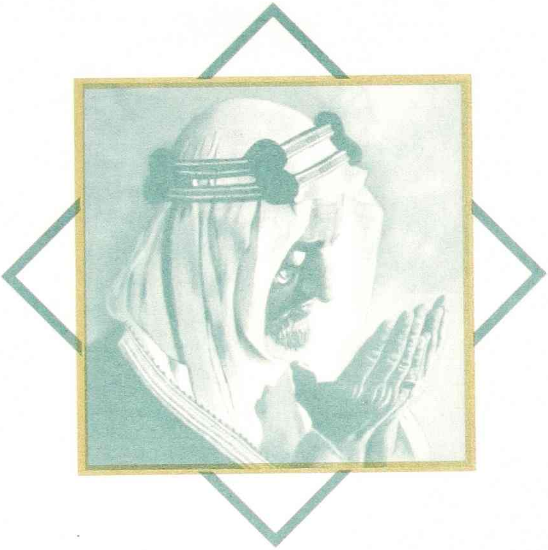
جائزة الملك فيصل العالمية
King Faisal International Prize

# ARTICLES IN MEDICINE AND SCIENCE V

## THE 2004 and 2005 KING FAISAL INTERNATIONAL PRIZE



# ARTICLES IN MEDICINE AND SCIENCE V

THE 2004 and 2005
KING FAISAL
INTERNATIONAL PRIZE

# ARTICLES IN
# MEDICINE AND SCIENCE V

## THE 2004 and 2005
## KING FAISAL
## INTERNATIONAL PRIZE

Custodian of the Two Holy Mosques
KING FAHD IBN ABD AL-AZIZ AL-SAUD
Patron of King Faisal Foundation

*Since its inception, Islam has stressed the importance of knowledge and thought; hence the great encouragement and honour that scholars in Muslim countries have enjoyed over the centuries. Therefore, when the King Faisal Foundation enhanced its activities by establishing the King Faisal International Prize, it was following a well-established Islamic tradition.*

*It is my hope that such activities spread throughout the Arab and Islamic worlds and that these countries unite in order to realize the highest scientific and intellectual objectives.*

Custodian of the Two Holy Mosques
King Fahd bin Abdul Aziz

(From King Fahd's address at the second annual ceremony of the King Faisal International Prize, 12 February 1980)

# Contents

# INTRODUCTION

The King Faisal Foundation continues the traditions of Arabic and Islamic philanthropy, as they were revitalized in modem times by King Faisal. The life and work of the late King Faisal bin Abd Al-Aziz, son of Saudi Arabia's founder and the Kingdom's third monarch, were commemorated by his eight sons through the establishment of the Foundation in 1976, the year following his death. Of the many philanthropic activities of the Foundation, the inception of King Faisal International Prizes for Medicine in 1981 and for Science in 1982 will be of particular interest to the reader of this book. These prizes were modeled on prizes for Service to Islam, Islamic Studies and Arabic Literature which were established in 1977. At present, the Prize in each of the five categories consists of a certificate summarizing the laureate's work that is hand-written in Diwani calligraphy; a commemorative 24-carat, 200 gram gold medal, uniquely cast for each Prize and bearing the likeness of the late King Faisal; and a cash endowment of SR750,000 (US$200,000). Co-winners in any category share the monetary award. The Prizes are awarded during a ceremony in Riyadh, Saudi Arabia, under the auspices of the Custodian of the Two Holy Mosques, the King of Saudi Arabia.

Nominations for the Prizes are accepted from academic institutions, research centers, professional organizations and other learned circles worldwide. After preselection by expert reviewers, the. Short-listed works are submitted for further, detailed evaluation by carefully selected international referees. Autonomous, international specialist selection committees are then convened at the headquarters of the King Faisal Foundation in Riyadh each year in January to make the final decisions. The selections are based solely on merit, earning the King Faisal International Prize the distinction of being among the most prestigious of international awards to physicians and scientists who have made exceptionally outstanding advances which benefit all of humanity.

(Excerpt from Introduction to 'Articles in Medicine and Science 1"
by H.R.H. Khaled Al Faisal,
Chairman of the Prize Board and
Director General of King Faisal Foundation)

## 2004 and 2005 Prize Awards in Medicine and Science

The 2004 Prize for Science (Biology) has been awarded to Professor Semir Zeki.

The 2005 awards were presented in April 2005

The Prize for Medicine (Topic: Tobacco Risks on Human Health) has been awarded to Professors Sir Richard Doll and Sir Richard Peto of the Clinical Trial Service Unit (CTSU) at Oxford University, U.K., for their pioneering and profoundly valuable epidemiologic research that has unequivocally established the link between tobacco and various diseases, such as vascular diseases and cancers, and has, in addition, served to propagate further research elucidating the molecular mechanisms of tobacco mediated cellular damage and DNA mutations. The impact of their studies has been so great has the impact of their studies been that several national health policies have been modified as a result of these findings. The WHO itself changed its position on smoking which culminated in a demonstrable decline in deaths related to cancer and atherosclerotic vascular diseases in several developed countries. Such significant benefits have transcended to large populations of developing countries as well, proferring an immeasurable contribution to mankind.

The Prize for Science (Physics) has been awarded jointly to Professors Federico Capasso (USA), Frank Wilczek (USA) and Anton Zielenger (Austria).

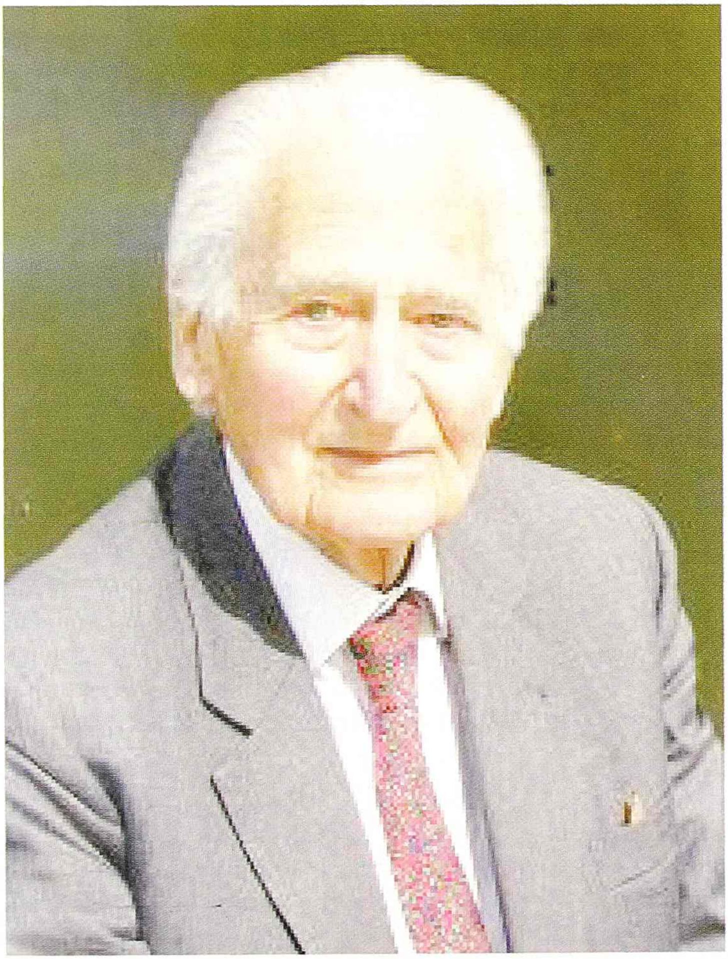
Professor Capasso of Harvard University is one of the most creative and influential physicists in the world having achieved international recognition through his design and demonstration of the Quantum Cascade Laser. This revolutionary approach, perhaps the most important development in laser physics during the last decade, signifies an imaginative breakthrough in this field enabling a remarkable contribution of excellent solid-state science and laser physics with new solid-state technology.

Professor Wilczek, a broadly accomplished and creative theoretical physicist, at Massachusetts Institute of Technology, has made a whole host of important contributions to several arenas. The most important of these has been the elucidation of Quantum Chromodynamics as the correct model for the Strong Force, one of the four known forces in nature. This groundbreaking work, alongside his other seminal achievements, elevates him to the ranks of the world's most prominent scientists.

With contributions ranging from epistemological and foundational research to the forefront of modern quantum technology, Professor Anton Zeilinger of the University of Vienna, has served and advanced mankind in both the cultural and technological domains. His impressive body of work includes that of applying the laws of quantum mechanics for the teleportation of the properties of a particle, heralded as a scientific milestone. In addition to this, he has successfully identified Quantum Cryptography as the only current method guaranteeing the confidentiality of a transmitted message as governed by natural laws.

# WINNERS OF THE 2005 KING FAISAL INTERNATIONAL PRIZE FOR MEDICINE

Medal: King Faisal International Prize for Medicine



**Professor Sir Richard Doll**
**Co-Winner**
**2005 King Faisal International Prize**
**(Medicine)**

## Synopsis of Achievements
PROFESSOR SIR RICHARD DOLL

### CAREER

Sir Richard Doll was born on 28th October 1912 in Hampton, Middlesex. He studied medicine at the University of London, qualified in 1937, and obtained his Membership of the Royal College of Physicians in 1939. After 6 years service in the Royal Army Medical Corps, as medical officer and physician in France, Cyprus, Egypt, and on a hospital ship, he started medical research, as an assistant to Dr Avery Jones. In 1948 he joined the Medical Research Council's Statistical Research Unit and became its Director 13 years later. In 1969 he was appointed Regius Professor of Medicine at the University of Oxford and held that position until 1979 when he became the first Warden of Green College, a new graduate college in Oxford with a special interest in clinical medicine. Since his retirement in 1983 he has worked as an honorary member of Sir Richard Peto's Epidemiology Studies Unit in the same University. In 1947 he married Dr Joan Faulkner (died 2001) and has two children.

### RESEARCH

Sir Richard's research has been principally in the field of epidemiology, seeking the causes of disease by observations on affected patients or on people exposed to different agents. His early research also included study of the effects of different treatments for gastric and duodenal ulcers by means of controlled trials with random allocation of treatments.

### SMOKING

His most important work began in 1948 when, with Professor Bradford Hill, he sought a reason for the enormous increase in the mortality from cancer of the lung that had occurred in men in the UK and, to a smaller extent elsewhere. With support from the Medical Research Council they were able to compare the characteristics and past histories of patients with lung cancer with those of other patients. The findings showed that by far the biggest difference between them was in the number of cigarettes that they had smoked. The difference could not be attributed to chance or to any bias in the design or conduct of the study and evidence from within the study combined with external evidence of the distribution of the disease over time, between men and women, and in different communities enabled them to conclude that smoking was the cause of the vast majority of all cases of the disease.

This conclusion was not widely accepted and Doll and Hill tested it by seeing whether knowledge of individuals' smoking habits would make it possible to predict their risks of lung cancer. They obtained information about the smoking habits of 34,000 male British doctors, followed them up and within 5 years found that the risks of lung cancer in men smoking different amounts were almost precisely those predicted from their earlier study.

By then, the study had also found that smoking might cause myocardial infarction and it was continued, with periodic updates of changes in smoking habits, for fifty years. The findings published at intervals, in conjunction with Sir Richard Peto, showed that more than 20 other diseases could be caused, in part, by smoking, that the total effect of consistent cigarette smoking from youth was to more than double mortality from all causes combined in middle and old age, that giving up smoking at any age reduced subsequent mortality, and that the earlier smoking was stopped the greater the reduction.

The study that Doll and Hill published in 1950 has subsequently been cited as a model for that type of research and the nature of the evidence that led them to conclude that the statistical association observed between smoking and lung cancer implied causality has become a standard method for determining causality from epidemiological studies.

## IONISING RADIATION

Doll's interest in the effect of ionising radiation was aroused when a test explosion of a hydrogen bomb in the Pacific caused radioactive fall-out throughout the northern hemisphere. Small doses were known to cause genetic mutations and hence possibly hereditary disease, but they were not thought to cause cancer. When, the government asked the Medical Research Council to assess the effects of nuclear radiation, the MRC asked Doll and Court Brown to try to determine the quantitative relationship between exposure to ionising radiation and the risk of leukaemia, the type of cancer that was most readily produced. They followed some 14,000 patients who had received radiotherapy for a benign condition and estimated the doses they had received from experiments on a model man. The risk of leukaemia was found to be proportional to the dose to the marrow, the organ from which the disease derives, and the size of the risk was estimated to be similar to that subsequently found by following up the survivors of the atomic bomb explosions over Hiroshima and Nagasaki.

Recent studies with Sarah Darby include a study of lung cancer in houses with different concentrations of naturally occurring radon in their atmosphere. This work, combined with the results of 12 similar studies in Europe, has provided a clear estimate of the risk. In conjunction with smoking radon in houses is estimated to cause some 9% of all lung cancers in Europe.

## OCCUPATIONAL HAZARDS OF LUNG CANCER

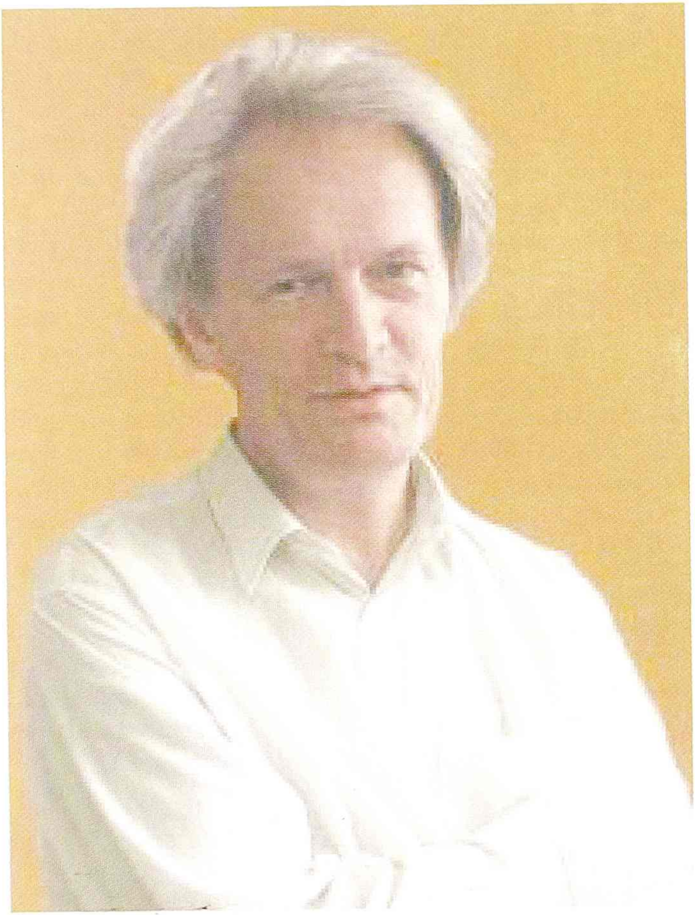
In 1955 Doll obtained the first firm evidence asbestos could cause lung cancer from observations on heavily exposed workers. Subsequent studies enabled him and Julian Peto to estimate the risks from the much lower exposures likely to be experienced in houses. Other occupational causes studied include the manufacture of coal gas and the refining of nickel.

## AVOIDABLE CAUSES OF CANCER

A notable contribution was the report Doll published, in collaboration with Sir Richard Peto, on the avoidable causes of cancer in the US. In the mid 1970s the belief had arisen that chemical pollution from industrial activity was the principal cause of cancer in that country and this was strengthened by the publication of a report under the auspices the Occupational Safety and Health administration and publicly accepted by the Secretary for Health, Education, and Welfare, stating that between 23 and 38% of all cancers were due to occupational exposure to just six industrial chemicals. This caused so much concern that the Office of Technology Assessment was asked to produce a report on what the causes of cancer actually were and Doll and Peto were asked to prepare it. Their report demonstrated the fallacies in the Department's estimate and showed that the causes included smoking, diet, microbiological infection, natural hormones, alcohol, ionising radiation, and sunlight. Occupation was a minor factor and chemical pollution of trivial importance. This report has subsequently become a reference book world-wide.

## HONOURS AND AWARDS

Doll was knighted in 1971 and made Companion of Honour in 1996. He was elected a Fellow of the Royal Society in 1966 and has received honorary degrees from 15 Universities, Other honours include the Royal Society's Royal Medal and the British Medical Association's Gold Medal, the United Nations Award for Cancer Research, the Mott Prize for Cancer Prevention , the Prince Mahidol Award, and the Shaw Prize in Medicine.



**Professor Sir Richard Peto**
**Co-Winner**
**2005 King Faisal International Prize**
**(Medicine)**

## Synopsis of Achievements

### PROFESSOR SIR RICHARD PETO

Professor Sir Richard Peto was born in 1943. He studied natural sciences and mathematics at Cambridge and then statistics at London University. After working for two years at the MRC Statistical Research Unit in London, he moved with Richard Doll in 1969 to Oxford where he became Lecturer (1972-75), then Reader (1975-92), and then Professor of Medical Statistics and Epidemiology. As well as continuing to collaborate with Richard Doll, Richard Peto is co-director (with Professor Rory Collins) of the Clinical Trial Service Unit (CTSU) at Oxford.

Professor Peto is one of the world's leading epidemiologists. His work has included studies of the causes of cancer in general, and of the effects of smoking in particular. He has also helped establish large-scale randomized trials of the treatment of cardiovascular diseases, cancer and other conditions and has been instrumental in introducing combined 'meta-analysis' of results from related clinical trials.

For more than 30 years, Professor Peto has worked (and continues to work) alongside Professor Doll on the detrimental effects of tobacco. Together, they are the best known and the most consistently productive tobacco epidemiologists in the world. Their scientific contribution to this field is matched by their ability to communicate their results with simple and effective messages that the public can understand.

In addition to his partnership with Doll in the cohort study on British doctors, Peto has initiated a series of very large studies of tobacco, blood pressure, obesity and death in China, India and other developing countries, including Cuba, Egypt and Mexico. These studies, which involved retrospective investigations of the smoking habits of more than a million dead people and interviews with more than two million people, have shown that tobacco is already causing even more deaths in developing than in developed countries and that the health risks of smoking will rise. Peto's landmark study with Alan Lopez (WHO, Geneva) concluded that about one billion people will be killed by tobacco this century if current smoking patterns persist. In recent years, Peto has extended his research to reveal the beneficial effects of smoking cessation. His ongoing international studies are having a major impact on health policies of nations.

Professor Peto has published more than 400 papers in leading scientific journals and conference proceedings. His scholarship has been widely recognized by numerous awards and prizes, honorary degrees, fellowships, visiting professorships, named lectures and memberships of academic institutions and learned societies both in the U.K. and abroad. He was elected a Fellow of the Royal Society of London (for introducing meta-analyses of trials, particularly in breast cancer treatment) in 1989, and was knighted by Queen Elizabeth in 1999 for his services to epidemiology and to cancer prevention.

# THE HAZARDS OF SMOKING AND THE BENEFITS OF STOPPING

**Richard Doll and Richard Peto**
Clinical Trial Service Unit &
Epidemiological Studies Unit
Harkness Building,
Radcliffe Infirmary, Oxford OX2 6HE
U.K.

## INTRODUCTION

The smoking of tobacco, particularly in the form of cigarettes, is now generally recognized to be an important cause of disease and to cause a substantially increased risk of death both in middle and in old age. What is not generally appreciated is how great the increased risk is and the extent to which it can be avoided by stopping smoking.

## SIZE OF RISK

The severity of the risk and the large number of diseases that smoking helps to cause are illustrated by the results of two large studies — one in the UK and one in the US — in which people with known smoking habits have been followed up and the mortality rates of different categories of smokers among them — regular smokers of different numbers of cigarettes, ex-smokers, and life-long non-smokers — are compared. The findings show that the mortality of continuing cigarette smokers is, on average, at least twice that of life-long non-smokers throughout middle and old age. In other words, for those who smoke cigarettes regularly smoking is, in adult life in the UK and the US, as hazardous as all other causes of death combined, resulting in 1 in 4 of continuing smokers dying because of their habit in middle age (that is, from 35 to 69 years of age), with those killed by tobacco losing on average 21 years of life, and 1 in 4 dying in old age, with those killed by tobacco losing an average of 8 years of life.

This enormous effect is not due to just one chemical or the production of just one disease, but to a wide range of different chemicals — some 4000 having been identified in cigarette smoke — and to a wide range of different diseases. Some 40 diseases are now known to be increased in incidence by smoking, varying from cancer of the lung, the incidence of which is increased so much that at one time over 90% of cases in the UK were due to smoking, through 15 other types of cancer, some of which are only occasionally caused by smoking, myocardial infarction, stroke, and gastric and duodenal ulcers, to the obvious and utterly miserable chronic obstructive lung disease (or chronic bronchitis and emphysema as it is often called).

Two Tables illustrate the findings for the diseases most closely related to smoking. The first shows the ratio of the mortality rates in continuous cigarette smokers to those of life-long non-smokers observed in the two studies to which we referred earlier. One study is of 34000 male British doctors who reported their smoking habits in 1951 and have been followed for 40 years reporting changes in their habits periodically throughout, (Doll *et al*, 1994). The other is of a million men and women who reported their smoking habits to the American Cancer Society and have been followed for 10 years (Thun, personal communication). The first two years observations in this study are, however, omitted to reduce the so-called 'healthy respondent effect': that is, the effect of excluding people already known to be ill at the start of a study. Women, it can be seen, died just like men from diseases caused by smoking when they had smoked as long as men, as they have now done in the US. The second Table, limited to data from the British study, shows, for seven of the same eight diseases (or groups of diseases), the differences in mortality between continuing smokers and ex-smokers and between continuing cigarette smokers with different daily consumptions of cigarettes. The rates are consistently higher in continuing smokers than in ex-smokers and in heavy smokers than in light smokers. For pulmonary heart disease we have had to show mortality rates rather than ratios for smokers compared with life-long non-smokers as there was no death from this disease in a life-long non-smoker.

The many other diseases caused in part by smoking are shown in Tables 3 and 4. Some are too rare or too seldom fatal in Western populations nowadays to have been detected in these follow-up studies of mortality, while some are only weakly related to smoking, so that very large numbers and evidence to exclude confounding have been needed, both of which are most easily obtained in specially designed case-control studies. Ten further types of cancer are listed in Table 3. Their relative importance varies from country to country depending on the background incidence of the disease, as smoking interacts with other causes producing, for example, about a 50% increase in the risk of cancer of the stomach irrespective of whether this disease is uncommon, as in the US, or very common, as in China.

Fifteen other diseases or groups of disease caused in part by smoking are listed in Table 4, together with four effects on reproductive health. Some are of great importance, because the background incidence is so high, and the outcome potentially so serious, as in the case of ischaemic heart disease and stroke. For some the relationship varies greatly with age, the risk of myocardial infarction, being, for example, increased by about 400 percent under 55 years of age, when the disease is rare among non-smokers, but by only about 20 per cent over 80 years of age, when the disease is common in non-smokers.

Much less evidence is available from these studies of the effects of smoking tobacco in other forms, as at the time the studies were conducted cigarettes had largely replaced other tobacco products in most economically developed countries. It is clear, however, that pipe and cigar smoking, as practised in these countries, were, relatively much less harmful. In our study of British doctors they caused only a fifth as much risk as cigarettes, (and cigarette smoking did not become common anywhere until the early years of the last century) so that goes a long way to explain why so little attention was paid to the effects of smoking until the middle of the century. For, the smoke from pipes and cigars is more irritating than the smoke from cigarettes and its different chemical constitution enables the nicotine in it to be absorbed from the mouth. Hence the smoke tends not to be inhaled and the noxious contents other than nicotine, not being carried into the lungs, are not absorbed and distributed throughout the body to anything like the same extent. The effects of pipe and cigar smoking were consequently seen principally in the mouth, pharynx, and oesophagus, where they are as capable of causing cancer as cigarette smoking is. Bidis, on the other hand, which are smoked predominantly in India and some other Eastern countries, are like hand-rolled cigarettes and are just as harmful. (Gajalakshmi *et al* 2003)

When all the effects of different methods of smoking are taken into account, Peto and his colleagues (2005) have estimated that thirty years ago smoking was responsible for 34% of all deaths in men in the UK. This enormous mortality could not simply be reduced by encouraging a switch from cigarettes back to cigars and pipes, because the ex-cigarette smoker, who has learnt to inhale, might well continue to do so despite the greater irritation that pipe and cigar smoke is likely to cause and the only really effective method of reducing the risk of smoking is by reducing the proportion of smokers who continue. Few people, however, realize how great a benefit cessation can achieve or that some benefit is obtained by stopping at any age, no matter how old. In fact, it is never too late to stop (given that, for example, lung cancer or some other potentially fatal disease has not already been induced); but the sooner it is stopped the greater the benefit.

## BENEFITS OF STOPPING

One set of observations that makes the benefits of stopping smoking very clear, at least for lung cancer, was obtained in a study aimed at assessing the effects of exposure to the radon that is present in the air in all ordinary buildings (Peto *et al*, 2000). The study was carried out in collaboration with Professor Sarah Darby in Devon and Cornwall, where the highest concentrations of radon in house air in the UK tend to occur. Detailed information had to be obtained about people's smoking habits, as these would nearly always cause a much higher risk than the radon and, unless taken into account, would obscure the effect of the radioactive gas that we were seeking to study. Interviews were obtained with nearly 1000 men and women with lung cancer and over 3000 controls, drawn from hospital patients suffering from diseases not caused by smoking and from a random sample of the general population, matched appropriately by sex and age groups.

The study was carried out in the early 1990s by which time a high proportion of men in the UK had stopped smoking for many years and we were able to estimate, from the results of this study and knowledge of the mortality from lung cancer recorded in the national mortality data, just what the risk of dying from lung cancer by 75 years of age would be in the UK if men and women continued to smoke or gave up at different ages. The results for men are shown in Figure 1. For life-long non-smokers the risk (in the absence of other causes of death), was about 0.4%; for men who stopped smoking about at 30 years of age, about 1.7%, for men who stopped at about 50 years of age, it was 6%, while for men who continued to smoke it was 16%. These figures, it should be noted, do not imply that only about 20% of the population are liable to develop the disease. On the contrary; twin studies have shown that hereditary factors are of relatively little importance in determining the susceptibility of smokers to lung cancer (Floderus *et* al, 1988; Kaprio and Koskenvuo, 1983; Carmelli and Page, 1996). Among smokers with similar patterns of cigarette consumption the difference between those who do and do not develop lung cancer is largely a matter of chance, depending on whether relevant mutations (which are constantly being caused by cigarette smoke in the stem cells of the bronchi) happen to be scattered across different cells or whether one of the stem cells happened, by chance, to accumulate enough mutations to change it into the seed of a growing cancer.

The effect of stopping smoking on the subsequent risk of lung cancer can be seen on a larger scale in the British national data. These show that the prevalence of smoking by men at ages 35-59 years has been progressively reduced since 1950, while it has increased in women of these ages until 1970, before beginning to fall *pari passu* with men (Figure 2) and that the trends in mortality have followed a few years later until the time came when men and women of these ages had been smoking cigarettes regularly since youth, as is shown for a slightly broader age group (35-69 years) in Figure 3.

## *Other diseases*

The effect of stopping smoking is not, of course, limited to the risk of lung cancer. We have less precise data for the effect of giving up at different ages on the risk of other diseases. It is seen, however, in Figure 4 which shows the relative risk of dying from some other diseases or groups of disease in relation to that in non-smokers for men who had given up for different lengths of time. The relative risks are shown for the cancers very strongly related to smoking (lung and larynx) for the other cancers strongly related to smoking (mouth, pharynx, oesophagus, pancreas, and bladder) for five cancers weakly related to smoking (nose, stomach, liver, kidney, and myeloid leukaemia) and for ischaemic heart disease. All decline progressively, but still remain slightly raised 20 years after stopping. For chronic obstructive lung disease (or chronic bronchitis and emphysema as it used to be called) the trend appears to be different. This, however, is due to the fact that some smokers with this horrible lung disease stop when disability gets severe, so that the mortality in the first 10 years after stopping is distorted by the inclusion of a number of men who stopped because they were already seriously ill and near to death. We know, however, from a study of transport workers (Fletcher *et al*, 1976; Fletcher and Peto, 1977) who were examined every six months over a period of 8 years that the rate of decline of lung function with age among smokers reverts, on average, to the slower rate of decline in non-smokers immediately smoking is stopped.

## *Total benefit*

The total benefit achieved by reducing the risk of all smoking attributable diseases is seen very clearly in the observations we made of British doctors, whose fate we have followed for 50 years (Doll *et al, 2004)*.

Those who stopped smoking by 35 years of age, in this population at an average age of 29 years, having smoked for not much more than 10 years, had a pattern of survival that did not differ significantly from that of life-long non-smokers. Those who continued smoking lost, on average. 10 years of life expectancy, but those who stopped at ages 60, 50, 40, and 30 years gained by doing so about 3, 6, 9, or the full 10 years respectively.

The effect of so many men stopping is seen clearly in the national data. The trend in the prevalence of smoking by men aged 35-69 years of age was shown in Figure 2 and the trend in the number of deaths at the ages attributed to smoking is shown in Figure 5. Between 1970 and 2000 the annual number of UK male deaths from smoking fell from about 70,000 to about 20,000 and is still falling, so it is now (2005) only about a quarter of what it was 40 years ago. That among females rose to 17,000 in 1985, since when it too has fallen, and is now (2005) about half of what it was 20 years ago.

## PUBLIC POLICY

The findings we have shown have mostly been obtained from observations on men and women in the UK and the USA. There is now, however, ample evidence that much the same overall hazards have been, or will be, observed in all other countries when men and women have been smoking substantial numbers of cigarettes for equally long. This is already clear, for example from studies in China (Lui *et al*, 1999) and India (Gajalakshmi *et al*, 2003) They should, therefore, have a major effect on public policy everywhere.

After the first lecture one of us ever gave on smoking and lung cancer, sometime in the early 1950s, there was a good discussion in the course of which someone said that, if confirmed, our findings with regard to lung cancer were clearly very important. It would, he said, be no good trying to persuade adult smokers to stop as they were already irreversibly addicted to nicotine. The one important thing, he insisted was to discourage young people from starting. On enquiring, this person turned out to be a representative of the tobacco industry. This remains, of course, the industry's policy today. Why? Because as long as adults smoke, children will want to do so too, to show that they are grown-up. It is, however, possible to get many adults to stop smoking. Among adults aged over 50 in Britain, two-thirds of the cigarette smokers have stopped, two-thirds of those who continue say, when interviewed, that they wish the hadn't started and would like to stop, and many of these will eventually succeed in doing so. Most can do it without great difficulty if they really want to. Those most addicted, however, do find it very difficult; but much can now be done to help them. Stopping smoking produces great benefits and, if it can be achieved when smokers are in their 30s, only little harm will have been done. It is, of course, better never to start and if children's role models set the example, there is a good chance that children never will. But adults have to set the example by not smoking themselves.

## ACKNOWLEDGEMENT

We are very grateful to our colleague Dr Jillian Boreham who has computed for us, over the last 10 years, many of the statistics we have cited and has prepared the Figures we have used.

## Table 1

## RATIO OF MORTALITY
## CIGARETTE SMOKER AND LIFE-LONG NON-SMOKERS:
## DOSEASES CLOSELY RELATED TO SMOKING

| Cause of death | | British doctors 1951-91 | US population 1984-91 | |
|---|---|---|---|---|
| | | men | men | women |
| Cancers of mouth, pharynx, and Larynx | (0.4)[*] | 24.0 | 11.4 | 6.9 |
| Cancer of oesophagus | (1.0) | 7.5 | 5.6 | 9.8 |
| Cancer of lung | (5.6) | 14.9 | 23.9 | 14.0 |
| Aortic aneurysm | (1.6) | 4.1 | 6.3 | 8.2 |
| Peripheral vascular disease | (0.1) | - | 9.7 | 5.7 |
| Chronic bronchitis & emphysema | (4.5) | 12.7 | 17.6 | 16.2 |
| Pulmonary heart disease | (0.3) | [**] | - | - |
| Peptic ulcer | (0.7) | 3.0 | 4.6 | 4.0 |

[*] Per cent of all deaths England & Wales, 1993

[**] No death reported in non-smokers

Table 2

DISEASES CLOSELY RELATED TO SMOKING:
EX-SMOKERS AND CURRENT SMOKERS BY AMOUNT

| Cause of death | Mortality compared to non-smokers | | | |
|---|---|---|---|---|
| | Ex- | Current smoking per day | | |
| | | any amount | 1-14 cigarettes | 25 or more cigarettes |
| Cancers of mouth, pharynx & larynx | 3.0 | 24.0 | 12.0 | 48.0 |
| Cancer of oesophagus | 4.8 | 7.5 | 4.3 | 11.3 |
| Cancer of lung | 4.1 | 14.9 | 7.5 | 25.4 |
| Aortic aneurysm | 2.2 | 4.1 | 2.5 | 5.4 |
| Chronic bronchitis & emphysema | 5.7 | 12.7 | 8.6 | 22.5 |
| Peptic ulcer | 1.5 | 3.0 | 1.4 | 4.5 |
| Pulmonary heart disease | (7)* | (10) | (5) | (21) |

*Mortality rate per 100,000 per year as the rate in non-smokers was zero and the ratios consequently infinite

## Table 3

## OTHER CANCERS RELATED TO SMOKING

Cancers of

lip
nasopharynx
liver
kidney
cervix
nose
stomach
pancreas
bladder
bone marrow*

*acute myeloid leukaemia

## Table 4

## OTHER DISEASES AND CONDITIONS RELATED TO SMOKING

Ischaemic heart disease

Hypertension

Myocardial degeneration

Other heart disease

Cerebrovascular disease

Arteriosclerosis

Pulmonary tuberculosis

Asthma

Pneumonia

Other respiratory disease

Crohn`s disease

Osteoporosis

Periodontitis

Tobacco amblyopia

Macular degeneration

and

Impotence

Reduced fecundity

Reduced fetal growth

Perinatal mortality

Capitations to figures

| | |
|---|---|
| Figure 1 | Cumulative risk of lung cancer by age 75 years at UK male rates 1990 for men who continued to smoke or stopped at different ages. |
| Figure 2 | Prevalence of smoking in UK 1950-2002: men and women aged<br>35-59 years. |
| Figure 3 | Mortality from lung cancer UK 1900-2002: males and females aged 35-69 years. |
| Figure 4 | Relative risk dying of chronic obstructive lung disease, ischaemic heart disease, and three categories of cancer in cigarette smokers who continued to smoke and those who had stopped for different periods of time compared to the risk in life-long non-smokers. |
| Figure 5 | Annual numbers of deaths attributable to cigarette smoking in males and females aged 35-69 years in the UK between years 1950 to 2000. |

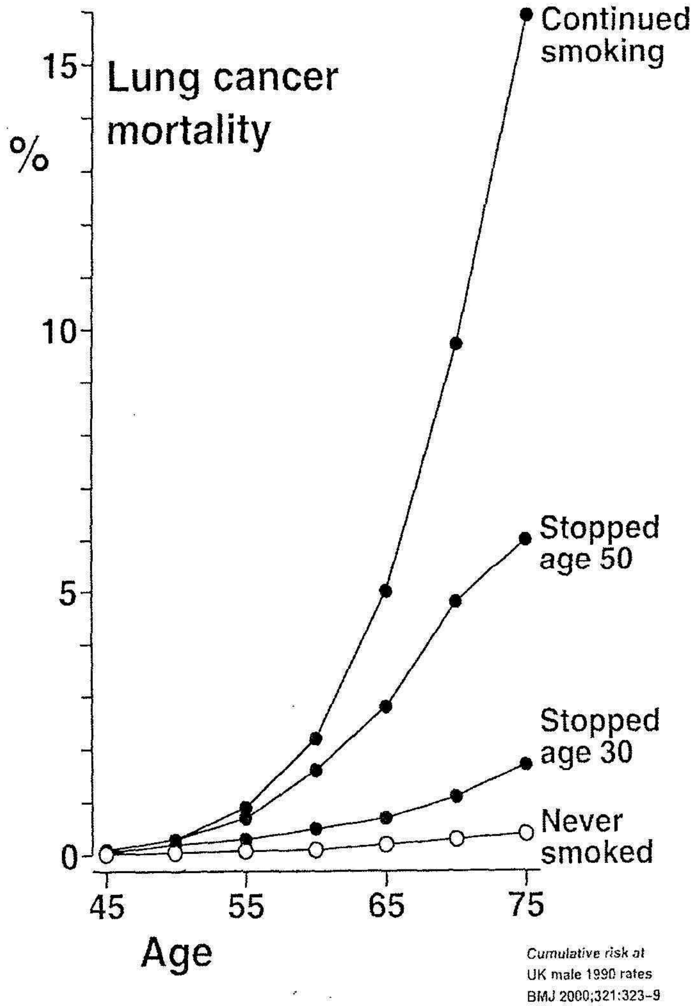

15
10
5
0
%
Lung cancer mortality
Continued smoking
Stopped age 50
Stopped age 30
Never smoked
45
55
65
75
Age
Cumulative risk at
UK male 1990 rates
BMJ 2000;321:323-9

Figure 1

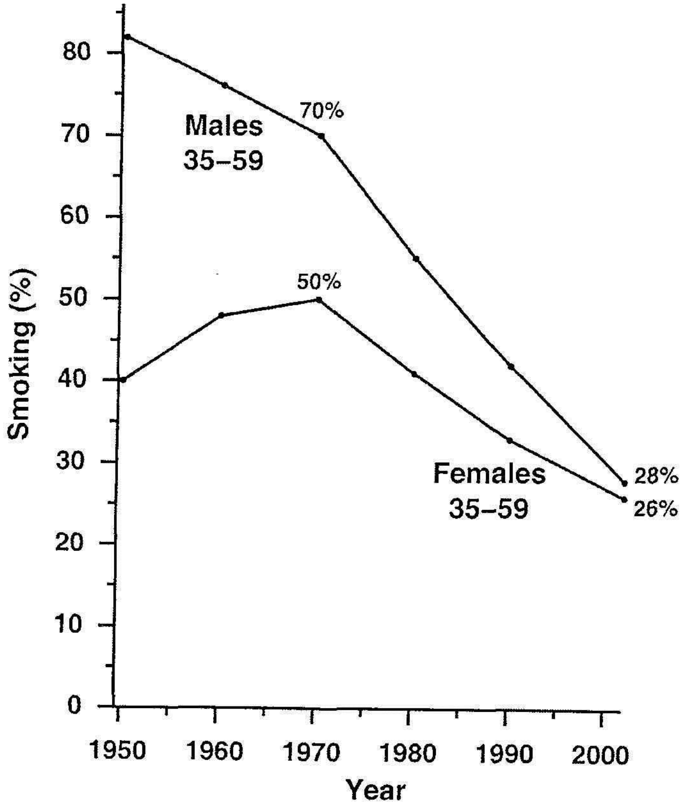
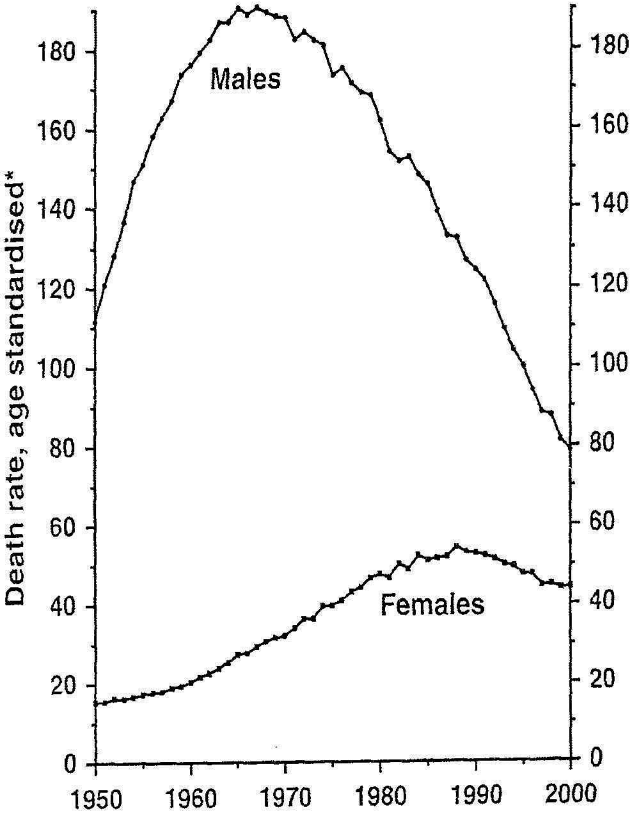

Smoking (%)
0
10
20
30
40
50
60
70
80
Males
35–59
70%
50%
Females
35–59
28%
26%
1950
1960
1970
1980
1990
2000
Year

Figure 2


Lung cancer mortality at ages 35–69

*Mean of annual rates per 100 000 men & women in component 5–year age groups

Source: WHO mortality & population estimates

Figure 3

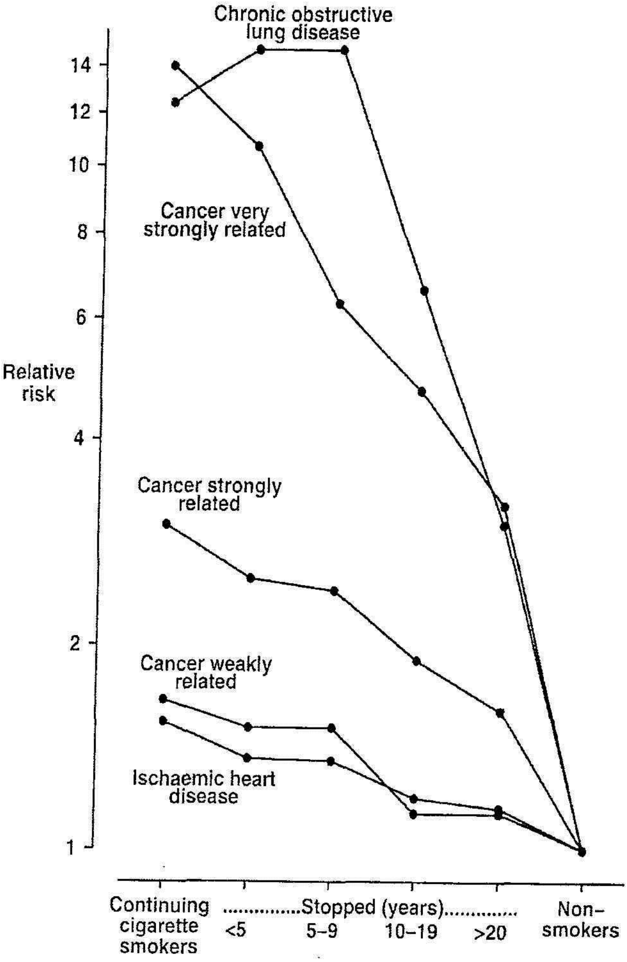

Chronic obstructive lung disease
Cancer very strongly related
Cancer strongly related
Cancer weakly related
Ischaemic heart disease
Relative risk
14
12
10
8
6
4
2
1
Continuing cigarette smokers
..............Stopped (years)..............
<5
5–9
10–19
>20
Non-smokers

Figure 4

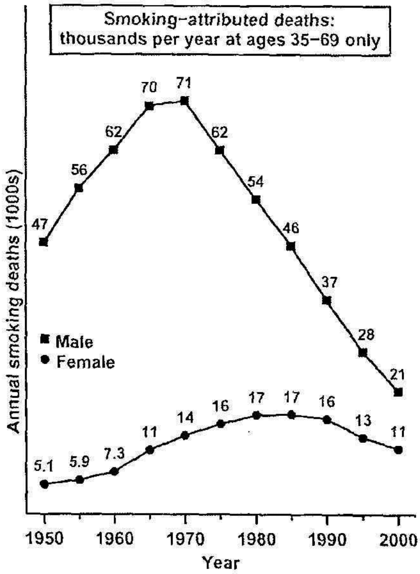
Smoking-attributed deaths:
thousands per year at ages 35-69 only
Annual smoking deaths (1000s)
■ Male
● Female
47
56
62
70
71
62
54
46
37
28
21
5.1
5.9
7.3
11
14
16
17
17
16
13
11
1950
1960
1970
1980
1990
2000
Year

Figure 5

## REFERENCES

CARMELLI D, PAGE WF. (1996) Twenty-four year mortality in World War II male veteran twins discordant for cigarette smoking. *Int J Epidemiol*, **25**, 554-559.

DOLL R, PETO R, WHEATLEY K, GRAY R, SUTHERLAND I.(1994) Mortality in relation to smoking: 40 years' observations on male British doctors. *Brit med J*, **309**, 901-911.

DOLL R, PETO R, BOREHAM J, SUTHERLAND I (2004) Mortality in relation to smoking 50 years' observation on male British doctors. *Brit Med J*, **319**, 1519-1528.

FLETCHER CM, PETO R. (1977) The natural history of chronic airflow obstruction. *Br med J*, **1**, 1645-1648.

FLETCHER C, PETO R, TINKER C and SPEIZER FE (1976). *The natural history of chronic bronchitis and emphysema.* Oxford University Press, Oxford.

FLODERUS B, CEDERLÖF R, FRIBERG L. (1988) Smoking and mortality: a 21 year follow-up based on the Swedish Twin Registry. *Int J Epidemiol*, **17**, 332-340.

GAJALAKSHMI V, PETO R. (2003) Smoking and mortality from tuberculosis and other diseases in India. Retrospective study of 43 000 adult male deaths and 35 000 controls.*Lancet*, **362**, 507-515.

KAPRIO J, KOSKENVUO M. (1985) Twins, smoking and mortality: a 12 year prospective study of smoking discordant pairs. *Soc Sci Med*, **29**, 1083-1089.

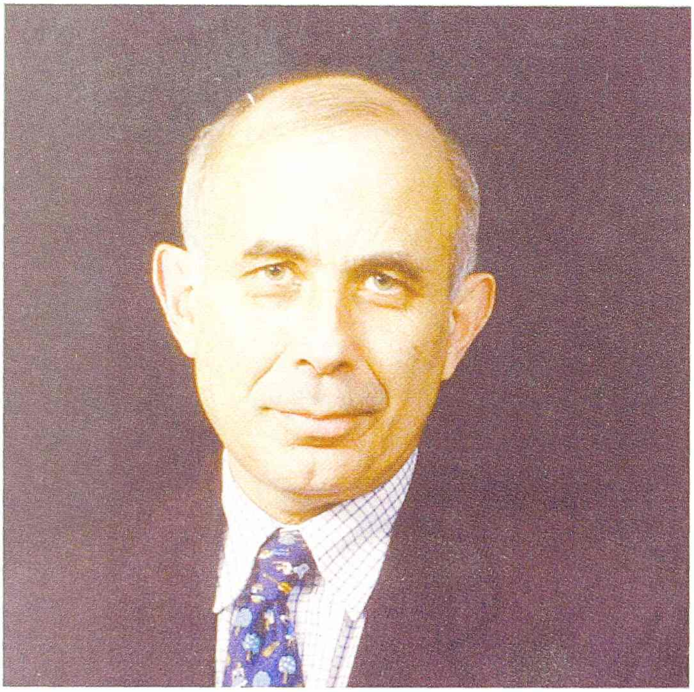
LIU BQ, PETO R, CHEN ZM, BOREHAM J, WU YP, LI JY, CAMPBELL TC, CHEN JS. (1998) Emerging hazards in China: proportional mortality studies of one million deaths. *Brit Med J*, **317**, 1411-1422

PETO R, DARBY S, DEO H, SILCOCKS P, WHITLEY E, DOLL R. (2000) Smoking, smoking cessation and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. *Brit Med J*, **321**, 323-329.

PETO R, LOPEZ AD, BOREHAM J, THUN M (2005) Mortality from smoking 1950-2000. In press Oxford University Press, Oxford

# WINNER OF THE 2004 and 2005 KING FAISAL INTERNATIONAL PRIZE FOR SCIENCE

Medal: King Faisal International Prize for Science



Professor Semir Zeki
Winner
2004 King Faisal International
(Science)

## Synopsis of Achievements
### PROFESSOR SEMIR ZEKI

Semir Zeki is one of the world's most respected authorities on how art is handled by the brain, as well as being an eminent Professor of Neurobiology and co-head of the Wellcome Department of Cognitive Neurology.

Professor Zeki's current area of research is the organisation of the primate visual brain from single cell physiology to higher cognitive problems. The research programme encompasses basic physiological and anatomical methods, as well as brain imaging techniques. One of his books, Inner Vision, even proposes a theory about the neuropbysiology of love.

"Vision," Professor Zeki argues, "is an active process that depends as much upon the operations of the brain as upon the external, physical environment."

According to Professor Zeki, the brain is selective about visual knowledge, comparing what it selects with a stored record of all that it has seen.

A lifetime's work on the visual brain has also led to an interest in art, about which he has published several articles and also a book, written jointly with the French painter, Balthus, and *entitled La Quête de l'Essentiel.*

In 1973, Professor Zeki identified a separate area in the brain called V4, which was full of cells that crackled with activity when exposed to different colours. This has been acclaimed as a keystone in this area of research and led to Professor Zeki being described as 'one of the founders of modem science,' (V Ramachandran, University of Califomia)

In 1985 he won the first Golden Brain Award, a prize set up by the Minerva Foundation, based in Berkeley, California, to honour original discoveries in vision and brain research. Other honours include the Prix Science pour L'Art, LVMH Paris (1991), Rank Prize (1992), Zotterman Prize, Swedish Physiol Society (1993) and the Electronic Imaging Award (2002).

Born in 1940 of Lebanese origin, Professor Zeki started his career by studying anthropology at University College London, then switched to medicine before settling down to a career in neurobiology and obtaining his PhD in anatomy from University College in 1967. He did his postdoctoral work at St. Elizabeth's Hospital, Washington DC, and spent a further year in the United States as Assistant Professor at the University of Wisconsin. He returned to London in 1969 lecturing in Anatomy at UCL until 1975.

Professor Zeki became professor in the Department of Anatomy and Embryology in 1981. He is now co-head of the Wellcome Department of Cognitive Neurology. He was elected to the Royal Society's Fellowship in 1990 and has lectured in universities around Europe throughout the 1980s and 1990s.

Positions held: Fellow of the Royal Society, Professor of Neurobiology, University College London, since 1981, Fellow of the Institute of Neuroscience, New York, since 1985, Member of the Board of Scientific Governors, Scripps Scientific Research Institute, California, since 1992, Member of the National Science Council of France, since 1998, Co-head of Wellcome Department qf Cognitive Neurology, 1996-2001, Henry Head Resident Fellow, Royal Society 1975-1980, Fellow of the Academy of Medical Sciences, Foreign Member of the American Philosophical Society, 1998, Member of the Academia Europaea, 1990, Member of the European Academy of Sciences and Arts, 1992

# The Organization of the Visual Brain and Visual Consciousness

**Semir Zeki**
Professor of Neurobiology
Dept. of Anatomy & Developmental Biology
University College London
Gower Street, London WC1E 6BT
U.K.

## *General organization of the visual brain*

The visual brain constitutes roughly about one quarter of our entire brain. The best studied part of it lies at the very back of the brain and is known as the primary visual cortex or area V1 (figure 1). V1 receives the visual signals from the retina and damage to it causes irreversible blindness.

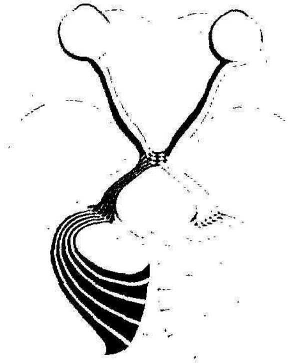


Figure 1: Schematic representation of the connections from the retina of the eye to the primary visual cortex, situated at the back of the brain

Surrounding area V1 is a large cortical zone which used to be known as "association" cortex. Studies over the past thirty years have shown that this cortex consists of several visual areas (Figure 2). These have been numbered V1, V2, V3...etc., to reflect their distance from V1. Each one receives its signals from V1, but each one receives a different set of signals. V1 thus acts like a post office, pigeon-holing and distributing signals to different areas lying in the cortex surrounding it.

Because these different visual areas receive different signals from V1, they are specialized to process different attributes of the visual scene, such as motion, colour, form and so on. There is, in other words, a functional specialization in the visual brain (Zeki 1978).

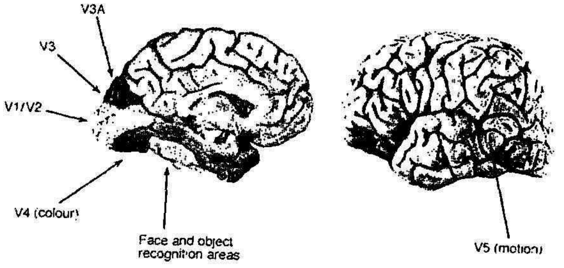


Figure 2: Some of the visual areas in the human brain

We may summarise this by saying that the visual brain consists of many, functionally specialized areas, each one of which has a different task. A consequence is that the brain does not analyse all the visual information that it receives in the same cortical area. Instead it does so simultaneously, and in parallel, in several different areas. As work progresses, we are beginning to discover more and more specialized areas in the visual brain. Parallel processing has now been shown to be an essential characteristic not only of the visual brain but of the brain in general.

## *Colour and Motion*

Of all the visual attributes, it is the study of colour and motion that has given us the best clues into the organization of the visual brain and its functioning. The cortical areas that are specialized for these two attributes are well separated from each other geographically (figure 3). They also receive easily distinguishable anatomical inputs. Moreover, the study of colour has contributed greatly to our understanding of how the brain, far from being a mere passive chronicler of external events, is an active participant in creating what we see.

Light is electromagnetic radiation and is colourless. It is what the brain does to the lights of different wavebands that, in a sense, creates colour. Colour may be said to be an interpretation, a visual language, that the brain gives to certain combinations of wavelength composition that the eye receives. This was recognized many years ago by Isaac Newton, who wrote in 1704, "For the Rays, to speak properly, have no Colour. In them there is nothing else than a certain power and disposition to stir up a sensation of this Colour or that". In very general terms, the brain does this by comparing the wavelength composition of the light (in terms of the amount of red, green and blue light) reflected from one surface and that reflected from surrounding surfaces. Each surface has a unique and immutable efficiency for reflecting lights of different wavelengths. This property, known as reflectance, is the percentage of light of a given waveband (eg red) reflected from it, as a function of the amount of light of that same wavelength incident upon it. This is a property that never changes. By contrast, the actual amount of light of any waveband incident upon a surface and reflected from it changes from moment to moment. The light in which surfaces are viewed can be more intense in one set of conditions than in another, or its wavelength composition can change (there is, for example, more red light at dawn and at dusk than there is at noon on a cloudy or sunny day). Such variations do not affect our colour perception to any substantial degree, because different surfaces always reflect the same percentage of light of any given waveband. By comparing the efficiencies of different surfaces for reflecting light of different wavebands, the brain is able to tell which surface is more efficient for reflecting, for example, red light. It then attaches a unique property – colour – to that characteristic. It is for this reason that the colours of surfaces do not change (except in shade) with quite large changes in the wavelength composition of the light in which they are viewed. A leaf, for example, appears green whether viewed at noon on a cloudy or sunny day or when viewed at dusk or dawn. Scientists had speculated that this remarkable ability of the brain depends upon learning, judgment and memory of what colour a surface should have (Helmholtz 1911, Hering 1877). We now know that it is done computationally (Land 1974) and automatically within area V4 (Zeki 1980). When V4 is damaged, such a computation becomes impossible and hence the patient becomes colour blind, even though the rest of the visual apparatus is intact (Zeki 1991).

## *Functional specialization in the human visual brain*

The development of techniques for imaging activity in the human brain has allowed us to locate these visual areas with precision. When cell in an area are especially active, their metabolic requirement increases and this necessitates an increase in blood flow to the relevant area. The principle on which brain imaging techniques work is to detect the change in blood flow to an area.

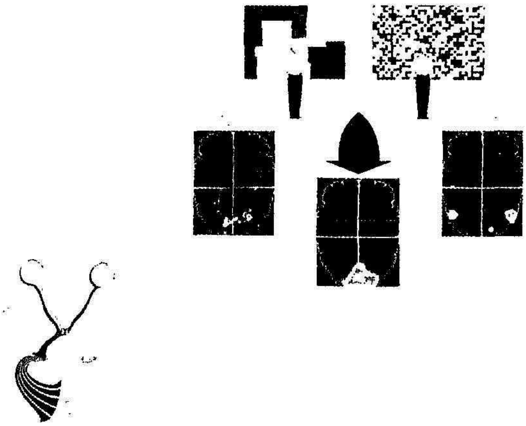


Figure 3: A simple experiment to demonstrate functional specialization in the human brain. When humans view a multicoloured abstract scene (above left), brain activity (shown in white and lighter shades of grey on brain slices) is restricted to the primary visual cortex (V1, centre panel) which lies at the back of the brain, where the input from the retina terminates (see inset at lower left). Activity also involves an area lying in front of it, area V4 (left panel). When they view a pattern in motion (above right) the activity is in V1 (centre panel) but in another region of the cortex surrounding it, area V5 (right panel).

Among the areas that have been charted and intensively studied using brain imaging are V4 and V5 (Zeki et al., 1991) (Figure 3). Area V4 is specialized for colour and V5 for visual motion. The specialization of these two areas for two different attributes of the visual scene explains two medical syndromes of cerebral origin, namely a specific incapacity to see colour or visual motion following cortical damage that is restricted to either of these areas. Such syndromes are quite different from the one which results after damage to V1. Because V1 receives all the visual signals before sending them on to the specialized visual areas, damage to it results in total blindness. By contrast, damage to the specialized areas results in an incapacity to perceive the visual attribute for which they are specialized and not to total blindness. Patients who, through a stroke or some other injury, have sustained damage to the colour centre (V4) are unable to see the world in colour but only in black and white or "dirty" shades of grey. But their capacity to perceive other attributes of vision is not affected. They can, for example, see visual motion or shapes. By contrast, patients with damage to V5 (Figure 4) become specifically incapable of seeing objects when in motion (Zihl 1981; Zeki 1991. They can however tell their colour and their shape, but only if the objects are stationary. There are other specialization in the visual brain. There is, for example, an area that, when damaged, leads to an incapacity to recognize people through their faces and there are areas that have been strongly implicated in the perception of objects.

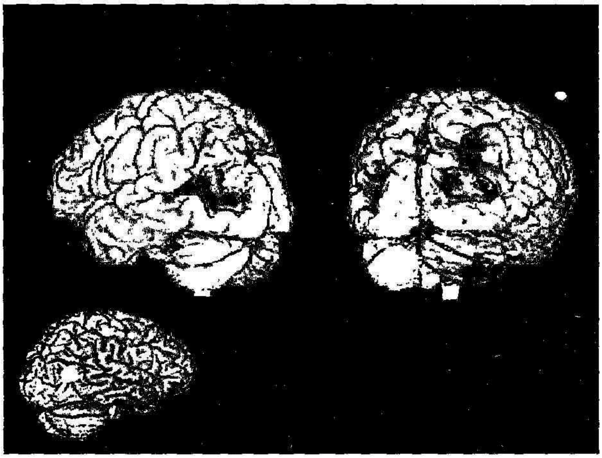


Figure 4: Damage to area V5 as in this patient leads to a condition in which the patient is incapable of seeing any visual stimulus when it is in motion but can only perceive them when they are stationary. Inset at lower left shows the position of area V5 in the normal human brain.

## *The hierarchy of visual perception*

Because we seemingly see all visual attributes at the same time (for example we see the colour, shape, and direction of motion of a bus simultaneously), we have tended to assume that all visual areas terminate their work at exactly the same time. Detailed experiments show that this is not so (Moutoussis and Zeki 1997; Zeki and Moutoussis 1997; Arnold et al., 2001). In fact, we see colour before form, and form before motion. The perceptual advantage that colour has over motion is between 80-100 milliseconds. Given that there are 1000 milliseconds in a second, this may not seem much. But in terms of neural time, it is a huge difference. It takes about 1 millisecond for the nervous impulse to cross from one nerve cell to the next, which makes of an 80 millisecond difference a very large one. This demonstration of an asynchrony in visual perception leads to an important conclusion, namely that different visual areas terminate their perceptual tasks at different times and hence that there is a perceptual hierarchy in vision. Such differences in perceptual times have strange consequences. One can present humans with a continuous visual motion and colour task, a task in which the colour and the direction of motion of a stimulus, say a spot, changes. If one then asks them to determine at any given moment the direction of motion and the colour of the stimulus, one finds that at that at any given moment (measured in blocks of 100 milliseconds) they ascribe the correct colour to the direction of motion that had occurred during the previous 80 milliseconds, and therefore the "wrong" direction of motion. This is because, at the chosen moment, the brain had completed working out (processing) the colour signals but had not yet had time to complete the processing of visual motion signals.

### *Processing sites in the visual brain are also perceptual sites*

The above demonstration leads one to suspect that there is no single perceptual site in the brain into which all the visual areas report the results of their work. A better alternative would be to test the proposition that a processing site in the visual brain is also a perceptual site. In somewhat complex experiments we presented human subjects with the same stimuli in two conditions. In one condition, they could perceive the stimulus correctly; in the other, they could not perceive the stimulus, although the same visual signals as in the first condition were still entering the brain. We then imaged the activity in the brain of the subjects under these two conditions. The results showed that, whether the subjects perceived the stimulus correctly or whether they reported seeing nothing (even though signals from the stimulus had reached their eyes and entered their brain), the same specific areas of the visual brain were active (Moutoussis and Zeki 2003). The difference between the two is that, when a stimulus is processed without being consciously perceived, the activity in the relevant brain area is much weaker than when it is (consciously) perceived. This leads us to conclude that a processing site in the brain is also a perceptual site. There are, therefore, many different perceptual sites in the visual brain, not just one as we had previously supposed. It leads us to a further conclusion, derived from imaging experiments and the study of patients with visual disorders, namely that a processing site acquires a conscious correlate and hence becomes a perceptual site through a local increase, within it, in the strength of activity, without necessarily engaging other cortical visual areas (Zeki and ffytche 1998).

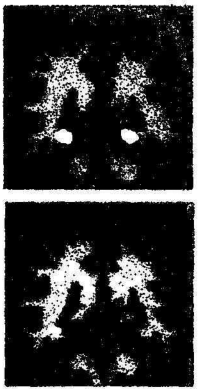


Figure 5: Activity in the brain of a human subject when (above) they perceive consciously a stimulus (in this example a face) and when the input to their eyes from the same stimulus is the same but arranged in such a way that they cannot perceive the face (below). Brain activity shown as white blobs on these slices through the brain is much more intense in the former (above) than in the latter instance.

## *The nature of visual consciousness*

The above experiments naturally lead us to enquire into the nature of visual consciousness. In the past, we have all though of consciousness as a unified entity. But if we become conscious of different attributes of vision such as colour and motion because of activity in different parts of the brain, then it follows that there are many sites in the brain for consciousness. If, moreover, we become conscious of different attributes of vision at different times, it follows that the activity in different parts of the brain reach a conscious state at different times. It follows from this that there isn't a single unified consciousness. There are instead many micro-consciousnesses, each one due to activity in a different part of the visual brain (Zeki 2003).

## *The Unity of consciousness*

This leaves us with the problem of the unified consciousness. The German philosopher Immanuel Kant (1781) spoke of the unified consciousness in specific terms. He referred to it as the "transcendental consciousness" and thought of it as "the consciousness of myself as original apperception", that is the source of all thought. It is, I think, in this sense and this alone that we can speak of a unified consciousness. Moreover, a unified consciousness in this sense is made possible only through the use of language and communication, the former being a uniquely human faculty. In simpler terms, animals are aware but only humans are both aware and aware of being aware.

The study of the visual brain has thus provided us with rich insights, not only into the functional organization of a relatively large part of our brain, but also into the mysteries of seeing, of specific blindnesses produced by specific lesions, and into the nature of consciousness itself.

## References:

Arnold, DH et al. (2001). Asynchronous processing in vision: color leads motion, *Curr. Biol.*, **11**, 596-600

Helmoholtz, H. von (1911). *Handbuch der Physiologischen Optik*, Leopold Voss, Hamburg

Hering, E. (1877/1964) *Outlines of a Theory of the Light Sense*, Translated by LM Hurvich and D. Jamieson, Harvard University Press, Cambridge.

Kant, I. (1781). *Critique of Pure Reason*, Translated into English by WS Pluhar (1996), Hackett, Indianapolis

Land, EH (1974). The retinex theory of colour vision, *Proc. R. Inst. Gt. Brit.*, **47**, 23-58.

Moutoussis, K. and Zeki, S. (1997). A direct demonstration of perceptual asynchrony in vision, *Proc. R. Soc. Lond. B.*, **264**, 393-399

Moutoussis, K. and Zeki, S. (2002). The relationship between cortical activation and perception investigated with invisible stimuli, Proc. Natl. Acad. Sci. USA, **99**, 9527-9532.

Newton, I. (1704) *Opticks*, Dover Press, New York

Zeki, S. (1978). Functional specialization in the visual cortex of the rhesus monkey, *Nature*, **274**, 423-428

Zeki, S. (2003). The disunity of consciousness, *Trends in Cogn. Sci.*, **7**, 214-218.

Zeki, S. et al. (1991). A direct demonstration of functional specialization in human visual cortex, *J. Neurosci.*, **11**, 641-649.

Zeki, S. and Moutoussis, K. (1997). temporal hierarchy of the visual perceptive systems in the Mondrian world, *Proc. R. Soc. Lond. B.*, **264**, 1415-1419.

**Professor Federico Capasso**
**Co-Winner**
**2005 King Faisal International Prize**
**(Science)**

## Synopsis of Achievements
PROFESSOR FEDERICO CAPASSO

Federico Capasso (b. 1949) is the Robert Wallace of Applied Physics and Vinton Hayes Senior Research Fellow in Electrical Engineering at Harvard University, which he joined at the beginning of 2003. He holds a Doctorate in Physics, summa cum laude, from the University of Rome, Italy (1973) and an Honorary Doctorate in Electronic Engineering from the University of Bologna (2003).

After teaching physics for six months in a Rome high school he held a research appointment from 1974 to 1976 at Fondazione Ugo Bordoni in Rome. From 1976 to 1977 he was a postdoctoral fellow at Bell Laboratories, Holmdel, NJ. He was then hired as a regular Member of Technical Staff at Bell Labs, Murray Hill, NJ, and subsequently promoted to Distinguished Member of Technical Staff. In 1986 he became head of the newly formed Quantum Phenomena and Device Research Department. In 1997 he was named head of Semiconductor Physics Research and a Bell Labs Fellow for his scientific contributions. From 2000 to the end 2002 he was Vice President for Physical Research at Bell Laboratories, Lucent Technologies.

Professor Capasso's research has cut across several disciplines in applied and basic physics and electrical engineering. These include quantum electronics and optics, semiconductor physics, mesoscopic physics, solid-state electronics, optoelectronics and micromechanics. A unifying theme of his research has been the quantum design and study of new artificial materials and nanostructures with man-made electronic and optical properties, an approach that Professor Capasso pioneered and dubbed bandstructure engineering. These studies include both the investigation of quantum effects in lower dimensionality systems and the invention of photonic and electronic devices in which quantum effects on a mesoscopic scale (a few to ~ 100 nm) play a dominant role. An excellent example of this research is the invention by Capasso and his collaborators of the quantum cascade (QC) laser, a fundamentally new light source whose emission wavelength can be designed to cover the entire spectrum from the mid to the far infrared by tailoring the active region layer thickness. These lasers are revolutionizing science and engineering in the mid-infrared and far-infrared spectrum. They are finding widespread use in scientific and industrial applications: high-resolution spectroscopy, chemical sensing and trace gas analysis, atmospheric chemistry, combustion and medical diagnostics. New transistors and photodetectors pioneered by Capasso have found applications in telecommunications and have opened up new directions of research in photonics.

In recent years Dr. Capasso's research has branched off to the study of the quantum electrodynamical (QED) phenomenon known as the Casimir effect, i.e. the attractive force between uncharged metals associated with quantum fluctuations of vacuum. This effort includes the high precision measurement of Casimir forces using MEMS (Micro ElectroMechanicalSystems). Capasso and his group are also focusing on designing geometries that will alter in non-trivial ways the Casimir force and on the search for the new QED phenomena such as the dynamic Casimir effect. Other current activities include the investigation of attractive radiative forces between small optical components as well the study of nanowire lasers and spintronic devices.

He has co-authored over 300 papers, edited four volumes, and holds 54 US patents.

He is a member of the National Academy of Sciences, the National Academy of Engineering, the American Academy of Arts and Sciences and an honorary member of the Franklin Institute.

His awards include the American Physical Society (APS) Arthur Schawlow Prize in Laser Science, the Institute of Electrical and Electronics Engineers (IEEE) Edison Medal, the Wetherill Medal of the Franklin Institute, the Wood Prize of the Optical Society of America (OSA), the Duddell Medal of the Institute of Physics (UK), the Materials Research Society Medal, the William Streifer Award of the IEEE Laser and Electro-optic Society, the Rank Prize in Optoelectronics (UK), the IEEE David Sarnoff Award in Electronics, the Willis Lamb Medal for Laser Science and Quantum Optics, the "Vinci of Excellence" Prize (France), the Welker Memorial Medal (Germany), the New York Academy of Sciences Award, the Newcomb Cleveland Prize of the American Association for the Advancement of Science. He is a Fellow of OSA, APS, IEEE, SPIE and AAAS.

# Adventures of a Quantum Designer: the Birth of the Quantum Cascade Laser

**Federico Capasso**
Robert Wallace Professor of Applied Physics and Vinton Hayes Senior Research Fellow in Electrical Engineering
Division of Engineering and Applied Sciences
Harvard University
Cambridge MA
USA

## The birth of a fundamentally new laser at Bell Laboratories

Semiconductor lasers have entered our daily lives in many applications ranging from compact disc players, laser printers and fax machines, laser pointers and last but not least as the most important light source for fiber optic communications, sending on land and across the oceans for thousand of kilometers up to 10 billions of bits per second (Gb/s) of information on a single color (or wavelength) of light. As such they have been at the center of the communications revolution. Semiconductors, with electrical properties intermediate between those of metals and insulators (which carry no electricity), are the materials that have probably had the greatest societal impact by shaping the information age. Suffice to say that all computer chips are made of Silicon, a non-light emitting semiconductor, while all semiconductor lasers are made of so-called compound semiconductors consisting of various elements of the periodic table. For example the light -emitting semiconductor used in all lasers for long-distance communications is an alloy made of four chemical elements: Indium, Gallium, Arsenic and Phosphorous. The central role of semiconductors in revolutionizing computers and communications was recognized by the awarding of part 2000 Nobel prize in physics to Jack Sinclair Kilby for his realization of the first integrated circuit and to Zhores Alferov and Herbert Kroemer for their role in the invention and demonstration of diode lasers used in modern communications.

A semiconductor laser comprises an active region sandwiched between semiconductor layers of opposite electrical polarity capable of supplying an electric current in the form of negative charges (electrons) and positive charges, known as holes, which flow into the active region when a suitable external voltage is applied to the laser via wires. Electrons and holes of course attract each other and in the process lose energy neutralizing each other and generating light. Each electron hole pair generates a photon of energy equal to the energy difference between that of injected electrons and holes, a quantity known as the energy bandgap, which is uniquely determined by the choice of the semiconductor of the active region. The latter therefore also determines the emission wavelength because the energy of a photon is equal to the fundamental constant of quantum mechanics (Planck's constant) times the photon frequency (frequency = speed of light divided by the wavelength). The external layers of the semiconductor laser sandwich have the important function of not only supplying electrons and holes but also of spatially confining them and the photons they generate to the active region. This allows the efficient neutralization of electrons and holes and the guiding of laser light along the active region to the external facets of the laser bar (of length of the order of one third of a millimiter). These semitransparent surfaces, defined by cleaving the semiconductor along crystalline planes perpendicular to the layers, serve the function of both laser windows, by allowing the external collection of the light, and of laser mirrors, by providing the optical feedback required for laser action. The latter works as follows: a fraction of the light generated in the active region is reflected back into the active region where it causes more electrons and holes to neutralize and generate photons by a process, first postulated by Einstein in 1917 and known as stimulated emission. This is a multiplicative and regenerative process since each photon in turn stimulates the emission of another one and the mirrors (facets) partially reflect the light back and forth thus allowing its amplification, which is central to the laser concept (LASER an acronym that stands for **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation).

Semiconductor lasers of the type described above, which goes under the name of diode laser, have demonstrated high performance and have been successfully commercialized in the wavelength range from the red to the near infrared (1.3 to 1.6 μm) employed in optical communications, using a variety of semiconductors alloys with different bandgaps to achieve the desired wavelength. However, their operation has been much less successful in the mid-infrared spectrum due to the lack of suitable materials with the required characteristics (reliability, ease of fabrication and insensitivity to temperature recycling).

The scientific and technological interest of the mid-infrared region of the light spectrum, often called the molecular fingerprint region, which covers the range of invisible light from approximately 2 to 20 μm wavelength, stems from the fact that gases and vapors have their tell-tale absorption features associated with vibro-rotational transitions of their constituent molecules in this part of the spectrum. These absorption lines can be probed by a number of spectroscopic techniques in a wide range of scientific, industrial and military applications centered on chemical sensing that will be discussed in a subsequent section.

While mid-infrared semiconductor diode lasers made of materials known lead salts have been commercially available for quite some time, they have very limited power (at most a few mW of peak and continuous wave power), a small continuous single mode tuning range and have not yet been operated at room temperature (RT), an essential requirement for widespread commercial penetration.[1] They also suffer from spectral degradation and severe reliability problems associated with thermal re-cycling that have greatly limited their use and commercialization.

The Quantum Cascade (QC) laser is a semiconductor laser which is fundamentally different from diode lasers in several respects which will be discussed in depth in the next section, its most revolutionary one being that unlike other all other lasers (with the exception of the so-called free electron laser) its wavelength is not determined by the chemical composition of the active region but by the thickness of nanometer thick layers contained in the latter. By controlled changes of that thickness the wavelength can be selected in a huge wavelength range [2-5]

The QC laser is a compact relatively high-power light source that can be designed to emit essentially any wavelength in the mid-infrared spectrum and in a large portion of the infrared spectrum form wavelength ~ 50μm as long as 150μm.[6] A decade after their invention QC lasers have therefore fulfilled and exceeded their initial promise,[7] greatly exceeding in performance mid-ir diode lasers and reaching wavelengths well beyond those of conventional semiconductor lasers. They have become commercially available, starting to fill the increasing need for compact and high performance mid-infrared semiconductor lasers, enabling also a wide range of new applications in this spectral region of paramount scientific and technical importance. There are presently over 45 groups worldwide working on QC lasers. Companies that are developing and or commercializing QC lasers or QC laser based sensor include: Thales Inc., Agilent Technologies, Alpes Lasers, Maxion, Teradyne, Physical Sciences Inc. to mention just a few.

QC lasers nowadays operate pulsed at RT with peak optical powers of up to hundreds of mW concentrated in a single laser mode and can be continuously tuned with currents over a significant wavelength range. Peak power levels of a few Watts in pulsed mode can be easily achieved with QC lasers and about 0.5 W of output power has been demonstrated in devices operating in continuous wave. [8] QC lasers have been successfully used in trace gas analysis with parts-per-billion in volume (ppbv) sensitivity and in portable QC-laser-based sensors opening up new market opportunities.[9,10] Figure 1 is a demonstration of a working high-power mid-infrared QC laser.

The invention and demonstration of the quantum cascade laser (QCL) [11] by my group occurred at Bell Labs in January of 1994, culminating a fifteen years effort, aimed at designing, growing, and testing artificially structured semiconductor materials and devices of ever increasing complexity and functionality.[12]

I had arrived at Bell Laboratories, then the research branch of AT&T, at the end of 1976 on leave from a leading telecommunication laboratory (Fondazione Bordoni) in Rome, Italy, where I had been working for two years on fiber optics, following a semester teaching in a high school, immediately after finishing my Ph.D. in Physics at the University of Rome in 1973. At that time Bell Labs was still in its "golden age", an amazing "caldron" where exciting discoveries and inventions were made on an almost daily basis, in a highly interdisciplinary environment, with relatively small barriers between different fields and departments and remarkable possibilities for collaborations. A hallmark of Bell Labs research has always been the tight interplay and cross-fertilization of science and technology, which led to so many revolutionary advances. This is precisely the type of interplay that the late Hendryk Casimir, distinguished physicist and director of the Philips Research Laboratories in the Netherlands, another world-renowned institution, described often as the "spiral of science and technology".

I would like here to mention just a few examples of this synergy in the physical sciences division at Bell labs: the discovery of the wave nature of the electron which sprung out from work focused on improving cathode surfaces in valves; the transistor, emerging from basic understanding of semiconductor physics motivated by the foresight that the time had come for a revolutionary device capable of overcoming existing projected bottlenecks in communications; the seminal work on lasers including the first useful semiconductor laser for fiber optics communications; the serendipitous discovery of the microwave back ground radiation, thought to be the remnant of the big-bang, by scientists attempting to minimize the noise of an antenna intended for satellite communications, and the relatively recent invention and first demonstration of functional magnetic resonance imaging by Dr. Ogawa, then a member of the biophysics department, which is revolutionizing brain sciences.

Bell Labs is also world renowned for its many breakthroughs in mathematics, information and network science; suffice to say that information theory was born there through the seminal and definite work of Claude Shannon and has continued with the research of Dr. Peter Shor in the nineties on quantum computing algorithms for prime number factorization, work recognized not long ago by the King Faisal Prize.

In summary the breadth of research in the physical sciences division at Bell Labs was just staggering, and basically remained that way until the early nineties, covering fields as diverse as condensed matter physics, materials science and chemistry, lasers and communications, physical optics, device physics, biophysics and astrophysics. Young scientists and even not so young ones were strongly encouraged to open up new areas of research and take risks. Most managers in the research area had been or were still top-notch scientists, who had come up through the ranks primarily on the basis of their scientific achievements. Some like the world famous Kumar Patel, inventor of the carbon dioxide laser, managed to continue to do research even in high-management positions.

I found the "Bell Labs spirit" highly contagious and that is why I still consider Bell Labs as my "Alma Mater". I have no doubt that it has deeply influenced my research "taste" and choice of problems, often at the interface between basic research and technology and straddling several disciplines. The Quantum Cascade laser, for which I am receiving this very prestigious prize, certainly falls in that category.

At the time when I arrived at Bell Labs its physical sciences division was "bubbling" with excitement about the capabilities of a new semiconductor thin film growth technique, pioneered by my colleague and friend of so many years, Dr. Alfred Y. Cho, and known as Molecular Beam Epitaxy or simply MBE. MBE is a crystal growth technique capable of depositing thin films down to a thickness of one molecular layer, nowadays widely used in the manufacturing of optoelectronic and high-speed electron devices. In this way for the first time artificial semiconductor materials with tailorable electronic and optical, designed using the laws of quantum mechanics, could be synthesized and be used in the fabrication of novel optical and electronic device.[12] To this date most the laser for optical recording (compact disc players) are manufactured by MBE, in addition to low-noise transistors for cellular phones and, of course, quantum cascade lasers!

In what turned later to be recognized as a stepping stone towards the invention of the QC laser, in the early eighties I had proposed devices that used an energy staircase created by applying an electric field to a semiconductor alloy with a periodic saw-tooth like band gap profile, for a variety of applications such as solid-state photomultipliers and ultra high speed " repeated velocity overshoot" devices.[12] This staircase energy diagram is the forerunner of the one used in the QCL over ten years later!

We demonstrated another important ingredient of the QC lasers in 1986, through our observation of sequential resonant tunneling through an AlInAs/GaInAs multiquantum well structure.[12] It is worth pointing out that in 1971 Rudolph Kazarinov and Robert Suris of the Ioffe Institute in St. Petersburg, proposed that laser amplification could occur between the quantized electronic states of a multiquantum well structure under a high electric field which caused resonant tunneling of electrons.[13]

## Programmable wavelengths and photon cascades by quantum design

The chemical composition of the active medium controls the emission wavelength of lasers because it determines the energy levels between which laser action takes place. This simple but important point is reflected in semiconductor lasers by the fact that very different band-gaps, and therefore different active region materials, must be selected to achieve substantially different laser wavelengths. QC lasers are in this sense fundamentally different since their wavelength is programmable over an unprecedented range by choosing the thickness of the materials in the active region, *without changing their composition.* Using the same semiconductor alloys ($Al_{0.48} In_{0.53} As$ and $Ga_{0.47} In_{0.53} As$), for the layers in the active regions, and by suitable choice of their thickness, QC lasers with an emission wavelength ranging from 4 to 24 μm have been demonstrated.[2-6]

To understand how it is possible to design the emission wavelength over such a wide range, note that in contrast to semiconductor lasers, in QC lasers the photon is emitted when an electron "jumps" between energy levels of quantum wells without the participation of a positive charge (the hole). As a result in QC lasers the bandgap plays no role in determining the emission wavelength, which is instead fixed by that energy level difference. These levels are an excellent example of quantum engineering using man-made materials. They arise from the well-known quantum size effect in nanometer thick layers, called quantum wells, in analogy with the well-known particle-in-a-box problem of introductory quantum mechanics. Only discrete energies are possible. By adjusting the quantum well thickness one can therefore design the energy level difference and accordingly the emitted wavelength. Likewise one uses similar strategies to maximize the strength of the optical transition, and to optimize the lifetimes of electrons in the energy levels, in order to achieve the population inversion required for laser action.

Designing QC lasers is therefore a sophisticated exercise in *band-structure engineering*. This technique combined with Molecular Beam Epitaxy (MBE) has played a key role in the invention and development of many modern semiconductor devices and materials with tunable electronic and optical properties.[12]

Since in QC lasers the electron is not neutralized by a hole after emitting a laser photon, it can be easily recycled by injecting it into an adjacent identical active region, where it emits another photon, and so forth. To achieve that, active regions are alternated with electron injectors and an appropriate bias voltage is applied, which gives rise to an energy staircase in which photons are emitted at the "steps". Thus for drive currents greater than that required for achieving laser action each electron injected into the structure will emit a number of laser photons equal to the number $N$ of cascaded stages. Typically $N$ ranges from 20 to 35 for wavelengths in the 4 μm to 8 μm range, but working lasers with a single stage and up to 100 stages, have been demonstrated. This cascading effect is responsible for the very high power attainable in QC lasers, up to 1 Watt at room temperature in pulsed mode at wavelengths near 5 μm and 8 μm and similar power levels in continuous wave (cw).
Fig 2a shows a micrograph, obtained by transmission electron microscopy, of a cross-section of a QC laser, designed to emit at 4.65 μm. Each stage comprises an injector and active region made of nanometer thick layers consisting of GaInAs quantum wells separated by AlInAs barriers.

QC lasers are fabricated in a standard rectangular mesa waveguide configuration (Fig. 2b) consisting typically of 1 to 3 mm long and 5 to 20 μm wide mesas, depending on the wavelength. The laser facets, as in standard diode lasers, are defined by cleaving the structure along crystalline planes normal to the laser bar. Dielectric layers protect the sides of the mesas and low resistance contacts are provided by suitable metallizations to the top and bottom of the structure. In this configuration QC lasers run typically on multiple closely spaced wavelengths, called longitudinal modes, separated in frequency by the reciprocal of the round-trip time of light in the cavity. To select a single mode one fabricates a grating either on the top of the mesa (as in Fig. 2b) or on top of the N stacks followed by a re-growth of the top waveguide cladding layer. The period of the grating $d$ selects the wavelength which satisfies the Bragg condition $\lambda_B = 2\ n_{eff}\ d$, where $n_{eff}$ is the effective refractive index of the waveguide. Radiation of this wavelength experiences a higher reflectivity than neighboring ones because of Bragg reflection and is therefore selected for laser action. These structures called distributed feedback lasers (DFB), are widely used in telecommunications and have allowed QC lasers to achieve single mode operation with extremely high rejection ratio (greater than $10^3$ to 1) of the other modes.[10] Wavelength tuning is achieved by changing the temperature of the active region either by a temperature controller or by linearly increasing the current injected into the laser.[10] This has the effect of increasing the refractive index $n_{eff}$ thus increasing the emission wavelength.

Fig. 3a is the energy diagram of a QC laser designed to emit at a wavelength of 4.65 μm. The applied electric field (70 kV/cm), which represents the slope of the energy diagram, is required to inject electrons from the ground state **g** of the injector into level **3**. The thinnest well in the active region provides a state that resonantly couples to level **3**, thereby enhancing the resonant tunneling of electrons from the injector into the upper state of the laser transition. The latter is defined by the energy level separation $\mathbf{E_3} - \mathbf{E_2} = 278$ meV, which is primarily determined by the thickness of the two largest wells in the active region. A necessary condition for laser action is that the electron population in state **3** exceeds that of state **2**. This population inversion is achieved if the electron relaxation time from state **3** to state **2** exceeds the electron lifetime in the latter state, i.e. ($\tau_{32} > \tau_2$). The relaxation time between the states is largely controlled by the emission of optical phonons (optically active vibrational modes of the lattice), when the level separation equals or exceeds the energy of the latter ($h\nu_{\mathrm{opt}} \cong 35$ meV). Its value increases with increasing separation and is $\tau_{32} = 2.6$ ps for electron scattering by optical phonons between states **3** and **2** of the structure of Fig. 3a. To maximize $\tau_{32}/\tau_2$ the energy separation $\mathbf{E_2} - \mathbf{E_1}$, is chosen to be equal to $h\nu_{\mathrm{opt}}$, a central design tenet for most QC lasers. An electron in state **2** therefore will scatter very fast ($\tau_{21} \sim 0.3$ ps) into level **1** because of the resonant nature of this process, which gives the dominant contribution to the lifetime $\tau_2$.

To prevent accumulation of electrons in level **1,** the exit barrier of the active region is made very thin, which allows fast tunneling into the adjacent downstream injector region. Injectors consist of a superlattice (SL), with quantum wells coupled by very thin barriers so that the electronic states extend over many layers forming narrow energy bands, called "minibands", separated by regions with negligible density of states, known as "minigaps" (≥100 meV wide). The miniband that faces state **2** collects electrons exiting the active region and allows their fast transport across the injector and their relaxation into state **g**. Electrons are then re-injected into the next stage. The SL is also designed so that a minigap faces state **3,** thus suppressing electron escape by tunneling. As a result enough electron population builds-up in that level to achieve laser action at a reasonable current density (the so called laser threshold). Injectors are intentionally doped, while the active regions are generally undoped, to minimize unwanted broadening of the gain spectrum that would lead to large threshold currents. This design of the active and injector region can be scaled to different wavelengths by appropriate choice of the layers' thickness. In this way very high performance QC lasers have been demonstrated up to a wavelength of 10 μm. [2-5, 10]

The depth of the quantum well, i.e. the conduction band discontinuity $\Delta E_c$ between the two materials, limits the shortest possible wavelength achievable in QC lasers. For $Al_{0.48}In_{0.53}As$ / $Ga_{0.47}In_{0.53}As$ heterostructures (which are lattice matched to the Indium Phosphide substrate) $\Delta E_c$ is about 520 meV, which limits the shortest wavelength to about 4.3 μm. $\Delta E_c$ can be made larger (~ 700 meV), by choosing strained layer compositions (with tensile and compressive strain compensation period-by-period) in order to accommodate shorter wavelengths (down to 3.4 μm).

The device of Fig. 3a operated pulsed at room temperature (RT) with a peak power in excess of 100 mW in a single mode near the design wavelength of 4.65 μm, when fabricated in a DFB configuration. The device had 24 stages. Even higher power levels (200 –300 mW) are achieved when the lasers are operating without a grating (multimode emission).

The group of Jerome Faist at the University of Neuchatel in Switzerland was the first to achieve RT cw operation of a QC laser, which now has achieved record power levels.[8]

For longer wavelengths (≥ 10 μm) it becomes advantageous to achieve laser action between spatially extended states such as of the structure depicted in Fig. 3b, which lases at a wavelength of 24μm, corresponding to the **2'–1'** optical transition. For this device we computed a very favorable ratio of the electron relaxation time (2.4 ps) between levels **2'** and **1'** and the lifetime of **1'** (0.7 ps), thus ensuring a strong population inversion.[5] The fabricated lasers had 65 stages and operated in pulsed mode emitted up to 2 mW of optical power at 100 K.

In a major breakthrough Alessandro Tredicucci and his collaborators at the Scuola Normale Superiore in Pisa, Italy, a few years ago reported the first operation of a QC laser at far-infrared wavelengths (≅ 70 μm), using a multiquantum well active region made of Aluminum Gallium Arsenide/Gallium Arsenide heterostructure.[14] Far infrared QC lasers now operate at wavelengths as long as 150 μm [6] and up to temperatures in excess of 100 K in pulsed mode and peak power levels of a few tens of mW at cryogenic temperatures.

For wavelengths greater than approximately 20 μm QC lasers can't use conventional dielectric waveguides to confine light along the layers due to the prohibitively thick sequence of layers required for this task, but rather metal semiconductor waveguides. The laser mode in this waveguide propagates along the semiconductor/metal interface with the intensity peaked at the latter.[2-5] This interface electromagnetic wave is known as a surface plasmon. In this way much less material is used and a near unity optical confinement factor (i.e. the fraction of the laser mode overlapping the active regions) can be achieved.

Band-structure engineering offers the opportunity to design laser materials with properties not easily found in bulk semiconductors. We have been able to design and demonstrate QC lasers capable of emitting simultaneously two far apart wavelengths and a broad quasi continuum of wavelengths.[15] This is essentially impossible to do with diode lasers because the shorter wavelength would be absorb by the active region. In QC lasers instead quantum wells with different energy level separations can be designed to emit and absorb a narrow range of wavelengths centered, which will not be absorb by other quantum wells designed for different wavelengths.

## Quantum Cascade Laser Applications

The advantages of laser based optical methods in trace gas analysis and chemical sensing are their non-invasive nature, high sensitivity and selectivity, and real-time detection. In point sensing a sample of air is introduced into a chamber in which the light from a laser undergoes many reflections to increase the optical path. This geometry combined with direct-absorption-spectroscopy or wavelength-modulation-spectroscopy in which the first or second derivatives of the spectrum are measured, greatly enhances sensitivity and allows one to determine the local concentration of a species with high accuracy and sensitivity (ppbv).

From a technological perspective the two regions in which the atmosphere is relatively transparent (3-5μm and 8-13 μm) due to lack of water absorption, are particularly important for chemical sensing applications. These "windows" can be used to detect small concentrations of traces of environmental and toxic gases and vapors by a variety of spectroscopic methods with very high sensitivity (ppbv) that would otherwise be masked by the large atmospheric background. Industrial uses of chemical sensing include: combustion diagnostics in the power and automobile industries, medical diagnostic, such as breath analysis for the early detection of diseases such as ulcer, diabetes and colon cancer, industrial process control. Military applications and law enforcement applications range from the detection of chemical and biological weapons of mass destruction to that of explosives and fugitive emissions from sites of illicit drug production and countermeasures, such blinding the infrared sensor of a heat-seeking missile. Earth atmospheric science also requires accurate spectral data in the mid-ir for the determination of chemical concentration profiles, which are important for the development of reliable global models of the climate.

Most of the above applications require tunable single mode mid-ir lasers in addition to reasonable power levels for improved signal-to-noise. Commercial applications such as field point sensors also need portability, which requires compact, room temperature and battery-operated light sources. All these requirements can hardly be met by existing mid-ir sources while QC lasers have satisfied them and have already been used in many trace gas analysis applications.[7,9,10]

In a typical spectroscopic measurement employed in these applications the temperature of the laser is adjusted to position the wavelength on the short wavelength side of a particular absorption peak and the wavelength is then repetitively scanned across using a low frequency (100 Hz–10 kHz) current ramp.

Chris Webster and coworkers at the Jet Propulsion Laboratory, flew a QC DFB laser aboard a NASA ER-2 high altitude aircraft (Fig.4) to conduct measurements of the concentration of trace gases in the earth's atmosphere.[15] Using a cryogenically cooled QC laser operating near 8 µm during a series of 20 aircraft flights from September 99 to March 2000 they took measurements of tracers of atmospheric circulation such as Methane ($CH_4$) and Nitrous Oxide ($N_2O$) up to ~ 20 Km in the stratosphere over North America, Scandinavia, and Russia. The surrounding air was sucked into a multi-pass gas cell aboard the plane and analyzed by derivative spectroscopy as the plane made 8 hrs long flights to map the vertical profile of the trace gases. The noise-equivalent sensitivity limit of the apparatus (minimum detectable mixing ratio) was ~ 2 ppbv for $CH_4$.

Anatoly Kostirev, Frank Tittel and collaborators at Rice University and Physical Sciences Inc. (PSI), have demonstrated the first application of a thermoelectrically cooled, distributed feedback quantum cascade laser for continuous monitoring of CO in ambient air based on absorption spectroscopy at $\lambda = 4.6$ µm (Fig. 5). The achieved sensitivity is much greater than that obtained with standard non-optical methods and holds promise for the commercial development of QC laser based CO sensors. The Rice group has pioneered trace gas detection in open air with near room temperature QC lasers, measuring concentrations near 1 ppmv for $CH_4$ and HDO, and of sub-ppbv for $N_2O$.[9]

Measurements of $NO_x$ concentrations at sub-ppm levels in vehicle exhaust are needed for emissions certification of future ultra-low emission vehicles. The group of Bill Weber at the Ford Research Laboratory has recently used wavelength-modulated QC lasers from our group, combined with a multipass absorption cell, to measure $NO_x$ concentrations at the few parts-per-billion (ppb) range in diluted exhaust-gas bag samples collected in the vehicle certification process.

We conclude this review by discussing the potential of QC lasers for optical wireless communications in the eye-safe (3-5)µm and (8-13)µm atmospheric transmission windows. In these regions losses associated with light scattering are many orders of magnitude lower with respect to the near infrared, particularly in conditions of fog or pollution. My group at Bell Labs demonstrated the first atmospheric transmission of complex data (multimedia satellite channels comprising 800 TV channels and 100 radio channels) with a QC laser ( = 8.1 m) over a distance of 200 m.[5]

## Acknowledgements

I am grateful to the many talented collaborators who have made this work possible. In particular I would like to emphasize the key contributions made by Jerome Faist and Carlo Sirtori to the design and demonstration of the first QC laser as well as their subsequent work in my group at Bell Labs from 1994 to 1996. Al Cho and Deborah Sivco grew the MBE material without which QC lasers would not have become a reality. Claire Gmachl played a central role in the development of single mode and multiwavelength QC lasers and in collaborations with many groups worldwide.

I am grateful to Bell Labs management for its support of this work over many years and to DARPA for providing partial funding of our research from 1996 to 2003.

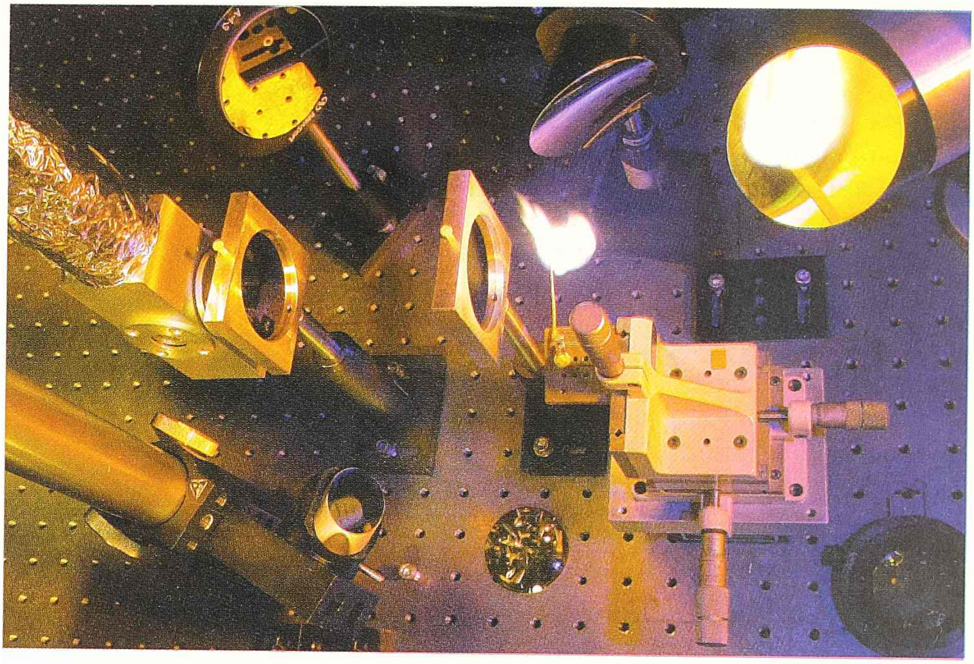
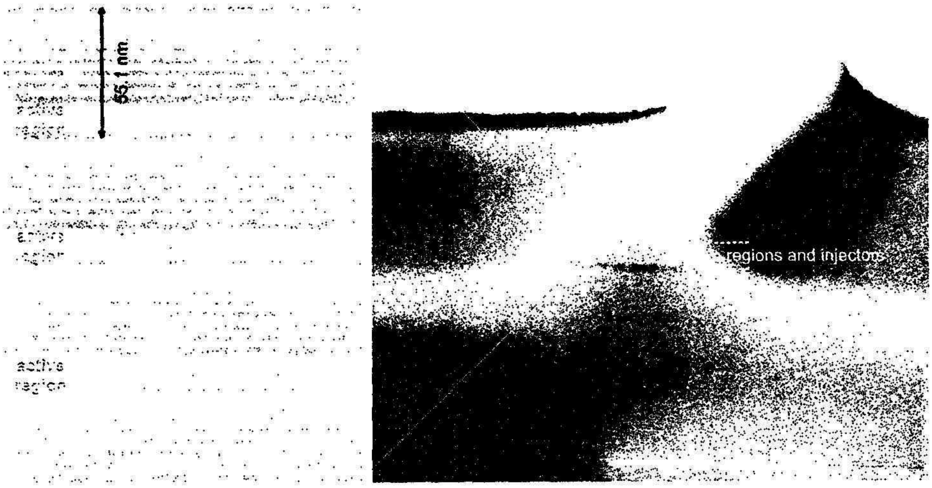


Figure captions

**Fig. 1**

A high power (200 mW) continuous wave mid-infrared quantum cascade laser emitting at a wavelength of 8 μm lights-up a match held in its path. The laser is housed in a cryostat and the invisible beam is made parallel and focused by the two lenses.



a. b.

Fig. 2

**(a)** High-resolution transmission electron microscope image of four stages of a quantum cascade laser designed to emit at a wavelength $\lambda \cong 4.65$ μm, grown by Molecular Beam Epitaxy. The white layers are the quantum wells, made of Gallium Indium Arsenide (GaInAs), a semiconductor alloy. They are separated by barrier layers made of Aluminum Indium Arsenide (AlInAs). The wavelength is primarily determined by the thickness of the two largest quantum wells (4.4 and 4.8 nanometer thick) in the active regions.

**(b)** Electron microscope image of a portion of the quantum cascade laser. The geometry is that of an optical waveguide shaped in the form of a ridge, fabricated by optical lithography and wet etching. Its width is 8 μm. The waveguide core comprises 25 stages, each consisting of an active region and an electron injector as shown in **(a)**. It is sandwiched between Al In As/GaInAs cladding regions of lower average refractive index, which guide the light parallel to the layers. Shown at one end of the waveguide is one of the two laser facets, which define the optical cavity. Its length is 2.5 mm. A voltage is applied across the device so that the electron current flows from top to bottom through the stacks of active regions and injectors. Shown is the top metallic contact, which stops short of the laser facet and exposes the grating, etched on top of the mesa. The grating selects the precise laser wavelength within the emission spectrum of the active region.

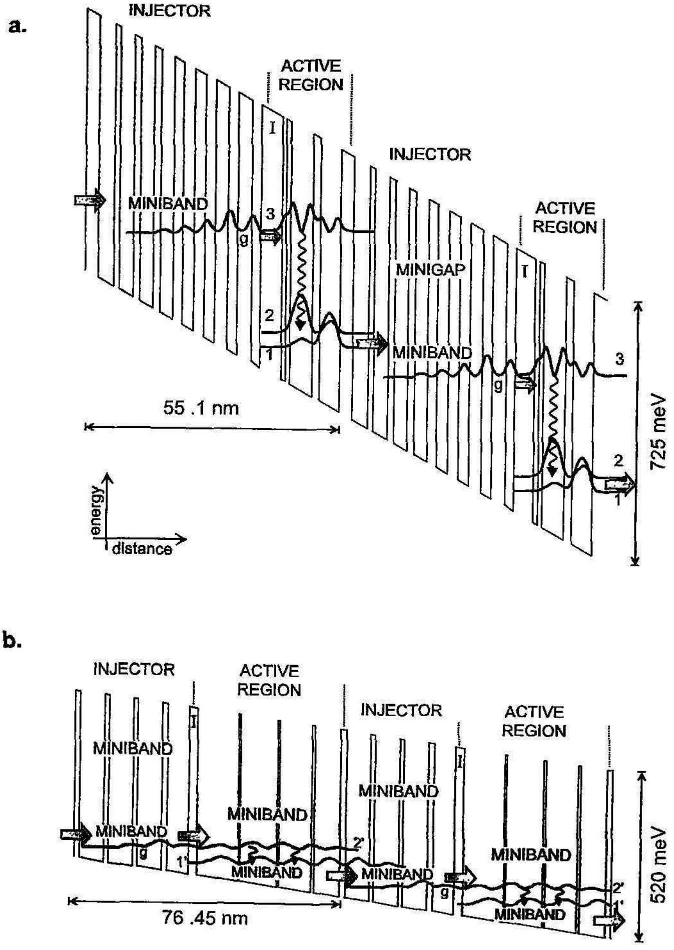


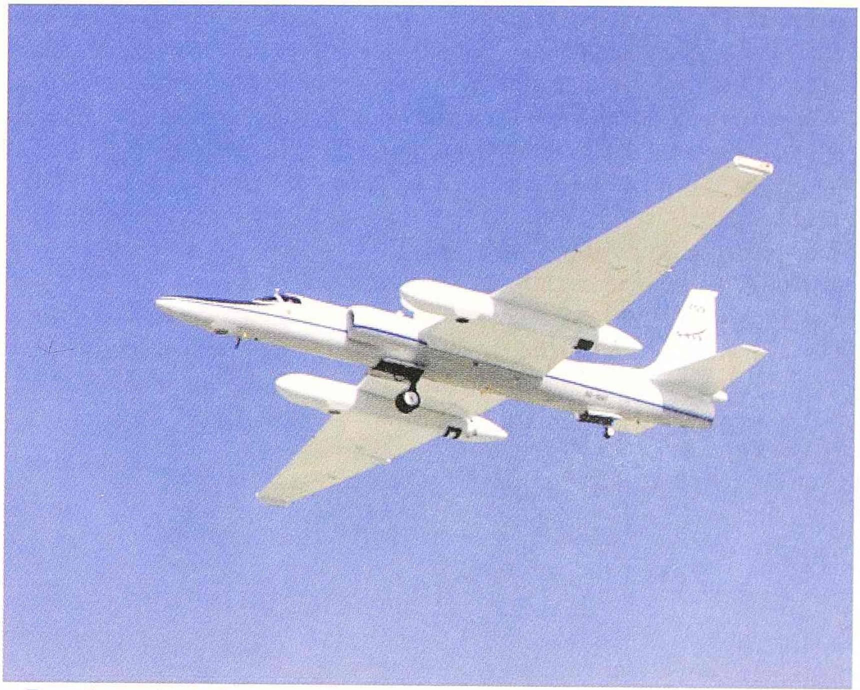
**Fig. 3**

Energy diagrams of quantum cascade lasers designed for emission at different wavelengths.

**(a)** Two stages of a QC laser emitting at wavelengths near 4.65 μm (wavy arrows) corresponding to energy difference between states **3** and **2**. The energy barriers and the quantum wells are made of AlInsAs and

InGaAs alloys, respectively. The alloy composition is chosen to provide high-energy barriers (725 meV). The slope of the band diagram represents the applied electric field required for resonant tunneling electron injection into the upper state of the laser transition. The wavy curves represent the modulus squared of the relevant wavefunctions.. "I" indicates the injection barrier. The three quantum wells of the active regions are 0.9. 4.8 and 4.4 nm-thick separated by 0.9 and 1.7 nm barriers. Electrons are injected from left to right and emit a laser photon per stage as they cascade down the structure.

**(b)** Two stages of a quantum cascade laser designed to emit near 24 μm (wavy arrows). Unlike the structure of **(a)** the laser transition is between delocalized electronic states **2'** and **1'**. The GaInAs quantum wells in the active region have a thickness ≅ 10 nm and are separated by ≅ 0.4 nm thick AlInAs barriers.



Dryden Flight Research Center EC98-44435-2 Photographed 2MAR1998
ER-2 #709 take off (NASA photo/Tony Landis)

**Fig. 4**

NASA's ER-2 high-altitude aircraft takes off from Dryden Flight Research Center (NASA photo courtesy Tony Landis). The ALIAS (Aircraft Infrared Laser Spectrometer) containing the QC laser is located in the superpod seen on the right wing. This QC laser based instrument carried high sensitivity measurements (parts per billion in volume) of the vertical profile of trace gases ($CH_4$ and $N_2O$) in the stratosphere

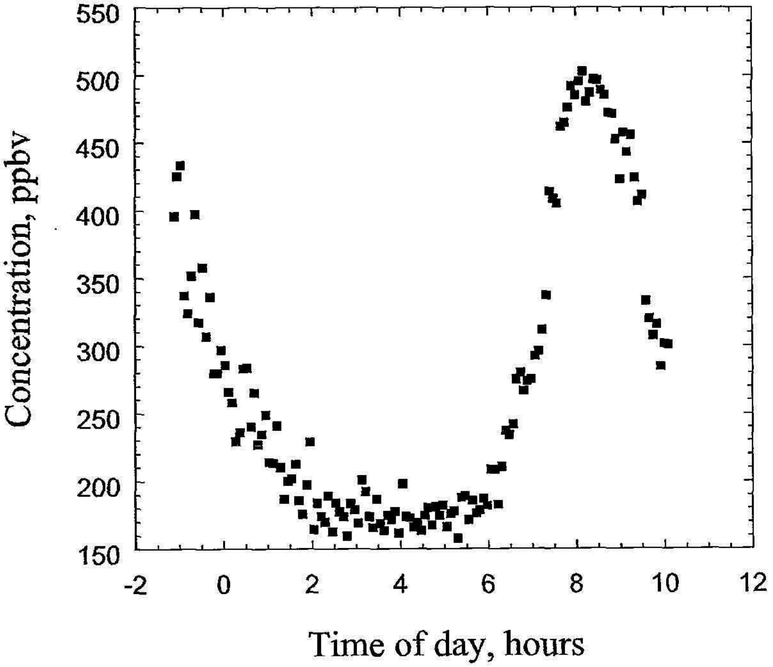
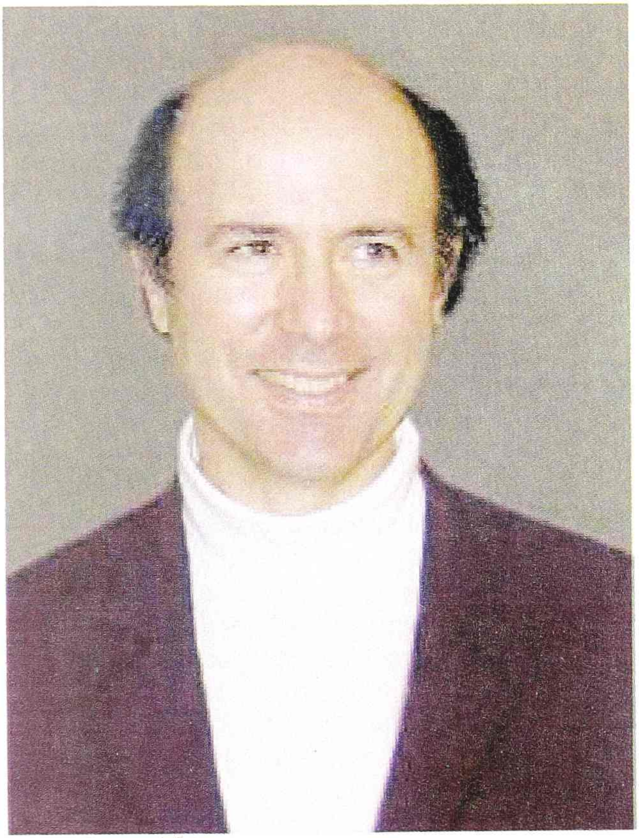


**Fig. 5** Concentration of Carbon Monoxide (CO) in ambient air monitored with a gas sensor based on a pulsed room temperature single mode QC laser operating at λ =4.6 μm. A detection limit of 12 part-per-billion in volume was achieved with a one meter optical path length. Two characteristic maxima of CO concentration were observed during a typical day corresponding to morning and evening rush hour traffic in Houston, Texas. (Courtesy of A. Kostirev and F. Tittel, Rice University).

## References:

1. M. Tacke, *Infrared Phys. Technol.*, **36**, 447 (1995); M. Hodges and U. W. Schiessl, *SPIE Proc.*, **3628**, 113 (1999)
2. F. Capasso, et al. *Opt. Photon. News,* **10**, 31, (October 1999); *ibid.* **12**, 40 (May 2001); F. Capasso, C. Gmachl, D. L. Sivco, and A. Y. Cho, *Physics Today* **55**, 34 (May 2002)
3. F. Capasso et al., *IEEE J. Selected Topics in Quantum Electronics*, **6**, 931 (2000).
4. C. Gmachl, F. Capasso, D. L. Sivco, and A.Y. Cho, *Reports on Progress in Physics* **64**, 1533 (2001)
5. F. Capasso, et al. *IEEE J. Quantum Electron.* **38**, *511* (2002)
6. B. S. Williams, S. Kumar, H. Callebaut, Q. Hu, and J. L. Reno, *Appl. Phys. Lett.* **83**, 2124 (2003
7. J. Ouellette, *The Industrial Physicist*, **7**, 8 (April 2000)
8. J. S. Yu et al. *IEEE Photonics Technology Lett.* **16**, 747, 2004
9. A. A. Kosterev and F. K. Tittel, *IEEE J. Quantum Electron.* **38,** 582 (2002)
10. C. Gmachl et al., *IEEE Circuits and Devices* **16,** 10 (2000).
11. J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A.Y. Cho, *Science*, **264**, 553 (1994); F. Capasso et al., US patent 5457709.
12. F. Capasso, *Science*, **235**, 172 (1986); F. Capasso and A. Y. Cho, *Surface Sci.* **299/300**, 878 (1994); F. Capasso and S. Datta, *Physics Today*, **43**, 74, February 1990.
13. R. F. Kazarinov and R. A. Suris, *Sov. Phys. Semicond.* **5**, 707 (1971)
14. R. Köhler, et al. , *Nature (London)* **417**, 156 (2002)
15. C. Gmachl, et al. , *Nature (London)***415**, 883 (2002)
16. C. Webster et al. *Appl. Opt.* **40**, 321 (2001)



**Professor Frank Wilczek**
**Co-Winner**
**2005 King Faisal International Prize**
(Science)

## Synopsis of Achievements
PROFESSOR FRANK WILCZEK

Frank Wilczek is a theoretical physicist. He was born in New York City, in 1951. He was educated in the public schools of New York, until he entered the University of Chicago in 1967. He graduated in 1970 with a B.S. in Mathematics. From there he went on to Princeton University, where he received Masters degrees in Mathematics and in Physics in 1972, and a Ph. D. in Physics in 1974. He stayed at Princeton, becoming Professor in 1979, before joining the new Institute for Theoretical Physics at Santa Barbara in 1980, where he became the Chancellor Robert Huttenbach Professor. He moved to the Institute for Advanced Study at Princeton in 1990, where he became the J. Robert Oppenheimer Professor. He moved to the Massachusetts Institute of Technology in 2000, as the Herman Feshbach Professor.

Frank Wilczek has received many awards for his work, including the Sakurai and Lilienfeld Prizes of the American Physical Society, the Dirac Medal of UNESCO, the Lorentz Medal of the Dutch Physical Society, and the 2004 Nobel Prize.

Working with David Gross, he discovered the central dynamical feature unique to non-abelian gauge theories, that their strength diminishes to zero at short distances or high energy due to quantum antiscreening of charge (asymptotic freedom). They explored many examples, including some with scalar fields and some with non-trivial weak-coupling infrared fixed points. They proposed one such gauge theory, now known as QCD, as the fundamental theory of the strong interaction. They proposed specific experimental tests of this proposal, which were performed successfully.

Wilczek has done pioneering work on the properties of matter in extreme conditions. He clarified the nature of the phase transition of QCD at high temperature, which is important for early universe cosmology and for the understanding of recent work on heavy ion collisions. Working with Mark Alford and Krishna Rajagopal, he discovered the color-flavor locked phase of QCD at high density. In this phase, which might be realized in the interior of neutron stars, the confinement of quarks and chiral symmetry breaking can be demonstrated analytically.

He discovered that the Higgs particle couples to ordinary matter through quantum fluctuations (top quark loops) into gluons. This observation is central to the ongoing search for that fundamental particle.

Wilczek has made important contributions to the formulation of unified field theories. He showed, in particular, that the hypothesis of low-energy supersymmetry improves the agreement between predictions for coupling constant unification and experiment – a result that continues to guide work in this field. This hypothesis will be tested definitively at the Large Hadron Collider (LHC).

Wilczek pioneered in the application of new insights from high-energy physics to cosmology. He showed how the asymmetry between matter and antimatter might be explained based on unified field theories. He introduced, named, and analyzed one leading candidate to provide the dark matter, axions. Axions have been a fruitful source of new cosmological ideas, notably by providing a concrete framework in which anthropic reasoning is demonstrably valid.

Another major theme of Wilczek's work has been to use ideas developed for the description of fundamental quantum fields for the description of macroscopic materials, and vice versa. He established the theoretical possibility of fractional quantum numbers and fractional quantum statistics for solitons and collective excitations. We now know that these possibilities are realized in quantum Hall effect phases. Recently, inspired in part by developments in high density QCD, he has predicted new exotic forms of superfluidity that can be realized experimentally.

Wilczek discovered that gauge fields appear in a beautiful and surprising way in the description of mechanical systems, notably including small creatures that swim at low Reynolds number.

In recent years Wilczek has devoted considerable effort to reflecting on the broader philosophical meaning of results in modern physics, and to communicating these results to a broader scientific public.

# New Physical Laws Suggested by Symmetry

**Frank Wilczek**
Herman Feshbach Professor of Physics
Massachusette Institute of Technology
Center for Theoretical Physics
Cambridge, MA 02139
USA

Over the course of the twentieth century, symmetry proved immensely fruitful as a source of insight into Nature's basic operating principles. Quantum chromodynamics (QCD), in particular, is constructed as the unique embodiment of a huge symmetry group, local SU(3) color gauge symmetry within relativistic quantum field theory. As we try to discover new laws, that improve on what we know, it seems good strategy to continue to use symmetry as our guide.

I feel that this is an especially appropriate subject for this volume, in light of the central role that symmetry has played in Islamic art and culture for centuries.

## 1. Unified Field Theory

Both QCD and the standard electroweak standard model are founded on gauge symmetries. This combination of theories gives a wonderfully economical and powerful account of an astonishing range of phenomena. Just because it is so concrete and so successful, this rendering of Nature can and should be closely scrutinized for its aesthetic flaws and possibilities. Indeed, the structure of the gauge system gives powerful suggestions for its further fruitful development. Its product structure SU(3)×SU(2)×U(1), the reducibility of the fermion representation (that is, the fact that the symmetry does not make connections linking all the fermions), and the peculiar values of the quantum number hypercharge assigned to the known particles all suggest the desirability of a larger symmetry.

The devil is in the details, and it is not at all automatic that the superficially complex and messy observed pattern of matter will fit neatly into a simple mathematical structure. But, to a remarkable extent, it does.

Most of what we know about the strong, electromagnetic, and weak interactions is summarized (rather schematically!) in Figure 1. QCD connects particles horizontally in groups of 3 (SU(3)), the weak interaction connects particles vertically in groups of 2 (SU(2)) in the horizontal direction and hypercharge (U(1)) senses the little subscript numbers. Neither the different interactions, nor the different particles, are unified. There are three different interaction symmetries, and five disconnected sets of particles (actually fifteen sets, taking into account the threefold repetition of families).

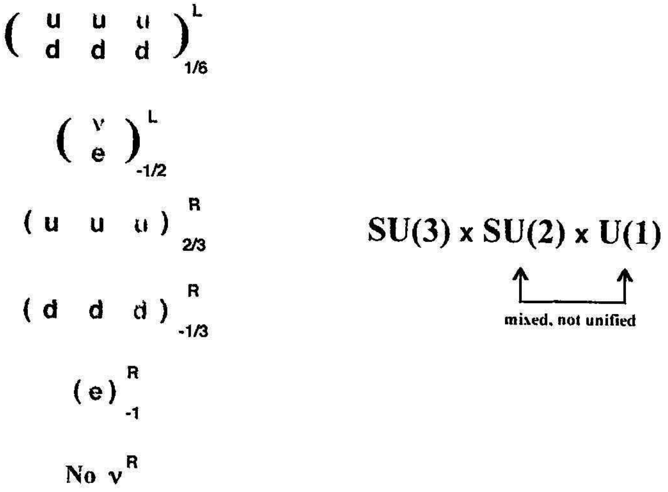


Figure 1: A schematic representation of the symmetry structure of the standard model. There are three independent symmetry transformations, under which the known fermions fall into five independent units (or fifteen, after threefold family repetition). The color gauge group SU(3) of QCD acts horizontally, the weak interaction gauge group SU(2) acts vertically, and the hypercharge U(1) acts with the relative strengths indicated by the subscripts. Right-handed neutrinos do not participate in any of these symmetries.

We can do much better by having more symmetry, implemented by additional gluons that also change strong into weak colors. Then everything clicks into place quite beautifully, as displayed in Figure 2.

There seems to be a problem, however. The different interactions, as observed, do not have the same overall strength, as would be required by the extended symmetry. Fortunately, asymptotic freedom informs us that the observed interaction strengths at a large distance can be different from the basic strengths of the seed couplings viewed at short distance. To see if the basic theory might have the full symmetry, we have to look inside the clouds of virtual particles, and to track the evolution of the couplings. We can do this, using the same sort of calculations that underlie Figure 3, extended to include the electroweak interactions, and extrapolated to much shorter distances (or equivalently, larger energy scales).

It is convenient to display inverse couplings and work on a logarthmic scale, for then the evolution is (approximately) linear. When we do the calculation using only the virtual particles for which we have convincing evidence, we find that the couplings do approach each other in a promising way, though ultimately they don't quite meet This is shown in the top panel of Figure 4.

Interpreting things optimistically, we might surmise from this near-success that the general idea of unification is on the right track, as is our continued reliance on quantum field theory to calculate the evolution of couplings. After all, it is hardly shocking that extrapolation of the equations for evolution of the couplings beyond their observational foundation by many orders of magnitude is missing some quantitatively significant ingredient. In a moment I'll mention an attractive hypothesis for what's missing.

## 1.1 Large Energy Scale and Rare Processes

A very general consequence of this line of thought is that an enormously large energy scale, of order 1015 GeV or more, emerges naturally as the scale of unification. This is a profound and welcome result. It is profound, because the large energy scale – which is far beyond any energy we can access directly – emerges from careful consideration of experimental realities at energies more than ten orders of magnitude smaller! The underlying logic that gives us this leverage is a synergy of unification and asymptotic freedom, as follows. If evolution of couplings is to be responsible for their observed gross inequality then, since this evolution is only logarithmic in energy, it must act over a very wide range.

The emergence of a large mass scale for unification is welcome, first, because many effects we might expect to be associated with unification are observed to be highly suppressed. Symmetries that unify SU(3) × SU(2) × U(1) will almost inevitably involve wide possibilities for transformation among quarks, leptons, and their antiparticles. These extended possibilities of transformation, mediated by the corresponding

|  | **R** | **W** | **B** | **G** | **P** |
|---|---|---|---|---|---|
| u | + | - | - | + | - |
| u | - | + | - | + | - |
| u | - | - | + | + | - |
| d | + | - | - | - | + |
| d | - | + | - | - | + |
| d | - | - | + | - | + |
| $u^c$ | - | + | + | - | - |
| $u^c$ | + | - | + | - | - |
| $u^c$ | + | + | - | - | - |
| $d^c$ | - | + | + | + | + |
| $d^c$ | + | - | + | + | + |
| $d^c$ | + | + | - | + | + |
| v | + | + | + | + | - |
| e | + | + | + | - | + |
| $e^c$ | - | - | - | + | + |
| N | - | - | - | - | - |

**Hypercharge Y = -1/6 (R+W+B) + 1/4 (G+P)**

Figure 2: The hypothetical enlarged symmetry SO(10) [2] accommodates all the symmetries of the standard model, and more, into a unified mathematical structure. The fermions, including a right-handed neutrino that plays an important role in understanding observed neutrino phenomena, now form an irreducible unit (neglecting family repetition). The allowed color charges, both strong and weak, form a perfect match to what is observed. The phenomenologically required hypercharges, which appear so peculiar in the standard model, are now theoretically determined by the color and weak charges, according to the formula displayed.

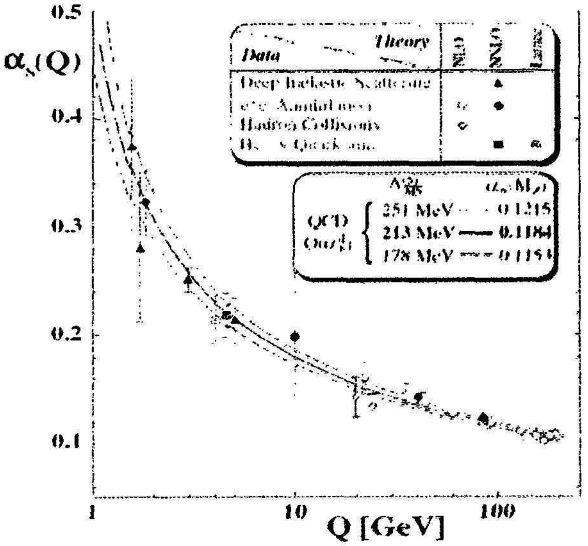


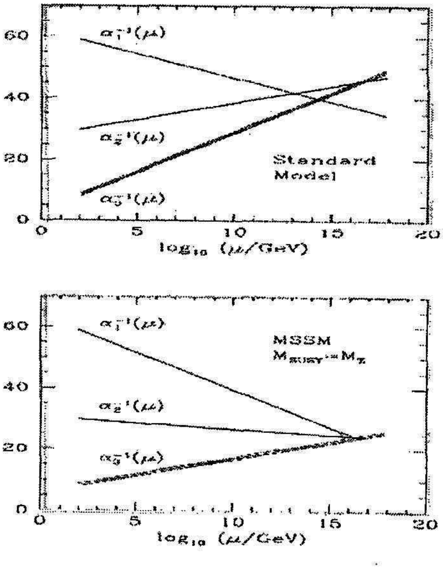
Figure 3: Many quite different experiments, performed at different energies, have been successfully analyzed using QCD. Each fits a large quantity of data to a single parameter, the strong coupling $\alpha_s$. By comparing the values they report, we obtain direct confirmation that the coupling evolves as predicted [1].

gauge bosons, undermine conservation laws including lepton and baryon number conservation. Violation of lepton number is closely associated with neutrino oscillations. Violation of baryon number is closely associated with proton instability. In recent years neutrino oscillations have been observed; they correspond to miniscule neutrino masses, indicating a very feeble violation of lepton number. Proton instability has not yet been observed, despite heroic efforts to do so. In order to keep these processes sufficiently small, so as to be consistent with observation, a high scale for unification, which suppresses the occurrence of the transformative gauge bosons as virtual particles, is most welcome. In fact, the unification scale we infer from the evolution of couplings is broadly consistent with the observed value of neutrino masses and encourages further vigorous pursuit of the quest to observe proton decay.

## 1.2 Large Energy Scale and Unification Including Gravity

The emergence of a large mass scale for unification is welcome, secondly, because it opens up possibilities for making quantitative connections to the remaining fundamental interaction in Nature: gravity. It is notorious that gravity is absurdly feebler than the other interactions, when they are compared acting between fundamental particles at accessible energies. The gravitational force between proton and electron, at any macroscopic distance, is about $Gm_em_p/\alpha \sim 10^{-40}$ of the electric

**Unification of gauge couplings**



**Gravity fits too!**
**(roughly)**

Figure 4: We can test the hypothesis that the disparate coupling strengths of the different gauge interactions derive a common value at short distances, by doing calculations to take into account the effect of virtual particle clouds [3]. These are the same sort of calculations that go into Figure 3, but extrapolated to much higher energies, or equivalently shorter distances. Top panel: using known virtual particles. Bottom panel: including also the virtual particles required by low-energy supersymmetry [4].

force. On the face of it, this fact poses a severe challenge to the idea that these forces are different manifestations of a common source – and an even more severe challenge to the idea that gravity, because of its deep connection to space-time dynamics, is the primary force.

By extending our consideration of the evolution of couplings to include gravity, we can begin to meet these challenges.

- Whereas the evolution of gauge theory couplings with energy is a subtle quantum-mechanical effect, the gravitational coupling evolves even classically, and much more rapidly. For gravity responds directly to energy-momentum, and so it appears stronger when viewed with high-energy probes. In moving from the small energies where we ordinarily measure to unification energy scales, the ratio $GE^2/\alpha$ ascends to values that are no longer absurdly small.
- If gravity is the primary force, and special relativity and quantum mechanics frame the discussion, then Planck's system of physical units, based on Newton's constant G, the speed of light c, and Planck's quantum of action h, is privileged. Dimensional analysis then suggests that the value of naturally defined quantities, measured in these units, should be of order unity. But when we measure the proton mass in Planck units, we discover

$$m_p \sim 10^{-18} \sqrt{\frac{hc}{G}} \tag{1}$$

On this hypothesis, it makes no sense to ask "Why is gravity so feeble?". Gravity, as the primary force, just is what it is. The right question is the one we confront here: "Why is the proton so light?". Given our new, profound understanding of the origin of the proton's mass, which I've sketched for you today, we can formulate a tentative answer. The proton's mass is set by the scale at which the strong coupling, evolved down from its primary value at the Planck energy, comes to be of order unity. It is then that it becomes worthwhile to cancel off the growing color fields of quarks, absorbing the cost of quantum localization energy. In this way, we find, quantitatively, that the tiny value of the proton mass in Planck units arises from the fact that the basic unit of color coupling strength, *g*, is of order ½ at the Planck scale! Thus dimensional reasoning is no longer mocked. The apparent feebleness of gravity results from our partiality toward the perspective supplied by matter made from protons and neutrons.

## 2 Supersymmetry

As I mentioned a moment ago, the approach of couplings to a unified value is suggested, but not accurately realized, if we infer their evolution by including the effect of known virtual particles. There is one particular proposal to expand the world of virtual particles, which is well motivated on several independent grounds. It is known as low-energy supersymmetry [5].

As the name suggests, supersymmetry involves expanding the symmetry of the basic equations of physics. This proposed expansion of symmetry goes in a different direction from the enlargement of gauge symmetry. Supersymmetry makes transformations between particles having the same color charges and different spins, whereas expanded gauge symmetry changes the color charges while leaving spin untouched. Supersymmetry expands the space-time symmetry of special relativity.

In order to implement low-energy supersymmetry, we must postulate the existence of a whole new world of heavy particles, none of which has yet been observed directly. There is, however, a most intriguing indirect hint that this idea may be on the right track: If we include the particles needed for low-energy supersymmetry, in their virtual form, in the calculation of how couplings evolve with energy, then accurate unification is achieved! This is shown in the bottom panel of Figure 4.

By ascending a tower of speculation, involving now both extended gauge symmetry and extended space-time symmetry, we seem to break though the clouds, into clarity and breathtaking vision. Is it an illusion, or reality? This question creates a most exciting situation for the Large Hadron Collider (LHC), due to begin operating at CERN in 2007, for this great accelerator will achieve the energies necessary to access the new world of of heavy particles, if it exists. How the story will play out, only time will tell. But in any case I think it is fair to say that the pursuit of unified field theories, which in past (and many present) incarnations has been vague and not fruitful of testable consequences, has in the circle of ideas I've been describing here attained entirely new levels of concreteness and fecundity.

## 3 Axions [6]

As I have emphasized at the start, QCD is in a profound and literal sense constructed as the embodiment of symmetry. There is an almost perfect match between the observed properties of quarks and gluons and the most general properties allowed by color gauge symmetry, in the framework of special relativity and quantum mechanics. The exception is that the established symmetries of QCD fail to forbid one sort of behavior that is not observed to occur. The established symmetries permit a sort of interaction among gluons – the so-called $\theta$ term – that violates the invariance of the equations of QCD under a change in the direction of time. Experiments provide extremely severe limits on the strength of this interaction, much more severe than might be expected to arise accidentally.

By postulating a new symmetry, we can explain the absence of the undesired interaction. The required symmetry is called Peccei-Quinn symmetry after the physicists who firstproposed it. If it is present, this symmetry has remarkable consequences. It leads us to predict the existence of new very light, very weakly interacting particles, *axions*. (I named them after a laundry detergent, since they clean up a problem with an axial current.) In principle axions might be observed in a variety of ways, though none is easy. They have interesting implications for cosmology, and they are a leading candidate to provide cosmological dark matter.

## 4 In Search of Symmetry Lost [7]

It has been almost four decades since our current, wonderfully successful theory of the electroweak interaction was formulated. Central to that theory is the concept of spontaneously broken gauge symmetry. According to this concĕpt, the fundamental equations of physics have more symmetry than the actual physical world does. Although its specific use in electroweak theory involves exotic hypothetical substances and some sophisticated mathematics, the underlying theme of broken symmetry is quite old. It goes back at least to the dawn of modern physics, when Newton postulated that the basic laws of mechanics exhibit full symmetry in three dimensions of space despite the fact that everyday experience clearly distinguishes 'up and down' from 'sideways' directions in our local environment. Newton, of course, traced this asymmetry to the influence of Earth's gravity. In the framework of electroweak theory, modern physicists similarly postulate that the physical world is described by a solution wherein all space, throughout the currently observed Universe, is permeated by one or more (quantum) fields that spoil the full symmetry of the primary equations.

Fortunately this hypothesis, which might at first hearing sound quite extravagant, has testable implications. The symmetry-breaking fields, when suitably excited, must bring forth characteristic particles: their quanta. Using the most economical implementation of the required symmetry breaking, one predicts the existence of a remarkable new particle, the so-called Higgs particle. More ambitious speculations suggest that there should be not just a single Higgs particle, but rather a complex of related particles. Low-energy supersymmetry, for example, requires at least five "Higgs particles".

Elucidation of the Higgs complex will be another major task for the LHC. In planning this endeavor, QCD and asymptotic freedom play a vital supporting role. The strong interaction will be responsible for most of what occurs in collisions at the LHC. To discern the new effects, which will be manifest only in a small proportion of the events, we must understand the dominant backgrounds very well. Also, the production and decay of the Higgs particles themselves usually involves quarks and gluons. To anticipate their signatures, and eventually to interpret the observations, we must use our understanding of how protons – the projectiles at LHC – are assembled from quarks and gluons, and how quarks and gluons show themselves as jets.

## 5. A Great Lesson

It is truly awesome to discover, by example, that we humans can come to comprehend Nature's deepest principles, even when they are hidden in remote and alien realms. Our minds were not created for this task, nor were appropriate tools ready at hand. Understanding was achieved through a vast international effort involving thousands of people working hard for decades, competing in the small but cooperating in the large, abiding by rules of openness and honesty. Using these methods – which do not come to us effortlessly, but require nurture and vigilance – we can accomplish wonders.

## References

[1] Figure courtesy S. Bethke, hep-ex/0211012.

[2] Unification based on $SO(10)$ symmetry was first outlined in H. Georgi, in *Particles and Fields - 1974*, ed. C. Carlson (AIP, New York, 1975).

[3] H. Georgi, H. Quinn, and S. Weinberg, *Phys. Rev. Lett.* **33**, 451 (1974).

[4] S. Dimopoulos, S. Raby, and F. Wilczek, *Phys. Rev.* **D24**, 1681 (1981).

[5] A standard review is H. P. Nilles, *Phys. Rep.* **110,** 1 (1984).

[6] A standard review is J. Kim, *Phys. Rep.* **150**, 1 (1987). I also recommend F. Wilczek, hep-ph/0408167.

[7] I treat this topic more amply in F. Wilczek, *Nature* **433**, 239 (2005).

**Professor Anton Zeilinger**
**Co-Winner**
**2005 King Faisal International Prize**
(Science)

## Synopsis of Achievements
### PROFESSOR ANTON ZEILINGER

Anton Zeilinger, born on May 20 1945 in Austria, received his PhD in Physics and Mathematics in 1971 from the University of Vienna. He works on the foundations of quantum physics, both experimentally and theoretically. His early work in that field has realized a number of very basic quantum predictions for the first time in experiment. He has also identified and experimentally realized a number of new phenomena which laid the basis for a novel quantum information technology. Building on these experiments his group did ground breaking work in quantum communication, quantum cryptography, quantum teleportation, and quantum computation. His work has also contributed to a new understanding of the fundamental issues in the interpretation of quantum mechanics. There he develops the Copenhagen interpretation of quantum mechanics into an interpretation where information is the central concept.

His scientific work started in the 1970s with neutron interferometry, originally at the Technical University of Vienna and later at the Massachusetts Institute of Technology. Among these early achievements are the first realization of the rotational spinor symmetry of fermions and tests of the linearity of the Schrödinger equation.

At Massachusetts Institute of Technology he became interested in the physics of entanglement as proposed in 1935 by Einstein, Podolsky and Rosen (EPR) and by Schrödinger. With M. A. Horne, he invented in 1985 the two-photon interferometer, which for the first time exploits an EPR state entangled in an external degree of freedom, momentum, rather than in spin. This proposal greatly expanded the area where entanglement can be experimentally realized and thus broadened the basis for quantum information research.

Together with D. Greenberger and M. A. Horne he invented in 1988 novel multi-particle entangled states. Today these are called GHZ (Greenberger-Horne-Zeilinger) states. They provide an all-or-nothing contradiction between local realism and quantum physics and carry the implications of Bell's theorem far beyond a statistical issue. These GHZ states and the multi-particle W-states also invented by them became central tools in quantum computation.

When in 1990 Zeilinger became Chair of Experimental Physics at Innsbruck University, he started experimentally both in atom interferometry and in entangled photon research. In atom interferometry, with his group, he demonstrated novel coherent effects with atomic beams like phase modulation and pendellösung of de Broglie waves.

In the field of entanglement, his group developed novel sources of entangled photons and applied them to the world's first demonstrations of a number of fundamental quantum information and quantum communication concepts. His realization of superdense coding in 1996 was the first case where using entanglement one was able to achieve a communication task impossible in classical communication. This was followed by the world's first quantum teleportation of individual qubits (1997), by the first quantum teleportation of entangled states (1998), by the first realization of GHZ states (1998) and a direct demonstration of the GHZ contradiction between quantum mechanics and local realism (1999) and by the first entangled-state quantum cryptography (1999).

His recent work on entangled photon communication focuses on carrying their application into the real world and thus developing a novel quantum communication technology. His group demonstrated the first outdoors transmission of entangled photons by sending them across the river Danube (2003) and confirming a violation of Bell's inequality even after transmission through air. This was followed by the world's first business application, a bank transfer, using quantum cryptography (2004), a long-distance experiment on quantum teleportation using underground optical fibers (2004) and a long-distance quantum cryptographic link across the City of Vienna (2005). He is presently working towards establishing world-wide quantum communication links via satellites

In parallel his group also works on optical quantum computation. They have realized all-optical quantum gates like a photonic nonlinear sign-shift gate and a probabilistic two-photon controlled-NOT gate and entanglement purification in 2004. Most recently, he provided with his group the first experimental demonstration of the "one-way quantum computer" concept. This is a completely novel concept which may change our notions of what actually constitutes computation (Nature, 2005, in press). In all computer concepts so far the computation proceeds by having a sequence of gates operate on the input. In contrast, the one-way quantum computer starts in a highly entangled "cluster" state which may be seen as the enabling "hardware" on which the computation is performed. The computation then actually proceeds by a sequence of measurements applied to that state. The specific measurement sequence chosen plays the role of the "software" implementing the specific computational problem to be solved.

Together with Markus Arndt, Anton Zeilinger is also interested to extend quantum interference to increasingly larger objects and to explore the quantum-classical boundary. His group holds the world record for the largest molecules (variants of fullerenes) for which quantum interference has been seen, including the biological macromolecule Tetra-phenylporphyrine. Detailed investigation of macromolecule interference at temperatures up to 3000 Kelvin and of the loss of coherence due to thermal radiation allows the prediction that it should be possible to observe quantum interferences of small viruses and even of nanobacteria.

Among Anton Zeilinger's distinctions include elected memberships of a number of learned societies and academies, most notably the German Order Pour le Mérite. He is Fellow of the American Physical Society and he carries many international prizes, like the Senior Humboldt Award, the Klopsteg Memorial Award of the American Association of Physics Teachers and the prestigious Lorenz-Oken-Medal of the German Gesellschaft Deutscher Naturforscher und Ärzte. He has held professorships and visiting positions at many institutions world-wide, including M.I.T., Melbourne University, the Technical University in Munich, Merton College at Oxford University, the Collége de France and the Humboldt University Berlin.

Zeilinger is also interested in conferring his excitement about quantum physics to a broader public. His book (in German) "Einsteins Schleier" (Einstein's Veil), published in 2003, and currently being translated into a number of languages, has been on German-language bestseller lists for many months.

Zeilinger is currently Director of the Institute of Experimental Physics at the University of Vienna and Director of the Vienna Department of the Institute of Quantum Optics and Quantum Information of the Austrian Academy of Sciences.

# From Quantum Curiosity to Quantum Technology

**Anton Zeilinger**
Professor of Physics
Institute of Experimental Physics
University of Vienna, and
Institute of Quantum Optics and Quantum Information
Austrian Academy of Sciences, Vienna

When I first heard of quantum mechanics during my student days, I was immediately fascinated. It was both the mathematical beauty of the theory which I marveled at and the interpretive questions which caught my attention. At my final Ph.D. exam, Professor Herbert Pietschmann was responsible for finding out if I know enough theory. His question about the mathematical description of spin by Wolfgang Pauli turned out to point towards a major research direction of myself in the future. This came to fruition many times in my scientific life.

## Wave experiments: form neutron interferometry …

When Helmut Rauch, together with Wolfgang Treimer and Ulrich Bonse[1] developed a new neutron interferometer, I became completely fascinated by the possibilities it opened up, particularly those connected to spin. Very soon, we performed a direct verification of the spinor symmetry of fermions.[2] There, one demonstrates that the quantum state of a neutron rotated by 360° acquires a negative sign and therefore interferes destructively with the original state. Neutron interferometry then became my main field of work for some time. The verification of other novel spin interference phenomena followed.[3] To this day, I am still proud of a very theoretical paper[4]where I brought the machinery of Pauli spin operators to full fruition in neutron interferometry; today, one can view it as interference between two qubits (quantum bits).

As a consequence of these early achievements, I was lucky enough to be invited to join the group of Professor C.G. Shull (Nobel Prize in Physics, 1994) who had just started to build up a neutron interferometry laboratory at MIT. There, we developed various new neutron interferometry types. These, we applied to study the Fizeau effect for neutrons. This is the change of phase shift due to motion of a medium[5] among other interesting neutron diffraction phenomena.

All these experiments used thermal neutrons, which at a typical speed of about 2000 m per second. For some experiments, it is desirable to have longer interaction times, and therefore I was happy to have been invited by Roland Gähler at the Institut Laue-Langevin in Grenoble to join him in working at the set-up for interference of cold neutrons which he had developed. There, we performed a precision test of the linearity of the Schrödinger equation.[6] This experimental series we topped off with precision measurements of the single- and double-slit diffraction of neutrons.[7] Some of this work entered textbooks in quantum mechanics, particularly the double-slit pattern observed. The data were taken automatically over a period of 5 days. We started the experiment and decided to go skiing while it was running. So, during the experiment no observer whatsoever was present besides people occasionally passing by not knowing what went on. So, to put it ironically, the experiment tells us something about the role of the observer in the collapse of the wave function.

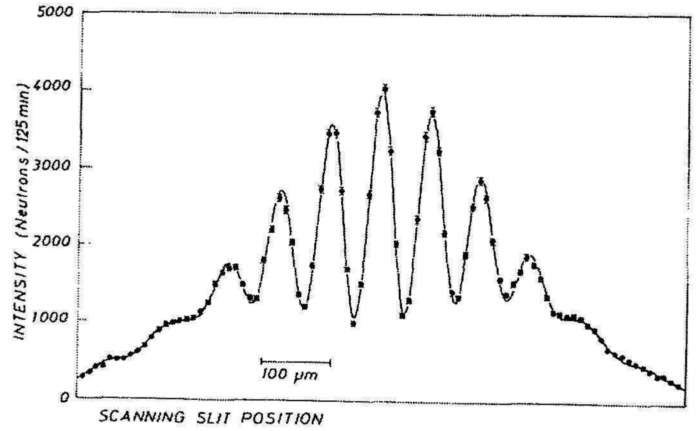


**Figure 1**

Double-slit diffraction pattern of neutrons. The solid curve represents the first-principles theoretical prediction. Note the intensity scale. It implies that while one neutron was registered, the next one to be registered in general was not even born yet. It was still residing inside its Uranium nucleus in the reactor waiting for fission to happen. The fact that the intensity pattern was thus obtained counting individual neutrons is one indication that quantum mechanics applies to individual systems and not only to ensembles.

Subsequently, we developed a new neutron interferometer for very cold neutrons.[8] But over the time it had became more and more apparent that neutron interferometry is seriously limited because of the low phase-space-density of neutron sources.

### ... via atoms to macromolecules.

In 1990 I accepted the chair in experimental physics at Innsbruck University. I took this as an opportunity to start new fields of research. Going beyond neutron interferometry, but still interested in matter waves we began experiments on atom interferometry. David Pritchard with his group had built an atom interferometer based on material diffraction gratings,[9] and he had also demonstrated diffraction of atom waves at standing light waves. I decided to combine these two, and we built the first non-material interferometer where all components were just standing light waves.[10] The atoms used were Argon. We were also able to observe some novel ways to modulate atomic deBroglie waves coherently. Most interesting phenomena arose from a detailed use of the possibilities of modulating atomic matter waves using concepts from holography. But while conventionally the hologram is made of atoms which diffract light, in our case the hologram was made of light which diffracts atoms. And the ready changeability of light waves opens up new experimental possibilities.[11,12] Some of our techniques have found their continuation in experiments with Bose-Einstein-Condensates (BEC).

Together with my then post-doc and now colleague Markus Arndt, we decided to follow a different road and investigated quantum interference of increasingly large and massive molecules. Our early demonstration of diffraction of fullerenes at material nano-fabricated gratings[13] was followed by development of an interferometer for massive molecules. The demonstration of the wave nature of biomolecules[14] using a macromolecule interferometer, and, most recently, detailed investigations of decoherence at temperatures up to 3000 Kelvin.[15]

### From entanglement theory...

I still remember the day when in 1985 during a visit to MIT - I had returned to Vienna to a new position - my colleague, Michael A. Horne, showed me an announcement of a conference in Finland held for the celebration of the 50th anniversary of the famous Einstein-Podolsky-Rosen paper of 1935.[16] He simply asked "Do you want to go to

Finland?" So we decided to search for a connection between quantum interference and quantum entanglement, which is the essence of the EPR paper. The result was the first ever proposal of a Bell inequality experiment using an external variable instead of spin[17] (Figure 2). Our idea was to use the momentum-entangled photons generated in pair annihilation of a positron with an electron. But we soon realized in collaboration with Abner Shimony[18] that the photon pairs[19, 20] generated in parametric down-conversion are ideally suited to perform our experiment. The experiment was performed by John Rarity and Paul Tapster. [21] This extension of entanglement beyond an internal degree of freedom greatly expanded our perspective for possible novel experiments, and this became important in some of our subsequent thinking in quantum computation and quantum communication.

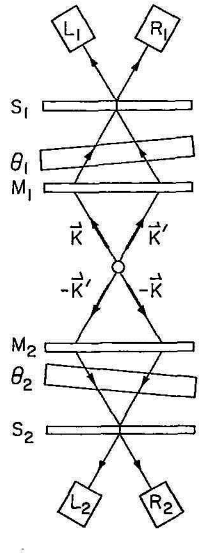


**Figure 2**
A two-particle interferometer. A source emits momentum anticorrelated pairs of particles into a superposition of both particles either carrying the primed or the unprimed momenta. Crystal plates redirect the beams and create outgoing superpositions heading for the detectors. Coincidence counts between the detectors of particle *1* and particle *2* exhibit violations of Bell's inequality and thus a conflict with the local realistic EPR point of view.

But our most interesting result of these early days came when Dan Greenberger of the City College of New York spent a sabbatical with me in Vienna 1986/87. There, we analyzed the status of entanglement and of tests of quantum nonlocality and realized that nobody had worked on entanglement beyond two particles at that time. So we decided to focus on more complicated situations, specifically on four-particle entanglement. We soon discovered[22] that a contradiction exists between quantum mechanics and local realism even for the situations where both approaches make definite, yet completely contradictory predictions for individual measurements[23] (Figure 3). This is the maximally possible contradiction between local realism and quantum mechanics. The multi-particle entangled states which exhibit this phenomenon are now called GHZ states. The quintessential one is

**Equation 1**

$$|\Psi\rangle_{GHZ} = \frac{1}{\sqrt{2}}\left(|0\rangle_1 |0\rangle_2|0\rangle_3 + |1\rangle_1|1\rangle_2|1\rangle_3\right)$$

Here, we have a superposition of three qubits, all three either in the quantum state $|0\rangle$ or in the quantum state $|1\rangle$. Very soon we also discovered that another class of multiparticle entangled states,[24] today called W-states, exists which cannot be turned into GHZ-states by local unitary operations. Little did we know at that time that both these W-states and the GHZ states would turn out to play central roles in a new filed which at that time was just starting to emerge, quantum computation.

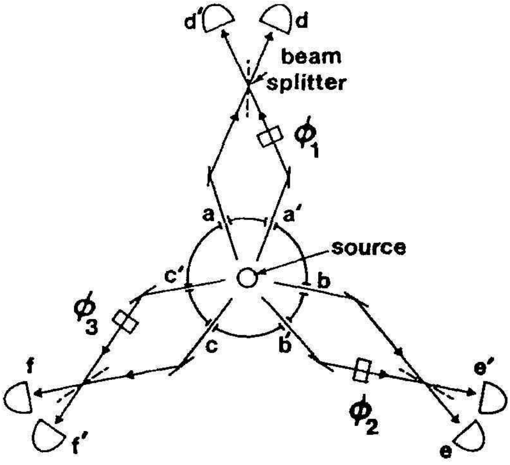


**Figure 3**
A gedanken three-particle interferometer. The source emits a triple of particles, *1*, *2*, and *3*, into six beams. A phase shift □1 is imparted to beam *a'* of particle *1*, and beams *a* and *a'* are brought together on a beam splitter before illuminating detectors *d* and *d'*. Likewise for particles *2* and *3*. The counts at the detectors are in maximal conflict with local realism.

### ... to fundamental entanglement experiments

After we had discovered the extreme GHZ contradiction between local realism in quantum mechanics, it became my goal to observe such states in the laboratory and to show the predicted contradiction explicitly and in great detail. This took 10 years. Thus, when I moved to Innsbruck to become chair of experimental physics there in 1990, besides the laboratory on atom interferometry I also built up a laboratory to do experiments with entangled photons. For that laboratory I received very useful advice from Ray Chiao at Berkeley and Leonard Mandel at Rochester. And I was happy to secure the collaboration of Harald Weinfurter, an excellent experimentalist already at that time who also had worked with neutrons until then, and Marek Zukowski, an upcoming, very promising theoretician from Poland. We had significant help by a number of visitors at this early stage, including Artur Ekert and John Rarity. Also Paul Kwiat, having just finished his Ph.D. with Ray Chiao in Berkeley joined us as a post-doctoral fellow. Collaboration with these colleagues turned out to be crucial for our subsequent successes.

The most significant result in that early period on the theory side was a proposal for entanglement swapping,[25] also called teleportation of entanglement, where we showed explicitly how entanglement between photons generated by independent sources could lead to entanglement between photons that did never interact in their past. This paved the way for many of our subsequent experiments, particularly after we were able to identify precisely the nontrivial experimental conditions how photons from independent sources can become entangled. [26]

It turned out that the crucial condition is to arrange the experiments such that from two detector-clicks created by two photons, it is impossible to distinguish, even in principle, from which of two sources which photon comes. Experimentally, this required very short pump-pulses and narrow bandwidth filtering so we entered the field of femtosecond laser physics. Today, these conditions seem obvious and they are applied in many laboratories world wide, but back at the time they were far from trivial.
On the experiment side, our most important breakthrough at that time was the development together with Alexander Sergienko and Janhua Shih, of a new high-intensity source for polarization-entangled photon pairs. [27] This source (Figure 4) produces photons in the state

**Equation 2**

$$|\Psi\rangle = \frac{1}{2}\left(|H\rangle_1 |V\rangle_2 + e^{i\chi}|V\rangle_1 |H\rangle_2\right)$$

where H and V describe horizontal and vertical polarization. This source turned out to be the work-horse in most of our subsequent experiments.

At about that time, ideas came up that entanglement might be of use in communication. The first explicit proposal by Artur Ekert[28] suggested that entangled states can be perfectly applied in quantum cryptography. There, Alice and Bob use entangled states to create the same random key at their respective locations. Using this key, they can then encrypt secret information. We set out to do that experiment but the first application of entanglement to an information protocol was hyperdense coding[i] following the proposal of Bennett and Wiesner. [29]

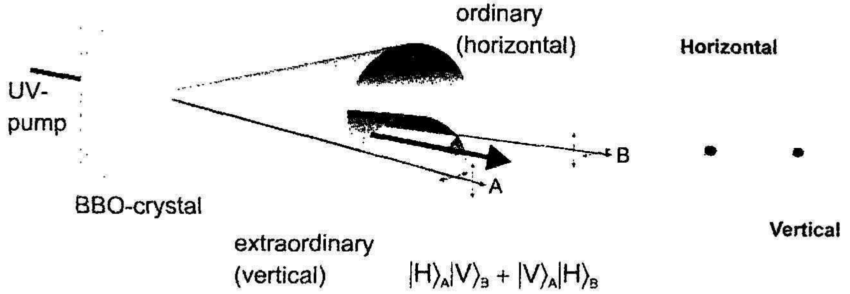


**Figure 4:**

Source of polarization-entangled photon pairs. A nonlinear crystal is pumped by a UV-laser. Spontaneous down-conversion with type-II phase matching generates entangled photon pairs (left). A photograph of the down-conversion photons (right).

In the experiment (Figure 4) Bob is able to transmit more than one classical bit of information per qubit he manipulates.

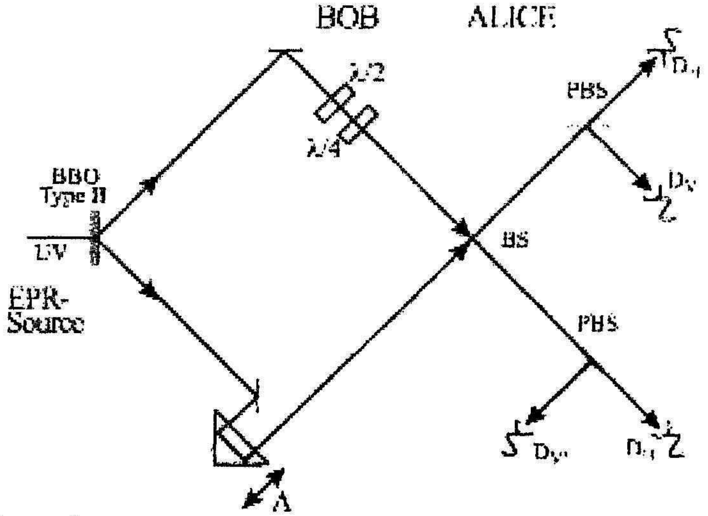


**Figure 5:**

Experimental setup for hyperdense quantum coding[i]. It consists of three distinct parts: the EPR source generating entangled photon pairs; Bob's station for encoding the messages by a unitary transformation of his particle; and,

> finally, Alice's Bell-state analyzer to read the signal sent by Bob. The setup allows one to transmit one of three messages, i.e., 1 "*trit*" ≈ 1.58 bit, by manipulating only one of two entangled photons.

### Interaction-free measurement

A most interesting direct application of fundamental quantum phenomena is interaction-free measurement.[31, 32] The idea was that one can detect an infinitely sensitive bomb which would explode even if only one photon hits it. The point is to place the bomb into one of the two beams in an interferometer. This is a simultaneous application of both the wave nature and the particle nature in one and the same experiment. It explicitly exploits the difference between probability amplitude and probability.[33]

In our experiment[31] we used a detector instead of a bomb and were able to show that it is indeed possible to detect the presence of that detector in the interferometer without the detector firing. In the original proposal the procedure had a success probability of at most 50%. We were able to demonstrate in the experiment that one can go above that limit and we discovered that the success rate can be made arbitrarily high.

### Quantum Teleportation

In 1992, a group of six theoretical colleagues had proposed the teleportation of a quantum state of a particle using entanglement.[34] I still remember that, at that time, our immediate reaction was that this experiment is impossible. It required the projection of two independently created photons into one joint entangled state. But, as I mentioned above, we were working on realizing GHZ-states. Doing that, we ourselves developed - both theoretically and experimentally - the necessary tools to do quantum teleportation. With my colleagues, particularly with Harald Weinfurter and Marek Zukowski I had developed the strategy and the essential components and ingredients of a teleportation experiment. When Dik Bouwmeester joined my group as a post-doc the resulting collaboration of a newcomer with experienced people resulted in the first realization of quantum teleportation of an independent photon in 1997[35] (Figure 6). At about that time, a young theorist from China, Jian-Wei Pan, also joined my group as a graduate student, poised at learning all tricks necessary to do a good experiment. In continuation of the earlier teleportation experiment, we were able to realize in 1998 entanglement swapping. This is the teleportation of a photon state which is itself entangled, and thus it results in entanglement of photons which never interacted.[36]

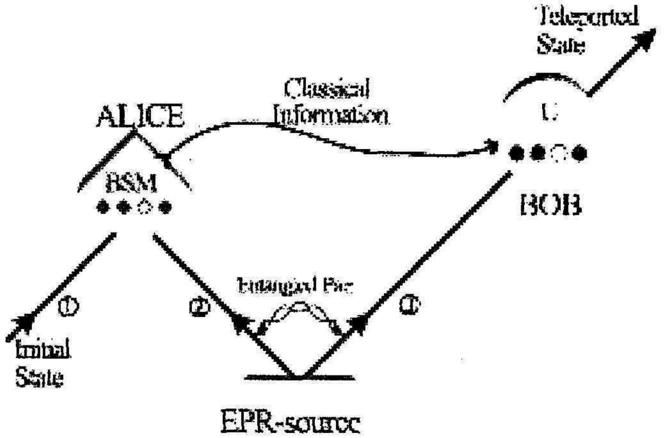


**Figure 6:**

Scheme showing principles involved in quantum teleportation. Alice has a quantum system, photon *1*, in an initial state which she wants to teleport to Bob. Alice and Bob also share an ancillary entangled pair of photons *2* and *3* emitted by an Einstein–Podolsky–Rosen (*EPR*) source. Photon *4* (created together with photon *1*) is used as a trigger indicating that photon 1 is under way to be teleported. Alice then performs a joint Bell-state measurement (*BSM*) on photons *1* and *2*, projecting them also onto an entangled state. After she has sent the result of her measurement as classical information to Bob, he can perform a unitary transformation (*U*) on the photon *3* resulting in it being in the state of the original photon *1*.

## The quest for GHZ states

This was the title of a paper we presented in 1997 at a conference in Poland[37] where we now gave in detail the necessary conditions to observe the GHZ states predicted 10 years earlier. Our experiment on quantum teleportation had been the first application of that strategy. In 1999 we finally observed[38] these states in the laboratory. Subsequent improvements allowed us to directly see and verify the extreme contradiction between local realism and quantum mechanics for the case of these states[39] (Figure 7).

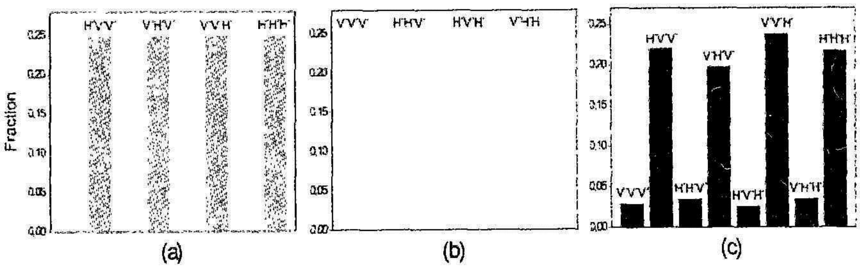


**Figure 7:**

Predictions of quantum mechanics (*a*) and of local realism (*b*), and observed results (*c*) for a set of measurements in the GHZ experiment. The maximum possible conflict arises between the predictions for quantum mechanics (*a*) and local realism (*b*) because the predicted correlations are exactly opposite. The experimental results in (*c*) clearly confirm the quantum predictions within experimental error and are in conflict with local realism.

## Towards a quantum cômmunication technology...

The earlier experiments by Freedman and Clauser[40] and the subsequent ones of Aspect and his group on two-photon entanglement[41] had left two loop-holes for local realistic theories open. One loophole was that one could still assume that some unknown communication might be responsible for the observed correlations. The only possibility to close that loophole is to change the settings of the polarizers measuring the two photons at a time after the photons have been emitted from the source. Also, in order to exclude any predictability of what will be measured the choice of the setting of the polarizers must be done in a fully random way. Following these ideas, my student Gregor Weihs performed an experiment where the two experimental stations were separated by 400 m and where ultra-fast switching of polarizers, driven by a fast quantum random number generator,[42] excluded any communication (Figure 8).

This experiment opened up the possibility for entangled-state quantum cryptography. The idea[43] there is to use the perfect correlations exhibited by the measurements on the two entangled photons to establish a random key which is subsequently used to encode a secret message. At about that time Thomas Jennewein joined my group and he was and is very interested in finding concrete applications of quantum phenomena. The same experimental set-up of Figure 8 was subsequently used to provide the first experimental demonstration of entanglement-based quantum cryptography.[44]

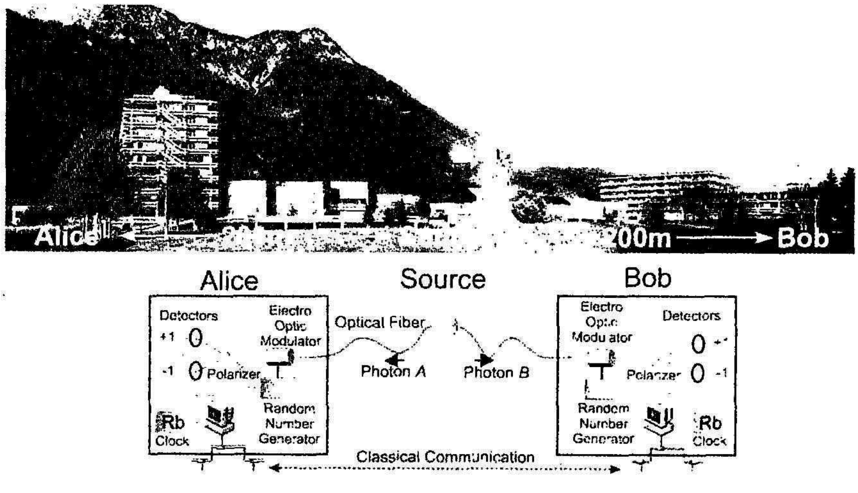


**Figure 8:**

The Innsbruck science and engineering campus at the west end of the city. The Bell-observer stations Alice and Bob are at the eastern and western ends of the site separated by about 400 meters. The down-conversion source is very close to the geometric center between the observers (top). The measurement processes on each measurement station (including randomly choosing a polarization measurement direction, setting the analyzer and detecting the photon) is fast enough that any possibility of classical (slower than the speed of light) communication between the observers could be excluded and the detection events are spacelike separated (bottom).

Presently we are in the process of developing turn-key quantum cryptography systems. The intention is to have systems which directly connect to existing classical cryptography systems already in use today. In order to both emphasize that goal and to demonstrate the present technical status we recently performed the World's first bank money transfer using quantum cryptography.[45] In the long run one can envisage a variety of goals for quantum communication. One would be to establish quantum cryptography links between any two locations on Earth; another is to imagine future quantum computers to exchange information using quantum teleportation, in some cases with immense gain in the speed of calculation.[46] Thus it is crucial to establish methods for the distribution of high-quality entanglement over large distances. After an initial experiment distributing entanglement[47] over free-space links of the order of a couple of hundred meters we were recently able to send entangled photons over a distance of 8 km across the city of Vienna (Resch et al.)[48] In a parallel development we are working at establishing quantum teleportation over large distances. Recently this includes the first long-distance quantum teleportation (Ursin et al.[49]) outside the protected environment of the laboratory. The ultimate goal is to establish quantum cryptography links on a world-wide scale and to distribute entanglement between locations separated by global distances. In view of the unavoidable losses of photons both in free space and in glass fibers and in view of the fundamental impossibility to amplify quantum states which is due to the no-cloning theorem, the only possibility is to establish quantum communication links via satellites. The experiment just mentioned[49] demonstrates that this is in principle possible. The reason is that the attenuation due to the atmosphere for a link up to a satellite should be the same as a horizontal connection over a distance of about 5 km. One can therefore reasonably expect such satellite links to be experimentally demonstrated within the next decade.

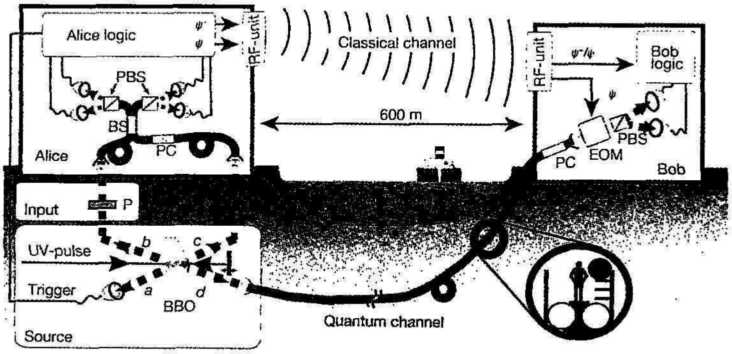


**Figure 9**

Long-distance quantum teleportation across the River Danube. The quantum channel (fiber *F*) rests in a sewage-pipe tunnel below the river in Vienna, while the classical microwave channel passes above it. A pulsed laser (wavelength, 394 nm; rate, 76 MHz) is used to pump a *β*-barium borate (*BBO*) crystal that generates the entangled photon pair *c* and *d* and photons *a* and *b* (wavelength, 788 nm) by spontaneous parametric down-conversion. The state of photon *b* after passage through polarizer *P* is the teleportation input photon; *a* serves as the trigger. Photons *b* and *c* are guided into a single-mode optical-fiber beam splitter (*BS*) connected to polarizing beam splitters (*PBS*) for Bell-state measurement. Polarization rotation in the fibers is corrected by polarization controllers (*PC*) before each run of measurements. The logic electronics identify the Bell state and convey the result through the microwave channel (*RF* unit) to Bob's electro-optic modulator (*EOM*) to transform photon *d* into the input state of photon *b*.

## ... and towards quantum computation

Quantum computation is an application of fundamental quantum phenomena, particularly entanglement and complementarity, in order to obtain novel ways to process information. The general idea is to have the information represented by a quantum state. This allows quantum superposition of different kinds of information. The quantum computer then consists of logical quantum gates which act unitarily on the state representing the information.

Yet recently it was realized that such work need not necessarily be done with localized "processors" employing atoms, ions or solid state devices. Quite surprisingly, using entangled photons and linear optics elements, one can also achieve a number of quantum computation procedures.[50]Following these suggestions in 2004, my young Japanese post-doc, Kaoru Sanaka, was able to directly observe the non-linear sign shift for a photon incident on a beam splitter.[51] This is the essential ingredient of a sign-shift gate. My Italian student, Sara Gasparoni, was then able to demonstrate a photonic CNOT gate,[52] again with linear optics. Using similar ideas, we were also able to realize entanglement purification.[53]

While all these linear-optics quantum gates operate probabilistically, a very important proposal was made by Raussendorf and Briegel.[54] They realized that deterministic quantum computation can be obtained if one also incorporates measurements of qubits in such a way that the result of earlier measurements determines which measurements are subsequently performed at other qubits. This is an entirely novel way of computation which is only possible in the quantum regime. One starts with a sufficiently complex entangled state. That state must be general enough to contain in essence all results of all calculations one would like to perform. The specific calculation then to be done implies a specific sequence of measurements on that state such that in the end the remaining quantum state represents the desired result. Since any quantum measurement is, by necessity, irreversible, one speaks of "one-way quantum computation". Our first experimental realization of a one-way quantum computer scheme has very recently been published (Walther et al.,[55] Figure 10).

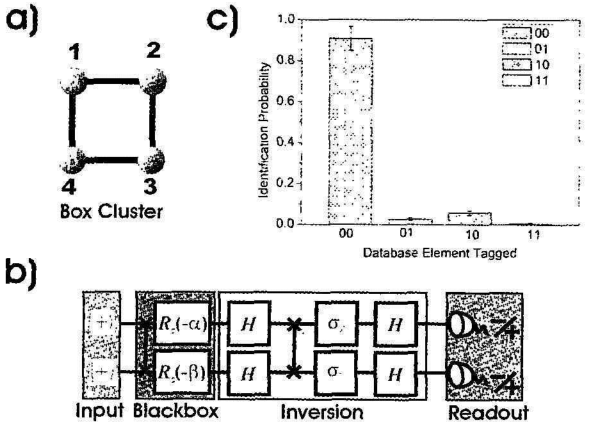


**Figure 10**

Grover's algorithm in a one-way quantum computer. *a)* The box cluster of four physical qubits was realized in the polarization state of four photons. The box cluster can implement the quantum search algorithm shown in *b)* on two input qubits. The circuit implements the labeling of one of the four inputs, denoted by the Blackbox. The search process itself is the operation "inversion about the mean", where quantum interference allows to find the labeled input by asking only once. The experimentally measured outputs of this quantum computation are shown in *c)*. For the *00* input state, there are four possible outcomes of the computation. The correct result is obtained on the average in 91% of the cases.

## Concluding Comments

I would like to conclude by expressing my astonishment over the fact that our early work on the foundations of quantum mechanics apparently helped to open up a new avenue towards novel ways of communication and computation. It is my personal conviction that right now we see a new information technology in the making. Nevertheless, the fascination with the mysteries of quantum mechanics still remains. Even there, the experiments helped to get a deeper understanding. With my student and now colleague, Caslav Brukner, I was able to develop a new understanding of the Copenhagen interpretation where we suggest that information is *the* fundamental notion of quantum mechanics.[56,57,58,59]

## References:

---

[1] H. Rauch, W. Treimer, U. Bonse, *Test of an single crystal neutron interferometer*, Phys. Lett. 47 A, 369 (1974).

[2] H. Rauch, A. Zeilinger, G. Badurek, A.Wilfing, W. Bauspiess, U. Bonse, *Verification of Coherent Spinor Rotation of Fermions*, Phys. Lett. A **54**, 425-7 (1975).

[3] G. Badurek, H. Rauch, A. Zeilinger, W. Bauspiess, U. Bonse, *Phase Shift and Spin Rotation Phenomena in Neutron Interferometry*, Phys. Rev. D **14**, 1177 (1976).

[4] A. Zeilinger, *General Formulation of Spin Rotations in Neutron Interferometry*, Zeitschrift für Physik **B 25**, 97-100 (1976).

[5] M. A. Horne, A. Zeilinger, *Fizeau Effects for Thermal Neutrons*, International Workshop on Neutron Interferometry, Grenoble 1978, Proceedings, Oxford University Press, 350-4 (1979).

[6] R. Gähler, A.G. Klein, A. Zeilinger, *Neutron Optical Test of Nonlinear Wave Mechanics*, Phys. Rev. A **23**, 1611 (1981).

[7] A. Zeilinger, R.Gähler, C.G.Shull, W.Treimer, W.Mampe, *Single and Double Slit Diffraction of Neutrons*, Rev. Mod. Phys. **60**, 1067-73 (1988).

[8] M. Gruber, K. Eder, A. Zeilinger, R. Gähler, W. Mampe, W. Drexel, *A Phase-Grating Interferometer for Very-Cold Neutrons*, Phys. Lett. A **140**, 363 (1989).

[9] D.W. Keith, C.R. Ekstrom, Q.A. Turchette, D.E. Pritchard, „An Interferometer for Atoms“, Phys. Rev. Lett. **66** (1991).

[10] E.M. Rasel, M.K. Oberthaler, H. Batelaan, J. Schmiedmayer, A. Zeilinger, *Atom Wave Interferometry with Diffraction Gratings of Light*, Phys. Rev. Lett. **75**, 2633-7 (1995).

[11] M.K. Oberthaler, R. Abfalterer, S. Bernet, J. Schmiedmayer, A. Zeilinger, *Atom Waves in Crystals of Light*, Phys. Rev. Lett. 77, 4980-83. (1996).

[12] S. Bernet, M.K. Oberthaler, R. Abfalterer, J. Schmiedmayer, A. Zeilinger, *Coherent Frequency Shift of Atomic Matter Waves*, Phys. Rev. Lett. 77, 5160 (1996).

[13] M. Arndt, O. Nairz, J. Voss-Andreae, C. Keller, G. Van der Zouw, A. Zeilinger, *Wave-particle duality of C60 molecules*, Nature **401**, 680-682 (1999).

---

[14] L. Hackermüller, St. Uttenthaler, K. Hornberger, E. Reiger, B. Brezger, A. Zeilinger, M. Arndt, *Wave Nature of Biomolecules and Fluorofullerenes*, Phys. Rev. Lett. **91**, 090408 (2003).

[15] L. Hackermüller, K. Hornberger, B. Brezger, A. Zeilinger, M. Arndt, *Decoherence of matter waves by thermal emission of radiation*, Nature **427**, 711-714 (2004).

[16] A. Einstein, B. Podolsky, N. Rosen, *Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?* Physical Review **47**, 777-780 (1935).

[17] M.A.Horne, A. Zeilinger, *A Bell-Type Experiment Using Linear Momenta*, Symposium on the Foundations of Modern Physics, Joensuu, P. Lahti and P. Mittelstedt (Eds.), World Scientific Publ. (Singapore), 435-9 (1985).

[18] M.A. Horne, A. Shimony, A. Zeilinger, *Two-Particle Interferometry*, Phys. Rev. Lett. **62**, 2209 (1989).

[19] D.C. Burnham, D.L. Weinberg, *Observation of Simultaneity in Parametric Production of Optical Photon Pairs*, Phys. Rev. Lett. **25**, 84 (1970).

[20] C. O. Alley, Y. H. Shih, in *Proc. Symp. on Foundations of Modern Physics*, P. Lahti, P. Mittelstaedt, eds., (World Scientific, Singapore), p. 435 (1985).

[21] J. G. Rarity, P. R. Tapster, *Experimental violation of Bell's inequality based on phase and momentum*, Phys. Rev. Lett. **64**, 2495–2498 (1990).

[22] D. Greenberger, M.A. Horne, A. Zeilinger, *Going beyond Bell's Theorem*, "Bell's Theorem, Quantum Theory, and Conceptions of the Universe", M. Kafatos (Ed.), Kluwer, Dordrecht, 69-72 (1989).

[23] D.M. Greenberger, M.A. Horne, A. Shimony, A. Zeilinger, *Bell's Theorem Without Inequalities*, Amer. J. Phys. **58** (12), 1131-43 (1990).

[24] A. Zeilinger, M.A. Horne, D.M. Greenberger, *Higher-Order Quantum Entanglement*, Proceedings "Squeezed States and Quantum Uncertainty", College Park, D. Han, Y.S. Kim, W.W. Zachary (Eds.), NASA Conference Publication 3135, National Aeronautics and Space Administration, (1992).

[25] M. Zukowski, A. Zeilinger, M.A. Horne, A.K. Ekert, *Event-Ready-Detectors Bell Experiment via Entanglement Swapping*, Phys. Rev. Lett. **71**, 4287-90 (1993).

[26] M. Zukowski, A. Zeilinger, H. Weinfurter, *Entangling Photons Radiated by Independent Pulsed Sources*, Fundamental Problems in Quantum Theory, D.M.

---

Greenberger, A. Zeilinger (Eds.), Annals of the New York Academy of Sciences vol. 755, 91 (1995).

[27] P.G. Kwiat, K. Mattle, H. Weinfurter, A. Zeilinger, A.V. Sergienko, Y.H. Shih, *New High-Intensity Source of Polarization-Entangled Photon Pairs*, Phys. Rev. Lett. **75**, 4337-41 (1995).

[28] A.K. Ekert, *Quantum Cryptography based on Bell's Theorem*, Phys. Rev. Lett 67, 661 (1991).

[29] K. Mattle, H. Weinfurter, P.G. Kwiat, A. Zeilinger, *Dense Coding in Experimental Quantum Communication*, Phys. Rev. Lett. 76, 4656-59 (1996).

[30] C.H. Bennett, S.J. Wiesner, *Communication via One- and Two-Particle Operators on Einstein-Podolsky-Rosen States*, Phys. Rev. Lett. 69, 2881-2884 (1992).

[31] P. Kwiat, H. Weinfurter, T. Herzog, A. Zeilinger, *Interaction-Free Measurement*, Phys. Rev. Lett. **74**, 4763-6 (1995).

[32] A. C. Elitzur, L. Vaidman, *Quantum Mechanical Interaction-Free Measurements*, Found. Phys. **23**, 987 (1993).

[33] P.G. Kwiat, H. Weinfurter, A. Zeilinger, *Quantum Seeing in the Dark*, Scientific American **275**, Nr. 5, 52-58 (1996).

[34] Bennett C.H., G. Brassard, C. Crépeau, R. Josza, A. Peres, W. K. Wootters, Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels, Phys. Rev. Lett. **70**, 1895 (1993).

[35] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, A. Zeilinger, *Experimental Quantum Teleportation*, Nature **390**, 575-579 (1997).

[36] J.-W. Pan, D. Bouwmeester, H. Weinfurter, A. Zeilinger, *Experimental Entanglement Swapping: entangling photons that never interacted*, Phys. Rev. Lett. **80**, 3891-94 (1998).

[37] M. Zukowski, A. Zeilinger, M. A. Horne, H. Weinfurter, *Quest for GHZ states*, Int. Conference on 'Quantum Future', Wroczlaw, Poland, 1997, Acta Phys. Polonica **93**, 187 (1998).

[38] D. Bouwmeester, J.-W. Pan, M. Daniell, H. Weinfurter, A. Zeilinger, *Observation of Three-Photon Greenberger-Horne-Zeilinger Entanglement*, Phys. Rev. Lett. **82**, 1345-49 (1999).

[39] J.-W. Pan, D. Bouwmeester, M. Daniell, H. Weinfurter, A. Zeilinger, *Experimental test of quantum nonlocality in three-photon Greenberger-Horne-Zeilinger entanglement*, Nature **403**, 515-519 (2000).

---

[40] S. J. Freedman, J. S. Clauser, *Experimental Test of Local Hidden-Variable Theories*, Phys. Rev. Lett. **28**, 938-941 (1972).

[41] A. Aspect, P. Grangier, G. Roger, *Experimental Tests of Realistic Local Theories via Bell's Theorem*, Phys. Rev. Lett. **47**, 460 (1981); **49**, 91 (1982).

[42] T. Jennewein, U. Achleitner, G. Weihs, H. Weinfurter, A. Zeilinger, *A Fast and Compact Quantum Random Number Generator*, Rev. Sci. Instr. **71**, 1675-1680 (2000).

[43] A.K. Ekert, *Quantum Cryptography based on Bell's Theorem*, Phys. Rev. Lett **67**, 661 (1991).

[44] T. Jennewein, Christoph Simon, Gregor Weihs, H. Weinfurter, A. Zeilinger, *Quantum Cryptography with Entangled Photons*, Phys. Rev. Lett. **84**, 4729–4732. (2000).

[45]A. Poppe, A. Fedrizzi, R. Ursin, H. R. Böhm, T. Lörunser, O. Maurhardt, M. Peev, M. Suda, C. Kurtsiefer, H. Weinfurter, T. Jennewein, A. Zeilinger, *Practical quantum key distribution with polarization entangled photons*, Opt. Express **12**, 3865-3871 (2004).

[46] Č. Brukner, J.-W.. Pan, C. Simon, G. Weihs and A. Zeilinger, *Probabilistic instantaneous quantum computation*, Phys. Rev. A **67**, 034304 (2003).

[47] M. Aspelmeyer, H.R. Böhm, T. Gyatso, T. Jennewein, R. Kaltenbaek, M. Lindenthal, G. Molina-Terriza, A. Poppe, K. Resch, M. Taraba, R. Ursin, P. Walther, A. Zeilinger, *Long-Distance Free-Space Distribution of Quantum Entanglement*, Science **301**, 621-623 (2003).

[48] K.J. Resch, M. Lindenthal, B. Blauensteiner, H.R. Böhm, A. Fedrizzi, C. Kurtsiefer, A. Poppe, T. Schmitt-Manderbach, M. Taraba, R. Ursin, P. Walther, H. Weier, H. Weinfurter, A. Zeilinger, *Distributing entanglement and single photons through an intra-city, free-space quantum channel*, Optics Express **13**, 202 (2005).

[49] R. Ursin, T. Jennewein, M. Aspelmeyer, R. Kaltenbaek, M. Lindenthal, P. Walther, A. Zeilinger, *Quantum teleportation across the Danube*, Nature - brief communications **430**, 849 (2004).

[50] E. Knill, R. Laflamme, G.J. Milburn, A scheme for efficient quantum computation with linear optics. Nature **409**, 46-52 (2001).

[51] K. Sanaka, T. Jennewein, J.-W. Pan, K. Resch, A. Zeilinger, *Experimental Nonlinear Sign Shift for Linear Optics Quantum Computation*, Phys. Rev. Lett. **92**, 017902 (2004).

---

[52] S. Gasparoni, J.-W. Pan, P. Walther, T. Rudolph, A. Zeilinger, *Realization of a photonic CNOT gate sufficient for quantum computation*, Phys. Rev. Lett. **93**, 020504 (2004).

[53] J.-W. Pan, S. Gasparoni, R. Ursin, G. Weihs, A. Zeilinger, *Experimental entanglement purification of arbitrary unknown states*, Nature **423**, 417-422 (2003).

[54] R. Raussendorf, H.-J. Briegel, *A one-way quantum computer*, Phys. Rev. Lett. **86**, 5188 (2001).

[55] P. Walther et al., "Experimental one-way quantum computing", Nature **435**, 169-176 (2005).

[56] A. Zeilinger, *A Foundational Principle for Quantum Mechanics*, Found. Physics **Vol. 29** no. 4, 631-643 (1999).

[57] Č. Brukner, A. Zeilinger, *Operationally Invariant Information in Quantum Measurements*, Phys. Rev. Lett. **83**, 3354-3357 (1999).

[58] Č. Brukner, A. Zeilinger, *Information and Fundamental Elements of the Structure of Quantum Theory*, in *Time, Quantum, Information*, edited by L. Castell and O. Ischebeck (Springer, 2003).

[59] Hans Christian von Baeyer, *In the beginning was the bit*, New Scientist **169**, 26 (February 2001).

PRINTED BY
KING FAISAL CENTR FOR RESEARCH AND ISLAMIC STUDIES