been undertaken in a number of disciplines that have developed largely independently: condensed matter physics, superconductivity and strongly correlated systems, atomic, molecular, and optical physics, and nuclear magnetic resonance, among others. Remarkably, in each of these disparate fields experiments have demonstrated that it is possible to realise and manipulate quantum mechanical degrees of freedom — such as photon polarization, the flow-direction of supercurrents, the location or spin-state of an electron, or the vibrational mode of individual trapped atoms. Typically, all of these phenomena, while described by quantum theory, are lost in the bulk properties of materials.

As such, it has been possible to directly observe, in the laboratory, some of the strangest phenomena that quantum mechanics predicts, including quantum superposition, entanglement, and even teleportation.

These experimental capabilities have led to a renaissance in quantum physics, allowing us to realize in the lab some of the most exotic gedanken (or “thought experiments”) from the early days of quantum theory: wave-particle duality, the Einstein-Podolsky-Rosen (EPR) paradox, and even an experimental version of the famous “Schrödinger’s cat”. Moreover, being able to create and control coherent quantum systems has led to a convergence of previously independent fields: a rediscovery of atomic physics in artificial atoms and other condensed matter systems, the modeling of superconducting circuits using the principles of quantum optics, the development of micro and nanofabricated atom-chip platforms.

More importantly, this recent work has led to the prediction of the imminent arrival of a second quantum revolution, one in which we shift away from implicitly leveraging the quantum-mechanical nature of matter, towards explicitly exploiting and controlling the quantum realm. We are now seeing the early development of a new class of technologies that use the strangest phenomena in the quantum world as resources to enable fundamentally new ends. Early research focused almost exclusively on computation and communications, but it has been realized that the landscape of potential applications is far broader than originally conceived.

Pioneering researchers ushering in this second quantum revolution have historically emerged from the various disciplines described above, but today, Quantum Science has emerged as a worldwide initiative and a substantive discipline of physics. The field now has hundreds of active research faculty across the globe, and literally thousands of young researchers and students. It spans experiment and theory, and focuses on new techniques and applications involving the realization and coherent control of quantum mechanical variables, often from mesoscopic systems — for example, ensembles of atoms behaving in unison, or superconducting currents composed of many electron-pairs.

WHERE IS QUANTUM SCIENCE GOING?

The second quantum revolution appears to be well on its way. The past decades of research have positioned the field such that today there exists a vast array of research areas with the potential to dramatically alter the course of technology. Here we present a few topics that appear poised to make a splash.

HYBRID QUANTUM SYSTEMS

In its primary stages, Quantum Science has focused on the demonstration and control of quantum coherent phenomena in a variety of platforms, often using technology-specific techniques in the process. Frequently, a particular experimental platform has great strengths traded off against intrinsic limitations. For instance, the integration of solid-state quantum devices (quantum dots, for example) provides benefits in scaling and communications, but limitations in coherence time — the effective lifetime of the system’s “quantumness” — due to the strong interaction with “dirty” substrate materials.

Recent technological developments have shown that it is possible to engineer hybrid quantum systems that, in many instances, marry the best features of two or more worlds. Ultimately, quantum systems are likely to employ an array of interacting quantum components, each contributing specific capabilities and producing new behaviour at the system level.

This is something that appears all the time in the macroscopic world around us — for instance, a CD player is a classical example of a complex engineered system that uses mechanics, optics and opto-electronics, precision measurement and feedback control, information processing, and error correction.

Similarly we may imagine complex engineered quantum systems, providing tremendous new capabilities, but requiring coupling between different platforms in a hybrid package. The unification of seemingly unrelated technologies thus holds great promise, and will be generating a lot of attention in the next few years.

QUANTUM CONTROL AND MEASUREMENT FOR ROBUST, LARGE-SCALE QUANTUM SYSTEMS

As contemporary technologies evolved from individual devices to complex systems, it became necessary to move beyond coarse and "manual" control methods. Similarly, these systems must be made to function in the presence of noise and faulty parts. Robust, efficient, and automated control mechanisms are an absolute requirement at the system level, needs that have given rise to the modern field of optimal control theory.

To illustrate the concept of optimal control theory, consider a car traveling on a straight line through a hilly road. How should the driver press the accelerator pedal in order to *minimize* the total traveling time? Here, the "control" rules might refer specifically to the way in which the driver presses the accelerator and shifts the gears. The "system" consists of both the car and the road, and the quantity to be optimized is the minimisation of the total traveling time. Control problems usually include other constraints — for example, the amount of available fuel, the amount you can press the accelerator pedal, speed limits, and so on.

Control theory is an ideal starting point for the creation of a quantum analogue, since the same motivations emerge in the quantum realm: we need ways to control and optimize quantum systems, taking into account quantum constraints and quantum rules.

Methods of quantum control and measurement are at the core of our capacity to construct new technologies from quantum building blocks. A key research area in quantum science focuses on the theoretical and experimental development of protocols providing fast, resource-efficient manipulation and readout of quantum systems. Techniques from modern control theory can be adapted to quantum systems, as long as we take care to properly include the effects of quantum measurement. Such considerations dramatically impact the landscape of accessible control techniques.

A good example is hardware robustness against error. Error-resistance is desirable in any technology, whether in relation to information processing or metrology (the science of measurement). Classically, error correction typically involves storing information several times over (known as redundancy), or coded in different ways — and checking that these different versions agree; if they do not, then the different versions of the data can be used to reconstruct the original. CDs and DVDs have this sort of error correction.

In its simplest version, classical error correction might look like this: say you want to transmit a single bit of information, with the value **1**. To guard against error, you send it three times — **1**,**1**,**1**. But along the way noise messes up one of the copies, and the message received is **1**,**0**,**1**. The receiver can simply look at the message and, since two of the copies are the same, deduce that one error has been made: the message should have been **1**,**1**,**1**. It is less likely that two errors were made in transit, so the receiver safely deduces the bit was a **1** and not a **0**.

In quantum information theory, rather and a **bit** we use a **qubit**: a quantum system that can be in a state corresponding to "**0**", or a state corresponding to "**1**", or — because of quantum superposition — *in both states at once!*

The trouble with taking these error correction ideas to the quantum realm is that they rely on measuring the system as you go along: you look at the message when it is received to figure out what it was supposed to be. In quantum systems that rely on a superposition of states, such measurement would force the system into one of these states: the "wavefunction collapses", and the whole benefit of the quantum effects is gone.

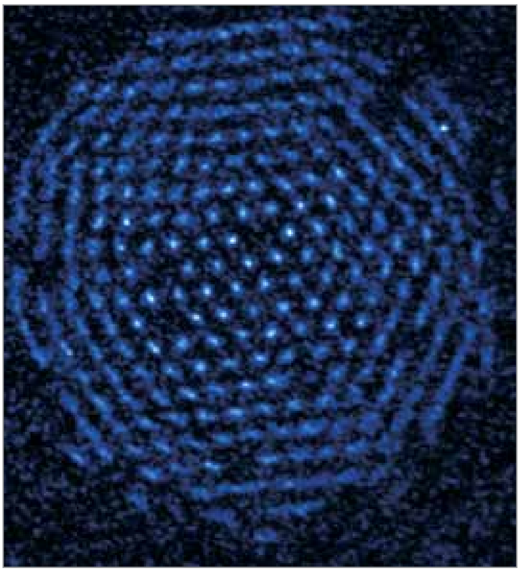


Figure 2. A two-dimensional Coulomb crystal of laser-cooled Beryllium ions. The regular crystal structure, spin-control mechanisms, and techniques for producing tunable coupling make this system an ideal candidate for studies of quantum magnetism and exotic large-scale entangled states. Image courtesy US NIST and the Quantum Control Laboratory, Sydney.

However, there are ways to do error correction without upsetting the quantum state. Analogous to making copies of a bit as shown above, quantum error correction involves spreading out one qubit of information across an entangled system of several qubits. To avoid the measurement issue, algorithms have been developed that can identify the *kind* of error that has occurred without actually *measuring* the qubit itself — this is somewhat like knowing, in the example above, that the noise has flipped the middle copy of information in transmitted message, but not telling you whether it is a **0** or a **1**. Knowing only the error still allows you to adjust the final message to correct it. Quantum Error Correction is a new and important field within the broader area of quantum science.

Beyond developing "useful" procedures, however, there is also a need for the development of control and measurement techniques operating at the absolute limits imposed by quantum mechanics. For example, noise due to quantum statistics — the inherent, fundamental fuzziness of quantum systems — sets limits on the achievable precision of a measurement. However, by manipulating systems into more exotic quantum states and using the approaches of control theory, we can improve the precision of measurements by effectively pushing this quantum noise onto other degrees of freedom of the system. Again a simple analogy: imagine measuring the size of an object — the measuring device has a particular inherent uncertainty or error. The idea mentioned above would be like improving your measurement of the *length* of an object, by pushing all the experimental error into the *width*. You obviously can't do that in the classical world, but something like it is possible with quantum systems.

SYNTHETIC QUANTUM SYSTEMS AND SIMULATION

With the ability to create and control quantum systems, we can then consider using these systems as a set of elementary, microscopic building blocks. We can combine different quantum systems, each with tunable interactions, to build up new and exotic systems to explore their properties — essentially, we can create new forms of "quantum matter".

Theory can guide us to understand some equilibrium properties of such many body systems, but amazingly, the physics of such strongly coupled, individually addressable, many-body quantum systems is largely unknown — and mostly inaccessible to existing numerical simulation techniques. This includes many of the most exciting conditions, such as highly entangled many-body states, and exotic states of matter with hidden, long-range quantum order. Ultimately, producing and characterising this "quantum matter" will lead us to new regimes of physics; the phenomena and behaviour arising in such interacting systems may be very different from those observed in their component systems.

Aside from simply accessing new regimes of physics, however, we have the potential to produce controllable simulators of other complex quantum systems — that is, using simple quantum building blocks to build models to simulate more complex quantum systems. In fact, by leveraging tunable interactions and independent control over constituent quantum systems, one may realize what amounts to a universal quantum simulator. This has the potential to impact the field of condensed matter physics and strongly correlated systems, but may also simply provide a useful computational tool for modeling the interactions of large ensembles. For example, a quantum simulator may be used to simulate the various candidate models for high-temperature superconductivity, which elude existing numerical calculations. Such simulations may ultimately be a fundamental tool in the production of new designer materials.

QUANTUM METROLOGY AND SENSORS

Physical systems that are strongly governed by quantum effects can serve as exquisitely sensitive detectors. "Quantum enhanced" sensing and metrology, including the ability to probe or image single electron and nuclear spins, or the measurement of single quanta in mechanical systems, is a fundamental and enabling capability that could lead to breakthroughs including the noninvasive imaging of proteins and drugs in-vivo, probing of biological and quantum mechanical phenomena in liquids and solids, and ultimately the development of a deep understanding of our world at the atomic scale.

A variety of systems are now capable of being employed for sensing and metrology, with applications from precision time and frequency standards, to deployable field sensors and bio-imaging. And the utility of exploiting quantum effects and control techniques for accessing new measurement capabilities is already proven: using quantum logic methods on trapped ions, a new frequency standard has recently been developed in the USA, with a precision more than one hundred times better than the current standard techniques for defining and measuring frequencies.

Similarly, techniques from quantum computing for phase estimation have been applied to a variety of precision measurements in proof-of-principle experiments, and may soon find application in demanding scientific and technological areas such as gravitational wave detection.

QUANTUM COMPUTING

The field of Quantum Science has diversified significantly since the publication of Shor's algorithm — an algorithm for factoring numbers using entangled qubits — set in motion the rapid development of what is now the field of quantum computation. Despite an expansion of applications and interests, the production of a functional and useful quantum computer remains a primary goal in the field of Quantum Science.

The field has been making huge strides not just in theory, but also in experimental implementations of quantum logic. The pace of experimental progress has been astounding, with robust and useful qubits made using trapped ions, neutral atoms, photons, spins in semiconductors, and superconducting circuits. Beginning with the first two-qubit logic gate in the 1990s using trapped ions, we've seen several experimental demonstrations of algorithms for factoring and searching, simple implementations of quantum error correction, and most recently a fully programmable two-qubit quantum processor. Similarly, theoretical developments have accrued rapidly, with novel concepts such as topological quantum computing, subspace error correction codes, and measurement-based (one-way) computing making significant progress.

Open challenges remain, largely around the development of efficient and scalable quantum error correction protocols, the continued realisation of high-precision quantum operations, and the scale-up of quantum logic hardware — two qubits can only get you so much computation!

The progress is extremely positive, and at this time no fundamental roadblocks have been discovered suggesting that a realistic quantum computer is out of reach. But as mentioned above, the impacts of developments in the field are felt outside of the quantum computing community — quantum logic and algorithms are now being employed in precision metrology, with interdisciplinary work likely to grow in the future.

QUANTUM SCIENCE IN THE 21ST CENTURY

Our newly discovered abilities to control and manipulate quantum coherent systems provide access to a host of physical phenomena that for decades were considered nothing more than mathematical oddities. Quantum science permits us to use these basic phenomena underlying physical reality as a resource in the construction of new technologies, and the expansion of our knowledge of the world around us. The world will be a very different place at the end of this century, and Quantum Science will likely have had a lot to do with the changes we see.

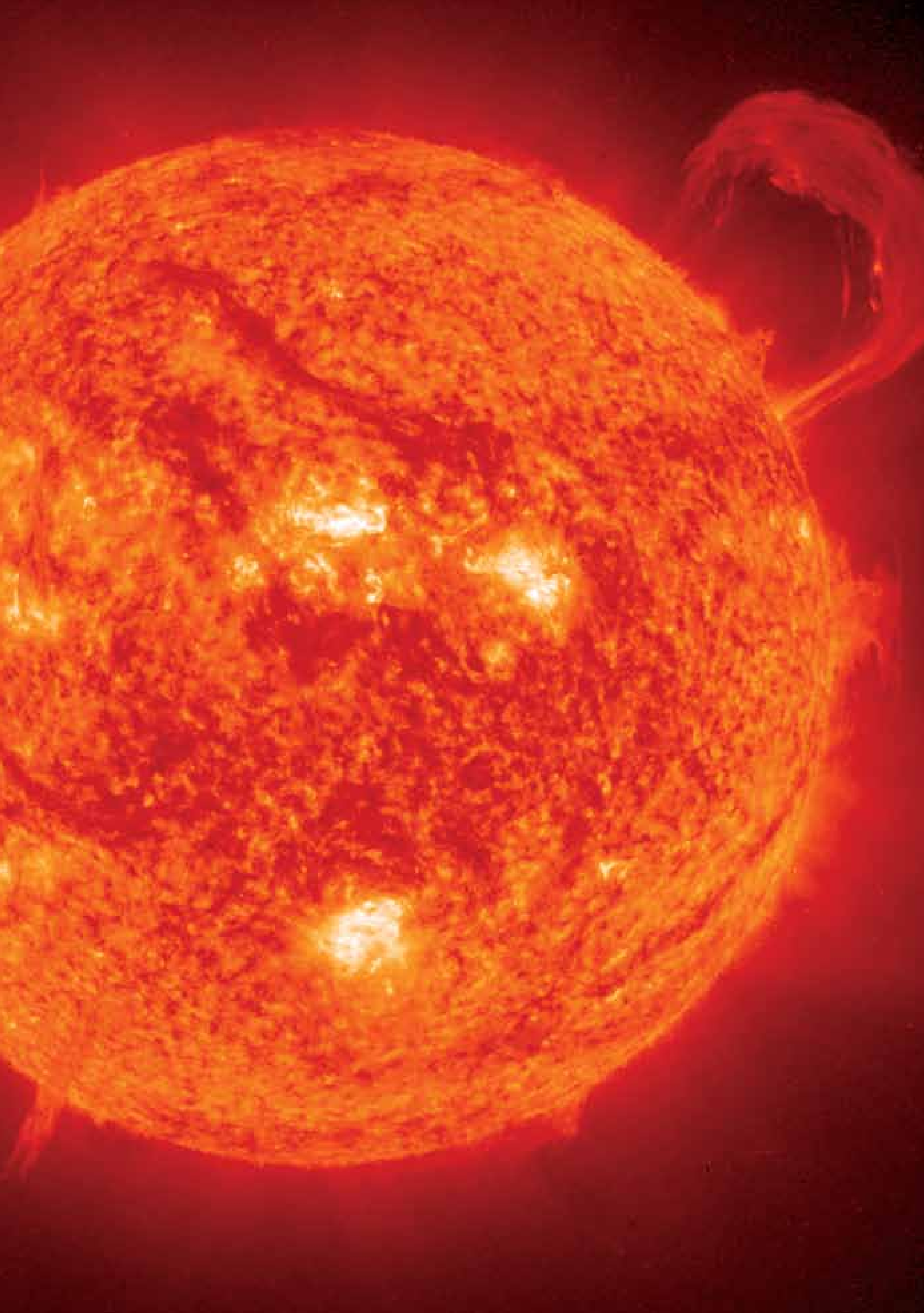
Further Reading:

Quantum Science Research Group at the University of Sydney: http://www.physics.usyd.edu.au/research/quantum/

Quantum Computing for Everyone, Michael Nielsen: http://michaelnielsen.org/blog/quantum-computing-for-everyone/

Quantum Technology: The Second Quantum Revolution, Jonathan P. Dowling, Gerard J. Milburn: http://arxiv. org/abs/quant-ph/0206091

Quantum Computation and Quantum Information, Michael A. Nielsen and Isaac L. Chuang. Cambridge University Press, 2000.



CHILDREN OF THE STARS, PLASMA IS THE FOURTH STATE OF MATTER

PROFESSOR CHRISTINE CHARLES

INTRODUCTION

Everywhere we look, there is plasma. But we stand on solid earth and this solid state accounts for less than one percent of the total mass of the Universe. Almost all of the rest is plasma, a hot ionised gas containing positive and negative charges (except, perhaps, for dark matter). Plasmas have existed since the very first moments of the Universe. It is the stuff of stars; it even fills the space between stars. It gives us the beautiful northern and southern aurorae. By properly harnessing the plasma state in the laboratory we can make plasma lights (fluorescent tubes), microchips for computers and mobile phones, plasma space engines and hydrogen fuel cells. Plasma applications also include arc welding, waste treatment and plasma coatings in the manufacture of biomaterials.

In this chapter, I will introduce plasmas and their applications in various fields of physics. Here we are not dealing with blood plasmas, but instead with a hot ionized gas containing atoms, ions, electrons and photons which carry the light to your eyes. First we will use water to explain the basic plasma constituents and reactions, including fusion reactions that occur in the Sun's core. After identifying some of the plasmas occurring 'naturally' around us, I will show some methods we can use to create and control plasmas in the laboratory. I will then discuss a few applications of laboratory plasmas in the fields of lighting, space travel, renewable energy, welding and biotechnology. At the end I will show some low cost experiments that can be constructed in high schools to illustrate the basic properties of plasmas.

PART I: THE FOUR STATES OF MATTER: SOLID, LIQUID, GAS AND PLASMA

Sunlight and water are two crucial ingredients for life, but an excess of sunlight and heat can lead to drought while an excess of water can lead to floods. Both usually have devastating effects on livestock, soil, residential areas etc… Since water is so abundant on Earth, let us use it to describe the four states of matter.

The water molecule, H_2O, contains one oxygen atom linked via OH covalent bonds (called the Coulomb force) to two hydrogen atoms, as shown in Figure 1. Between two water molecules there is a hydrogen bond (also known as the Van De Waals force) as shown in Figure 2. At atmospheric pressure (i.e. at sea level at the surface of the Earth) and at temperatures below zero degrees Celsius ($T \leq 0^{\circ}C$), water molecules will arrange themselves in a repetitive crystalline

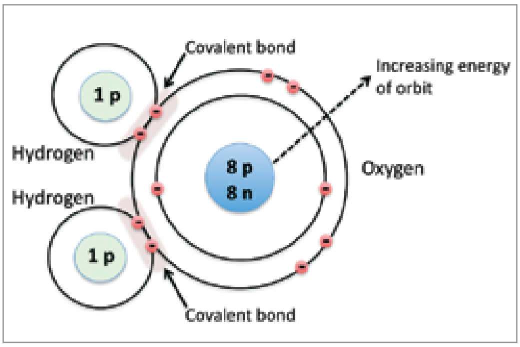


Figure 1: The Bohr model of the water molecule H_2O showing the OH covalent bond.

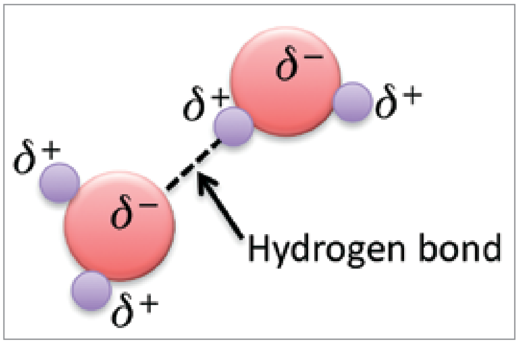


Figure 2: Hydrogen bond between water molecules, which is an electrostatic attraction resulting from a negative charge excess (δ^-) near the oxygen due to the unshared electron pairs (Figure 1) and a positive charge excess (δ^+) near the hydrogen.

structure to form a solid which we call ice. Ice is found naturally in the Earth's ice caps at the North and South poles and in smaller amounts in glaciers, icebergs and snow. At home, we can control the transition between the liquid and solid states of water by using a freezer to make ice cubes for our drinks.

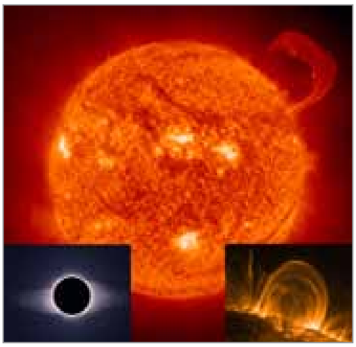
The solid state is called the **first state of matter** — to reach the other three states we simply need to find a source of energy, such as heat or motion. If we heat up a couple of ice cubes (for example against the skin in the palm of our hand) or rub them together, the temperature will rise and water droplets will form, leading to the formation of water as a liquid: the **second** state of matter.

We all know that snow melts at just above zero degrees Celsius at atmospheric pressure. Although the crystalline structure of ice or snow is now broken, water molecules are closer to each other but less tightly bound (in most cases, molecules in the liquid phase, e.g. liquid oxygen or liquid hydrogen, are further apart than in the solid phase, but the polarity of water makes it an exception).

If we add more energy (such as heat) into liquid water, we can make the transition to the **third** state of matter: by boiling water in a saucepan until steam or water vapour forms at 100 degrees Celsius, the transition to the gas state of matter is reached and we observe the formation of water vapour rising upwards. All the hydrogen bonds between the H_2O molecules are now broken, individual molecules are 'floating' around and are 'free' to escape the saucepan: unless there is a lid, this gas will expand to fill up space in your kitchen. A naturally occurring example is steam from natural hot springs (such as those found in New Zealand or Japan).

If we continue putting energy or heat into our water vapour, we will eventually start breaking the bonds within the H_2O molecules, producing oxygen and hydrogen atoms together with some additional particles known as 'radicals' (for example OH radicals). We also start removing electrons from those molecules, radicals and atoms to form positively charged ions: by raising the temperature the gas has become ionised and the **fourth** state of matter, called **plasma**, is obtained. Scientists currently believe that 99 percent of the matter in the Universe is in the plasma state.

To get some insight into the plasma state, it is necessary to have a closer look at its atomic structure. Consider the Bohr Model of H_2O shown in Figure 1: every atom contains a dense and heavy nucleus with protons (positively charged) and neutrons (no charge). The fast and light electrons (negatively charged) rotate around the nucleus on orbits of various 'altitudes', which are known as energy levels. If one H_2O molecule is stripped



Opposite: Figure 3: Satellite images of the Sun (Courtesy of NASA), a very magnetically active star; the left inset is the solar corona seen during a full solar eclipse and the right inset shows magnetic loops at the surface of the Sun.

of one electron (usually from an orbit furthest away from the nucleus), it becomes a positively charged ion H_2O^+, which will react to electric and magnetic fields (as will any charged particle).

Electrons can also be moved from low to high energy levels without being separated from the molecule to form an excited state: we write this symbolically as H_2O^*. These excited states do not last long and the electrons will fall back to a low energy level — and when this happens, a photon or ray of light is emitted.

A plasma can be thought of as a 'hot soup' of ions, electrons, atoms, radicals, molecules and photons. We cannot see individual ions, electrons, atoms, radicals and molecules with our eyes … but we can see the photons. To the naked eye, and depending on the surrounding conditions (temperature, pressure and chemistry), a plasma usually looks like a ball of light: the Sun is a prime example, as is a streamer or arc such as a lightning strike. The intensity of the emitted light increases with the density of charged particles within the plasma: the Sun is a very dense and bright star with around 10^{31} charged particles per cubic metre. In comparison, lightning strikes typically have 10^{21} to 10^{26} charged particles per cubic metre.

A water plasma is a complex system with various types of molecules (H_2O, H_2, O_2), atoms (H, O), radicals (OH), positive ions (H_3O^+, H_2^+, O_2^+, H^+, O^+, H_3^+), electrons and photons. To add to the complexity, electron attachment from 'free' electrons contained in the plasma can also occur, leading to an excess of negative charge with the formation of negative ions (OH^-, O^-, H^-). Most of these constituents can also be in various excitation states or energy levels: H_2^*, O^*, O_2^{+*} etc.

It becomes difficult to define a temperature for which the plasma state is reached since each species (each molecule, atom, ion …) has its own temperature which depends on the energy input into the plasma, the relative amounts of each species, the collisions between these species and much more. Therefore, the temperature of the electrons is often used when comparing plasmas: in laboratory plasmas, it is typically 50,000 degrees Celsius. Being heavier, the ions have a temperature typically around a few thousand degrees Celsius. Most importantly, though it contains charged ions and electrons, a plasma remains 'globally' neutral with equal numbers of positive and negative charges.

Having now explained the four states of matter with the water molecule, we go back to a simpler case: a hydrogen plasma, which would contain H_2 molecules, H atoms, H_2^+, H^+, H_3^+, H^- ions, electrons and photons (there is only a relatively small number of negative ions so we will not worry about them for now). Let us continue putting energy into this plasma to create one of the hottest plasmas known today. In this plasma, all electrons will be removed, and the ions (essentially the nuclei comprised of protons and neutrons) will go fast enough or be sufficiently 'hot' to overcome Coulomb repulsion, the force that makes two positive ions repel each other. Ions will be able to collide and **fuse**, forming heavier nuclei. This is known as a 'fusion reaction'. A perfect example of a very hot plasma is our Sun, which 'burns' hydrogen to produce helium through a series of steps (known as proton-proton chain) with the release a large amounts of energy of the order of $4x10^{26}$ watts:

$$\text{4H nuclei} + \text{2 electrons} \rightarrow \text{Helium nucleus} + \text{energy} \qquad (1)$$

A hydrogen nucleus contains 1 proton since it is a hydrogen atom stripped of its single electron (H^+). A helium nucleus contains 2 protons and 2 neutrons bound together since it is a helium atom stripped of its 2 electrons ($^4_2He^{++}$); it is called an alpha particle. In terms of weight 75 % of the Sun is hydrogen and 23% is helium. The rest is oxygen, carbon, neon, iron and other minority elements.

The coolest temperature at the surface of the Sun is about 4,000 degrees Celsius, which is significantly lower than that at the core where fusion reactions occur: there the temperature reaches an astounding 15 million degrees Celsius. The energy produced in the core as very high-energy photons (known as gamma rays) takes 10 to 170 thousand years to reach the Sun's surface, after a conversion into an enormous amount of low energy photons of visible light. This light or radiation escapes into space and takes another 8 minutes to travel from the Sun to the Earth. Part of the energy is also released in the form of a plasma, which is called the solar wind.

Our solar system was born about 4.5 billion years ago (the Big Bang occurred about 13.7 billion years ago) and the Sun has been burning its hydrogen fuel for all that time — and will do so for another 4.5 billion years or so, when it reaches the end of its lifetime. Before then, it will continue to supply us with 'free' sunlight and heat before eventually fading.

PART II: OUR SOLAR SYSTEM AND ENERGY RESOURCES

There is a lot of activity inside and at the surface of the Sun, as shown in Figure 3. Most of the radiation or light originates from a thin layer called the photosphere (300 km thick, with a particle density of around 10^{23} per cubic metre). This radiation expands out into space, and as the Earth is 150 million kilometres away from the Sun, only about 1.4 kilowatts per square metre impacts the top of Earth's atmosphere. It is this 'free' energy supply which can be converted to electricity using solar panels, or to hot water using solar hot water systems. Plants grow and make food by harnessing sunlight and the oxygen in our atmosphere in a process known as photosynthesis.

The second solar component, the solar wind, is essentially a plasma (with protons, alpha particles, electrons and other minority ions) born on the surface of the Sun and accelerated in the solar corona, a layer further out from the surface in the Sun's atmosphere. The corona is seen as a halo during a total solar eclipse (left inset on Figure 3) since its density is lower at about 10^{15} to 10^{16} particles per cubic metre. Interestingly the solar wind has a temperature between 1 and 20 million degrees Celsius, which is much hotter than the coolest surface layer (4,000 degrees) located 500 km above the photosphere.

The Sun is a magnetically active star with sunspots, flares of various sizes (magnetic loops, shown on right inset of Figure 3) and coronal mass ejections which determine what we call space weather: in these ejections, the solar wind expands into space and impacts all planets in our solar system. Since Earth is also a big magnet, the Solar wind strongly interacts with Earth's magnetosphere to produce effects such as electric storms and auroras, with maximum activity at the North and South poles. Although these are beautiful lights and draperies in the sky, they can cause significant damage since they occur in the two plasma layers in the outer part of the atmosphere (the magnetosphere and the ionosphere) where most of the telecommunications satellites are placed in orbit (Figure 4). Closer to the Earth's surface are gas layers (the troposphere, the stratosphere and the mesosphere), which contain the ozone layer that protects us from harmful high frequency ultraviolet light or UV radiation from the Sun.

From space, Earth appears as a blue sphere with green, red and white patches while Mars is a smaller red sphere with some grey patches (Figure 5). Oceans cover a large proportion of our planet and Earth's atmosphere contains mostly nitrogen (78%) and oxygen (21%). In contrast, Mars is a

Figure 4: Representation of a typical telecommunication satellite orbiting the Earth (Courtesy of ESA) showing solar panels and radiation shield.

Figure 5: Satellite images of planet Earth (blue) and planet Mars (red) (Courtesy of NASA).

dry planet and its atmosphere is essentially carbon dioxide (95%) with a bit of nitrogen (3%). It is interesting to note that most of the water on Earth is in the liquid state and there is an equilibrium between the Earth's oceans and atmosphere which provides us with a relatively continuous supply of water.

The first unicellular organisms appeared on Earth about 3.8 billion years ago, and the oxygen that we depend upon now was produced and released from photosynthesis by blue-green algae (cyano-bacteria) or stromatolites: the latter are in fact an assembly of limestone sheets resulting from the process. Stromatolites are still around today, and can be seen at a place called Shark Bay in Western Australia.

Human evidence can be traced back a few million years, and Homo Sapiens about 160 thousand years. Our evolution is directly linked to the Earth's evolution (its volcanic activity, meteorite impacts, atmosphere content, ice ages ...) and we rely on the presence of sunlight, water and oxygen as our three main 'fuels'. Our fourth 'fuel' is food, and we convert it via chemical processes to heat (to maintain our body temperature and allow us to breathe, think etc) and also to mechanical energy (so that we can move, work, study and play). Each of us is in fact a 'thermal machine', a reservoir of energy that needs a constant supply of energy throughout the day. We each take an average of about 10,000 steps a day and need around 2,500 kilocalories per day: this is energetically equivalent to a 200 watt light globe turned on for 24 hours, or the usable energy in half a litre of petrol.

As we go about our daily lives, we also need to find shelter, go shopping and drive to work — and in a developed country such as France, the equivalent daily consumption is nearly 10 litres of petrol. This corresponds to 15 litres of petrol extracted from the Earth since transformation, transport and storage of energy is always subject to heat or energy loss back to the environment. In fact the usual repartition is not 15 litres of petrol, but more like 50% petrol, 20% electricity (coal, hydro, thermo nuclear, solar, wind), 20% natural gas (methane, CH_4), and the rest from solar, wind or other sources. There is no 'energy perfect system': for example, 25 percent of the petrol burnt in a car serves to move it, the rest is lost in heat. Although energy loss cannot be suppressed, it can be substantially decreased.

On Earth, we have many different forms of energy available to us. At the Earth's core is a large reservoir of heat that is maintained by radioactive decay. This reservoir dates from the birth of the Earth itself. Fossil fuels with a large hydrocarbon content (coal, petroleum, natural gas) are the fossilized remains of dead plants and animals formed over millions of years. In only two centuries, human activity on this planet has expanded by using the non-renewable supplies of fossil fuels: first with coal and steam engines, then with petrol and combustion engines, and also with electricity. With the Earth's population and our individual energy needs both increasing, it is clear that we will eventually run out of our fixed supply of hydrocarbon fuel. We also know that we are already altering Earth's fragile equilibrium both at the surface and in the atmosphere, in ways such as deforestation, soil erosion and carbon dioxide (CO_2) emission, and creating the well-known greenhouse effect and global warming.

In the future, how and where will we find energy? How will we transport, store and manage this energy? What will we do with the waste products? Plasma, the fourth state of matter, is not a solution to all these important challenges, but may contribute to a better management of our energy resources. As will be shown in the next section, the diversity of plasma applications implies that the fourth state of matter will be a key player in maintaining the health of planet Earth.

PART III: LABORATORY PLASMAS AND THEIR APPLICATIONS

Gaseous plasmas or electrical discharges have been studied since the 1920s. Examples of naturally occurring plasmas include the Sun, the solar wind, the Earth's magnetosphere and ionosphere, and lightning strikes. There are various ways of creating a plasma in the laboratory and there are numerous plasma applications which we use in our everyday lives. Depending on the application, one plasma species, or a combination of species, can be used; for example, the photons emitted by a plasma are used in fluorescent tubes or plasma display panels. Plasma ions are used in the manufacturing of small structures in integrated circuits for the production of computer chips. They can also be used in space engines as a source of thrust to maintain satellites or spacecraft onto correct orbits, and in the manufacturing of new materials such as hydrogen fuel cell electrodes. The reactive molecules, atoms or radicals, combined with the high plasma temperatures, can be used for plasma/arc welding, spraying, waste treatment and much more.

LIGHTING

A fluorescent tube is a plasma lamp that uses electricity to produce visible light. The tube contains neon gas with a small amount of mercury vapour. Excited mercury atoms produce short wavelength (high frequency) ultraviolet light, and a phosphor coating on the inner surface of the tube converts the ultraviolet radiation to light that is visible to the human eye (roughly 400 to 800 nanometres in wavelength, or frequencies between 800 and 400 terahertz). A fluorescent tube only costs a few dollars and has electrodes at both ends: the plasma (or 'discharge') is established by applying 240 V from the alternating current (AC) outlet.

Plasmas are also used for light-production in plasma display panels, though the market for these has been considerably reduced due to the exceptional progress in LCDs (Liquid Crystal Displays).

MICROCHIPS

In developed countries around the world (and, increasingly, in developing countries too), most people have access to laptop computers, mobiles phones, digital cameras, digital watches, MP3 players, GPS systems, and of course much more. Nearly all industries have efficient and reliable computerised control systems to provide their service to the highest quality and the lowest price. The transistor, a critical component of many modern-day electronic devices, was invented in 1948, and in the 50s and 60s transistors and integrated circuits were manufactured using a method known as 'wet chemistry'. The 1970s saw a microelectronics revolution, with plasma processing enabling the development of 'dry' processing, and this lead to the miniaturization of integrated circuits.

A dry plasma process typically uses both the active molecules and atoms in the plasma together with the ions; the latter can be 'directed' and energised using electric fields.

In 1975, the present home computer would be the size of a house and cost millions of dollars. From 1970 to 2006, plasma technology has reduced the size of electronic components by a factor of 250,000 and created most new high-tech industries in the world!

It takes about 30 plasma processing steps to create a transistor on a silicon wafer, and around two months for a complete wafer with millions of transistors to go from entrance to exit in a production line. The full suite of processes includes the etching and/or deposition of metals (for the connections), semi-conductors (for the heart of the transistors) and insulators (for transistor gates, or isolation between components).

A simple example in this quite complicated process is plasma etching. To etch silicon wafers, gases such as sulfur hexafluoride (SF_6) are dissociated (broken apart) by the plasma to produce active halides:

$$SF_6 \text{ (gas)} \rightarrow SF_2 + 4F \text{ (plasma)} \quad (2)$$

These then form volatile products with the silicon at the wafer surface:

$$4F \text{ (plasma)} + Si \text{ (wafer surface)} \rightarrow SiF_4 \text{ (volatile)} \quad (3)$$

In this way, the silicon (Si) is 'etched' from the wafer surface, and the volatile product SiF_4 is pumped out of the plasma system.

To obtain the necessary patterns corresponding to transistors and more complex circuits, photolithographic processes are used: a patterned mask is placed onto the silicon wafer, the plasma process is carried out and the mask is stripped. A processed wafer would typically have hundreds of identical circuits, so that each circuit could be sliced off, integrated in a 'black' chip with connecting legs and mounted onto printing boards to supply power, transport information and obtain interconnectivity between various circuits of functions. Schematics of some of these steps are shown in Figure 6.

SPACE ENGINES AND PLASMA ROCKETS

Based on Newton's third law, the principle of a rocket is to emit particles via an open exhaust or nozzle thereby creating a force (called the thrust) in the opposite direction: these particles can be in a gas phase (as in conventional chemical rockets) or in the plasma phase (creating what are known as 'plasma engines').

Although a fluorescent tube is designed to be powered by the 50Hz AC domestic mains power supply, it is possible to ionise the neon gas inside the sealed tube with radiofrequency power (RF, corresponding to MHz frequencies) applied to an emitting antenna, similar to a radio station antenna. Bringing a fluorescent tube close to the tip of a Tesla coil (30,000 Volts at 3 MHz, with a pulse rate of 10 percent) will light up the fluorescent tube — so will a lower signal (1,000 volts at 13.56 MHz) applied to a metallic loop wrapped around the tube.

Once the neon plasma is created, opening one end of the tube results in the plasma exiting and expanding, creating an opposite push on the tube — and turning it into a rocket! Not a very good rocket, though: doing this with a fluorescent tube in the open air, nothing much would happen since the thrust would be very small and air friction or drag would prevent any movement. But the thrust is there!

Conventional rockets are based on combustion of propellant in the nozzle, and are fundamentally limited chemically by the amount of energy that can be stored in the chemical bonds of the fuel, and thermally by the maximum temperature that the walls of the combustion chamber can withstand. Chemical rockets are well-suited to high thrust, low specific impulse (high propellant usage) missions like a shuttle launch, but an alternative is needed for interplanetary missions that require high specific impulse (low propellant usage) but only require low thrust because they occur in microgravity. Gravity causes the density and pressure of air to decrease exponentially as one moves away from the surface of the Earth; in space or in the interstellar medium there is microgravity, and very few neutral and charged particles. The field of electric propulsion is dedicated

Fig. 6 a

Fig. 6 b

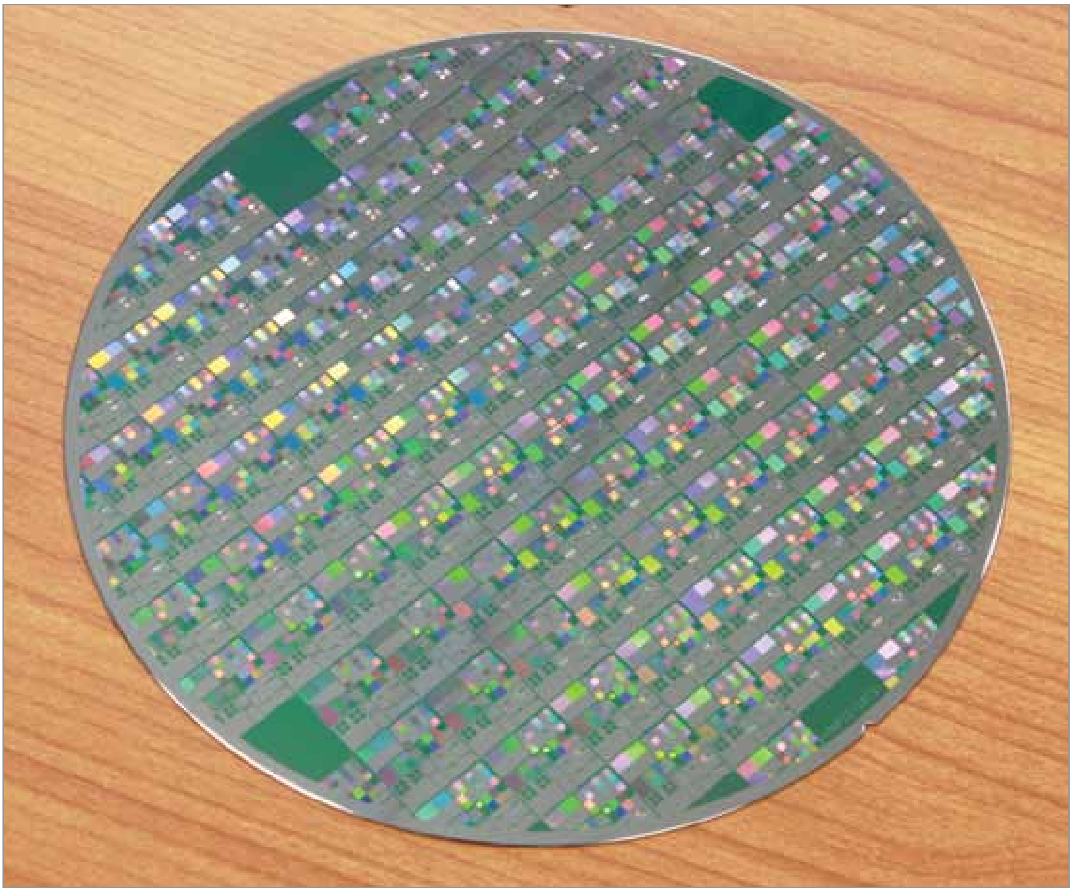


Fig. 6 c

Figure 6: A summary of the microelectronics steps: (a) Bare silicon wafer; (b) A typical high aspect ratio trench obtained by plasma etching of silicon (Courtesy of LAM RESEARCH); (c) Plasma processed silicon wafer showing the integrated circuits; (d) Microchips encapsulating integrated circuits mounted on a printed board; (e) A modern phone showing an application of the microelectronics revolution.

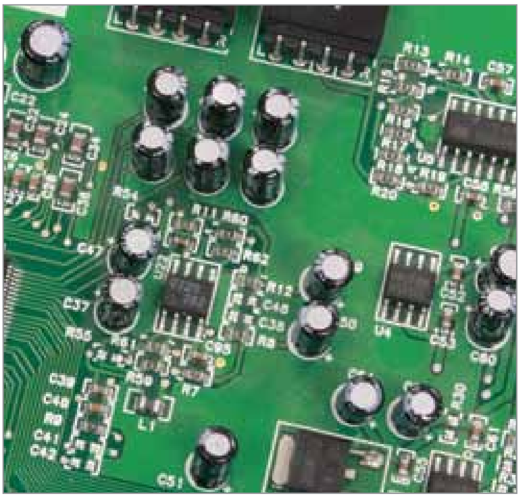


Fig. 6 d

Fig. 6 e

to developing the higher specific impulse technologies necessary for such missions.

Plasma propulsion is a branch of electric propulsion that uses the properties of plasma to overcome the basic barriers that limit conventional rockets. Because plasma propulsion uses its propellant more efficiently, it can be used on missions that require greater changes in velocity once in space. Ion gridded thrusters (for example, NASA's DEEP SPACE 1 mission) and Hall effect thrusters (such as ESA's SMART 1 mission to the Moon) both eject ions that are initially created in a plasma, and both need a second device called a neutraliser to emit a sufficient number electrons to neutralise the high velocity emitted ion beam. The ions are accelerated by potentials applied to an anode or grids, allowing the thrusters to achieve low fuel consumption and provide good thrust (typically 100 mN for 1 kW, which corresponds to the force of Earth's gravity on a mass of 10 grams). The drawback of this efficiency is the erosion of the grids or thruster walls by the ions.

For ambitious interplanetary missions where safety and reliability are more important, magneto-plasma thrusters are of interest since a neutral plasma is emitted (so there is no need for a neutraliser), and where the source of thrust is essentially from the ions. Other applications also include low cost missions where spacecraft residues can be used to put satellites into graveyard orbits at the end of their lifetime. These thrusters have yet to be tested in space.

One of the thrusters in development, the Australian Helicon Double Layer Thruster (HDLT) concept, is based on the discovery of a current-free electric double layer in an expanding magnetised plasma (see Reference 1): electric double layers are like cliffs of potential (like a river waterfall) that can energise charged particles falling through them. A large area, low divergence high velocity ion beam (the source of thrust) is measured in the nozzle for various propellants and the plasma itself provides a sufficient amount of neutralizing electrons. A cartoon of the HDLT in Figure 7 shows that the HDLT consists of a Pyrex tube, an RF copper antenna surrounding the tube and connected to a radiofrequency (13.56 MHz) generator, two solenoids (connected to a current power supply) creating a divergent magnetic field, and a gas feed (connected to a gas tank) positioned at the 'closed' end of the thruster. There are no moving parts and no electrodes or grids immersed into the plasma. The HDLT is basically like a fluorescent tube with a gas inlet at the closed end (the open end is the exhaust), an antenna wrapped around the tube to couple power into the plasma, and some copper windings to form a solenoid that generates a diverging magnetic field.

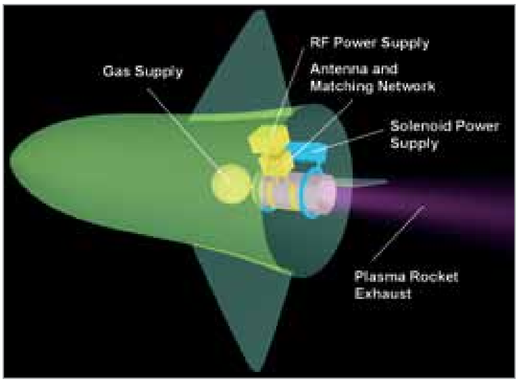


Figure 7: A cartoon of the Helicon Double Layer Thruster.

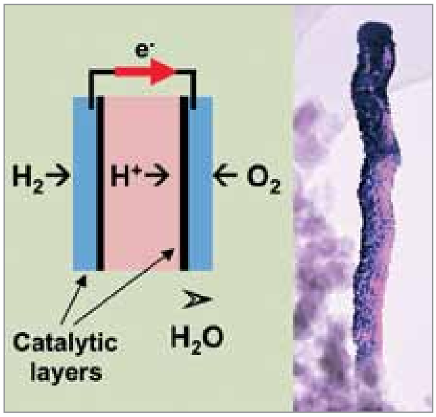


Figure 8: Schematic of a hydrogen fuel cell with a typical carbon nanotube supporting nano-clusters of platinum (the catalyst).

The HDLT works with a variety of propellants (argon, xenon, oxygen, hydrogen, ammonia...), including carbon dioxide — which would allow refuelling on Mars or Venus, both of which have an atmosphere containing mostly CO_2.

HYDROGEN FUEL CELL CARS

Since the 1960s, NASA has been using fuel cells in space and military operations, and fuel cells are now being developed to power cars, homes and businesses. A hydrogen fuel cell uses the chemical energy stored in hydrogen to make water, heat and electricity — in effect, the H_2 molecule is being used as a battery with stored potential energy in its covalent bonds, and this energy is greater than the energy in the covalent bonds of water:

$$H_2 \text{ (gas)} + \tfrac{1}{2} O_2 \text{ (gas)} \rightarrow H_2O \text{ (liquid)} + \text{energy} \qquad (4)$$

For this 'global' reaction to occur, hydrogen needs to be decomposed into two protons and two electrons at one electrode (anode). This is done with the aid of a catalyst, say platinum (Pt):

$$H_2 + Pt \rightarrow 2\,H^+ + 2 \text{ electrons} + Pt \qquad (5)$$

A membrane transports the two protons to the cathode and an external circuit transports the two electrons to the cathode where the second reaction occurs:

$$\tfrac{1}{2} O_2 + 2\,H^+ + 2 \text{ electrons} + Pt \rightarrow H_2O + Pt \qquad (6)$$

A schematic of this membrane-electrode assembly is shown in Figure 8. The catalyst, usually platinum, needs to be present to trigger the reactions at the two electrodes: it consists of a carbon cloth covered with a porous coating containing carbon particles (for conductivity), platinum clusters and some polytetrafluoroethylene to evacuate water out of the cell (polytetrafluoroethylene is hydrophobic; it is also used as a non-stick coating for pans and cookware). Unfortunately there is a limited supply of platinum on Earth, so we need to use what we have efficiently and effectively.

In commercial hydrogen fuel cells manufactured with chemical processes, the platinum represents around one-third of the cost — and 90 percent of the platinum is not 'active'. Hence research on fuel cell manufacturing with plasma processes aims at reducing the amount of catalyst, and also at developing a complete 'plasma fuel cell' technology to deposit the catalytic electrodes and the proton exchange membrane in a series of plasma steps (similar to the steps involved in integrated circuits).

By placing a negatively biased platinum target in a plasma (for example an argon plasma) and a fuel cell electrode next to it, the plasma ions are accelerated towards the target and sputter the platinum atoms away from the target. These platinum atoms get transported in the plasma and re-deposit onto the electrode in the form of small nano-clusters, each just a few nanometres in diameter but with a large effective surface area. In the laboratory, this process increases the catalytic efficiency by a factor of 50 at the anode (hydrogen side) and by a factor of 5 on the cathode (oxygen side).

More recent work focuses on depositing the catalyst nano-clusters onto carbon nanotubes also grown in a plasma (for example in a methane-hydrogen plasma). An example of a plasma-grown carbon nanotube, plasma coated with platinum nano-clusters, is shown in Figure 8.

In principle, the energy conversion efficiency in a fuel cell can be as high as 80 percent, compared with about 30 percent for a combustion engine. Although many car manufacturers have developed fuel cell cars, a network of hydrogen cars would need an entirely new infrastructure to provide the hydrogen fuel. The city of Perth in Australia had a successful hydrogen fuel cell bus trial (see Reference 2) running for four years between 2004 and 2007.

OTHER APPLICATIONS

In laboratory plasmas, the operating pressure varies from atmospheric pressure to 10^{-7} atmosphere, electron temperatures go from one thousand to a few hundred thousand degrees Celsius, and charged particle densities are in the 10^{12} to 10^{20} per cubic metre range. Fusion plasmas try to achieve particle confinement with typical charged particle temperatures of 10^{8} degrees Celsius. Figure 9 and Figure 10 show photos of a low-pressure plasma and an atmospheric pressure plasma respectively.

Industrial plasmas are used over a broad range of pressures using a variety of reactive or non-reactive gases. In microelectronics most plasma processes are at low pressure, to obtain large volume plasmas and process large wafers as uniformly as possible to reduce manufacturing cost. It is also necessary to use clean rooms to prevent dust contamination at all times. In the food industry, silicon oxide and aluminium oxide plasma coatings are deposited to provide a better airtight barrier for food packaging (for example, in milk and juice tetrahedral containers where the raw materials for the package itself are paperboard, plastic layers on both sides, and sometimes aluminium foil). In medicine, titanium dioxide plasma coatings are often used on orthopaedic implants to improve bone adhesion.

Thermal plasmas (plasmas at or near atmospheric pressure, in which both the electrons and heavy species are at high temperature), such as electric arcs and plasma jets, are very widely used in industry. For example, arc welding is used to join metals by melting their edges together with the intense heat of an electric arc. Plasma cutting employs similar arcs to cut through metals, while steel and other metals are produced in electric arc furnaces.

Plasma spraying deposits wear-resistant or thermally resistant coatings to surfaces by injecting particles into a plasma jet. Here the particles are partially melted and then blasted onto the substrate — a typical application is the coating of jet engine turbine blades with zirconia.

In plasma waste treatment, a relatively new thermal plasma application, the intense heat of the arc plasma converts waste (ranging from household garbage to highly-toxic chemicals) into harmless — and in some cases useful — products, such as syngas (a mixture of hydrogen and carbon monoxide), which can be combusted to produce electricity. For example, an Australian system called PLASCON (see Reference 3) is used around the world to convert organic liquids, ozone-depleting chemicals and fluorocarbon greenhouse gases into harmless substances.

The recent development of non-thermal, atmospheric-pressure plasma sources has found applications in biomedical fields. While in the past only the thermal properties of plasmas (temperatures above 80°C) were useful for cauterisation or sterilisation, current research is directed primarily at the non-thermal effects of plasma for the treatment of heat-sensitive

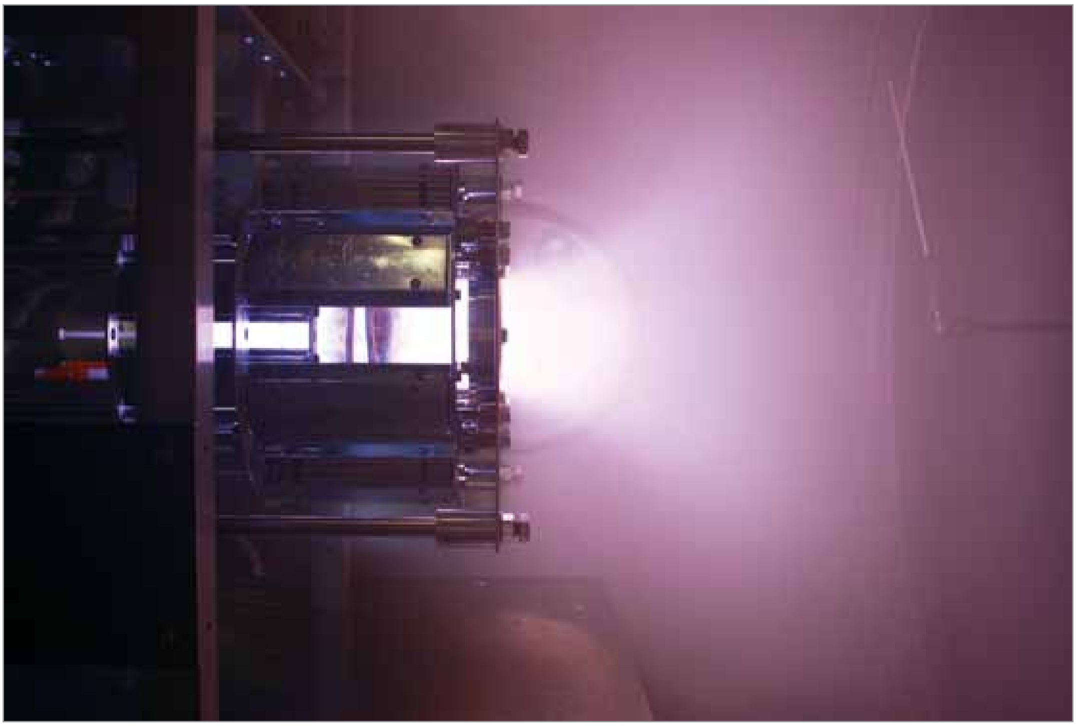


Figure 9: Photo of a typical low pressure laboratory plasma (a radiofrequency excited plasma thruster using argon propellant).

materials and even living tissues: a new field in plasma chemistry called 'Plasma Medicine'. Without damaging surrounding healthy tissue, cold atmospheric-pressure plasmas at room temperature produce diverse reactions in tissue, with physical as well as therapeutic effects. Numerous components of plasmas, such as reactive species (for example, O and NO_x), charged particles, electric fields and even UV light can trigger a complex sequence of biological responses in tissues and cells.

Some examples of plasma medical applications are sterilisation, healing wounds and even treating dental cavities, skin diseases and some cancers. Obviously the complexity of both the plasma and the cell biology makes the study of the interaction of plasma with living system quite complicated. The understanding of those interactions is the basis of plasma medicine research, which lies at the frontier of plasma science, chemistry and biology.

The list of plasma applications expands every day.

PART IV: EXPERIMENTAL WORKSHOP

THE TESLA COIL

During my lecture we are going to perform plasma experiments with the Tesla Coil, as shown in Figure 11.

The Tesla Coil (Figure 11a) was invented by Nikola Tesla about one hundred years ago. Nikola Tesla also gave us AC power rather than DC (direct current or continuous) power. The Tesla Coil is very similar to an automobile spark plug, producing little sparks of very short duration — consequently, from the tip you get little bursts of high frequency (about 3 MHz) that last about ten oscillations, around a microsecond or two (Figure 11b). However, it operates at very high voltage — around 30 thousand volts on the tip — so it acts like a radio antenna. The Coil produces electromagnetic waves that will broadcast all around the building; so if anyone is nearby with a radio, it will affect their reception.

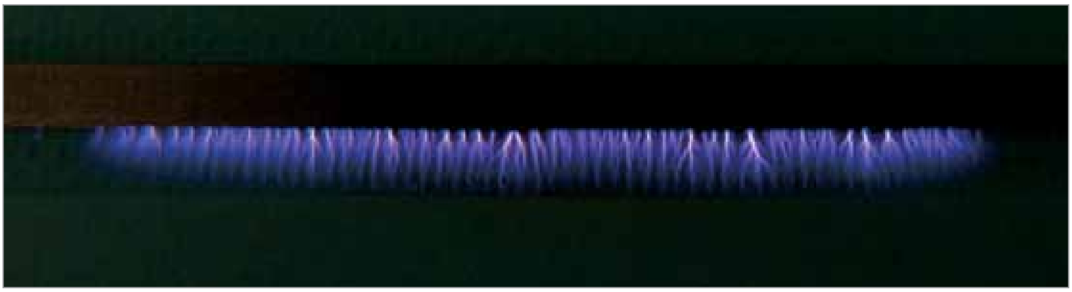


Figure 10: Photo of a typical atmospheric pressure laboratory plasma (a radiofrequency excited surface filamentary discharge in air).

Experiment: The leg of a chair
To see how high the voltage is, we turn the coil on and we will look for something that is earthed (some metallic rod such as the leg of a chair). If you look closely (see Figure 11c) you will see that we have a spark of one to one and a half centimetres and there is a current path to earth via the metallic rod: a pink 'streamer', like a pink lightning strike, similar to Figure 10. You can see the photons emitted by the plasma; the ions and electrons are in there, but your eyes can not 'see' them. In the laboratory we use electrostatic probes to collect them in the form of an ion or electron current.

What you can't see in the picture is the smell that accompanies the streamer: it comes from ozone, O_3, produced when the spark dissociates an O_2 molecule into two oxygen atoms, each of which can attach to another of the O_2 molecules to form O_3.

Experiment: The spoon
Interestingly, at this high frequency you do not get a shock from the Tesla coil, even though the voltage is so high — you can touch the tip of the coil, but we will not do that since you can make much more spectacular demonstrations.

If you hold a piece of metal, such as a spoon, you become an earth because the radiofrequency will go through your body (Figure 11d). Holding the spoon with a loose grip, you can feel a kind of tingling sensation: this is quite an interesting effect called the 'skin' effect, where the radiofrequency current from the spoon to earth is carried by a thin surface layer of your skin. We commonly use this 'skin' effect in our plasmas in the laboratory, for example in the Australian Helicon Double Layer Thruster described earlier (Figure 9).

Experiment: The fluorescent tube
Because the electric field is so high (remember, the coil is at 30,000 volts), when it reaches a fluorescent tube it will be enough to produce a glow.

In air at atmospheric pressure we need about 10,000 volts/cm to produce a spark (as we saw with Experiment 1) — but if you decrease the pressure you need much less voltage: at a tenth of an atmosphere you need only 1,000 volts. At one-hundredth of an atmosphere you only need less than 100 volts.

Now fluorescent tubes normally are at much lower pressure than air — that is why with the normal 240 V applied to the electrodes at both ends, we can get a discharge to light up the tube. But we are not going to use these electrodes to light up the fluorescent tube — we are going to use a human.

Holding the tube with one hand, we have a similar effect to the spoon experiment. The current flows from the tip of the Tesla coil through the air, then through the neon plasma that is generated and finally through you to the Earth (Figure 11e). So you can see that the 'plasma light' is a maximum where your hand is — with two hands on the tube the plasma stays between the hands (Figure 11f).

Fig. 11 a

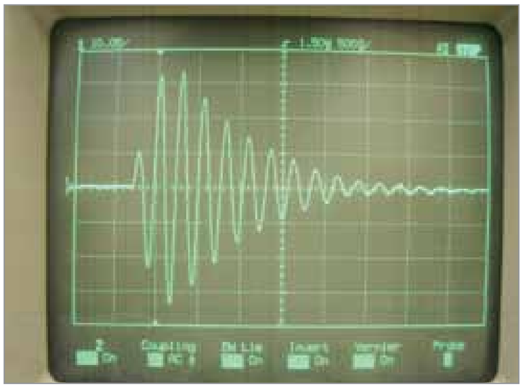


Fig. 11 b

Fig. 11 c

Fig. 11 d

Fig. 11 e

Fig. 11 f

Figure 11: Examples of plasma experiments: (a) Tesla coil; (b) Tesla coil signal; (c) The leg of a chair or metallic post experiment; (d) The spoon experiment; (e) The fluorescent tube experiment with one hand; (f) The fluorescent tube experiment with two hands; (g) The incandescent lamp experiment

Fig. 11 g

Experiment: The incandescent lamp (an old fashioned globe)

Inside the old-style light globes, the space is filled with gas at low pressure — it is argon, which we often use in our laboratory plasmas because it is cheap and inert. Holding the globe with your fingers on the glass (Figure 11g) and the Tesla Coil close to the terminals of the bulb, you see that streamers are produced inside that travel to your fingers.

With argon the streamers usually appear pink; if you see other colours, it suggests that there is a small amount of other gases in the lamp. The globe also gets hot because the discharge is heating the glass (in fact most of the heat loss is due to the ions being accelerated towards the walls in a plasma 'sheath'). If you hold it long enough, it will break — we won't try that. You may have seen the 'plasma spheres' in science museums or stores that run in the MHz frequency range. Those apply the same effect: the discharges follow where you push with your fingertips.

Experiment: Holding hands

Another thing we can try is the 'holding hands' trick — we will give this a try during the lecture. Student One holds a spoon in their right hand, their left hand holds the right hand of student Two, who in turn holds a fluorescent tube with their left hand. This 'experiment' works very well if the air is dry, and you may also have to adjust how hard you are holding hands because of the 'skin' effect. Remember that the current goes through your skin — we can turn the plasma on in the fluorescent tube by putting the tip of the Tesla Coil near the spoon and using that skin effect through your bodies.

In South Australia in 2009 we had a chain of over 10 people ... and the fluorescent tube still lit up!

PART V: CONCLUSION

This chapter was a brief description of plasma, the fourth state of matter. Plasmas occur naturally pretty much everywhere in the universe, and by harnessing the plasma state in the laboratory it is possible to use its broad range of properties in a multitude of applications.

Studying plasmas also means combining various disciplines such as physics, chemistry, engineering, even biology. This opens new and exciting scientific pathways for students and researchers alike. How we use energy plays an important role in our daily lives, and the sustainable consumption of the Earth's resources will ensure both the health of our planet and our civilization. How do we make a sustainable future that we can live and thrive in? Plasma research will help.

ACKNOWLEDGMENTS:

I would like to thank Dr Tim Wetherell for his help with Figures 5, 6, 7, 9 and 11, Dr Amael Caillard for Figure 8, James Dedrick for Figure 10, Dr Tony Murphy for discussions on thermal plasmas, and Dr Christian Sarra-Bournet for discussions on plasma medicine.

Footnotes

1. For more information on the author, the Space Plasma, Power and Propulsion Laboratory at the Australian National University and the HDLT visit http://sp3.anu.edu.au
2. http://www.transport.wa.gov.au/alt_finaloperation_report.pdf
3. http://www.plascon.com.au/

A VERY LARGE MICROSCOPE TO PROBE VERY SMALL DISTANCES — PART I

PROFESSOR ALLAN G. CLARK

INTRODUCTION

In the following two lectures I hope to share the excitement of our current research in Particle Physics at the CERN Large Hadron Collider (LHC), and to explain why the experiments at the LHC are expected to result in new understandings of fundamental interactions at high energies (small distances) in the laboratory, as well as insight into the conditions of the early Universe.

In our first lecture, we will first motivate the reason to explore very small distance scales and why this is important. We will then summarize our current understanding of the interactions between fundamental particles as well as what some of the open questions. We will see that this is also relevant to our understanding of the early Universe. Finally, we will show that collisions at the LHC probe much shorter distance scales than previously possible, and then conclude by describing the LHC machine and what makes it so special.

Our second lecture we will discuss the particle physics experiments at the LHC, with particular emphasis on the ATLAS experiment where Australia is very actively involved. We will show how to obtain some important physics results from the LHC experiments, and we will make a menu for the future.

HOW DO WE PROBE VERY SMALL DISTANCES?

The human hair, with a diameter of typically 50 μm, is easily visible with the naked eye, because the eye is sensitive to light waves of wavelength λ ~ 0.5 μm. We can see better using a transmission microscope where distance scales of typically 0.5 μm can be separated (Figure 1a).

Now, a fundamental tenant of quantum physics is that light can be thought of as either a beam of particles (we call these particles *photons*) or as a light wave. This is called wave-particle duality and in the case of photons, this duality can be expressed by the famous equation:

$$E = h\nu = \frac{hc}{\lambda} \qquad (1)$$

where E is the energy of the photon, λ is the photon wavelength and h[1] is a fundamental constant of nature called the Plank constant. Einstein introduced this equation in his explanation of the photoelectron effect. It is easy to see that by increasing the energy of a photon, its effective wavelength is decreased (for example the wavelength of x-rays is in the range 10^{-9} to 10^{-11} m), and this allows the resolution of objects much smaller than visible light.

Wave-particle duality also turns out to be true for particles with mass, and this is expressed

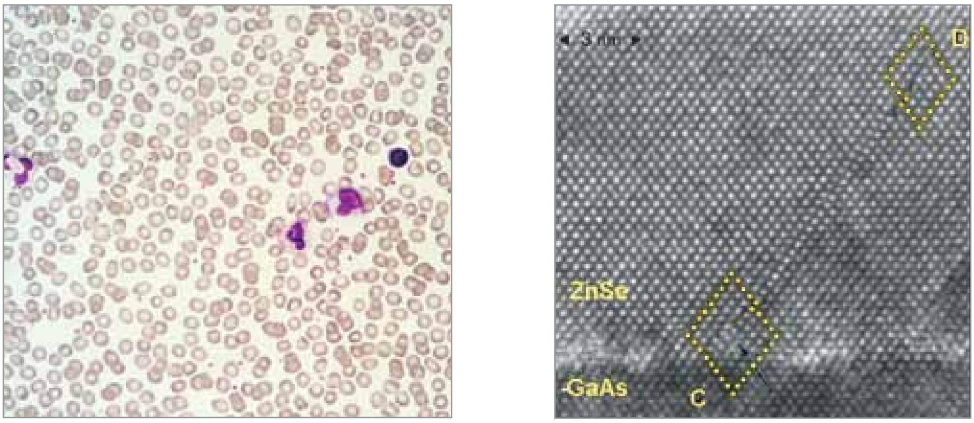


Figure 1. a) Identification of biological cells using visible light in a transmission microscope. b) Individual atoms identified in an electron microscope using an electron beam.

by the De Broglie equation:

$$p = \frac{h}{\lambda} \qquad (2)$$

where *p* is the particle momentum.

Imagine an electron of kinetic energy 10 eV (what are these units!)[2]. Using classical mechanics, the electron momentum and effective wavelength become:

$$p = m_e \, v = (2m_e T)^{1/2} = 17.08 \times 10^{-25} \, kg.\mu/sec \quad (3)$$

$$l = h/p = 3.88 \times 10^{-10} m$$

As a result, electron microscopes are able to probe atomic distances, as shown in Figure 1b.

Of course, as we go higher in energy, relativistic equations must be used. But you can see that for a particle of 1 TeV (10^{12} eV), the effective wavelength is $\lambda \sim 10^{-18}$ m. By accelerating particles to very high energies, we therefore have probes of very short wavelength. The importance of that comes next.

FUNDAMENTAL PARTICLES AND THEIR INTERACTIONS: WHY WE WANT TO PROBE VERY SMALL DISTANCES

A fundamental particle is defined as a particle without structure that is indivisible. Our understanding of what is indivisible has evolved with time and I would like to explain it. Even the ancient Greeks considered the idea that all matter was composed of fundamental building blocks. Just 100 years ago, the notion was that atoms were indivisible (hence the name: the word "atom" means indivisible). Then in 1911 Rutherford directed a beam of a-particles from a radioactive polonium source, with energies around 5 MeV, at a thin gold foil. As shown in Figure 2a, he observed scattering of the a-particles off the foil, characteristic of a hard core with radius of $\sim 10^{-14}$ m, much less than the radius of $\sim 10^{-10}$ m expected for a gold atom (the effective wavelength of his probe was $\lambda \sim 4 \times 10^{-14}$ m). The atomic nucleus had been discovered — and atoms, evidently, do indeed have constituent parts.

By the 1930s experimental evidence showed the existence of three fundamental particles: protons and neutrons were understood to make up the core of each atom, and a cloud of electrons surrounded the atom as shown in Figure 2b. At that time, with advances in quantum mechanics, Dirac developed a relativistic wave equation for a particle with spin (the electron)[3]. A consequence of the equation was that there should be another particle with exactly the same mass of the electron, but opposite charge (called the anti-particle). Anderson discovered this "positron" in cosmic rays in 1932. The same physics predicted anti-particles for the other particles too, and the anti-proton was discovered in 1954 using the most powerful accelerator at that time.

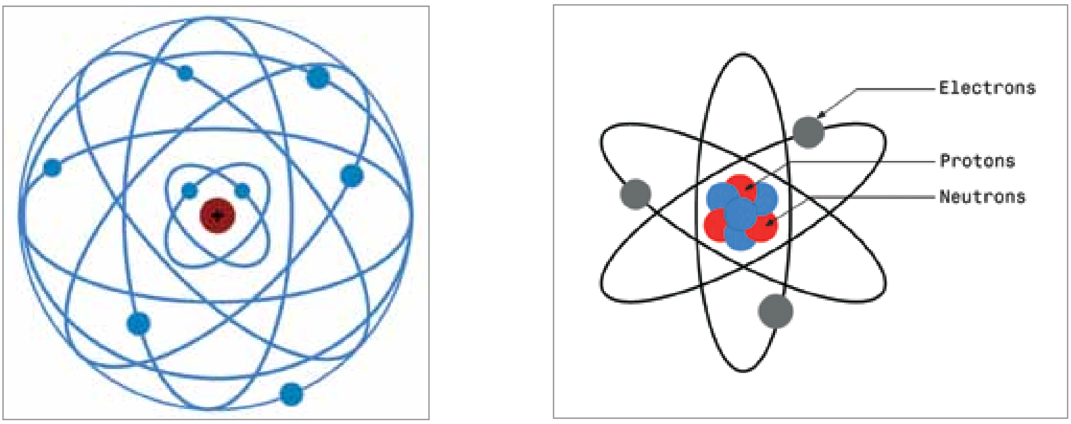


Figure 2. (a) The atom as interpreted by Rutherford in 1911. (b) Our experimental understanding of fundamental particles in the 1930s, with the neutron, proton and electron the fundamental building blocks.

With the invention and development of particle accelerators of increasing energy, physicists were able to probe smaller and smaller distances. The basic ideas was to put the probing particle into the machine, accelerate it to high momentum with a velocity close to the speed of light, and then direct it into a target to observe to result.

I will now introduce a second formula from Einstein, the very famous:

$$E = mc^2 \qquad (4)$$

In this formula, mass is no longer inertial (as it is in Newton's law, $F = ma$), but also has an intrinsic energy. A consequence of this equation is that particles of sufficient energy may produce new massive particles when they collide; conversely, energy can be created from mass. This is a fundamental concept in physics, and is indeed what we observe: in the 1950s and 1960s, a very large number of new particles were identified in the products of collisions in high-energy accelerators.

These particles were organised into two groups, based on their spin angular momentum. One group, the **baryons**, includes all particles with half-integral (1/2, 3/2, ...) spin, also known as fermions — this group includes the electron, proton and neutron. The other group, called **mesons**, contains particles with integral (0, 1, 2, ...) spin, also known as bosons — this group includes the particles known as pions.

Each particle has a characteristic mass, spin and charge. By looking at patterns between the particle properties, and the way the particles behaved (for example, their decays), Gellman and Ne'eman in 1961 proposed a pattern of classification for the known particles. Then in 1964, Gellman and Zweig showed that this pattern could be naturally explained and simplified if the proton and neutron — and, by extension, many other particles — were not fundamental, but instead composite systems of smaller particles called quarks.

In 1968, Friedman, Kendall and Taylor experimentally demonstrated the existence of quarks by firing electrons of energy between 7 and 17 GeV ($\lambda < 2 \times 10^{-16}$ m) at a hydrogen target. As in the case of Rutherford, they measured the scattering of electrons from the target, and the result was compatible with the existence of quarks within the proton; see Figure 2c.

So in 1968 we had a picture of our world where matter is made up of electrons, up-type and down-type quarks. There is also the neutrino, predicted by Pauli in 1932 and discovered in 1956. This is a very good approximation to our Universe today, 13.7 billion years after the Big Bang (though it was not always like that — more on this later). But there are a few rough edges to this physical picture that I should clarify.

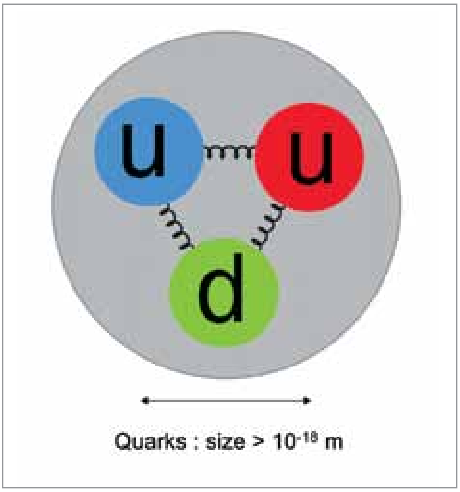


Fig. 2(c) Our experimental understanding of fundamental particles in 1968: the electron was joined by the up-type quark and the down-type quark within the proton and neutron.

Figure 3 (opposite) Fundamental particles of the Standard Model.

The first rough edge concerns the mesons. Physicists observed that mesons are not stable, and are only observed in high-energy collisions with accelerators or cosmic rays. The simplest meson is the pion (π^+, π^0, and π^-) that we now know is composed of a quark and an anti-quark (antiquarks are written with a bar above the symbol, so pions are

$u\bar{d}$, $d\bar{u}$, $u\bar{u}$ *or* $d\bar{d}$).

Furthermore, a special class of particles were identified, initially in cosmic rays and then in accelerators, that became known as "strange" particles (for example, the K-meson). These particles decayed quickly into other particles made of up and down quarks. These strange particles could be explained by the existence of a third quark, dubbed the "strange" or s-quark, with the same properties as the down quark, except for a larger mass. The muon ($\mu^\pm$) was also discovered in cosmic rays in 1947 — this particle is identical to the electron, but again with a larger mass. The s-quark and the muon, so similar to the d-quark and electron respectively, gave an initial hint that there might be several generations of quarks and electron-like particles.

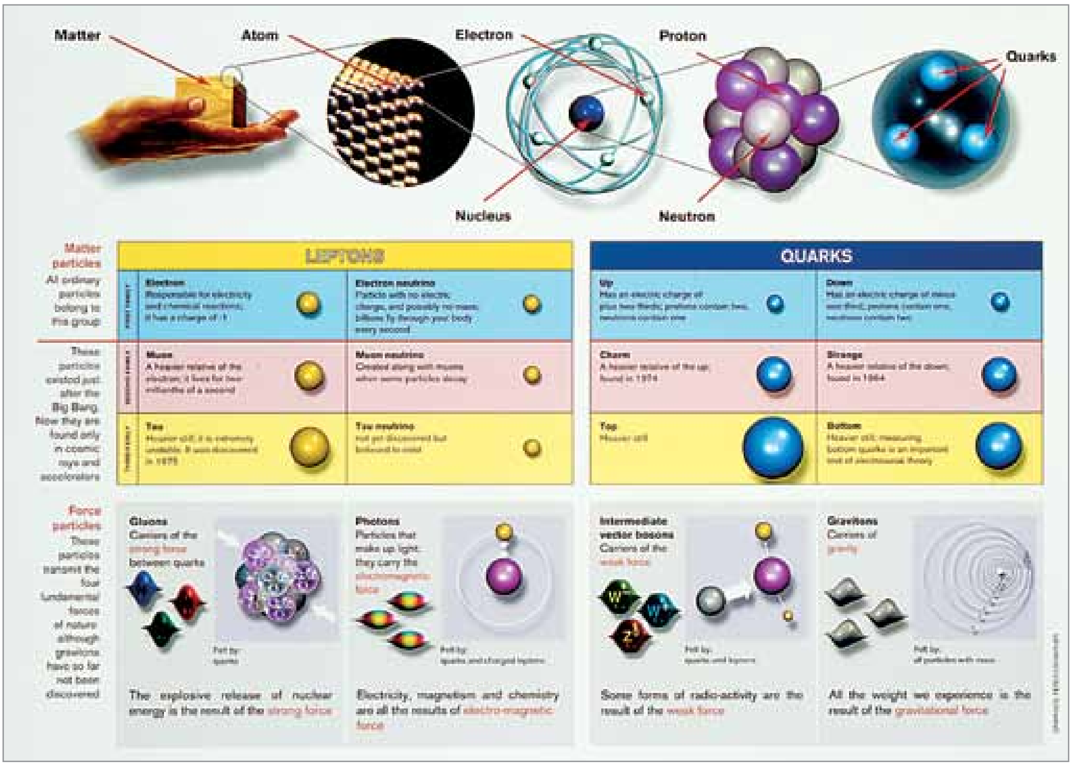
The second rough edge concerns the photon, itself a boson and a fundamental particle. It turns out that the photon is associated with the electromagnetic force, and we explain this in the next section.

THE ZOOLOGY OF THE STANDARD MODEL

The Standard Model (SM) as we know it today is shown in Figure 3. In the first generation of particles, we have a pair of fermions called quarks (u-quark and d-quark) and a pair of fermions called leptons (electron and electron-neutrino). These are the basic building blocks at the typical energy scale of our world, and they form the basis of all chemical and biological systems.

However, experiments over the last 40 years have identified two other generations, also shown in Figure 3, each with a pair of quarks and a pair of leptons. The different generations differ only by their mass. To a good approximation, they do not exist in our world and can only be created in high-energy collisions. When they are created, they don't last long: the second and third generations decay quickly into first generation constituents. In addition, there are the antimatter particles: associated with each fundamental quark or lepton there is a corresponding anti-quark or anti-lepton.

The quarks are sensitive to the four known forces: the gravitational, electromagnetic, weak nuclear and strong nuclear forces. The leptons, however, are not sensitive to strong nuclear forces. The SM explains the interaction of the electromagnetic, weak and strong forces with the quarks and leptons,



but so far gravitational forces are not included in the SM — which is reasonable to some degree, since gravity is so weak at the scale of particles.

In classical physics, a force such as the electromagnetic force is said to act between two particles through a *field*. In the SM, the quarks and leptons interact by the exchange of a so-called gauge boson. The action of an electromagnetic force on a system of charged particles results in the electromagnetic interaction, and the associated *gauge boson* is the photon, as mentioned above. The particles interact a little like they are throwing photons back and forth, like a ball (see Figure 4a). The strength of the interaction is determined by the particle charge. A typical interaction is shown in Figure 4b.

The quarks inside protons and neutrons are held together by the strong nuclear force, which acts on a different kind of particle charge known as "colour" — the action of a "strong field" on a system of "coloured" particles results in the strong interaction, and the associated gauge boson is called the gluon (see Figure 4d). Each quark has one of three colours, and the strength of the strong interaction is determined by the particle's colour.

In the Standard Model, leptons have no colour, and therefore do not couple to the gluon — they are like electromagnetically neutral particles, such neutrons, that do not feel the electromagnetic force.

If we look back to the protons of Figure 2, we see that the u- and d-quarks are held together by the gluons. It is a characteristic of the strong force that the further the quark separation — that is, the more you pull the quarks apart — the stronger the binding force of the gluons. As a result, the quarks are never found individually: you can never pull a quark out by itself. This contrasts with other particles like electrons, which are quite easily ionised from atoms.

Gluons are thought to have zero mass, and the mass of the u- and d-quarks is small — and

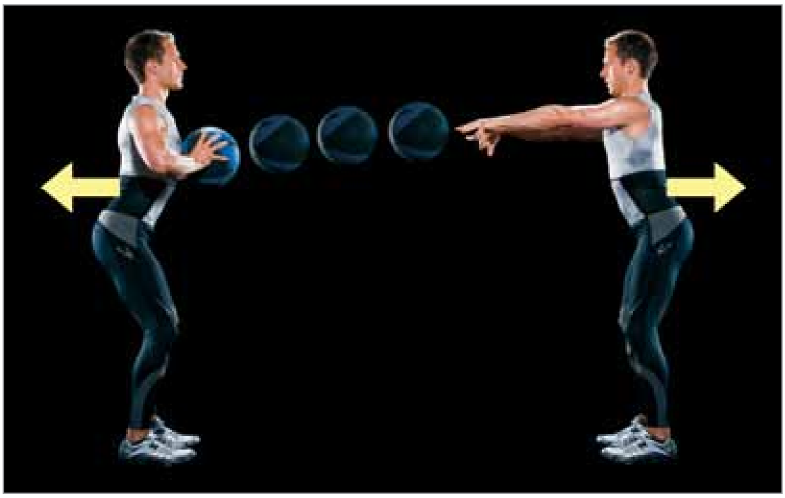


Figure 4. (a) The basic interaction "vertex" is a bit like two people exchanging a basketball: each person experiences a change of energy and momentum when the ball is passed. Opposite we show the basic particle interactions (vertices) for the electromagnetic (b), weak (c) and strong (d) forces.

yet the mass of the proton, which is comprised of quarks and gluons, is around 2,000 times larger than the electron. How can this be? Let us look at Einstein's equation again:

$$m = \frac{E}{c^2} \qquad (5)$$

It is the energy associated with the binding together of the quarks by the gluons that gives the proton its mass. Because the strong force is so strong, the quarks and gluons are held together with so much energy, so the mass of the proton is much greater than the intrinsic mass of its parts.

Finally, there is the "weak nuclear force" that controls the decay of radioactive elements. This interaction is somewhat more complicated: we now know that there are two gauge bosons associated with the weak force, called the $W^{\pm}$ boson and the Z boson (Figure 4b). Unlike the photon and the gluons, the W and Z bosons are very massive, more than 80 times the proton mass.

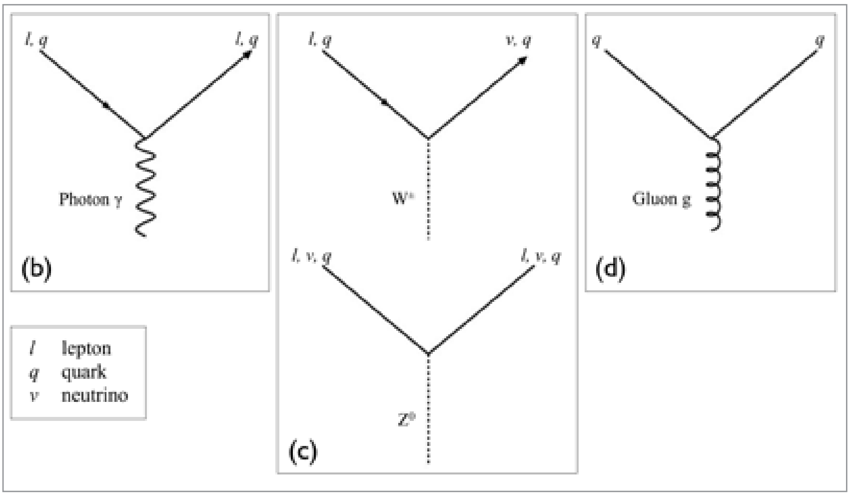
SYMMETRIES AND CONSERVATION LAWS IN THE STANDARD MODEL

An interesting property of this zoo of fundamental particles is that the heavy quarks or leptons always decay to a lighter partner via one of these interactions — *unless some conservation law superimposed on the theory prevents it*.

Conservation laws govern the laws of nature; for example, you are all aware of the conservation of momentum and energy. It turns out that a conservation law is always associated with an underlying *symmetry* in the mathematics of the physical theory — and *vice versa*: a symmetry in the equations always comes with some physical quantity that is conserved.

For example, the conservation of momentum results from what is called the Lorentz symmetry, which is concerned with the coordinate system against which measurements of position and time are taken. You can shift the origin of your coordinates and the physics stays the same: the Universe does not care where your ruler and stopwatch are, or what scale they use. That *symmetry* in the physics leads to conservation of linear momentum. Similarly, you could rotate your coordinates by any angle you like and the physics would not alter: that symmetry gives you conservation of angular momentum.

There are different, subtler kinds of symmetry, called "gauge symmetries", that occur when physica; interactions are independent of the phase of the quantum mechanical wave functions describing the particles. That's hard to describe, but one simple example involves electric potential: you may know that it is always possible to define the zero-point of electric potential where you like, and doing



so does not change the electric or magnetic *fields* involved. This *symmetry* under changes in electric potential leads to a very well known result: the conservation of electric charge.

Several of these gauge symmetries underlie the Standard Model, each ensuring the conservation of a different "charge": the electric and weak charges, and strong "colour".

In electromagnetic and strong interactions, another crucial conservation law is called "flavour conservation". This has the consequence that quarks, for example, must be created as quark-antiquark pairs; similarly, an electron can't be created without an anti-electron (positron).

BREAKING SYMMETRIES AND THE HIGGS PARTICLE

Symmetries are obviously central to physics — much of the richness of physics, however, results from the *breaking* of these symmetries. A broken symmetry refers to the case where a physical system is theoretically symmetrical in some way, but ends up in an apparently non-symmetric state — the fundamental symmetry is hidden, or "broken".

A simple analogy can help explain symmetry breaking: imagine a ball sitting on top of a symmetrical hill, like a Mexican hat. Here the ball is in a symmetrical state — but the slightest bump will cause the ball to roll down the hill. And when it does, the ball will roll down in a specific direction and end up at the bottom of the hill, in a place that is easily identified from all the other possible positions: the original symmetry has been "broken". But the particular direction it rolled was just as likely as any other direction it could have gone, and so the system still retains the basic underlying symmetry; it has simply been hidden!

Symmetry breaking in particle physics seems to be responsible for giving massive particles their mass. The Standard Model's symmetries predict that the weak force's W and Z bosons, as well as all the quarks and

Table 1: Properties of the four fundamental forces of Nature.

	Intensity of Force	Binding Particle (Field Quantum)	Occurs In:
Strong Nuclear Force	~1	Gluons (no mass)	Atomic Nucleus
Electromagnetic Force	$\sim 10^{-3}$	Photons (no mass)	Atomic Shells
Weak Nuclear Force	$\sim 10^{-5}$	Z_0, W^+, W^- (heavy)	Radioactive Beta Decay
Gravitational Force	$\sim 10^{-38}$	Gravitons (?)	Heavenly Bodies

leptons, should have zero mass, just like the photon and gluon. And yet the W and Z are obviously, experimentally, quite massive, so something is wrong. One explanation for this calls on broken symmetry — the underlying mathematical symmetry of the weak force (giving zero-mass W and Z) is broken, putting the world in a state where these particles do have mass. The question is: *how does this symmetry breaking happen?*

A theoretical mechanism proposed by Higgs and others introduces a new field — the Higgs field — that permeates the Universe and interacts with all the fundamental particles. The interaction gives these particles mass, but the underlying gauge symmetry is maintained overall. The effect of the Higgs field can be imagined as a spoon (a particle) in a pot of honey (the Higgs field): movement of the spoon in the honey is difficult, making the spoon appear massive.

In its simplest form, this theory implies the existence of a new, as yet undetected particle: the Higgs boson. To show that the Higgs mechanism is indeed responsible for giving particles their mass, we need to find that Higgs particle — and it is possible we might find it amongst the data we collect at the LHC.

Other conservation laws include invariance of space inversion (a "mirror image" symmetry, called parity), particle-antiparticle transformations (swapping all + charges for – charges, called charge conjugation) and time reversal (running time backwards instead of forwards).

The Electromagnetic and strong interactions respect all of these symmetries, but for some reason both parity (P) and charge conjugation (C) are broken in weak interactions — and it isn't immediately obvious why this would be the case. Importantly, the violation of C, P and the combination of C and P (simply called CP) in weak interactions implies a fundamental difference between particles and antiparticles in nature. We are able to explain the weak interaction, but it is more difficult to explain the physical reasons behind these subtle symmetry problems.

Another consequence of the breaking of flavour symmetry in weak interactions is that quarks and leptons of the second and third generations can decay to the lighter first generation via the weak interaction — but not via the strong or electromagnetic interactions.

From experiments at accelerators since the 1970s, the SM has been verified to very good accuracy, and there is a detailed understanding of each of the interactions with their associated conservation laws, at distance scales of around 10^{-18} m. All of the required fermions and gauge bosons have been identified; there is no serious experimental discrepancy with the Standard Model.

SO WHY BUILD THE LARGE HADRON COLLIDER?

One answer, as we noted above, is the Higgs Particle: this is the only particle of the Standard Model that has not been discovered. If this is the correct mechanism, we believe that it should be discovered at the LHC (see my second chapter).

But that is not the only reason. At first glance, each of the four physical forces seems independent — however, just as Maxwell showed in the 1880s that electric and magnetic forces had a common origin, Glashow, Weinberg and Salam showed that the electromagnetic and weak forces are related in the Standard Model: they are unified into a single *electroweak* force. It is because of the breaking of the electroweak gauge symmetry below a given energy that the W and Z bosons gain their mass.

This success with the electroweak unification leads us to wonder whether there are more surprises hidden in the laws of physics that can't be accounted for in the Standard Model.

THE STANDARD MODEL IS INCOMPLETE

Even if the Higgs mechanism turns out to be correct, the Standard Model throws up other questions that must be answered — to mention just a few:

- The up-quark is very light, the top quark is very heavy — why do the quarks and leptons have such a wide range of masses?

- Why are there three families of quarks and leptons? And do other fundamental particles exist at even higher masses?
- Why are quarks and leptons fermions, while the gauge particles are bosons?
- Is it possible that quarks and leptons are themselves composite, with even smaller "fundamental particles" inside?
- Why is the weak force so different from the strong and electromagnetic forces?
- Why is gravity so much weaker than the other three forces? And can it be unified with other forces?
- Matter and antimatter are reated together, but the universe we live in seems to be almost entirely matter — why is there an asymmetry between matter and anti-matter in the Universe?

Beyond these questions, there are deeper issues. At very high energies (or equivalently at distance scales less than around 10^{-19} metres), the mathematics of the Standard Model becomes unstable; the parameters of the theory require a very artificial degree of fine-tuning to retain a sensible theory.

This indicates that the SM isn't complete, and that there is likely an extension of the SM in the energy range observable by the LHC. One favoured extension is called super-symmetry (SUSY): this is a new kind of symmetry that associates with each fermion (quark or lepton) a new — and as yet unseen — particle of integral spin, and *vice versa*; see Figure 5.

Super-symmetry would resolve many problems of the Standard Model:

- SUSY stabilises the mathematics of the Standard Model, making it work better at high energies; and
- If SUSY particles are produced in particle-antiparticle pairs (another symmetry), then the lightest neutral super-symmetric particle would be a natural candidate for dark matter in the Universe (see the chapter by Fred Watson in this book), since it would be massive, yet would not interact with normal matter other than through gravity.

So far SUSY particles have **not** been detected, and so if they do exist, they must have large masses. But if SUSY particles exist within the mass scale accessible to the LHC, then we should be able to find them — and we will talk about this in the next chapter.

What else is missing in the SM? Though electricity, magnetism and the weak nuclear interactions have all been unified into one "electroweak" theory, the mathematical structure of the Standard Model does not at this stage unify the strong and electroweak interactions, and does not include gravitational interactions. The gauge boson for a quantum theory of gravity is thought to be the graviton, but such a theory would only be observed at the impossibly small distance scales around 10^{-35}m.

A holy grail of physics is to unify the General Relativity of Einstein with a quantum theory of gravity, and to join quantum gravity with the other interactions — so that, beyond some energy scale (see Figure 6), all of the forces would be brought together into a single, complete, unified physical theory. Various unified theories (called Grand Unified Theories, or GUTs) have been developed, but so far none have been fully successful.

One promising framework that aims for this kind of unification is so-called "string theory" that pictures fundamental particles to be string-like vibrating objects in a multi-dimensional space. Such tiny strings would be point-like when projected onto our 3-dimensional space. String theories show much promise, though they have yet to be experimentally confirmed — however, versions of SUSY arise naturally in such theories, and it is a major research activity of theoretical physicists (see below).

THE CONNECTION BETWEEN PARTICLE PHYSICS, THE LHC AND THE UNIVERSE

We already have in the LHC two protons colliding at a combined energy of 7 TeV (7 x 10^{12} eV). To put that in perspective, a 1 TeV collision corresponds to 0.16 x 10^{-6} joules — roughly the total kinetic energy of a flying mosquito. That doesn't sound like much, until you consider that the mosquito is comprised

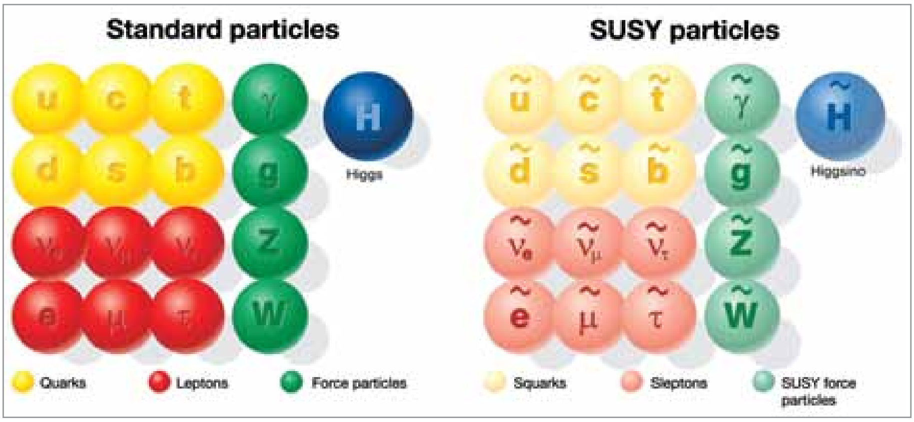


Figure 5. SUSY particles

of about 10^{24} atoms! When that energy is concentrated into a proton collision in a very small volume, then it approximates the conditions of the very first moments of the Universe.

Figure 7 shows a schematic of the evolution of the Universe as a function of time. The *Big Bang* model, modified to include *inflation* in the first moments of the Universe, quantitatively explains the main features of this evolution (though there are, of course, many unresolved questions in this theory). While much of the history of the Universe involves astrophysics, particle physics is relevant in period up to 10^{-10} seconds after the Big Bang.

We are unable to say how the Big Bang started — but we can say that it involved an enormous energy density on a small distance scale, and was therefore inherently quantum in nature. We do not know whether we are a unique Universe or one of many, whether the initial condition was point-like with infinite energy density or a small region of high energy density. Einstein's General Theory of Relativity, which explains the existing data — for example the rate of expansion of the Universe — breaks down at small distance scales because it is not quantised, and as we noted above, a theory of quantum gravity does not yet exist. Hence we need to push the energy frontier in particle physics to higher and higher energies to test our understanding of the Universe.

INFLATION AND DARK MATTER

This chapter is not the place for a description of the Big Bang model, but I would like to say a few more words, because of the strong connection between cosmology and particle physics — and in particular because of two questions relevant to the LHC: the lack of anti-matter in the Universe, and the existence of Dark Matter in the Universe.

In the first 10^{-36} seconds, the Universe was extremely dense and energetic, expanding thanks to a kind of negative pressure. We don't know the details, but we believe that all of the forces were unified as a single GUT field, and all the interactions were massless. At that time, the strong force separated from the rest, and there was a very short period of enormous "inflation": space-time expanded a staggering amount — by a factor of 10^{78} — in the short period of 10^{-36} to 10^{-32} seconds.

We do not fully understand what caused the inflationary field, but the model has been successful. In the period that followed this expansion, the elementary particles and the heavy Higgs, W and Z bosons of the electroweak field were created. This expansion continued until around 10^{-10} or 10^{-11} seconds — and at this age of the Universe, the conditions match those that we can produce with proton collisions at the LHC.

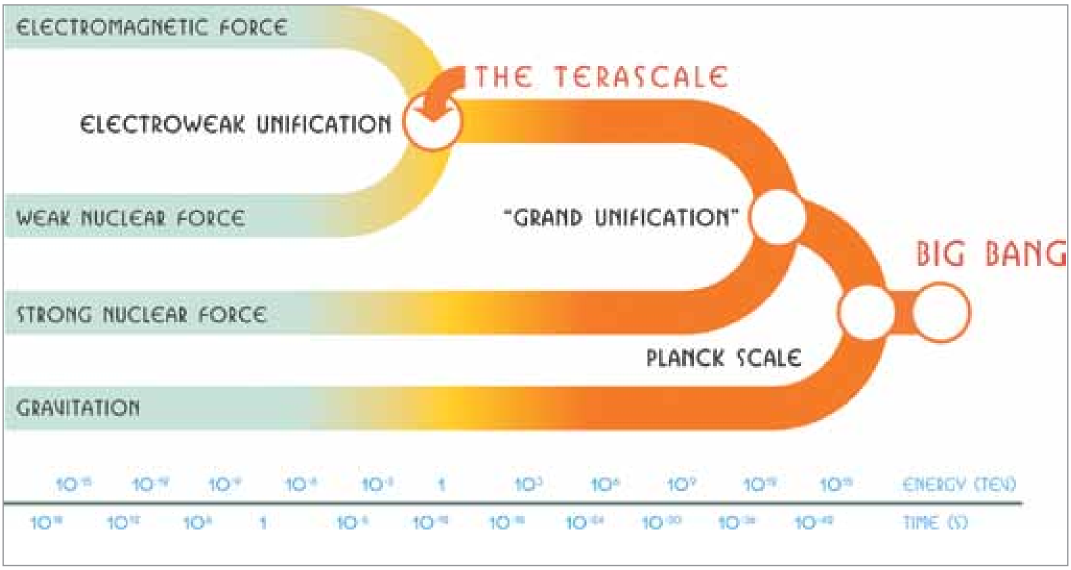
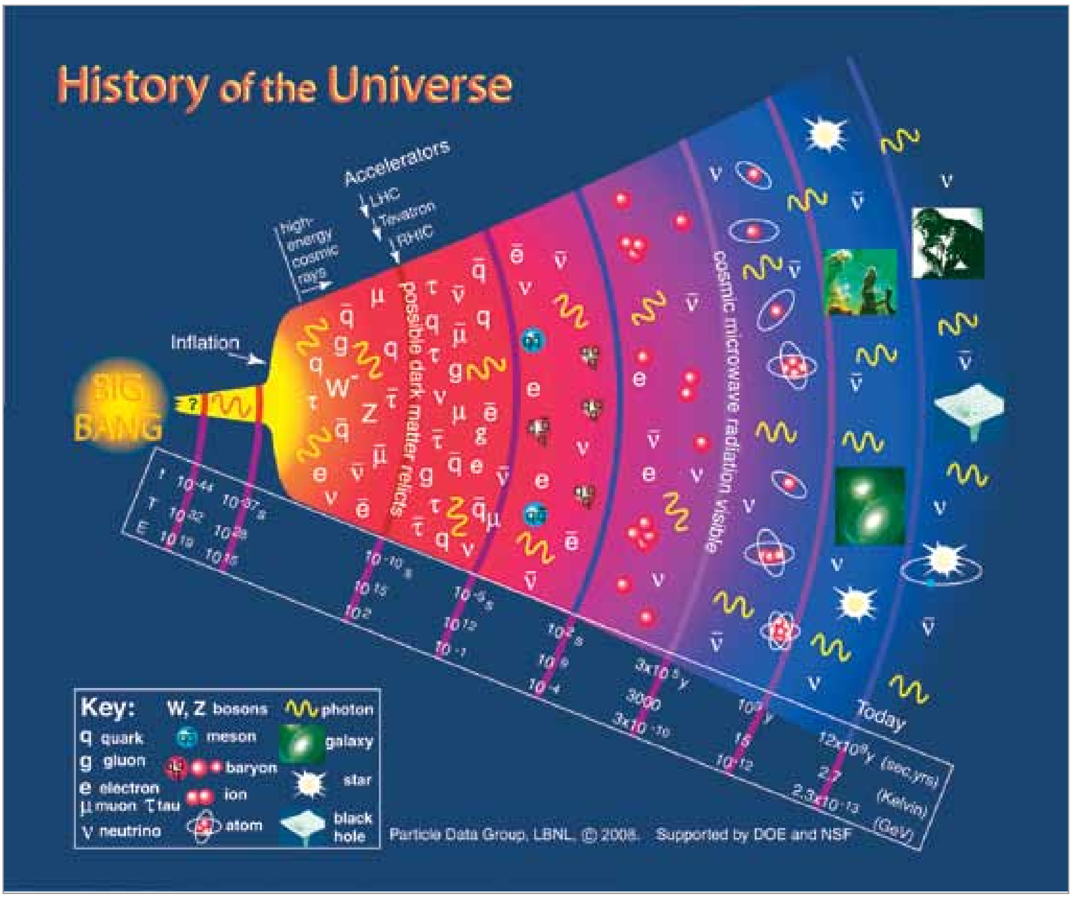


Figure 6. SUSY and coupling constant evolution

Figure 7. Schematic showing the major evolutionary steps of our Universe, and in particular the period up to 10^{-10} seconds after the Big Bang — where collisions at the LHC are thought to produce similar conditions to that of the evolving Universe.



So what can the LHC tell us about these early epochs? First, we know that the strong and electromagnetic interactions force the creation of matter-antimatter pairs — and so the early Universe should have contained equan amounts of matter and antimatter. But so far as we know, the Universe today is dominantly made of matter. What happened to the anti-matter? A partial answer may come from CP-symmetry violation in weak interaction decays, but the numbers do not really fit, and so one of the LHC experiments (called LHCb) will concentrate on CP-violation studies in the b-quark system.

The second question involves Dark Matter. Certain astronomical measurements, such as the rotation of galaxies and clusters of galaxies due to gravity, indicate that an astounding 80% of the matter in the Universe is not visible (see the chapter by Fred Watson). This "Dark Matter" must be massive, but with only a very small probability of interacting otherwise with normal matter. As mentioned above, the lightest neutral particle predicted by Super Symmetric theories would be a good candidate to explain the discrepancy — and other extensions to the Standard Model can also produce compatible results. Experiments at the LHC should be able to shed some light on these (dark) matters.

HOW TO ACCELERATE AND COLLIDE PARTICLES

From equations (1) and (2), you can see that particle accelerators have two main purposes: to provide probes of very short distance (like an incredibly fine-scaled microscope), and to create new massive particles from the available energy during collisions.

The acceleration of a charged particle can occur in an electric field *E*

$$\vec{F}_{elec} = q\vec{E}, \qquad (6)$$

while a magnetic field $\vec{B}$ can be used to guide (bend) the particles,

$$\vec{F}_{mag} = q\vec{v} \times \vec{B}, \qquad (7)$$

where $\vec{\mathbf{v}}$ is the particle velocity (see Figure 8).

In practice, accelerator science is very sophisticated and sits at the extreme limit of available technology — but it is based on just those two equations. An accelerator can be linear or circular; for technical reasons, at the LHC complex the particles start off in a linear accelerator, then progress through two circular accelerators (known as "synchrotrons") before finally being injected into the LHC itself.

A synchrotron accelerator consists of a circular vacuum tube surrounded by a series of bending magnets (dipoles) and focussing magnets (quadrupoles) — together these magnets guide the particles around in a circular path. At one or more places along the circle at each turn, an electric field accelerates the particles in the direction of motion (equation 6) — this gives them a boost of speed (energy). The electric and magnetic fields are synchronised with the particle velocity so that the particles always get a kick in the right direction.

Since the particles travel in a bunch, some are moving slightly slower or faster than others — so to prevent collision losses and to keep the bunch together, the acceleration kick is provided by travelling electromagnetic waves in an accelerating cavity called a klystron. Figure 9 shows a bunch of particles "riding" one of these waves: slower particles are more strongly accelerated and pass the centre of the bunch, while faster particles do not get such a big acceleration and return towards the centre of the bunch. As a result the particles are always kept close to the middle of the bunch, oscillating front-to-back — but the bunch as a whole gets accelerated. For each turn, since the particles are going faster, the magnetic field is increased to keep them in a fixed-radius orbit.

Once the required energy has been reached, the particles can either be ejected from the machine and directed onto a target, or focussed into a very small bunch and collided with another bunch travelling in the opposite direction. The advantage of the latter is that higher collision energies can be obtained

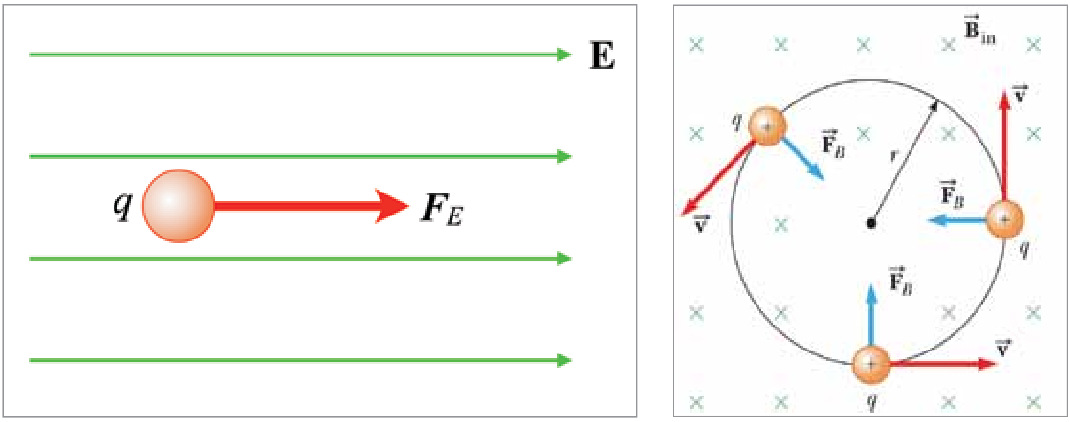


Figure 8. Acceleration of particles in an electric and a magnetic field.

(Figure 10), and this is the principle used in the Large Hadron Collider.

Colliders have been constructed for electron-positron (e^+e^-) collisions, proton-proton (pp) collisions such as the LHC, proton-antiproton ($p\bar{p}$) collisions such as at the Fermilab Tevatron near Chicago, and $e^{\pm}p$ collisions. In the case of e^+e^- or $p\bar{p}$ colliders, a single vacuum tube can be used because the accelerating particles are of opposite charge. However, for pp colliders such as the LHC, two beam pipes are needed to accelerate the two proton beams, which are then brought together for the collisions. From the first particle colliders in the 1960s, the energy involved in the collisions has steadily increased, and the accelerators have grown much larger (Figure 11).

When the beam particles collide at very high energy, many new particles are produced. The “exotic” particles, for example the W, Z and Higgs particles, or mesons and baryons containing second- or third-generation quarks, quickly decay into particles composed of first-generation fundamental constituents. These are observed and recorded in a large detector surrounding the collision point. The particles created in any given collision cannot be predicted ahead of time: while some collisions occur relatively frequently, others — such as the creation of a Higgs particle — are expected to be very rare.

BANG FOR BUCK: GETTING THE MOST OUT OF YOUR COLLIDER

While the energy of collision is one design criteria of a collider, the other design aim is to maximise the rate of collisions, especially if one is looking for rare processes. Intuitively, this is determined by the number of bunches in the machine (n_b), the number of particles in each bunch (N_p), the transverse size (area) of each bunch (A), and the rotation frequency of the bunch (f_{rev}).

Consider a collision of two protons, in which a Higgs particle is produced. This is a rare process, and we measure the probability of it occurring using a quantity called the *cross-section*, *s*. The concept of the cross-section can be understood by analogy: think of trying to collide a very small ball, like a marble, with a billiard ball of radius *R* — in that case the cross-section for the collision will be

$$s = \pi R^2,$$

which is the billiard ball’s area seen front-on by the marble. Here the cross-section is measured in units of area (for example, cm^2).

Now, for a collision between two protons, the concept of the cross-section is the same, except of course it will be a very small number: typical particle collision cross-sections are measured in nanobarns, where 1 barn = 10^{-24} cm^2.

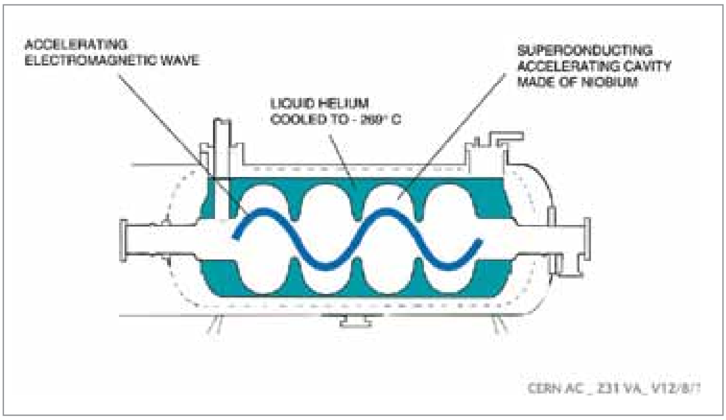


Fig. 9a

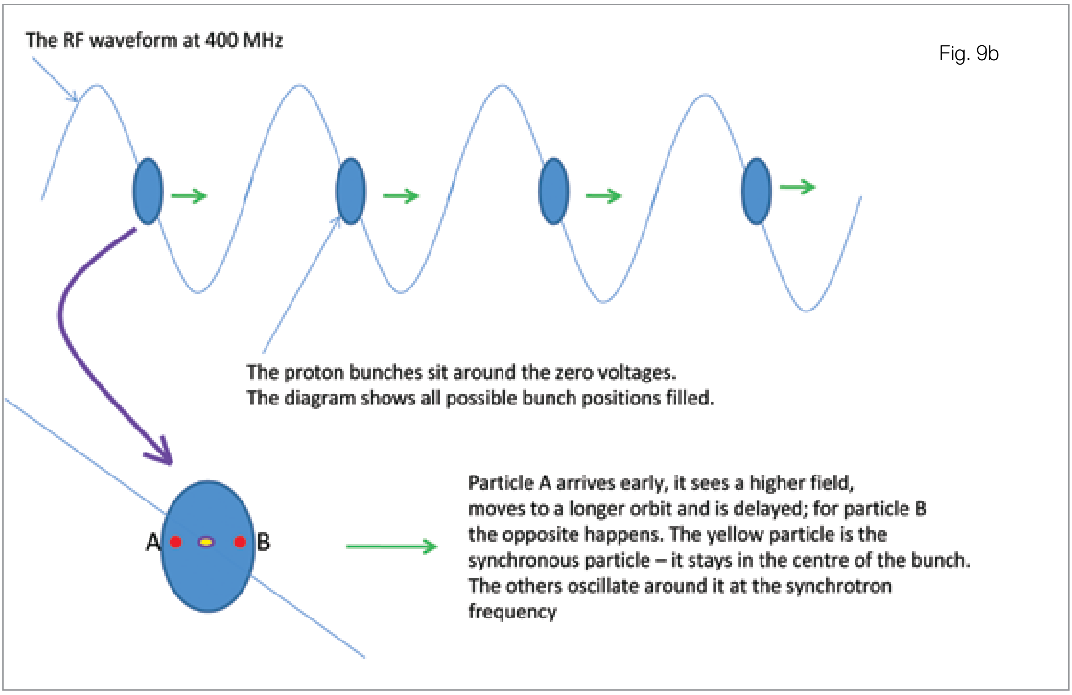
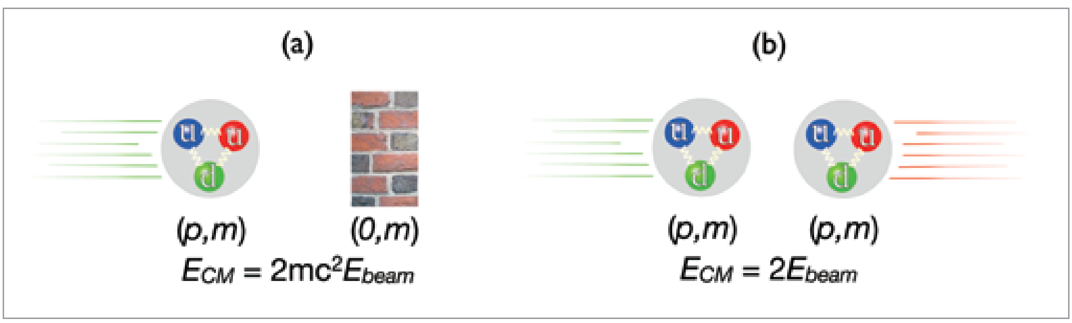


Figure 9. a) An acceleration cavity (klystron) as used for acceleration of protons in the LHC. b) Principle of acceleration for a bunch of particles in an accelerating cavity.

Figure 10. Comparison of the collision energy for (a) a particle of momentum *p* and mass *m* striking a fixed target and (b) two particles of momentum *p* and mass *m* hitting head-on.



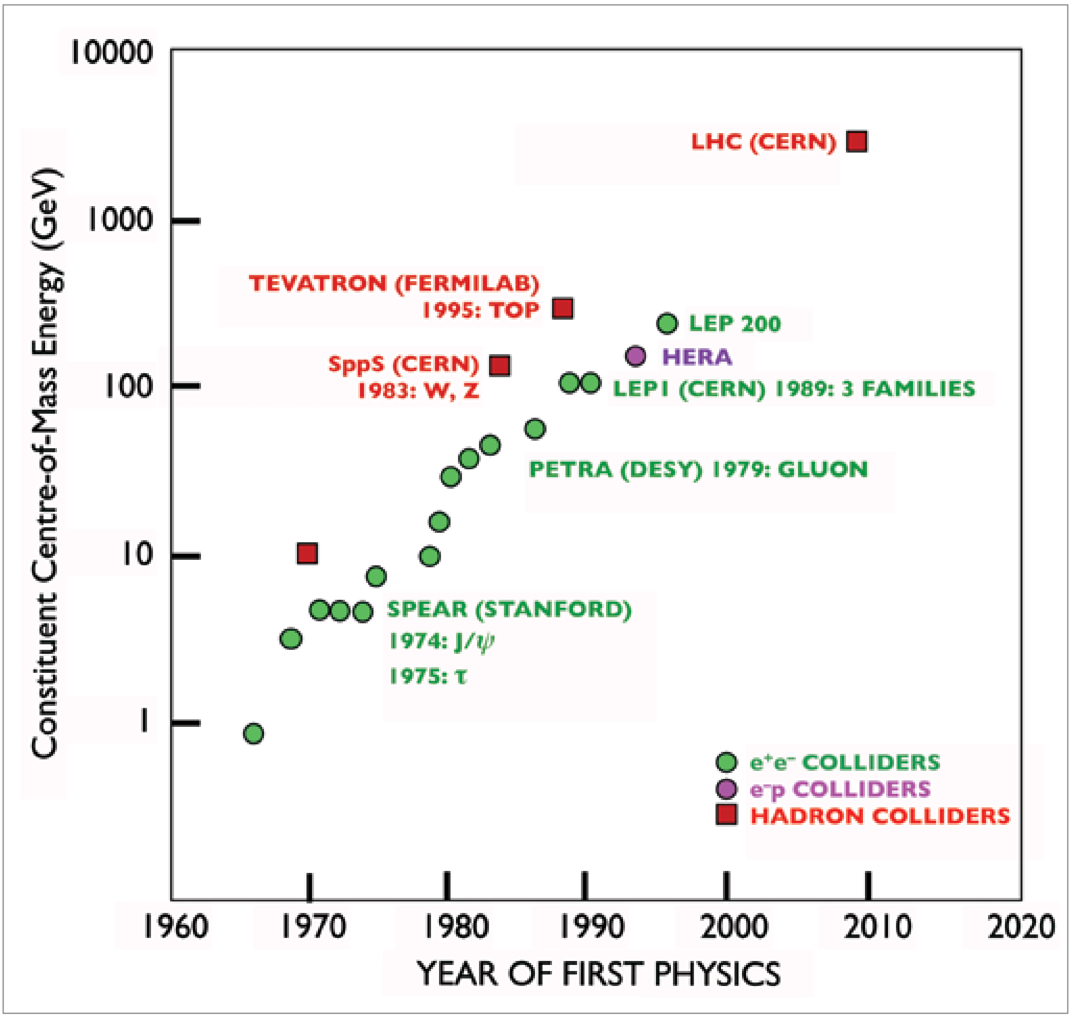


Figure 11. Construction date for successive colliders, shown as a function of their collision energy.

Given a cross-section *s*, the number of interactions per second will be

$$N = Ls \qquad (8)$$

where *L* is called the luminosity (with units of cm^{-2} sec^{-1}), a measure of the number of colliding particles from the two beams (in a unit area, per unit time). Assuming that the beams have an equal number of bunches, each with an equal number of particles and the same bunch surface,

$$L = n_b f_{rev} N_p^2 F / 4\pi A^2. \qquad (9)$$

F is a small correction for when the beams do not collide completely head-on, as is the case at the LHC. The luminosity *L* is completely determined by the machine conditions — and so our aim is to maximise it.

THE LARGE HADRON COLLIDER (LHC)

The LHC is the largest accelerator ever built. It is located in a tunnel 26.7 km in circumference at the Franco-Swiss border, buried between 45 and 170 metres underground, and tilted at 1.4° from the horizontal for geological reasons. The tunnel was constructed in the 1980s for the LEP (Large Electron-Positron) collider (see Figures 12 and 13). The ring circumference determines the beam energy, and the magnets have

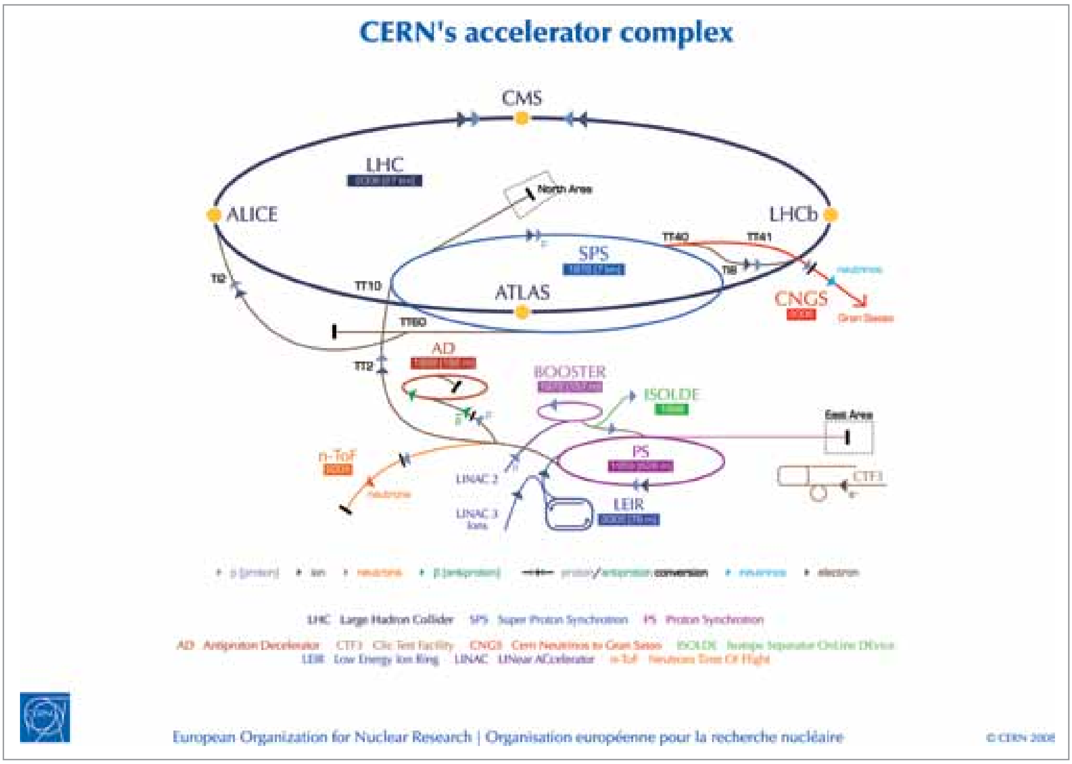


Figure 12. An illustration of the CERN accelerator complex.

been built to have the highest possible (robust) magnetic field using niobium-titanium superconductor[4] technology.

A total of 1,232 magnetic dipoles, each 16.5 metres long with a maximum magnetic field of 8.33 Tesla[5], operate at 1.9° above absolute zero — this makes the LHC magnets one of the coldest places in the Universe, requiring the world's largest refrigeration plant. The magnets guide the protons in a circular path, and will eventually allow a maximum beam energy of 7 TeV (so a proton-proton collision energy of 14 TeV). In addition, more than 4,800 additional focussing and correction magnets are installed — and all of these magnets need to work in unison!

The protons are accelerated in two vacuum tubes, in which the vacuum is maintained around ten times lower than that found on the moon. To reduce the cost, single dipole magnets are designed to have opposing fields at the position of the two beam pipes, as shown in Figure 14. At one position along the machine, the protons enter a cavity with a 400 MHz radio-frequency electric field — this gives the protons an acceleration boost at each turn as they circulate. At design intensity, 2,808 bunches will be circulating in each beam — each bunch containing more than 10^{11} protons, and separated by just 25 nanoseconds. An illustration of the machine is shown in Figure 15.

To maximise the luminosity, the transverse size of the bunches is reduced at each collision point, by squeezing the beams with the field from the quadrupole magnets, similar to the way light is focussed by a lens. After squeezing, the transverse beam size is around 17 μm — but transverse squeezing ends up stretches the bunches lengthways, to around 7.5 cm. Eventually, the LHC will achieve instantaneous luminosities of more than 10^{34} $cm^{-2}s^{-1}$ — which is a very large number of colliding protons!

CONCLUSION TO PART I

The LHC started operation in December 2009, and in March 2010 the first collisions were produced at an energy of 7 TeV. This

Figure 13. The LHC superimposed on the terrain of the Geneva region.

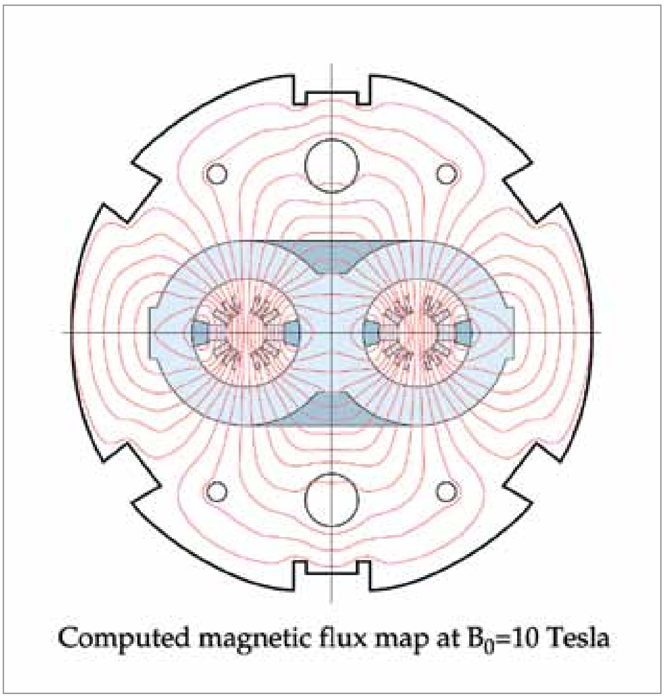


Figure 14. A schematic of the magnetic field of the LHC dipole magnets.

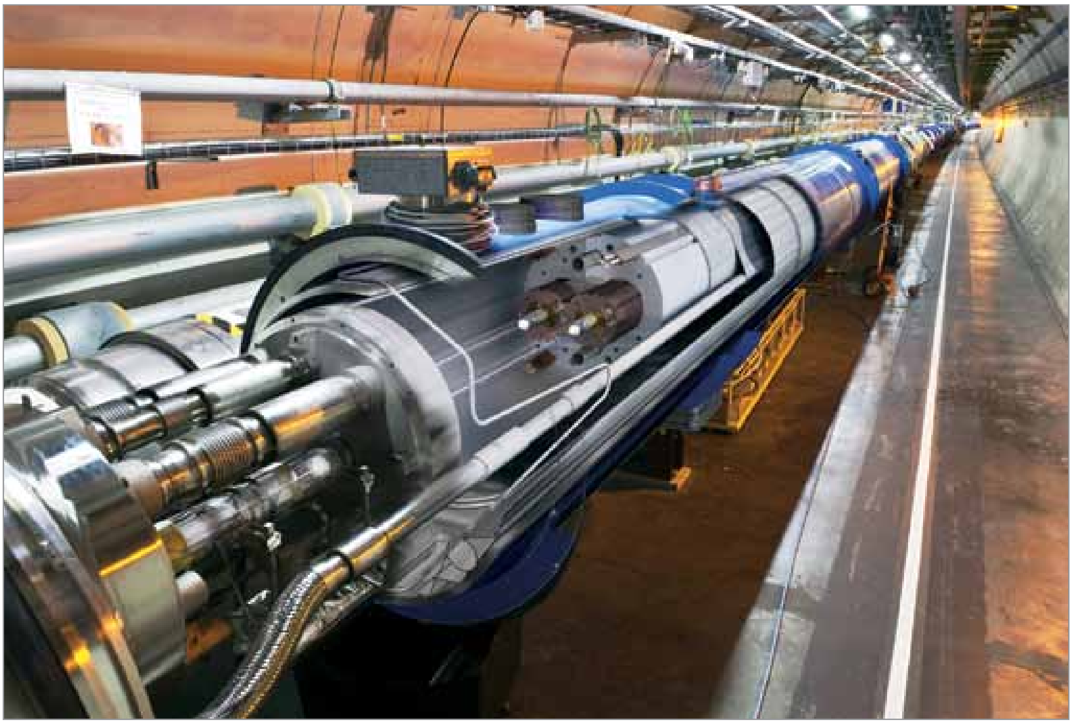


Figure 15. Photograph of the LHC collider installation, with illustrated cut-away showing the particle beams.

was a historic day, as the collision energy was 3.5 times higher than any previous collisions at an accelerator. Early on, these collisions involved a single small bunch of protons in each direction, but over the following months both the number and intensity of the bunches steadily increased.

Luminosities greater than 10^{32} $cm^{-2}s^{-1}$ have already been achieved — let's put that into perspective. Say you want to study a process, some kind of particle event resulting from a proton collision, that has a cross-section of 10^{-32} cm^2 (or 10 nanobarns) — so you expect to see one event per second of that process occurring at this luminosity.

But for all possible processes occurring from the proton collisions, the total cross-section is nearly 10^{-25} cm^2 — 10 million times higher than the one process we're interested in. This means that, to study such an event, we have to identify it from 10 million other events. That is the challenge of searches at the LHC: finding an incredibly small needle in a very large haystack of collision data!

The collection of data at 7 TeV collision energy will continue until the end of 2012. The LHC will then stop for about one year for modifications to increase the collision energy to 14 TeV.

For one month until early December 2010, the LHC was also used to collide bunches of lead ions (Pb) together — at an energy of 287 TeV per beam! We will talk about the analysis of data from the LHC in our next lecture. As a preview, however, remember the equation $E=mc^2$. The energy of collision can result in the production of many new particles — Figure 16 shows the particles reconstructed by the ATLAS experiment in just one Pb-Pb collision that was recorded.

To conclude, the basis of particle accelerators and colliders is simple —charged particles are accelerated with an electric field, and magnetic fields both guide the particles in circular paths, focus them for collision. The reality, though, is much more complex: the technology of creating and synchronising all of these fields is an amazing technological achievement.

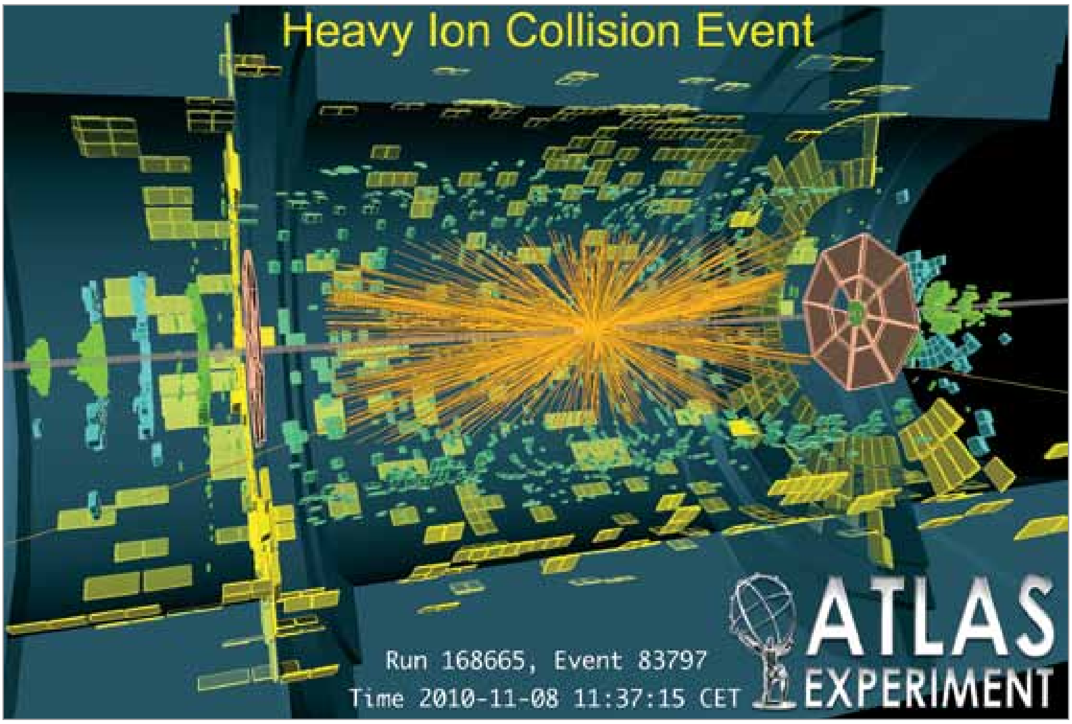


Figure 16. Reconstruction of a lead-lead collision recorded by the ATLAS experiment at the CERN LHC

Footnotes

1 h = 6.626 x 10^{-34} joule second

2 An electron volt (eV) is the energy gained by an electron as it is accelerated between two plates with a voltage difference of 1 volt (1 eV = 1.602 x 10^{-19} joules)

3 Spin is a property associated with fundamental particles that is similar to the angular momentum of an electron orbiting around a nucleus, except that it is intrinsic to the particle, a bit like a spinning top. In quantum mechanics, it takes discrete values,

$$\vec{s} = \frac{nh}{2}\hat{n}; n = 1,3,...$$

for fermions and

$$\vec{s} = nh\hat{n}; n = 0,1,...$$

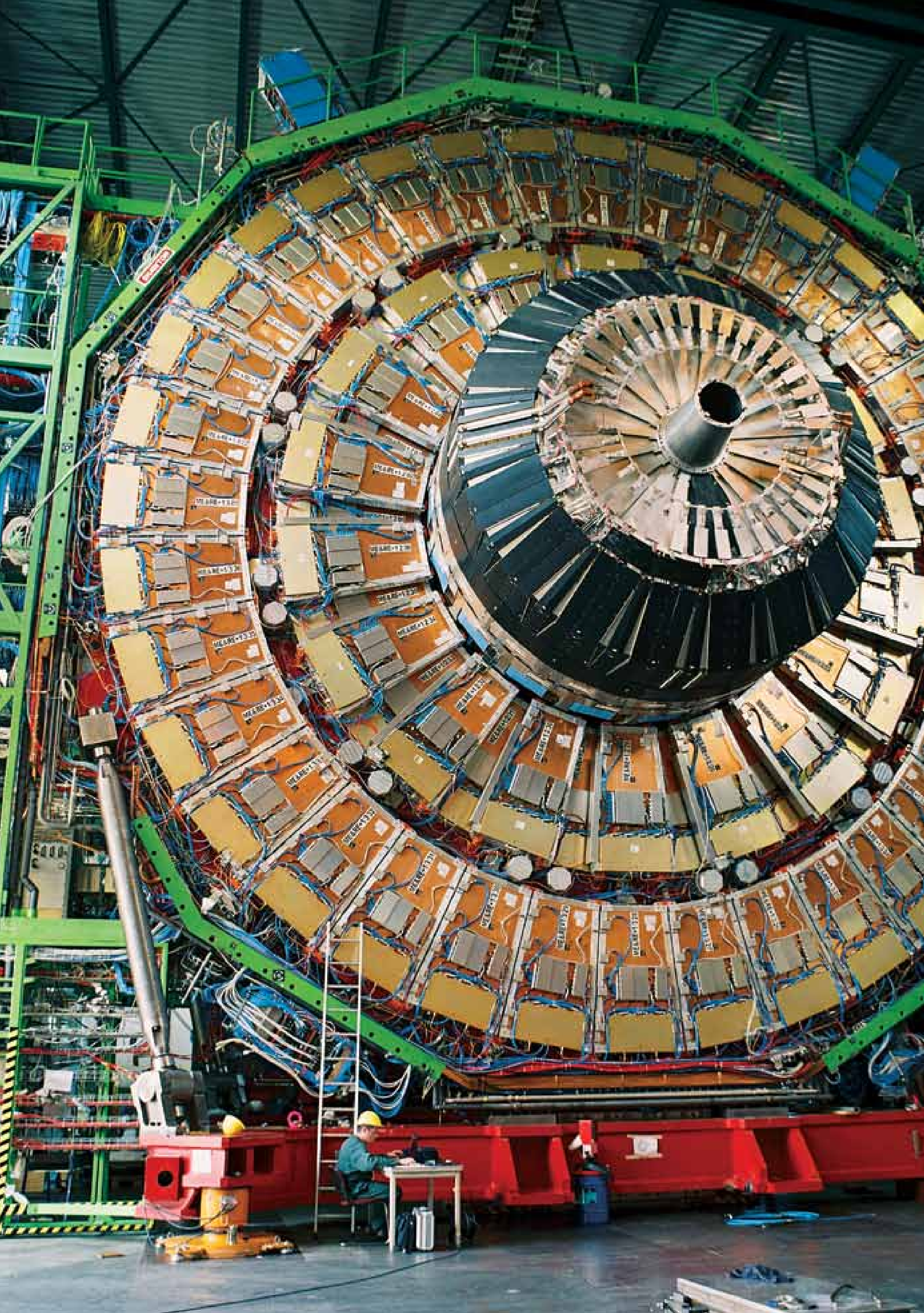
for bosons.

4 Below a critical temperature, some materials undergo a phase transition to become superconductors, allowing electric current to flow with almost no resistance. These superconductors can be used to create very large magnetic fields.

5 This is about 170,000 times stronger than the magnetic field at the Earth's surface.

References & Further Reading

1 *The Particle Adventure*, http://www.particleadventure.org/
2 The Large Hadron Collider web-site:
http://public.web.cern.ch/public/
http://lhc.web.cern.ch/lhc/
http://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/
3 *The Large Hadron Collider: a Marvel of Technology*, ed. Lyndon Evans, EPFL Press (2009).



A VERY LARGE MICROSCOPE TO PROBE VERY SMALL DISTANCES — PART II

PROFESSOR ALLAN G. CLARK

INTRODUCTION

In this chapter, we discuss the particle physics experiments at the CERN Large Hadron Collider (LHC), with particular emphasis on the ATLAS experiment in which Australian researchers are very actively involved. After a tour around the LHC to identify the experiments, we will show in more detail how the experiments work, and then describe the road from a collision to an experimental result.

A major initial aim of the LHC physics exploration will be to discover and study the Higgs particle — if it exists. A second aim will be to search for extensions of the Standard Model that might explain Dark Matter in the Universe (see also the chapter by Fred Watson in this book). We will show how these searches are carried out, the results already obtained, and a menu for the future.

EXPERIMENTS AT THE CERN LHC

As we described in the previous chapter, Figure 1 shows the LHC ring that sits between 45 and 170 metres below the Geneva landscape. When we follow the LHC ring around its 26.7 km length, we find 4 places where quadrupole magnets focus the proton bunches in each ring to a very small frontal area and cause them to collide. When the LHC reaches its specified design intensity, every 25 nanoseconds (or 40 million times a second) two proton bunches will collide, each containing more than 10^{11} protons. On average, about 20 protons within each bunch will collide; the remaining protons in each bunch continue unaffected.

The particles produced in each collision are tracked in very large particle detectors surrounding each collision region. The two largest detectors are ATLAS and CMS, whose schematics are shown in Figures 1 to 4. These two general-purpose detectors have been designed to be able to discover the Higgs particle (if it exists), and are also optimised to search for new physics at the TeV scale. Another experiment, called LHCb, is specifically designed to study the physics of *b-quark* production, and the ALICE experiment is studies the collision of lead ions. There are in addition several smaller, specialised experiments.

HOW DETECTORS WORK: THE EXAMPLE OF ATLAS AND CMS

Imagine a *pp* collision in the ATLAS or CMS experiments, as shown schematically in Figure 6. Indeed two *partons* (a name referring to either quarks or gluons) within each proton will interact and the energy of the collision — by $E=mc^2$ — will allow the creation of

Figure 1: The CMS detector before testing using cosmic rays. After assembling CMS, a section of the detector was operated to check the alignment and functionality of the magnets and detector systems (copyright © CERN, 2011, CERN-EX-0607006).

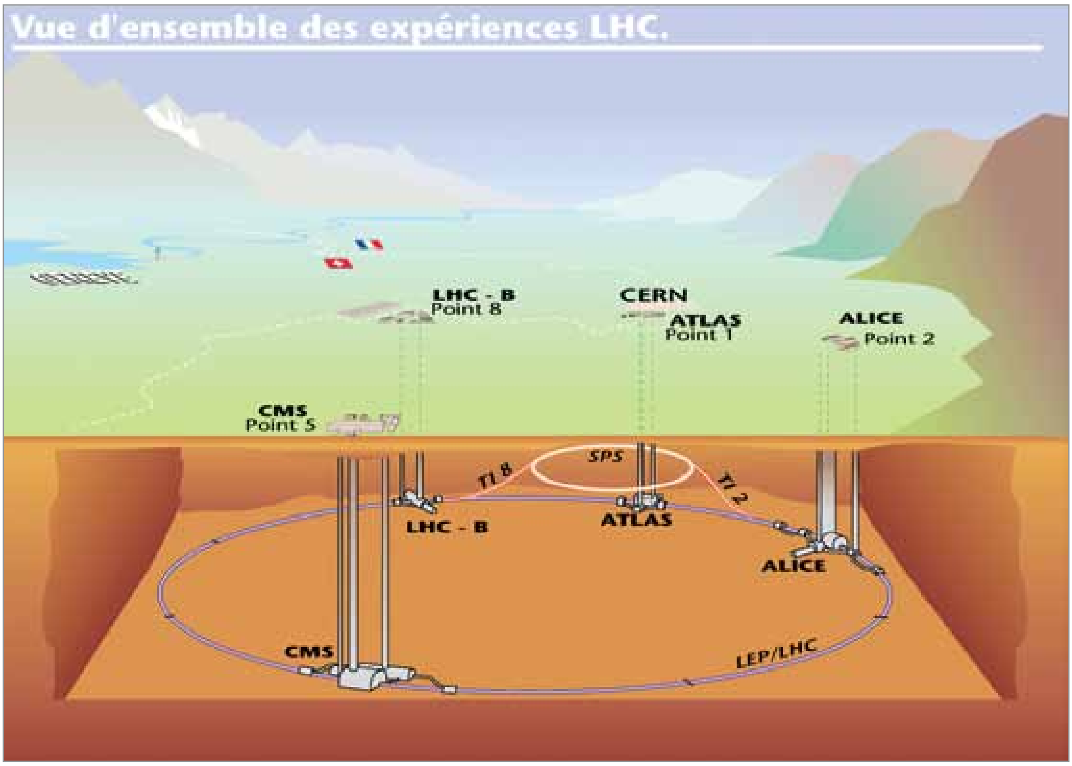


Figure 2. The four main experiments — ALICE, ATLAS, CMS and LHCb — at the LHC, located between 45 m and 170 m underground. Huge caverns have been excavated to house the detectors. The SPS, the final link in the pre-acceleration chain, and its connection tunnels to the LHC are also shown (copyright © CERN, 2011, LHC-PHO-1998-349).

new final state fundamental particles. These will (mostly) travel at close to the speed of light outwards from the collision point. The extraordinary energy density at the collision is very much like the early moments of the Universe.

Any neutrinos or charged leptons produced in the interaction will traverse the detectors of the experiment, provided that they do not decay before reaching the detectors — for example, most muons will traverse the detectors before they decay, but tau leptons will decay before reaching the detectors.

Any quarks or gluons that are produced will be sensitive to strong forces and must "dress up" into bound states (baryons or mesons) as they leave the interaction point — quarks and gluons are **never** found on their own. The baryons or mesons are typically unstable, and will rapidly decay via strong, weak or electromagnetic interactions, to the lightest known baryons or mesons.

Only particles that have a sufficiently long lifetime will traverse the detector. Within the particle zoo of the Standard Model, these are:

- baryons : protons and neutrons;
- mesons: pions ($\pi^{\pm}$) plus a small fraction of kaons ($K^{\pm}$, K^0), but not the neutral pion, π^0 (see below);
- leptons: electrons and positrons ($e^{\pm}$), muons ($\mu^{\pm}$) and neutrinos (ν_e, ν_m, ν_t);
- photons: the π^0 mainly decays into 2 photons after an average distance of just 25 nm.

It is the task of the experimenter to reconstruct the initial collision from what is recorded in the detectors of the experiment by the traversed particles. To understand if it is an interesting event (for example, the production of a Higgs particle), it is necessary to identify and measure as many of the outgoing particles as possible.

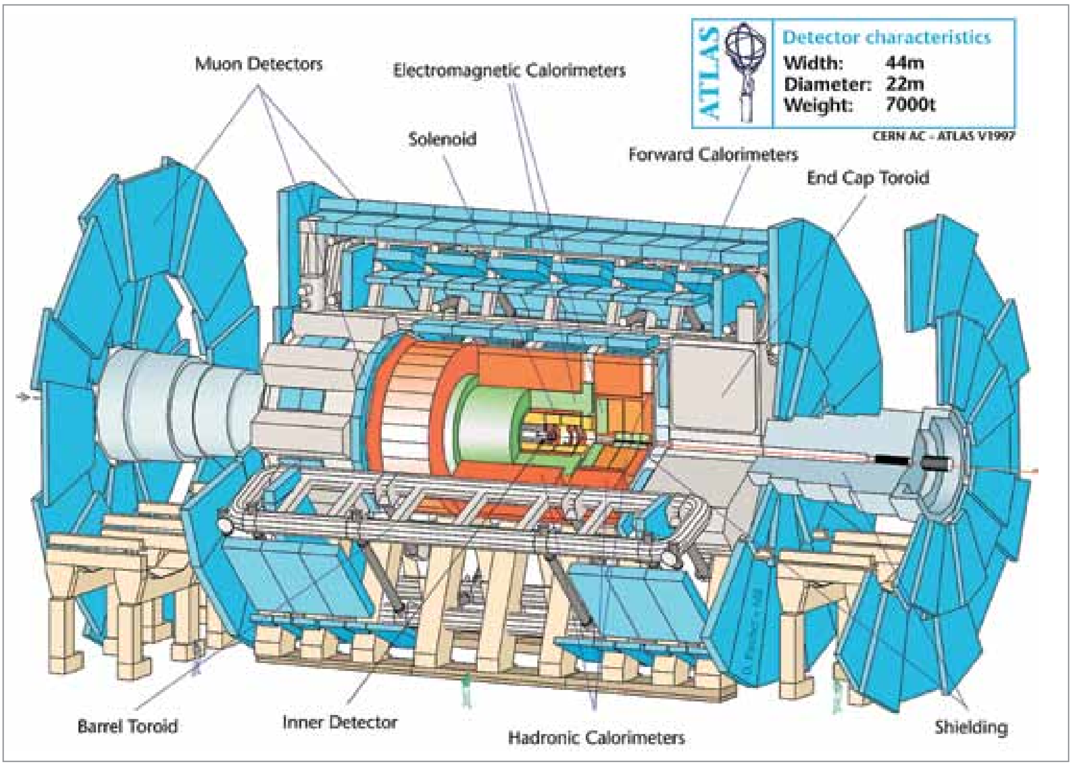


Figure 3. A cutaway 3-dimensional drawing of the ATLAS experiment showing the different detector components. The size of the detector can be judged by the scale drawings of people in the diagram, at the base of the detector (copyright © CERN, 2011, CERN-DI-9803026).

Another interesting thing to note from Figure 6 is that only part of the *pp* collision energy is involved in the *parton-parton* collision, and event-by-event this is not deterministic. Although every effort (Figures 3 and 4) is made to cover the full solid angle around the interaction with detectors, there will always be some particles that do not traverse the detector (for example, continuing along the beam pipe, or through gaps and joins of the detectors) and it is the task of the experimenter to understand all of those details.

On average, there will be around 20 proton collisions in a bunch crossing. Every collision will be different, and the relative probability of any given collision is defined by the cross-section. Figure 7 shows the cross-section s (y-axis) expected from the Standard Model for various processes, as a function of the collision energy (x-axis). The total cross-section (that is, *pp* → *anything*) is at the top; the cross section for *pp* → *b-quark* + *anything* is in red; the W and Z boson cross-sections (*pp* → W or Z + *anything*) are in green, and so on.

At the bottom of the figure, the brown lines are theoretical predications for the cross-sections for the Higgs particle (*H*): three different lines, for three different possible Higgs masses (M_H = 120 GeV, 200 GeV and 500 GeV).

Carefully reading this graph shows how relatively likely the different events are. For example, at 7 TeV collision energy (the middle vertical dotted line), the relative chances of the following interactions are shown in equation 1:

$$\mathrm{s}\,(pp \rightarrow anything) : \mathrm{s}\,(pp \rightarrow Z + anything) :$$
$$\mathrm{s}\,(pp \rightarrow H_{m=120GeV} + anything)$$
$$\approx 1 : 2\times10^{-7} : 5\times10^{-11} \qquad (1)$$

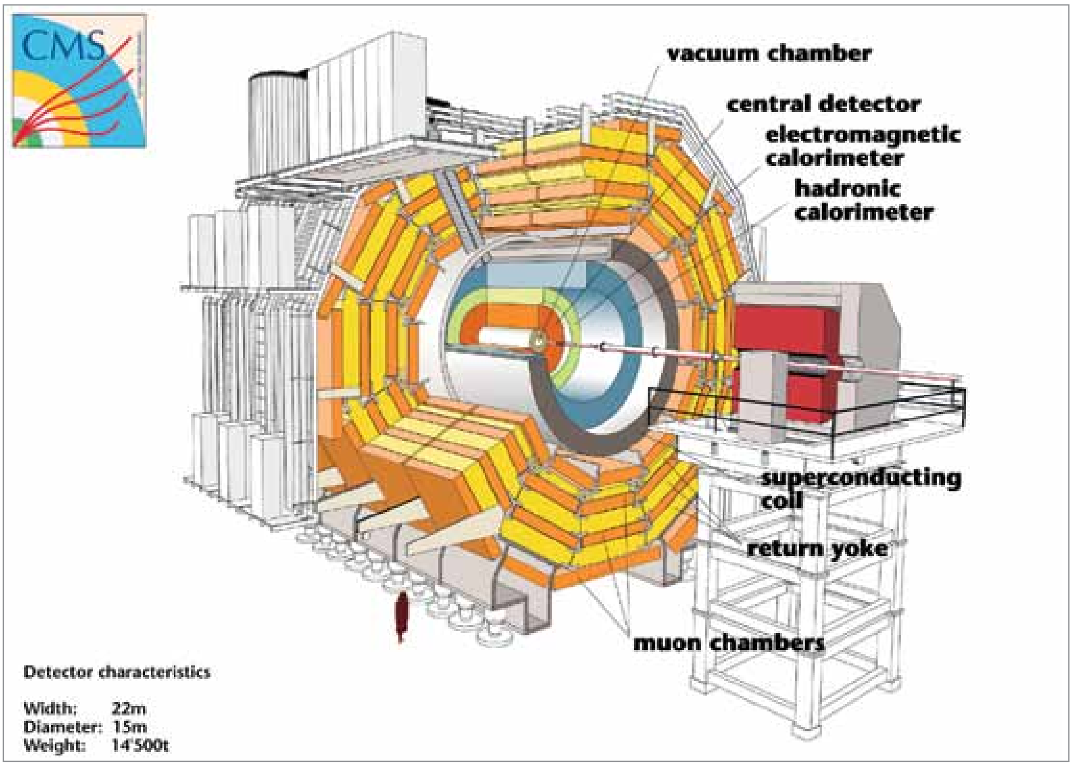


Figure 4. A cutaway 3-dimensional view of the CMS experiment, showing the outer four layers for detecting muons (interleaved with three layers of iron), the central calorimeters and the inner tracking system. Note the height of the people around this detector (copyright © CERN, 2011, CERN-DI-9803027).

As you can see, the most interesting processes are often extremely rare.

Taking as an example the ATLAS detector, Figure 8 shows a slice of the detector perpendicular to the direction of the beam particles. Surrounding the collision point are successive layers of detectors, very much like a "Russian doll". The different layers have very specific functions that I describe below.

Figure 9 shows a circular tube filled with gas (this is exactly the case for the muon detectors of ATLAS and CMS, and for the Transition Radiation Tracker of ATLAS). The passage of a charged particle ionises the gas. Now, suppose there is a thin wire inserted in the tube. Keeping the tube at ground or negative voltage and the wire at a positive voltage, the electrons are attracted to the wire and the charged ions drift to the wall of the tube — a current flows and an electrical signal is produced. If, instead of a gas, the detector medium is solid or liquid, the same principle applies.

In the case of neutral particles (photons or neutrons), there will be no ionisation. But photons of sufficient energy, in the presence of matter, may produce an electron-positron pair, and neutrons will interact if they collide with the nucleus of an atom (a strong interaction), producing charged particles. In either case, the resulting charges cause an electrical signal.

Now let us look again at the "Russian doll" of Figure 8. Every 25 nano-seconds at least one collision will occur, with the production of particles in the final state. These particles traverse the different layers of the detector. For both the ATLAS and CMS detectors, the layers are as follows :

- A very highly granular inner tracking detector, immersed in a magnetic field. In the case of ATLAS, the three inner layers

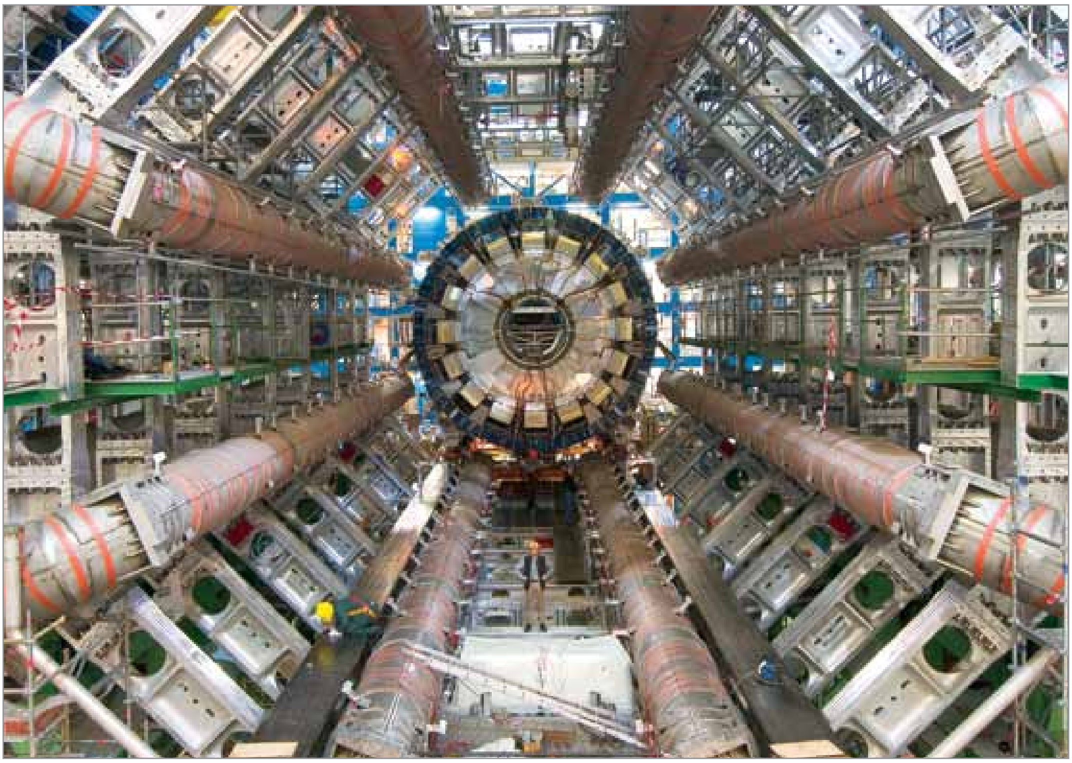


Figure 5. The ATLAS detector during assembly. A total of 8 superconducting toroidal magnets surround a calorimeter, which was later inserted into the detector and is used to measure the energy of particles produced in pp collisions (copyright © CERN, 2011, CERN-EX-0511013).

are made from around 80 million silicon pixels, each 0.05 x 0.4 mm^2 in size. These are followed by 8 layers of silicon strip detectors, with a granularity of 0.08 x 120 mm^2, and then by many gaseous tube detector layers that operate as described in Figure 9.

Figure 10a shows a 3-dimensional scale drawing of the different tracker elements, and Figure 10b shows a photograph of the silicon strip tracker during assembly. Charged tracks leave a signal in the traversed elements of each layer. The electrical signals (see below) are digitised and the information transferred to a computer where the hits are used to reconstruct all the charged tracks. Since the tracks are bent by the magnetic field, the associated track momentum can be calculated. The accuracy of the measurements depends on the precision of individual measurements, and the strength of the magnetic field. Neutral tracks will not (to first approximation) be detected.

- The particles then enter what we call an *electromagnetic calorimeter*. The details of the CMS and ATLAS calorimeters differ, but the principle is the same. Electrons or photons encounter a heavy material (thin lead plates for ATLAS, and a lead-tungstenate glass for CMS), resulting in a shower of secondary particles and the loss of all the particle energy, as shown in Figure 8. A measurement of the electrical signal (ATLAS) or the optical signal (CMS) allows a measurement of the particle energy. Of course, in both cases the calorimeters are very segmented, allowing the energy reconstruction of many individual particles of the collision.
- Following the electromagnetic calorimeter, a segmented hadronic calorimeter (usually made from thin iron plates interleaved with

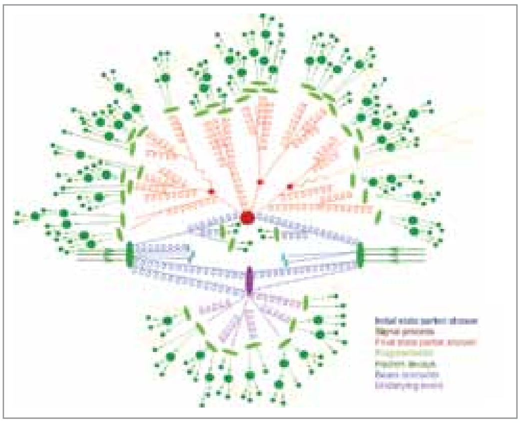


Figure 6. An interaction of 2 protons

a detection medium) is used to measure the energy of hadrons, such as charged pions, protons or neutrons. These initially interact in the iron plates via strong interactions and then create a shower of secondary particles whose ionisation is detected in the detection medium.

- Finally, muons normally traverse the full detector, leaving the signal of a single particle, before making a signal in gas tube detectors installed behind the hadron calorimeter. In the case of ATLAS, the 8 large superconducting coils shown in Figures 2 and 4 provide an additional independent measurement of the muon momentum.

As seen in Figure 8, the different detector signals enable the identification of photons, electrons, muons, protons, neutrons and charged pions. But one important known particle is missing: the neutrino. The neutrino leaves no signal in the detectors; since it is neutral, there is no signal in the tracking detectors, and since it is sensitive only to weak interactions, the probability of an interaction in the calorimeters is very small. However, a fundamental tenant of particle physics is the conservation of momentum in an interaction, and so provided all the other particles are measured, the contribution from neutrinos can be interpreted[1].

The missing energy of reconstructed objects like neutrinos in the direction transverse to

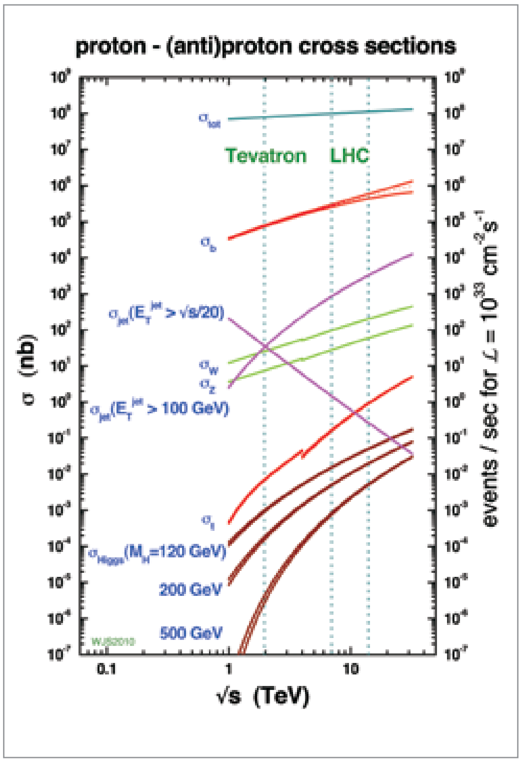


Figure 7. Cross sections for some interesting and proposed physics processes as a function of the collision energy. Cross sections for $p\bar{p}$ colliders are shown for collision energies below ~4 TeV, and for pp collisions above that value. The vertical lines indicate the collision energy of the Fermilab Tevatron (1.96 TeV), the current running energy of the LHC (7 TeV), and two possible future running energies of the LHC (10 and 14 TeV). The cross sections shown are, from largest to smallest, the total scattering cross section (σ_{tot}), the bottom-quark production cross section (σ_b), the production cross sections for hadronic jets above several different energy values (σ_{jet}), the W, Z, and top-quark production cross sections (σ_W, σ_Z, σ_t), and the production cross sections for a theoretical Standard Model Higgs boson at several masses (σ_{Higgs}). From W. Stirling, 2009.

the beam axis is called E_T^{miss}. Of course, if there are two neutrinos in the final state, only the summed effect can be measured — we can't tell how much energy was with each nuetrino. And as we will discover later when we search for new particles at the TeV scale, any massive new particle that traverses the experiment undetected will also, in most cases, be misidentified as a neutrino.

There is one very spectacular detector signal that I have not so far discussed: *hadronic*

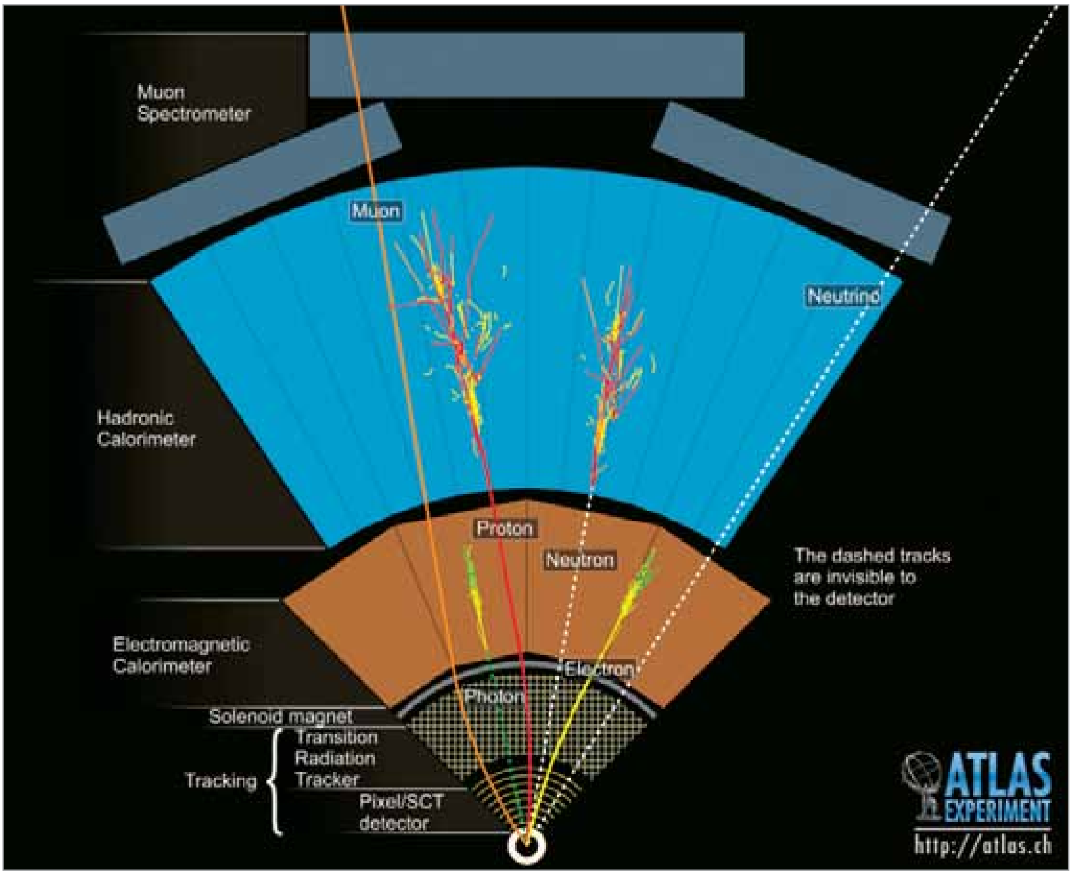


Figure 8. A cross-section of an azimuthal slice of the ATLAS detector showing successive subdetectors as described in the text. Most of the particles traversing the detector have a distinctive signature that can be recorded electronically (copyright © CERN, 2011).

jets. We already described how the strong interaction does not allow partons (quarks or gluons) to exist naked; they must be dressed up as baryons or mesons. Imagine the processes:

$$q + q \rightarrow q + q$$

or

$$q + g \rightarrow q + g \qquad (2)$$

Here, two colliding quarks (within the original protons) produce two new quarks, or a quark and a gluon produce a new quark and a new gluon — each typically with lots of energy.

As these product quarks or gluons separate, the attractive strong force between them increases very quickly (trying to stop them becoming naked!). The potential energy of the system rises quickly too — this energy results in the creation of another gluon or a $q\bar{q}$ pair … and this process of converting energy to quarks and gluons continues until the remaining energy is too low to create any new particles.

The result is a *hadronic jet* of baryons and mesons that is collimated along the direction of the initial quark or gluon, and reconstructed in the tracking detectors and the calorimeters. A typical interaction of this type that has been recorded by the ATLAS detector is shown in Figure 11.

Now the fun really starts! Every collision of two proton bunches (as noted, this will eventually be every 25 nanoseconds!) is likely to result in at least one *pp* collision (and, as shown in Figure 10, it is often more). However, only a small fraction of the collisions are of interest.

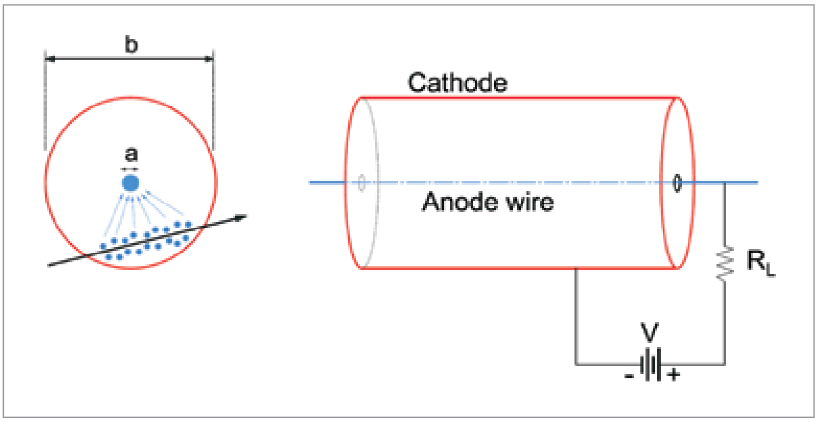


Figure 9. The basics of charged particle detection. On the right is a drawing of one of the first Geiger-Mueller counter tubes. A thin conducting wire (Anode) at positive voltage is surrounded by a metal tube at negative voltage (Cathode). The tube is filled with an inert gas. When a charged particle traverses, it ionises the gas, and the electrons drift towards the wire (left drawing) giving an electrical signal (courtesy C. Joram, CERN)

I will use the example of the process of Figure 6:

$$pp \rightarrow Z + anything$$

then

$$Z \rightarrow e^+e^- \text{ or } Z \rightarrow \mu^+\mu^- \qquad (3)$$

to show how interesting events are retained (this is called the trigger).

The cross-section for Z-production with subsequent decays of the Z to an electron or muon pair is ~0.8 nanobarns, compared with the summed cross-section of ~100 millibarns for all collision processes, and ~60 mbarns for inelastic collisions (collisions with at least one produced particle observable in the experiment) — that is, Z-production is about one event in 10 million: a needle in a haystack.

Initially, a search is made using custom-made electronics on the detector for either two energetic signals in the electromagnetic calorimeter (the electron and positron), or two high-momentum muons. If no suitable combination is found, the information of that bunch crossing is discarded and the detector is prepared for the next bunch crossing; but if the event passes the selection criteria, parts of the data are transferred to a farm of computers where algorithms are used to reconstruct the event (including the particle tracks, from the reconstruction of hits in the tracking detectors). This is illustrated in Figure 12.

We have used the Z production as an example, but as you can imagine, there is a *large menu* of interesting event topologies that we would like to retain. It is, of course, very important to ensure that interesting physics processes are not discarded unintentionally. When the machine and detectors operate at full capacity:

- up to 100,000 bunch crossings per second (of the 40 million total) are retained, using custom on-detector electronics ("level 1 trigger");
- information is transferred from the detector and analysed in a PC farm to retain up to 3,000 bunch crossings per second ("level 2 trigger");
- all of the information is then used for detailed event reconstruction, and the information of several hundred bunch crossings are retained ("Event Filter"); and finally
- the selected events are transferred world-wide for physics analysis.

Figures 13 and 14 show the signature for the example of Z production in the ATLAS detector, after final event selection. The details differ, but the same basic technique exists for event selection in all of the LHC experiments.

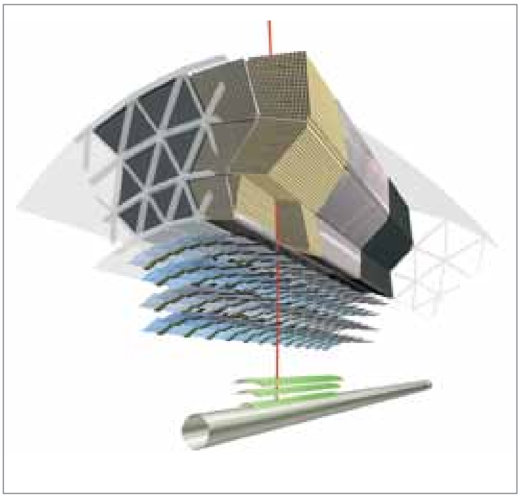


Fig. 10a

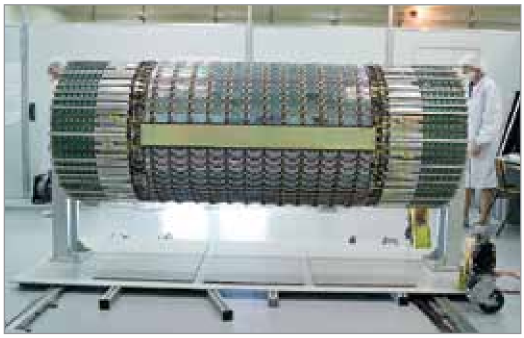


Fig. 10b

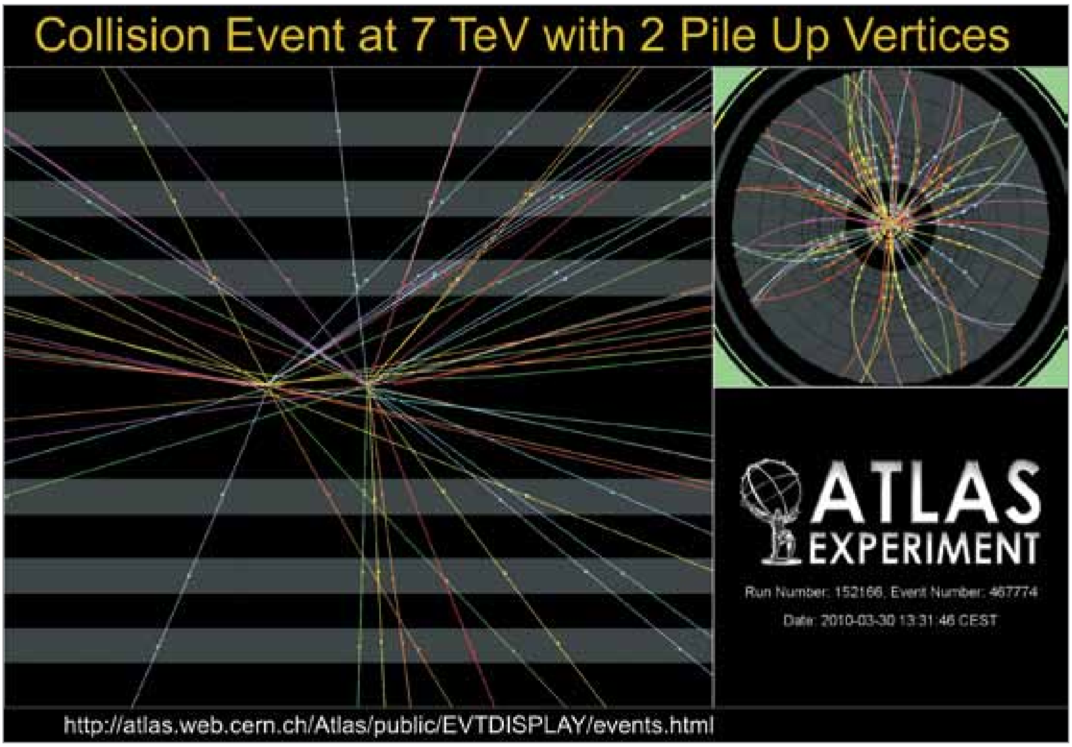


Fig. 10c

Figure 10. (a) A 3-dimensional cutaway of the central part of the ATLAS inner tracking detector used for the reconstruction of charged tracks. The elements include a silicon pixel detector at small radius, followed by layers of a silicon microstrip detector and, further out, successive layers of proportional tubes. (b) Photograph of elements of the silicon microstrip tracking detector. There are a total of 4088 silicon modules of the type shown. (c) Reconstruction of tracks in a *pp* collision recorded by the ATLAS experiment, showing the hits on each track. Two pp collisions were recorded in this proton-proton bunch crossing. (Copyright © CERN, 2011)

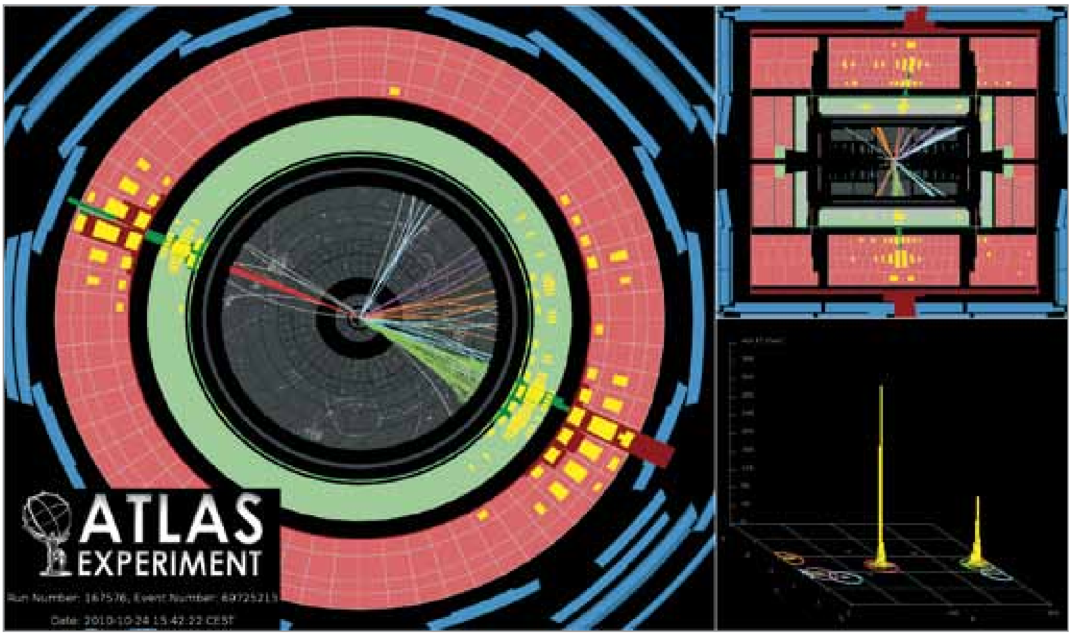


Figure 11. The two-jet event, with the highest-E_T jet reconstructed by the ATLAS experiment in 2010. The figure shows that the energy recorded by the calorimeter is very localized, and is matched to highly collimated charged tracks (copyright © CERN, 2011).

(Opposite) Figure 12. Technical details of the transfer of detector information from the ATLAS experiment, and the reconstruction of pp collisions in computer farms at the surface (copyright © CERN, 2011).

SEARCHING FOR THE HIGGS PARTICLE AT THE CERN LHC

We will now follow the steps of searching for the Higgs particle, one of the most important ongoing search activities at the LHC.

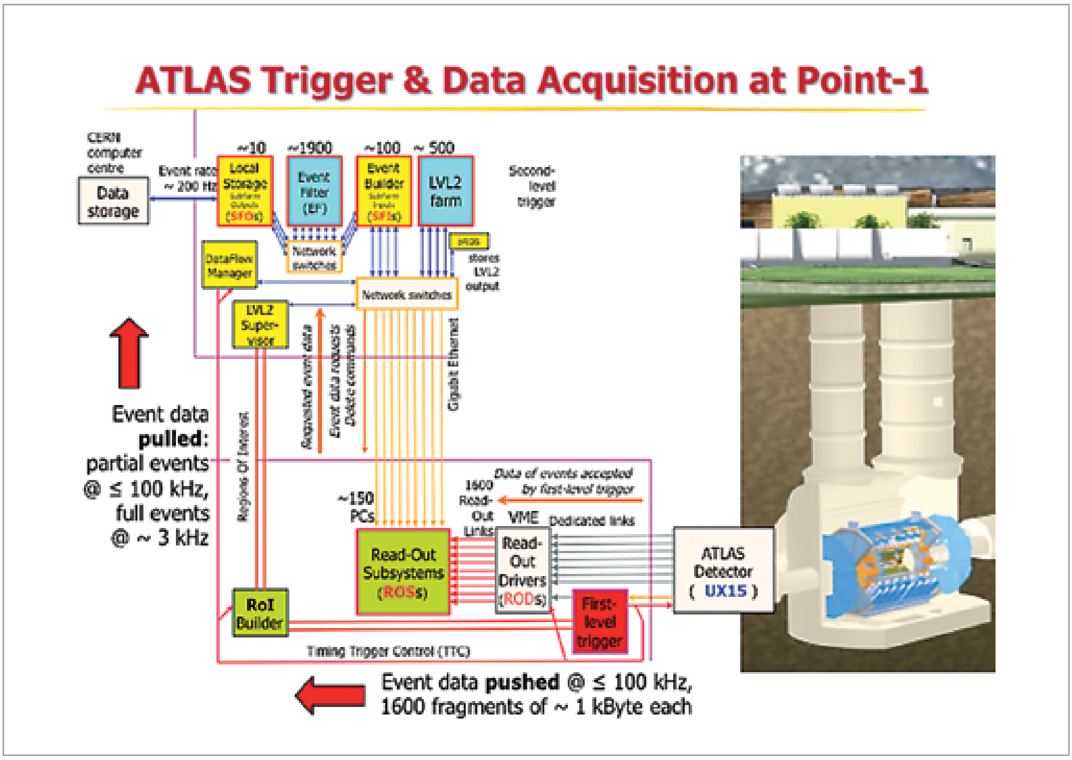
A first step of our search is to know what we are looking for! If the Standard Model is correct, a single Higgs particle will exist — and if we knew the mass of the Higgs we could compute the production cross-section with good accuracy. This is how the results of Figure 7 are obtained, using different possible Higgs masses.

We also know, from previous searches (which we will not go into here) at the CERN LEP e^+e^- Collider, that it is at least 95% likely that the Higgs mass is greater than 114.4 GeV. In addition, competing experiments[2] at a $p\bar{p}$ collision energy of 1.96 TeV have excluded, with 95% probability, the range between 158 and 175 GeV. Figure 15 shows some of the main ways to create the Higgs at the LHC, and the expected cross-section for these processes in *pp* collisions as a function of the Higgs mass.

Of course, it might be that the Standard Model is incorrect — and there we are on shifting sands. But in most extensions to the Standard Model, several Higgs particles are expected, with at least one of them being very like the Standard Model Higgs (though possibly with a different Higgs production cross section).

We now expect that the LHC will operate at a collision energy of 7 TeV until the end of 2012, before modifications to increase the collision energy to 14 TeV. How much data will we collect? We need to introduce some more terminology here: the concepts of integrated luminosity, measured in inverse femtobarns (fb^{-1}, femto = 10^{-15}). The integrated luminosity adds up the luminosity over time, giving effectively a measure of the amount of collisions over a period of time. It tells us, for a given event's cross-section, how many actual events could occur.

Here's an example. The expected cross-section for a Standard Model Higgs of mass 120 GeV is (from Figure 15) s ~ 19 picobarns



(pb). So if we collect a data sample of $1fb^{-1}$, we can expect to see around 19,000 Higgs particles produced over that time.

Combining data from the Atlas and CMS detectors, we can predict a data sample corresponding to an integrated luminosity of 10 fb^{-1} — which is very large.

However, as we said earlier, we do not detect the Higgs directly; it decays almost immediately, and we can only deduce it from the particles that are detected by the experiment. For a Standard Model Higgs, it turns out that the Higgs prefers to decay into the heaviest possible particles and, for a given Higgs mass, the fraction of decays into different decay products can be predicted from the theory; this is shown in Figure 16, where you can see that, as the Higgs mass increases, it becomes more likely that it will decay into heavier and heavier particles (the W, Z and top quark). Searches are being made in each of these decay channels at the ATLAS and CMS experiments. Next, we will look two examples of these searches.

Search 1*: Low mass Higgs, detecting* $pp \rightarrow H +$ *anything :* $H \rightarrow \gamma\gamma$

If m_H < 130 GeV, the dominant decay mode is $H \rightarrow b\bar{b}$, and the final state would be usually $pp \rightarrow H + anything\ ; H \rightarrow b\bar{b} \rightarrow hadronic\ jets$. There would be two high-energy hadronic jets ($E_T \sim m_H/2$), possibly identifiable as b-jets, plus other collision particles. But the Higgs process is rare, and other Standard Model processes can produce very similar *background* final states that look identical to the Higgs much occur more frequently — for example $pp \rightarrow b\bar{b} + anything;\ b\bar{b} \rightarrow 2\ jets$.

The easiest decay to detect is $H \rightarrow \gamma\gamma$ because the photons give a very clear signature and the energy of the photons can be accurately measured, but only a very small fraction of the Higgs decay in this way (about two per thousand). There is also a significant background from the Standard Model process $pp \rightarrow \gamma\gamma + X$ as shown in Figure 17a, but this can be partially separated as described below.

We initially require the identification of two signals in the electromagnetic calorimeter,

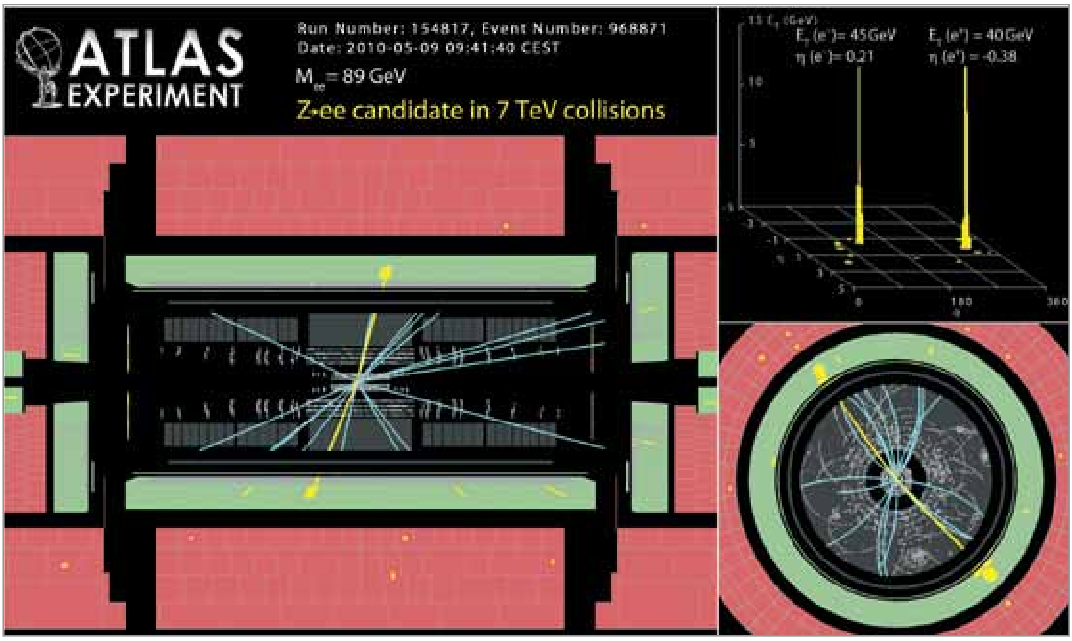
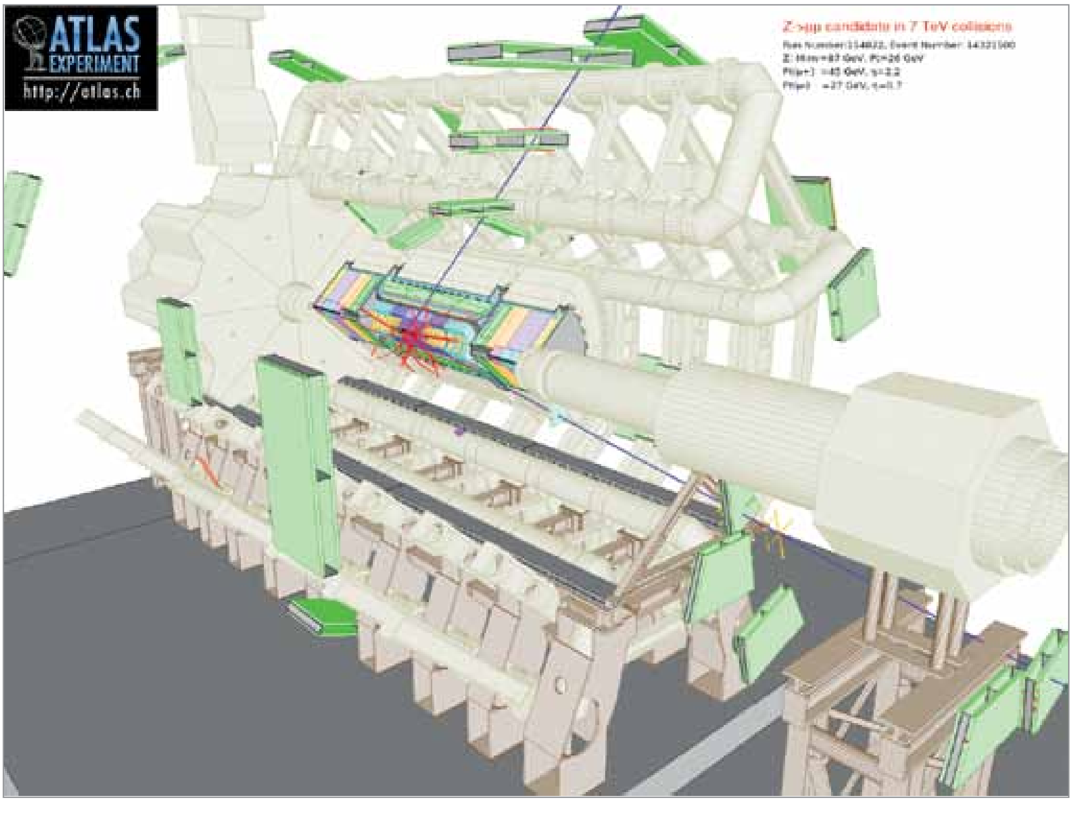


Figure 13. Event display of a candidate event for the process $p + p \rightarrow Z + anything : Z \rightarrow e^+e^-$. The electron and positron have energies transverse to the beam axis of 45 GeV and 40 GeV respectively. The reconstructed Z mass is 89 GeV. (Copyright © CERN, 2011.)

Figure 14. Event display of a candidate event for the process $p + p \rightarrow Z + anything : Z \rightarrow m^+m^-$. The reconstructed Z mass is 87 GeV. (copyright © CERN, 2011.)



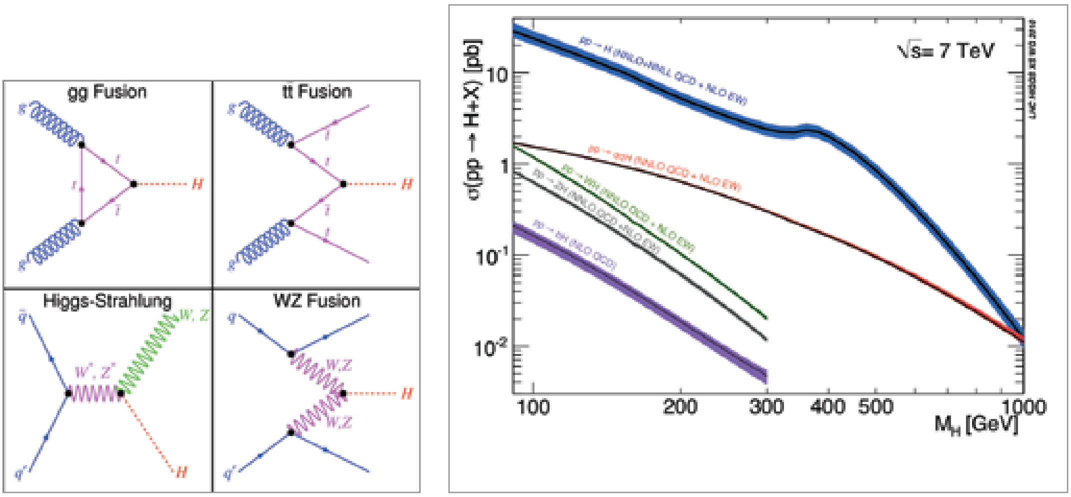


Figure 15. (a) Some lowest-order diagrams for the creation of the Standard Model Higgs particle; (b) Cross-section for the Standard Model Higgs, at a collision energy of 7 TeV, as a function of m_H.

each consistent with the experimental signature of a high-energy photon, as we discussed in Figure 8. Assuming that the two photons originated from a single massive particle, we can evaluate the mass of that particle from the energy of the photons.

If we plot the reconstructed mass for each such recorded two-photon event, the expected result for a Higgs mass of m_H = 120 GeV is shown in Figure 17b — on this graph, superimposed on a continuous background from Standard Model processes or from detector misidentifications, a small peak will be visible (the red bump). That peak would be a signature of the Higgs.

The key to Higgs identification is to understand the shape and magnitude of the background and to measure significant excess in the peak above the background. ATLAS expects to see an excess of around 200 events superimposed on a background of around 85,000 events in the mass range of Figure 17b. The width of the peak is largely determined by the experimental accuracy of the measurement, and both the CMS and ATLAS experiments have given special attention to making the most accurate measurement possible.

Search 2:

$pp \rightarrow H + anything;\ H \rightarrow W^+W^-;\ W \rightarrow l^{\pm}\nu_l$ *where* $l^{\pm}$ *is an* $e^{\pm}$ *or* $\mu^{\pm}$

If $m_H > 130$ GeV, the main decays are $H \rightarrow W^+W^{(*)}$ and $H \rightarrow ZZ^{(*)}$. The W and Z will themselves decay: for example $W^{\pm} \rightarrow l^{\pm}\nu_l$ or $W^{\pm} \rightarrow quark - antiquark$.

The fraction of decays is around 10% for each of the e± and μ± final states, and so only around 4 % of the W⁺W⁻ events will be in this category (in a data sample of 1fb^{-1}, we would expect the production of around 100 $H \rightarrow W^+W^{(*)}$ events with $W^{\pm} \rightarrow l^{\pm}\nu_l$ decay). But, as with Search 1 above, there are other Standard Model processes that mimic the Higgs production and decay, including direct WW and WZ production.

We require two signals in the detector, each consistent with a charged lepton (e± or μ±) having a high energy transverse to the beam axis (typically more than 15 GeV). Since high-energy neutrinos from the W-decays traverse the detector without leaving a signal, we require a significant E_T^{miss}.

Other Standard Model processes that might form a background with the same final state as the Higgs signal include :

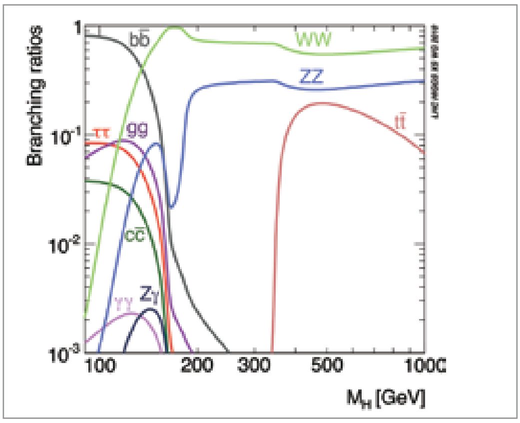


Figure 16. Branching ratios for H vs m_H

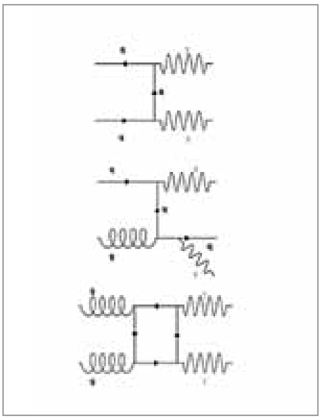


Figure 17. (a) Diagrams showing simple processes for creating photon pairs from Standard Model background processes to the Higgs.

$$pp \to t\bar{t} + anything;\ t \to W^+ b \text{ and } \bar{t} \to W^- \bar{b}\ ;\ W \to l^{\pm}\nu_l$$

$$pp \to W^+ W^- + X;\ W^{\pm} \to l^{\pm}\nu_l$$

$$pp \to Z + X;\ Z\ \ l^+ l^-$$

$$etc \qquad (4)$$

as well as

$$pp \to W^{\pm} + \geq 1 \text{ hadronic jet}$$

$$pp \to Z + \geq 1 \text{ hadronic jet} \qquad (5)$$

where the hadronic jet is misidentified as a charged lepton. All of these processes occur more often than the Higgs, but their contributions can be minimised by removing events that do not have significant E_T^{miss}. Nevertheless, a detailed understanding of the Standard Model processes and of the detector performance is necessary.

The case of $H \to ZZ^{(*)}$ for $Z \to e^+e^-$ or $Z \to \mu^+\mu^-$ decays of each Z gives a very distinctive signature, as shown in Figure 18 — but it is very rare. The signal will become important when very large data sets become available.

Putting the optimised analyses of all the different possible Higgs signals together, the ATLAS and CMS experiments hope to be able to identify the Higgs particle if it exists, using the data of 2011 and 2012. In case the Standard Model Higgs does not exist, Figure 19a shows the expected region of Higgs mass that can be excluded with 95% probability — the vertical axis effectively shows the amount of data needed to rule out a Higgs for a particular mass (horizontal axis). The graph shows that, with sufficient data, the experiments will be able to exclude a Higgs mass below 500 GeV.

If a signal from the Standard Model Higgs particle does exist, Figure 19b shows the statistical significance (s, vertical axis) that can be achieved as a function of the Higgs mass — normally a five-standard-deviation significance is required to confirm a signal, corresponding to a probability of less than 1 part in a million that it does not exist. The graph shows that, with sufficient data, ATLAS and CMS should be able to find the Higgs with high certainty … if it exists.

WHERE IS SUSY AT THE CERN LHC

We introduced the idea of supersymmetric (SUSY) particles in the previous lecture. This extension to the SM solves many of that model's inadequacies, and can provide an explanation for Dark Matter in the Universe. However, there are two main problems that I have not explained to you.

First, SUSY is a very general theory, and many arbitrary parameters and simplifications are needed to be able to compare experimental searches with the theory. But far worse, there is no direct experimental evidence for supersymmetry — we have no hints on where to search for SUSY particles, we can only

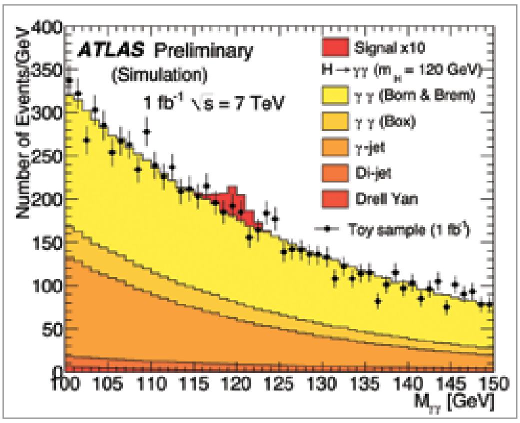


Figure 17.(b) The expected invariant mass distribution from signal and background events, for Standard Model Higgs production *pp→H+ anything : H → γγ* with m_H= 120 GeV at a pp collision energy of 7 TeV. The plot has the signal contribution (red) enhanced by a factor of 10. (Copyright © CERN, 2011.)

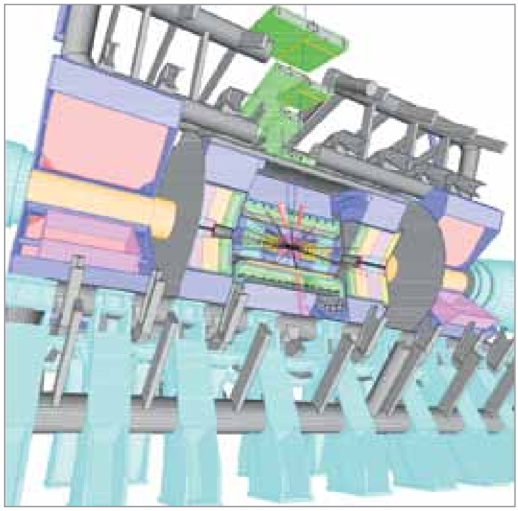


Figure 18. A 3-dimensional display of an event with $H \to ZZ^* \to e^+e^-\mu^+\mu^-$ decay (m_H = 130 GeV). Selected slices of the detector elements are shown. Reconstructed tracks with E_T > 1 GeV are shown in the inner tracker. The 2 electron tracks extrapolate to their energy depositions in the electromagnetic calorimeter (red trapezoids). The 2 muon tracks reconstructed in the muon chambers are shown as blue lines extrapolated back to the beam-line through the calorimeters. A recoiling hadron jet is visible with its energy deposited in the electromagnetic (red) and hadron calorimeters (magenta).
(Copyright © CERN, 2011)

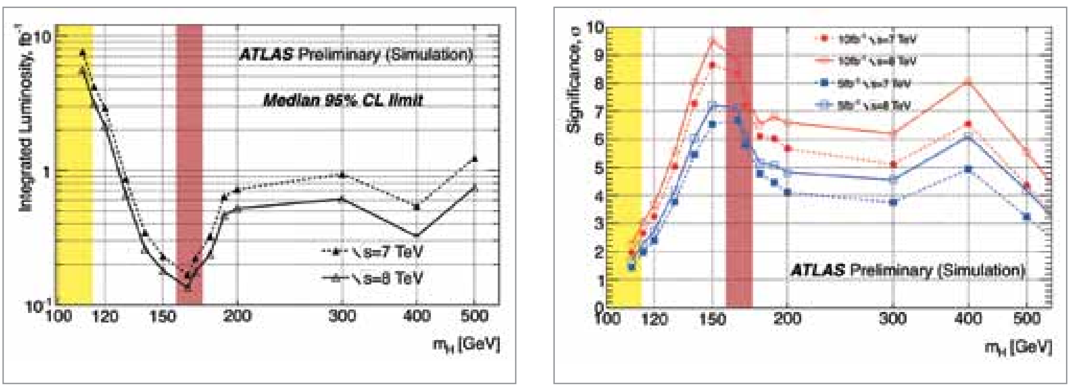


Figure 19. (a) The *integrated luminosity* required, as a function of m_H, to exclude with an expected 95% confidence a SM Higgs for *pp* collision energies of 7 TeV or 8 TeV. The shaded regions are the regions excluded at the LEP e^+e^- Collider (yellow) and the Tevatron $p\bar{p}$ Collider (brown). (b) The expected discovery significance as a function of m_H for data samples corresponding to an integrated luminosity of either 5 or 10 fb^{-1}, for a collision energy of either 7 or 8 TeV. (Copyright © CERN, 2011.)

‘guess’ where to look. This is why searches are very important at the LHC: a massive SUSY and anti-SUSY pair should be frequently produced via strong interactions if their mass is less than one-half of the *pp* collision energy.

The searches being carried out can then be largely independent of the details of the SUSY models, with a major caveat: a produced SUSY particle (whichever one it might be) will decay in a cascade until the lightest neutral or charged SUSY particle is created. In the case of a neutral SUSY particle, and assuming some basic conservation laws within the theory, the lightest neutral particle should be stable, and will pass through the experiment undetected. It will look exactly like a neutrino, but the measured E_T^{miss} will be very large because the mass of the SUSY particle is large.

A comparison of the resulting data with theory of course needs a model for SUSY, and there are many. The production cross-section will depend on the model, and on the mass of the produced SUSY pair. Figure 20 shows a possible *pp* collision producing a SUSY pair at the LHC. In this particular case (based on a SUSY model esoterically called mSUGRA), the lightest neutral particle is the neutralino (written as $\tilde{\chi}_1^0$). From Figure 20 you can see that we should be looking for a final state consisting of several high-energy hadronic jets, and large E_T^{miss}. Figure 20 also shows the simulation of such an event in the ATLAS experiment.

Both the ATLAS and CMS experiments make such searches — they look for events with several high-energy jets, large E_T^{miss}, and possibly high-E_T leptons or photons. Of course, many Standard Model processes will also produce final states identical to the SUSY signal, and these must be carefully understood — the signal of SUSY will be in any excess events of this type over what the SM predicts, and this excess will be small.

And because the SUSY signal will be small, even cracks in the detector could produce fake a fake signal — a particle passing through a small crack would go undetected, leading to a large E_T^{miss}. So these possibilities must also be understood very well.

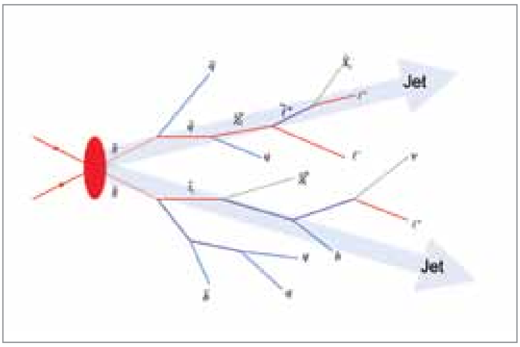


Figure 20. Example of gluino-antigluino production at the LHC. The gluino will in this case, provided R-parity is conserved, cascade towards the lightest neutral super-symmetric particle, that will be stable. The details are dependent on the exact super-symmetric model used.

Figure 21 shows the E_T^{miss} distribution from ATLAS data collected in 2010, after selecting at least 2 high-E_T jets — it’s a complicated graph, concentrate on the black dots representing the ATLAS data (with error bars), and the red line along the top of the vertical bars representing the expected result from the Standard Model. The dotted line shows a prediction from SUSY models — and you can see that the actual data reveals no excess above the SM expectation so far that could be interpreted as SUSY (or other new physics). But the search is continuing.

Even with existing data, the LHC can exclude the existence of the supersymmetric partners of the gluon (called the gluino, $\tilde{g}$) and quark (called the squark, $\tilde{q}$) significantly beyond the limits obtained previously. With the data collected in 2011, the ATLAS and CMS experiments should be able to detect, with a five-standard deviation significance, masses of the gluino and squark of up to 700 GeV if they exist. By the end of 2012, that might be extended to around 1200 GeV.

This is the story of SUSY and its relation to Dark Matter so far. There is no signal to date, but an army of young researchers is sifting the data, looking for signals of large E_T^{miss}, sometimes with associated leptons, photons, etc. A key issue is to understand and minimise the background of Standard Model processes. These searches are at a 7 TeV collision energy;

In 2014, collision energies of up to 14 TeV will open a new search region for SUSY particles, and also for other new phenomena.

SOME CONCLUDING COMMENTS

In this second lecture, we have introduced the physics research that is taking place at the CERN Large Hadron Collider. In particular, two experiments, called ATLAS and CMS, are searching evidence of the Standard Model Higgs particle.

In addition, one of the biggest mysteries in cosmology is the nature of the Dark Matter that is known to exist in the Universe. One of the most credible explanations for Dark Matter would be the existence of 'super symmetric partners' to each of the fundamental particles (quarks, leptons and gauge bosons), as described in SUSY models.

Data collected in the period until the end of 2012 will — we hope — cast light on both of those research programs. However, these are just two aspects of the physics research at the ATLAS and CMS experiments. Physicists expect that the Standard Model will need to be extended for very high-energy collisions at the LHC, but of course the experimental signals for that may be very rare. From 2014, the *pp* collision energy at the LHC will be 14 TeV, and the ATLAS and CMS experiments expect to collect a very large amount of data (an integrated luminosity of 3,000 fb^{-1}).

There is no guarantee that the Higgs and SUSY models will be correct, and other physics directions are being explored. To give one example, we have assumed in this talk that quarks and leptons are fundamental, based on the existing experimental data. But there is no guarantee that this is the case, and many unified theories of strong and electroweak forces require that quarks and leptons are themselves composite objects made from preons. Both ATLAS and CMS are repeating the experiment of Rutherford in 1911 for the atom, and of Friedman with his colleagues in 1968 for the quark, by probing deeper into the proton for evidence of deeper structure, and new limits have recently been published.

The opportunities for the discovery of new physics at the LHC are numerous, and the ATLAS and CMS detectors have been optimised very systematically towards maximum sensitivity to a wide range of new physics signals, based on detailed benchmarks from many models. It will, however, take some time, and I am confident that several of you will, in a few years, enjoy this fabulous adventure.

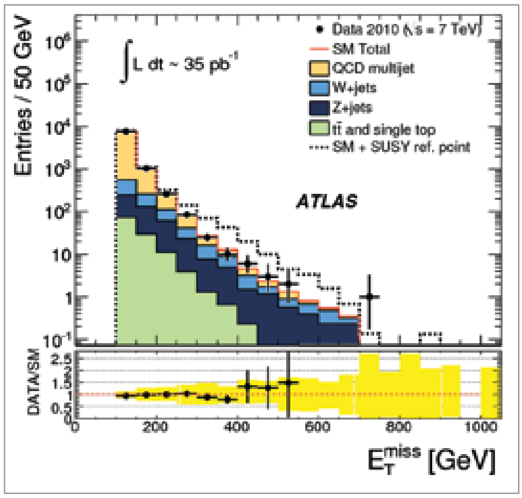


Figure 21. The distributions of E_T^{miss} after requiring two hadron jets of E_T > 20 GeV, using the full ATLAS data sample in 2010. The lower plot shows the ratio of the data to the SM expectation. Black error bars show the statistical uncertainty from the data, while the yellow band shows the uncertainty of the Standard Model prediction. For comparison, the plot also shows the expected contribution from a SUSY particle just beyond the limits already obtained at the Fermilab $p\bar{p}$ Collider. (Copyright © CERN, 2011.)

Footnotes

1 Strictly, this is only true for e^+e^- colliders, where the colliding particles are fundamental. For pp colliders, the component along the beam direction is not known (Figure 6), and only the transverse components are measured.

2 The CDF and D0 experiments at the Fermilab $p\bar{p}$ Tevatron Collider near Chicago.

Additional Reading

The ATLAS and CMS Public web-pages:
http://atlas.ch/
http://cms.web.cern.ch/cms/

Colliding Particle: Hunting the Higgs:
http://www.collidingparticles.com/

The Particle Adventure:
http://www.particleadventure.org/

FORK

The one true utensil

BENDING SPOONS FOR FUN & PROFIT

DR KARL KRUSZELNICKI

The year 1972 was a good one for something "new". In 1969, we had three amazing technological events — we walked on the Moon, the 747 Jumbo jet was introduced, as was the supersonic passenger jet, the Concorde. But on the social/human side, the Vietnam War, which had begun back in 1955, was going badly for the West (but not so badly for Vietnam).

So when the Israeli-born Uri Geller claimed that he could bend spoons, purely with the Power of his Mind — the Western World was ready to believe. He rocketed to fame.

URI 101

Soon Uri's mental skills had expanded so much that he could project the Power of his Mind through the TV set (or so he claimed). While on live TV, he told the audience that he could make their broken machinery at home start working again. And sure enough, the TV switchboard would light up with the audience telling amazing stories of how old watches had started working again. So now his repertoire included both "watch starting" and "spoon bending".

Uri's abilities soon expanded even further. He claimed that he had solved the mystery of déjà vu. It was simply his mind sending telepathic waves ahead of him while he walked, so that he would "know" what was around the corner.

Then he got into "Map Divining". He received $350,000 from the Australian mining company Zanex for advising them where to find gold around the old gold-mining town of Maldon in Victoria, and in the Solomon Islands. We don't know what resulted from his Solomon Islands advice. But we do know that Mr Geller spoke with Alan Svansio, a Zanex geologist, about Maldon. Apparently, he didn't tell Zanex anything new, just the standard geological information for the area that they already knew. It would have been ridiculously easy (especially if you had a budget of $350,000) to find out what the geology reports for any area are.

OR NOT ...

URI GELLER HAS LOTS OF FANS IN THE "PSYCHIC" COMMUNITY. ONE OF THEM, GARY WISEMAN, WROTE IN *PEOPLE* MAGAZINE THAT MR GELLER HAD SUCCESSFULLY FOUND MINERALS FOR THE MINING COMPANY RTZ (RIO-TINTO ZINC CORPORATION). HOWEVER, RTZ WROTE TO THE AUSTRALIAN SKEPTICS SOCIETY SAYING THAT THEY HAD NEVER ACTUALLY EMPLOYED MR GELLER.

AND AFTER IT WAS REPORTED THAT A DIRECTOR OF THE MINING

COMPANY ZANEX HAD EMPLOYED MR GELLER, SOME OF THE LARGER SHAREHOLDERS SOUGHT TO HAVE TWO OF THE DIRECTORS OF ZANEX REPLACED.

Mr Geller later wrote newspaper columns in which he got his readers to "improve" their lives with special "energised" orange paint or ink. More recently, he has moved into the lucrative motivational-improvement self-help personal-growth New Age industry. You can buy a Mind-Power Kit, which includes a crystal and a book which can help you develop your ability to perform dowsing, or Extra-Sensory Perception.

Curiously, Mr Geller still flies in aeroplanes and drives around in cars — rather than using teleportation.

THE SECRET

The well-known skeptic and magician James Randi, also known by the stage name of "The Amazing Randi", summarised the secret of bending a spoon in just seven words: "BEND IT WHEN NO ONE IS LOOKING".

PSYCHIC URI GELLER?

URI GELLER CALLS HIMSELF A PSYCHIC AND CLAIMS THAT HE HAS HAD VISIONS FOR MANY YEARS. HE SAYS THAT IT'S POSSIBLE THAT HIS POWERS CAME FROM "EXTRATERRESTRIALS". ACCORDING TO THE THE SKEPTIC'S DICTIONARY,"… HE HAS SUED PEOPLE FOR MILLIONS OF DOLLARS FOR SAYING OTHERWISE. HIS PSYCHIC POWERS WERE NOT SUFFICIENT TO REVEAL TO HIM, HOWEVER, THAT HE WOULD LOSE ALL THE LAWSUITS AGAINST HIS CRITICS".

IN 2000, MR GELLER SUED THE NINTENDO VIDEO GAME COMPANY FOR US$100 MILLION FOR DEVISING A CHARACTER THAT HE CLAIMED WAS A THEFT OF HIS IDENTITY. THE CASE WAS THROWN OUT OF COURT. OVER THE YEARS, HE HAS MADE MANY PUBLIC PREDICTIONS ABOUT SPORTING EVENTS. OVER THE YEARS, THE PLAYERS AND TEAMS THAT HE MOST OFTEN CHOOSES TO WIN WILL ACTUALLY LOSE.

There is nothing paranormal or psychic about bending a spoon — it's just plain old Stage Magic (if you can ever call something as wonderful as "Magic" plain or old). More precisely, Spoon Bending is an example of "sleight-of-hand". "Sleight" means "the use of dexterity or cunning, especially with the aim of deception", and so "sleight-of-hand" means "manual dexterity, usually related to performing magic tricks". (The word "sleight" comes from the Old Norse word for "sly".)

Now here's a clue to what's going on. Please note that Uri Geller is a strong and fit man. But, even so, whenever he bends a spoon only with his mind, he has to hold that spoon with two hands. This is quite surprising, as I have seen people in their nineties who are strong enough to carry a spoon (sometimes even three!) in just one hand. But Mr Geller, who is much younger, needs two hands to hold a spoon. What is going on?

If Mr Geller really were using the power of his mind to bend the spoon, he wouldn't need to touch it. He could just place it on a table and use the supposed power of his mind. But, surprisingly, he uses his two hands to hold the spoon, and after some entertaining and distracting stage patter and a few hand movements, Bingo, another bent spoon …

Leonardo da Vinci's paintings are more than just "paint on canvas", and a good magician's hand movements are more than just "hand movements". The skill of the execution of the hand movements is everything — and Uri Geller is reasonably skilled.

If you are a Card-Carrying Cutlery Mutilating Stage Magician, it's important that you distract your audience by standing up, sitting down, telling jokes — anything will do. While they are distracted, bend your spoon that you are holding with your two hands. Next, cover the bend with one hand and release the other hand and wave it around. Get a random innocent member of the audience to hold the other end of the spoon. Then, gradually transferring all the weight of the spoon to the volunteer, wave your free hand above the centre of spoon, and slowly uncover the bend that you added a few moments earlier. Voilà, instant psychic powers!

A nice variation to this trick is to take the spoon back and (when nobody is looking) bend it even more. Then, while you distract the audience again, dump the spoon on a bench. Casually tell the audience that often the spoon will keep on bending after your mind has given it the initial impulse. On cue, everybody turns to look at the spoon all by itself on a bench, and — gasp! — it has bent even more.

If you want to learn how to Bend a Spoon, check out "Spoon Bending" on Google and YouTube. You will find vast numbers of detailed instructions and movies showing exactly how to do it.

NEUROSCIENCE OF MAGIC – 1

For thousands of years, magic has been one of the most popular types of performance art. It succeeds only when the magician can manipulate our attention (what we see, hear, smell and so on), our perception (what we actually notice), our trust and our awareness. And of course, magicians can get away with practically anything if they can get the audience laughing.

So magicians have some rather special insights into human consciousness. These powerful yet subtle insights are what some neuroscientists are now exploring by working with magicians.

SHAKE = BENT?

THERE IS A WELL-KNOWN ILLUSION CALLED THE "DANCING BAR" OR "RUBBER TREE". HOLD A METAL (OR WOODEN) BAR SO THAT IT POINTS AT THE PERSON YOU WANT TO TRICK THEN WOBBLE IT RAPIDLY UP AND DOWN. AMAZINGLY, THE BAR APPEARS TO BEND IN THE MIDDLE, THE CLOSER END MOVING INDEPENDENTLY FROM THE END THAT YOU ARE HOLDING.

WELCOME TO THE "DANCING BAR" ILLUSION. RECENTLY, NEUROSCIENTISTS HAVE WORKED OUT WHAT'S REALLY GOING ON.

IN YOUR BRAIN THERE ARE REGULAR MOTION-SENSING NEURONS THAT SPECIALISE IN LOOKING AT THINGS THAT MOVE. THEY SPRING INTO ACTION AS SOON AS YOU START MOVING THE BAR UP AND DOWN. BUT THERE ARE ALSO OTHER CELLS IN YOUR BRAIN CALLED "END-STOPPED NEURONS". THEY LOOK AT THINGS THAT MOVE AND THEY ALSO PERCEIVE THE BOUNDARIES OF OBJECTS, SUCH AS CORNERS AND EDGES. THE END-STOPPED NEURONS AND THE REGULAR MOTION-SENSING NEURONS RESPOND DIFFERENTLY TO MOVING OBJECTS.

SO TWO PARTS OF YOUR BRAIN RESPOND DIFFERENTLY TO THE SAME MOVING OBJECT. THIS DIFFERENCE WARPS (OR SHIFTS) YOUR ESTIMATION OF WHERE THE EDGES OF THE MOVING BAR REALLY ARE. AND THIS MAKES YOUR BRAIN "SEE" THE SOLID BAR APPEAR TO FLOP LIKE BOILED SPAGHETTI.

YOU MIGHT NOTICE THAT ABOUT HALF OF THE PROPONENTS OF THE BENDING SPOON WILL AIM THE SPOON RIGHT AT THE CAMERA AND RAPIDLY OSCILLATE IT UP AND DOWN – AND YES, IT REALLY DOES LOOK LIKE THE HEAD OF THE SPOON IS MOVING RELATIVE TO THE HANDLE.

In general, most magic tricks use one or more of the five following "devices".

The first device is "Visual Illusion". An example of this is the after-image — the image left in your vision when there is a sudden change of brightness. A woman stands on stage wearing a skin-tight, form-fitting, white dress. The magician tells us that he will magically turn her dress red. The house lights dim and then he theatrically turns on a very bright red spotlight. The white dress, and the woman's face, turn red. He gets a weak laugh from the audience. The magician apologises. The timing of what happens next is critical. He turns off the very bright red spotlight (the house lights are already off). There will be total darkness for one-tenth of a second. Trap doors open around the woman. Hidden cables pull strongly on her white dress, which turns out to be held on with Velcro. The white dress vanishes through the trap doors, which immediately shut. During this time, thanks to the very bright red spotlight which is now switched off, an after-image of the woman has been "burnt" into the audience's vision. They "see" the woman, even though it's total darkness. After 1/10th of a second, the house lights snap on to full brightness. The woman is now seen to be wearing a bright red dress — which was hidden all the time under her white one.

The second device is "Optical Illusion". A few thousand years ago, the great Greek thinker Aristotle noticed that if he stared at the moving water in a river for a while and then looked at the rocks on the side of the river, something strange happened. He wrote, in his work the Parva Naturalia, that the rocks appeared to move in the opposite direction to the flow of the water.

This illusion is used when the magician gets the audience to stare at (for example) spinning discs that have concentric rings of slanted lines. Depending on how the lines are drawn, they might appear to simultaneously expand and contract.

HOLDING HANDS

IF YOU GO LOOKING ON THE INTERWEB FOR "HOW TO BEND SPOONS", YOU'LL MOSTLY FIND EITHER SKEPTICS OR BELIEVERS.

SURPRISINGLY, MOST BELIEVERS ADMIT THAT THEY HAVE TO TOUCH THE SPOON TO MAKE IT BEND. BUT THEY SAY THAT THIS IS ONLY TO "CHANNEL" THE ENERGY. THEY ALSO CLAIM THAT WHILE THEY DO APPLY "SOME" FORCE, THIS FORCE ISN'T ENOUGH TO BEND THE SPOON. MIND YOU, THEY OFTEN INSIST THAT YOU SHOUT LOUDLY WHILE TOUCHING THE SPOON, AND/OR DO THIS WITH A GROUP OF PEOPLE WHO ARE SIMULTANEOUSLY SHOUTING AND "TOUCHING" THEIR SPOONS …

NEUROSCIENCE OF MAGIC – 2

The third device is "Cognitive Illusion". This is where the magician aims to misdirect the attention of the audience. As a result, they don't notice that something has changed. This is called "Inattention Blindness".

A magician might move his hands slowly in a curve, or quickly in a straight line. Your eyes will accurately follow the slow-moving hand in a smooth pursuit. But when your eyes follow a fast-moving object, they don't move smoothly — instead, they jump from one location to the next. These sudden eye movements are called "saccades" — and during a saccade, you are effectively blind. You actually see "reality" only at the beginning and end of a saccade — and the stuff in between you "make up" and insert into "reality".

Another "neuroscientific" fact to add to the mix is that our brains treat curves differently from straight lines. Straight lines are fairly predictable — they just go straight. But curved lines (and corners as well) are less predictable. So our brains are "wired" to pay more attention to curved lines. And this is something a magician can do to get you to not notice something, or to redirect your attention.

Another way of getting the audience to not notice something is the famous "Gorilla Illusion", devised by the American Drs Simons and Chabris. Here, a person in a gorilla suit walks across a basketball court while the audience is trying to count how many times the basketball players pass the ball to each other. Just to make himself really obvious, the gorilla stops in the middle of the players, turns to fact the audience, crouches, beats his chest, straightens, turns again, and finishes his walk across the basketball court. Half the audience does not see the gorilla — regardless of their age, gender, level of education, nationality, religion and so on. Why? Two reasons — first, you see only a microscopic part of reality, and second, "attention".

The only "bit" of reality that you actually "see" in sharp, full colour is an area the size of your thumbnail when held at arm's length. It's tiny. The vast majority of what you see is just "virtual reality". When you enter a new location, your eyes quickly scan around and build up a fake virtual world. There are motion detectors to get you to respond to moving objects in your peripheral vision, and to shift your gaze and stare at them. But in general, all you really see and pay attention to is that tiny area the size of your thumbnail.

Also, your "rational" brain has been given the job of counting. So, for half the audience, this instruction overrides what passes right through their field of view.

The fourth device used by magicians is the field of "Special Effects". These include false gunshots, impressive explosions and so on.

The fifth device is the "Gimmick" — mechanical and secret devices such as a spoon made of a special metal alloy (for example, Nitinol) that can change shape at a certain temperature.

FOOLING THE BRAIN

WHY IS SO EASY TO FOOL THE "VISUAL" PARTS OF THE BRAIN? IN SHORT, BECAUSE THERE IS TOO MUCH "REALITY" FOR US TO SEE. SO OUR BRAIN COMPRESSES "REALITY", ALTERS "REALITY" TO MAKE IT "REASONABLE", AND TAKES SHORTCUTS SO THAT IT CAN HANDLE THE MASSIVE TORRENTS OF DATA POURING IN.

SUPPOSE SOMEBODY WALKS TOWARDS YOU. AS THEY GET CLOSER, THEIR IMAGE FILLS MORE AND MORE OF YOUR FIELD OF VISION ON YOUR RETINA. BUT ARE THEY REALLY GETTING BIGGER, OR IS THEIR HEIGHT STUCK AT 1.7 METRES? NOPE, THEY JUST STAY THE SAME HEIGHT. SO OUR BRAIN HAS TO DO "SOMETHING" SO THEY "APPEAR" THE SAME SIZE.

IS THIS WHAT THE HIPPIES MEANT WHEN THEY SAID, "REALITY IS FOR PEOPLE WHO CAN'T HANDLE DRUGS"?

WHAT MAGIC CAN DO IS EXPOSE THE COMPRESSIONS OF REALITY, AND THE SHORTCUTS OUR BRAIN TAKES, IN ORDER TO TRY TO MAKE SENSE OF THE BIG CRAZY WORLD AROUND US.

TO CATCH A THIEF ...

The right person to catch out a magician is another magician — not a scientist. A scientist is definitely the wrong person.

First, a scientist is expecting everything to be honest and above board — while a magician is aiming to trick you. So scientists are no different from the rest of us — they are easily tricked. Second, there's an old saying, "You set a thief to catch a thief". Only a thief knows the tricks of a fellow professional thief. So, in trying to catch a fellow magician, only a professional magician would know what to look for, and where and when.

Back in 1973, when the TV host Johnny Carson invited Uri Geller onto his show to bend some spoons and other metal items, he also invited James Randi along. Johnny Carson had been an amateur stage magician earlier in his career. James Randi knew that it was easy to "prepare" and weaken a spoon: all you had to do was bend it back and forth many times until it got hot, and the force needed to bend the spoon suddenly decreased. Before Mr Geller went on air, Johnny changed all the spoons and other metal items that Mr Geller had planned to use. Mr Geller responded by saying that his "power" couldn't be turned on and off like a switch — and that he simply didn't feel strong enough to try.

And what about the broken watches spontaneously restarting at home for the TV audience when Mr Geller told them to? Actually, they didn't spontaneously restart — they were picked up and given a bit of a shake … And for how long did they keep going — a few minutes, or the next 20 years?

Mr Geller started his climb to fame with the Bending Spoon trick in 1972. But in 1968, a conjuring magazine published in Israel gave "the instructions for a spoon trick that was indistinguishable from the Geller demonstration".

Uri Geller is not a psychic with supernatural powers. He's just a skilled magician who turned a few basic magic tricks into a career. To quote The Skeptic's Dictionary, "… Geller is a fraud, he has no psychic powers, and what Geller does amounts to no more than the parlour tricks of a conjurer".

And as James Randi said, "If Uri Geller bends spoons with divine powers, then he's doing it the hard way".

CHOICE BLINDNESS

A SURPRISING RESULT CAME OUT OF A SIMPLE CARD TRICK. THE CARD TRICK WAS THE DOUBLE-CARD PLOY, WHERE THE MAGICIAN USES SKILFUL SLEIGHT-OF-HAND TO SWITCH CARDS FROM ONE HAND TO THE OTHER. THE TECHNICAL NAME FOR THIS SLEIGHT-OF-HAND IS THE MAGIC PALMING TECHNIQUE.

DR JOHANSSON AND HIS COLLEAGUES SET UP AN EXPERIMENT. THEY HAD SMALL CARDS, EACH OF WHICH HAD A MARKEDLY DIFFERENT FACE ON IT (FOR EXAMPLE GLASSES OR NO GLASSES, LONG OR SHORT HAIR AND SO ON). VOLUNTEERS WERE ASKED TO CHOOSE WHICH OF TWO FACES THEY FOUND MORE ATTRACTIVE. THEN, AFTER THE VOLUNTEERS HAD MADE THEIR CHOICE, THE EXPERIMENTERS SURREPTITIOUSLY SWITCHED THE TWO CARDS FROM ONE HAND TO THE OTHER.

HERE COMES THE WEIRD STUFF. FIRST, ONLY ONE-QUARTER OF THE VOLUNTEERS NOTICED THAT THE "MORE ATTRACTIVE" FACE WAS NOW IN THE EXPERIMENTER'S OTHER HAND.

SECOND, WHEN THE VOLUNTEERS WERE ASKED WHY THEY LIKED THEIR CHOSEN FACE, THEY CAME UP WITH A WHOLE BUNCH OF REASONS. AND THEY DID THIS EVEN WHEN (AS IN THREE-QUARTERS OF THE FACES) THEY WERE REFERRING TO THE FACE THEY HAD PREVIOUSLY CHOSEN AS LESS ATTRACTIVE! THE VOLUNTEERS WERE NOW JUSTIFYING A CHOICE THAT WAS THE EXACT OPPOSITE OF WHAT THEY ORIGINALLY MADE.
WHY ON EARTH DO WE HUMANS DO WEIRD STUFF LIKE THAT? WHY ARE WE SO IRRATIONAL? MAYBE THIS EXPERIMENT WILL HELP THE NEUROSCIENTISTS UNDERSTAND THE HUMAN MIND A LITTLE BETTER – AND POSSIBLY BRING US CLOSE TO WORLD PEACE. (JUST KIDDING.)

CAPTURING CO_2

DR DEANNA D'ALESSANDRO

INTRODUCTION

Carbon dioxide is an essential component of the carbon cycle that maintains life on earth as we know it today. As the prime mover of carbon through the atmosphere, CO_2 plays a vital role in enabling the cycle of carbon from the Earth's crust (where it is found in elemental graphite and diamond, carbonates and fossil fuels such as coal and oil) to our oceans (where it occurs in carbonate minerals formed by the action of coral-reef organisms and aqueous CO_2). The biological processes of photosynthesis, respiration and decay are major components of the cycle: photosynthesis by terrestrial plants and marine plankton use sunlight to fix atmospheric CO_2 into carbohydrates, while animals eat the plants and release CO_2 by respiration and decay. Natural fires and volcanoes also release CO_2 into the air.

WHY THEN IS CO^2 CONSIDERED SUCH A PROBLEM?

For hundreds of millions of years, the carbon cycle has maintained a relatively constant amount of CO_2 in the Earth's atmosphere (approximately 0.04% by volume). While the contribution from human industry is relatively small, its recent growth has shifted this natural balance: since the start of the Industrial Revolution around 1760, the concentration of CO_2 in the atmosphere has risen dramatically from 280 ppm (parts per million) to 385 ppm today.[1,2] This significant rise has been attributed to an increasing dependence on the combustion of fossil fuels (coal, petroleum, and natural gas) which account for 86% of man-made greenhouse gas emissions, the remainder arising from land use change (primarily deforestation) and chemical processing.

While CO_2 transmits the sunlight that shines on the Earth, it also traps some of the heat that is radiated back from the land and oceans in a process called the "greenhouse effect". Modern climate science projects that the accumulation of greenhouse gases in the atmosphere — most importantly CO_2, but also methane (CH_4), nitrous oxide (NO_2), hydrofluorocarbons and even water vapour — will contribute to a net increase in surface air temperatures of 5.2°C between the years 1861 and 2100.[3] Possible consequences include higher temperatures, melting of the polar ice caps, changes in the dry and rainy seasons, as well as increasing ocean acidity with corresponding changes in carbonate equilibria. Compounding these effects is the destruction of forests for agriculture and wood, which exhausts the environment of its natural ability to absorb CO_2. Thus, the balancing effect of the natural carbon sinks that once maintained the complex interplay between components of the carbon cycle and helped to regulate the Earth's temperature is being destroyed — the net effect being an accumulation of CO_2 in the atmosphere at 30% above pre-industrial levels.[4]

REDUCING ATMOSPHERIC CO_2 LEVELS

The prospect of a worsening climatic situation due to global warming is one of the most urgent environmental concerns for our generation. Efforts to delay and prevent the further rise in temperatures that could lead to irreversible destruction of the Earth's biosphere have focused on reducing the rate of CO_2 emissions to the atmosphere.

The need for strategies to reduce greenhouse gases has prompted action from many of our governments and industries, and a number of high profile collaborative programs have been established to tackle the issue including the Intergovernmental Panel on Climate Change (IPCC), the United Nations Framework Commission on Climate Change and the Global Climate Change Initiative. The capture and sequestration of CO_2 — the predominant greenhouse gas — is a central strategy in these initiatives, as it offers the opportunity to meet increasing demands for fossil fuel energy in the short to medium term, whilst reducing the associated greenhouse gas emissions in line with global targets.[5] Carbon capture and storage (CCS) schemes embody a group of technologies for the capture of CO^2 from power plants, followed by compression, transport and permanent storage. Of course, CCS is not the only solution to the problem, and it must be used to complement other crucial strategies such as improving energy efficiency, switching to less carbon-intensive fuels and phasing in the use of renewable energy resources such as solar energy, wind and biomass.

Having established the need for CO_2 capture, this chapter aims to introduce the issues associated with capture from major man-made sources, discusses the scientific and technical challenges and highlights recent advances in materials and emerging concepts.

CURRENT STATUS OF CO_2 CAPTURE WORLDWIDE

To date, no large-scale power plant operates with a full carbon capture and storage system, although three major pilot projects are currently underway in Weyburn (Canada), In Salah (Algeria) and Sleipner West (Norway) — where the Norwegian oil company Statoil has been stripping one million tons of CO_2 per year from its natural gas production and re-injecting it into offshore saline formations beneath the North Sea since 1996. In addition to these, several smaller projects have commenced on the Dutch continental shelf (Netherlands), Snøhvit (Norway), La Barge (Wyoming, United States), Fenn Big Valley (Canada), Ketzin (Germany) and Schwarze Pumpe (Germany). A further 40 CCS projects have already been proposed worldwide between 2008 and 2020.

Closer to home, a number of smaller demonstration projects are currently underway here in Australia[6] (Figure 1): the Hazelwood coal-fired power station in Victoria captures around 50 tonnes of CO_2 per day using a solvent-based process (see below) while the Otway Basin project off the south-western Victorian coastline is the country's first demonstration of the deep geological storage or geosequestration. A number of other projects are undergoing feasibility studies including the Gorgon project off the coast of Western Australia.

POST-COMBUSTION CO_2 CAPTURE

Coal is one of the world's cheapest and most widely used sources of fossil fuel for power generation. Coal-fired power plants provide 41% of the world's electricity and currently contribute 42% of anthropogenic (man-made) CO_2 emissions.[7] In the Australian context in particular, there is enough brown coal (or lignite) to last us hundreds of years. Improvements in how power is generated from coal, as well as the ability to capture the CO_2 produced as a by-product, will provide us with a secure energy future in the short- to medium-term while more sustainable forms of energy are realised.

How much CO_2 is emitted? A typical 300 megawatt (MW) coal-fired power station releases approximately 86.3 kg of CO_2 every second, which amounts to 7500 tonnes of CO_2 per day and 2.65 megatones (Mt) per year. To put this into perspective, every year a typical plant emits a volume that is equivalent to 360 Olympic size swimming pools!

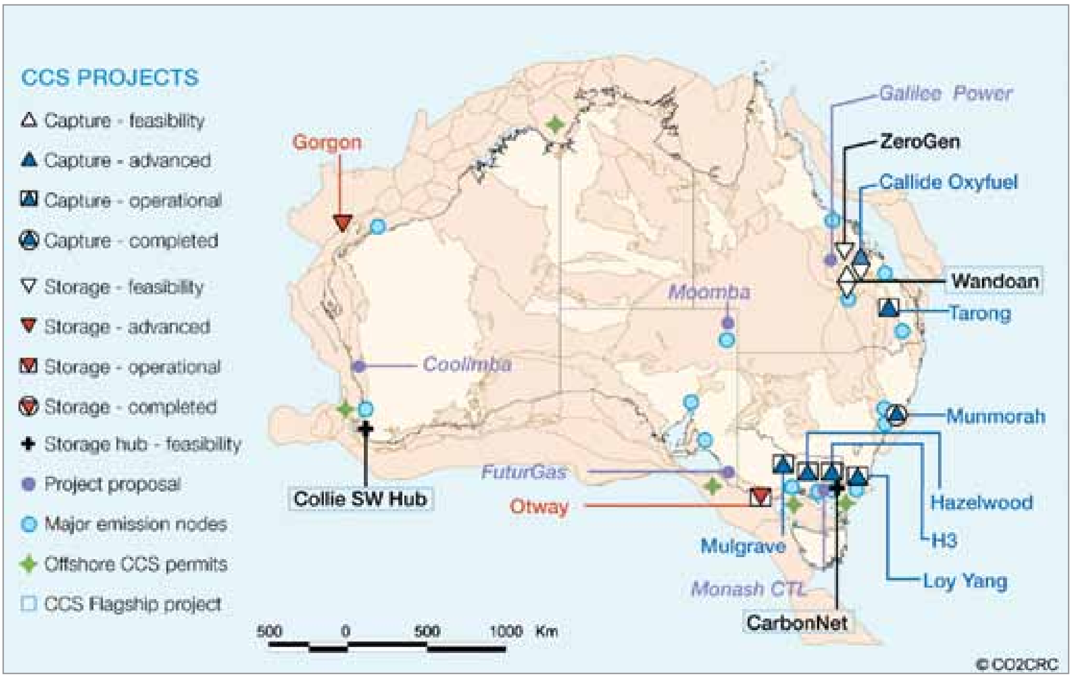


Figure 1: Map showing the current status of CO_2 capture and storage projects in Australia. Image courtesy of CO2CRC, http://www.co2crc.com.au.

Post-combustion capture involves the removal of CO_2 after the coal is burned and has the advantage that it can be retrofitted to existing plants. Indeed, the first patent outlining a process for CO_2 capture using amine-based solvents was filed in 1930, and post-combustion 'wet-scrubbing' CO_2 capture technologies have been employed industrially for over 50 years.[8] Why then do so few power stations integrate carbon capture, and why has the deployment of CCS been so slow?

To answer these questions, we need to consider the chemical processes that are involved in post-combustion capture from a typical coal-fired power station (see Figure 2).

Power plant flue gas streams consist primarily of N_2, H_2O, and CO_2 in a 13:2:2 ratio by weight. For post-combustion capture from flue gas, a major obstacle is the high volume and low pressure of the flue gas (around 1 atmosphere). The CO_2 concentration is low (around 15%), and capture requires separation from a mixture of gases, as shown in Table 1.

Current industrial separation processes involve the use of alkanolamine solutions (or "scrubbers"). In the simplest case, the flue gas (typically at 50°C) is passed through an absorber column containing an aqueous solution of monoethanolamine (MEA) at high pressure (60-70 atmospheres). CO_2 is strongly absorbed by the amine, while N_2 passes through the column and is released into the atmosphere.

The reaction of CO_2 with the amine forms a carbamate (see Figure 3). Ideally, we want to absorb as much CO_2 as we can with the scrubber: the CO_2 loading capacity (the amount of the gas that the scrubber can contain) lies in the range 0.5 to 1 mole of CO_2 per mole of amine for primary and secondary amines. The reaction of CO_2 with tertiary amines (such as *N*-methyldiethanolamine, MDEA) occurs with a loading capacity of 1 mole of CO_2 per mole of amine, albeit with a relatively lower reactivity towards CO_2 compared with the primary amines. The carbamation reaction cannot proceed for

Composition	Postcombustion	Precombustion	Kinetic diameter (Å)
CO_2	15-16%	35.5%	3.30
H_2O	5-7%	0.2%	2.65
H_2		61.5%	2.89
O_2	3-4%		3.45
CO	20 ppm	1.1%	3.75
N_2	70-75%	0.25%	3.64
SO_x	< 800 ppm		
NO_x	500 ppm		
H_2S		1.1%	
Conditions			
Temperature	50-75 °C	40 °C	
Pressure	1 bar	50-60 bar	

Table 1: Typical compositions of gases (by weight) in post-combustion and pre-combustion processes, and the kinetic diameter of the gas molecules.

tertiary amines, leading instead to bicarbonate formation (Figure 3).

In practice, the addition of small amounts of primary and secondary amines increase CO_2 absorption rates for the tertiary amines. Sterically-hindered amines such as 2-amino-2-methyl-1-propanol (AMP) have recently received considerable attention due to the lower stability of their carbamates, which gives rise to CO_2 absorption capacities of 1 mole of CO_2 per mole of amine and lower solvent regeneration costs compared with conventional primary and secondary amines.

Following the absorption process, the liquid amine CO_2-rich solvent passes into a desorber column that is heated with steam to 110–140°C in order to release the CO_2 in high purity. The high heat of formation associated with carbamate production leads to a considerable energy penalty for regeneration of the solvent — that is, lots of energy is required to retrieve the amine again for reuse. Following regeneration, the amine solution is cycled back to the absorption tower for additional CO_2 absorption.

Whilst post-combustion methods are advantageous in that the technology is commercially mature and can be easily retrofitted into existing power plants, they suffer a number of drawbacks. These include:

- large energy requirements to regenerate the solvent;
- the need for inhibitors to control corrosion and oxidative degradation due to residual oxygen in the flue stream;
- the sensitivity of the solvents to chemical degradation from other by-products in the flue gas streams, such as SO_x and NO_x;
- solvent boil-off and carryover in the CO_2 removal and regeneration steps; and
- **25-35% energy requirement for CO_2 capture!**

This energy penalty is considered too high for large-scale carbon emissions reduction, and would lead to reduced efficiencies and increased costs for electricity production — and it is obvious why so few large-scale power plants worldwide perform post-combustion carbon capture. Improved strategies for post-combustion capture do exist, including the use of liquids with lower heats of adsorption, increasing the concentration of the adsorbent molecules, and improving the mass transfer and reaction kinetics. The use of other amine-based molecules with lower regeneration temperatures has been considered for chemical absorption. Clearly, there is an urgent need for more efficient and cost-effective alternatives to the conventional "wet scrubbing" methods.

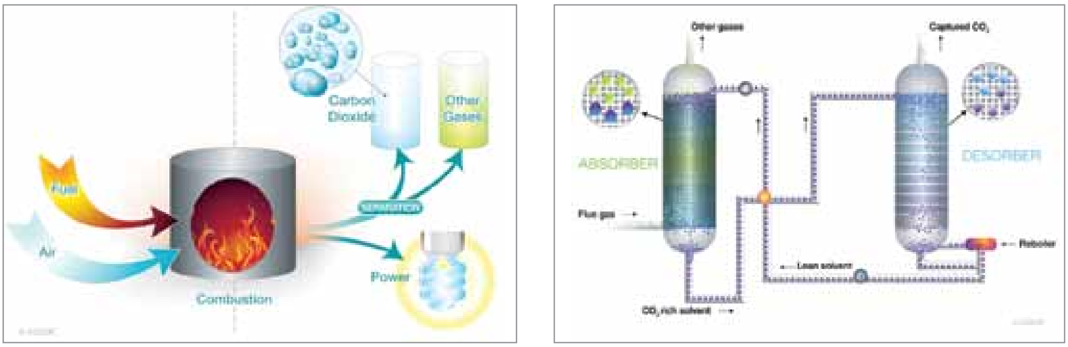


Figure 2: Schematic diagram of post-combustion capture from a typical coal-fired power station.

EMERGING METHODS FOR CO_2 CAPTURE

While the retrofitting of existing power plants using post-combustion capture methods presents the closest marketable technology, two major alternatives to post-combustion CO_2 capture processes have been proposed and are currently in the test stages of development:

- **Pre-combustion processes** involve a preliminary fuel conversion step using a gasification process and subsequent shift-reaction to form a mixture of CO_2 and hydrogen prior to combustion. The high pressure of the product gas stream facilitates the removal of CO_2.
- **In oxyfuel (or denitrogenation) processes**, fuel is combusted in oxygen instead of air by the exclusion of nitrogen, thereby producing a concentrated stream of CO_2 without the need for separation.

While the emerging technologies associated with pre-combustion and oxyfuel processes cannot be easily retrofitted into existing power plants, as can post-combustion CO_2 capture processes, the projections of the IPCC indicate that the extensive capital investments will be compensated by the relatively higher efficiencies of CO_2 separation and capture.[9] Each application involves different gas separations which impose distinct requirements and constraints for materials. Table 1 summarises the typical gas compositions and properties relevant to pre-combustion processes in comparison with the post-combustion processes that we discussed earlier.

Another important application for CO_2 capture technologies is the 'sweetening' of sour natural gas wells. Natural gas reserves (mainly CH_4) are typically contaminated with over 40% CO_2 and N_2, and the use of such gas fields is only acceptable if this additional CO_2 is separated and sequestered at the source of production. This application requires an efficient separation of CO_2 from the natural gas components at high pressures.

The capture of CO_2 from ambient air has also been suggested as a strategy to reduce atmospheric CO_2 levels, particularly with respect to small and mobile emission sources (for example, aircraft and home furnaces). Clearly, the low concentration of CO_2 in air (just 0.04%) presents a significantly higher thermodynamic barrier to capture compared with post-combustion methods, while the expense of moving large volumes of air through an absorbing material presents a further challenge.

NEW MATERIALS FOR CO_2 CAPTURE

The key factor that underlies significant advancements in capturing CO_2 is finding improved materials to perform the capture process.[10,11,12] The challenge for gas separation materials is that the differences in properties between the gases that have to be separated are relatively small, as is evident from the kinetic diameters provided in Table 1.

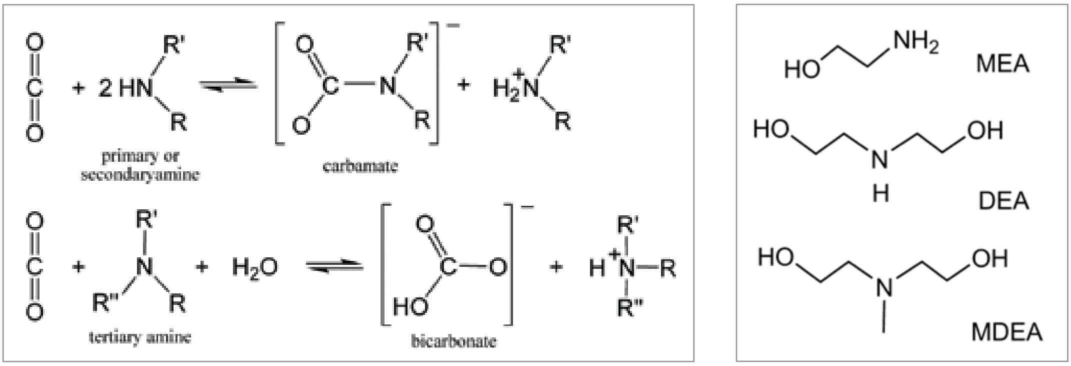


Figure 3: General reaction schemes for the chemical absorption of CO_2 by primary or secondary and tertiary amine-containing solvents (Left). Industrially important alkanolamines include the primary, secondary and tertiary alkanolamines methanolamine (MEA), diethanolamine (DEA) and *N*-methyldiethanolamine (MDEA), respectively (Right).

However, differences do exist in the electronic properties of the gases: for example, CO_2 has a large quadrupole moment (a measure of the asymmetry of the molecule's charge distribution) compared with N_2 and CH_4; and CH_4 adsorbs preferentially over N_2 due to its higher polarizability (how it reacts to an external electric field).

A diverse range of promising materials for CO_2 capture applications that could be employed in any one of the previously mentioned processes (post-combustion, pre-combustion or oxyfuel) have been proposed as alternatives to conventional chemical absorption. These include the use of physical absorbents, membranes, cryogenic distillation, hydrate formation, chemical-looping combustion using metal oxides and adsorption on solids using pressure and/or temperature swing adsorption (PSA/TSA) processes. An overview of some of these new materials is shown in Figure 4.

The key requirements for these new materials are that they must exhibit: air and water stability, corrosion resistance, high thermal stability, high selectivity and adsorption capacity for CO_2, low energy cost for regeneration, as well as adequate robustness and mechanical strength to withstand repeated exposure to high pressure gas streams.

1. Physical *Absorbents*

A promising alternative to chemical absorption using solvents such as MEA is the use of physical solvents, in which the solvent selectively binds CO_2 at high partial pressures and low temperatures. Physical absorbents such as Selexol™ (a mix of dimethylethers of polyethylene glycol) and Rectisol® (methanol chilled to −40°C), for example, have been used industrially for 40 years for natural gas sweetening and the treatment of synthesis gas. The advantage in this case is the lower heat consumption in the solvent regeneration step, as the stripping process can be driven by heat or a pressure reduction (i.e., 'flash distillation').

Ionic liquids are another class of physical solvents that are also known to be selective for CO_2 absorption. These comprise combinations of large organic cations (positive ions) and smaller inorganic anions (negative ions), and are typically viscous liquids near room temperature. In addition to their extremely low vapor pressures, they are non-flammable, immiscible (don't mix) with water, environmentally benign and very thermal stability. The mechanism for capture is physisorption: a weak association forms between the ionic liquid and CO_2 molecules through van der Waals forces rather than chemical bonds.

2. Physical *Adsorption* Materials

Solid physical adsorbents (in packed or fluidised adsorbent beds) possess significant advantages for energy efficiency compared

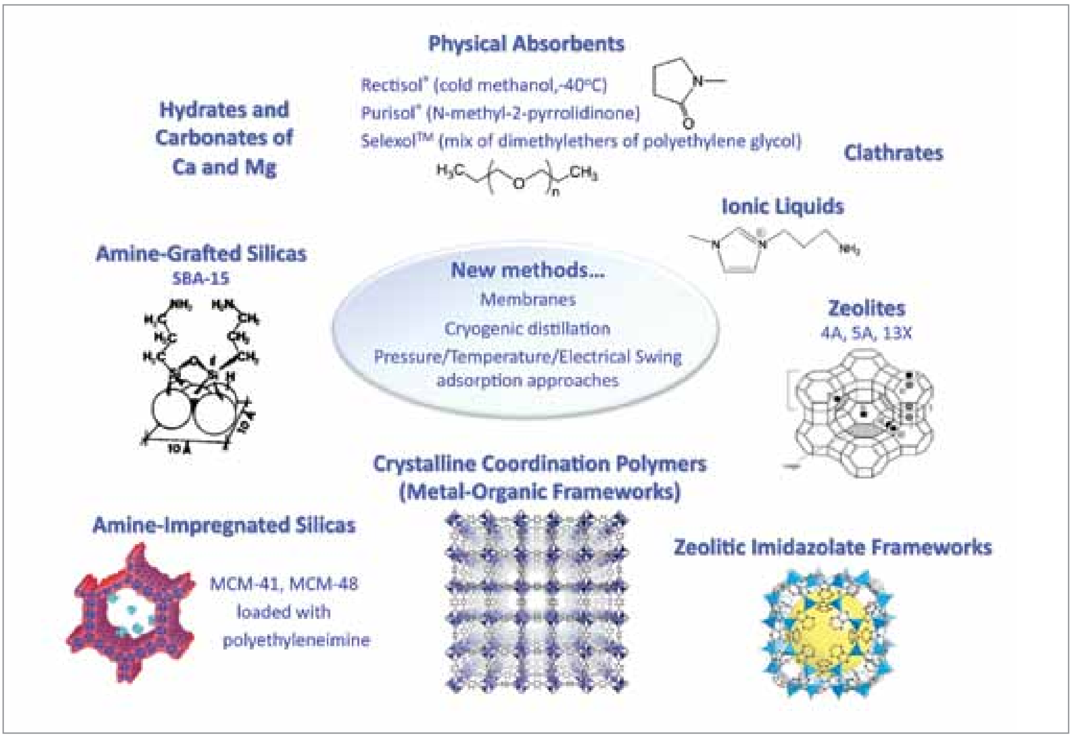


Figure 4: Overview of new materials for CO_2 capture.

with chemical and physical absorption approaches. Whereas CO_2 molecules dissolve into the bulk of the material in absorption, CO_2 adsorption involves either physisorption (described above) or chemisorption (covalent bonding) interactions between the gas molecules and the surface of a material. The CO_2 laden solid is purified in stages using pressure vacuum or temperature swing adsorption cycles to remove and concentrate the CO_2.

Adsorption capacity (the quantity of CO_2 adsorbed) and selectivity (ability to select out CO_2 for adsorption) are the principal properties relevant to adsorptive gas separation. While both factors are dependent on the operational temperature and pressure, as well as the nature of the adsorbent material and the gas being adsorbed, the factors which influence selectivity are more complicated. Possible mechanisms of adsorptive separation include: (i) the molecular sieving effect, which is based upon excluding certain components of a gas mixture based on their size and/or shape; (ii) the thermodynamic equilibrium effect, which depends on preferential adsorbate-surface or adsorbate packing interactions; (iii) the kinetic effect, due to differences in the diffusion rates of different components of a gas mixture; and (iv) the quantum sieving effect, due to the different diffusion rates of some light molecules in narrow micropores.

Microporous and Mesoporous Materials: Zeolites — microporous crystalline solids of silicon, aluminium and oxygen — are amongst the most widely reported physical adsorbents for CO_2 capture in patents and research literature. They constitute the primary adsorption material for commercial hydrogen production (involving H_2/CO_2 separation) using pressure swing adsorption, with the most popular of these based on zeolite 13X.

Zeolites are typically employed at elevated pressures (above 2 atmospheres), and their adsorption capacity has been shown to be greatly reduced by the presence of moisture in the gas — this leads to very high

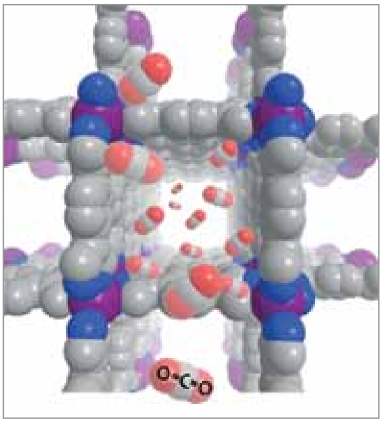
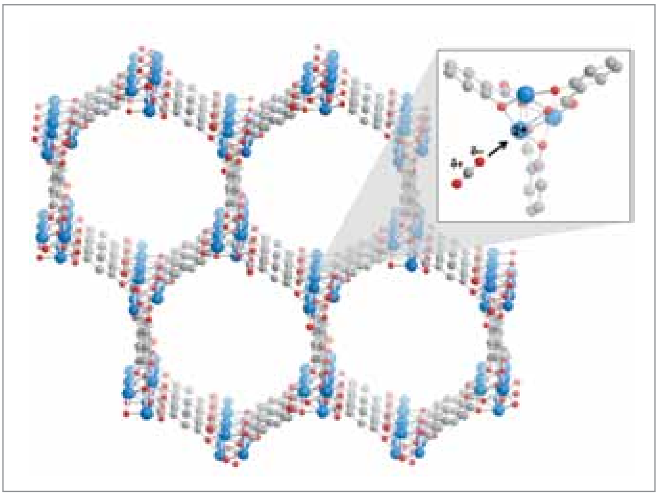


Figure 5: Example of CO_2 passing through a porous MOF structure.



regeneration temperatures (often in excess of 300°C). These additional recovery costs for their regeneration pose a significant disadvantage. The major concerns for the application of zeolites in low pressure CO_2 capture is their poor ability to handle high volumes of gas and the need to enhance their selectivity towards CO_2 adsorption. As a result, zeolites are not currently considered a serious contender for low-pressure capture applications.

An advantage of porous solid materials is the ability to modify their properties by impregnating or tethering active groups, such as alkyl-amines, onto their internal surfaces. This strategy has often been exploited to improve the gas sorption properties of porous materials for low pressure capture applications, such as those relevant in flue streams and for the capture from ambient air — and numerous amine-modified silica materials have been tried. The surface modification with primary amines allows the adsorption of CO_2 via the formation of carbamate species, reminiscent of the amine-CO_2 chemistry in conventional liquid phase scrubbing described earlier. The stripping of CO_2 can be achieved at lower temperatures (35°C) than those required for the regeneration of amine solvents (typically >100°C), thus decreasing the energy requirements of the process.

Carbonaceous Adsorbents: Activated carbon, charcoal, and virgin coal have been studies for their use in high pressure CO^2 capture applications. While the surface properties of the adsorbents can vary widely, the materials are advantageous in that they are inexpensive relative to other solid adsorbents (such as zeolite 13X), and are insensitive to moisture. The majority of studies on carbon-based sorbents are motivated by the significant industrial potential of enhanced oil recovery and enhanced coal-bed methane recovery schemes, which involve the CO_2-driven displacement of valuable oil and coal-bed methane.[9] Although these processes are already operational worldwide, the use of CO_2 (rather than the conventional technique requiring vast quantities of water) has been cited as a more environmentally feasible and attractive alternative. Another technique gaining increasing interest for use in CO_2 capture with carbon-based sorbents is electrical swing adsorption: examples include conductive monolithic mesoporous carbon adsorbents, which rapidly adsorb CO_2 when electric charge is applied and then desorbs when the charge is removed.

Metal-Organic Frameworks (MOFs) and **Zeolitic Imidazolate Frameworks**: Amongst several candidate groups of materials for CO_2 capture, highly porous three-dimensional solids known as MOFs and ZIFs show exceptional promise (Figure 5). These

Figure 6. (left) The crystal structure of the magnesium-based framework [Mg_2(dhtp)] viewed along the one-dimensional honeycomb-like channels. The inset shows the charge-induced interaction between one of the open Mg_2^+ coordination sites and a CO_2 molecule. Blue, red and gray spheres represent Mg, O and C atoms, respectively. Hydrogen atoms are omitted for clarity.

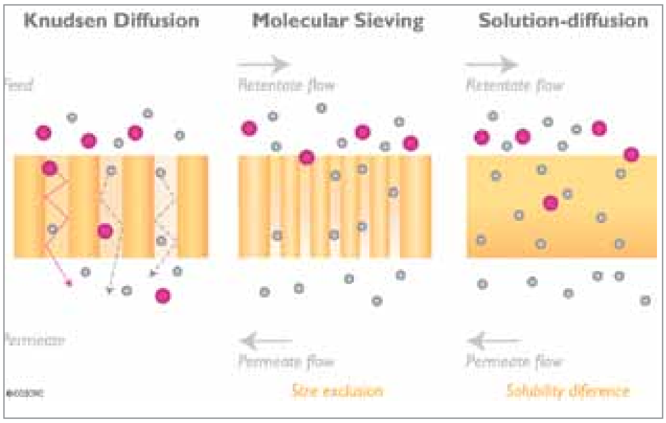


Figure 7. Schematic diagram of a membrane for CO_2 separation.

microporous crystalline solids are composed of organic bridging ligands coordinated to metal-based units (such as copper) to form highly porous sponge-like structures. Their significant gas uptake and release abilities are attributed to their extraordinarily high surface areas: one teaspoon of such a material can have a surface area equivalent to a football field!

ZIFs can be considered a sub-set of MOFs in which (i) transition metal ions such as zinc and cobalt replace the tetrahedral atoms such as silicon and aluminium, and (ii) imidazolates and benzimidazolates replace the bridging oxides in the aforementioned zeolites. These have been shown to exhibit exceptional selective capture and storage of CO_2.[13]

Current research is intense in the area of industrial applications of MOFs and ZIFs in gas storage, separation and catalysis, based on their unique structural properties, including robustness, high thermal and chemical stabilities, unprecedented internal surface areas (up to 5000 m^2/g), high void volumes (55-90%) and low densities (from 0.21 to 1.00 g/cm^3), which can even be maintained upon evacuation of the guest molecules from the pores. The regular, monodisperse (same size and shape) nature of the crystalline array of micropores is a key feature that distinguishes MOFs from other porous materials (such as polymers, mesoporous silicas and activated carbon). In addition, the ability to systematically modulate the pore dimensions and surface chemistry within MOFs is a feature that was previously largely absent in zeolite materials.

Figure 6 shows a portion of the crystal structure of a magnesium-based MOF along the one-dimensional channels, and the inset shows the charge-induced interaction between one of the open Mg_2^+ coordination sites and a CO_2 molecule.[14] There are other applications of using MOFs for capture and storage: for example, these materials are serious candidates for storing hydrogen in mobile applications such as cars and buses.

3. Membranes

Owing to their ability to selectively extract CO_2 from mixed gas streams and their low energy requirements, membrane separation technologies are a candidate for high-efficiency CO_2 capture (Figure 7) — and facilitated liquid membranes are an important class of these materials. These have several liquid absorption stages to separate CO_2 from flue gas streams, facilitated by preferential, reversible chemical reactions between CO_2 and 'carriers' dissolved in the porous membrane, such as carbonates, amines and molten salt hydrates, carbonic anhydrase or ionic liquids. Transport across the membrane is driven by the difference in the partial pressure of CO_2, which is higher on the side that contacts the flue gas than on the other. This gradient can be obtained by pressurising the gas on one side of the membrane, and/or applying a vacuum on the other — significantly, the pressure differential supplies the energy for separation and is the key to the low energy penalty of the process.

Clearly, membranes represent a promising technology for gas separation, however they suffer a number of drawbacks, particularly with regard to CO_2 capture from flue gas: in this case, the low CO_2 partial pressure provides a only a minimal driving force for gas separation, which creates an energy penalty due to the need for compression of the feed gas. Membrane materials also suffer from a decrease in permeability over time due to the build-up of particles on the surface.

WHAT DO WE DO WITH THE CO_2? MINERALISATION, FUELS AND BIOFIXATION

Two important points must first be made with regards to capture materials and potential capture technologies, given the sheer magnitude of global CO_2 emissions.[8] First, any chemical employed to capture CO_2 will rapidly exhaust its global supplies if it is used in a once-through manner; and second, any chemical produced from CO_2 as a reactant will rapidly saturate global markets for that chemical. These considerations underscore the need to be able to regenerate the capture materials; in this case, the energy input for regeneration is one of the key factors in determining the efficiency and cost of any new capture material.

Chemical fixation through the conversion of CO_2 into fuels, commodity chemicals, construction materials or mineral carbonates represents another promising alternative for CO_2 capture. The challenge here lies in the fact that CO_2 is a highly stable compound containing a low amount of chemical energy, and so natural conversion processes may be slow and inefficient as a result.

CO_2 could be used to produce fuels such as methanol, formic acid, dimethyl carbonate, methyl formate and higher hydrocarbons, as well as polymeric materials and pharmaceutical chemicals. Given the sheer magnitude of CO_2 emissions and the needs for effective catalysts, however, it appears unlikely that this approach will make a major contribution to reducing CO_2 emissions in the near-term.

A new approach for CO_2 capture involving clathrate or gas hydrate crystallisation is applicable to both post- and pre-combustion capture from flue gas or synthesis gas, respectively. The process relies on the ability of water to form non-stoichiometric crystalline compounds at high pressures and low temperatures in the presence of CO_2, N_2, O_2, H_2, and natural gas components. The gas molecules are trapped within a network of cavities formed by a hydrogen-bonded network of water molecules: a CO_2/N_2 or CO_2/H_2 mixture contacts water at a suitable temperature and pressure to form hydrate crystals, which are separated and decomposed to create a CO_2-rich stream. The preferential incorporation of CO_2 over the other gases into the hydrate crystal phase arises from the difference between the hydrate formation pressure for CO_2 relative to N_2 or H_2.

A "sunshine to petrol" approach has recently been proposed, harnessing photosynthetic energy to recycle CO_2 into biofuels by mimicking the natural processes of photosynthesis.[15] This biological fixation method uses microalgae (microscopic aquatic plants that carry out photosynthesis) to produce renewable transportation fuels, while also removing CO_2 from large point sources using either open ponds or enclosed systems such photobioreactors. Photosynthetic CO_2 mitigation is advantageous in that it does not require pure CO_2, and does not incur costs for separation, capture and compression of CO_2 gas.

STORING CO_2

The long term storage of CO_2 is a relatively untried concept, and is envisaged in deep geological formations such as saline aquifers or depleted oil/gas fields, by injection into oceans, or by sequestration in the form of mineral carbonates. A large amount of computational modeling work is currently underway to assess the safety of long-term CO_2 storage.

CONCLUSION: CHALLENGES FOR OUR GENERATION

Clearly, the development of more efficient, cost-effective, and industrially viable CO_2 capture materials is essential for the deployment of large-scale CCS schemes. Post-combustion capture from power plant flue streams provides one strategy towards reducing CO_2 emissions to the atmosphere; however, there is an urgent need for new methods and materials that perform this separation. In contrast to the low pressure, predominantly CO_2/N_2 separation required for post-combustion capture, materials for pre-combustion capture (high pressure, predominantly CO_2/H_2) and natural gas sweetening (predominantly CO_2/CH_4) have distinct requirements. Careful consideration must therefore be given to the specific technology and stage of the process at which capture occurs, to tailor the properties of a given material.

While numerous new methods and materials have been developed over the past two decades, it is clear that the ultimate integration of these into industrially useful platforms requires significant cooperation between scientists, engineers, venture capitalists, policy makers and governments. To date, no fully integrated, commercial-scale CCS projects are in operation; however, many of the component technologies are relatively mature. Progress has been made on testing new materials at the pilot plant stage, and there is now an urgent need for large-scale deployment of the most promising technologies. From this perspective, it is apparent why the problem of CO_2 capture is regarded as one of the grand challenges for the 21st century.

Despite the numerous issues surrounding CO_2 capture, and the many political, regulatory and economic drivers which will ultimately dictate the deployment timeframes for new CCS schemes, the time is ripe for scientists — in this International Year of Chemistry — to play a central role in solving the CO_2 capture problem.

Footnotes

1 A. Neftel, H. Friedli, E. Moor, H. Lötscher, H. Oeschger, U. Siegenthaler, B. Stauffer. *In Trends: A Compendium of Data on Global Change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy (1994).

2 R. F. Keeling, S. C. Piper, A. F. Bollenbacher, S. J. Walker. *In Atmospheric CO2 values derived from in situ air samples collected at Manua Loa, Hawaii, USA*, Scripps Institution of Oceanography, La Jolla, CA (2009).

3 A. P. Sokolov, P. H. Stone, C. E. Forest, R. Prinn, M. C. Sarofim, M. Webster, S. Paltsev, C. A. Schlosser, D. Kicklighter, S. Dutkiewicz, J. M. Reilly, C. Wang, B. Felzer, H. D. Jacoby. J. *Climate* **22**, 5175 (2009).

4 M. Z. Jacobson, *Energy Environ. Sci.* **2**, 148 (2009).

5 United Nations Framework Convention on Climate Change,

6 CRC for Greenhouse Gas Technologies, http://www.co2crc.com.au

7 International Energy Outlook 2010, U.S. Energy Information Administration; Online: http:/www.eia.doe.gov/oiaf/ieo/ (2010).

8 R. R. Bottoms (Girdler Corp.), "Separating acid gases," U.S. Patent 1783901 (1930).

9 B. Metz, O. Davidson, H. de Coninck, M. Loos, L. Meyer. *IPCC special report on Carbon Dioxide Capture and Storage*. Cambridge University Press, Cambridge, United Kingdom and New York (2005).

10 D.M. D'Alessandro, B. Smit, J.R. Long, Angew. *Chem. Int. Ed.* **49**, 6058 (2010).

11 D.M. D'Alessandro, T. McDonald, *Pure Appl. Chem.* **83**, 57 (2011).

12 Figueroa, Fout, Plasynski, McIlvried, Srivastava, Int. J. *Greenhouse Gas Control* **2**, 9 (2008).

13 R. Banerjee, A. Phan, B. Wang, C. Knobler, H. Furukawa, M. O'Keeffe, O. M. Yaghi. *Science* **319**, 939 (2008).

14 D. Britt, H. Furukawa, B. Wang, T. G. Glover, O. M. Yaghi, *Proc. Nat. Acad. Sci. U.S.A.* **106**, 20637 (2009).

15 C. Stewart, M.-A. Hessami, *Energy Conv. Manage.* **46**, 403 (2005).

PHOTOVOLTAICS: SOLAR ELECTRICITY BY COUPLING LIGHT AND MATTER

PROFESSOR MARTIN GREEN

INTRODUCTION

The Sun already provides 99.99% of the energy needed to support human life as we know it. Providing the final 0.01% has been problematic, since the traditional way of doing this for the past century — by burning fossil fuels — is increasing the amount of solar energy trapped by the Earth, increasing temperatures. A less damaging way of generating this relatively small amount of commercial energy is required.

Photovoltaic solar cells provide a potential solution by tapping into a small fraction of this large solar resource. This approach combines two of the most significant developments of the 20th century: the development of quantum mechanics, and the development of the semiconductor microelectronics industry.

How do solar cells work? What are the challenges in trying to displace a multi-billion dollar industry that has been entrenched for over a century? These are some of the issues explored in this chapter.

OVERVIEW

In just a few days, the Earth receives more energy from the Sun than from all the fuel burnt over the whole of human history. Three weeks of sunshine outweighs all known fossil fuel reserves.

The challenge for a sustainable future is to tap into a very small fraction of this huge energy source to provide the relatively small amount of commercial energy we presently produce by burning fossil fuels. Photovoltaic solar cells provide the most elegant way known of doing this, by converting sunlight directly into electricity. The underlying physics is quite complex, but the use of the cells is very simple: just place them in sunlight and they act like a battery, producing electricity at their output connections.

Figure 1 shows a typical present-day solar cell. The cell is about the size of a DVD disc, about 12 to 15 centimetres square, and made from a thin wafer of silicon — the same material as used in microelectronics. A silver "finger" pattern partly covers the side exposed to sunlight and makes electrical contact to this side. A second metal layer makes contact to most of the rear. When sunlight falls on a cell, electricity can be drawn from between these contacts in the same way as from a mobile phone battery — but here we have a battery that in principle can last for as long as the sun shines without recharging.

Generally, several dozen solar cells are connected together in a weatherproof package known as a solar module or solar panel (Figure 2). A toughened glass sheet forms the top surface of this module, protecting the cells from the harsh and often inclement outdoor environment where they must operate. Manufacturers warrant

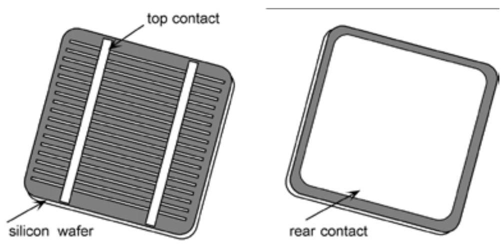


Figure 1: A silicon solar cell shown from the top and rear. Sunlight enters from the top and is absorbed within the thin silicon wafer that forms the body of the cell, giving an electrical output between the top and rear contacts.

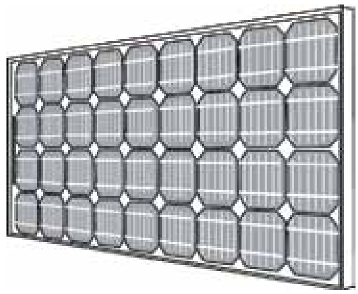


Figure 2: A standard solar cell module showing, in this case, 36 cells connected together in a weatherproof package with a strengthened glass cover sheet.

modules for 25 to 30 years, with saucepans one of the few other products that come with such a generous warranty. As more experience is gained, warranties could well grow to 50 years.

Modules are typically about 1 or 2 square metres in area. They are rated in terms of their electrical power output, expressed in Watts. A typical module will give about 150 Watts for each square metre of area. Ten modules are enough to provide half the electricity needs of a typical home — and more than enough for all the electricity needs of a well-designed home.

For many years, the use of solar cells was restricted to small, specialised uses, mainly in remote areas. Rapidly reducing costs mean that increasing numbers are now starting to appear on residential rooftops. At the end of 2010, around 8% of detached homes in New South Wales had a photovoltaic system on their roof, with this number rapidly increasing.

However, it is still early days for this technology with costs expected to reduce much further as the technology matures. In the not too distant future, instead of building large coal fired power stations to provide the world's rapidly increasing demand for electricity, we may build large fields of solar cells, producing clean energy with no polluting by-products.

THE SUN

The Sun is one of many billions of stars in our galaxy, and our galaxy is one of billions in the Universe. Despite this apparent insignificance, the Sun is by far the dominant force in our solar system. As already noted, light radiated from the Sun supplies 99.99% of the energy to support life on Earth — through direct heating, and through photosynthesis, which sits at the base of the food chain. The other 0.01% is the energy we use to power our vehicles, businesses, industries, devices and homes.

The Sun is a hot sphere of gas heated by a nuclear fusion reaction at its core (Figure 3). Internal temperatures reach a very warm 15,000,000 K. The intense radiation from the core is radiated to an outer convective zone where cooler temperatures (a mere 2,000,000 K) make the Sun's constituents more opaque. Energy is transferred through this optical barrier by convection. It is then re-radiated from the outer surface of the Sun, known as the photosphere; here temperatures are about 6,000 K, about twice the temperature of the filament in an incandescent light.

WHAT IS LIGHT?

Some of the greatest minds in history have tried to understand what light is. Isaac Newton (1642–1727) made giant leaps forward: he showed that sunlight could be

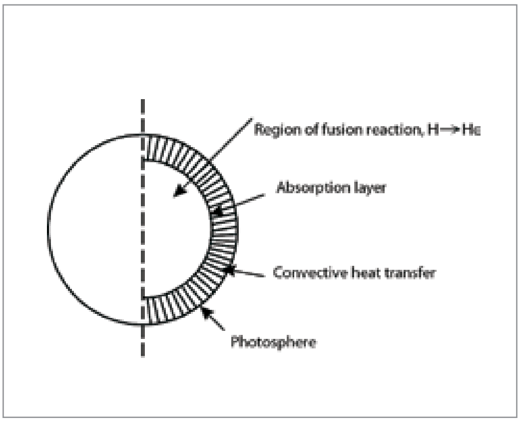


Figure 3: Cross-sectional view of the sun.

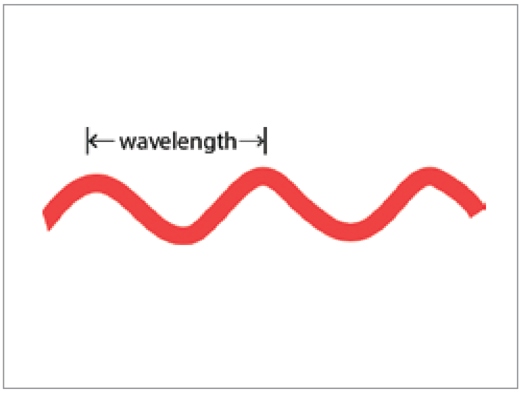


Figure 4: The wavelength of a wave is the distance between peaks or troughs. For visible light, the wavelength is very small, ranging from 0.44 micrometres (μm, also called a micron — a millionth of a metre) for violet light, to 0.65 microns for red light.

split into the colours of the rainbow, and that these colours could be recombined to give white light. Newton thought of light as a stream of tiny particles, like miniature billiard balls. Some of Newton's own experiments, and later ones in the 18th and 19th centuries, showed that this view had to be wrong — these experiments showed that light acted as a wave, like ripples on a pond. Waves explained what are known as 'interference' effects (think of the bands of colours often seen on the surface of soap bubbles, or in oil-slicks on the road). These effects could be explained if a "wavelength" could be associated with each colour, equal to the very small distance between the successive peaks or troughs in the wave (Figure 4).

BLACK BODIES

During the 19th century, techniques for measuring the properties of the light emitted by hot bodies, such as the Sun, improved enormously. Of particular interest was the nature of light emitted by a "black body" as it was heated. A black body is simply an object that absorbs light very well — a perfect black body absorbs all light that hits it. A popular choice of model black body studied at the time was metal coated with black soot.

A black body is not only a good *absorber* of light, but is also a good *emitter* of light when heated. We all know that a hot body first glows red, then orange, then white as it gets hotter and hotter. This was put on a quantitative basis in the late 19th century when emission curves were measured, such as in Figure 5.

A heated black body emits light with a continuous range of wavelengths, with the light that we can see (wavelengths between 0.4 and 0.8 microns) often forming only a small part of it. A black body heated to 3,000 K (the temperature of an incandescent lamp) emits only a small amount of its energy in the visible spectrum — these lamps are therefore not very efficient for lighting, which is why they are being replaced.

As the temperature rises, more of the energy lies in the visible spectrum. At 6,000K, the temperature of the Sun's photosphere, the emission is centred over the wavelengths to which our eyes respond — a result of evolution or grand design, depending upon your point of view. The wavelength distribution of sunlight reaching the Earth, however, is distorted by absorption, by causing atoms in water and, importantly, carbon dioxide atoms to jiggle around. The latter jiggling is even more important for the longer-wavelength light emitted by the Earth (at a much cooler 300 K).

QUANTUM PHYSICS

The above black body measurements posed a problem, because physicists couldn't understand them using conventional theories.

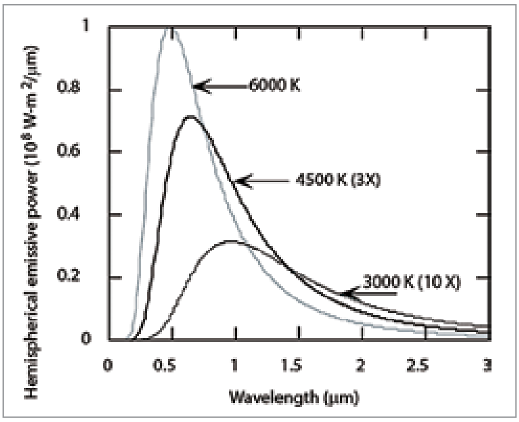


Figure 5: Emitted energy fluxes from perfect "black bodies" at different temperatures.

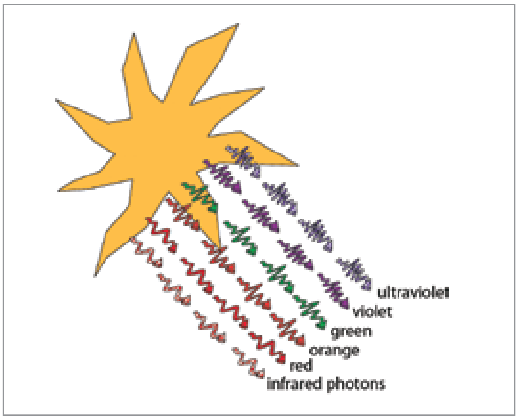


Figure 6: A quantum picture of light from the Sun. Light comes from the Sun in packets of energy known as photons, some of which are visible to the eye as colours.

Max Planck (1858–1947) found that the only way he could explain the measurements of blackbody radiation was if he assumed that changes in energy within the hot body were not continuous (as such changes were always assumed to be), but could occur only small steps, now known as 'quanta'. This clue triggered a revolution in 20th century physics.

Albert Einstein (1879-1955) is deservedly well known for his work on relativity; however, he made outstanding contributions to several other areas of physics, including the physics of quanta, or "quantum physics". His somewhat overdue Nobel Prize in 1921 was awarded for "his services to theoretical physics and especially for his discovery of the law of the photoelectric effect". This was Einstein's key contribution to quantum physics, described in the introduction to his 1905 paper on light quanta. Abandoning the classical picture of light as a wave, Einstein suggested that the energy of light was not spread continuously in space but (in English translation) 'consists of a finite number of energy quanta localized at points of space that move without dividing, and can be absorbed or generated only as complete units' — in other words, back to the corpuscular billiard balls of Isaac Newton.

Einstein pointed out that the well-known wavelike properties of light did not contradict this corpuscular interpretation. In its effects, light behaves neither as a wave nor a billiard ball, but as something foreign to everyday experience. Following Einstein, we can think of light from the Sun coming in tiny packets of energy, as shown schematically in Figure 6. These light quanta are now known as "photons".

The eye cannot see all the photons coming from the sun, only the visible light that ranges through the seven colours of the rainbow, from red through to violet. Each different colour is associated with photons of different energy, and photons of the same colour carry exactly the same amount of energy. This means the number of photons of each colour emitted by a heated body can be calculated by dividing the total energy at each wavelength, as in Figure 5, by the energy of each photon.

The energy of each photon is about twice as large for light at the violet end of the visible spectrum as at the red end. The packet size is even larger for ultraviolet light, which is the reason it causes sunburn. Beyond the red end of the visible spectrum there is low-energy infrared light; again, the eye cannot see this, although devices such as "heat-sensing cameras" can, and we also feel it as heat.

The evidence that Einstein relied upon to support his light quanta ideas came from experiments made by other scientists studying the photoelectric effect, which

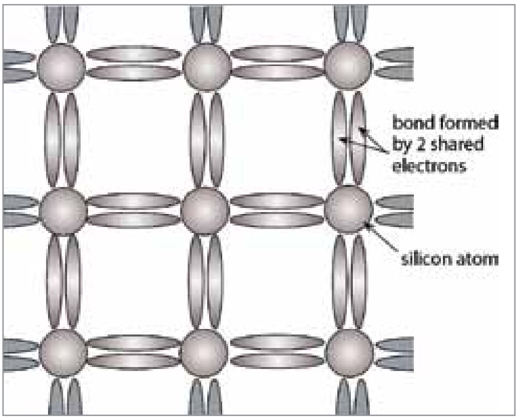


Figure 7: Sketch of how silicon atoms try to arrange themselves when cooled from a melt. Each silicon atom preferentially attaches to four neighbours (the actual arrangement is more interesting, involving the third dimension, but is more difficult to visualise and to draw).

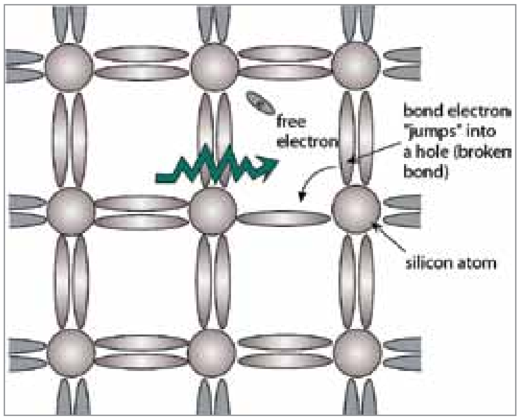


Figure 8: Silicon with one electron released from a bond, such as by the energetic photon shown. The released electron is free to move through the semiconductor, and so is the broken bond.

involves the interaction of light with metal conductors. The closely related "photovoltaic" effect, on which solar cells rely, involves light interaction with materials known as "semiconductors". These materials have properties in-between those of metals, which are good conductors of electricity, and insulators, which are very poor conductors.

SEMICONDUCTORS

Silicon is the most common semiconductor, underpinning microelectronics, computers, the Internet, mobile phones, HD TV and most of the other rapidly evolving areas of modern life. Silicon is made up from atoms with an atomic number of 14, which means that an isolated atom of silicon consists of 14 electrons surrounding a dense, positively charged central nucleus with 14 protons and 14 neutrons, like a miniature solar system. Ten of the 14 electrons are very tightly bound to this central core and are not of any further interest — at least, not for solar cells.

The four electrons left remaining determine how silicon atoms arrange themselves when they form solid silicon material, like the wafers in Figure 1. Solid silicon for solar cells is made by extracting silicon from quartz (a compound of silicon and oxygen), melting this extracted silicon in large bowls known as crucibles, and then slowly cooling it. Silicon freezes with the atoms doing their best to arrange themselves in one very particular pattern: each silicon atom tries to link with four neighbouring atoms, as in Figure 7. The "glue" bonding the atoms together is two shared electrons, one from each atom. Remembering that each silicon atom has four electrons that are not tightly bound, everything works out neatly if each silicon atom is surrounded by exactly four other atoms.

Since electricity is just the flow of electrons, silicon is a poor conductor of electricity when all electrons are locked up in bonds as in Figure 7. Pure silicon therefore is an insulator. However, these bonds can be broken if sufficiently jolted — for example, by an energetic photon from the sun. Once released from a bond (as shown in Figure 8), the electron can move through the silicon and contribute to electrical current flow; that is, silicon with broken bonds acts like a conductor.

This gives a clue as to why silicon is known as a semiconductor: sometimes it acts like an insulator, and sometimes like a conductor. We are now close to understanding how a solar cell works.

The electrons released from bonds are able to move through the semiconductor. Perhaps more surprisingly, the broken bonds themselves can move. This is because it is

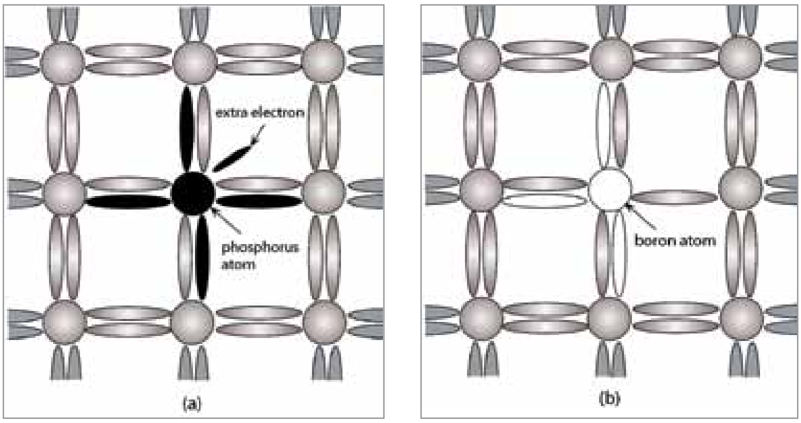


Figure 9: (a) Silicon with small amounts of phosphorus added; (b) with small amounts of boron.

very easy for an electron in a neighbouring bond to jump into the vacant spot left by the broken bond (see Figure 8). This jump restores the originally broken bond, but leaves a new broken bond behind.

In this way, the broken bond can move through the silicon. To visualize this motion, the broken bond can be thought of as a particle (called a "hole"), a bit like a bubble. Just as two negatives make a positive, the hole (an absence of negative charge) has an electrical charge opposite to that of the released electron — that is, holes are positive charges. When a photon breaks a bond in silicon, a negatively charged electron and a positively charged hole are created, known as an "electron-hole pair".

ELECTRON AND HOLE DOPING

The properties of silicon can be altered in ways other than by shining light on it. One important way is by adding small amounts of impurities known as "dopants". For example, if a small amount of phosphorus is added to molten silicon, the solidified silicon will contain phosphorus atoms in some positions where silicon would normally be, as shown in Figure 9(a).

Phosphorus has five electrons that are loosely bound to its central core. Four of these are used in the bonds between neighbouring silicon atoms, but the fifth one is at a bit of a loose end ("Nigel no-friends"). It is only weakly bound to the original phosphorus atom and can be very easily torn away. Once separated, it acts very much the same as an electron released by light absorption.

If a different impurity, such as boron, with only three electrons not tightly bound to its core, is introduced in the same way, full bonds will be formed with only three of the neighbouring silicon atoms, as shown in Figure 9(b). Adding, or 'doping', with boron is a good way to introduce broken bonds or holes into the semiconductor material.

Silicon doped with phosphorus is a reasonably good conductor because it has plenty of unbound or free electrons. Since it has plenty of these negative charge carriers, it is called "negative-type" (or "n-type") material. Silicon doped with boron is also a reasonably good conductor because it has plenty of holes, which are positive charge carriers. Not surprisingly, such material is known as "positive-type" (or "p-type") material.

One of the most important devices in all of microelectronics is formed by a junction between p-type and n-type material. In fact, such p-n junctions can be regarded as the basic building blocks of microelectronics, and could be considered one of the most important inventions in human history. Such junctions are also key elements of solar cells.

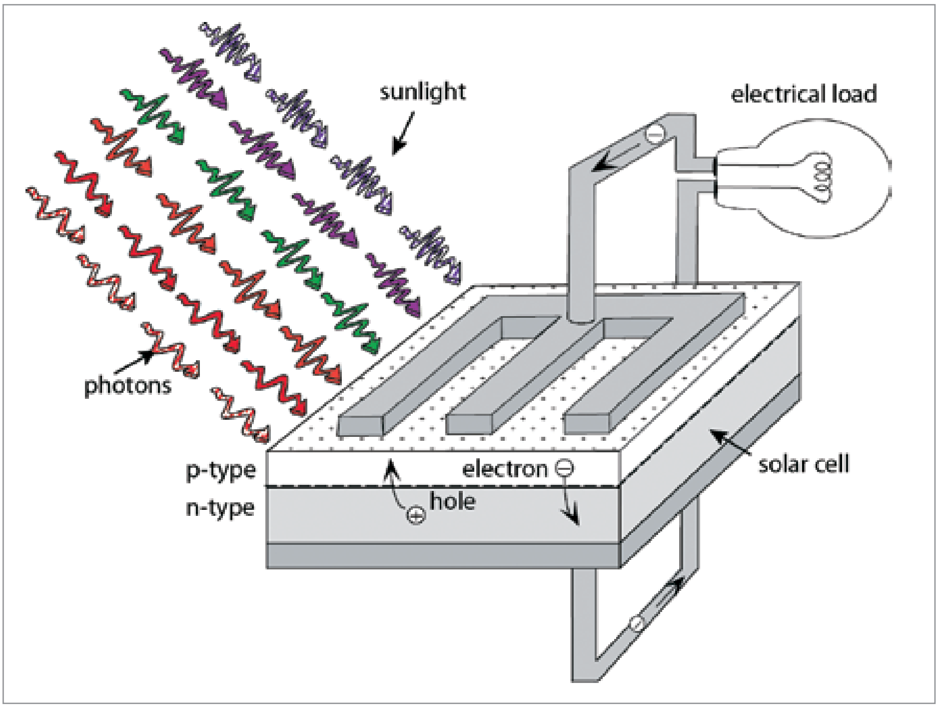


Figure 10: Photons in sunlight release electrons from the silicon bonds, giving mobile electrons and holes. The p-n junction causes these to go in opposite directions. The electrons flow through the external load and meet up with holes on their return.

INTERNAL WORKINGS

Based on what we now know, we can now understand what goes on inside a solar cell (Figure 10). Photons in sunlight enter the silicon through the spaces between the top metal contact. Once inside, the more energetic photons are absorbed by giving their energy to electrons, originally constrained in the bonds holding the silicon atoms together. This releases electrical charge carriers — the electrons and holes — within the silicon material (as in Figure 10).

The p-n junction is the final bit of technology required to make a solar cell work. Its job is to encourage all the released electrons to move off in the same direction, and all the holes to move in the opposite direction.

The released electrons flow more easily in the region where there are plenty of them (that is, the n-type side of the device); the holes flow more easily in the p-type region. This asymmetry causes a directional flow of electrons, released by the light, from the p-type to the n-type side of the p-n junction, and an opposite flow of holes. If an electrical load is connected between the n-type and p-type regions, the flow of electrons continues around through the load back to the p-type side of the device, where each electron will mate up with a hole. The bonds will be restored and the electrical circuit completed.

ROOFTOP PHOTOVOLTAICS

Electricity for urban homes in the past has generally been sent hundreds of kilometres through aboveground electricity wires connecting the home directly to large power stations. However, over the past few years a new way of supplying energy has emerged: through putting a solar system on the roof.

Figure 11 shows the basic concept. The house remains connected to standard electricity supply network. No storage is required. If the system is generating more electricity during the daytime than the home needs, the excess is sent back through the network to factories and offices that need daytime power. During the night, electricity flows in the opposite

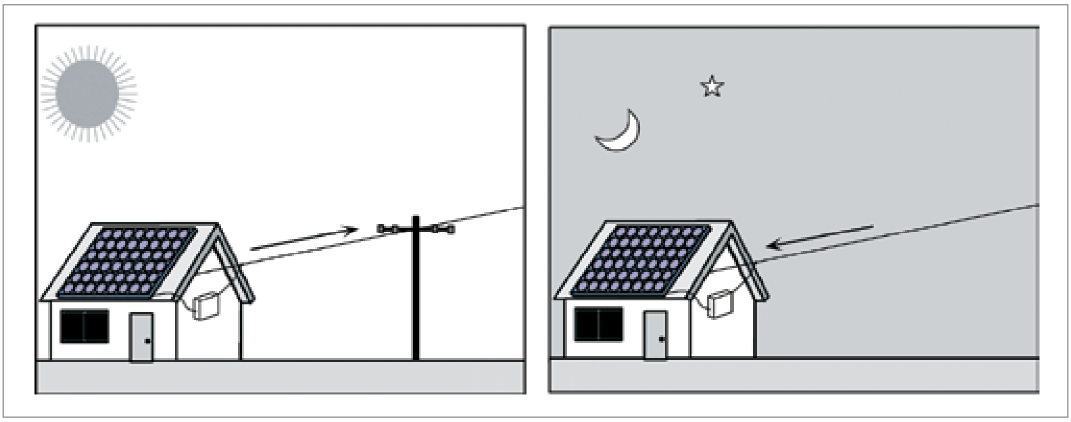


Figure 11: Residential use of photovoltaics — by day, excess power is sent to the grid, and by night, power is supplied to the home (drawing courtesy of CSG Solar Pty Ltd).

direction. This scheme works well in cities like Sydney, states like California and countries like Japan where the maximum demand for electricity is on hot summer afternoons.

No storage is required until more than about 15% of the total electricity supply comes from photovoltaics, since the electricity network is designed to be tolerant to sudden changes in the demand for electricity. These changes in demand are very similar to the changes in supply if, for example, a cloud suddenly blocks sunlight from the solar cell.

One of the reasons rooftop systems have been so popular is that electricity is at its most expensive at the point of use in a home: it is three to four times more valuable here than at the output of a large power station. As photovoltaic costs reduce, rooftop use was identified as the first large-scale application likely to become economically attractive. With the rapid decrease in photovoltaic costs over recent years, due to government subsidies worldwide encouraging such use, "grid parity" — parity with the price of electricity supplied from the grid — has been reached in some parts of the world, and is only a few years off in most of the rest.

The most successful scheme promoting this use has been based on "feed-in tariffs", whereby a homeowner is guaranteed an amount for each unit of electricity the rooftop system sends back to the electricity network. In the early days when solar cells were more expensive, this was far above the price homeowners would pay for conventional electricity, but enough to make the system an economically viable proposition (for example, enough to allow the homeowner to finance a bank loan for the system).

The best schemes have an inbuilt "degression rate": each year, the tariff offered to a homeowner installing a new system is reduced. This forces the companies supplying the systems to concentrate on reducing prices so that they can continue in business.

Schemes like this have driven down the price of photovoltaics dramatically, so much so that we are entering a new phase of "nett metering". In many parts of the world, rooftop photovoltaics are now cheap enough to compete with the costs of electricity normally supplied to the home by the grid network. All that is needed to make the system economically attractive is to pay the homeowner the same rate for electricity exported to the grid as for that bought from the grid.

There are two schools of thought on whether this is a fair deal or not. One is that the situation is very similar to a homeowner installing energy efficient appliances to reduce demand, which most people would agree is a good thing. Even companies that sell power to the homeowner would agree with this, in

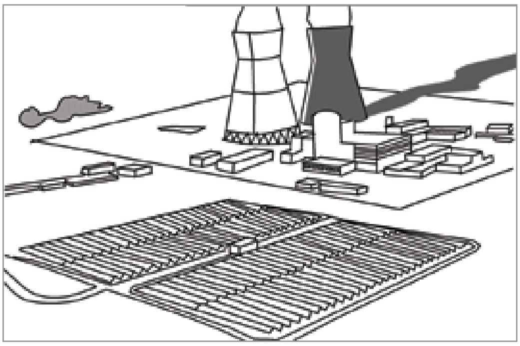


Figure 12: A sketch of the Rancho Secco power station in California with a large photovoltaic field (foreground) installed next to a nuclear power plant — the latter is being decommissioned after increasing safety concerns prompted a public referendum in 1989, sealing its fate.

Figure 13: Average cost of electricity supply in Sydney (1910 – 1988). Source: Rosenthal and Russ (1988).

principle at least, despite the fact their sales are thereby depressed.

The other school of thought is that this exported electricity is no more valuable than that supplied to the grid by a coal-fired power station, located hundreds of kilometres away. Hence it should only be valued at the price of this electricity, which is three to four times lower as previously mentioned.

This is probably a situation where issues, rather than short-term economics, need to be taken into account. Do governments have a role in encouraging what are essentially energy conservation measures financed by private homeowners? Do governments wish to support the development of technology that ultimately has the potential to reduce our dependence on fossil fuels?

These applications will greatly increase the demand for solar cells, which will help further reduce their costs. Ultimately these could become low enough for photovoltaics to compete for electricity installed right at the output of a conventional generator, as in Figure 12.

"LEARNING BY DOING" AND "TECHNOLOGY LOCK-IN"

It has been mentioned above that the more solar cells that are made, the cheaper they are expected to become. This "learning by doing" is quite common for manufactured goods. For example, Figure 13 shows the cost of electricity supply in Sydney over the early years of its supply by fossil-fuel generators. The cost of supply has decreased so much that, today, it is ten times cheaper than it was a century ago.

When you think about it, burning fossil fuels to make electricity is not an inherently cheap process. First the coal has to be mined, then transported to the power station, then burned to make steam that runs a turbine, that turns a generator, that makes electricity. A colossal 60 – 70% of the energy of the coal is wasted in the process, and special cooling systems are needed to get rid of the waste heat. Originally this waste was just dumped as hot water into a nearby river or stream, as was common for other industrial wastes; it was soon realised this was not a good idea, given the serious consequences for the river ecosystems. Similarly, soot and noxious gasses produced by burning coal could just be dumped into the atmosphere, but this caused health problems and deforestation. Companies now have the duty to dispose of all wastes responsibly, with the costs of this factored into the price of electricity.

Even with the initial cost savings provided by cheap waste disposal, fossil-fuel generated electricity was expensive, as seen in Figure 13. Only the very rich could afford it. But evolving technology and the economies of scale drove down prices as the industry grew and now we have cheap fossil-fuel generated electricity, despite the extra costs involved in more responsible handling of wastes than in the early days.

This cheapness poses problems; it is an example of "technology lock-in". Due to more than a century of ongoing development and the investment of billions of dollars, what is inherently a complex and expensive process is now cheap. This makes it very difficult for new technology to displace it, even if it is very much better or inherently cheaper.

We have seen how simple the solar conversion process is. Silicon is all around us, making up over 25% of the earth's crust in the form of compounds with oxygen in rocks and sand. Silicon wafers were expensive when they were only used in microelectronics, but are now much cheaper — and they will become very much cheaper again as unnecessary features needed for microelectronics are dropped and economies of scale kick in, as well as the other benefits of learning by doing.

Even better, my research team has shown that we can make the silicon one hundred times thinner, and yet only drop the performance by one half, by coating thin layers of silicon onto glass. It is very likely that such "thin-film" approaches will eventually exceed the performance of present modules, by increasing the power output by 50% or more while reducing costs.

The other important issue is that electricity production by burning coal is the single largest source of carbon dioxide emissions. There is now no reasonable doubt about the links between the rapidly increasing levels of carbon dioxide measured in our atmosphere, and the increasing global temperatures being measured . Carbon dioxide has to be considered a pollutant which cannot be dumped indiscriminately into the atmosphere, in the same way that we now know dumping excess heat and other wastes into rivers, and soot, sulphur dioxide and ozone into the atmosphere, are not good ideas. The cost of responsibly disposing of this carbon dioxide has to be factored into the costs of electricity production. Unfortunately, there is no known way of doing this cheaply and sustainably, which means we have to find alternatives to the way electricity is produced — and solar is one of the few that can sustainably do this.

SUMMARY

The previous sections have given a whirlwind tour of 20th century physics and microelectronics. We have seen how solar cells operate as 'quantum' devices: each photon in sunlight, if energetic enough, can give one electron flowing in the external circuit connected between the cell terminals. Photons in sunlight can be exchanged for such electrons, ideally on a one-to-one basis.

Solar cells have had fairly limited use in the past but are now starting to appear on the rooftops of private homes. This application, although subsidised in the past, has helped drive down costs to the stage where these systems can compete with the cost of conventional electricity supplied to the homeowner. If this use can be encouraged, solar cell costs will continue to fall rapidly.

Electricity production by the burning of fossil fuels is the single biggest source of carbon dioxide emissions to the atmosphere. Although the associated electricity production process is complex, more than a century of development has led to low costs, despite increasingly stringent requirements upon the handling of wastes from the process. However, no sustainable way of handling the carbon dioxide emitted in the process has been found, meaning that we will have to reduce our reliance on this approach. Solar cells are the best alternative that has yet been demonstrated, as a sustainable way of decreasing our reliance on fossil fuels.

ACKNOWLEDGEMENTS

The author thanks the UNSW School of Photovoltaic and Renewable Energy Engineering for permission to use several drawings from its "Catalogue of Photovoltaic Drawings".


THE AMAZING
COMBUSTO
MAN ON FIRE
ENTRY–
TWO BITS
WORLD'S
GREATEST
MYSTERY!
PLUS – THE BEARDED LADY

SPONTANEOUS HUMAN COMBUSTION

DR KARL KRUSZELNICKI

One of the most bizarre, and still unexplained, things that happens to our human flesh is a phenomenon called "Spontaneous Human Combustion" (*if* it's real). Somehow, a human body bursts into flames — for no apparent reason. Let me emphasise that in true Spontaneous Human Combustion (*if* it really exists) the human body bursts into flames *without touching an external flame or source of heat*. Often the entire body is reduced to ashes.

Now I have to say, right at the very beginning, that I am still not convinced that Spontaneous Human Combustion actually exists. Part of the reason is that nobody has been obliging enough to burst into flames before an expert audience of forensic pathologists.

However, I reckon that it is well-proven that a human body can burn, via the "Wick Effect" (more below). But to start burning without any external source of heat? Not impossible, but definitely not proven.

HISTORY OF SPONTANEOUS HUMAN COMBUSTION

In the Old Testament of the Bible, Ezekiel 28:18 it is written:

"By your many sins and dishonest trade, you have desecrated your sanctuaries,

So I made a fire come out from you, and it consumed you,

And I reduced you to ashes on the ground in the sight of all who were watching."

There have been about 200 reports of Spontaneous Human Combustion over the last 400 years. There is no obvious pattern. The cases are roughly equally divided between male and female. The victims can be any age between four months and 114 years. Some were drunkards, while others were teetotallers. Some had been underweight, while others had been overweight. There's just no simple pattern.

Charles Dickens included an instance of Spontaneous Human Combustion in chapter 32 of his 1853 novel, *Bleak House*, when he decided to get rid of one of his characters in the rather convenient way of having them burst into flames. He writes, describing the room after the paranoid and superstitious Krook has combusted, "There is a little fire left in the grate, but there is a smouldering, suffocating vapour in the room, and a dark greasy coating on the walls and ceiling". Dickens was convinced that Spontaneous Human Combustion existed. In the preface to the second edition of Bleak House he describes how he studied the subject carefully and found about 30 cases of Spontaneous Human Combustion. So he decided to use it in his story. However, his description does not really count as "Spontaneous". Spontaneous means there is no possible external source of combustion — in *Bleak House* there was "a little fire left in the grate".

A typical case of Spontaneous Human Combustion was that of Ann Martin in Philadelphia in 1957. It is said that only her feet and the top of her trunk remained intact, with everything in between reduced to ashes. Newspaper, just 60 centimetres away, was completely untouched.

In your average house fire, the extremities of the body (hands and feet, arms and legs) are burnt to ashes, while remains of the central trunk are always left behind. A forensic pathologist can examine the bones in the pelvis, and tell whether the person was a man or a woman, young or old. If they happened to be a woman, the pathologist can tell whether or not she'd borne children. Whichever the case, the central trunk is *always* left relatively intact in a house fire. But in Spontaneous Human Combustion (if it really exists), the central trunk is reduced to ashes, while the hands and feet are usually left untouched.

To burn a corpse in a crematorium, you need a *lot* of heat. They will typically burn the body at 1200 degrees Celsius for an hour and a half, and then at 1000 degrees Celsius for another hour and a half. Even then, they're not left with just dust and ashes but bone fragments — so they have to grind these down with a mortar and pestle. In even the worst house fire you won't get a temperature above 850 degrees Celsius — so where does the energy for this human fireball come from?

THE WICK EFFECT (TOTALLY PROVEN)

Dr J. D. DeHaan from the California Criminalistics Bureau carried out some rather convincing studies that showed that a pig body can actually burn. It's now popularly called the "Wick Effect".

He chose a pig carcass, because pigs are surprisingly similar to humans. Our diets and physiology are similar. In fact, we are so similar that a heart valve can be taken from a pig and successfully transplanted into a human heart, and work well for many years.

There is enough fat in a pig body (or a human body) to support combustion. Burning a gram of fat will release about 40 kilojoules of heat. If you set it up correctly, you can start a smouldering fire that is fuelled by the fat, and much of the central trunk can be consumed and turned into ashes. It's important that you burn the carcass only inside a room with a closed door. Air can get in, but not as quickly as it could through an open door or window. It turns out that for the Wick Effect to happen, you need to slightly limit the amount of air, so that you get a smouldering fire. The Wick Effect won't happen out in the open with ready access to air. For example, the clothing won't turn into a porous wick, but will burn totally into ash. No "wick" means no "Wick Effect".

First, he placed some clothing on a small pig (this would later act as a wick). Then he poured one litre of petrol on the clothing. Some soaked into the clothing, while some soaked into the carpet under the pig carcass. He lit the petrol. The petrol burnt strongly for about a minute, emitting about 60 kilowatts. Within that first minute, the top layer of the pig skin had dried, scorched and begun to split and curl. Liquid fat began to ooze out. The clothing had also scorched. It did not combust into powdery ash, but turned into a porous, stiff, blackened substance — in other words, a wick. It sat on top of the pork skin, soaking up the liquid fat.

Second, after about a minute, the petrol fire went out, but the carpet continued to burn. It burnt with about the same intensity as the petrol, but for about another 25 minutes, after which it extinguished. Like the petrol fire, the carpet fire emitted heat at about 60 kilowatts.

Third, within a minute of both the petrol extinguishing and the carpet beginning to burn, the pork fat began to burn. It was absolutely essential that a rigid, carbonaceous, char wick was present. The pork fat burnt where there was a "wick" (clothing), and did not burn where there was not a wick.

You can see this effect in a burning candle. A pool of molten oil surrounds the central burning wick. But the flame comes only from the wick. You can touch a burning flame (from a cigarette lighter or burning match) to the hot pool of molten oil — but it will not burn. (Apparently, the porous wick will "increase the effective vapour pressure of the liquefied fat fuel.") You can see the link to a human body burning on the trunk, but not the extremities — you need clothing, which can turn into a wick, to support the burning.

After about 10 minutes, the heat put out from the burning fat was great enough to keep itself burning. The pig carcass then burnt for about two hours, " … small 4-6 cm flames were sustained continuously under (the left side) of the carcass". The pig carcass lost mass at about 1 to 2 grams per second. The heat output was about 35 to 40 kilowatts.

Dr DeHaan tested several pigs — some lean, some fat. As you would expect, the more fatty pig carcasses burnt for longer, generated more heat and consumed more of the carcass.

In the case of a large 95-kilogram pig, the carcass generated about 60 kilowatts, burnt for 6½ hours, had flames up to 30 centimetres high for most of that time, and lost 60 per cent of its initial weight.

BROWN FAT (TOTALLY UNPROVEN)

Okay, so a human body will burn, thanks to its fat. But you need an external source of heat to start it burning. Is it possible to generate lots of heat from a human body without a flame?

It's not impossible (but it's totally unproven, so far).

First, when we adults want to generate heat, we have to move a muscle or a *bunch* of muscles. Either we pump iron and get warm that way, or else we run around the block, if we are feeling athletic. If we are pretty lazy, we will just sit there and shiver. Either way, if we want to generate heat we have to move muscles.

But newborn babies in the first few months of their life cannot make their underdeveloped muscles shiver. Luckily, they have a unique trick for generating heat. They have a substance called "brown fat".

Brown fat is different from white fat. White fat takes in energy and turns it into more white fat. (It turns out that white fat does a lot of other stuff as well, and is quite an active organ, and even makes hormones — but I won't go into that now.) Brown fat goes the other way — it turns stored brown fat directly into heat. It bypasses the muscle-shivering method of making heat.

It was wrongly thought until very recently that adults don't have brown fat. However, radiologists have been seeing brown fat on their MRI scans for years. They simply hadn't bothered to tell the hormone doctors (endocrinologists) who were wrongly convinced that adults don't have brown fat.

BIG FINISH

Maybe, in Spontaneous Human Combustion (if it really does exist), a person has brown fat, and *then* has an abnormal runaway reaction that generates enough heat to set the body alight. (I have no idea of what this abnormal runaway reaction might be. Okay, I have a suspicion. Look up "decoupling of oxidative phosphorylation", if you have a spare hour or two.)

But in the early 21st century, we really don't know. The closest medical phenomenon to Spontaneous Human Combustion that we know of is Malignant Hyperthermia. In this rare (and often fatal) disorder, under the right circumstances a person's body temperature can get as high as 47 degrees Celsius. But that's a long way from burning a body.

I really didn't think that Spontaneous Human Combustion might be possible, until I read a report in *New Scientist*. It was by a retired Scenes of Crime Police Officer, John Heymer, from Gwent in the UK. He described something that seemed to be a classic case of Spontaneous Human Combustion, including the total absence of a source of flame or heat in the room. But at the time, he couldn't describe it as such in his official report, or else he would have been laughed out of the Police Force.

So as for me, I will just remain vaguely dubious, until a really well-documented case turns up.

METAMATERIALS AND THE SCIENCE OF INVISIBILITY

PROFESSOR SIR JOHN PENDRY

A CLOAK OF INVISIBILITY

To be invisible is somehow magical. The possibility has been dreamt of for centuries, but only with the advent of transformation optics and metamaterials — artificially engineered materials with properties that arise from structure rather than chemistry — has the dream approached reality. To be scientific about the matter: invisibility requires two properties of a cloak. First the cloak must ensure that no external source of light is reflected, either from the hidden object or from the cloak itself. This requirement is relatively easy to achieve with a pot of black paint, or its equivalent for absorbing microwaves if we require invisibility to radar. In essence, conventional stealth operates in this fashion. The second requirement is that no shadow be cast; see Figures 1 and 2. Removing a shadow is a much greater challenge, but one that transformation optics has successfully addressed.

Our brains assume that light travels in a straight line to reach our eyes as if it travelled along a rigid rod. Suppose that the rod were not rigid, but that the middle bit was fluid and could be bent so as to curve around the hidden object (Figure 3). Our brains would be none the wiser because rays would reach our eyes exactly as before with no hint of the curved path that they had in fact taken.

Transformation optics enables us to bend light in this way by prescribing the exact metamaterial properties needed to achieve the required bending. Metamaterials can make light flow like water around the hidden object, smoothly closing in behind the object to leave no trace of its presence.[1, 2, 3]

There are naturally occurring instances of light moving in a curve. When the sun heats sand in the desert, air immediately above the sand is also heated. It becomes less dense and therefore less refracting. This effect tails off with height above the sand as the air temperature returns to a relatively cool value. This gradient in refractive index bends light from the sky, and if the refractive index gradient is strong enough the light is reflected as if the desert surface were covered with a layer of reflecting water rather than sand and hot air — hence the disappointment of a thirsty traveller who thinks he sees an oasis.

In manufacturing a cloak we exploit the potential of an index gradient to bend light, but in a precisely controlled fashion. Mirages are also commonly seen in the summer time on hot roads which can appear to be wet due to the effects described (Figure 4).

METAMATERIALS — WHAT ARE THEY?

The idea that internal structure could influence a material's response to light has been debated since before the time of James Clerk Maxwell. In the 1950s researchers built structures made from thin metallic wires to

Fig. 1a Fig. 1b

Figure 1. (a-b) Peter Pan is not invisible if we only paint him black. Seen against a bright background a black figure is perfectly visible.

Fig. 2a Fig. 2b Fig. 2c

Figure 2. (a-c) To be truly invisible Peter must lose his shadow.

mimic in the laboratory the properties of the plasma that constitutes the ionosphere. In the 1990s, when work I was involved with at the Marconi Company produced the so-called 'split ring' metamaterial pictured in Figure 5 [4], it was finally realised just how radically structure could influence properties. Manufactured by etching a copper circuit board with cell dimensions a few millimetres across, it had a resonant response to radar signals at around 10 GHz. This artificial structure gave for the first time a magnetic response in the opposite direction to the applied magnetic field.

The split ring structure enabled David Smith, then at the University of California San Diego, to demonstrate the first material with a negative refractive index, as prescribed many years previously by Victor Veselago[5,6]. The missing ingredient, negative magnetism, was now made possible by the new structure.

The response of conventional materials to light is dictated by their chemical composition, but the internal structure of a material can have an even stronger influence. A silver mirror is highly reflecting, but a black-and-white photograph owes its blackness to billions of nanometre-scale silver spheres embedded in the film. Restructuring the silver into spheres, much smaller than the wavelength of light, produces a dramatic effect.

Macroscopic electromagnetic fields are averages over fluctuating local fields — averages that are very well defined, as there are typically billions of molecules contained in one cubic wavelength of matter. Metamaterials extend this concept, replacing the molecules with man-made structures that might have dimensions of nanometres for visible light or, in the case of GHz radiation, may be as large as a few millimetres, but still much less than the wavelength. In this way properties are engineered through structure rather than through *chemical composition* — see Figure 6.

Possibilities are limited only by our imagination, not by the number of elements in the periodic table. In the past decade exploitation of metamaterials has exploded, making available optical properties such as negative refraction never found in nature, and devices such a cloaks of invisibility previously thought to be impossible.

In Figure 7 we see another example: a metamaterial from the Wegener group in Karlsruhe[7] (shown in false colour). Fashioned from microscopic helices of gold, its response depends strongly on whether the light is left or right circularly polarised, much more strongly than a solution of sugar molecules.

TRANSFORMATION OPTICS

Full exploitation of the potential of metamaterials requires a sophisticated design

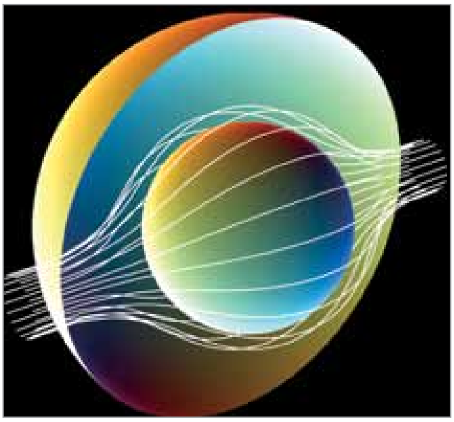


Figure 3. Trajectories of rays through the cloak showing how they avoid the cloaked region and return to their original path after traversing the cloak.

Figure 4. Mirages are also commonly seen in the summer time on hot roads which can appear to be wet due to the effects described.

tool. Conventional optics simply redirects light at a series of interfaces between homogeneous materials within which rays travel in straight lines. Metamaterials have the potential for more sophisticated control and can force light to travel along a curved path, opening the possibility of devices such as cloaks of invisibility. Transformation optics is the theoretical tool that tells us what kind of metamaterial will bend light along a given path[1, 8].

In 1919 a remarkable experiment conducted during a solar eclipse showed that the Sun acts like a giant lens, slightly bending rays of starlight that pass close to the Sun's disk. Einstein had predicted that the gravitational field of the Sun would distort space and, as far a light is concerned, the squashed space near the sun would appear to have a refractive index greater than unity. This experiment and its confirmation of the prediction made Einstein the legend we know today. Very helpfully for metamaterials, Einstein produced a formula relating the distortion of space to changes in effective refractive index.

What transformation optics seeks to achieve is to take a ray of light, which might originally be travelling in a straight line in free space, and distort its trajectory in any way we please. In principle we could do this by distorting space itself using extremely massive objects but, thanks to Einstein, this is not necessary — changing the refractive index will have the same effect as far as light is concerned. Figure 8A (top) shows a ray of light travelling through undistorted space. I have drawn in some coordinate lines so that when space is distorted as in Figure 8B (bottom) we can see the distortion bending the coordinates. The ray behaves as if it is nailed to the coordinate system. All we have to do to control the rays is to choose an appropriate distortion of coordinates. Mathematically, this is known as a coordinate transformation, and Einstein's formula, given the coordinate transformation, will dictate the refractive index.

ACTIVE METAMATERIALS

Metamaterials already offer a rich field of exploration for manipulation of light in entirely novel ways, but researchers want to go further. Current interest is focussed on whether their properties can be switched, either by passage of an electrical current or by a source of controlling light powerful enough to change the metamaterial's properties.

For example, if the metamaterial contains a semiconductor, illumination by a powerful light source will excite charge carriers in the semiconductor and increase its conductivity. This in turn will alter how the metamaterial responds to a second light beam, enabling us to modulate the intensity of the second

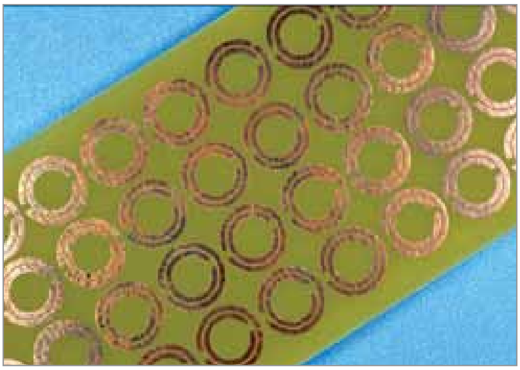


Figure 5. The first metamaterial to show 'negative magnetism', manufactured at the Marconi Company by Mike Wiltshire and designed to work at GHz frequencies. The response on this structure lies in the opposite direction to the applied magnetic field.

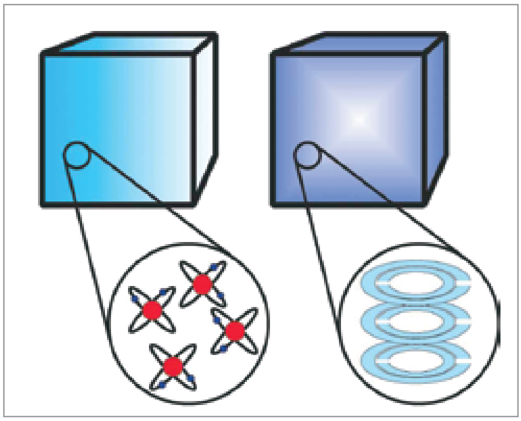


Figure 6. In conventional materials their properties derive from the constituent atoms; in metamaterials, they derive from the sub-units which may contain many atoms.

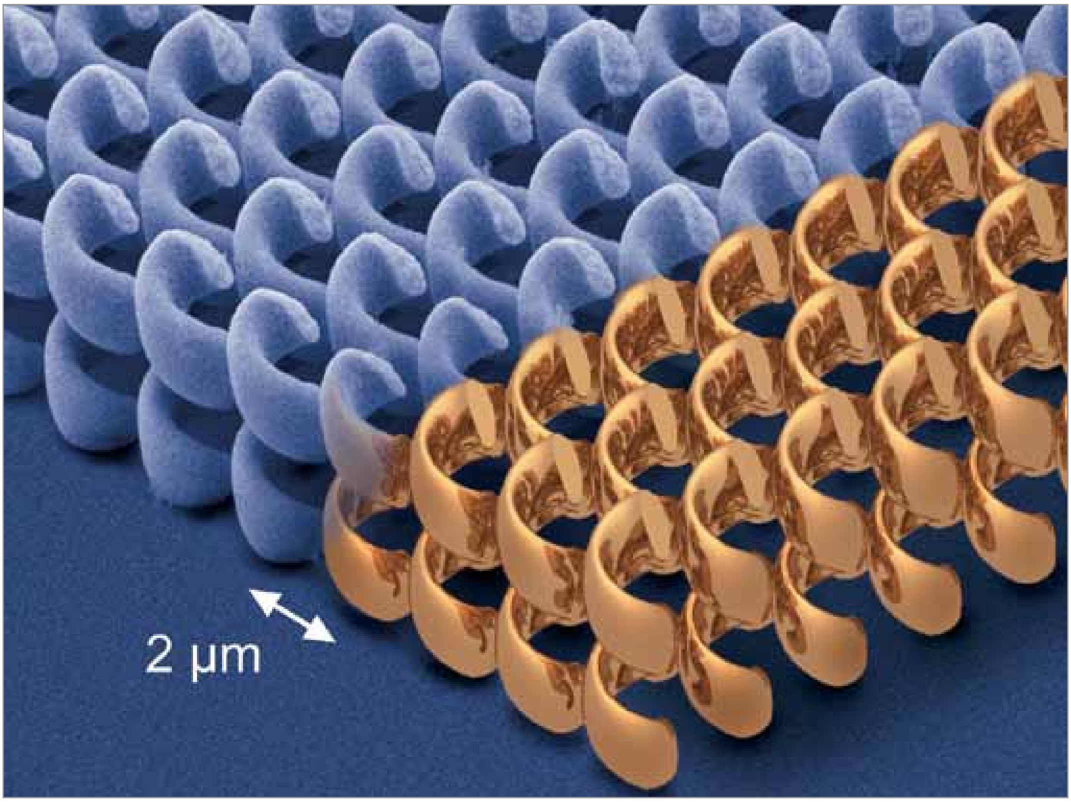


Figure 7. A more recent structure from the Wegener group in Karlsruhe working at visible frequencies; the structure shows extreme chirality— its response depends on whether the light is left or right circularly polarised.

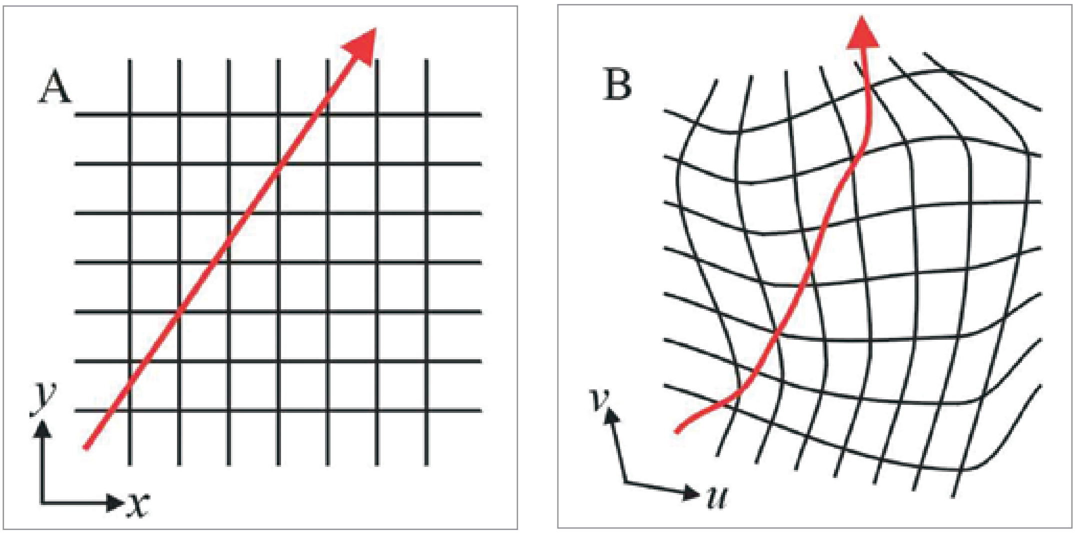


Figure 8. Left: in free space a ray of light travels in a straight line. Right: distortion of space as if it were a rubber sheet distorts the trajectory of a ray passing through. The coordinate system records the distortion.

beam, or alternatively to refract it into a new direction. Thus metamaterials with their superior control of light and the emerging possibility of switching and modulation are making light almost as controllable as electrons in computer chips.

A major challenge for active metamaterials is the amplification of light. Many of the more exciting predictions for novel properties assume ideal metamaterials made of lossless components. In reality there is always a finite amount of light lost in a metamaterial and in some instances this can have a serious effect on performance. For instance some of the more unusual properties, such as negative refraction, require that light is trapped inside the structure whilst it is manipulated. Usually this is done by building tiny sub-wavelength resonators into the structure. Figure 9 shows the so called 'fishnet' structure made by the Xiang group at the University of California, Berkeley[9], comprising alternating layers of silver and dielectric, and designed to have a negative refractive index for visible light. The silver-dielectric combination traps light between the layers of silver. However, the increased dwell time allows further opportunities for loss and reduces the performance of the metamaterial.

To deal with the issue of loss the Shalaev group at Purdue[10] has introduced into the structure an optical gain (amplification) material: a dye, as shown in the lower panel in Figure 9, where a cut-away diagram shows how the dye is introduced. A powerful light source pumps the dye molecules into an excited state, subsequently enabling them to give up their energy to a second light beam and thereby to amplify it. Here serendipity intervenes: the trapping effect that made the fishnet structure absorb light rather strongly now operates in the opposite direction. The longer that light can dwell in the structure the more time it has to benefit from the amplification process.

CURRENT CHALLENGES FOR METAMATERIALS

The advent of metamaterials has inspired renewed interest in the fundamentals of electromagnetism. In part our dreams have been realised, but to some extent they have outrun our capacity to make them happen. What are the roadblocks?

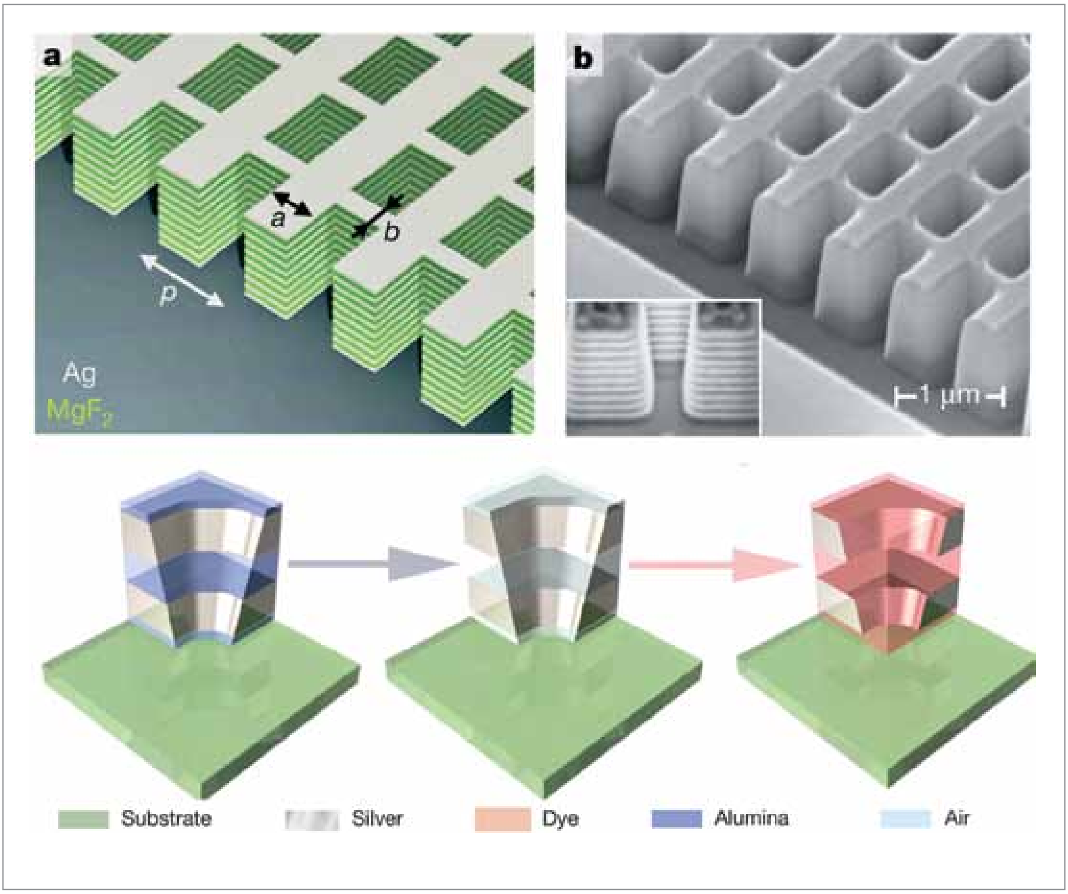


Figure 9. Top: diagram and scanning electron microscope image of a 'fishnet' structure fabricated by the Xiang group.

Top (a): diagram of the 21-layer fishnet structure with a unit cell of p = 860 nm, a = 565 nm and b = 265nm . Top (b): scanning electron microscope image of the structure with the side etched, showing the cross-section. The structure consists of alternating layers of 30 nm silver (Ag) and 50 nm magnesium fluoride (MgF_2), and the dimensions of the structure correspond to the diagram shown in a). The inset shows a cross-section of the pattern taken at a 45-degree angle.

Bottom: schematic of the fabrication process for the Shalaev structure into which gain is introduced. Left: one-quarter of the structure with an alumina spacer. Middle: after etching the alumina, the structure has air or solvent as the spacer with alumina pillars as support. Right: finally the gain material (an organic dye shown in pink) is introduced into the structure.

Fundamental to the metamaterial concept is the ability to structure a material on a scale less than the wavelength of radiation. This is not a problem when the radiation is a mobile telephone signal with a wavelength of around 30 cm, and this explains why much of the early experimental work was with radiation of this wavelength. Visible light presents a greater challenge: whereas traditional technologies have required manufacturing tolerances of better than 1 micrometre, metamaterials are now demanding nanometre-scale accuracy. Some technology exists, such as ion beam etching, but it is expensive and handles only small samples. Investment in nano-manufacturing is required for further progress.

To make a metamaterial we have to start from an ordinary material from which the nanostructures are built, and ultimately the quality of the metamaterial is limited by the ingredients from which it is made. At mobile telephone frequencies metals are excellent conductors, and the tiny metallic elements which constitute the fine structure of many metamaterials are relatively loss-free. However, the response of metals to visible light is limited by the mass of the metal's electrons: the response becomes more sluggish and losses, even in an excellent conductor such as silver, are much greater.

One approach accepts the fact of loss and seeks to compensate by introducing a gain medium, as mentioned earlier. This may be a medium containing dye molecules activated by an external source of energy, or an array of quantum dots pumped into excited states by an electrical current.

What is not in doubt is the excitement created by the new metamaterial world, which has liberated theorists to dream of what previously thought impossible — and challenged experimentalists to make it happen.

Footnotes

1 J.B. Pendry, D. Schurig, and D.R. Smith, *Science* **312** 1780-2 (2006).

2 U. Leonhardt, *Science* **312**, 1777 (2006).

3 D. Schurig, J. J. Mock, B. J. Justice, S. A. Cummer, J.B. Pendry, A. F. Starr, D. R. Smith, *Science*, **314**, 977-80 (2006).

4 Pendry, J.B., Holden, A.J., Robbins, D.J., and Stewart, W.J., 1999, *IEEE transactions on microwave theory and techniques*, **47**, 2075.

5 Veselago, V.G., 1968, *Soviet Physics USPEKHI*, **10**, 509.

6 Smith, D. R., Padilla, W.J., Vier, D. C., Nemat-Nasser, S. C., and Schultz, S., 2000, *Phys. Rev. Lett*., 84, 4184.

7 Gansel, J.K., Thiel, M., Rill, M.S., Decker, M., Bade,K., Saile, V. von Freymann, G., Linden, S., Wegener, M., 2009, *Science*, **325**, 1513.

8 Ward, A.J., and Pendry, J.B., 1996, *Journal of Modern Optics*, **43**, 773.

9 Valentine, J., Zhang, S., Zentgraf, T., Ulin-Avila, E., Genov, D.A., Bartal, G., & Zhang, X., 2008, *Nature*, **455**, 376.

10 Xiao, S., Drachev, V.P., Kildishev, A.V., Ni, X., Chettiar, U.K., Yuan, H.K., & Shalaev, V.M., 2010, *Nature*, **466**, 735.

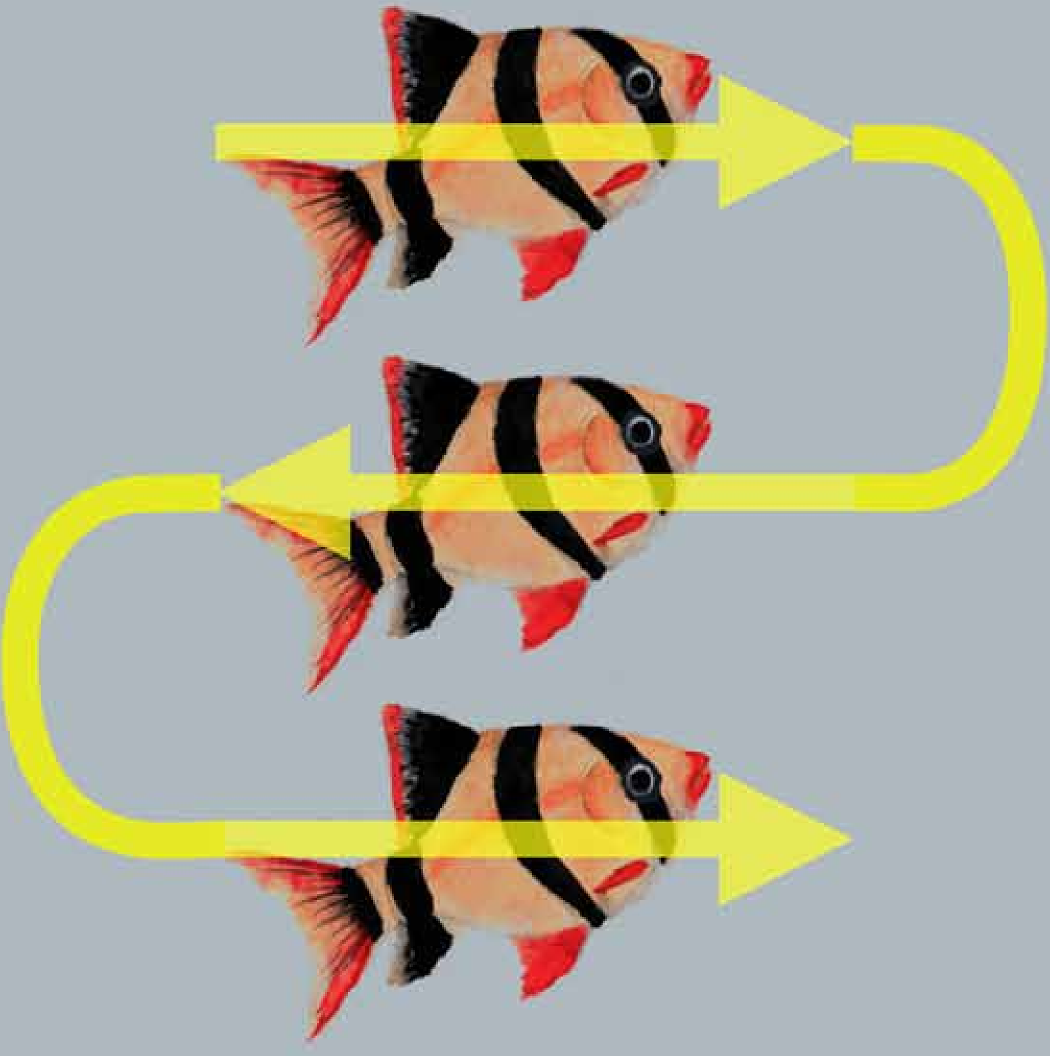
Other references and further reading

A selection of articles at both advanced and popular level can be found at my website: http://www.cmth.ph.ic.ac.uk/photonics/Newphotonics/refs.html

and at: http://people.ee.duke.edu/~drsmith/index.html

Optical Metamaterials: Fundamentals and Applications by W. Cai and V. Shalaev (Springer)

Physics and Applications of Negative Refractive Index Materials by S. Anantha Ramakrishna and Tomasz M. Grzegorczyk (CRC press)



NEGATIVE REFRACTION AND A PERFECT LENS

PROFESSOR SIR JOHN PENDRY

BENDING LIGHT THE WRONG WAY

One of the driving forces behind the early development of metamaterials was Victor Veselago, a Russian scientist, who predicted the possibility of a *negative refractive index*. Such a material would bend light the 'wrong' way at interfaces. Figure 1 shows negative refraction in action and demonstrates the ability of negative refraction to focus light without the aid of curved surfaces.

Objects immersed in water appear closer to the surface than they really are: swimming pools or ponds can appear to be shallow when they are not. The apparent depth is related to the real depth by the refractive index of the water: the larger the index, the closer an object appears relative to the surface. The logical consequence of this law is that a **negative** refraction would make the object appear to be **above** the surface: a fish pond filled with negatively refracting material would seem to have fish floating in the air above the pond. This prediction, first made in 1967[1], remained untested until the deployment of metamaterials.

The argument for negative refraction is technically complex. Light consists of both electric and magnetic fields, and energy is shared equally between the two fields. Light propagates by tossing energy back and forth between the two, as if the electric and magnetic components were in a dance.

In a material with a positive refractive index — water, glass or most other transparent materials — the material responds to each of the two fields by acquiring a polarisation (electric and magnetic) so as to slow down the dance: light propagates more slowly within the material. Veselago's idea was that a material that responded to fields by polarising in the opposite direction would reverse the dance and create the effect of negative refraction. His dream remained unrealised until the advent of metamaterials that could create this response.

There was something decidedly peculiar about waves travelling in a negatively refracting medium: theory predicted that the energy flow would be in the opposite direction flow of the wave fronts — the implication being that a wave packet travels in the opposite direction to the waves it contains (see Figures 2 and 3).

These novel concepts generated considerable interest at the time, but absence of any material realisation led to their neglect. Veselago's recipe for negative refraction requires that a material's response to an electric field, described by the permittivity ε, and response to a magnetic field described by the permeability μ, should both be negative, and the simultaneous realisation of negative values for these quantities proved elusive.

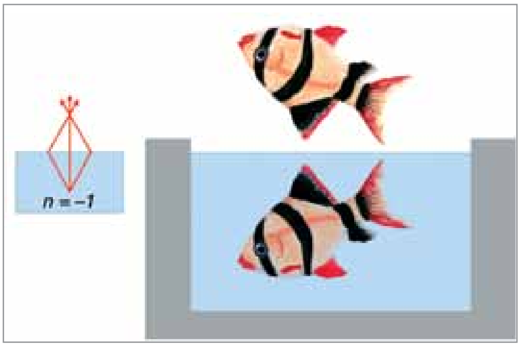


Figure 1. Negative refraction bends light 'the wrong way' at an interface (see inset top left) and as a result can refocus light as it emerges from a negatively refracting medium into air. This implies that objects such as fish swimming in a negatively refracting fluid appear to an external observer to be floating in the air above.

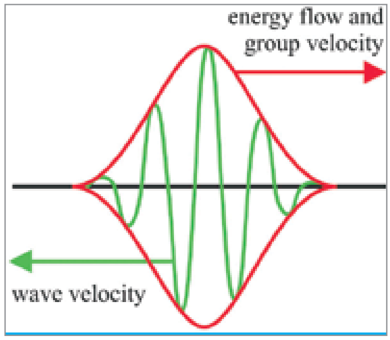


Figure 2. Wave packets in negatively refracting materials show a split personality: the envelope of the packet goes one way and the waves inside head off in the opposite direction

However, in recent times it has been proposed that novel electromagnetic properties can be realised by micro-structuring a material on a scale much smaller than the wavelength[2]. These advances, described in the previous chapter, enabled the field to move into an experimental phase; a short review of progress can be found in Reference 3. It was Smith *et al.*[4] who first constructed and demonstrated a material in which the sign of the electrical permittivity ε and magnetic permeability μ were simultaneously negative, and the same team verified the predicted negative refraction angle in a subsequent experiment[5]. Their work led to an explosion of interest in the field.

These conclusions on negative refraction have been disputed by some. We have seen that negative signs carry certain implications with them and have hidden consequences that are easy to confuse with inconsistencies. However, extensive computer simulations and further experiments have greatly clarified the picture. As a result, by mid 2003 I was able to write in a piece for *Nature* journal's "News and Views"[6] that these materials were now *Positively Negative*. This period of controversy ensured that the concept of negative refraction had a full and vigorous public debate before being accepted as valid.

The experiments of Parazzoli *et al.* at Boeing[7] were amongst several experimental and computational demonstrations of negative refraction that confirmed the earlier results. The Boeing team prepared two prisms shown in Figure 4: one negatively refracting, carefully constructed from low-loss material, the other positively refracting but of the same physical dimensions. The deviation of a microwave beam freely propagating in air was measured for each sample and the results are shown in Figure 5 — the clearest demonstration that these media refract radiation through negative angles.

The new materials have many strange properties: a reversed Doppler shift, and Cerenkov radiation that emerges in the opposite direction to conventional radiation. The strangest property, following directly from negative refraction, is the ability of the material to focus light (to which we have already alluded). We illustrate this effect in Figure 6 for the case $n = -1$ when focussing is free from aberration, and all rays radiating from a point source are brought to a double focus: once inside the slab, and once again outside. We also choose $\varepsilon = -1$, $\mu = -1$, so that the material's impedance matches the vacuum and there is no reflection at the interface. This is a rather good lens: free from aberrations and free from

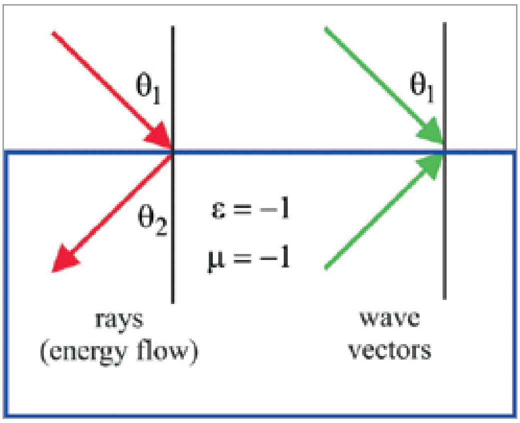


Figure 3. In a negatively refracting material, light makes a negative angle with the normal. To the left of the figure we show the direction of the wavepackets, to the right the flow of the waves.

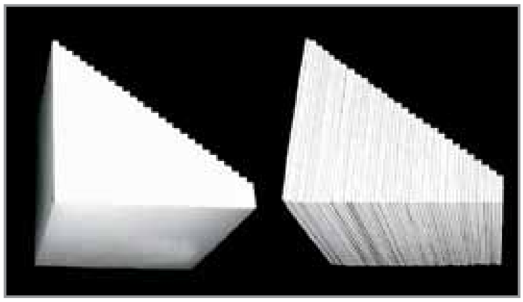


Figure 4. Boeing PhantomWorks 32° wedges. Right: negatively refracting sample; left: Teflon.

spurious reflections. In fact we shall see that it is even better than that!

A lens functions by acting on all rays radiated from an object, restoring each to its proper phase at the image plane. A conventional lens ensures that each ray has the same optical path length to the image plane and therefore all the phases are correctly reproduced, except for a constant additional term. Our new lens is even more efficient. As wave packets (see Figure 2) enter the negative slab the phase 'clock' is set in reverse and winds backwards in angle so that propagation through a thickness d of this medium unwinds the phase acquired passing through the same thickness of vacuum. Wave packets arrive at the image plane with exactly the same phase as in the object plane. This is another clue to the superior nature of this lens.

Now we know that we can build on this foundation and develop new applications based on the concept. Many applications are being tested for antennae and radar devices, where these novel properties give added flexibility in design. New radar lenses are being developed which offer compact and relatively aberration-free performance. Waveguides have been constructed that show novel dispersion characteristics as a consequences of negative fillings.

NEGATIVELY REFRACTING METAMATERIALS

Negative refraction requires a material in which, for both magnetic and electric fields, the response of the material lies in the opposite direction to the applied field. Permittivity ε and permeability μ of conventional materials derive from the response of constituent atoms to applied fields: ε and μ represent an average response of the system. On a length scale much greater than the separation between atoms all we need to know about the system is given by ε and μ.

Metamaterials carry this idea one step further: the constituent material is structured into sub-units, and on a length scale much greater than that of the sub-units properties are again determined by an effective permeability and permittivity. In the case of electromagnetic radiation this usually means that the sub-units must be much smaller than the wavelength of radiation, as described in the previous chapter.

In this way the properties of a complex structure can be summarised by the effective values ε_{eff} and μ_{eff}, which is a great simplification in our thinking. So the concept is a familiar one, but with the difference that the sub-units can take very many forms. This flexibility in design enables metamaterials

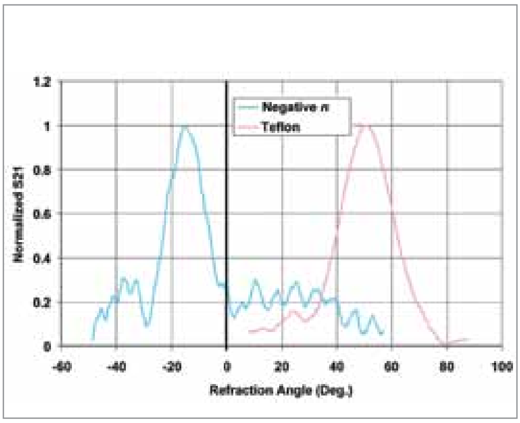


Figure 5. Angle of refraction by each of the two prisms shown in Figure 4: blue — data for negative sample; red — data for positive sample. The data clearly demonstrate that refraction is reversed in a negatively refracting medium.

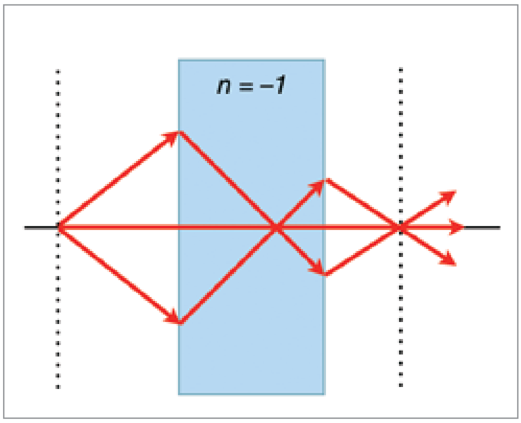


Figure 6. A negative refractive index medium bends light to a negative angle relative to the surface normal. Light formerly diverging from a point source in the object plane is set in reverse and converges back to a point. Released from the medium the light reaches a focus for a second time in the image plane.

to have values for ε_{eff} and μ_{eff} that are not encountered in nature, and in the present context that will mean one or both of these parameters being negative. Furthermore, these materials can give magnetic activity at frequencies where previously materials have been thought of as magnetically inert.

Particular flexibility is given to us in the design of these materials by the very large difference in conductivity between insulators and metals. Therefore all the metamaterials we discuss here will owe their activity to the metallic content. Perhaps the simplest possible example is seen in Figure 7.

This was the first structure designed by the Marconi collaboration [8] and simulates the properties of a low density plasma having a permittivity of the form,

$$\varepsilon_{metal} = 1 - \omega_p^2 / \omega^2$$

sketched in Figure 8. Why this should be the case can be understood as follows: a plasma comprises a gas of charged particles (take for example a metal where the particles are free electrons). The plasma frequency is dictated by the density of charges and by their mass.

$$\omega_p^2 = ne^2 / \varepsilon_0 m_e$$

In metals this gives a plasma frequency typically in the ultraviolet region. However, in our wire structure the electron density is much reduced relative to solid metal. Furthermore there is an additional contribution to the effective mass of the electrons from the inductance of the wires. Typical dimensions of a few microns radius for the wires, and a spacing of a few millimetres, give effective masses to the electrons of the same order as a nitrogen atom, and as a result the plasma frequency is depressed into the GHz region. Note how this very radical change in dielectric properties is brought about by an extremely small amount of metal. In our example the concentration of metal is only a few parts per million, comparable with the levels of dopants in a semiconductor.

An example of metal content playing quite a different role is shown in Figure 9 where we see a structure designed to be magnetically active at GHz frequencies. The structure has a negative magnetic permeability around 10GHz and has been the basis for many studies of negative materials. A magnetic field applied normal to the plane induces electrical currents around the magnetic ring, except that the ring is split so that at some point the current has to jump across the gap between the two rings — which it can do because of the capacitance

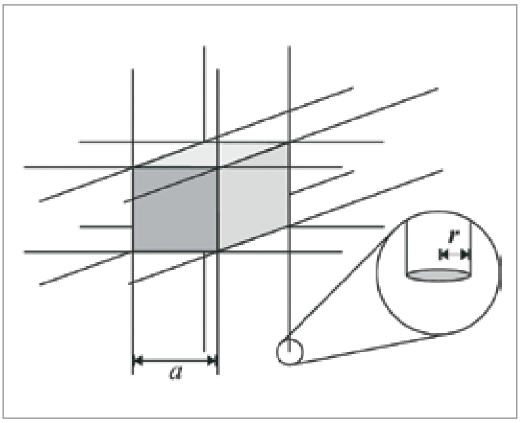


Figure 7. A metamaterial with ε < 0: a periodic structure composed of thin infinite wires arranged in a simple cubic lattice, which mimics the response of a plasma. In a typical example the wires might be a few tens of microns in diameter and spaced by a few millimetres, giving a plasma frequency in the GHz range.

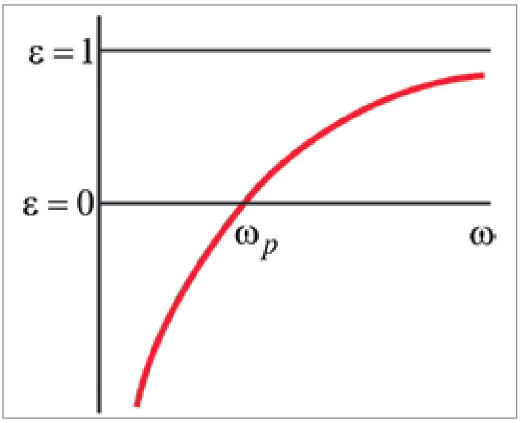


Figure 8. The schematic permittivity, ε, of a plasma: below the plasma frequency, ω_p, ε is negative.

between the two rings. The flow of current in turn creates a magnetic field. In fact the flow of current in an inductive ring through a capacitance forms a resonant circuit, and hence gives a characteristic resonant response to a magnetic field shown in Figure 10.

Loss is always an issue in negative materials because it attenuates the effects we seek to create. Curiously, the main source of loss encountered appears to be the surrounding dielectric material in the structure rather than the resistivity of the metal. High quality dielectric substrates are vital to good performance. With a highly active structure and low loss components the critical value of $\mu_r = -1$ can be associated with very small values of loss as measured by μ_i.

A PERFECT LENS

As I hinted previously, the focussing properties of a negative slab are rather unusual, especially for the case where ε = –1, $\mu = -1$. In this instance, rays are drawn to an aberration-free focus, and suffer no reflection from the surfaces of the slab. Yet the focussing properties are even more remarkable than this: further investigation shows that the slab is free from the wavelength restriction on resolution.[9] In the ideal limit of ε = –1 and $\mu = -1$, resolution increases without bounds.

First a few words on the wavelength limit to resolution. In Figure 6 an object emits electromagnetic waves of frequency w. Each wave has by a wave vector, **k**, where,

$$k_z = \sqrt{\mathrm{w}^2/c_0^2 - k_x^2 - k_y^2}$$

is responsible for driving the wave from object to image, and k_x, k_y define the Fourier components of the image. The larger the magnitude of k_x, k_y we can propagate to the image plane, the better the resolution. The problem is that making these transverse wave vectors too large gives k_z an imaginary value and the wave decays exponentially along the z-axis. These decaying components of the object field are often referred to as the 'near field'. They are confined to the vicinity of the object and serve to lock away high-resolution information. Hence the biggest Fourier component that we can capture has magnitude $k_0 = \omega/c_0$ and the wavelength restriction on resolution follows.

In the previous chapter we talked about transformation optics and how this tool enables us to bend light almost as we please. We imagined that the light rays were embedded in a rubber sheet so that, by squashing or stretching the sheet, the rays of light would also be distorted. In effect the

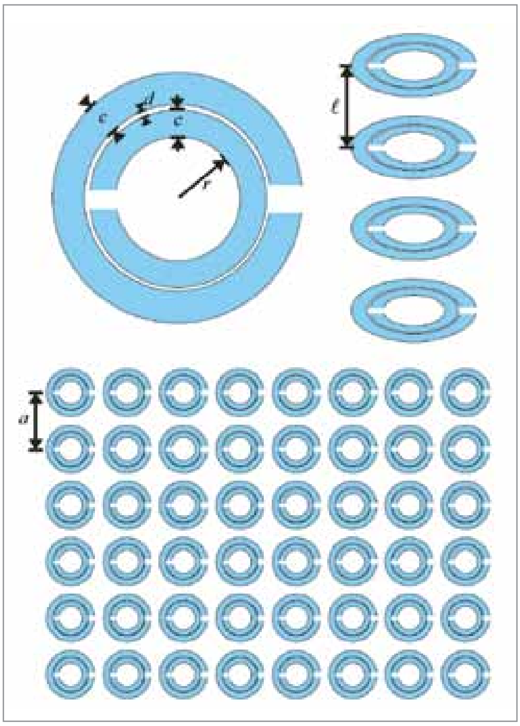


Figure 9. This metamaterial is designed to give a magnetic response to an external magnetic field in the GHz region of the spectrum: rings are manufactured in layers which are then stacked to form an array of resonant columns. Typically the lattice constant, a = 10mm, is much less than 30mm, the wavelength at these frequencies.

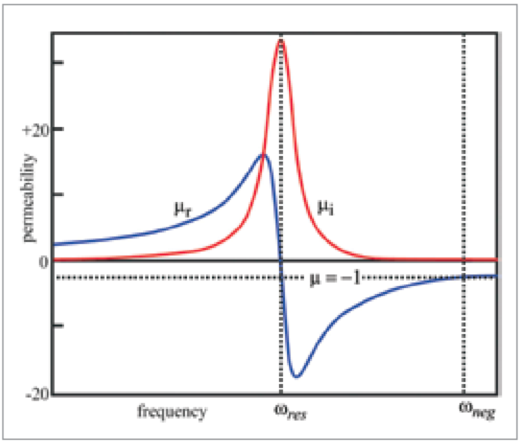


Figure 10. Schematic permeability of the magnetic metamaterial shown in Figure 9 showing the resonant response of the structure at ω_{res}.

squashing of the imaginary rubber material produces an increase in refractive index, whereas stretching has the opposite effect.

Suppose we start with ordinary flat space. Next we start to compress part of the space contained between two planes. The more we compress, the higher the effective refractive index of the squashed region. If we push hard enough we can squash this region of space into zero volume, with an infinitely high refractive index. We can even imagine squashing space further so that it folds back on itself, just like a sheet of metal would buckle under compression. The resulting folded space, shown in Figure 11, has some interesting geometry in that a single point exists on three separate layers (or manifolds) of space. Hence a fish on one side of the negative index materials comes into focus three times, the middle focus being inverted.

An alternative view of the negatively refracting material is as “optical antimatter”, in the sense that a layer of negative metamaterial cancels the effect of propagation through a piece of positive space. An ordinary camera lens partly achieves this, but not perfectly because it is limited by diffraction and can resolve details no smaller than the wavelength of light. The new lens deploying negative refraction is different: it is perfect. The image *is* the object because be have annihilated the space separating object and image.

The realisation in 2000 that a negatively refracting lens breaks the diffraction limit caused a furore, so entrenched was the notion that lenses cannot do better that the wavelength of light. Several experiments have subsequently shown my theoretical prediction to be correct. However a major technical challenge remains: the lens is

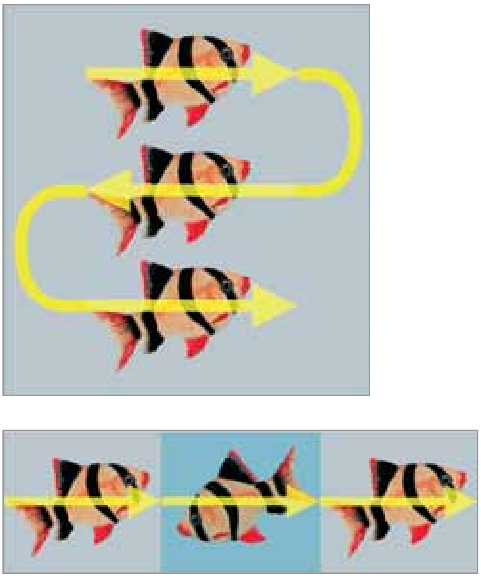


Figure 11. Top: a fish outside a slab of negatively refracting material is focused twice more: once inside the negative medium and once again on the far side. If we equate negative refraction with negative space, we retrieve the picture below in which light passes the same point three times, coming to a perfect focus on each occasion.

extremely intolerant of imperfections and is only perfect if it is made precisely according to prescription.

Sub-wavelength focussing was first realized by the Eleftheriades group at GHz frequencies[10], where the necessary metamaterials are available; however an optical version of the lens is much more difficult. Although metals such as silver have an intrinsic permittivity that is negative, magnetic responses at optical frequencies are rare and as yet we do not have a fully 3D negatively refracting optical material — at least not one that is capable of sub-wavelength operation.

An approximation to the perfect lens can be obtained if all the dimensions are much less that the wavelength. Under these conditions the electric and magnetic components of the field are almost independent and we are free to concentrate on the electrical part, ignoring magnetic effects. A purely electrical field is indifferent to μ and therefore it should be possible to construct a 'poor man's lens' of silver by tuning the frequency so that $\varepsilon \approx -1$. Losses in the silver prevent exact realization of this condition; nevertheless, a version of this poor man's lens was built by the Xiang group in Berkeley[11]. They demonstrated sub-wavelength imaging on a scale of a few tens of nanometres. More recently, silver has been replaced by silicon carbide where the low loss Restrahlen bands give a much better approximation to $\varepsilon \approx -1$, and hence better resolution as a fraction of the wavelength. A resolution of $\lambda/20$ was achieved in the silicon carbide experiment shown in Figure 12[12].

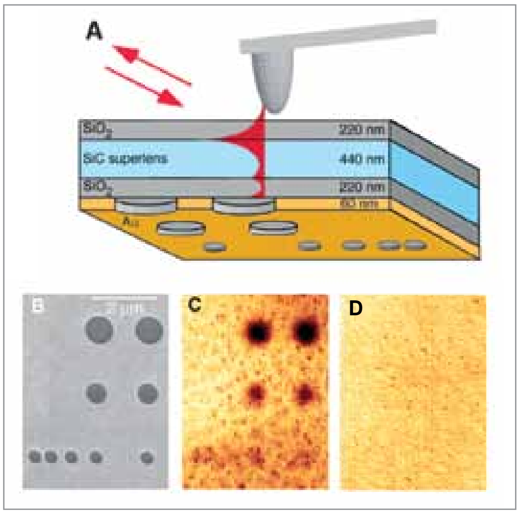


Figure 12. (A) A silicon carbide superlens is sandwiched between two layers of dielectric. Light incident on a scanning probe tip scatters and is focused on the gold layer below where it reflects from details of the structure inscribed in the gold. The lens refocuses this scattered light on the tip where it scatters a third time before reaching a detector. (B) SEM image of the gold layer. (C) Images at a frequency where $\varepsilon \approx -1$ showing a resolution as good as $\lambda/20$. (D) The lens is detuned to a different frequency violating the conditions for perfect lensing, and no images are seen.

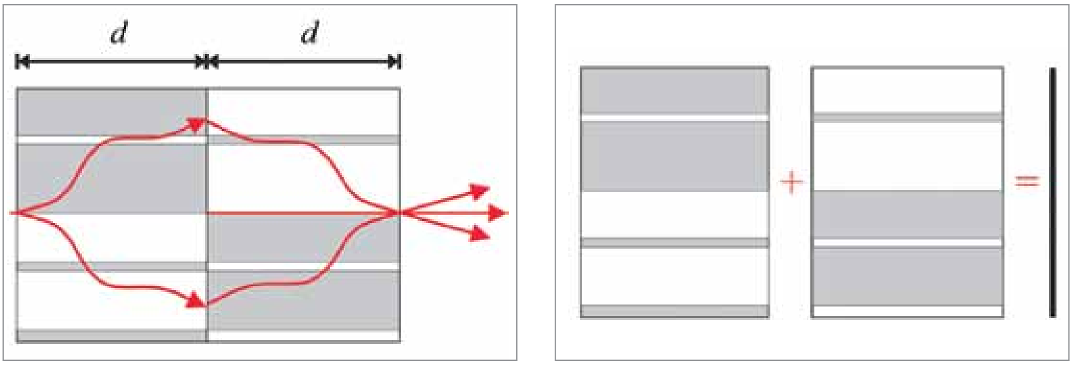


Figure 13a. Left: An alternative pair of complementary media, each annihilating the effect of the other. Light does not necessarily follow a straight line path in each medium, but the overall effect is as if a section of space thickness 2d were removed from the experiment. Figure 13 b Right: A graphical expression of our new theorem: complementary halves sum to zero. The optical properties of the rest of the system can be calculated by cutting out the media and closing the gap.

OPTICAL ANTIMATTER

We have discussed how the original Veselago lens is a far more remarkable structure than first thought. In fact the more we examine it the less like a lens it seems. For one thing it does not have a focal length; for another, it can focus only objects that are a finite distance from the lens. Gradually the realisation has dawned that something much more interesting is happening, and that an alternate way of understanding properties of a negative slab is as a piece of negative space. Optically speaking, a slab of material with $\varepsilon = -1$, $\mu = -1$ appears to annihilate the effect of an equal thickness of vacuum.

Having realised this, it is only one further step to prove a more general result: two slabs of material of equal thickness placed adjacent to one another optically annihilate if one is the negative mirror image of the other, the mirror being taken to lie on the interface between the two slabs. A simple instance of this is shown schematically in Figure 13.

Although this result is quite plausible where rays follow a simple distorted trajectory in each medium as illustrated above, in some instances the theorem has startling consequences. Consider Figure 14, a system drawn to my attention by David Smith: the mirror theorem applies, but a ray construction contradicts the theorem. Applying the laws of refraction to ray 2 implies that the ray is rejected by the system instead of being transmitted through to the other side — and a dark shadow behind the cylinders is predicted by the ray picture.

In fact, a full solution of Maxwell's equations shows that ray 2 is transmitted and emerges through the system just like ray 1, and no shadow is formed. The apparent paradox is resolved by recognising that a series of resonances form on the surfaces of the cylinders that enable radiation to tunnel across the gap between the cylinders. A clue to the nature of these resonances is given by the closed loop of dotted rays in the centre of the figure, which indicates the presence of a state that traps radiation — that is, a resonance.

This elaboration of the perfect lens greatly increases the richness of the subject as a far greater variety of lens geometries can now be considered.

One important restriction of the new lenses as currently formulated is that they produce images of exactly the same size as the objects. To do otherwise we must introduce curved surfaces, and this is most easily done through coordinate transformations. For example, if we transform between

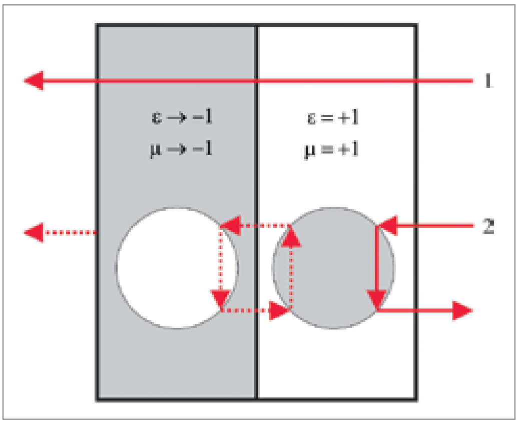


Figure 14. The left and right media in this 2D system are negative mirror images and therefore optically annihilate one another. However a ray construction appears to contradict this result. Nevertheless, the theorem is correct, and the ray construction is erroneous. Note the closed loop of rays indicating the presence of resonances.

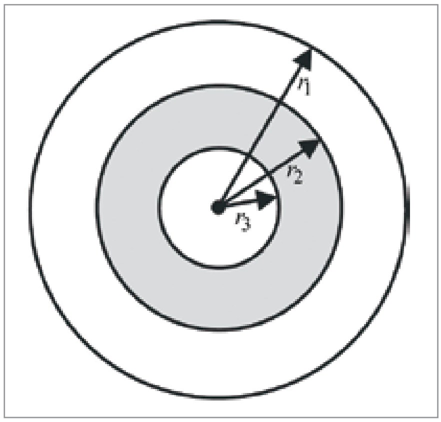


Figure 15. It is possible to design a spherical annulus of negative material lying between r_2 and r_3 that acts like a magnifying glass. To the outside world the contents of the sphere of radius r_3 appear to fill the larger sphere radius r_1 with proportionate magnification.

Cartesian coordinates (x,y,z) and spherical coordinates (l,θ,ϕ),

$$x = r_0 e^{l/l_0} \sin\theta \cos\phi$$
$$y = r_0 e^{l/l_0} \sin\theta \sin\phi$$
$$z = r_0 e^{l/l_0} \cos\theta$$

the original geometry of a slab of negative material is transformed to an annulus of material as in Figure 15. Furthermore a coordinate transformation leaves the form of Maxwell's equations unaltered, changing only the values of ε , μ and **E**, **H** (the electric and magnetic field vectors) appearing in the equations, leaving the basic topology of the image unaltered.

Detailed discussion of these powerful techniques is beyond the scope of this review, but the conclusions can be stated simply. For example a spherical structure, sketched in Figure 15, will behave like a magnifying glass when constituted as follows,

$$\varepsilon_x = \varepsilon_y = \varepsilon_z = +\frac{r_2^2}{r_3^2}, \quad 0 < r < r_3$$
$$\varepsilon_x = \varepsilon_y = \varepsilon_z \to -\frac{r_2^2}{r^2}, \quad r_3 < r < r_2$$
$$\varepsilon_x = \varepsilon_y = \varepsilon_z = +1, \quad r_2 < r < \infty$$
$$\mu_x = \varepsilon_x, \quad \mu_y = \varepsilon_y, \quad \mu_z = \varepsilon_z$$

In other words outside radius r_2 the structure is empty space, between r_2 and r_3 lies the negatively refracting material (though now more structured than before), and finally inside the annulus is a material of constant high permittivity and high permeability. This structure has the unusual property that, when viewed from beyond a radius of r_1, where

$$r_1 = r_2^2 / r_3 \ ,$$

the contents of the inner sphere (any electrical or magnetic sources of radiation) appear to be expanded to fill a sphere radius r_1 filled with $\varepsilon = \mu = 1$ material — that is, magnified by a factor r_2^2 / r_3^2. The region of space between r_1 and r_3 has vanished and is not visible.

Viewed from inside radius r_3, the world outside radius r_1 appears to be shrunken by a factor r_3^2/r_2^2 and now reaches to radius r_3, and appears to have $\varepsilon = \mu = r_2^2/r_3^2$ — the same as the filling of the internal sphere. Again the region of space between r_1 and r_3 has vanished and is not visible.

In the limit that the required values of e and m are perfectly realised this picture is exact; hence we have arrived at a prescription for a magnifying glass with perfect resolution.

As always the material constraints on performance will be critical but, in principle, it is now possible to think of designing such structures. Think of your favourite coordinate transformation, apply it to the original perfect lens and get a new lens for free!

CONCLUSIONS

Unusually for the physical sciences, negative refraction began life as a theoretical concept rather than an experimental discovery. It posed instead a challenge to experiment: to find materials with negative values of ε and μ. At the same time the theory had to be defended in debates on the validity of the concept. I think it is fair to say that 2003 saw this initial phase draw to a close with positive conclusions both on the concept and its experimental realisation. The future holds many new oportunities.

Theorists, now confident of the foundations, can move forward and exploit the concepts. New discoveries can be expected to follow.

Experimentally the goal is to improve design of the metamaterials, chiefly by reducing loss, but also by moving on from the cottage industry phase to viable volume manufacture of these materials. To exploit opportunities in all frequency ranges requires more work. Currently the microwave region of the spectrum is the most fertile for experiment, but THz frequencies beckon, and the optical region beyond is currently being probed by nanoscale near-field experiments.

Further in the future we can expect new devices. Novel waveguides and aerials operating in the GHz range will be first in the field, but good progress is being made at magnetic resonance imaging frequencies. Ultimately the concepts will integrate with the substantial effort being invested in plasmonic phenomena at optical frequencies, where the aim is to produce devices structured on a sub-wavelength scale.

Footnotes

1 Veselago, V.G., 1968, Soviet Physics USPEKHI, **10**, 509.

2 Pendry, J.B., Holden, A.J., Robbins, D.J., and Stewart, W.J., 1999, *IEEE transactions on microwave theory and techniques*, **47**, 2075.

3 McCall, M.W., Lakhtakia, A., and Weiglhofer, W.S., 2002, 'The negative index of refraction demystified,' *European Journal of Physics*, **23**, 353.

4 Smith, D. R., Padilla, W.J., Vier, D. C., Nemat-Nasser, S. C., and Schultz, S., 2000, *Phys. Rev. Lett.*, **84**, 4184.

5 Shelby, R.Ath, D. R., and Schultz, S., 2001, *Science* **292**, 77.

6 Pendry, J.B., 2003, *Nature* 'News and Views', **423**, 22.

7 Parazzoli, C.G., Greegor, R.B., Li, K., Koltenbah, B.E.C., and Tanielian, M., 2003, *Phys. Rev Lett.*, **90**, 107401-1.

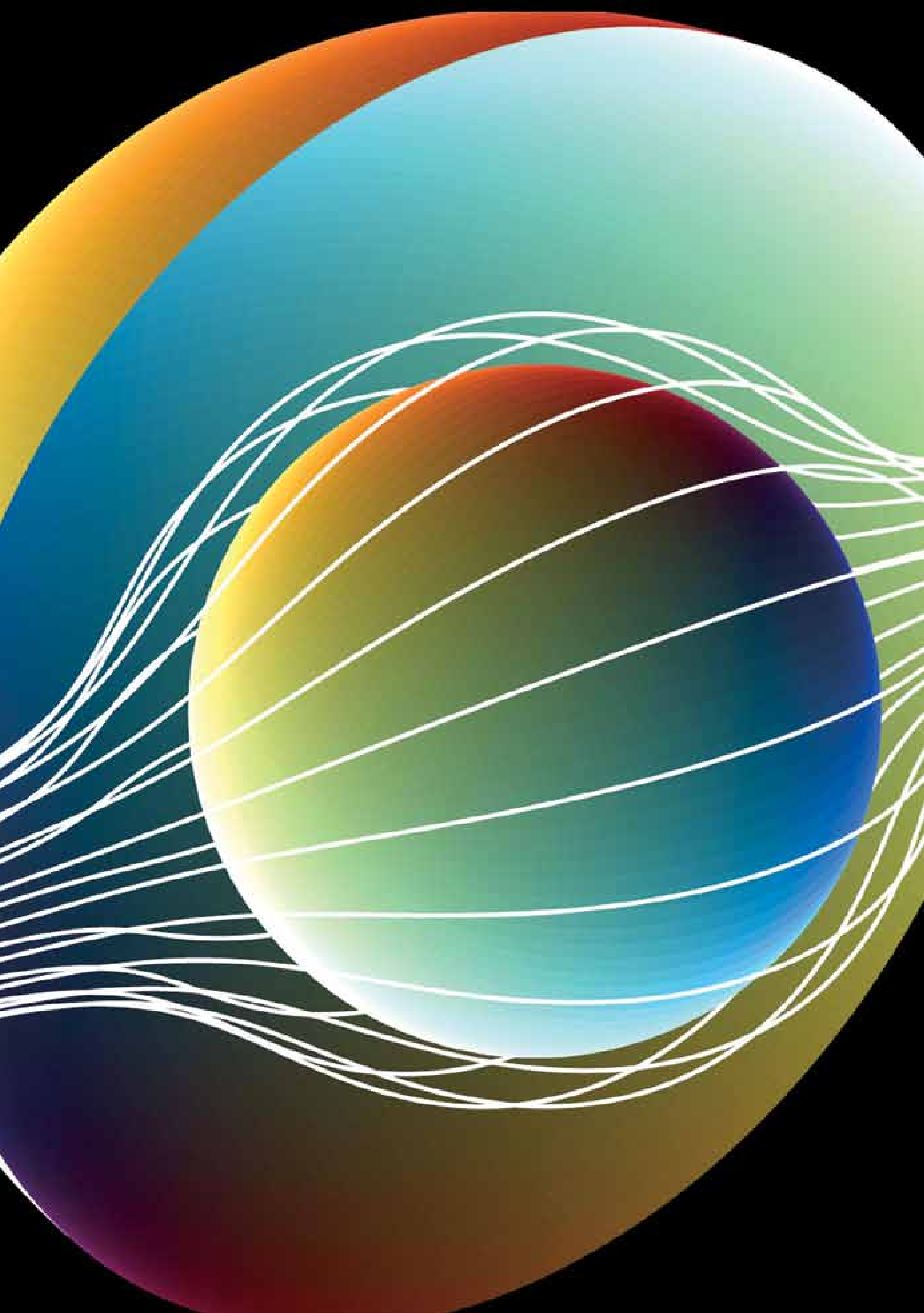
8 Pendry, J.B., Holden, A.J., Stewart, W.J., and Youngs, I., 1996, *Phys. Rev. Lett.*, **76**, 4773.

9 Pendry, J.B., 2000, *Phys. Rev. Lett.*, **85**, 3966.

10 Grbic, A., and Eleftheriades, George V., 2003, *Appl. Phys. Lett.*, **82**, 1815.

11 Fang, N., Liu, Z., Yen, T.J., and Zhang, X., 2003, *Optics Express*, **11**, 682.

12 Korobkin, D., Urzhumov, Y., and Shvets, G., 2006, *JOSA B*, **23**, 468.



PAINTBRUSHES, CANNIBAL CRICKETS AND HUMAN OBESITY

PROFESSOR STEPHEN SIMPSON

I am going to take you on a strange journey, and we are going to begin in the midst of a desert locust swarm in North Africa. This animal is one of a dozen or so species of grasshopper that are called locust, some of which live in Australia. They are defined by having one peculiar feature, which distinguishes them from all other grasshoppers, and that is that they are essentially two animals packed inside the same genotype.

To illustrate this let us take the two juvenile desert locusts shown in Figure 1. The one on the left has been reared on her own, and grown up to be camouflaged in colouration: she is a beautiful green colour, which allows her to blend in with the vegetation where she lives and to avoid the attention of predators. She would live a solitary existence, being shy in her behaviour and tending to avoid — indeed, to be repelled by — other locusts.

In contrast, her sister, the animal on the right, was reared in a crowd; she is brightly coloured and would be a very active animal, being attracted by other locusts and tending to aggregate. These aggregations would, at some point, start to march through the habitat, and when they become adults, fly as huge swarms that migrate hundreds of kilometres and cause devastation to agriculture in vast areas of Africa and Asia.

The process of change from one form to the other is initiated in response to a change in population density, and begins to occur really quite quickly. Behaviour is the first feature to change. If you took the animal on the left and you put her in a crowd, within an hour she would begin to be attracted rather than repelled by other locusts, and within an hour or two, she could be part of a marching band. This transition is known as 'density dependent behavioural phase change' and my research group has spent the last fifteen years trying to understand it.

One of the initial questions we had to answer was: what is it about being in a crowd that causes the change? What stimuli are provided by other locusts in a crowd that might trigger the transition? On reflection, it could be the sight of other locusts, their smell, being contacted by them, or perhaps their sound.

As it turned out, touch is critical. What happens is that the environment brings together solitary locusts, against their predisposition to avoid each other, at limited resource sites. The congregated locusts jostle with each other, and it is the jostling which causes the change from repulsion to attraction. But being touched just anywhere won't work, which is where the paintbrushes in this chapter's title come in.

TICKLING LOCUSTS

We spent many happy hours tickling locusts on various body parts with paintbrushes and demonstrated that, unless you stimulate

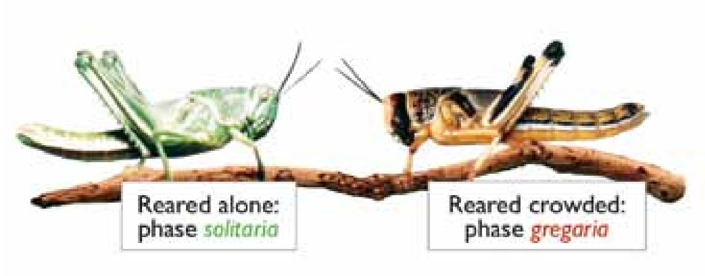


Figure 1. The two forms of the desert locust. The insect on the left was reared alone, while her sister in the right was reared in a crowd – and might contribute to a swarm.

hairs on the hind legs of the animal, you don't cause the change from solitary to gregarious behaviour.

This was rather an important discovery because it gave us access to the underlying neurochemistry of the process; it allowed us to delve into the nervous system, to seek the controlling neural pathways and to focus in on the molecular changes that were going on within the animal during the process of phase transition. An important finding was that the ubiquitous brain chemical serotonin was critical in triggering the change from solitary to gregarious behaviour. Tickling the hind leg of a solitary locust, or electrically stimulating the nerve pathways running from touch-sensitive hairs on the leg, leads to the release of a puff of serotonin in the central nervous system, which modulates neural circuitry to produce the change in behaviour.

ON THE MARCH

Let us return to a recently aggregated group of locusts, which at this point are simply milling around. They are very active, they are attracted to each other but are not going anywhere in particular — until suddenly, as if of one mind, the entire group becomes highly aligned and starts to march.

As it marches through the habitat, the mass of locusts takes on the appearance of a river flowing through the environment. Of course, up close it is not a fluid, but is made up of particles — and those particles are locusts, which gave us a clue as to how we might go about studying the collective behaviour of locusts.

We turned to our statistical physics colleagues, who have developed a set of mathematical 'self-propelled particles models' in which they take individual particles, program them to behave in rather simple ways with respect to their neighbours, and then simulate what happens in clouds or swarms of such particles.

Using these models we found that the collective decision to start marching emerges within the crowd from simple local interactions between locusts; there is no leader locust, there is no hierarchical control, it just happens because the locusts are following a very simple rule, and that rule is, 'align with your moving neighbours'. Once a critical density of locusts is reached, just adding one or two more to a local area will suddenly cause the transition to collective, aligned movement: the march has begun. We were able to explore the phenomenon in the laboratory by causing locusts to march endlessly round and round an arena shaped like a Mexican hat. We could measure the interactions between individual locusts in the crowd and, by adding more and more locusts into the arena, demonstrate really very good accordance between what the animals did

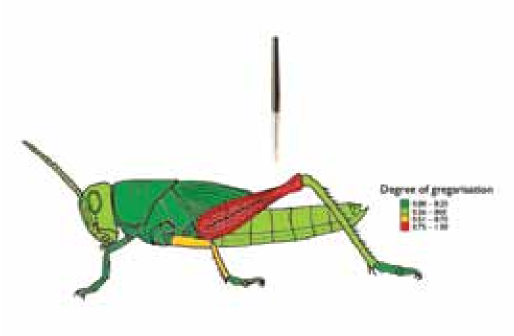


Figure 2. Stroking the hind legs, but no other body region, with a paintbrush causes a solitarious phase locust to change phase and start behaving gregariously.

Figure 3a. A fearsome Mormon cricket from Utah.

Figure 3b. Mormon crickets on the march.

and what the self-propelled particles model was predicting.

MARCH ... OR BE EATEN

This raises the next question, why on earth do they align? What causes alignment? Why do they do it?

The answer to this question came from a different, but similar animal called the Mormon cricket (Figure 3a), and it turned out that the answer was actually really rather sinister.

The Mormon cricket is a rather large flightless cricket that lives in Northwest America, and forms vast marching swarms (Figure 3b). We encountered one vast swarm as it was crossing a road: a steady stream of crickets marching across the road for five days!

Figure 4. Mormon crickets form huge marching bands, because if they don't keep moving they get cannibalised.

We found and demonstrated that the Mormon crickets are on a forced march to find protein. Individuals have to keep moving because they need protein, and that particular environment where these swarms occur, is deficient in protein. So if a cricket doesn't keep moving it won't find more protein — but worse than that, if it stops moving it becomes another cricket's protein meal! Were you to be incapacitated in the face of these animals I think you would become a protein meal as well.

It happens that the Mormon cricket has a specific appetite for protein, and if you satiate that appetite the animal stops cannibalising and stops marching.

PROTEIN OR CARBOHYDRATE?

A protein appetite is not actually restricted to Mormon crickets; it is found in many, many animals — the more animals we look at, the more we find. We discovered that the same is true for locusts. Were you to give a locust the opportunity to mix its own diet, you would find that over a period of time, it would ingest a very precise mix of protein and carbohydrate. If it could, the locust would defend that point when you try to perturb that environment.

But lets say you do not allow the locust to get to that precisely balanced diet, by feeding it with diets that contain a higher percent of carbohydrate and lower percent of protein, or a higher percent of protein and a lower percent of carbohydrate. Then you find that the animals prioritise regulation of protein over that of carbohydrate, such that on the high carbohydrate diets they over-consume carbohydrate to edge closer towards their target level of protein — and they do so to a greater degree than they will over-consume protein on high percent protein diets.

Now there are costs to doing either of those things, to not being at your target: if you eat too much carbohydrate as a locust you become obese and you die early. Granted it is pretty hard to tell that a locust is obese because it has an exoskeleton, but it is chubby on the inside. It becomes a little bit like an overweight knight, wedged into his suit of armour.

On higher percent protein diets than optimal you find that they end up very lean: they lose fat mass and become susceptible to starvation, which is ecologically very relevant.

WE'RE NOT SO DIFFERENT

So with locusts, protein intake is prioritised if there is a change in the diet that forces a trade off between the regulatory systems for protein and carbohydrate. And again, if you look at other animals you find a very similar system: we found it in rats, we found it in fish and birds — and we found it in humans, too.

We made an experiment that involved taking a group of ten people — undergraduate

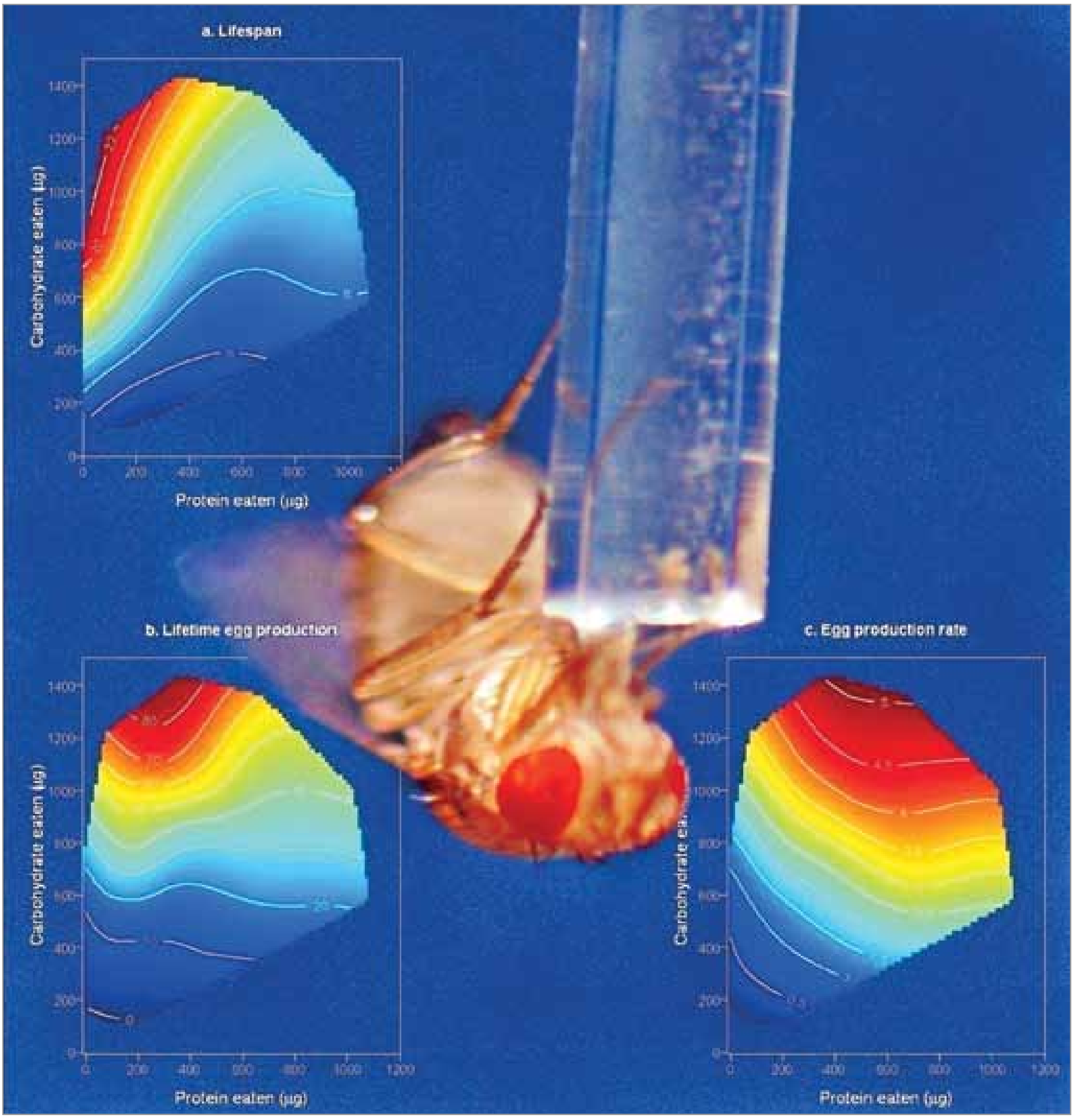


Figure 5. Effects of dietary composition on ageing and reproduction in fruit-flies were measured by confining 1,000 individual flies to one of 28 diets and measuring how much they ingested and how many eggs they laid throughout their lives.

students — up to a chalet in the Swiss Alps, and providing them with a buffet of food items over a particular period of time. And we manipulated their diet. When we confined the students to high carbohydrate and fat items they tended to over-consume carbohydrate and fat, and suffered a very small deficit of protein. But when we put them on a high protein diet, they were unwilling to over-consume protein and hence ingested substantially less non-protein energy.

Going back through the research literature over the past forty years, we found the same pattern. This is summarised, as we saw with the locust, by the statement that protein intake is prioritised when the diet forces a trade off between it and non-protein energy. And when we re-analysed nutrition survey data from Australia in 1995 we found very similar patterns.

OVERWEIGHT? YOU'RE OVER-COMPENSATING

What are the implications on human health for this set of regulatory responses that we and other animals demonstrate? Well, there seems to be a shift in our diets towards more high fat and high carbohydrate items — this could be because they are more accessible and more affordable, or we have forgotten how to cook, we eat out, there are abundant reasons documented for why this might happen. The net result is that human physiological responses will lead to the over consumption of fat and carbohydrate to gain limited protein. And of course unless you get rid of that pattern in your diet, body composition will change.

Just to give you an idea for the power of this response, were your diet to drop from 14% protein energy to 12.5% — which sounds a trivial reduction, and is what has occurred in the United States over the past forty years — then you would have to consume 14% more non-protein energy to get the same absolute quantity of protein. That is, a small drop in protein leads to a much larger increase in non-protein to compensate. To make things worse, the effect is greatly exacerbated if the requirement for non-protein energy is diminished as a result of falling levels of exercise.

In contrast, if you shift your diet towards a higher percent of protein, then your body will not let you over consume protein. The net result is that there is an under-consumption of non-protein energy, leading to a negative energy balance and the potential to lose weight. So protein has the power to drive obesity, it would seem, and also to ameliorate it.

WHY PROTEIN?

Why do humans and other animals possess such a strong response to protein regulation? It is easy to explain why eating too little protein is to be avoided: protein is the only source of dietary nitrogen for growth.

But why are we and many other animals so unwilling to eat excess protein? The answer came from a study that looked at the effect of nutrient intake on longevity and fecundity in fruit flies (one of the main model systems for studying ageing), as well as two other insect models.

The data from these studies showed that eating too much protein caused an early death, which explains why regulatory responses have evolved limiting its excessive intake — in flies at least. If the female fly was offered the opportunity to select her own diet, she behaved like a nutrient-seeking missile and tracked a diet that maximised lifetime egg production (Figure 5).

The idea that eating fewer calories makes organisms live longer (up to a point of course) is widespread and is one of the big ideas in biomedical research on ageing. Our results showed that this is not true for flies, but a critical analysis of existing data indicated that the caloric restriction dogma may not apply to mammals either.

A major study on mice is now underway in our laboratory, to test whether calories or nutrient balance is responsible for ageing and longevity. This study is the most extensive dietary analysis of ageing and its metabolic and molecular correlates ever undertaken. There even appear to be fundamental links between dietary macronutrient balance and immunity, cancer risk and eating disorders such as anorexia. The mouse study is allowing these ideas to be tested; we have also begun to extend these findings on ageing in mammals to study a group of elderly men, and to assess and seek novel targets for drug discovery.

DIET AND DISEASE

The link between diet and immunity is especially important, with implications for resistance to disease among human populations. We have shown in insects that there is a fundamental link between nutrient balance and disease resistance in insects — whether the same applies to mammals is an important question. We are now moving into this research area, with experiments on mice and fruit flies to quantify the link between nutrition, immunity and the community of microorganisms that live in the gut and play a critical role in the host's health and well-being.

Our research indicates that humans and many other animals are ill equipped to regulate nutrient and energy intake when they experience a rapid shift in their nutritional environment — such as the modern westernised diet — in part as a consequence of the powerful protein appetite. But will future generations evolve ways of dealing with excess ingested energy?

CAN WE EVOLVE RESISTANCE TO OBESITY?

We documented the evolution of resistance to obesity in animals with an experiment in which we restricted caterpillars to high-energy regimes over successive generations. Initially, eating excess carbohydrate to gain limiting protein led to obesity, but after eight generations the insects had evolved to rid themselves of excess carbohydrate without laying it down as body fat: they had become immune to obesity in an obesogenic world. This demonstration opens the way to discovering genes for resistance to obesity and has implications for explaining how organisms such as pest insects respond to environmental change.

We have also combined nutrition and collective behaviour by studying nutrient balancing in ant societies and soil dwelling slime moulds, which play a critical role in decomposition and nutrient cycling. These extraordinary experiments have implications that extend from solving multiple supply chain problems to understanding carbon sequestration in soils. We are currently extending the study of nutritional interactions beyond individuals and groups to model how these interactions fashion species assemblages, food webs and the functioning of ecosystems.

We have gone on a strange journey, from locusts to cannibals, to human obesity, to ageing, to a new paradigm for the study of nutritional ecology — a journey that provides a compelling example of the power of pure scientific discovery.

Further reading

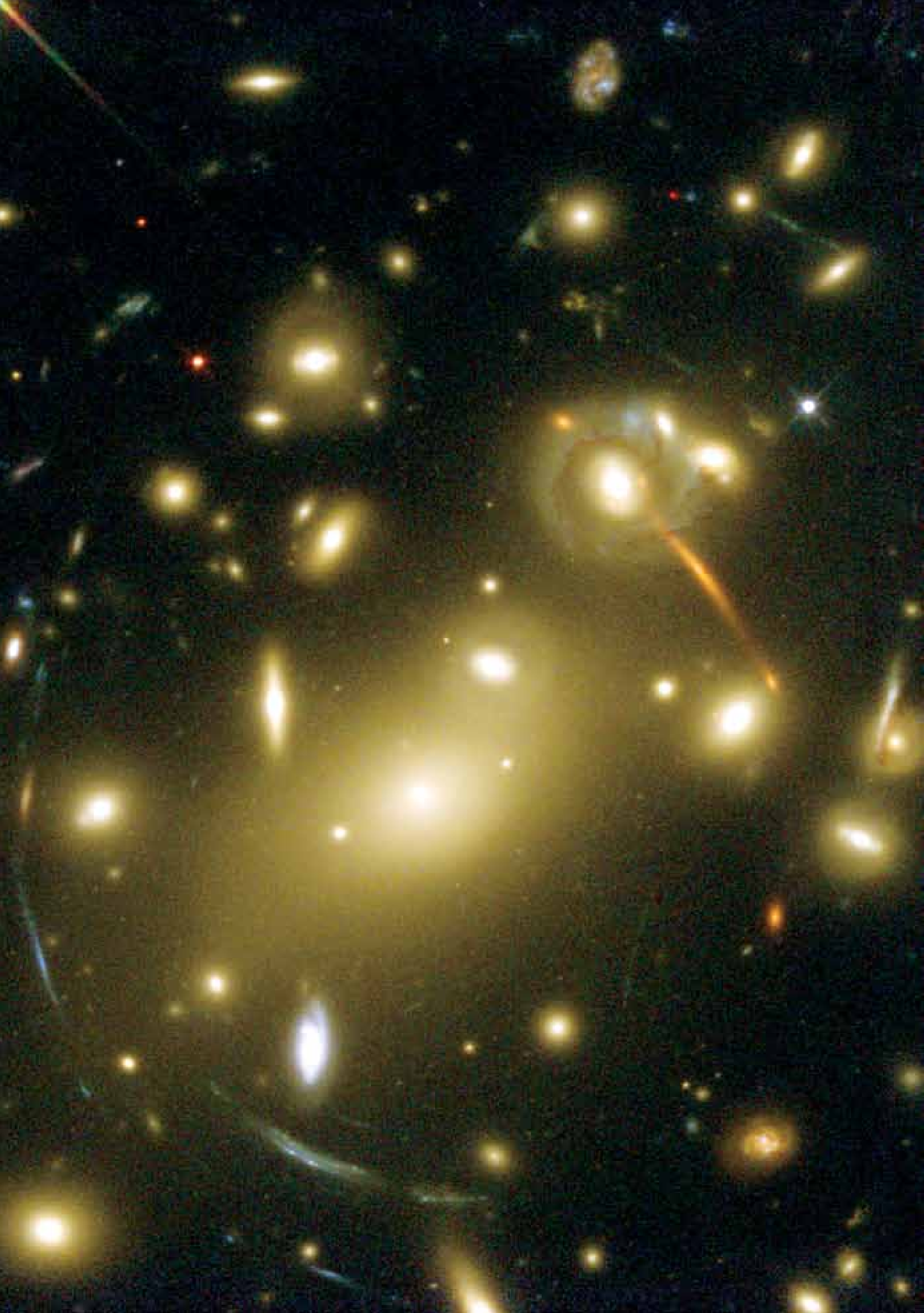
Anstey, ML, Rogers, SM, Ott SR, Burrows, M, and Simpon, SJ 2009. Serotonin mediates behavioral gregarization underlying swarm formation in desert locusts. *Science* 323, 627-629.

Lee, KP, Simpson, SJ, Clissold, FJ, Brooks, R, Ballard, JWO, Taylor, PW, Soran, N, and Raubenheimer, D 2008. Lifespan and reproduction in Drosophila: New insights from nutritional geometry. *PNAS* 105, 2498-2503.

Simpson, SJ and Sword GA 2008. Locusts. *Current Biology* 18: R364-R366

Simpson, SJ and Raubenheimer, D 2005. Obesity: the protein leverage hypothesis. *Obesity Reviews* 6: 133-142.

Simpson, SJ, Sword, GA, Lorch, PD, and Couzin, ID 2006. Cannibal crickets on a forced march for protein and salt. *PNAS* 103: 4152-4156.



DARK SECRETS: DARK MATTER, DARK ENERGY AND DARK SKIES

PROFESSOR FRED WATSON

OVERVIEW

Perhaps the single most embarrassing aspect of modern astronomy is that we don't know what most of the Universe is made of. The great American amateur astronomer, John Dobson, once described the Universe as 'mostly hydrogen and ignorance', but he was wrong. In fact it's mostly dark matter and dark energy — and ignorance. The hydrogen only accounts for about four percent of the contents of the Universe.

So what are dark matter and dark energy? That simple question encompasses two of the most puzzling problems in contemporary astrophysics. The mystery is deepened by the fact that between them, these two invisible components of our cosmic environment make up an overwhelming 95 percent of the mass-energy budget of the Universe. You can see why astronomers are embarrassed …

The quest to discover the nature of dark matter and dark energy presents challenges that demand all the ingenuity of modern astronomers and physicists. Moreover, it requires a formidable array of scientific hardware, much of it representing a major investment in astronomical technology. A lot of that instrumentation is still in the planning stage, but it is already clear that large ground-based telescopes designed to collect visible light and infrared radiation (so-called optical-infrared telescopes) will continue to play a major role.

The mission to understand nature at its most fundamental level also places strict limits on sky quality at observatory sites, since the measurement of exceedingly faint objects is required. Thus, the darkest secrets of astronomers also include their quest for dark skies.

This chapter gives a broad brush overview of some of the strategies that are being used to explore dark matter and dark energy. It also highlights one of the consequences of those strategies, namely the ongoing requirement for extremely low levels of background skyglow at the world's observatories.

INTO THE DARKNESS — A HISTORY OF ASTRONOMY'S DARKEST SECRETS

How do we know dark matter is present in the Universe? It reveals itself only by one thing — the effect of its gravitational attraction on matter that we can see. Other than that, we have no way of detecting it. We can, however, sense this gravitational 'smoking gun' by a number of methods.

While many scientists imagine that the investigation of dark matter is a contemporary topic with a fairly short history, it was as long ago as 1933 that the US-Swiss astronomer Fritz Zwicky (1898–1974) first noticed that something didn't add up. He was investigating clusters of galaxies — the largest concentrations of matter in the Universe. Most people know that galaxies are aggregations

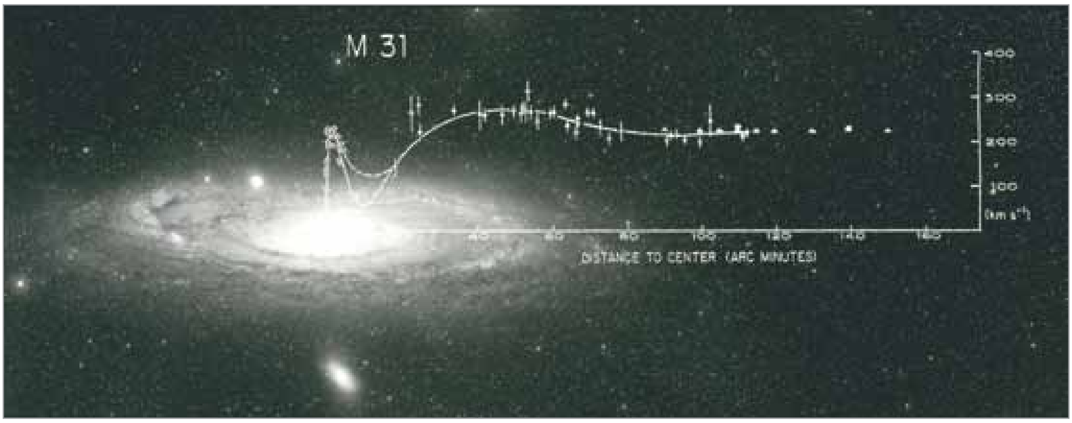


Figure 1: Vera Rubin's rotation curve for the nearby spiral galaxy M31 (the Andromeda Galaxy). It is superimposed on an image of the galaxy itself, and shows that the rotation remains roughly constant (at some 250 km/second) well beyond the visible limits of the galaxy. Only a massive dark halo can explain this. (Source unknown — try Carnegie Institution, Washington, DC for hi-res version and permission.)

of stars — hundreds of billions at a time — and that they often occur in the shape of a flattened disc with spectacular spiral arms. Such spiral galaxies also contain large amounts of gas (mostly hydrogen) and dust embedded in their discs. Other types of galaxies occur too, however, including the large football-shaped objects we call elliptical galaxies.

Zwicky was particularly interested in a very rich cluster of galaxies in the constellation of Coma Berenices (Berenice's Hair) in the northern hemisphere sky. Using a well-established astronomical technique that enables velocities along the line of sight to be accurately measured (by means of the Doppler effect), he determined the speeds of several members of the cluster. He was amazed to find that these galaxies seemed to be moving too fast for the cluster to hold onto them. Using a mathematical device known as the virial theorem, he established that the gravitational attraction of the visible matter in the cluster — the stars, gas and dust in the constituent galaxies — was insufficient to bind the cluster together. Given their velocities, the galaxies he was observing should have escaped from the cluster long ago. Zwicky inferred from this that something else was present, an invisible component that neither emitted light nor absorbed it from the radiation of background objects. Whatever it was, it exerted a gravitational pull on the members of the cluster. He was right on the money with that deduction, but little attention was paid to it at the time.

Forty-five years later, the US astronomer Vera Rubin made accurate measurements of the rotational velocities of galaxies, and discovered new evidence for the existence of dark matter. We know that individual galaxies rotate, and by selecting spiral galaxies that we see almost edge-on, can measure the way the rotational velocity changes with distance from the centre of the galaxy. Such velocity profiles are known as rotation curves.

If the matter in a given galaxy is concentrated where the light of its stars is concentrated, one would expect the rotation curve to show stars close to the centre of the galaxy moving rapidly, while those further away move more slowly. This was not what Vera Rubin's measurements showed, however. She discovered that far from falling off with distance from the galaxy's centre, the rotation curve stayed almost constant out to its extremities (see Figure 1). Rubin's observations were made with clouds of gas rather than stars, but the bottom line was the same. The rotation curves could only be explained if the galaxies were enveloped in giant halos of so-called dark matter. Thus was born the modern era of dark matter studies, and this has culminated today in new sky surveys designed to explore the properties of dark matter in fine detail.

Dark energy has a similarly long pedigree, although the observational evidence only began to emerge more recently. It was in Einstein's adoption of his so-called cosmological constant (introduced belatedly into his equations of general relativity early in 1917) that the idea first appeared. General relativity is a theory of the way gravity acts by distorting space, and it remains the best theory of gravity we have. It has been tested many times, and our observations fit the theory with an accuracy that is now better than one part in a hundred thousand. The only area in which relativity breaks down is on the scale of the very small, as it is incompatible with quantum theory.

In 1917, however, Einstein thought his new theory was in big trouble. If his equations were applied to the Universe as a whole, they became unstable, producing a Universe that would have to expand or contract. To the best of Einstein's knowledge, though, the Universe was static. So he did something very clever. He introduced a mathematical entity that he called the cosmological constant (usually given the symbol lambda, Λ). This was supposed to represent an inbuilt repulsive (or attractive) force in the fabric of space that would counteract any tendency for it to expand or contract. When the real expansion of the Universe was discovered twelve years later, however, Einstein quickly withdrew the idea in embarrassment, famously describing it as his 'biggest blunder'.

Except for a very few cosmologists, most people then simply assumed that the cosmological constant was zero, and that space had no inbuilt force field. But in 1998, two separate groups of scientists (including one led by Professor Brian Schmidt of the Australian National University) produced hard evidence that far from slowing down as expected, the Universe is expanding more rapidly today than it was seven or eight billion years ago. This is attributed to an inherent springiness of space — or dark energy — that is overcoming the tendency of the Universe to decelerate because of the mutual gravitational attraction of everything within it. Its action may be the same as Einstein's cosmological constant, but there might also be subtle differences. As we shall see, a number of different hypotheses could give rise to a Universe whose expansion accelerates and, in the best scientific tradition, it is important to devise tests that might differentiate between them.

Exploring the properties of dark energy to try to understand exactly what causes it ranks with the dark matter question as the most pressing problem facing today's astrophysicists. They, in turn, are contemplating ambitious strategies to enable the research, proposing experiments that demand instrumentation at the cutting edge of astronomical technology.

HOW CAN WE PROBE DARK MATTER?

It is principally through large scale surveys of the three-dimensional positions of galaxies in 'local' space (out to two billion light years or more) that we know as much as we do about the way dark matter behaves. Other, quite different techniques are also revealing more details of this elusive component of the Universe, however, and we will look at them in turn.

GALAXY SURVEYS

In order to establish the three-dimensional position of a galaxy, you need to know its distance, and that is best estimated from a property known as redshift. As its name implies, redshift is the extent to which the light emitted from a galaxy has been shifted towards the red end of the spectrum — not, as most people think, by the galaxy's motion through space (the Doppler effect), but by the expansion of space itself since the light began its journey to Earth. Thus, the light waves have been stretched to a longer wavelength as the Universe has expanded.

Measurement of galaxy redshifts demands the use of optical instruments known as spectrographs on large telescopes. The spectrograph splits the light from the galaxy into its rainbow colours, revealing the 'bar-code' of features arising from elements in the atmospheres of the galaxy's constituent stars. The amount by which that bar-code is shifted to the red is essentially a date-stamp

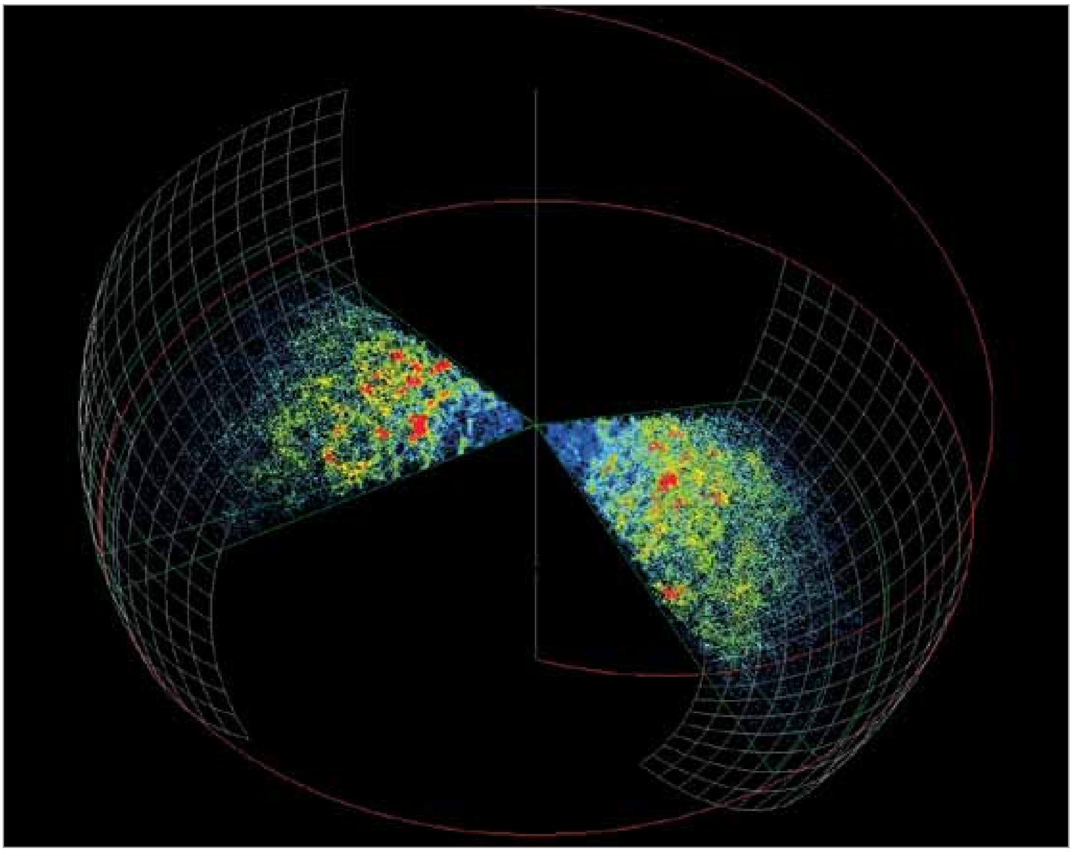


Figure 2: The 2dF Galaxy Redshift Survey (see text). In this remarkable slice through the local Universe, each dot represents a galaxy, and the colour coding represents the local density of galaxies at that point (red is the highest density, blue is lowest). The two halves of the diagram correspond to the views on either side of the disc of our own Milky Way Galaxy. The gradual petering-out of galaxies at the edge of the diagram (at a distance of some 2.5 billion light years from our Galaxy) is not real, but represents the sensitivity limit of the survey. (Courtesy Australian Astronomical Observatory.)

on the light—it tells us unequivocally what the size of the Universe was when it was emitted. The mathematics of the situation is rather easy, in fact. If z is the redshift of the galaxy (the fractional change in wavelength of its light) and R is the ratio of the size of the Universe when the light was emitted to the size it is today, then $R = 1/(1+z)$. Thus, the light from a galaxy with a redshift of 1 was emitted when the Universe was half its present size. Once R is known, then the distance to the galaxy can be determined.

If you want to get statistically-significant information on the large-scale structure of the Universe, however, you need not just one or two redshifts, but thousands. The Australian Astronomical Observatory (AAO) 2dF Galaxy Redshift Survey (2dFGRS) pioneered such studies in the late 1990s with a survey of 221,000 galaxies. The AAO was formerly the Anglo-Australian Observatory until its incorporation into the Australian Government's Department of Innovation, Industry, Science and Research (DIISR) in July 2010, while the reference to 2dF — 2 degree field — comes from the name of the instrument that was used to make the survey. It was a spectrograph on the 3.9-metre Anglo-Australian Telescope (AAT) with a field of view on the sky of 2 degrees, able to observe up to 400 galaxies at a time.

The 2dFGRS showed first and foremost that the distribution of galaxies in the Universe is far from uniform (see Figure 2). Galaxies

Figure 3: The dome of the Anglo-Australian Observatory's 1.2 m UK Schmidt Telescope at Siding Spring Observatory, home of the RAVE stellar radial velocity survey. The town of Coonabarabran is visible in the distance behind the telescope. (Author.)

seem to concentrate along the edges of great voids in space with a characteristic size of around 300 million light years. It's almost as if the Universe resembles a foam of galaxies, and indeed such a frothy structure is exactly what is predicted by our theories of the origins of galaxies.

However, much more can be learned about the state of today's Universe from large-scale surveys like the 2dFGRS. Matter of all kinds — visible and dark — affects the geometry of space (because of the distortion of space predicted by general relativity), and such large scale positional surveys allow the distortion to be investigated. That can then be combined with our knowledge of the distribution of visible matter in the Universe. This tells us, for example, that the ratio of dark matter to visible matter in the Universe is about four to one, and that dark matter and visible matter concentrate together. 'Beacons of light on hills of dark matter' is one eloquent description of clusters of galaxies.

A number of new galaxy surveys are exploring the geometrical distortion of space in great detail. One that involves observations of some 375,000 galaxies with a wide range of different telescopes (including the AAT) is GAMA, or 'Galaxy and Mass Assembly'. Its aim is to produce the most comprehensive database to date of galaxies and their properties as they have evolved over the past third of the age of the Universe. The dark-matter halos of galaxies are part of GAMA's stock-in-trade.

GALACTIC ARCHAEOLOGY

Knowing that dark matter concentrates wherever visible matter is found suggests another approach to its investigation. Our own Milky Way Galaxy, in which the Solar System is embedded, is a giant spiral system of 400 billion stars with associated gas and dust — and dark matter. So dark matter is all around us. How is it distributed? Does it occur in clumps, and if so, how big are they? And what might we learn from their size?

The fact that we are surrounded by stars whose histories reflect the history of our Galaxy as a whole has suggested that much can be learned about our Galaxy's past by studying them. This technique is sometimes known as 'near-field cosmology' to distinguish it from the more usual method of probing the history of the Universe, in which objects at great distances (and hence long look-back times) are studied. What is investigated in these near-field studies includes the velocities of the stars along the line of sight (their so-called radial velocities, this time due to the Doppler effect) and the chemical make-up of the stars, usually known as their 'metallicities', which are a reflection of the stars' ages. Other physical parameters are also determined, such as stellar temperatures and surface gravity.

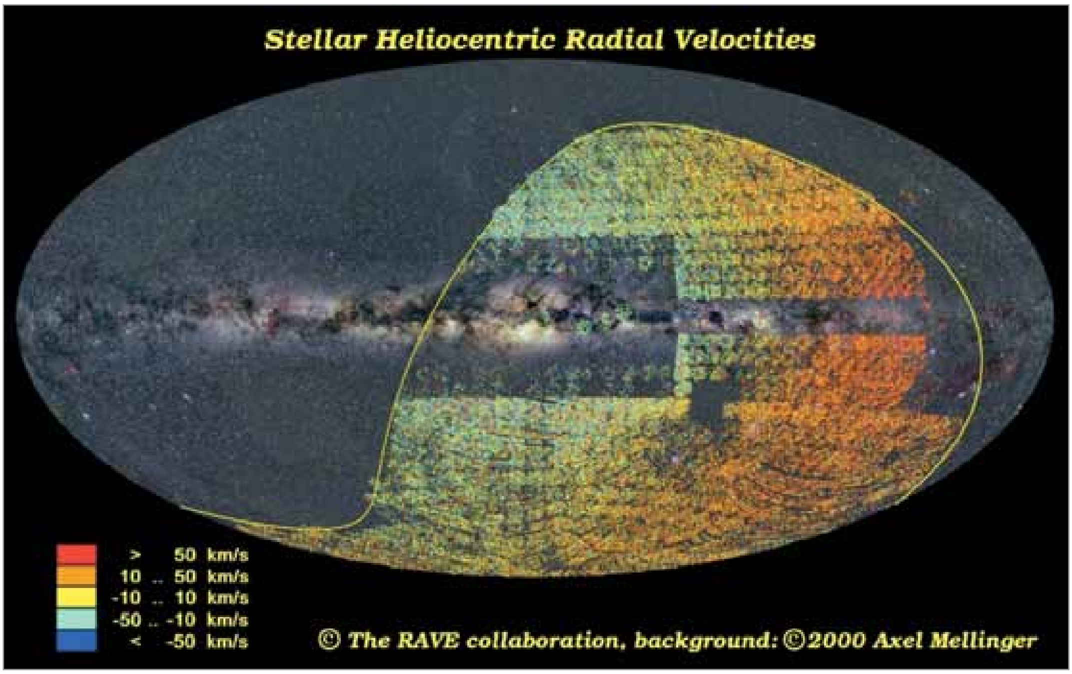


Figure 4: Progress chart for the first four years of the RAVE survey, showing the velocities of some 150,000 stars. The map represents the whole sky, with the Milky Way (the plane of our Galaxy) running across the centre. Each of the coloured panels represents one pointing of the UK Schmidt Telescope, whose six-degree field of view encompasses several hundred stars. The stars are colour coded in velocity, and the gradual change in overall tint is because of the Sun's own motion with respect to the average motion of the stars (the so-called reflex solar motion). The curved yellow line represents the celestial equator, only southern hemisphere objects being included in the survey. (Courtesy RAVE.)

These data are now being collected in great numbers by several large-scale spectroscopic surveys of stars, analogous to the 2dFGRS mentioned above. The most ambitious of them is the RAVE survey of a million stars being carried out on the AAO's 1.2 m United Kingdom Schmidt Telescope (Figure 3). (RAVE is an acronym for RAdial Velocity Experiment.) The survey began with pilot observations in 2003, but became the Schmidt Telescope's full-time task two years later. So far, almost half a million stars have been measured (see Figure 4).

One possible spin-off from the RAVE survey is a more detailed look at how dark matter is structured. Stars move under the influence of gravity, and by sampling large populations of stars, the underlying gravity field can be mapped, and the dark matter component identified. A likely outcome of this is an estimate of the minimum size of a clump of dark matter. Such estimates have already been made from more limited star velocity surveys, and suggest it may be around 1,000 light years across. If that is the case, then we can deduce that its temperature is a few degrees above absolute zero, rather than the few tenths of a degree that is usually assumed, presenting a problem for the theoretical astrophysicists who try to model dark matter. Currently, such values remain sufficiently inexact that they do not yet present serious difficulty. As the RAVE survey evolves, however, we will see better measurements being made, and it is possible that the temperature of dark matter will eventually be one of the high-confidence outcomes.

In the future, this technique of large-scale velocity mapping will be extended to large samples of individual stars in nearby galaxies using the coming generation of 20–30 metre class 'extremely large telescopes' (ELTs), described later in the chapter.

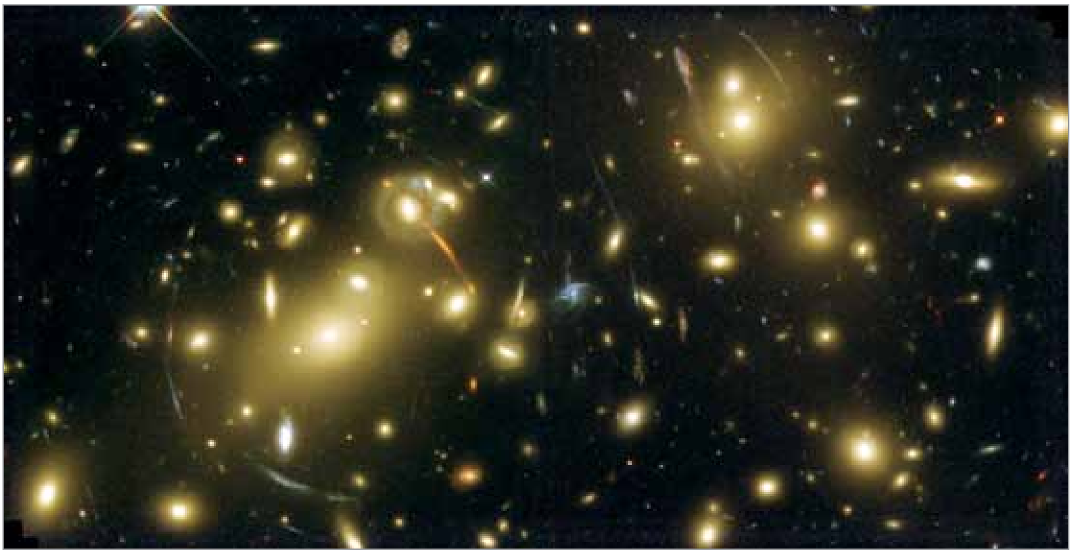


Figure 5: The rich galaxy cluster Abell 2218, showing its effect as a gravitational lens on a much more distant cluster of galaxies behind it. The distant galaxies (mostly bluish in colour) are distorted into arcs of light by the combined gravitational field of the nearby cluster galaxies (orange-yellow). In effect, the nearby cluster is acting like a telescope of monumental proportions. (NASA, A. Fruchter and the ERO team, STScI, ST-ECF.)

GRAVITATIONAL LENSING

A more difficult, but potentially more powerful method of probing dark matter relies on the distortion of space caused by the gravity of galaxy clusters, and its effect on the light from background objects. This is the so-called gravitational lensing technique. It has been known since the earliest days of general relativity that the distortion of space around a gravitating object — planet, star, galaxy or cluster of galaxies — has the effect of focusing light rays from more distant objects behind it.

The effect of a symmetric point mass in producing a gravitational lens is analogous to the way a glass lens acts on light rays — hence the name. Gravity produces an unusual form of lens, however, and its equivalent manufactured in glass would have a similar shape to the base of a wineglass. That means the focusing effect is nowhere near as crisp as a real lens, and usually results in arc-like images of the distant objects (see Figure 5)

Because a gravitational lens results from the action of both visible and dark matter, it can be used to investigate the distribution of both types of matter in the lensing object (a galaxy or a cluster of galaxies). The technique involves a statistical reconstruction of the distortion of space around the foreground galaxy or cluster, arrived at by making assumptions about the distribution and appearance of the background galaxies. The result is a detailed map of the geometry of the distorted space (Figure 6). From such a map, the actual distribution of matter in the cluster can be deduced, and it is found that the bright galaxies are themselves embedded in large volumes of dark matter, exactly as suggested by other techniques.

Because of the extreme faintness of the background objects, this work was the province of the Hubble Space Telescope HST) during the 1990s but today's generation of 8–10-metre class telescopes are capable not only of providing images of the faint lensed objects, but also obtaining spectra to establish their distances. In the future, it will be extended to very great distances (at which almost all objects are lensed by intervening matter) using the new ELTs. Meanwhile the HST continues to provide new insights into the distribution of dark matter over large distances (Figure 7).

Such observations as these involving a wide range of look-back times suggest the

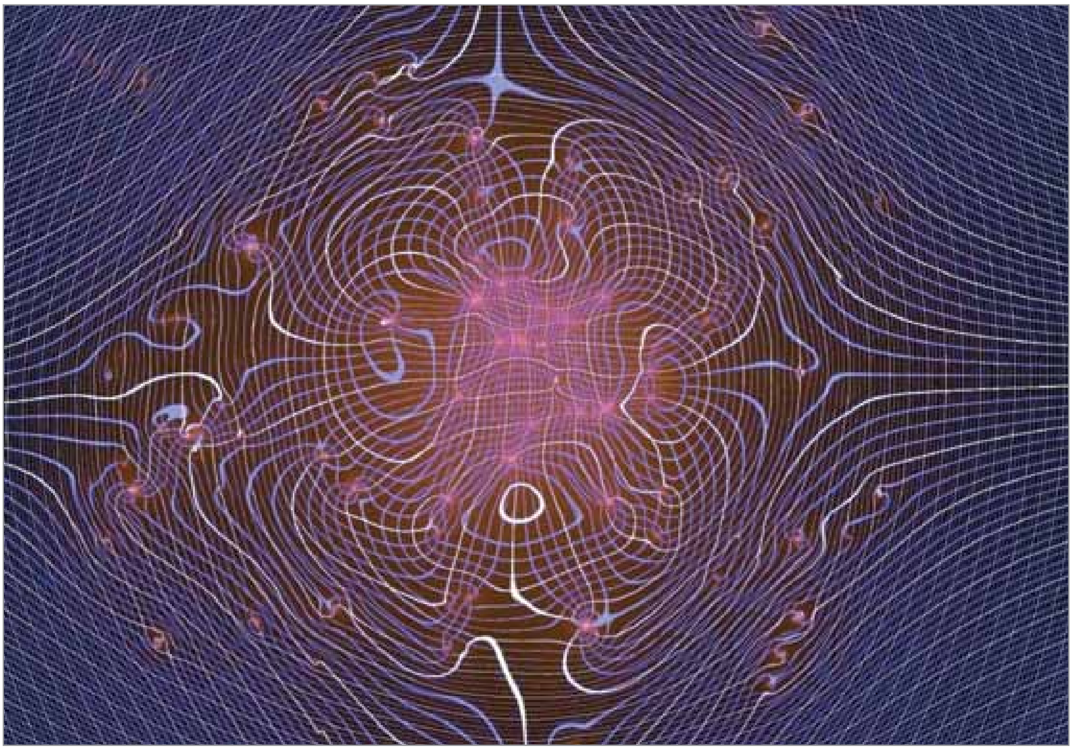


Figure 6: Simulation of the effect of a rich cluster of galaxies (shown pink) on the geometry of the surrounding space, as deduced by the appearance of lensed background objects. The distortion results from the gravitational effect of both visible and dark matter, and provides a powerful method of charting the distribution of dark matter. (Source: Large Synoptic Survey Telescope.)

possibility of testing our understanding of the vital role played by dark matter in the evolution of today's galaxies. The assumption is that dark matter clumped together in the early Universe, and its enhanced gravitational field then attracted concentrations of hydrogen from the Big Bang, which eventually collapsed into stars and galaxies. A consequence of this is that dark matter at greater look-back times should be less clumped than it is in the more recent past, and there are already indications that this is, in fact, the case (see Figure 7).

By combining all these various studies of dark matter, astrophysicists hope to learn enough about its behaviour for a clear leader to emerge from the various competing models of what constitutes dark matter that have been built by theoretical physicists studying a range of sub-atomic particles. Alongside these theoretical studies, the experiments being carried out at facilities like CERN's Large Hadron Collider will continue to narrow down the range of possibilities for dark matter. The best bet at the moment seems to be one of a number of hypothetical particles that essentially do not interact with normal matter, but do have appreciable mass, such as axions or neutralinos. Either way, it seems that the first announcement of the exact nature of dark matter will come not from a telescope, but from a particle collider.

TAPPING INTO DARK ENERGY

As recently as the mid 1990s, we believed that the expansion of the Universe — a consequence of the Big Bang some 13.7 billion years ago — was gradually slowing down as a result of the mutual gravitational attraction of everything within it. An entirely reasonable assumption, you'd think, but a few heretics in the astronomical community had some doubts about this model. They wondered whether a kind of dark energy, or 'springiness' of space, might be starting

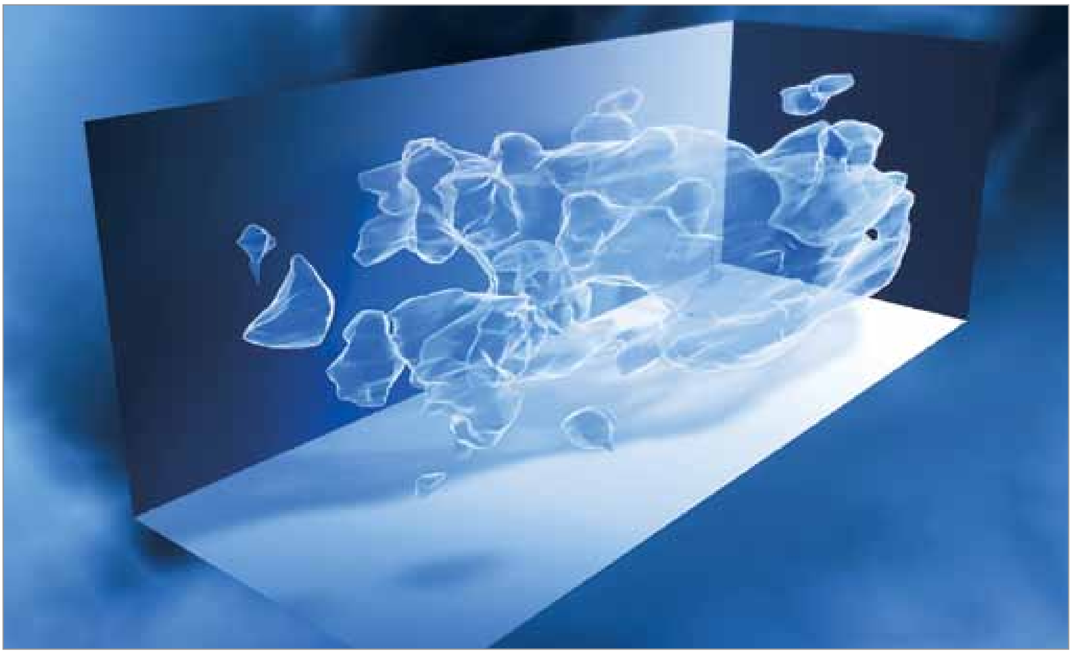


Figure 7: Schematic illustration of recent dark matter observations made with the Hubble Space Telescope using the gravitational lensing technique. The nearest galaxies measured (those closest to the depiction of the telescope) have a look-back time of about 3.5 billion years, while the look-back time of the most distant ones is around 6.5 billion years. There is some evidence that typical clumps of dark matter have become more fragmented in more recent times, suggesting that evolutionary processes are in action. (Source: NASA/ESA/Massey.)

to counter the deceleration, pushing the Universe into an era of accelerated expansion. It was a rather esoteric point until, quite suddenly in 1998, two separate groups of scientists produced hard evidence that the Universe is, indeed, expanding more rapidly today than it was seven or eight billion years ago.

POSSIBLE SOURCES OF DARK ENERGY

The evidence for an accelerated expansion came in the form of observations of a particular kind of supernova — the so-called Type Ia — at very great distances. Supernovae of this type are caused by old stars exploding violently as a result of matter being deposited onto them from a nearby companion star, and provide extremely bright standard candles, easily outshining their host galaxies. What caused all the excitement was that these remote supernovae were dimmer than they ought to be, given their estimated distances — and hence look-back times — from our Galaxy (Figure 8). This suggested that, yes, the expansion of the Universe is accelerating, a result that has now been confirmed by a number of different methods. Moreover, we now know that the dark energy driving the acceleration is the largest single component of the Universe, amounting to some 75 percent of its total mass/energy density. So what is it?

When consideration was given to this question, an intriguing possibility emerged. Could the dark energy be something to do with that long-neglected orphan of general relativity, the cosmological constant, described earlier in this chapter? If it had the form of a constant negative pressure, and Λ had a value small enough that it only began to overcome gravity when the characteristic distances separating galaxies had become very great, then it might just fit the bill.

There are other theoretical possibilities, too, but they require the introduction of various flavours of 'new physics' — phenomena that are not predicted by relativity, such as quantum gravity and string theory. They can be summarised as:

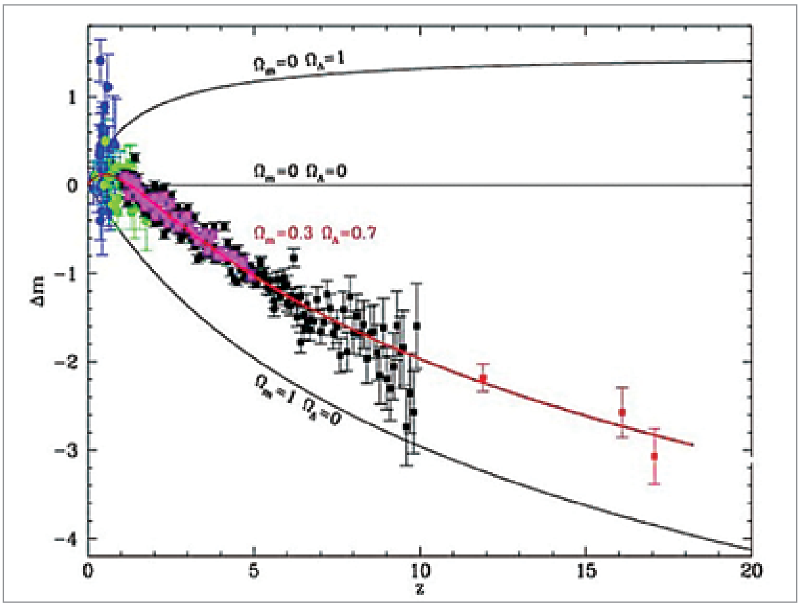


Figure 8: A simulated Hubble diagram, which shows the relationship between the brightness of an object and its redshift (which is analogous to distance, or look-back time). This diagram is plotted for supernovae of Types Ia, II and Population III (pink, black and red simulated data points), but includes the real observations of Type Ia supernovae (blue and green data points) that were used to establish the existence of dark energy (see text). The diagram is normalised to an empty Universe; i.e. it is plotted so that the points would lie along the horizontal line at Δm=0 if the Universe contained no mass or energy. The red track represents a Universe composed of 30 percent matter and 70 percent dark energy. (Source: Hook et al., The Science Case for the European Extremely Large Telescope. Permission OK.)

1. A new fundamental force — the whimsically-named quintessence, which echoes the fifth element of ancient Greek philosophy. Like the effect of the cosmological constant, this would have to be a dark energy with negative pressure, but a key difference is that quintessence evolves with time, leaving its imprint on the Universe only in the relatively recent past.

2. A modification of gravity, that makes it act differently on small scales from the way it acts on large scales. If correct, this could eliminate the need for an underlying negative pressure.

3. Abandoning the cosmological principle, which says that the Universe is the same in all directions. That would permit a Universe that has significant differences between one place and another, perhaps again eliminating the need for an overall repulsive pressure. This possibility seems to have been ruled out by recent work at Johns Hopkins University in Baltimore, which has examined the idea of a local 'void' in the Universe giving the illusion of accelerated expansion. Such a hypothetical void cannot be consistent with the most recent measurements of the expansion velocity.

No-one really knows what causes dark energy. The best guesses involve a seething foam of virtual particles at the quantum level, popping in and out of existence and imbuing the fabric of space with negative pressure. Theories on this bizarre area of quantum physics are many and various — and far from complete. Embarrassingly, even the best of them predict a repulsive force that is 120 orders of magnitude (yes, 10^{120}) bigger than what we observe. Such intense repulsion would tend to rip everything apart, and would certainly have prevented atoms from forming in the early Universe. This estimate is famously considered by many to be the worst prediction in the whole of science.

TYING DOWN THE EVIDENCE

A good start on understanding the issue might be to identify which model of dark energy best fits the astronomical observations. Cosmological constant or quintessence — or

Figure 9: The proposed Giant Magellan Telescope will use seven 8.4-metre diameter mirrors to synthesise a single mirror some 22 metres in diameter. It is being proposed by an international consortium, which includes the Australian National University. (Source: Todd Mason, GMT Project.)

something else? The problem is, though, that the required observations are hard to do, and are only just starting to come in. A number of groups throughout the world are now actively engaged on the process of obtaining them, typically by extending the supernova observations to greater distances and greater numbers of objects. But the best chance of really tying down the intimate details of dark energy comes from more subtle — and more difficult — observations. The fact is that the nature of dark energy has a significant influence on the large-scale geometry of the Universe, just as dark matter does. Therefore, if that geometry can be probed at different stages in the Universe's history, there is a real chance that the correct model of dark energy can be identified, and its characteristics precisely determined.

We already have an accurate knowledge of this large-scale geometry at two key times in cosmic history. One is in the very early Universe. Detailed observations of the cosmic microwave background radiation (the flash of the Big Bang, seen at a look-back time of almost 13.7 billion years) made with the WMAP and Planck spacecraft have revealed what the Universe was like in its infancy. And, as we have already seen, large-scale surveys of the three-dimensional distribution of galaxies in *today's* Universe have been made with the 3.9-m Anglo-Australian Telescope (the 2dF Galaxy Redshift Survey) and other telescopes.

Both these snapshots of the Universe reveal structure that is of great interest to cosmologists. In the cosmic microwave background radiation (CMBR), there are very slight (one part in 10^5) point-to-point temperature variations that originated in acoustic oscillations in the primordial fireball — the 'bang' of the Big Bang, if you like. Likewise, as we saw in Figure 2, the distribution of galaxies in today's Universe is spidery, resembling a honeycomb, or foam of galaxies. It turns out that this structure is exactly what you would expect to see if you could 'fast forward' the CMBR to the present time.

Comparison of the Universe's characteristics at these two periods has already allowed a wealth of information on its evolution to be deduced. The missing ingredient has been a similar 3-d survey of galaxies at great enough distances that they correspond to a look-back time of about half the age of the Universe — a time seven or eight billion years or so ago, when dark energy first began to make its presence felt. But to do the job properly required a survey of hundreds of thousands of faint galaxies, a hugely ambitious programme.

That has now been at least partially provided by a galaxy survey whimsically named WiggleZ, which has again been carried out on the AAT. The 'Z' in WiggleZ refers to redshift, of course, but the 'wiggle' part is all about the acoustic oscillations and the structure they have imprinted as the Universe has evolved. Completed in January 2011, WiggleZ has measured the redshifts of some 200,000 galaxies, mapping the cosmic structure across look-back times up to eight billion years.

Further down the track, studies such as this could be carried out with even greater precision if bigger telescopes were used, allowing fainter galaxies to be observed. That will be the province of the new generation of 20 to 30 metre telescopes mentioned earlier — the so-called ELTs (Extremely Large Telescopes). These represent the latest step in the evolution of the telescope, and are no more than a decade away — at least in terms of their construction. There are several projects under consideration, including E-ELT (the 40-metre European ELT), TMT (the US Thirty Metre Telescope project), and the international GMT (Giant Magellan Telescope), in which Australia is already a partner (Figure 9). The last of these is of particular interest to studies of dark energy, as it will have a sufficiently wide field of view to make it a formidable survey telescope rather than simply an instrument that studies single objects in great detail.

So, what are the results coming from studies such as WiggleZ and distant supernova measurements? Both these techniques show results on the nature of dark energy that have a decided preference for... wait for it... yes, the cosmological constant. Supernovae studies indicate that the negative pressure of dark energy seems to have changed by less than 20 percent since the Universe was about half its current size, a finding consistent with dark energy being due to the most conservative of the current models, Einstein's Λ. Likewise, the independent WiggleZ measurements are best fit by a model with a constant negative pressure. These are truly remarkable findings, not least because they show that even when Einstein thought he was blundering, he was actually right. What a guy.

STAYING IN THE DARK

A common theme in all these endeavours is the observation of extremely faint objects at visible or near-infrared wavelengths with very large telescopes. The major limitation of this technique is that the radiation from the target objects is usually much less than one percent of the natural sky background. Thus, very careful sky-subtraction is required in order to maximise the signal-to-noise ratio in the data (a vitally important yardstick in estimating the quality of astronomical observations). Background sky is, in fact, the dominant source of noise in such faint work, and for this reason, any artificial sky glow added to the natural background is fatal.

AS BLACK AS THE NIGHT?

The one thing that surprises most visitors to an observatory site is not how dark the night sky is, but how much light actually comes from it. When there is no moon in the sky (which is for approximately two weeks each month), the absence of light pollution renders the stars dazzling. A closer look, however, reveals that it's not just the stars that are bright. The sky itself is luminous, a phenomenon made more obvious by the presence of clouds, which appear jet black

Figure 10: The Earth at night. This montage of images made taken from the International Space Station clearly shows the effect of upward light-spill from the world's centres of population. Prominent features include the Nile Valley, the network of roads extending through the former USSR (all leading to Moscow), and the high levels of light pollution in North America, Europe, the Indian sub-continent and south-east Asia. (NASA.)

against their background. They contrast strikingly with their counterparts in a brightly-illuminated city sky.

What causes this natural luminosity of the sky? The brightest component is due to continuous auroral emissions in the upper atmosphere at particular wavelengths. Oxygen emits strongly at 557.7 nm, and there is an intense series of hydroxyl and other emission features redward of 630.0 nm. In addition to these discrete emissions, however, there is a continuous-spectrum background whose sources include illuminated dust in the solar system, faint stars within our own galaxy and, at the very faintest level, the light of distant galaxies that are indistinguishable from one another.

Any celestial object that an astronomer wishes to investigate is superimposed on this sky background, and usually the background completely dominates the observation. Often, as noted above, astronomers are looking for faint objects whose brightness is less than *one percent* of the natural sky background. Little wonder that any increase in the background glow due to artificial light simply renders these objects invisible. In that respect, there are few natural environments as sensitive to pollution as the night sky.

Modern electronic detectors such as TV-type charge-coupled devices (CCDs) allow astronomers to subtract the background from their observations. However, this process introduces further unwanted noise into the signal, and it remains true that the smaller the background level the more likely the astronomer is to succeed in his task. To be effective in observing such unbelievably faint objects as galaxies at distances in excess of six billion light-years (as required for dark energy studies, for example), telescopes need as dark a sky as possible.

KEEPING THE SKY DARK

The first requirement when you build an observatory is to find a site that is free from artificial light. There are many other considerations, but that first hurdle was what directed the overall strategy of astronomers in the 1960s and 70s when they identified the world's best places for building large telescopes. Sites in the USA (on the mainland as well as in Hawaii), Chile, South Africa, the Canary Islands and Australia all fulfilled these conditions, and an estimated $US3.5 billion has been invested over the years in observatories on these sites. The challenge is to keep them dark, given the ubiquity of urbanisation of our planet (Figure 10).

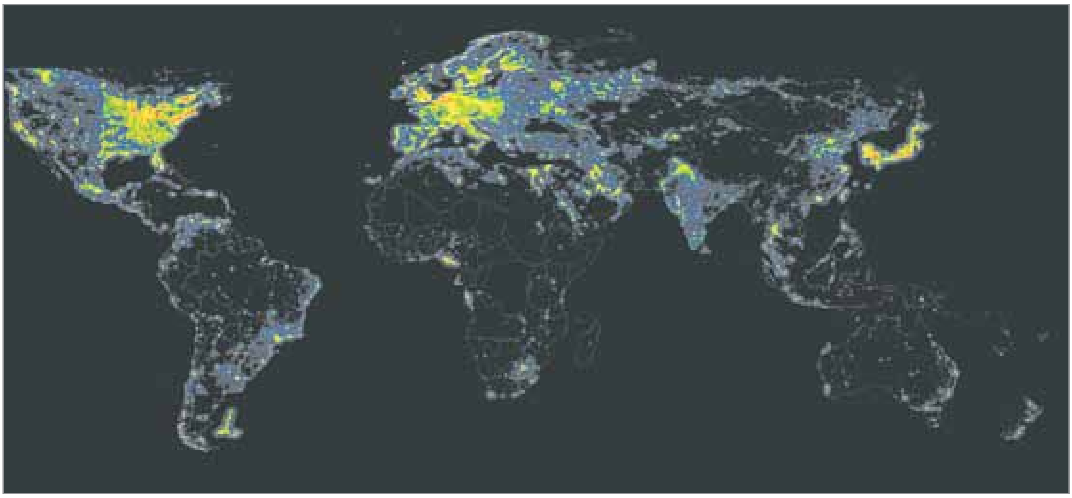


Figure 11: The world's light pollution, modelled from satellite images like Figure 10, combined with the light propagation characteristics of the atmosphere. The shading on the map represents the brightness of the night sky as seen from that point. (Source: Pierantonio Cinzano, University of Padua.)

It was the late Roy Garstang together with David Crawford in the USA who did fundamental work on the spread of urban sky glow in the 1980s, and that work has recently been taken further by the studies of Pierantonio Cinzano at the University of Padua in Italy (see Figure 11). Cinzano has mapped the world's light pollution using night-time images from space (to reveal the upward-pointing light sources) combined with the known propagation characteristics of the atmosphere by molecular and aerosol scattering. The process is highly refined, taking into account the curvature of the Earth, and shows that two-thirds of the world's population live in light-polluted conditions. More dramatically, about one-fifth of the population can no longer see the Milky Way.

Cinzano's studies also show that light pollution is encroaching on many of the world's major observatory sites. This leads to the conclusion that remoteness is not enough to protect an observatory, and steps have to be taken to reduce the spread of light pollution. The steps can be legislative (and the environment of most of the world's leading observatories is so regulated) or educational — to bring home the message that light-pollution is bad news not only for sky-watchers, but for anyone who is distressed by wastage of energy.

This message is now strongly reinforced by the need to minimise the carbon footprint of lighting installations, and improving outdoor lighting is a relatively straightforward way of doing that. But recent studies have also told us much more about the ill-effects of light pollution on humans and animal species. Human circadian rhythms are now known to be very sensitive to disruption by strong lighting, a finding that has significant implications for the health of shift-workers, for example. And among the detrimental effect of light pollution on the animal kingdom are the disruption of nocturnal species and the disturbance of migration routes.

AUSTRALIA'S DARK SKIES...

The small town of Coonabarabran in north-western NSW hosts Australia's national optical observatory. At Siding Spring Mountain, some 20 km from the town, the 3.9 m Anglo-Australian Telescope (AAT) provides Australian optical astronomers with their largest observing facility on home soil. It is likely to remain the largest telescope in Australia for the indefinite future. Such is the sensitivity of this instrument that in moonless conditions it is capable of recording a light-level equivalent to a 100 watt lamp-bulb at the distance of the Moon (384,000 km).

The AAT is accompanied by the Australian Astronomical Observatory's other telescope, the 1.2 metre UK Schmidt Telescope (Figure 3). As we have seen, both these instruments play a crucial role in studies of dark matter and dark energy. They have now been joined by the Australian National University's new 1.3 metre Skymapper telescope, which has been especially built to look for distant supernova explosions in the quest to identify the source of dark energy.

The Siding Spring Observatory site is protected by the Orana Regional Environmental Plan No.1 (REP), which originally covered an area of radius 100 km centred on the AAT. This New South Wales state legislation was enacted in 1990, but is currently being revised. The purpose of the revision is to bring the legislation up to date (particularly with regard to the important role of full cut-off fixtures — light fittings that permit no light to escape above the horizontal plane). It is also desirable that it be made easier to apply, and more understandable. In addition, the area of coverage is being increased to a radius of 200 km (the 'Orana Dark Skies Region') to include major population centres such as Tamworth. That city — the 'City of Light' — anticipated the provisions of the REP by adopting a very effective local lighting strategy in 2002, becoming the first regional centre in Australia to do so.

Beyond the 200 km radius, there is no overarching legislation to protect the skies of Siding Spring Observatory — even though the light of the Sydney–Newcastle metropolitan area (at a distance of about 350 km) can clearly be seen from the telescopes (as predicted by the studies shown in Figure 11).

The astronomers of Siding Spring Observatory take a pragmatic view of this problem. They do not necessarily want to see lights switched off — but they do want them better designed. For those astronomers, it is a question of winning hearts and minds — particularly outside the REP area — and the developers, makers and sellers of outdoor lighting fixtures are their prime targets. It is no exaggeration to say that in some degree at least, these people hold the future of Australian optical astronomy in their hands.

FUTURE PROSPECTS

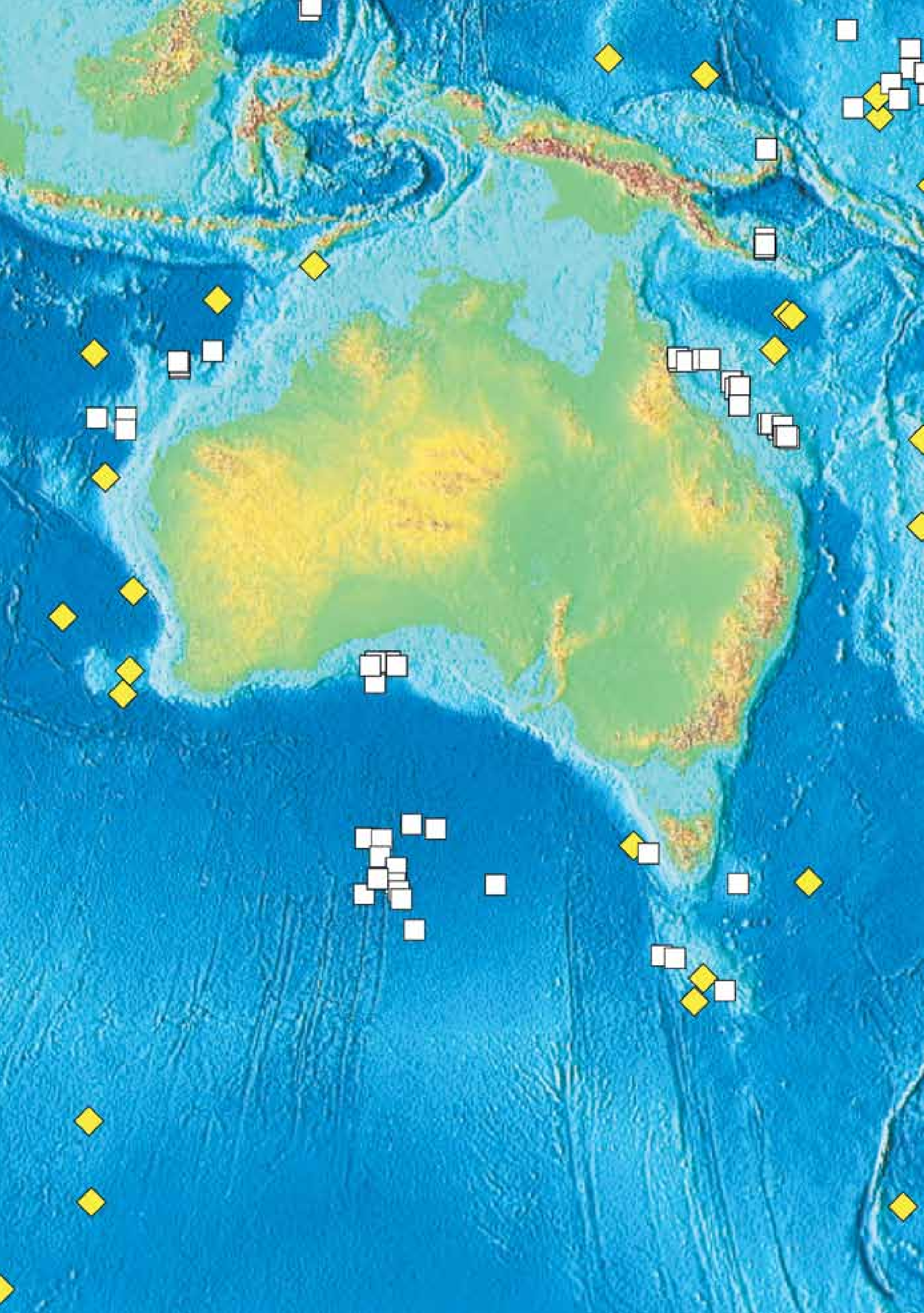
Today's astronomers — their sights firmly fixed on dark matter and dark energy — recognise that the need for dark skies imparts a pressing need for sustained public outreach. The aim is to win support for the endeavours of their observatories among local populations. This allows a direct link to be made between environmentally friendly outdoor lighting and the productivity of large telescopes in world-scale scientific projects.

For the coming generation of ELTs, the issue of sky background is even more acute. Such is their sensitivity that even the Earth's natural skyglow will provide an almost insurmountable challenge in recording and measuring faint objects. However, it is likely that most of these telescopes will be built on such remote sites that dark sky legislation and enforcement will be fairly tractable issues. Some, indeed, are already planned on existing sites with dark sky protection in place.

The campaign to save energy and at the same time keep our night skies dark is, as we have seen, a battle for hearts and minds more than anything else. But what greater ally can we have in this than the prospect of unlocking some of the deepest secrets of the Universe? By ensuring that the world's biggest telescopes are kept firmly in the dark, we may be on the brink of revealing the true nature of dark matter and dark energy, perhaps uncovering some of nature's deepest secrets.

ACKNOWLEDGEMENT

The information in this chapter comes from the work of a very large number of colleagues working in Australia and elsewhere. I should like to express my gratitude to them all for their papers, articles and personal communications.



EXPLORING THE EARTH'S VARIED AND DYNAMIC SEAFLOOR

DR JOANNE WHITTAKER AND PROFESSOR DIETMAR MÜLLER

INTRODUCTION

The floor of the ocean doesn't get that much attention, but when it does, the deep sea is often portrayed as an endless expanse of deep, smooth ocean floor inhabited by giant squids. Instead, ship and satellite geophysical data reveal steep cliffs and valleys over vast areas, sometimes over 3 kilometres in elevation (Figure 2). Other parts of the ocean floor are dead flat.

We know now that, thanks to plate tectonics, in geological terms the seafloor is very dynamic, being continually formed and entirely recycled over periods of approximately 200 million years (Figure 3). This is particularly the case when compared with the continents, which can remain relatively unchanged for millions or even billions of years. However, when German meteorologist and geologist Alfred Wegener first proposed that continents are drifting, he felt compelled to identify the driving force of this monumental motion. It was a quest he never completed: in 1930, Wegner froze to death during an expedition crossing the Greenland ice cap, after his theory had been dismissed as eccentric, preposterous and improbable.

Ocean basins cover 70% of the Earth's surface. What forces are at work and how the seafloor changes over time are secrets that have long haunted scientists since Dr Wegener's tragic fate. Today, more is known about the surface topography of the planet Venus than is understood about many areas of the seafloor. The study of the planet's evolution since its accretion 4.5 billion years ago comprises arguably the largest scale research project you can possibly undertake. But many of the secrets of continental drift and plate tectonics are to be found underwater.

The popular consensus is that sea level should always remain as it is now; and we think of a vast continent such as Australia as a stable geological feature, which probably always had an appearance more or less as it does today. It is difficult for us to fathom the meaning of geological, as opposed to human, timescales, and the manner in which the Earth's systems change over longer time periods. However, geological studies show us that the Earth's internal processes have led to changes in global and regional sea levels of hundreds of metres, long before humans existed. It is critical that we better understand processes and timescales we cannot observe directly. We might understand the idea that ice-ages may influence global climate and sea-level: melting ice caps may raise the global sea level, and a new ice age would lower it. But, in order to make useful predictions about future climate changes, it is essential to understand long-term global sea-level change and geochemical cycles.

Figure 1: Location of drill sites around Australia

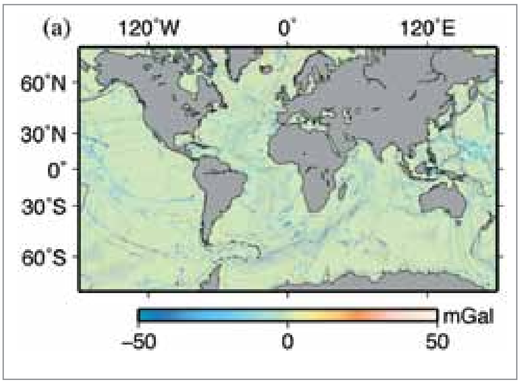


Figure 2. Global map showing downward continued satellite gravity. Note the varied textures of the seafloor

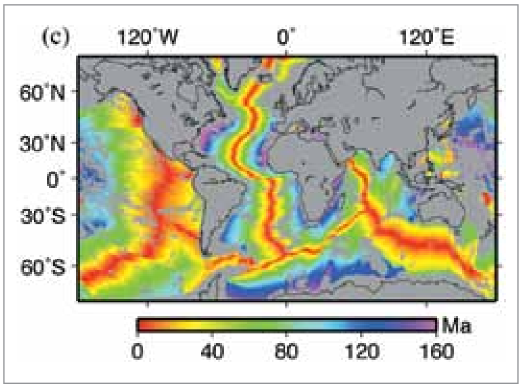


Figure 3. Global map showing the age of the seafloor

Studying the seafloor is critical for building reconstructions of the past position of the landmasses. In turn, these plate reconstructions are key to modelling long-term climate changes, glacial cycles and ocean currents. For example, the dramatic expansion of glacial cover across Antarctica 35 million years ago has been directly linked with the separation of Australia and South America from Antarctica. This left Antarctica isolated on the South Pole, surrounded by ocean. This allowed strong westerly winds to initiate the Antarctic Circumpolar Current, which thermally isolates Antarctica to the present day, and is believed to have contributed strongly to the extreme expansion of glaciers across Antarctica (Figure 4).

TECHNOLOGY

Just how you go about reconstructing the history of the ocean's depths is no small task, not least in the field of Earth sciences that often requires the modelling of phenomena for the entire globe.

The most important aspect of modelling where the continents were in the past, and what the oceans looked like millions of years ago, is imaging what the Earth's seafloor looks like now. Marine geo-scientific data are collected at various scales using a vast array of techniques. For example, recently the most comprehensive marine geophysical imaging and ocean drilling campaign in history around Australia has been completed, carried out by Geoscience Australia and the International Ocean Drilling Programs (IODP), resulting in the collection of geological data from more than 600 drill sites in Australasia/Antarctica (Figure 1), and thousands of kilometres of marine seismic, gravity and magnetic data, as well as high-resolution images of the sea floor. These data hold an enormous amount of information on the evolution of ocean basins and margins through time.

Over the past ten to twenty years, enormous progress has been made in methods for mapping the seafloor. Previously, the echosounding methods available relied on measuring a single acoustic signal. The travel time between the vessel to the seafloor and back gives scientists the 'depth' at a single point. Numerous such soundings add up to a depth profile. Prior to the advent of echosounders, mariners used a sounding line, essentially a rope with knots tied in it at regular intervals, to measure the depth of the seafloor. Obviously, this technique could only be used for relatively shallow depths and anything deeper than 100 metres or so was not routinely measured.

The relatively recent development of multibeam sonar systems has resulted in fairly sophisticated systems, which map a swath of seafloor underneath a ship's track, typically at least as wide as two times the

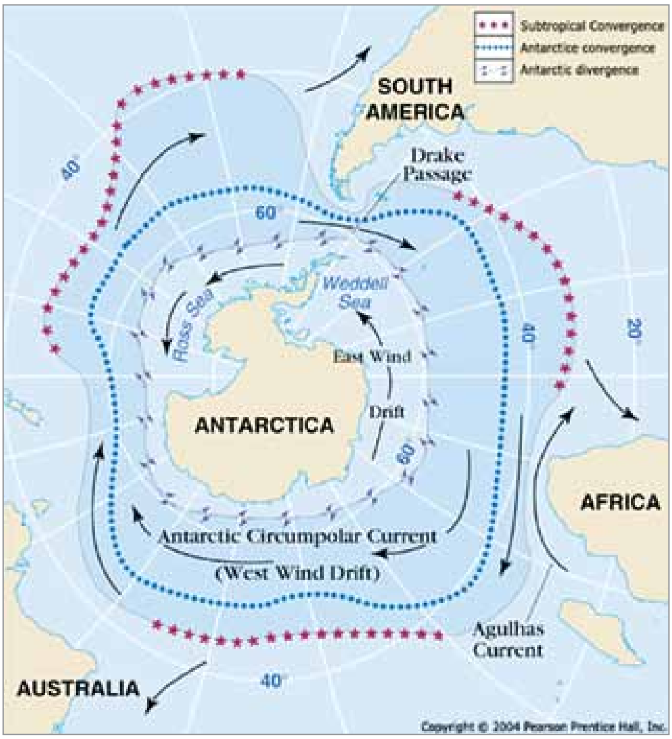


Figure 4: Antarctic Circumpolar Current

water depth. In Figure 5 we compare an old-fashioned seafloor topography (called bathymetry) map (Figure 5a), which was hand-contoured from profiles of ship depth soundings, with a modern multibeam bathymetry map of the same area southwest of Tasmania (Figure 5b). The multibeam image shows a nearly vertical cliff more than 1,500 metres tall dropping down from the South Tasman Rise, a submerged continental plateau, to the abyssal seafloor to the west. Also shown are previously unmapped volcanic cones. This example demonstrates how much we have learned about the seafloor from multibeam data.

This technology is critical. The ocean bed rests up to 7 km beneath the waterline, ruling out the possibility of humans seeing first hand this grand sunken underworld of moving plates and magma. While oceans might appear as 'big, blue, featureless bath tubs', in fact, they boast submarine mountain chains up to 3 km high, and trenches some 7 km deep. It's also a world of constant if slow activity, characterised by the spreading of the seafloor, and the formation of new ocean floors at mid-ocean ridges, where tectonic plates are moving apart. The multibeam sonar system helps to create images of what this world looks like now — but it remains painstaking work. Mapping multibeam bathymetry is a bit like mowing the lawn, you move backwards and forwards over the area you want to image. It's tedious, but it's relatively easy to make a static image of the ocean floor.

What's much more interesting is to then use these images to reconstruct the evolution of the seafloor over time. Then you can see the enormous geological time involved in plate tectonics, and the changing geometry and topography of oceans and continents.

THE SEAFLOOR AND PLATE TECTONICS

The Earth's surface is made up of a thin rigid skin (the tectonic plates, composed of the continents and the oceanic crust) sitting on top of a zone where solid rock is hot enough to flow like warm toffee, called the mantle.

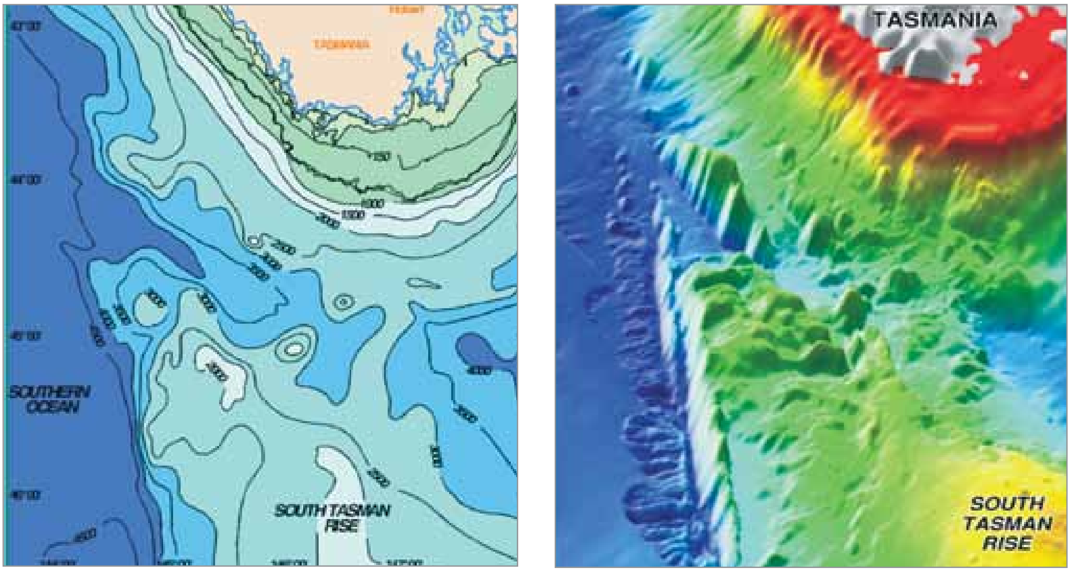


Figure 5. (a) Hand-contoured bathymetry map of the western South Tasman Rise and adjacent abyssal plain, compared with (b) a modern multibeam seafloor image of the same area (from Phil Symonds, Geoscience Australia).

Plate tectonics is an expression of the slow convection of the mantle over geological times, and the principal means by which the Earth loses its internal heat.

The formation of an ocean basin starts with continental rifting, the slow stretching of continental crust driven by the forces that result in the fragmentation of land masses (Figure 6). Continental rifting is the prerequisite to continental margin and basin formation, and is often stretched out over many tens of millions of years. When the crust eventually breaks, seafloor spreading starts, and the two conjugate, or opposite, margins become tectonically passive.

Once seafloor spreading occurs, new ocean crust is formed as the upwelling mantle reaches the surface and cools. Data gathered by oceanographic surveys starting in the 1950s led to the discovery of mid-ocean ridges, including a great mountain range on the ocean floor virtually encircling the Earth. Called the global mid-ocean ridge system, this immense submarine mountain chain is more than 50,000 kilometres long, in places more than 800 km across, and rises an average of about 4,500 metres above the sea floor. Though hidden beneath the ocean surface, the global mid-ocean ridge system is the most prominent topographic feature on the surface of our planet. Once formed, the seafloor subsides and deepens as a function of the square root of time.

The converse to seafloor spreading occurs where oceanic and continental plates converge: this location is known as a subduction zone. Here, the process of subduction recycles the oceanic plates back into the convecting mantle. Due to the unchanging size of the Earth, crust must be destroyed at the same rate as it is created. Subduction zones are where most of the largest and most deadly earthquakes occur, for example just offshore Indonesia, Japan and South America.

Plate tectonic processes control the production of magma, the generation and destruction of crust, and crustal deformation,

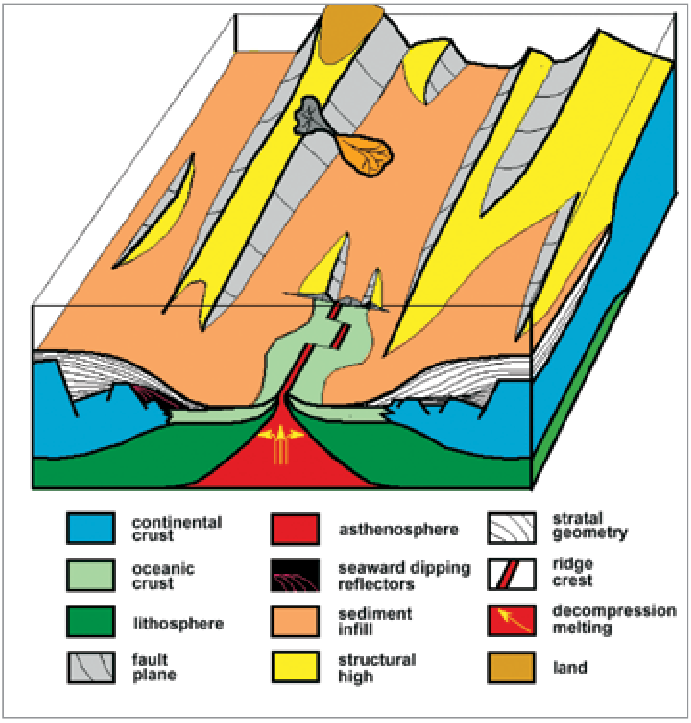


Figure 6. Sketch of a continental rift basin, formed by extension of continental crust. Once the crust breaks, seafloor spreading starts and breakup is completed.

driving the formation of basins and mountain belts with associated natural resources (Figure 7). The rates of plate creation at mid-ocean ridges and plate subduction in zones where plates converge exert major controls on the carbon cycle and other geochemical cycles, by means of seafloor weathering, mid-ocean ridge degassing, and subduction of sediments. Plate motions also determine the distribution of land and sea, and the volume of the ocean basins, which depends on the average global seafloor spreading rates (how fast tectonic plates are moving away from each other).

We use software called GPlates (see www.gplates.org) to combine plate tectonic modelling techniques and a multitude of datasets in order to reconstruct vanished ocean basins (Figure 8). The palaeo-oceans are modelled by creating “synthetic plates” whose locations and geometry are established on the basis of preserved ocean floor, regional geological data and the rules of plate tectonics. The resulting oceanic palaeo-age can then easily be converted into oceanic palaeo-depth maps, as the depth of ocean crust increases as the square root of its age, due to thermal cooling.

THE ROUGH ABYSSAL PLAINS

New seafloor created at a mid-ocean ridge forms what are known as abyssal plains. Abyssal plains cover approximately 50% of the Earth’s surface and lie at depths between 2500 m and 5000 m. Although Wikipedia describe abyssal plains as amongst ‘the flattest, smoothest and least explored regions on Earth’, they can actually exhibit varied and interesting textures and morphologies.

Abyssal plains contain two main features: fracture zones and abyssal hill fabric. Fracture zones (Figure 9) are linear features formed perpendicular to the mid-ocean ridge and can continue for hundreds of even thousands of kilometres across ocean basins. They can be up to 200 km wide with several kilometres of vertical relief. In contrast, abyssal hill fabric forms parallel to the mid-ocean ridge. These regions are less

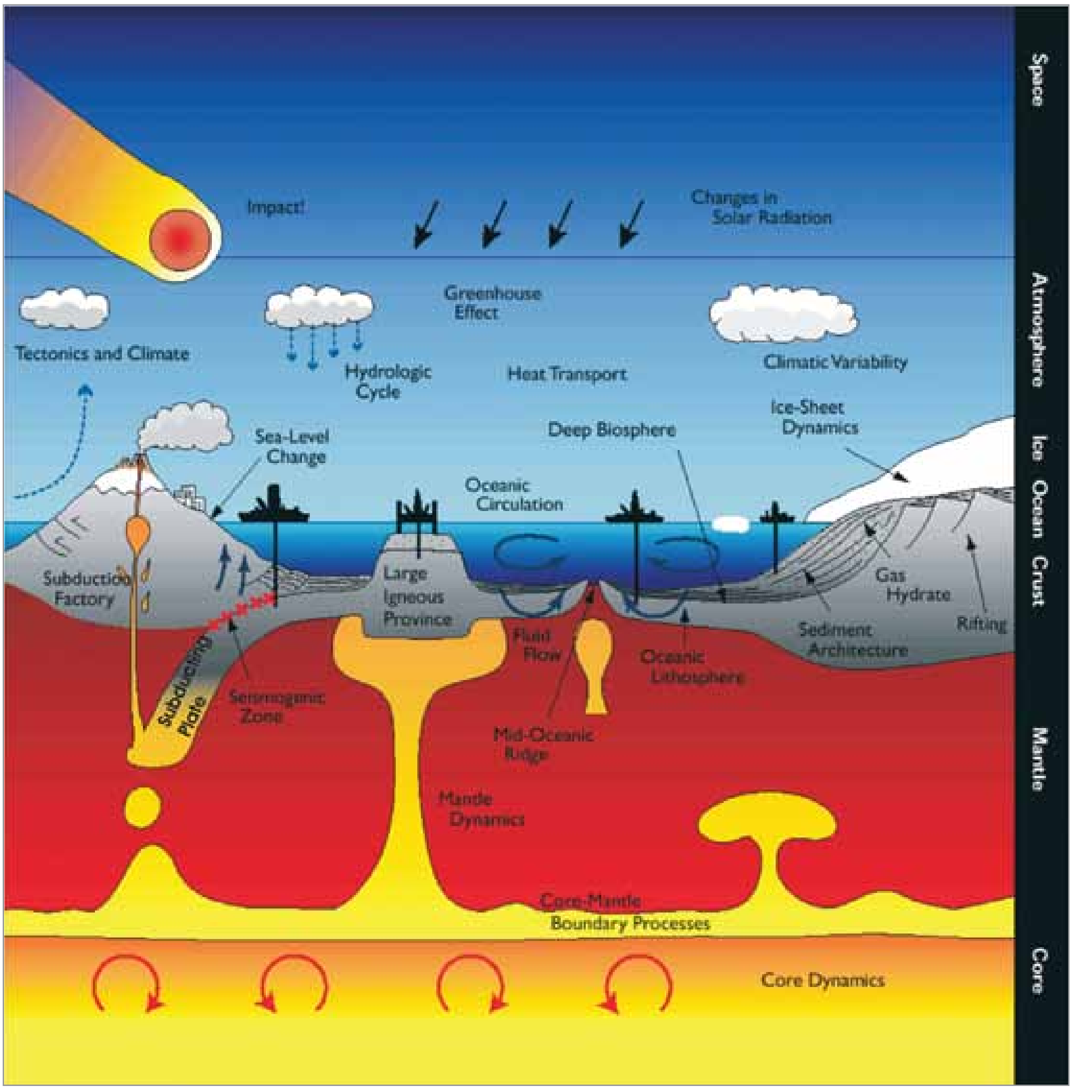


Figure 7. Cartoon of the dynamic Earth. Complex links among tectonics, global biota, ocean chemistry and circulation, and atmospheric composition control the long-term evolution of the climate system. The most fundamental process controlling this system is plate tectonics. Modified from *Integrated Ocean Drilling Program initial Science Plan*, unpublished report, 2000.

dramatic than fracture zones, with only a few hundred metres of vertical relief: think of sand dunes stretching out across a desert, one after the other.

From the above description it may be easy to think that all the abyssal plains look very similar; however, this is simply not true. Figure 10 shows the roughness of the seafloor, and you can see that there are large variations in how rough the seafloor is in different areas. For instance, in general the eastern Pacific is very smooth, punctuated only by the occasional roughness of a very long fracture zone. In contrast, the seafloor surrounding the mid-ocean ridge stretching from southern South America, beneath Africa to the middle of the Indian Ocean is very rough.

All these regions were formed by the same process of seafloor spreading at mid-ocean ridges. So, why should they be so different? It turns out that factors such as the speed and

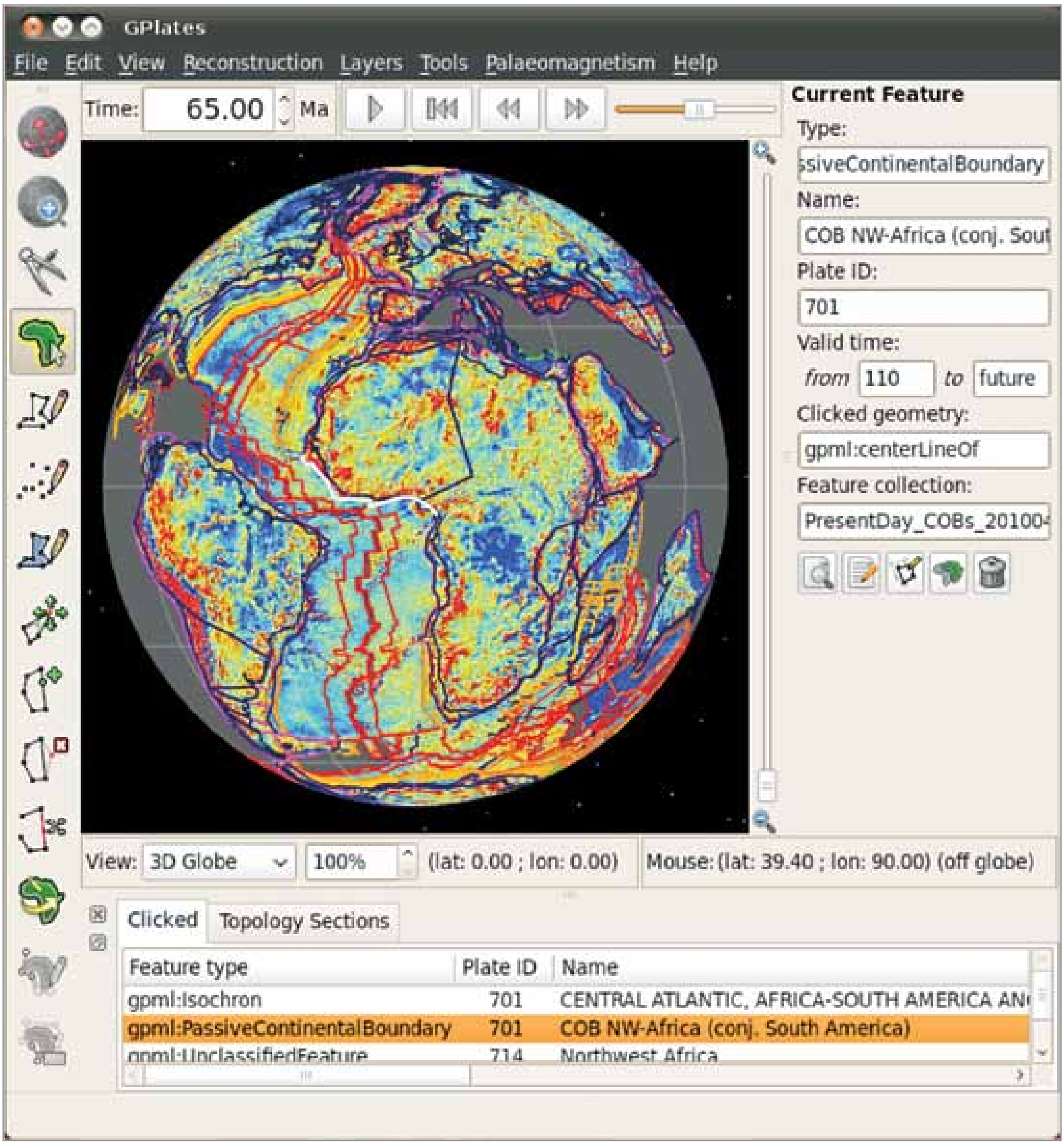


Figure 8: Screenshot of the software GPlates, used to reconstruct vanished ocean basins

angle at which the tectonic plates are moving apart from each other, and how hot the underlying mantle is, have a large effect on what the newly formed seafloor looks like.

Mid-ocean ridges, where there is rapid creation of new oceanic crust, have only a small or no median valley, with vertical relief up to only a couple of hundred metres. In contrast mid-ocean ridges, where new oceanic crust is formed gradually due to the tectonic plates slowly moving apart from one another, are characterised by a prominent median valley that may exhibit vertical relief up to 1–2 kilometres high and can be up to 20–30 kilometres wide.

The speed the plates move apart and subsequently cause new seafloor to form also affects the shape and density of fracture zone formation and the ruggedness of the abyssal hill fabric. The slower that continents move away from each other, the rougher becomes the ocean floor, due to new crust being

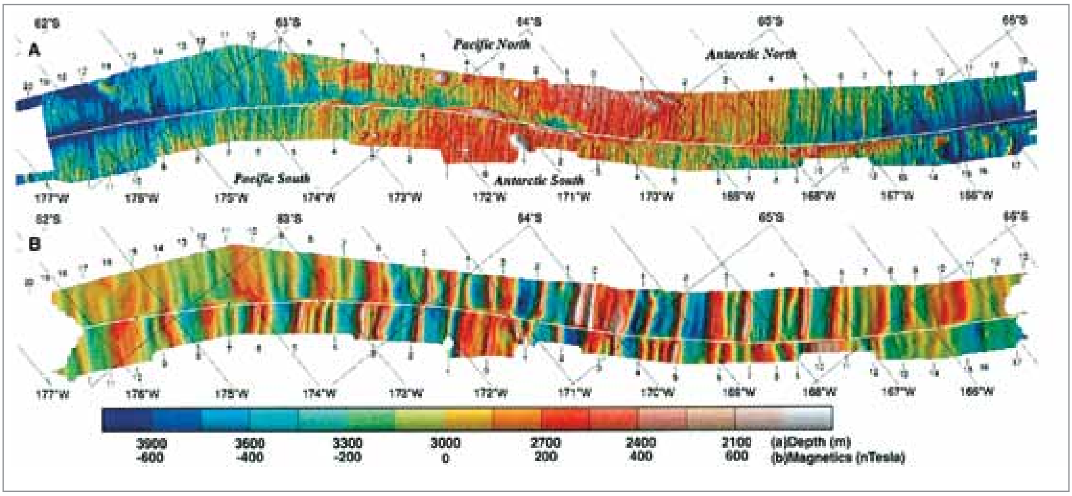


Figure 9. Pitman Fracture Zone in the South Pacific. (a) topography, (b) magnetic anomalies (Source: S. Cande)

formed and then stretched and deformed before a new batch of seafloor starts to form. This process results in the formation of much more rugged abyssal hill fabrics; with faster spreading however, new seafloor is being constantly created without the stretching and deforming stage. Slower spreading rates also result in the formation of more closely spaced fracture zones, although the mechanism by which this occurs remains poorly understood.

The temperature of the mantle beneath a mid-ocean ridge can also affect the roughness or smoothness of the newly forming seafloor. Take a close look at Figure 11: you will notice that the edges of the ocean basins, close to the continents in the Indian and Atlantic oceans, are often smoother (more blue in colour) than the centre of the basins, which are rougher (more red in colour). The reason for this is that the seafloor at the edges of the basins is older and formed following the breakup of the supercontinent Pangaea in the Jurassic, a time when dinosaurs roamed the Earth. At this time the (then) new ocean basins were created underlain by partially molten rocks much hotter than we observe under mid-ocean ridges today. The lavas erupting at these new mid-ocean ridges resulted in extremely smooth topography in the Atlantic and Indian oceans. After extracting molten rock from the mantle over 100 million years, this giant melt supply system had cooled down, making it much harder to create new ocean crust. At this point the ocean floor became more brittle, developing large faults, abyssal hills and submarine mountain ranges.

Understanding why and how the rough patches of ocean floor form is important as they affect the deep ocean circulation, with dramatic affects on the Earth's climate. Where it meets abyssal peaks and ridges, deep water circulating around the bottom of the abyssal ocean is forced to mix and rise to the surface, profoundly affecting ocean circulation and ultimately global climate.

HOTSPOT TRACKS

The vast majority of earthquakes and volcanic eruptions occur near plate boundaries, but there are some exceptions. For example, the Hawaiian Islands (Figure 12), which are entirely of volcanic origin, have formed in the middle of the Pacific Ocean more than 3,200 km from the nearest plate boundary. How do the Hawaiian Islands and other volcanoes that form in the interior of plates fit into the plate-tectonics picture?

In 1963, J. Tuzo Wilson, the Canadian geophysicist who discovered transform faults, came up with an ingenious idea that became known as the "hotspot" theory. Wilson noted that in certain locations around the world, such as Hawaii, volcanism has been active for very long periods of time. This could only happen, he reasoned, if relatively small,

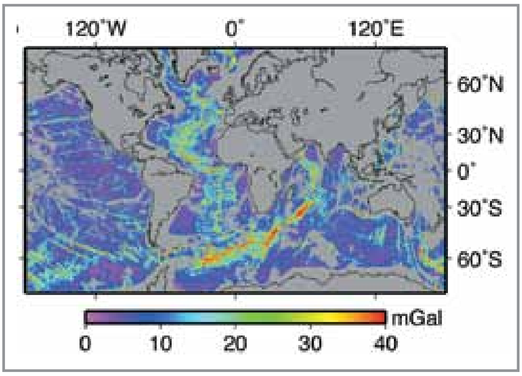


Figure 10. Roughness of the seafloor. Purple areas are relatively smooth, green through red areas are relatively rough.

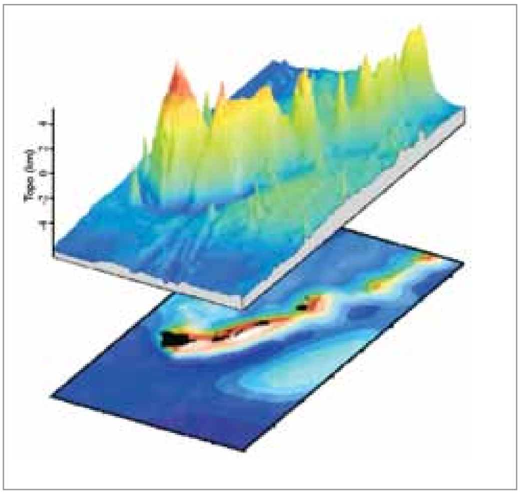


Figure 12: Bathymetry of the Hawaiian Islands

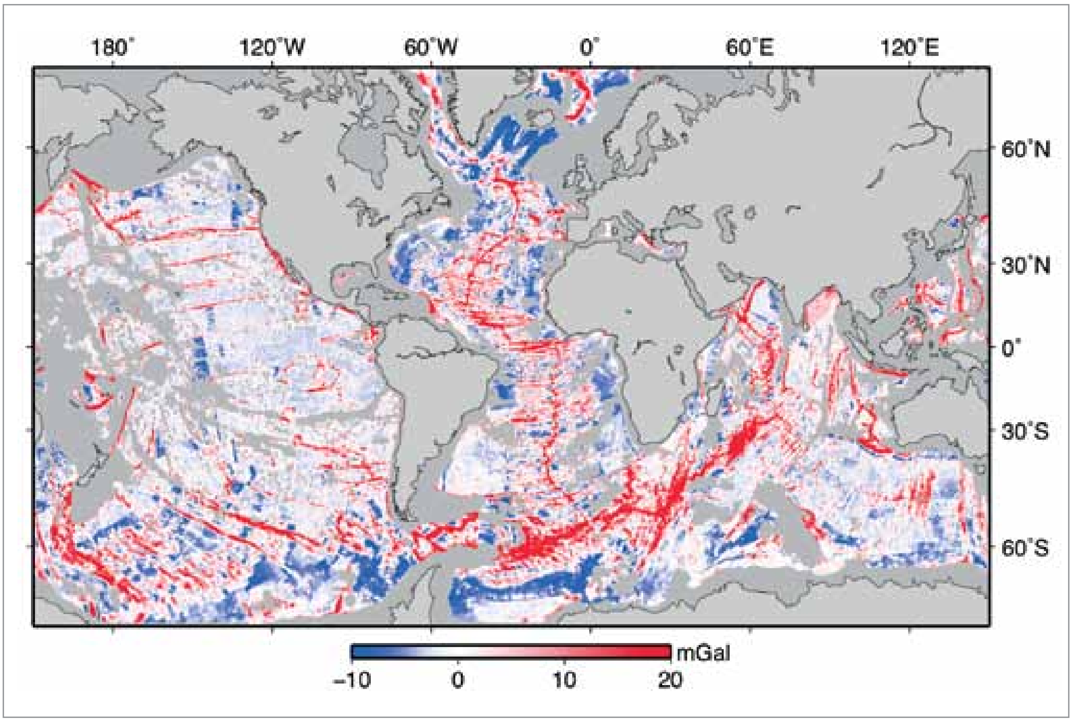


Figure 11. Global Residual Roughness

long-lasting, and exceptionally hot regions — called hotspots — existed below the plates, providing localized sources of high heat energy (thermal plumes) to sustain volcanism (Figure 13).

One of the most enigmatic hotspot chains is the Emperor-Hawaii chain that extends from the Aleutians to the island of Hawaii (Figure 14). The cause of the major bend in the chain has long puzzled scientists and been the source of much scientific debate. In 2007, it was proposed that the destruction of an entire mid-ocean ridge, known as the Izanagi Ridge, 50 million years ago in the Pacific

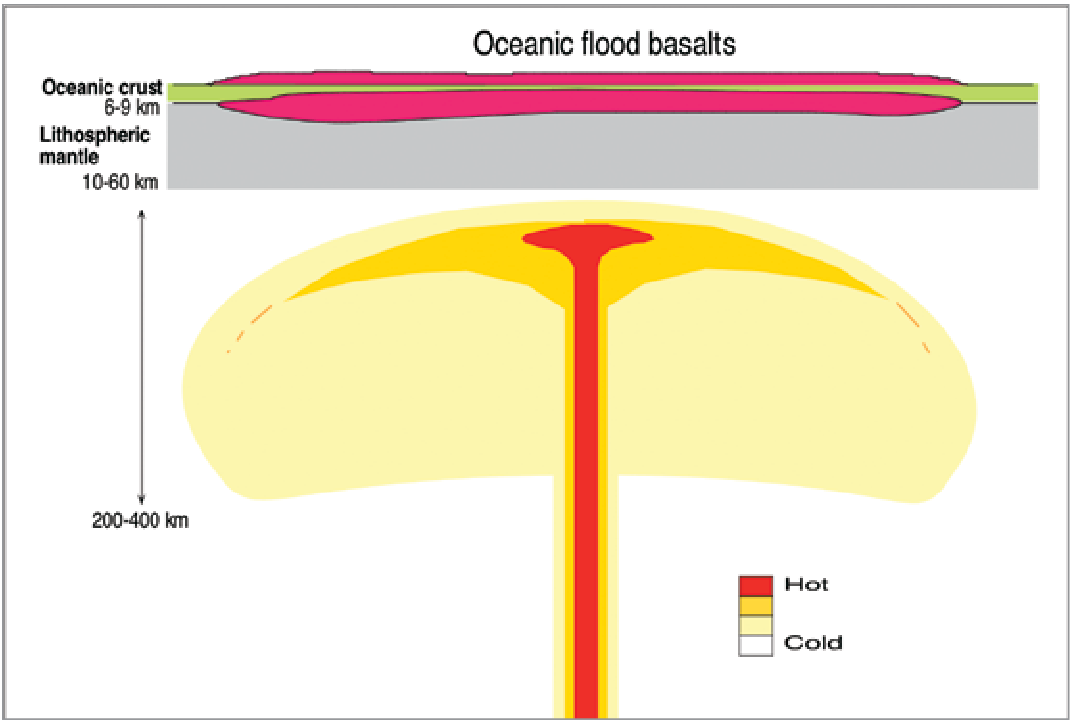
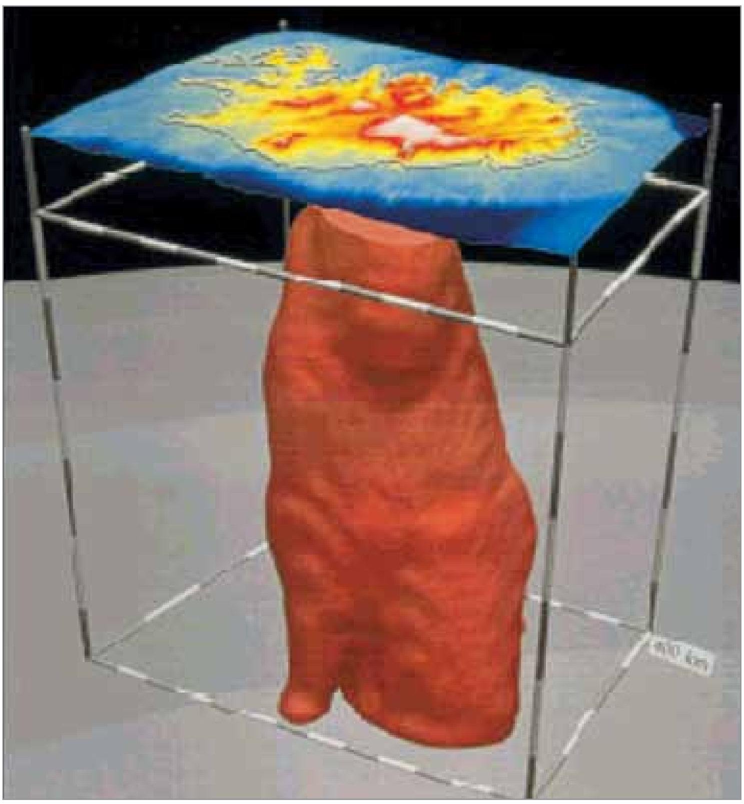


Figure 13: (a) Cartoon of Upwelling mantle plumes One theory suggests that a “plume head” develops, above a “plume stem”. (b) Iceland plume stem as imaged with seismic tomography.



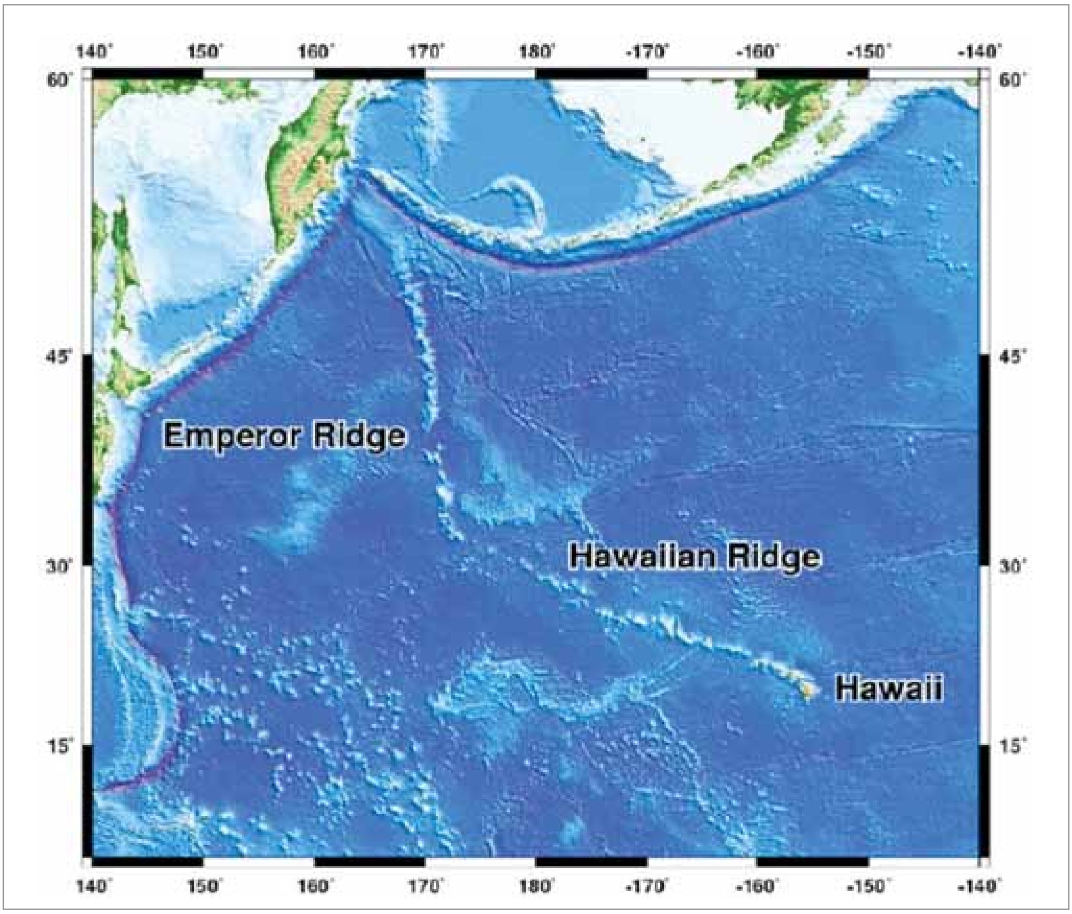


Figure 14: Emperor-Hawaii chain

Ocean initiated a chain reaction of geological events from Hawaii to Antarctica and caused the bend.

Think of the Earth as a giant recycling system, where new crust is continually created at mid-ocean ridges and crust is destroyed at the ocean edges as it plunges into the sticky rocks of the mantle below, through a process known as subduction. With a steady conveyor belt, crust continually moves away from the mid-ocean ridges towards the ocean edges and subduction. However, when crust is destroyed more rapidly than it is created at a particular conveyor belt then the mid-ocean ridge is headed for a catastrophe! When the mid-ocean ridge itself hits the edge of the system and is destroyed, it upsets the balance between creation and destruction of crust. After the death of a mid-ocean ridge subduction ceases, but may be jump-started again depending on the forces pushing and pulling the plate. The re-ignition of the conveyor belt sends a giant shock through the system of plates akin to a stockmarket crash.

Such a tectonic cataclysm and its aftermath occurred approximately 50 million years ago: the ancient Izanagi mid-ocean ridge was entirely destroyed in a dramatic fashion when it plunged beneath an area stretching from Korea to north of Japan (Figure 15). This event affected the motions of an entire hemisphere of plates surrounding the Pacific, with changes cascading as far away as the Southern Ocean between Australia and Antarctica.

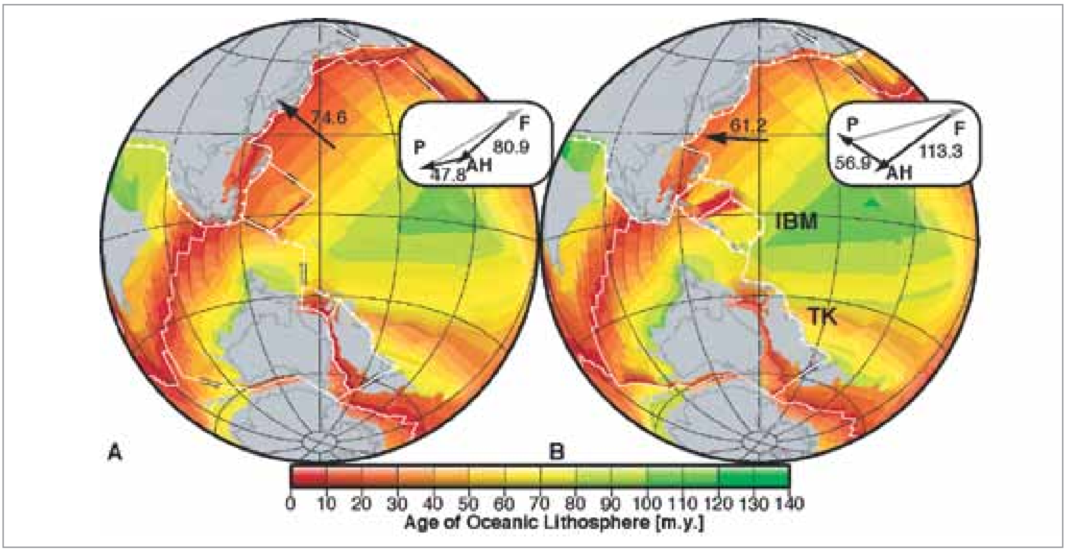


Figure 15: Global age grids at two time slices (a) 55 Ma, and (b) 45 Ma. Black arrows show western Pacific (30˚N, 150˚E) plate motion over the stages 60-50 Ma and 50-40 Ma respectively. Note that eastern and western Pacific directions and changes in plate motion are substantially different.

In fact, the reorganisation of plate motions caused by the Izanagi Ridge subduction played a crucial role in Australia's geological history — it resulted in a previously unmapped swerve and acceleration of Australia away from Antarctica. The rearrangement in plate motions also led to the birth of several new subduction zones in the southwest Pacific, forming islands such as Tonga that otherwise would not exist. The new model has far-reaching implications for understanding the formation of natural resources in frontier exploration areas, extending from the Great Australian Bight to the Lord Howe Rise east of the Tasman Sea, and the circum-Pacific volcanic "ring of fire".

DYNAMIC TOPOGRAPHY

We tend to look at the present state of the Earth as a static system. Popular consensus in the face of global warming is that sea level should always remain as it is now, and we think of a vast continent such as Australia as a stable geological feature, which probably always had an appearance more or less as it does today. It is extremely hard for us to fathom the meaning of geological, as opposed to human, timescales, and the powerful processes originating in the hot, churning interior of the Earth.

Yet these processes have changed regional sea levels by hundreds of metres, or even several kilometres, through geological time long before humans existed. A graphic example of the enormous magnitude of relative sea level variations due to geological forces can be found in fossilised worm burrows — originally formed near the sea surface on seamounts in the southern Pacific Ocean — that are now found under 2.5 kilometres of water. Such drowned, usually flat-topped seamounts were named "guyots" by Harry Hess of Princeton University during World War II, when he served as an officer aboard a naval vessel. Hess recognized that such seamounts must have undergone enormous subsidence, but he did not know the reason at the time. Today we know that oceanic crust subsides by several kilometres after it forms at mid-ocean ridges, simply because it cools and contracts.

Such changes occur so slowly that we cannot observe them directly — only the geological record gives us some clues for vast pre-historic basins, floods or mountains that may have existed at times when dinosaurs roamed the face of the planet. Most of us understand the idea that ice-ages may influence global climate and sea-level:

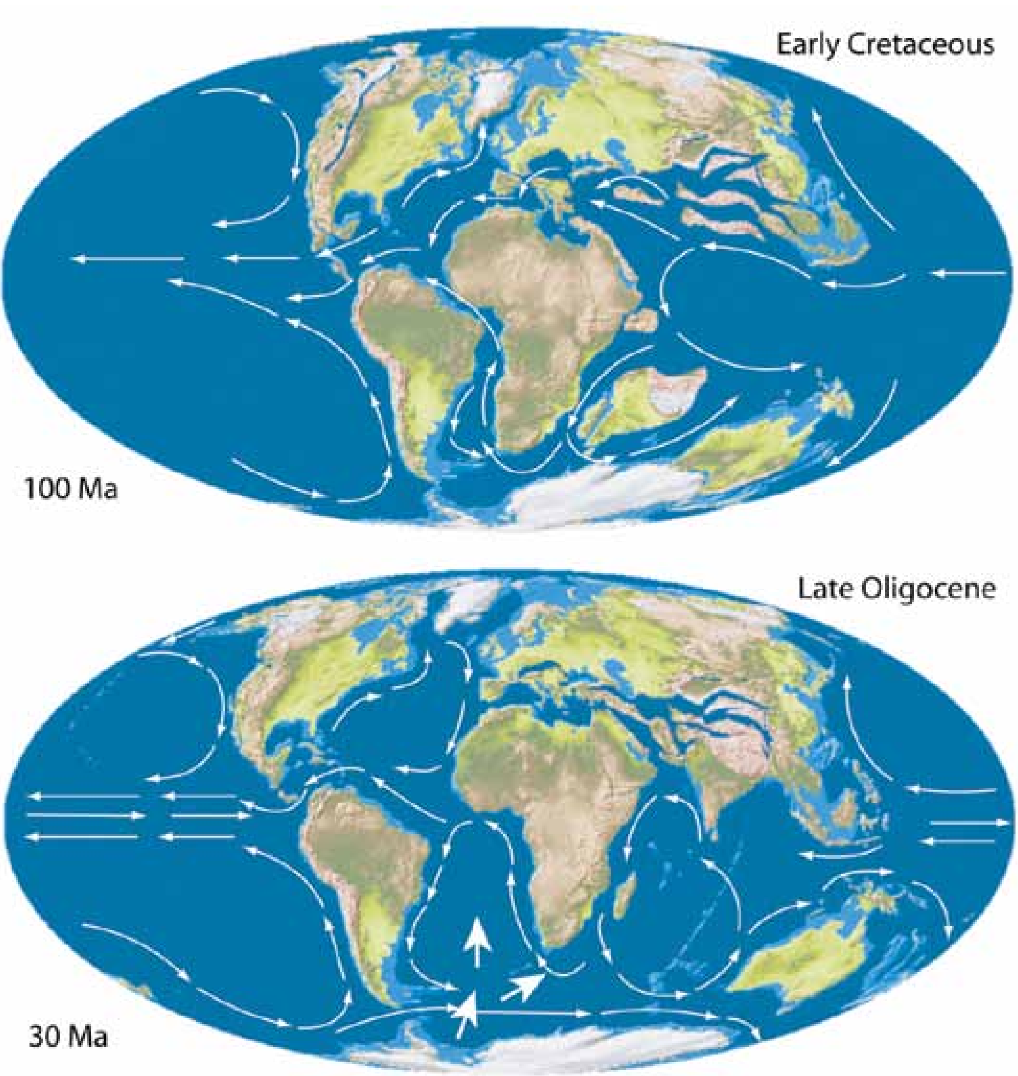


Figure 16. Reconstructions of continental topography for the Early Cretaceous at 100 Ma and in the Late Tertiary (Oligocene) at 30 Ma. Overlain is a sketch of oceanic surface circulation.

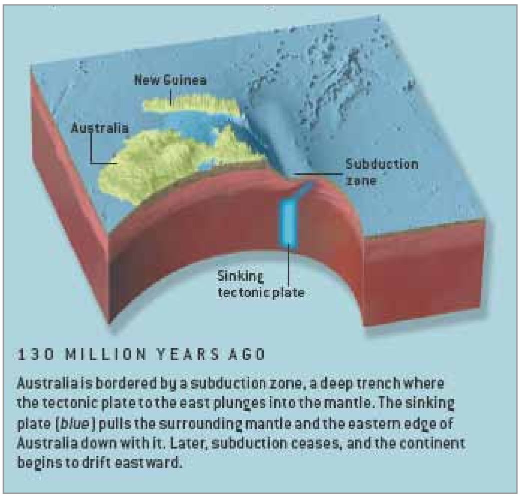


Fig. 17a

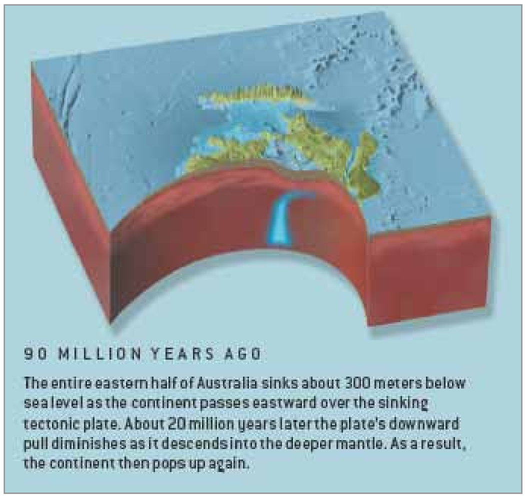


Fig. 17b

Figure 17. A computer model reveals how the ghost of an ancient subduction zone dragged down a continent. Modified from Gurnis (2001)
(a) In the Early Cretaceous, 130 million years ago, Australia is bordered by a deep oceanic trench in the east, where an oceanic plate is being recycled (subducted) back into the Earth's mantle. The box shown comprises a depth of nearly 3000 km, down to the boundary between the Earth mantle, shown in dark red, and the Earth's core (not shown). The vertically sinking oceanic plate in the mantle is shown in blue, representing cooler temperatures than the surrounding mantle.
(b) Reconstruction of Australia in the Late Cretaceous, 80 million years ago. Between 130 and 80 million years ago, the Australian plate moved east, overriding the sinking plate to the east. As a consequence, eastern Australia is drawn down by the "negative Buoyancy" of the sinking slab, resulting in marine inundation of large parts of eastern Australia by a shallow sea. After 80 million years ago, the downward drag of the subducting plate on the surface slowly diminished due to the increasing vertical distance between the sinking slab and the continent, resulting in a rebound and uplift of eastern Australia.
(c) Present day location and topography of Australia. About 45 million years ago, Australia started moving north, resulting in convergence with Indonesia. The entire continent is now drawn down once again, but not as much as in the Cretaceous. This is caused by a system of subducting slabs around Indonesia and Papua New Guinea (not shown here), which Australia is approaching.

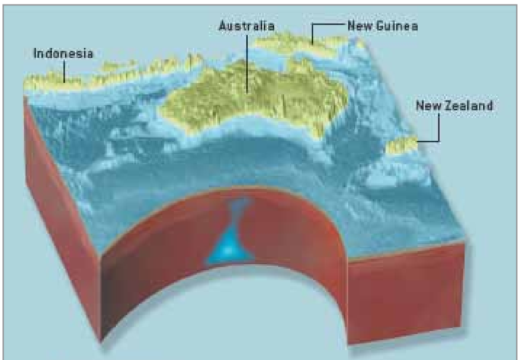
melting ice caps may raise the global sea level, and a new ice age would lower it. However, gigantic forces driven by sinking or rising branches of the global mantle convection system not only help drive the horizontal motion of tectonic plates, but they also lift and lower the continents, resulting in dramatic regional sea-level changes that have nothing to do with the history of ice caps, or with human impact.

During the Cretaceous period, Australia, South America, Africa, India, Antarctica and New Zealand were assembled into a vast supercontinent called Gondwanaland, which existed for more than 400 million years before it fragmented into the landmasses we know today, due to seafloor spreading (Figure 16).

The dispersion of Gondwanaland together with the seafloor spreading in the Pacific Ocean changed the palaeogeography of southern Eurasia and Australasia substantially.

Some of these changes were due to Australia separating from its neighbouring landmasses of India and Antarctica and the coeval formation of new ocean basins and coastlines. However, some of the changes were due to Australia's motion relative to the comparatively stable underlying mantle.

Downwelling or upwelling mantle can have a large affect on the overlying crust. Downwelling mantle causes downward vertical motion of a continent, also called subsidence, which may result in flooding and widespread marine inundation. Vertical motion



PRESENT DAY

Australia lies north of its former site, having been pushed there by activity in adjacent tectonic plates beginning about 45 million years ago. The entire continent has dropped relative to its greatest elevation as the result of a downward tug in the mantle under Indonesia—a landmass that is also sinking.

Fig. 17c

of the Earth's surface associated with mantle convection is extremely slow: It typically has rates less than 100 m/million years, or 0.1 mm/year.

Australia shows many oddities in its distribution of inland seas through time. During the Cretaceous period (145-65 million years ago), the marine inundation of Australia and global sea-level curve were out of phase. Maximum flooding of Australia occurred between 120-110 million years ago, when a large fraction of the continent, especially in the east, was covered by a shallow sea (Figure 17) — paradoxically, when global sea levels were relatively low.

During the Late Cretaceous, eastern Australia became progressively uplifted and exposed, with a flooding minimum 80-70 million years ago, again in disagreement with global sea levels that are inferred to have been near a maximum. The puzzle presented by this Australian example can be explained by dynamic topography, caused by upwelling and downwelling mantle material beneath the Australian continent.

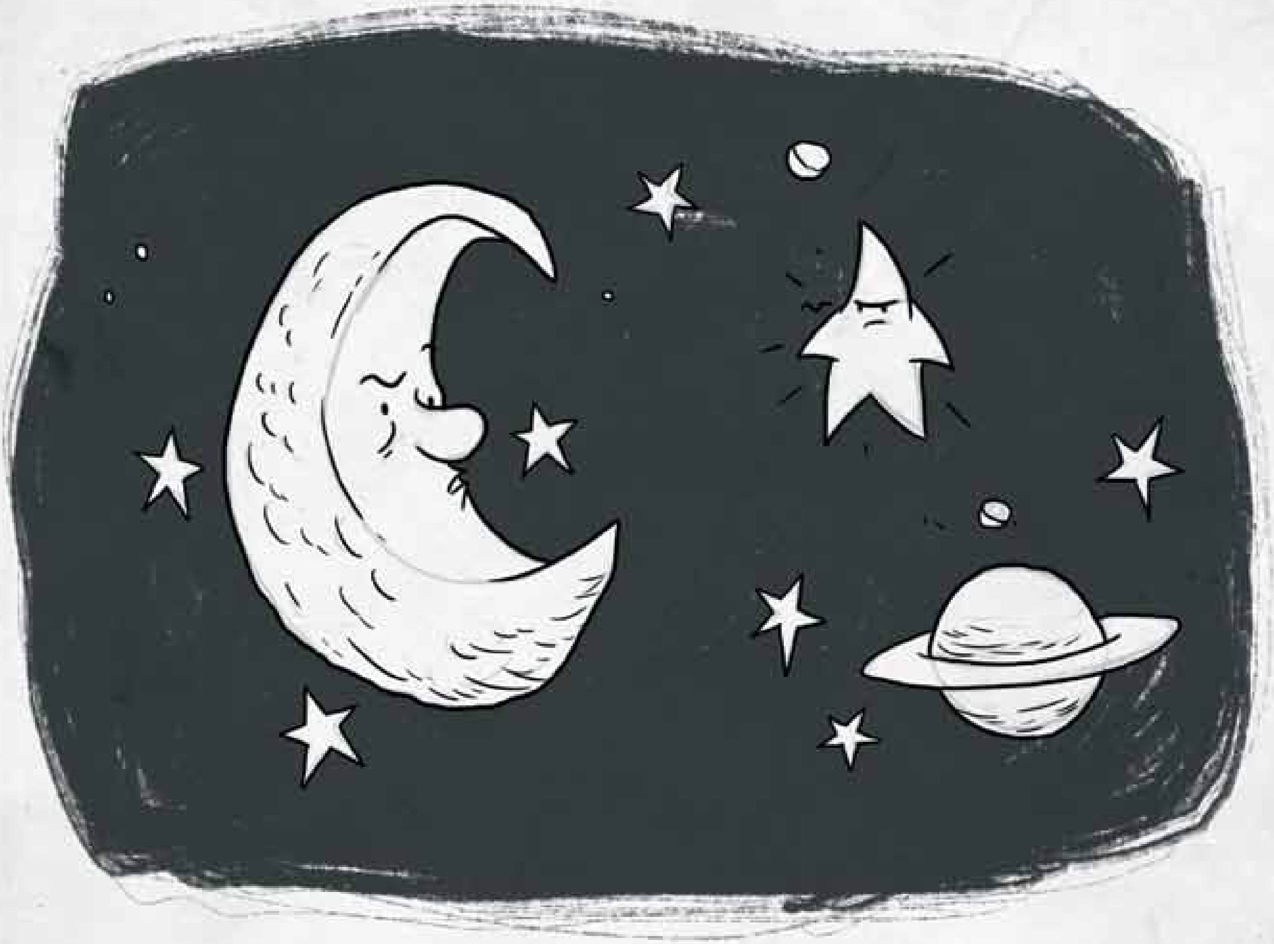
CONCLUSION: FULL CIRCLE

Before the advent of plate tectonics, scientists believed that the history of the Earth's surface could be understood entirely based on vertical motions: it was assumed that ocean basins are deep because they have sunken, and mountain belts exist because they have risen — without any involvement of horizontal motions.

Then the theory of plate tectonics came along, and it seemed that its framework allows us to account for most geological observations by means of horizontal motions between plates, and the ensuing sea-floor spreading, subduction and collisions, accompanied by continental aggregation and dispersal. Changes in the elevation of continents were thought to be a direct consequence of horizontal plate motions, creating mountain belts that are subsequently eroded by physical and chemical weathering.

Now, based on sophisticated geodynamic models that include mantle convection and horizontal plate motions, we have learned that the vertical ups and downs of large landmasses —"elevator tectonics" — may be only indirectly related to plate tectonics, and not an immediate consequence of horizontal plate motions after all.

The rich diversity of the Earth's landscape, from the deepest ocean trenches to the highest mountain chains, from vast flat plains to the volcanic "ring of fire", arises not just from the crust lifting and falling, and not just from spreading, subducting and colliding — it comes from all of these things. Our theory of the Earth has come full circle.



"I think I preferred you when you were a collapsing cloud of material composed primarily of hydrogen, along with helium and trace amounts of heavier elements," said the moon.

TWINKLING STARS

DR KARL KRUSZELNICKI

One of the best-known rhymes in the English language is "Twinkle twinkle little star, How I wonder what you are". The rhyme's author, Jane Taylor, first published it way back in 1806, in a book called *Rhymes for the Nursery*. She originally called the poem "The Star" for its first publication. And sure enough, if you look up at the night sky, the stars really do seem to twinkle.

But the reality is that stars don't twinkle.

STARS ARE NOT CONSTANT CANDLES …

The stars we see at night are huge objects, of the order of a million kilometres in diameter. All the stars that we can see with the naked eye are quite close to us, and well and truly inside our galaxy, the Milky Way.

For a star to actually "twinkle", it would have to get brighter and duller by a noticeable amount, and do this a few times per second.

Stars can in fact vary their brightness, but not a few times per second.

For example, our star, the sun, will change its brightness by 0.1 per cent (or one part in a thousand) over the Solar Cycle of about 11 years. We've been mapping this for about 400 years. Recent peaks were in 1991 and 2002, with the next one expected around 2013/2014, and so on.

There are some uncommon stars, called Variable Stars, that can vary in brightness not by one part per thousand, but up to 10,000 times. They can do this over periods from as short as a few hours to as long as a year.

Some stars will destroy themselves in a gigantic explosion called a supernova, becoming millions or billions of times brighter than normal. Of course, they can do that only once.

But no stars can vary their brightness as much and as quickly as the twinkling star in the nursery rhyme would have us believe.

… BUT THAT'S NOT WHY THEY TWINKLE

So what makes stars seem to twinkle, to the joy of stargazers, but to the intense annoyance of astronomers?

The answer is our atmosphere — that turbulent, fragile, translucent ocean of five trillion tonnes of air that is our window to the heavens.

Imagine that you are looking at the bottom of a shallow pond, with the surface of the water totally smooth. Everything on the bottom of the pond is easy to see. Suddenly, a wind springs up, rippling the surface of the water — and the floor of the pond now appears fuzzy. Then, to make matters worse, a stream of turbulent water suddenly begins to flow into the pond exactly where you are looking. Now it's impossible to pick out small details on the bottom. The "twinkling" water blurs all the details.

You can see another example of "twinkling" on a country road on a hot day — the shiny "lake" in the distance. The sun heats up the dark road, which heats up the air immediately above the road. The hot air next to the road bends the light more than the cooler air in the next layer higher up. The effect of all this "physics" is to make a virtual mirror out of this hot air. This "mirror" reflects the sky, and so you see a shiny patch on the road off in the distance. So that "water" in the road ahead is just the sky.

In the same way, our atmosphere can make the stars above appear to "twinkle".

In our atmosphere there are hot and cold areas. They appear as millions of columns or "bundles" of warmer or colder air, tens of centimetres across, and at an altitude of several kilometres. As the light from a star passes through these columns, on its journey down to the ground, it gets bent this way and that. The actual point size of the star is very small, but the turbulent air makes it look much bigger.

HOW TO FIX TWINKLING

The fact that air can bend light was first noted by the ancient Greek astronomer Cleomedes. In 1706, the great thinker Sir Isaac Newton in his work *Opticks* proposed that a solution to air turbulence interfering with telescope viewing might be found "... in a most serene and quiet Air, such as may be found on the tops of the highest Mountains above the grosser Clouds". Indeed, many of today's great telescopes are in such locations.

The whole point of the Hubble Space Telescope was to get a biggish telescope totally above the turbulent air. Indeed, the astronauts on the International Space Station get to see the stars without any added twinkling. The astronaut Jap Apt, who helped launch the Compton Gamma Ray Telescope said in 1991, "for one thing, the stars don't twinkle because you're above the atmosphere". And the astronaut Dr Edgar Mitchell, who went to the moon in Apollo 14, said, "without the atmosphere to block, the stars don't twinkle".

But there is a second way for astronomers to get around the twinkling of the stars. First, they use a laser to measure exactly how much the atmosphere is interfering with the incoming starlight. Second, they use this information to distort the mirror of the telescope, hundreds of times each second, to compensate for this interference. The US Air Force in 1982 first used this technology to peek at Soviet satellites. When the technology was declassified, the astronomers began using it. By 1990, it was built into the 3.6 metre telescope at La Silla, in Chile.

A third way for astronomers to get a clear image, unperturbed by twinkling, is to take a very short exposure, say one-hundredth of a second. But this will work only with very bright objects.

So the light from a distant star flies unperturbed for thousands of years, to get massively distorted only in its last dozens of microseconds of flight before it lands, as a twinkle, in our eyes.

TWINKLE LITTLE STAR

ALL THE GOLD THAT YOU WILL EVER TOUCH IN YOUR LIFE WAS "MADE" INSIDE A STAR AS IT EXPLODED. IT WAS BLASTED INTO SPACE, BECAME PART OF OUR PLANET ABOUT FOUR AND A HALF BILLION YEARS AGO, WAS FOUND AND ENDED UP IN YOUR HAND.

HELP US TO HONOUR EXCELLENCE

The Professor Harry Messel International Science School recognises, rewards and honours excellence in talented senior high school students, and to encourage them to pursue careers in science. The ISS is a two-week, fully residential program of lectures, workshops, tours and special events, which participating scholars attend at no cost. Each scholarship is valued at approximately A$3,000.

Alumni of the International Science Schools can be found in senior positions in all walks of life, with many of them acknowledging that 'their' Science School was responsible for changing their lives, and recalling its two weeks as an exciting developmental experience.

The Science Schools, renamed the Professor Harry Messel International Science Schools in 1999 to honour Harry for his foresight, have continued uninterrupted since then. In order to ensure the continuation of the International Science Schools, so that future students may also benefit, the Science Foundation for Physics established the Messel Endowment. The Endowment aims to raise sufficient donations to fund the International Science Schools in perpetuity.

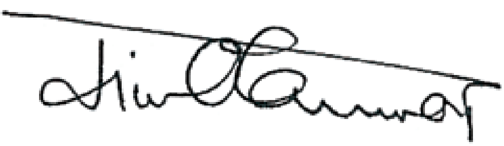
Over the last decade, the fundraising campaign has come a long way, and ISS programs in from 2005 to 2011 have received considerable benefit from the endowment. As acknowledged with gratitude in the front of this book, many individuals and companies have already contributed to the Messel Endowment. All our endowment donors are also acknowledged on our web site, in all books of the lectures presented at future International Science Schools, and on a permanent display in the School of Physics.

Despite this success, the Endowment still has a long way to go, and the Foundation is determined to achieve the target. We hope you have enjoyed your time at the International Science School, or had your mind expanded by reading the chapters submitted by the ISS lecturers and their on-line podcasts. To help us continue the work of the Foundation, we also hope you that you might consider making a tax-deductible gift to the Messel Endowment, and make a positive impact on the life of a future scientist!

A contribution of A$30,000 to the Messel Endowment will ensure the participation of one student in perpetuity, but every donation — large or small — helps us reach our goal.

A donation form has been included on the next page, and more donation forms are available from The Messel Endowment at www.physics.usyd.edu.au/foundation/

Please join us today in our vision for the young scientists of tomorrow through the Messel Endowment.



Mr Jim O'Connor
President, Science Foundation for Physics

THE UNIVERSITY OF SYDNEY

Science Foundation for Physics

Donation Form

PERSONAL DETAILS

Title: ☐ Mr ☐ Ms ☐ Mrs ☐ Miss ☐ Dr ☐ Other_______________________

First name/s: Last Name:

Mailing
Address:____

City: State: Post Code:

Country: Email:

Tel Home: ( ) Tel Business: ( )

Mobile:

DONATION DETAILS

Please accept my tax deductible gift of $

The gift is to be used for the following:

Gifts of $2.00 and over are tax deductible

PAYMENT DETAILS

☐ *Cheque* (please make payable to: The University of Sydney)

Credit Card: ☐ Visa ☐ MasterCard ☐ American Express ☐ Diners Club

Card No:

Expiry Date: / Cardholder Name:

Signature:

☐ I would like my gift to remain anonymous

☐ Please send me information about the Science Foundation for Physics and/or the School of Physics

☐ I am interested in receiving information about making a bequest to the Science Foundation for Physics.

You can make your gift on-line go to www.usyd.edu/gift

Thank you for your generous support

Send or Fax to: Science Foundation for Physics
School of Physics (A28)
The University of Sydney, NSW 2006 Australia
T: + 61 2 9036 5194 Fax 61 2 9351 7726

PRIVACY ASSURANCE: The information you provide on this form is collected by the University of Sydney to maintain contact and keep you up-to-date with information about the University, its services, events and achievements. It may be passed on to groups affiliated with the University, such as alumni organisations and foundations (local and overseas), SU Sport and residential colleges. Your name may be published in the annual honour roll. If you wish to remain anonymous or do not wish to receive information, please contact Advancement Services (fax 9351 5688, or email: alumni.office@sydney.edu.au).The University abides by the "NSW Privacy and Personal Information Protection Act". ABN:15211513464 Charitable Fundraising No: 10369

ISS2011

LIGHT & MATTER

THE UNIVERSITY OF SYDNEY

LIGHT & MATTER

36TH PROFESSOR HARRY MESSEL INTERNATIONAL SCIENCE SCHOOL

2011