Technical University of Denmark

Changing energy systems to prevent climate change

Halsnæs, Kirsten; Larsen, Hans Hvidtfeldt

Published in:
Engineering challenges

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Halsnæs, K., & Larsen, H. H. (2009). Changing energy systems to prevent climate change. In C. B. Hansen (Ed.), Engineering challenges: energy, climate change & health (pp. 138-147). Kgs.Lyngby : Technical University of Denmark (DTU). (DTU research series).

DTU Library
Technical Information Center of Denmark

Engineering challenges

energy, climate change & health

DTU

Engineering challenges

energy, climate change & health

DTU

Engineering challenges

energy, climate change & health

Preface

Action, Boldness & Creativity

Action, boldness and creativity are the ABC of engineering.

The refuge of academia, around which we all must stand guard, is a prerequisite for the free flow of thought, the critical solitary reflection and the inner secluded contemplation from which new ideas are born. Ideas of the individual that must be integrated with the ideas of others to form groundbreaking trains of thought, not only lifting the spirit of humanity but also bettering the human condition, indeed giving promise of a better future for humankind.

The academic refuge can of course be abused in wasteful daydreaming, idle pastimes or even academic arrogance. All of us who detest such abuses must, in our defense of academic freedom, continuously fight to secure its ideals. However, even uncompromised, academic freedom cannot ensure the delivery of what academia can offer society, neither to the academic society nor – more importantly – to the society at large. Especially in science and technology, more, much more, is required. This "much more" is not to be solicited from society; we must solicit it from ourselves in the form of dedicated action. Passive reflection must be paired with active observation; interaction between theory and practice must be promoted; iteration between building models and experimental design must be employed; and dialogue between deduction and induction must be practiced.

To be constructive, these endeavors require not only the abilities of the professionally trained researcher and the eager curiosity of the natural-born scholar but also a certain boldness. If limits are not tested, if well-known truths are not challenged, if mental faculties are not exercised vigorously, then the challenging questions posed – even the most important ones – are reduced to being solely academic.

Scientific independence, a sibling to academic freedom, is an intellectual ideal very sensitive to its circumstances, and it may similarly wither away if it is left to self-indulgence and vain pride.

Science and technology as research fields must responsibly and ethically live up to the highest standards of academic freedom and scientific independence to remain the justly admired disciplines they are. Nevertheless, the intercourse between science and technology bears a much larger promise. Adding genuine creativity to action and boldness brings us to the pinnacle of engineering.

Perhaps the fascinating wonders of engineering can only be grasped if the ABC ingredients are recognized and understood in this context. Further, perhaps this insight is also the key to realizing why the wonders of engineering harbor the only possible answer to the pressing challenges posed by today's mounting problems and the only credible hope for a sustainable and livable future on this planet.

I hope the readers of this book will be impressed, as I have been, by how DTU researchers actually unfold the ABC of engineering. We will join forces with colleagues from all over the world to meet the engineering challenges of tomorrow, especially those related to energy, climate change and health.

Lars Pallesen
President

Contents   Page

Marine structures:
consuming and producing energy

Preben Terndrup Pedersen
& Jørgen Juncher Jensen

Seventy percent of the surface of the Earth is water. This space is immensely important for maintaining a reasonable standard of living for an ever-increasing population. Utilization of the sea is vital for the following reasons.

- The sea has huge importance for global exchange of goods; current global trade could not take place without using oceans for transport.
- The sea has great potential for the exploration and future cultivation of living resources in the form of fish and plankton.
- The sea and the seafloor have large reservoirs of raw materials and hydrocarbons.
- The sea receives 70% of Earth's primary sustainable energy source: radiation from the sun. This thermal energy can be harvested in the form of thermal, wind, current or wave energy, salt gradients and other forms.

Exploiting this potential requires marine structures.
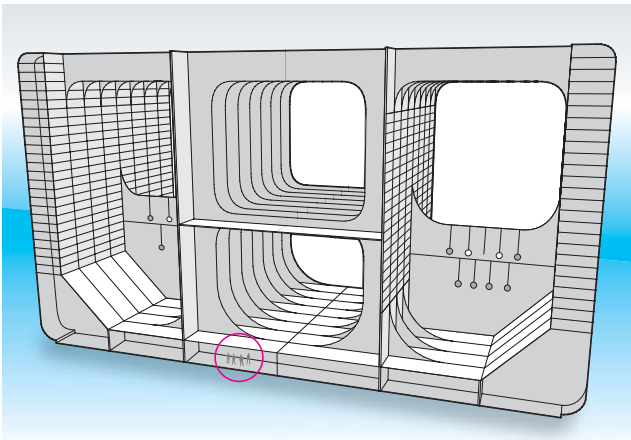
Fig. 1.1. Cross-section of a very large crude oil carrier

Shipping: energy and environment

Ships are by far the largest and most complex mobile structures humans build. Their mission determines their size and main dimensions. In addition to the basic functional considerations influenced by cargo and routes, such requirements as low resistance, high propulsive efficiency, navigational limitations on draft and beam and stability influence the choice of dimensions and layout. Ship structures are therefore under constant develop-

ment. Due to progress in research and continuing efforts towards more optimal solutions, ship transport has become the most environmentally friendly form of transport measured in tons per kilometer. An example is the very large crude oil carriers with cargo capacity between 220,000 and 320,000 tons (Fig. 1.1). For these vessels, the mass of the hull plus machinery and equipment is less than 15% of the mass of the oil carried (deadweight). No other means of transport has such a low ratio between self-weight and deadweight.

Ships transport about 90% of global trade in raw materials and finished goods, accounting for about 4% of total CO_2 emissions. In this global picture, Danish shipping plays an important role. Ships registered in Denmark carry 10% of global trade measured in value. Even though marine transport is relatively very energy efficient, as demonstrated above, the ships registered in Denmark emit as much CO_2 as the rest of Denmark.

The energy efficiency (grams of CO_2 per ton-mile) strongly improves not only by reducing speed but also by increasing the size of the vessel (Fig. 1.2). Larger vessel sizes also strongly reduce the capital cost and the personnel cost per deadweight ton.

This has been a main driver in the trend towards larger and larger ships. Container ships are an example. The first specially designed container ship appeared in 1960 and could carry 610 20-foot equivalent units (TEU). In 1988, the size of container ships increased to about 4500 TEU, requiring transcending the existing maximum width of the Panama Canal (Panamax). Odense Lindø Shipyard was the first to design and build a series of post-Panamax-size ships that could carry about 8000 TEU. These ships were the largest container ships until the delivery in 2006 of Emma Maersk, also from Odense Lindø Shipyard, as the first in a series of seven sister ships. Emma Maersk can carry more than 11,000 TEU; the deadweight is 157,000 tons, the main engine 80,080 kW and the maximum speed about 25.5 knots. These ships are the largest container vessels (length 398 m, width 56.4 m, depth 30.0 m) in the world fleet and have an unrivalled fuel efficiency index of 125 grams of CO_2 per TEU-mile.

The structural design of such large vessels poses several technical challenges. The size has increased

Container ships
Service engine loading = 85% of maximum continuous rating

Legend:
- Panamax – normal speed
- Panamax – 10% speed reduction
- Panamax – 20% speed reduction
- Post-panamax – normal speed
- Post-panamax – 10% speed reduction
- Post-panamax – 20% speed reduction

Y-axis: Energy efficiency design index (grams of CO_2 per ton-mile)
X-axis: Scantling deadweight (tons)
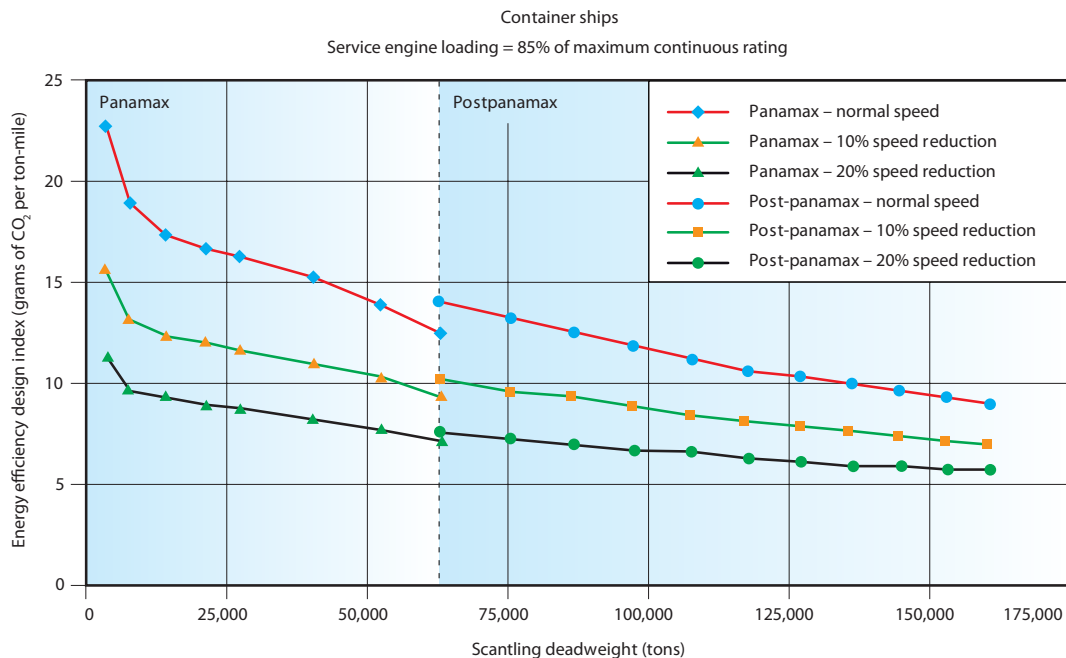
Fig. 1.2. Energy efficiency design index in grams of CO_2 per ton-mile as a function of size for different speeds. Panamax: maximum width of the Panama Canal.

so rapidly that there has been no feedback from service experience. Thus, these ultra-large steel structures have to be designed solely by direct calculation.

Several unsolved challenges are associated with the design based on the basic mechanical principles of such large and fast ships.

All the important structural loads on ships vary with time. The most important loads are caused by wave-induced motion of the water around the ship and the resulting motion of the ship itself.

When ships increase in size and materials with greater strength are used, hull flexibility plays an important role in the response of the vessel. The natural frequencies of the hull girder become so small that considering the ship hull as a rigid body is not sufficient in calculating the wave-induced loads. The effects of structural hull deformation become important. That is, several hydroelastic effects must be considered for ultra-large container ships.

Since container vessels have very open deck areas to facilitate the easy loading and unloading of containers below deck, the hull girders of these vessels have very low torsional rigidity. The hydrodynamic torsional loads may therefore cause con-

siderable torsional deformation when the ships are sailing in oblique waves. The result is deformation of the hatch openings and large stress where the deck beams are attached to the side shell and especially large stress in front of the accommodation module. Existing ships often develop fatigue cracks in these areas. Linear hydrodynamic hull girder loads can be based on calculations in regular sinusoidal waves, but no consistent mathematical procedures can predict the torsional stress of flexible hulls in stochastic seas.

Another serious wave-induced effect on the fatigue strength of flexible ship hulls is springing: a continuous excitation of the lowest natural frequencies of the hull girder due to the high-frequency components in the wave spectrum and due to nonlinear excitation effects. The natural frequencies of hull girders on ultra-large container ships are as low as 0.40 Hz (or 2.5 seconds in natural period), and springing can cause structural hull girder vibration in moderate seas with stress magnitudes that can cause significant fatigue damage to the hull within just a couple of years of operation. Again, no reliable procedure can predict these springing loads on large ships.

To reduce the resistance and to increase the deck

Fig. 1.3. Bow flare slamming and pressure distribution from computational fluid dynamics calculation
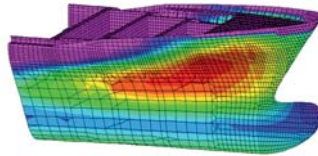
area, the hull forms of large container ships are unique, with excessive bow flare and stern overhang. The result is that slamming loads become crucial for the integrity of the ship hull in harsh weather conditions. The calculation results shown in Fig. 1.3 are for slamming impact in a regular si-

nusoidal wave. No consistent procedure has yet been developed to predict the long-term distribution of slamming loads in irregular waves.

Slamming loads cause large local impact forces on the hull, which may set up the plating and buckle internal frames. A second significant effect of slamming is transient vibration of the entire hull. This slam-induced hull girder vibration is called whipping (Fig. 1.4). The combination of the still-water loads and the wave-induced loads on the hull and the vibration response of the hull girder to bow or stern slamming impact can induce bending moments and shear forces that may cause the hull to collapse. One such example is the large container ship Napoli, which broke during a storm in the English Channel in 2006 due to whipping loads. This case has not been settled yet and poses a scientific challenge and a benchmark case for further developments.

Fig. 1.5 summarizes some of these new challenges associated with the structural design of large flexible container ships such as the Emma Class series.

Further research is strongly needed to design safe, large, fuel-efficient ships. Reliable probabilistic prediction of the maximum loads on large flexible ship structures during their lifetime operation in different seas is a major challenge that requires wave data, advanced nonlinear hydrodynamic codes, structural analysis models and advanced probabilistic tools.
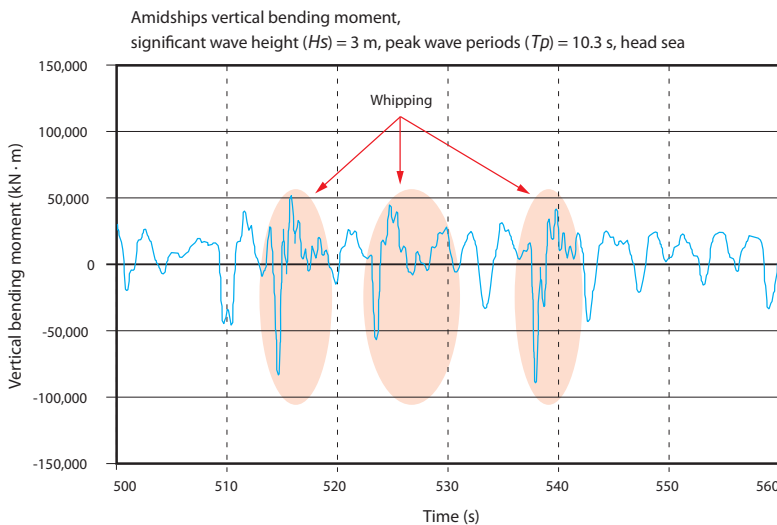
Fig. 1.4. Measured full-scale hull girder stresses showing slamming-induced hull girder whipping bending moments

Stern slamming

Fatigue stresses at hatch corners due to:
• Vertical bending
• Torsion
• Springing
• Whipping

Hatch-opening distortions due to wave-induced loads

Non-linear wave-induced stresses caused by:
• Vertical bending
• Torsion
• Springing
• Whipping

Whipping acceleration due to slamming
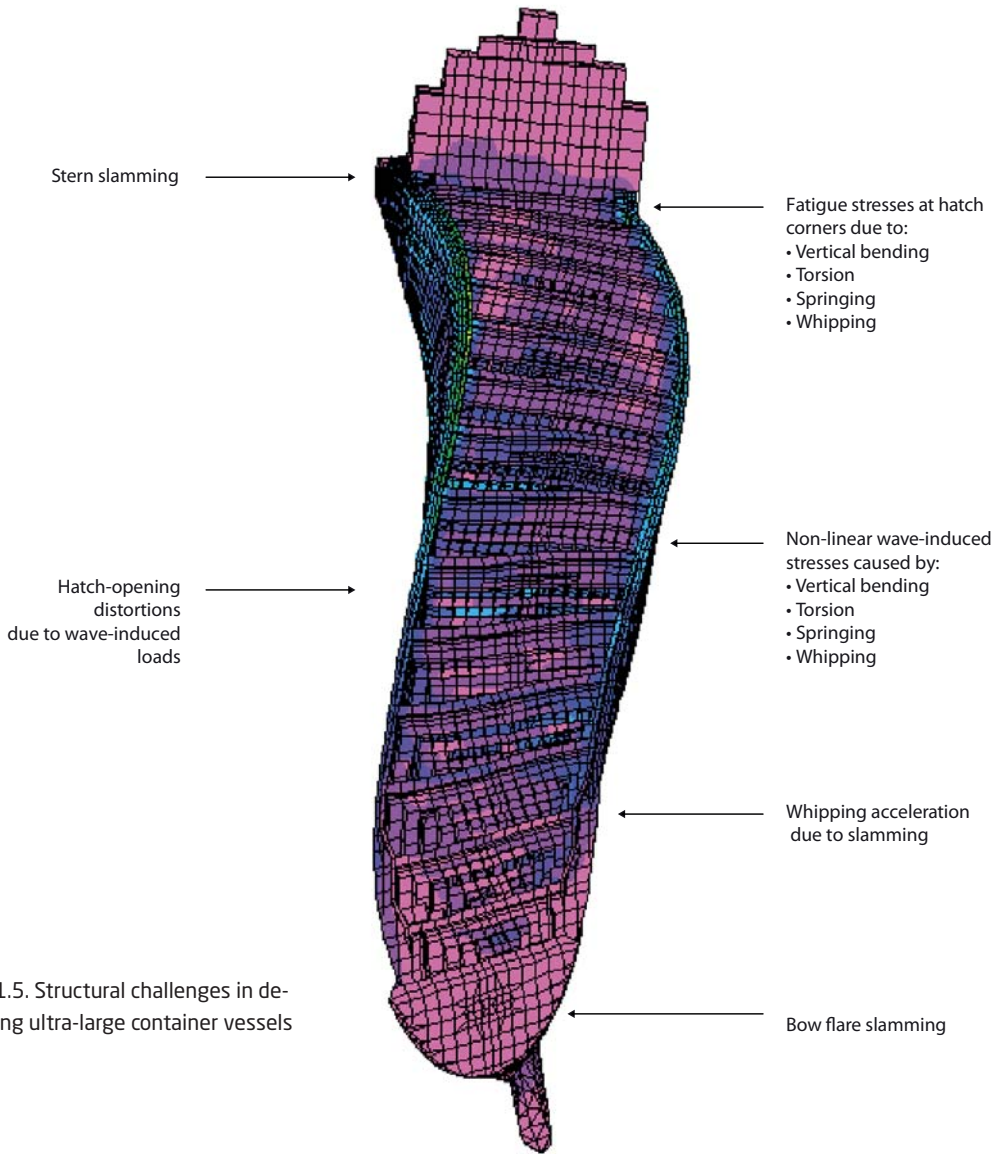
Fig. 1.5. Structural challenges in designing ultra-large container vessels

Bow flare slamming

The enabling analysis tools most needed include the following.

- *Improving computational procedures for calculating fluid dynamics*. These numerical calculation tools include free surfaces for predicting, for example, the flow field around the hull in calm water as well as in waves, calm water resistance and scale effects, viscous and scale effects on wave-making, propeller inflow, propeller torque, thrust and cavitation phenomena, seakeeping ability, wave-induced pressure distributions in nonlinear confused sea, slamming pressures and maneuvering.

- *Developing consistent theory on fluid structure interaction*. Consistent procedures are needed for calculating coupling between impulsive sea loads and structure, the influence of hull flexibility on wave-induced loads, the structural effect of sloshing in tanks and wave-induced springing excitation of ship hulls.

- *Procedures for nonlinear probabilistic design*. An important characteristic of sea loads is their variability over time. Like the sea itself, the loads imposed by the sea are random in nature and can therefore only be expressed in probabilistic terms. Determining a single value for the maxi-

mum loading to which a ship will be exposed during, for example, 25 years of operation in the north Atlantic is therefore not generally possible. The loads must be probabilistically represented instead. Further, because of the complexity of the structure itself and the limitations of analysis capability, the structural response of the structure cannot be absolutely accurately predicted even if the loading were known exactly. The design process must be based on reliability-based methods, and the degradation of the structures over time due to corrosion, fatigue cracks and other factors must be considered.

Analytical tools within these research fields need to be developed further to ensure the safe design of completely new concepts such as ships for transporting CO_2 from power plants to the deep seafloor or offshore oil and gas fields and for the design of a possible future generation of even larger ships such as Malacca-max vessels carrying 18,000 TEU.

As mentioned above, the most effective way to reduce CO_2 emissions is to increase the ships' size and to reduce speed. However, additional measures can be applied for any given main dimensions and speed. Fig. 1.6 depicts several current research and development fields that explore opportunities for further reducing fuel consumption and emissions.

The development of computational fluid dynamics in combination with tank testing has made progress in optimizing hull shapes to reduce calm water resistance. Other means of reducing the calm water drag resistance currently being analyzed include air lubrication of the ship bottom and new types of paint. Research on propellers has led to new optimized propeller shapes, devices to recover the rotational losses from the propeller stream and improvements in the water flow to the propeller. Computational fluid dynamics is also being used to explore alternatives to propeller propulsion such as fin propulsion and optimal strategies to use the wind as a possible auxiliary propulsion system in future ships either as rigid wings on the deck or as high-flying kites attached to the foremast.

Even if the large two-stroke engines have very high thermal efficiency (about 50%), further fuel efficiency can still be gained. One possibility is to install a system to utilize the exhaust heat to generate power for a shaft generator.

The integrated potential of technical solutions such as those mentioned above for improving energy efficiency and thereby reducing fuel consumption and CO_2 emissions is estimated to be 15-20% for existing ships.

Kite-assisted propulsion

Cargo and port operations

Operations decision support

Emissions

Bow design

Hull design
Anti-fouling paint
Drag reduction
Air lubrication

Main engine
Auxiliary engine
Shaft generator

Aft hull design
Propeller
Rudder
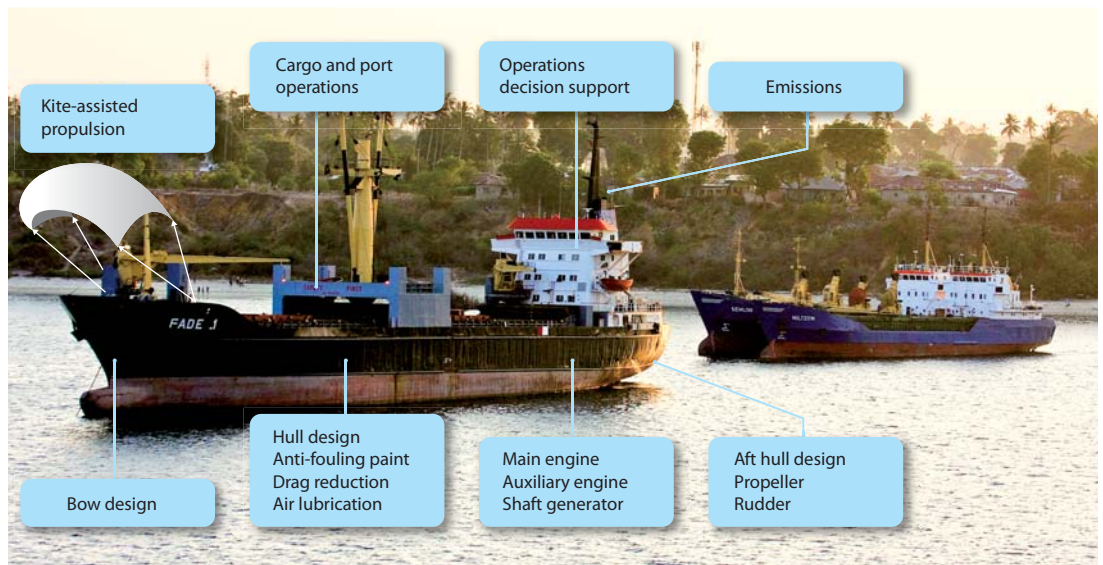
Fig. 1.6. Design elements for improving environmental performance with any given main dimensions
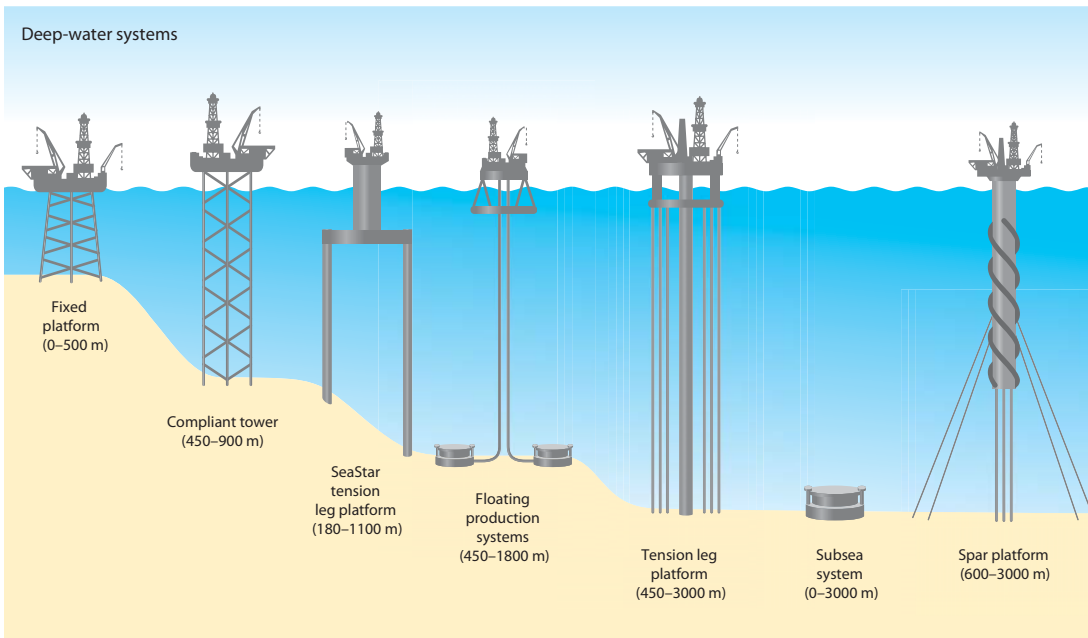
Fig. 1.7. Concepts for deep-water oil and gas production platforms

Offshore oil and gas production

For the next many years, oil and gas will still be the main energy resources. Extracting the newly discovered oil and gas resources situated below the seabed at great water depths or in Arctic waters constitutes an enormous technological challenge.

Offshore oil and gas production already comprises a significant portion of total production. Oil can currently be exploited at water depths up to about 3000 m (Fig. 1.7) using either tension leg platforms or spar platforms (the name spar comes from the mast analogy for sailing boats). Both concepts are huge floating units and require focus on hydroelasticity, fluid-structure interaction and probabilistic design procedures.

The overall design goal of these platforms is to prevent direct resonance with dominant sea wave frequencies. Independent of the prevailing characteristic wave height, the wave energy is concentrated at wave periods between 5 and 25 s (Fig. 1.8). The tension leg platforms have natural vibration periods in the vertical direction of less than 5 s and exceeding 25 s in the horizontal plane; the spar platform has natural periods above 25 s for vertical and horizontal rigid body motions.

The tension leg platforms have good motion characteristics from the tendons in tension, reaching down and connected to the sea floor. The caisson of the tension leg platforms must have enough buoyancy to yield the necessary tension in the tendons in extreme waves. The tendons are typically made of high tensile steel tubes about 0.6-0.8 m in diameter with a wall thickness of 30-50 mm. The design of the tendons constitutes a main challenge, as fatigue damage may occur due to high-frequency resonant vibration, conceptually similar to springing vibration in ships. Because the tension leg platforms move horizontally, the analysis has to include nonlinear or at least second-order forces.

The spar platform is basically a long tubular column 25 to 50 m in diameter and 200-250 m long. The bottom of the tubular column is at a depth where the pressure fluctuation due to the prevailing waves is very small. A more standard slack mooring system can therefore be used to keep the platform in place. The stability is ensured by having the center of gravity well below the center of buoyancy, using ballast in the bottom of the platform. As the diameter is fairly large, vortex-induced vibration can be a problem and helical strakes (Fig. 1.7) are usually needed especially in areas with strong currents. Estimating the vortex-induced vibration is a design challenge being solved by computational fluid mechanics codes.
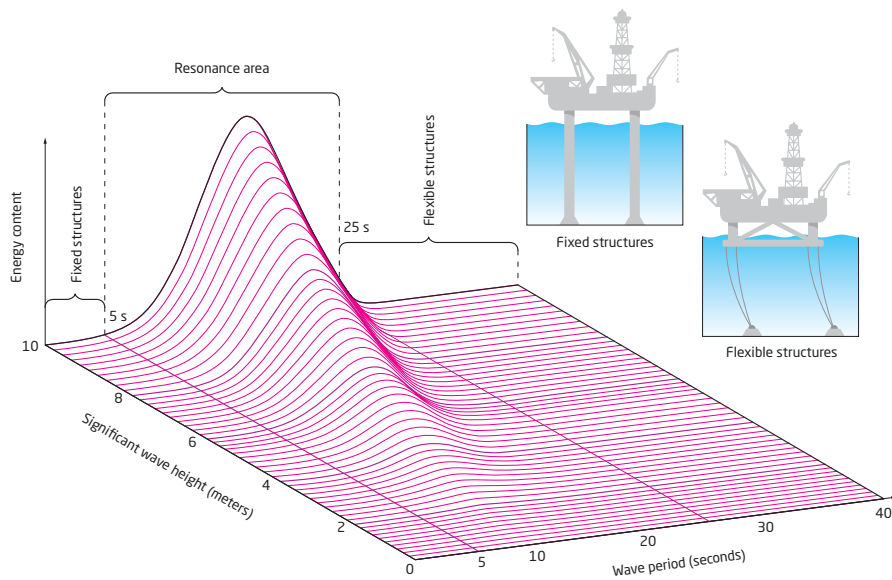
Fig. 1.8. Wave spectra for the north Atlantic

Properly estimating wave loads and wave load effects on these structures requires the inclusion of non-linear effects and, similar to ships, considering the randomness of the sea states. However, in contrast to ships, which are designed for worldwide trade, these offshore structures are designed for a specific location. Due to the high cost, wave buoys are typically used to collect detailed information on the wave environment at the intended location. A more accurate statistical description of the sea states can thereby be expected. The nonlinearities in the wave loads can often only be dealt with by time domain simulation in short-term sea states, which then has to be integrated into predicted long-term extreme values and estimated fatigue damage.

A direct procedure can be too time-consuming even considering the huge cost of such structures. Several approximate although rather accurate procedures have been developed and validated. For short-term analysis, the first-order reliability methods developed within structural mechanics have proved to be very useful for predicting stochastic wave and wind loads. One outcome of the first-order reliability methods is the most probable wave scenario leading to a certain response. This is illustrated in Fig. 1.9 for the extreme overturning moment for a jack-up drilling rig. This type of drilling rig is used for oil exploitation in water depths up to 120 m. Dynamic effects are important, as the fun-

damental period of the jack-up drilling rig at the maximum water depth and the zero-up crossing wave period are both about 9 s. The dynamic effects are responsible for the shape of the wave scenario showing the largest wave crest one period before overturning (at time = 60 s) and not, as in a quasi-static case, at the instant of over-turning. The critical wave scenario can be used in model testing to validate the calculation procedure, thus avoiding long model testing series with random waves.

The feasibility of producing oil from offshore Arctic fields is being intensively assessed. For areas covered by multi-year ice, gravity-based concrete structures are believed to be the most feasible solution for water depths up to about 60-80 m. For greater water depths, no feasible solutions exist today, as floating structures cannot be applied due to the ice. In areas with 1-year ice, steel structures as jack-up drilling rigs can be applied in areas with lighter ice. Subsea completion systems with wellheads situated on or below the seafloor are also being explored.

The main research challenge related to Arctic off-shore exploration is related to estimating the ice loads and the movements of the icebergs and ice sheets. Ice scour on the sea floor is also a serious concern, as it determines the necessary coverage of pipelines and subsea completion systems.
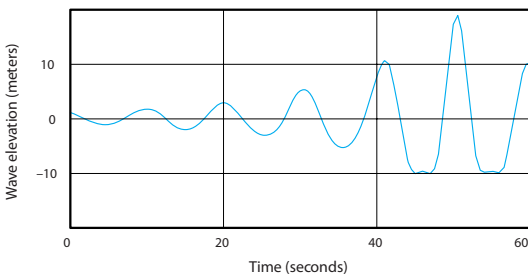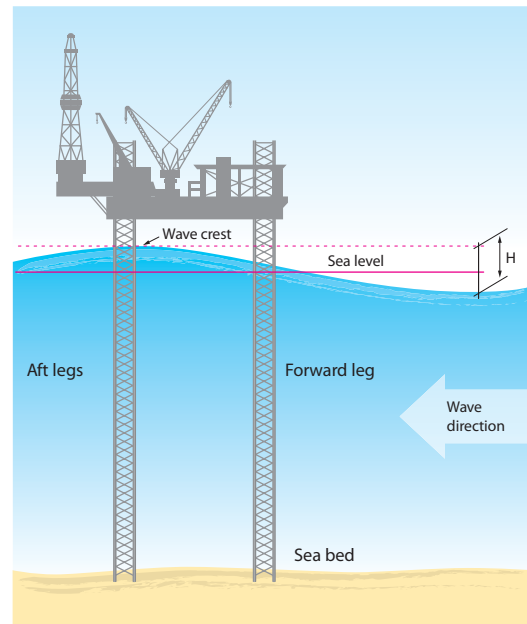
Fig. 1.9. Most probable wave scenarios leading to the overturning of a jack-up drilling rig at time = 60 s

Sustainable offshore energy production

Offshore wind turbines have multiplied rapidly in recent years. In shallow water, the standard monopile or jacket concepts from offshore oil production are usually applied. Scour and resonance frequencies are the main design problems but do not differ much from oil production platforms.

For deeper waters, floating offshore wind turbines are being considered, although no full-scale construction is operating. The first to come will probably be the spar concept (Fig. 1.10). The height of the cylinder is about 165 m, of which 100 m is below the sea surface. It will hold a 2.3-MW wind turbine and can be positioned at water depths between 120 and 700 m. A new challenge compared with the platforms used in the oil industry is re-

strictions imposed on the acceleration of the nacelle and on the yaw and pitch motions of the platform.

Other offshore wind turbine concepts being evaluated are based on the tension leg platform. Concepts based on floating islands hosting several wind turbines are being developed.

Finally, sustainable energy can be harvested from the waves. This field is still in infancy, and although many designs have been presented and evaluated in model scale, a real breakthrough awaits. One problem is that wave forces vary much more than wind forces. Designing wave energy devices such that they can cope with extreme wave loads and still be economically feasible for exploiting energy at the usual low wave heights thereby becomes a challenge. Another problem is related to the motion of such structures if they are floating. Generally, the motion will limit the effectiveness of systems based on waves running down a turbine. One way to circumvent this problem could be to combine floating wave energy devices with offshore wind turbines, such as the Poseidon plant (Fig. 1.11).
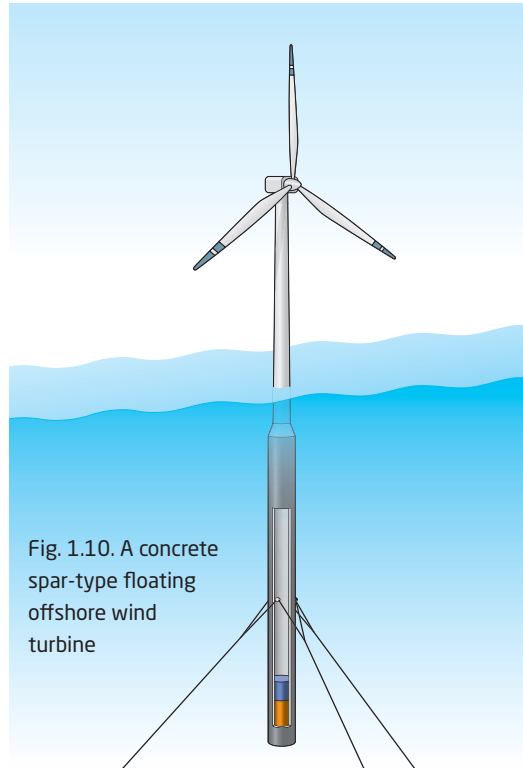
Fig. 1.10. A concrete spar-type floating offshore wind turbine
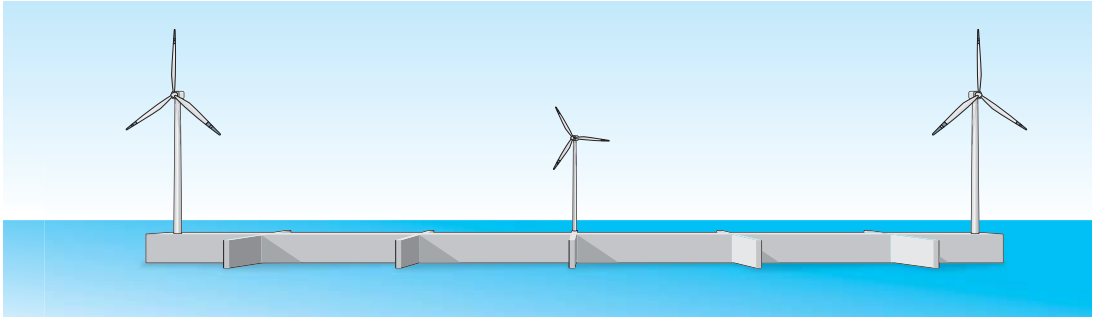
Fig. 1.11. A proposed project for combining offshore wind turbine and wave energy plants

However, such large structures have quite substantial wave loads and risk colliding with ships. Finally, marine growth on the water intake of such devices might pose serious maintenance problems.

The future

Climate change poses both risks and opportunities to all parts of the maritime sector. Since climate change requires urgent and extensive action by all parts of society to avoid the risk of serious damage to global prosperity and security, maritime research naturally strongly focuses on reducing emissions and on promoting new renewable energy sources.

Shipping in general is the most energy-efficient mode of transport, and transferring land-based transport of goods to water-based transport has substantial climate benefits. However, due to increasing volumes transported, greenhouse-gas emissions from maritime activities are projected to comprise a significant percentage of total human emissions in the future without technological improvements.

Oil and gas are still needed in the next 40 to 50 years, and an increasing proportion of the new di-
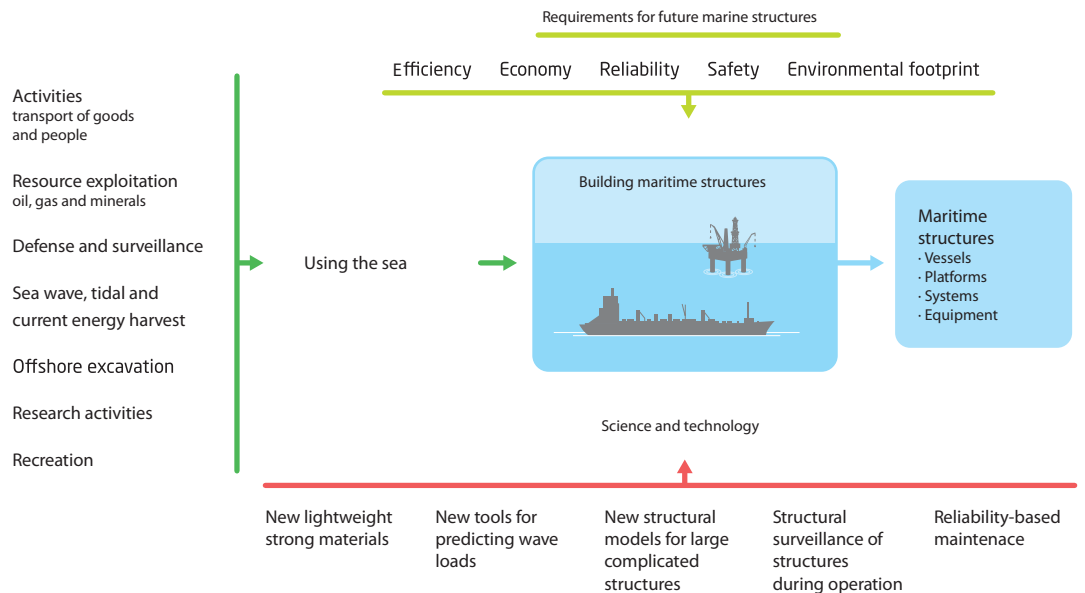
Fig. 1.12. Activities, requirements and scientific and technological disciplines that constitute the complex framework for the industrial exploitation of marine structures

coveries will be from offshore drilling in much harsher environments than are customary today.

Since the seas concentrate all the sustainable energy received from the sun, harvesting offshore wind energy and wind-driven wave energy will be emphasized more in the future.

All this technological development requires maritime structures. The following apply to maritime structures (Fig. 1.12).

- The sea is a special environment that poses unique requirements through loads, response and materials.
- Prototype testing is normally not possible, and analysis and design must be based on first principles.
- Loads and response can only be expressed in probabilistic terms.
- This is a global activity, and international laws and regulations govern it.
- The investment required is huge.

Denmark has a maritime cluster, the Blue Denmark, and the infrastructure and engineering background to participate in the future technological development of maritime structures. This could be a future growth field with immense importance for the global environment.

More to explore

Mansour A, Liu D. *Strength of ships and ocean structures*. Jersey City, NJ, Society of Naval Architects and Marine Engineers, 2008.

Jensen JJ. *Load and global strength*. Amsterdam, Elsevier, 2001.

Moan T. Marine structures for the future – a sea of opportunities. *Marine Systems and Ocean Technology*, 2005:1:1–19.

The authors

Preben Terndrup Pedersen: MSc (1966) and Lic. Techn. (1969), Technical University of Denmark. Professor in Strength of Marine Structures, DTU Department of Mechanical Engineering. Sabbaticals: Harvard University and DNV, Norway. Head, DTU Department of Mechanical Engineering, 2000–2007. Research: loads and response of ships and offshore structures, ship collisions and grounding, risk analysis.

Jørgen Juncher Jensen: MSc (1972) and Lic. Techn. (1975), Technical University of Denmark. Professor in Marine Engineering, DTU Department of Mechanical Engineering. Head, Section of Coastal Maritime and Structural Engineering. Sabbaticals: regularly visiting researcher at the University of California at Berkeley. Research: hydroelastic and stochastic response of marine structures.
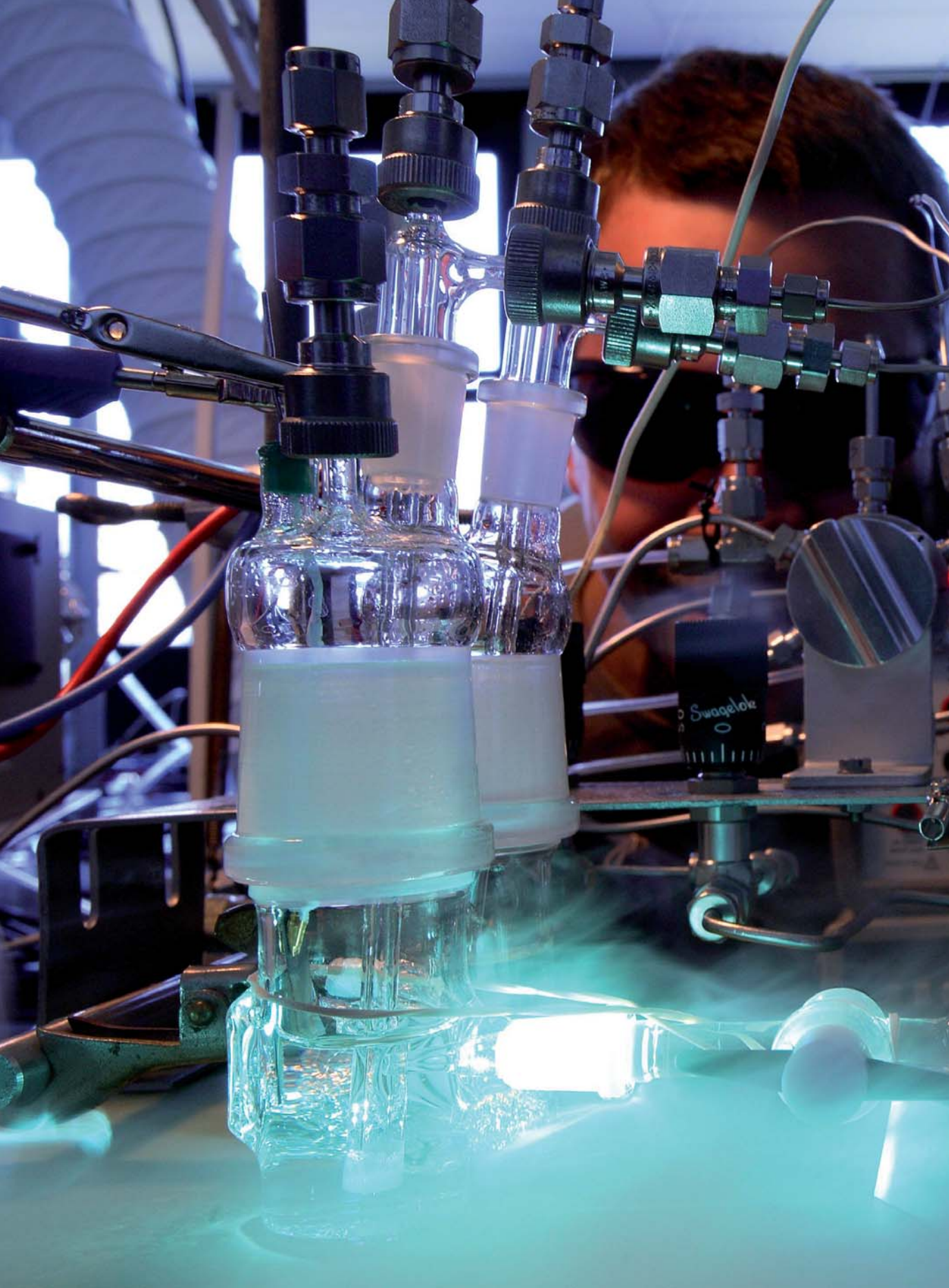
Sustainable energy

– catalysis is part of the solution

Ib Chorkendorff

Today most industrialized countries rely heavily on fossil fuel (more than 80%) for the energy for a modern and convenient lifestyle. A great challenge humanity faces is partly or fully shifting our energy supply from fossil fuels to sustainable energy sources. There are many reasons for this shift: lack of fossil resources, negative environmental effects, security of supply, independence and prices. Regardless of the reason, humanity living in balance with nature and trying to minimize negative environmental effects by promoting sustainability are desirable in the long term.

Solving these issues and those connected to fossil fuels requires new and improved catalysts: materials that increase the efficiency of the current energy technology and increase the energy output from sustainable sources such as wind turbines, biomass and solar energy. Catalysts thus play a key role in securing both future energy supplies and the environment. Improving and designing new catalysts are the major objectives for the DTU Center for Individual Nanoparticle Functionality.

Sustainable energy sources are limited and inefficient

So why did humanity not shift to sustainable energy long ago? First, sustainable energy sources are inefficient with today's technology. Denmark is currently getting 13% of its energy from sustainable resources (10% from biomass and 3% from wind energy). Although considerable increases in capacity are planned, they cannot cover Denmark's entire energy demand. Wind energy is unstable because the wind fluctuates. Energy from biomass, the energy stored in plants by photosynthesis, is also a limited source because photosynthesis is inefficient. Even under the best conditions, no more than 0.5% of the solar energy supplied to the plants can be harvested. Many researchers and engineers, both nationally and internationally, are therefore searching for other, more efficient and stable sustainable energy supplies. In this quest, they have found one shining source.

The Sun: an unlimited but expensive energy source

The Sun is an everlasting (at least on our time scale) and plentiful energy source. In less than 2 hours, the Earth receives enough solar energy to cover global human energy consumption for 1 year. This energy can be harvested as electricity by photovoltaic (solar) cells and used locally or transmitted to consumers farther away. Experiments with solar cells have demonstrated much higher efficiency per land area than can be obtained by growing biomass. If a realistic 10% efficiency is assumed, this only requires 5% of Denmark's land area and not necessarily farmland, enabling such a solution without influencing food production.

Solar cells, however, have several problems. The amount of sunlight arriving at the Earth's surface is not constant but depends on location, time of day and year and weather conditions. Further, solar cells are very expensive and thus currently not viable economically, although improved technology and rising energy prices may alleviate such challenges. The relationship between the price of fossil fuels and the viability of new sustainable but also more expensive energy technologies applies to all new energy approaches.

Both fossil fuels and more sustainable alternatives pose several limitations and concerns for which there are no perfect solutions. Fossil fuels are limited and harm the environment; sustainable energy sources are inefficient, expensive and unstable. Thus, the present use of fossil fuels needs to be improved, surplus sustainable energy needs to be stored and new renewable energy technologies need to be developed. This work of improvement and invention highly depends on the ability of catalytic materials to accelerate and lower the energy consumption of the chemical reactions involved in energy conversion and storage, and this requires new and improved catalysts. The following sections focus on the use of and research in catalytic materials for existing and future energy supplies (Box 2.1).

Fossil fuel reserves are still extensive

Fossil fuels will be the primary energy resource for decades to come since no economically competitive or sufficiently efficient sustainable energy technology can take over. The reported estimates of reserves amount to about 40 years of oil, 60 years of natural gas and hundreds of years of coal. These numbers have been fairly constant during the past

Box 2.1. What is catalysis?

A catalyst is a material that speeds up a chemical reaction without being consumed in the reaction itself. Most people benefit daily (and perhaps unknowingly) from the use of catalytic converters in cars that remove harmful emissions of CO and NO by facilitating chemical conversion into CO_2 and N_2 (Fig. 2.1).

This makes the reaction pathway from reactants to products much easier, and the reaction runs faster and uses less energy. Since chemical reactions only take place on the surface of materials, one way to improve a catalyst's efficiency is to increase its total surface area. The smaller the particles, the larger the fraction of the material is on the surface. Thus, materials in the shape of nanoparticles (on the order of 10^{-9} meters in diameter) have a much larger total surface area than the same amount of material in the shape of microparticles (on the order of 10^{-6} meters in diameter) and are thus much more efficient than microparticle catalysts for the same amount of metal.
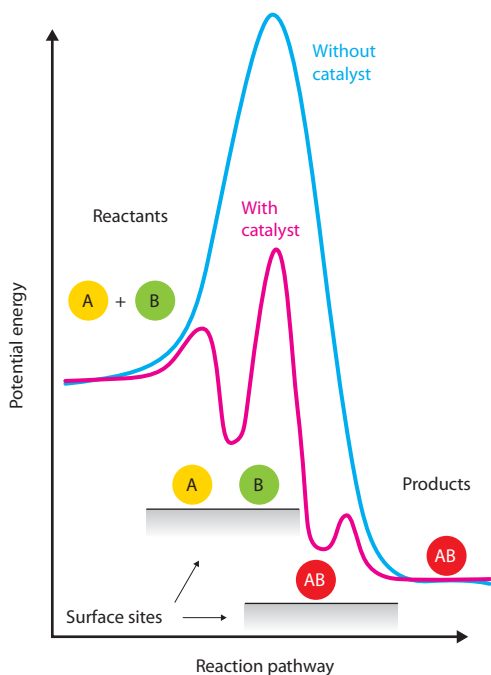
Catalysts have important roles in improving existing energy technologies and in creating new ones. Catalysis is involved in 20–30% of total economic production and plays an essential part in many industrial processes including pharmaceuticals, food, catalytic conversion and 90% of all chemical production. Unfortunately, a predominant problem is that most important catalysts comprise expensive and scarce materials such as platinum and palladium. Thus, improving catalytic processes requires finding less expensive and more abundant materials with the same or higher catalytic activity than current catalysts.

Fig. 2.1. Catalysts typically consist of active sites on metal or oxide particles. The catalyst reduces the energy needed for the chemical reaction by adsorbing the reacting gases (reactants) onto the active surface sites on the particle, thus bringing the gas molecules into closer contact.

decades, indicating that new resources are still being found. The argument of peak oil production is often raised, but the primary product produced from oil – gasoline for transport – may be produced from natural gas or even coal. As oil prices go up, catalytically converting first natural gas and later coal into liquid fuel will become profitable, extending gasoline resources well beyond 40 years. Thus, fossil resources are available, although the price will increase as the reserves shrink and they become increasingly difficult to extract.

Nanocatalysts improve existing technologies

Catalysis is the key technology for efficiently converting fossil fuels into useful chemicals or converting natural gas and coal into gasoline. As this needs to happen for many years to come, there are good reasons to improve and optimize these processes as much as possible. The processes require catalytic materials that have a huge surface area per mass to maximize the active area of metals that are often scarce and expensive. Nanoparticles are therefore the preferred size for catalytic materials.

Fig. 2.2. The experimental hall of the DTU Center for Individual Nanoparticle Functionality funded by the Danish National Research Foundation. The experimental facilities rely heavily on a surface science approach for studying the active surface of catalytic materials at the atomic level under well-controlled conditions while linking to real operating conditions. The hall hosts eight instruments specially designed for various investigation, each representing an investment of €400,000 to €1 million. The instruments are used for basic research, often in close collaboration with companies manufacturing or using catalysts such as Haldor Topsøe A/S and IRD A/S.

The shape and size of the nanoparticles are often essential for their functionality, and large particles may not work at all.

Optimizing catalytic reactions saves energy

Understanding what optimizes the nanoparticle reactivity for a specific reaction as a function of size and shape is a scientific challenge but also the key to optimizing many catalytic reactions and to designing new processes. The importance of this understanding is illustrated by recent work in which the Center for Individual Nanoparticle Functionality (Fig. 2.2) and Center for Atomic-scale Materials Design of the DTU Department of Physics closely collaborated with Haldor Topsøe A/S in studying two essential steps in hydrogen production both experimentally and theoretically in great detail.

Producing hydrogen is key in producing both liquid fuels such as gasoline and alcohols and ammonia (fertilizer). The world uses about 100 million tons of nitrogenous fertilizer per year, consuming at least 1% of total world energy production. Thus, improving the process of producing hydrogen is very significant.

Hydrogen is manufactured by converting methane (CH_4) and water into a mixture of CO and hydrogen by the steam-reforming process. This is also the first step in producing liquid fuels from natural gas and the route for producing large amounts of ultra-pure hydrogen for synthesizing ammonia. In the latter case, CO can be further converted by water into CO_2 and hydrogen through a low-temperature process called the water-gas shift reaction. CO_2 is subsequently removed, and hydrogen is the product. However, even parts per million (ppm) levels of residual quantities of oxo-compounds such as CO and CO_2 in the synthesis gas will harm the ammonia synthesis catalyst substantially and must therefore be removed by the inverse steam-reforming process. This converts both gases back to methane

$$
\begin{array}{lll}
CH_4 + H_2O & \longrightarrow & CO + 3H_2 \\
CO + H_2O & \longrightarrow & CO_2 + H_2 \\
CO + 3H_2 & \longrightarrow & CH_4 + H_2O \\
\\
CO_2 + 4H_2 & \longrightarrow & CH_4 + 2H_2O
\end{array}
$$

Steam-reforming process
Water-gas shift reaction
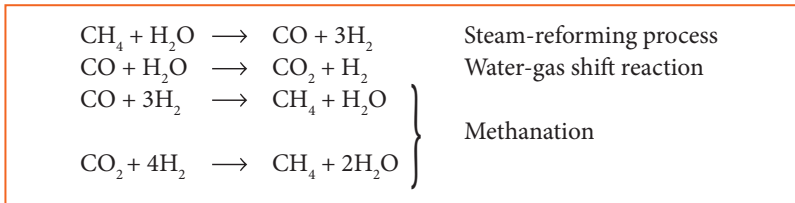
Methanation

and water, and then the oxygen can easily be removed from the gas. This process is called the methanation reaction and is represented by the last two reactions. Only a small proportion of the hydrogen is sacrificed in this manner to produce ultraclean hydrogen.

Shape and size matter

Catalysts are involved in all these steps. By lowering the activation energy, the catalyst makes the chemical reactions run faster and use less energy. We examined the methanation reaction and found that the first rate-limiting step of the reaction is the dissociation of CO into C and O atoms bound to the active site on the catalyst surface. Speeding up this step requires understanding exactly what influences the reactivity of the active site. Surprisingly, the reactivity turned out to depend on the geometry of the active site on the catalytic nanoparticle.

Instead of nanoparticles, single crystals are often used as model systems when studying catalytic reactions. A single crystal has all atoms positioned in identical surroundings in a crystal lattice. It is therefore a better model system than the more disordered nanoparticles for investigating the influence of the local atomic arrangement on the catalytic activity. Comparing measurements of CO dissociation on flat and stepped nickel (Ni) single-crystal surfaces accordingly (Fig. 2.3) clearly showed that the monoatomic steps play an essential role in the reaction. When the stepped crystal surface is exposed to a fixed dose of CO, a certain amount will dissociate, leading to carbon on the surface at a rate that can be monitored. However, when the step sites are blocked with sulfur, carbon formation is reduced, clearly demonstrating the importance of the step sites.

Theoretical calculations backed the experimental data, showing the exact same picture: CO cannot dissociate on the planar crystal terrace but need ensembles like those created along a monoatomic step to proceed. This observation should naturally also affect the real catalyst, and this was confirmed by investigations of the influence of the particle size and the reactivity on real catalysts measured at Haldor Topsøe A/S. These investigations of how the structure of the catalytically active site influences a step in the methanation reaction mainly illustrate the importance of identifying the active site on a catalytic material to determine exactly what to modify to improve the catalyst. This also illustrates another important point: fabricating catalytic nanoparticles requires not only making small particles but also particles with the right shape and size to optimize the catalytic activity of the sites essential for the reaction. This insight is very important for manufacturing new and more efficient catalysts, not only for ammonia production but for all catalytic reactions including more efficient conversion of fossil fuels.

Power after 4:00? Learn how to store

Most sustainable energy is in the form of electricity and fluctuates depending on sun, wind and other
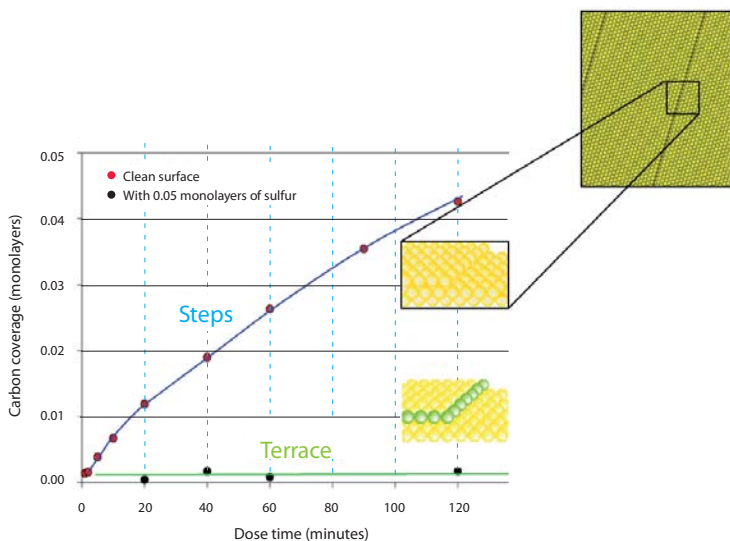
Fig. 2.3. The dissociation probability of CO on a flat and stepped surface of an Ni single crystal is measured as a function of dose time. The crystal is cut so the surface exposes a monoatomic step for every 26 rows of atoms (insert, upper right). The CO dissociation rate is measured as the amount of C deposited on the surface. The blue curve shows CO dissociation on unblocked steps and the resulting C deposition. When the steps are blocked by sulfur (green flat curve), no C and thus no dissociation is seen at all: the terrace atoms are not active. This observation clearly shows that steps play a crucial role in dissociating CO.

weather conditions. If sustainable energy is to constitute most energy in the future, ways to average out this source are strongly needed to secure a stable energy supply 24 hours a day. When there is a surplus of electricity due to favorable production conditions such as windy weather or low energy consumption at night, the energy needs to be stored. When electricity generation decreases or consumption increases, energy can be released from storage.

The simplest way of storing electrical energy is in batteries. Unfortunately, batteries are limited by capacity and charging time for such huge amounts of energy. Other storage methods than batteries must therefore be considered, preferably transforming electrical energy into energy stored in chemical bonds in liquid fuels such as alcohol or hydrogen. In theory, converting electrical energy into chemical by splitting water into oxygen and hydrogen is very simple. This is called electrolysis. The hydrogen produced can be stored and converted back to electricity when needed or, like natural gas, be transported out to the consumers and there oxidized back to water in a fuel cell, thereby generating electricity. This sounds very simple and tempting, especially if the two processes could take place in the same cell: a combined electrolysis (H_2O to H_2 and O_2) and fuel cell (H_2 and O_2 to H_2O). In this way, hydrogen could be produced, stored and used again right on the spot. From water to water – it does not get any better.

Inadequate fuel-cell efficiency

Unfortunately, such processes lose energy, and only about 30% of the energy is available to consumers. Understanding this requires examining how a proton-exchange membrane fuel cell operates (Fig. 2.4).

The major energy losses in this process appear because the hydrogen and oxygen on the anode and the cathode (respectively) are not ideally adsorbed and desorbed. Thus, the acid fuel cell only has an efficiency of 45%. Electrolysis is the reverse process, and alkaline electrolysis has an efficiency of about 70%. Combining the two types of cells leads to an overall efficiency of 30%, as mentioned above. Unfortunately, the two types are not compatible, and a cell therefore needs to be developed that can operate reversibly, reducing capital investment. This requires new efficient catalysts for the electrolysis.

Less expensive catalysts are needed

Fortunately, fuel cell efficiency can be improved (Fig. 2.5). In contrast to power plants, which have a maximal theoretical operating efficiency for electricity of about 45% at reasonable operating temperatures, the theoretical efficiency for hydrogen oxidation is about 90% at 80°C, which is the maximum operating temperature for a proton-exchange membrane fuel cell. The fuel cells are currently not much more efficient than the power plants, as indicated by the blue band in Fig. 2.5, but they are not limited by thermodynamics in the same manner.

1 Hydrogen fuel is fed to the anode on one side of the fuel cell while oxygen is channeled to the cathode on the other side.

2 At the anode, a platinum catalyst causes the hydrogen to split into positive protons and negatively charged electrons.

3 The proton-exchange membrane allows only the protons to pass through it to the cathode. The electrons must travel along an external circuit to the cathode, creating an electrical current.

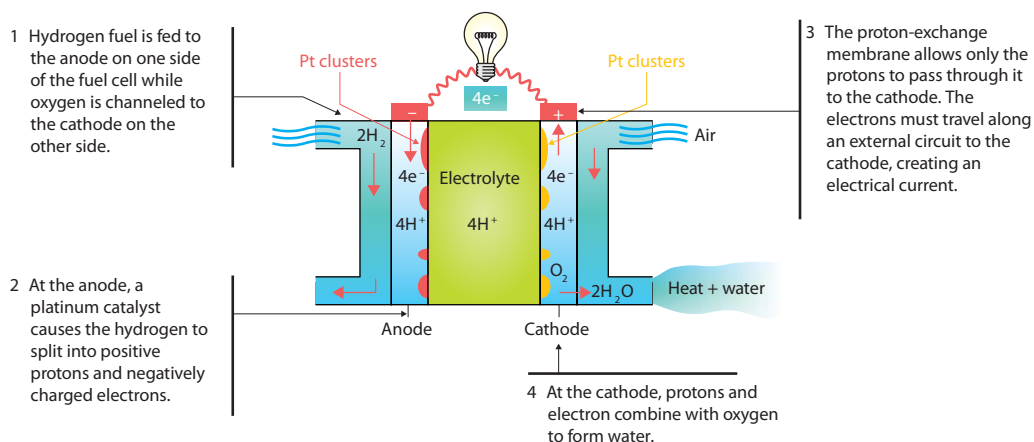4 At the cathode, protons and electron combine with oxygen to form water.

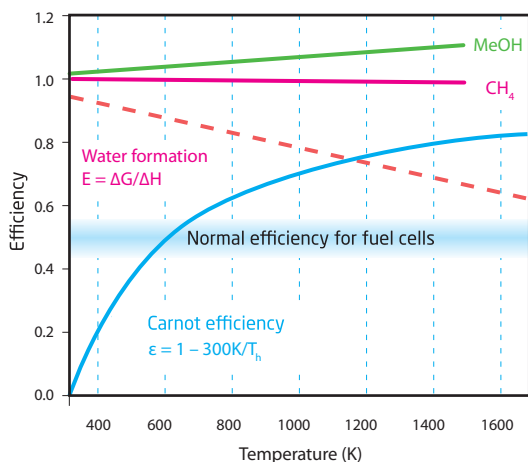Fig. 2.4. Main operating mechanism of a proton-exchange membrane fuel cell

Fig. 2.5. Theoretical efficiency of fuel cells and power plants. Low-temperature fuel cells have theoretical maximum efficiency at 80-90°C (~360 K), with efficiency declining as the temperature increases (red line). Power plant efficiency is restricted by the limitations of Carnot efficiency (blue line) and shows the opposite temperature relationship: increasing temperature raises the efficiency. Due to corrosion at rising temperatures, the operating efficiency of power plants is about 45%. In contrast, the thermodynamic limit is much higher in fuel cells, and better catalysts can achieve higher efficiency at low temperature. Currently fuel cells are not much better than power plants, as indicated by the blue band, but they have the potential for higher efficiency.

Approaching the theoretical maximum efficiency under real-life working conditions requires understanding the fundamental problems and improving the catalysts for these processes. The fundamental problem in fuel cells is that oxygen and hydrogen react too strongly with the catalyst, and energy is required to get them off the surface again. Similar to heterogeneous catalysis, a material is required that can provide appropriate active sites with the correct binding energy that lower the barriers on the reaction pathway. For electrochemistry and fuel cells, these barriers are called overpotential because higher potential (and subsequent loss of energy) may help overcome the barrier. Most of the overpotential (energy lost) is on the oxygen side, but the hydrogen side also has problems. We recently studied this effect for the hydrogen evolution process on the cathode side.

All the good catalysts are the platinum metals, in accordance with the fact that platinum is actually used as the catalyst in a proton-exchange membrane fuel cell. Platinum is rather good and close to the maximum, so there is not so much improvement to gain here, although we demonstrated that catalysis can be improved if bismuth can be alloyed into the platinum surface. The major problem, however, is relying on scarce and expensive metals such as platinum, palladium and rhodium. Current fuel cell technology requires roughly 0.5 grams of platinum per kW of capacity. Producing hydrogen-fueled cars with 100 kW of capacity requires that each car have fuel cells containing 50 grams of platinum. With the current platinum production of roughly 200 tons per year, only 4 million cars could be manufactured annually, an order of magnitude less than the current annual car production rate. Platinum is more expensive than gold, and extensive use will immediately increase the price. This indicates that relying on this technology not only for cars but also for averaging out sustainable energy supplies requires figuring out how to use platinum much more efficiently or designing catalysts based on more abundant elements.

Inspiration from nature

Inspired by nature, we recently investigated the catalytic activity of molybdenum. Plants and bacteria use enzymes to fixate nitrogen into ammonia, and in this process hydrogen is released through the action of an assisting coenzyme containing sulfides in combination with iron and molybdenum. In both theoretical and experimental approaches, we have found that several of these compounds are very active in evolving hydrogen. Similar to the catalytic nickel nanoparticles described earlier, the shape and size of molybdenum sulfide (MoS_2) nanoparticles greatly influences their catalytic activity (Fig.2.6). Both theoretical predictions and experimental results have demonstrated that the edges are the reactive sites. Identifying the active site is of fundamental importance, because it provides deep insight into how structure and reactivity are related and thereby how to design new improved catalysts.

MoS_2 is a semiconductor, but changing it into nanoparticles makes the edges conductive and thus
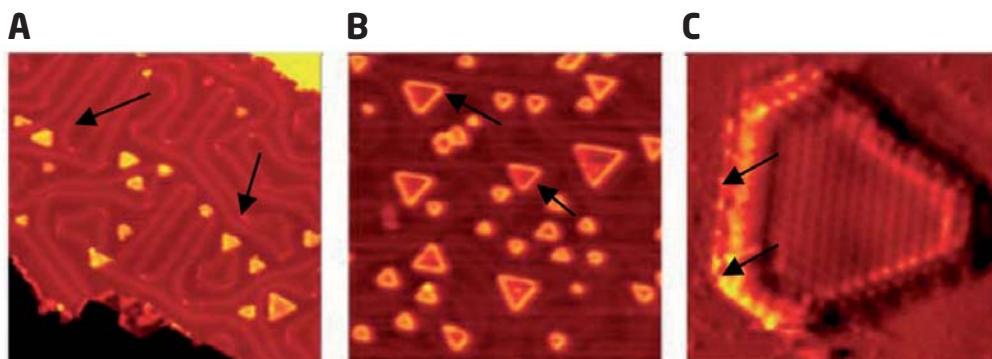
Fig. 2.6. Single layered S-Mo-S nanoparticles (arrows A and B) grown on a gold substrate and characterized on the nanoscale with scanning tunnel microscopy (STM). By varying the particle size, the number of sites on the top facet and on the edges can be controlled. By combining our nanoscale information with electrochemical measurements of the hydrogen evolution, it was possible to identify the edge sites as the active sites for hydrogen evolution. Images A and B are 47 nm by 47 nm, and image C is a close-up of 6 nm by 6 nm revealing the atomic arrangement and the important edge state (arrows C).

usable as an electrode for hydrogen evolution. Some overpotential (an activation barrier) still needs to be overcome to obtain a good and inexpensive substitute for platinum, but this work now enables improvement of the electrocatalyst to be predicted by making it bond hydrogen more strongly.

Photocatalysis

No single energy technology can meet the enormous global energy demand (about 14 TW and rising) while reducing the strain on the environment and fossil fuel reserves. A series of new energy technologies therefore needs to be improved and developed that can fulfill these needs. The preceding sections have described ways to improve existing fossil fuel–based technologies by catalysts that are more efficient. Further, we have indicated the usefulness of fuel cells as a way of securing stable sustainable energy supplies, again requiring catalysts that are more efficient. Finally, a third way to be explored is developing photocatalytic cells, in which sunlight directly splits water into hydrogen and oxygen (Fig. 2.7).

Semiconductor photoelectrocatalyst

The principle of a semiconductor photoelectrocat-

Fig. 2.7. The photocatalytic cell combines harvesting of light into electron-hole pairs and using them to split water into oxygen and hydrogen directly. This requires carefully designed materials to use a large part of the solar spectrum and a good catalyst for the electrochemical process where hydrogen and oxygen are evolved.

alyst is that solar radiation excites electron-hole pairs. In reality, this happens by one of the valence electrons absorbing a photon. This electron is excited into the conduction band, leaving a hole behind in the valence band. If the material is designed correctly, the electrons and holes can be separated to two different catalytic active surfaces. On one side – the anode – the holes oxidize H_2O into O_2 and H^+; on the other side – the cathode – the electrons reduce the generated H^+ into H_2: hydrogen evolution. Photocatalysis thus combines capturing the energy of sunlight and converting it into electricity to be used for electrolyzing water (Fig. 2.8). In theory, it is a combined version of a photovoltaic cell and subsequent electrolysis, except that this all takes place in one unit.

Many demands have to be fulfilled simultaneously to make a viable photocatalyst. First, it must efficiently absorb sunlight and generate electron-hole pairs. Second, the energy of the electron-hole

pair must be great enough to drive a chemical reaction such as water-splitting. Finally, sufficient energy must be available to overcome the significant overpotential to actually get the electrolysis process running at any appreciable rate. Several materials need to be designed to fulfill these requirements, although they lead to increased complexity and thus cost.

Better storage methods needed

Which form of energy is most suitable for storage? Hydrogen is not a particularly convenient energy form to store. Although it has a record high energy density per weight, it is a gas that takes up a large volume and cannot achieve high energy density as a liquid – and thus smaller volume – without considerable energy loss. It is therefore highly desirable to convert it into storable molecules such as methanol or ammonia, either in a two-step process
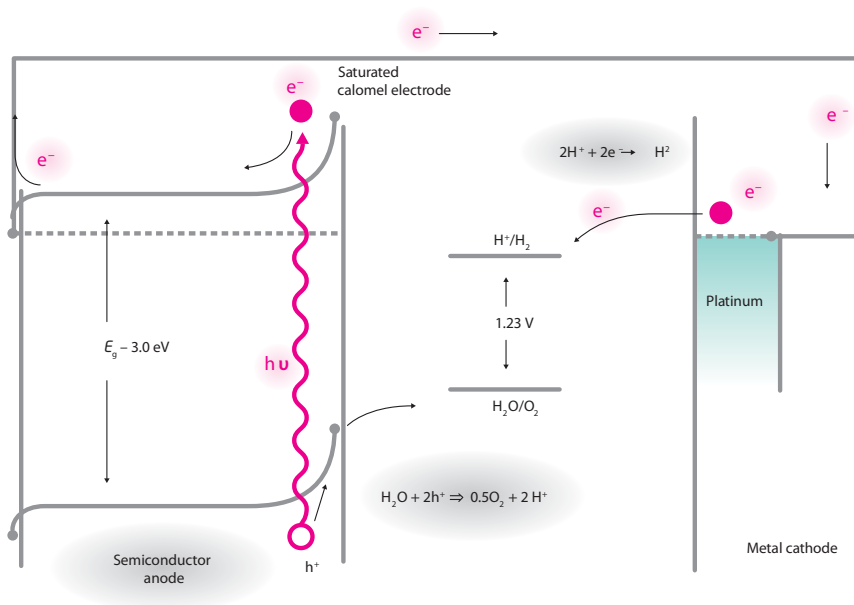
Fig. 2.8. The energy levels for splitting water. To the left is the energy diagram of a semiconductor shown with the fully occupied valence band and the empty conduction band. Irradiation with light of sufficient energy excites electrons from the valence band into the conduction band. Due to band-bending in the semiconductor, electrons are transported away from the surface and led to the platinum cathode and used for evolving hydrogen from solvated protons shown on the right side. The holes of the electron-hole pairs move in opposite direction, to the surface of the semiconductor, where they split water molecules and generate molecular oxygen and solvated protons. The cycle is hereby closed. Note that the energy of the water redox couple must lie well within the band gap of the semiconductor. The nanoparticles catalyzing the reactions on the semiconductor are left out for simplicity.
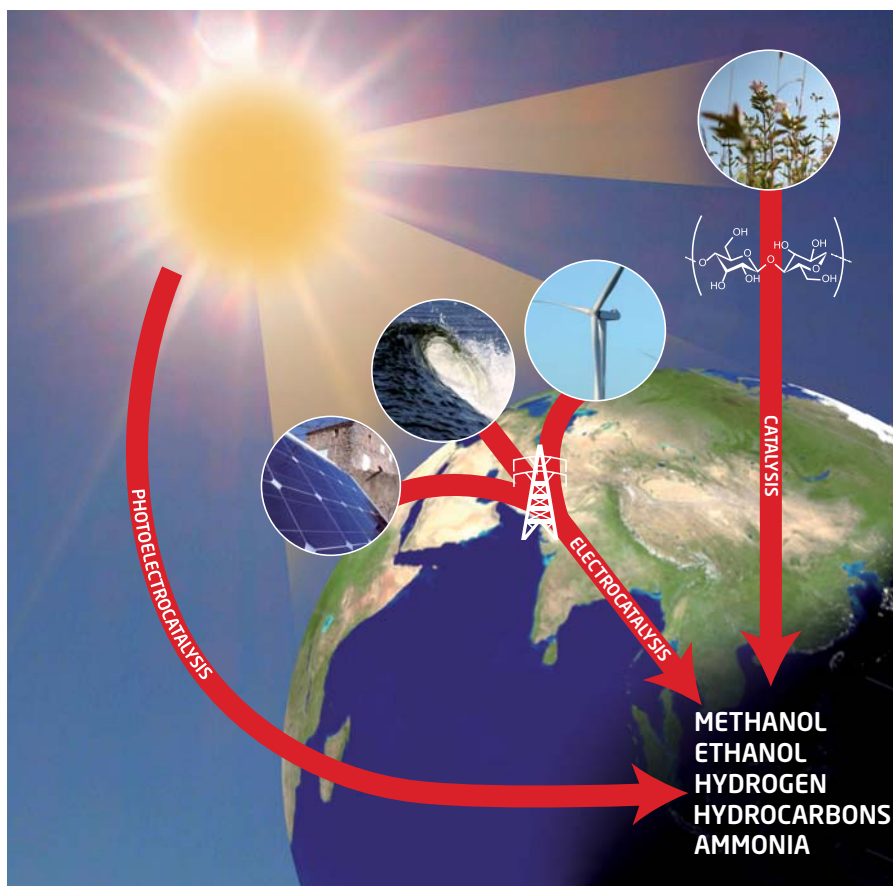
Fig. 2.9. The various routes from capturing sunlight to producing various forms of energy that can average out temporal variation. A new €16 million research initiative – Catalysis for Sustainable Energy (CASE) – at DTU, funded by Denmark's Ministry of Science, Technology and Innovation – is working towards discovering new solid catalysts and processes for energy conversion for a spectrum of energy sources such as those shown.

or directly coupled into, for example, the photocatalytic process (Fig. 2.9). The photocatalytic process would mimic nature if methanol were considered a simple version of sugar, which is nature's favorable way of storing energy in photosynthesis.

Nature handles these materials problems by a multistep process, where the high potential is avoided at the price of very low efficiency of the photosynthesis process. Although there are hopes that genetically manipulated plants may improve this, improvement by an order of magnitude is unlikely. This means that biomass can never become more than a partial solution to human energy consumption, as the need for energy and food

will compete for the same arable land (an average person in Denmark uses about 50 times as much energy as his or her own metabolism). Photocatalysis, in contrast, holds the promise of much higher efficiency. Unfortunately, this promise is far from being realized, and so far, only 2–3% efficiency has been realized.

By identifying and exploiting important properties in catalysis, such as the importance of the shape and size of catalytic nanoparticles, we hope to be able to design and realize simple and inexpensive systems for energy harvesting, conversion and storage that can ease the transition to a society relying on sustainable energy.

The author
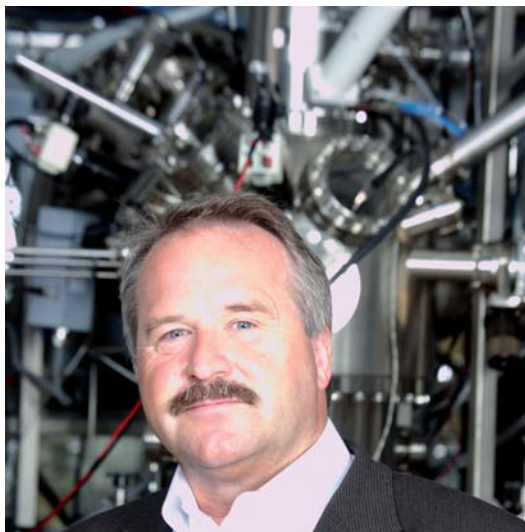
Ib Chorkendorff: PhD in Physics, Odense University, 1985. Director, DTU Center for Individual Nanoparticle Functionality (CINF), funded by the Danish National Research Foundation. He has many years of experience in surface physics, surface chemistry and heterogeneous catalysis, and especially how they relate to sustainable energy conversion.

More to explore

Walker G, King D. *The hot topic: how to tackle global warming and still keep the lights on.* Mariner Books, 2008.

Basic research needs for solar energy utilization: report of the Basic Energy Sciences Workshop on Solar Energy Utilization, April 18–21, 2005. Washington, DC, United States Department of Energy, 2007.

Lewis NS, Nocera DG. Powering the planet: chemical challenges in solar energy utilization. *Proceedings of the National Academy of Sciences of the United States of America* 2006: 103: 15729–15735.

Johansson TB, Goldemberg J, eds. *World energy assessment report: 2004 update*. New York, United Nations Development Programme, UN-DESA and World Energy Council, 2003.

Basic research needs: catalysis for energy: report from the Basic Energy Sciences Workshop on Solar Energy Utilization, August 6–8, 2007. Washington, DC, United States Department of Energy, 2008.

International energy outlook 2008. Washington, DC, United States Department of Energy, 2008.

Andersson MP, Abild-Pedersen F, Remediakis IN, Bligaard T, Jones G, Engbæk J, Lytken O, Horch S, Nielsen JH, Sehested J, Rostrup-Nielsen JR, Nørskov JK, Chorkendorff I. Structure sensitivity of the methanation reaction: H_2-induced CO dissociation on nickel surfaces. *Journal of Catalysis* 2008: 255: 6–19.

Jaramillo TF, Jørgensen KP, Bonde J, Nielsen JH, Horch S, Chorkendorff I. Identification of active edge sites for electrochemical H_2 evolution from MoS_2 nanocatalysts. *Science* 2007: 317: 100–102.

Wind energy research:
crucial for takeoff

Erik Lundtang Petersen & Jens Nørkær Sørensen

In March 2009, the annual European Wind Energy Conference and Exhibition took place in Marseilles, France, with more than 7000 people attending. Ten years earlier, the same event took place in Nice with 2000 participants and with almost no wind energy activity in France. In 2009, France had entered the gigawatt club: the 10 countries with more than 1 GW of installed wind energy capacity. Denmark has been in the club for many years.

What really differentiated the 2009 Conference from the previous conferences was the historic agreement reached in December 2007 and made into European Union (EU) law a year later. The European Parliament and the Council of the European Union agreed to a proposal from the European Commission for 20% of all energy consumed in the EU to come from renewable sources by 2020. For the wind energy community (and renewable energy in general), this was a great leap forward. For the first time, Europe's countries have a legally binding renewable energy target. The target forms part of the EU's 20-20-20 climate and energy package, which includes a 20% reduction in CO_2 emissions and a 20% improvement in energy efficiency by 2020.

The 20% target means that 30–40% of the EU's electricity will come from renewable energy in 2020, and wind energy is expected to overtake hydroelectric generation as the EU's largest source of renewable electricity.

Focusing on wind energy, the challenge is to advance the 155 TWh generated by wind energy today to 419 TWh in 2020, from 4.2% to 12–15% of total electricity production. Of this, one third should come from offshore wind farms. Is this realistic and feasible? According to the European Wind Energy Association, this is not ambitious enough. The reason for this claim is the surprising status (for some) of specific European wind energy trends and global trends in general. In 2008, wind energy turned out to be the EU's dominant electricity-generating technology in terms of generating capacity installed. In leading the EU power sector for the first time, wind accounted for 43%, or 8484 MW, of new capacity, beating all other technologies including natural gas, oil and coal. Almost the same amount of wind power was installed in 2008 in the United States and China but not with the same po-

sition relative to the other generation techniques.

The European installation rate for wind in 2008 amounted to 20 wind turbines per working day, 8.5 GW in installed capacity and 65 GW in total capacity, an increase of 15% from the year before. The Directive's target is 180 GW in 2020, which "only" requires 9.5 GW to be installed every year until 2020: only 1 GW more than the present installation rate. The European Wind Energy Association finds this insufficiently ambitious; they propose a target of 230 GW installed by 2020.

To put the many numbers in perspective, the installed capacity by the end of 2008 would produce 142 TWh of electricity in a normal year, equivalent to about 4.2% of the electricity used in the EU and the electricity used by 30 million households. This avoids the emission of 108 million tons of CO_2 per year, the equivalent of taking more than 50 million cars off Europe's roads.

Wind energy in Denmark

Denmark has been in the global forefront in industrial development and research since the the mid-1970s. Today, Denmark is the home of Vestas Wind Technology, the largest wind turbine manufacturer; LM Glasfiber, the largest turbine blade manufacturer; Siemens Wind Power, also a large global player; and several hundred subsuppliers. Denmark's wind energy research has grown together with the industry, mostly concentrated at the Risø National Laboratory, the Technical University of Denmark (DTU) and University of Aalborg. After the merger in 2007 between DTU and several research institutes, including Risø, the wind energy research at DTU encompasses more than 180 employees. The program of the 2009 European Wind Energy Conference indicates the strong impact of the DTU research: 10 chairs, 25 oral presentations and 25 poster presentations, numerous committee members and participants in numerous meetings.

These developments clearly show that wind energy has entered a critical phase of technological maturation. Research objectives do not solely focus on increasing the size of the largest turbines, which has been the mantra for many years, but reflect a broader vision. The wind sector has to focus on efficiently and professionally delivering, installing and connecting large amounts of wind power to the

electricity grid, with strong concerns about the reliability, availability and accessibility of the turbines.

The four highest priority research areas as identified by the European Technology Platform for Wind Energy and the European Wind Initiative:

- wind resource assessment;
- wind turbine technology;
- wind energy integration; and
- offshore deployment and operation.

These research areas reflect the overall objective of the wind energy sector: reducing costs, which will help to secure the global objectives of security of energy supply, reducing CO_2 emissions to alleviate climate change and helping to ease poverty.

The wind energy research at DTU is deeply involved in all aspects of the four priority areas described above, including basic research, strategic research and industrial research and consultancy. A large proportion of the work takes place in national and EU research programs. The following description focuses on two research areas that have had the greatest impact: wind resource assessment and aerodynamics and aeroelasticity. Due to the increasing importance of offshore wind energy, especially for Denmark and the EU, this subject has been given special attention.

Assessing wind resources
A recent publication from the European Wind Energy Association, *The economics of wind energy*, states the following.

The local wind resource is by far the most important determinant of the profitability of wind energy investments. Just as an oil pump is useless without a sizable oilfield, wind turbines are useless without a powerful wind field. The correct micro-siting of each individual wind turbine is therefore crucial for the economics of any wind energy project. In fact, it is beyond dispute that, during the infancy of modern wind industry in 1975–1985, the development of the European Wind Atlas methodology was more important for productivity gains than advances in wind turbine design. The European Wind Atlas method developed by Risø was later formalized in the WAsP computer model for wind resource assessment.

Henrik Stiesdal, Chief Technology Officer of Siemens Wind Power, says that the Wind Atlas Analysis and Application Program has had colossal importance for the wind energy industry. The availability of wind resources at a turbine site and the ability to accurately calculate them over the 20-year lifespan of the turbine are essential for the developments described above. Is there actually enough energy in the wind to support the anticipated global wind energy development? Yes, and if this were not the case, the global climate would be quite different from the present climate. The atmosphere is very effective at transforming the incoming solar radiation into kinetic energy. It has to be, because it has to transport large amounts of energy from the equatorial regions to the polar regions due to the offset of radiation balance: positive at the equator and negative at the poles. During this process and due to the rotation of the Earth, the large weather systems are created, such as the Westerlies in the Northern Hemisphere and the Roaring Forties in the Southern Hemisphere.

The wind dissipates between 30 and 100 times more energy per second than all of humanity.

The objective for the research on wind resources today is to be able to predict the resources anywhere on the globe within 5%, and we are close to meeting this for many regions. Regions with very complicated topography and climate are still a serious challenge.

The wind atlas method published in 1980 for Denmark developed directly into the methods currently used. The wind atlas method was developed here, enabling a user to calculate the energy output from a specific wind turbine at a specific site by including local conditions in the calculations. The publication of the method, the statistical tables and the examples profoundly influenced the take-off of Denmark's wind industry. The method was brought over into the work with the *European wind atlas* (12 countries) from 1982 to 1989. During this period, the personal computer was invented and the whole method was then put into WAsP. Today more than 2600 licensees in more than 100 countries use WAsP.

The *European wind atlas* relied heavily on the availability of a dense network of high-quality meteorological stations. These are often not available, and today's research therefore focuses on establish-

The numerical wind atlas method

The wind atlas method
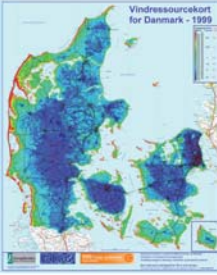
The conventional method used to estimate wind resources on regional scales is to analyze wind measurements made at several sites around the region, such as the *European Wind Atlas*. This method requires a sufficient spatial distribution of high-quality data.

Numerical wind atlas methods have been devised to solve the issue of insufficient wind measurements. One such method is the KAMM/WAsP method developed at Risø National Laboratory.

The KAMM/WAsP method uses statistical-dynamical downscaling. The basis for the method is that meteorological situations on a large scale and weather situations on a small scale are robustly related.

Global wind climate

Information about the large-scale weather situation is freely available from various reanalysis datasets (such as NCEP-NCAR reanalysis or NNRP2 on the right).
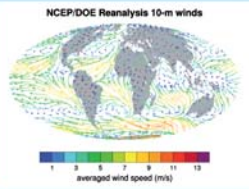
This data set has been created by assimilating measurement data from around the globe consistently for many years (1979 to present). The data set was generated to provide a reference for the state of the atmosphere and to identify any features of climate change. The data set also provides a long-term record of large-scale wind climate.
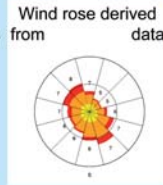
The reanalysis data is used to determine about 100 large-scale wind situations, called wind classes, that represent the regional large-scale wind climate.

Regional wind climate

To make these wind classes meaningful at a smaller scale, a mesoscale model is used to determine how regional scale topography modifies the large-scale wind forcing. For each wind class, a mesoscale model simulation is therefore performed using the Karlsruhe Atmospheric Mesoscale Model.
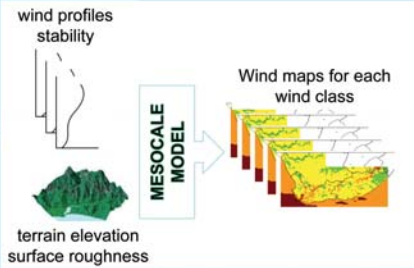
The atmospheric flows from all mesoscale simulations (the mesoscale wind pattern generated by KAMM for each wind class) are then recombined according to their frequency of occurrence in the atmospheric reanalysis.

This generates a regional wind climate. The mesoscale winds are then normalized for a standard height level over a flat surface of homogeneous roughness. Such an example is seen on the left for the wind atlas of Egypt.

Microscale wind climate
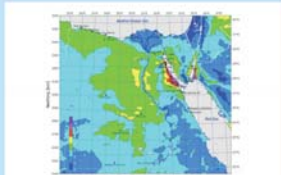
Wind resource map of Egypt: mean wind speed at 50 m above ground level

Regional wind climate maps, however, are still not directly applicable to the real environment. As shown on the right, wind climates are used within the Wind Atlas Analysis and Application Program (WAsP), together with detailed knowledge of the local topography and surface characteristics, to establish micro-scale wind climate. These results are then directly transferable to the local wind energy estimation.

Mesoscale wind climate

microscale wind climate
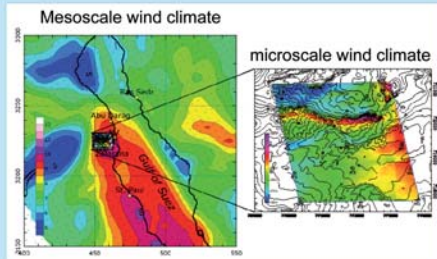
Work is currently underway to transition from the KAMM/WAsP method to other more sophisticated downscaling methods of generating regional wind atlases. Some of these include:
• clustering algorithms of the large-scale synoptic pattern plus dynamical mesoscale modeling using various mesoscale models; and
• full dynamical downscaling: fully simulating 20–40 years of mesoscale model simulations using the reanalysis as boundary or initial conditions.

Local wind resource

Simple, fast and inexpensive → Complex, slow and expensive

| Horizontal interpolation | Risø DTU Wind Atlas (KAMM/WAsP) | Statistical-dynamical | Fully dynamical |

Fig. 3.1. The numerical wind atlas method

ing methods that use meteorological data to verify the wind atlas calculation rather than being the basis. This method is called the numerical wind atlas method (Fig. 3.1).

A numerical wind atlas is produced by a numerical weather forecasting model, the same type of model meteorological offices use to produce daily weather forecasts. In theory, the model is run for several years for a region, for example of the size shown in Fig. 3.2 with a resolution of a few kilometers, and the wind statistical quantities are collected at the grid points and then a micro-siting model such as WAsP is applied to specific wind turbine sites. In practice, due to limited computer resources, climatological data are used to create 100 different large-scale wind situations (wind classes) representing the large-scale wind climate. Then the mesoscale model is run for each of the wind classes to determine how regional scale topography modifies the large-scale wind forcing. Fig. 3.1 further describes and illustrates the application.

Three factors facilitated this development:
- increasing computer power;
- the reanalysis data set, in which 50 years of global meteorological data and weather forecasts have been analyzed, resulting in a global dataset, originally intended for global warming analysis, but that has proven to be a unique dataset for establishing wind atlases all over the world; and
- the detailed mapping of the surface of the world.

All three datasets are available on the Internet free of charge or at low cost.
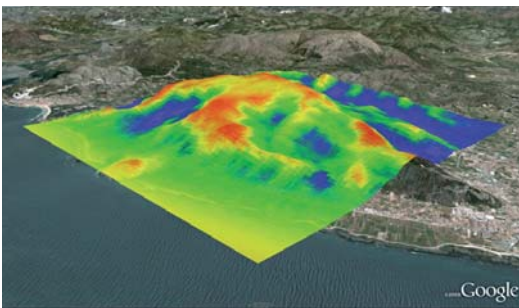
Fig. 3.2. Wind resource map for Serra Santa Luzia in northern Portugal. WAsP projects the expected generation of a wind turbine and shows it in color. Generation varies by more than 7 times from the lowest values (blue) to the highest ones (red)

DTU is currently creating wind atlases for India, three provinces in China and South Africa. Great challenges ahead include producing a new *European wind atlas* for all 27 EU countries and developing a new numerical wind atlas method with the following fundamental research challenges to overcome the limitations of the current method described above.
- The mesoscale models are used in a steady-state mode and are unable to represent time-dependent thermally induced circulation, such as diurnal and seasonal cycles.
- Important output from mesoscale models (such as boundary-layer height and stability) is not used as input to the microscale model.
- Methods of downscaling mesoscale models to microscale models are not well developed.
- The assessment method cannot provide time-dependent parameters.

Many of these issues arose from limitations in computer and networking power that might be overcome by the rapid increase in computing power and storage. Further, these developments should enable trends in wind power potential due to climate change to be estimated.

Aerodynamics and aeroelasticity

Aerodynamics and aeroelasticity comprise crucial disciplines in designing wind turbines. The rapid increase in size, from rotor diameters of 10–15 meters in the early 1980s to today's rotor diameter of more than 120 meters, has made the use of dynamic aeroelastic analysis an intrinsic and necessary part of the design process. Universities and research laboratories have mainly developed the aeroelastic tools, which consist of different elements. As input, a time history of the wind inflow is needed, as seen by the rotor, containing the main statistical properties of the wind. Besides inflow data, the aeroelastic code consists of an aerodynamic part to determine the loading of the wind and a structural part describing the dynamic response of the structural parts of the wind turbine.

The standard approach for the aerodynamic part used by wind turbine manufacturers is based on the blade-element momentum theory. This assumes that the flow takes place in independent

stream tubes and that the load can be determined from two-dimensional sectional airfoil characteristics. Since the model has to cope with all kinds of operating conditions, it is extended by many add-ons to simulate such factors as yaw misalignment, wind shear and blade pitch. Despite the shortcomings of the model, it runs very fast and forms the basis for all design and optimization of wind turbines. An alternative to the blade-element momentum technique is to use advanced full-field methods based on solutions of the Euler or Navier-Stokes equations. Such models contain much more information, since they describe in detail the flow characteristics around the wind turbine. Nevertheless, they are very computing-intensive, and the available computer capacity limits their range of applicability. The structural parts of a wind turbine may be modeled using various methods. The simplest is to use a modal approach in which the blade and tower deflection comprises a linear combination of basic eigenmodes of the individual parts of the wind turbine. An alternative method is the multi-body formulation, in which the various rigid parts are connected through springs and hinges. Finally, the dynamics of the structural parts can be modeled using full finite element discretization. In many cases, the aeroelastic codes are coupled to optimization and control algorithms, which makes the required computing efficiency even more severe.

Today, the research in aeroelasticity is directed towards the challenges of the design trends of new wind turbines, which are generally governed by the development of wind turbines that are larger and more slender. Important research areas include nonlinear structural dynamics, strong coupling between loading and deformation (fluid–structure coupling), the structural stability of wind turbine blades, operation in wakes, complex inflow, the aerodynamics of parked rotors and floating offshore wind turbines.

DTU has been engaged in research and development related to aeroelastic tools as a main activity in wind energy for many years. DTU developed the aeroelastic codes Flex5 and HAWC, which are among the industrial design tools wind turbine manufacturers most commonly use.

The numerical wind tunnel

Today, basic studies on wind turbine aerodynamics and wind fields are carried out using computational fluid dynamics (Fig. 3.3). DTU has conducted research on computational fluid dynamics and wind energy for more than two decades. This has resulted in the development of the EllipSys3D program, which today is comprehensive software for simulating and optimizing airfoils and rotor blades and for predicting optimum siting of wind turbines and wind farms. EllipSys3D is a computing code for solving the Navier-Stokes equations in general curvilinear coordinates. The equations are formed in a cell-centered multi-block topology, and the code is valid for both laminar solutions and for turbulent flow cases by solving the Reynolds-averaged equations or by using sub-grid-scale models for large eddy simulations. The code is optimized by using multi-grid acceleration, and special treatment of the block topology ensures very rapid convergence. A further application of the model is simulating the generation and emission of acoustic waves (noise) from wind turbines (Fig. 3.4).
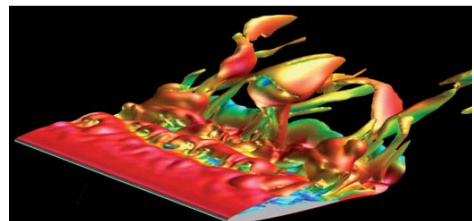
Fig. 3.3. Detailed computational fluid dynamics computation of the flow around a segment of a wind turbine blade
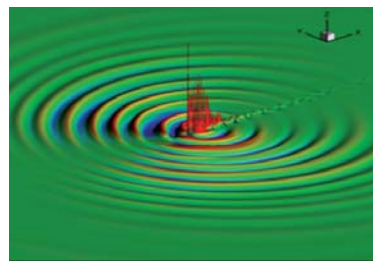
Fig. 3.4. Aero-acoustic computation showing the emission of acoustic waves about an airfoil

Wind energy on the high seas

Although wind turbines have become a natural part of Denmark's landscape, they are still a relatively new feature in much of the European countryside. In densely populated countries such as the Netherlands, United Kingdom and Denmark, however, wind energy is expected to primarily be developed further at offshore locations. Today, offshore wind accounts for a small proportion of the total installed wind power capacity. Offshore wind power has mainly been developed in northern European countries, and the total offshore wind power installed worldwide is about 1500 MW, less than half the total wind power installed in Denmark. Although the total capacity is limited, the growth rates are very high, with more than 35,000 MW planned to be installed in Europe in the next one to two decades. Generally, offshore wind power poses specific challenges in manufacturing, transport, erection and maintenance. On land, the cost of the wind turbines comprises about 75% of the total cost, with the plant hardware plus maintenance costs accounting for the rest. Offshore, this split may very well be reversed, as offshore turbines are likely to experience more downtime than those onshore, and maintenance therefore plays a major economic role. The research on offshore wind power has mainly targeted wave loads, wake effects, foundation and grid transmission. With the rapid development in offshore wind power, important fields include developing tools for predicting the mutual interaction between wind farms and modeling the detailed characteristics of the atmospheric boundary layer. For large offshore sites, wake losses are likely to be higher than at onshore locations, due to less mixing between the flow inside the wind farm and the ambient turbulence (Fig. 3.5 and 3.6). Measurements have shown that offshore wind farms generate more power in the summer than in the winter, indicating that the stability of the atmospheric boundary layer plays an important role in generation.

Fig. 3.5. A view of the Horns Rev wind farm in the North Sea 14 kilometers west of Denmark
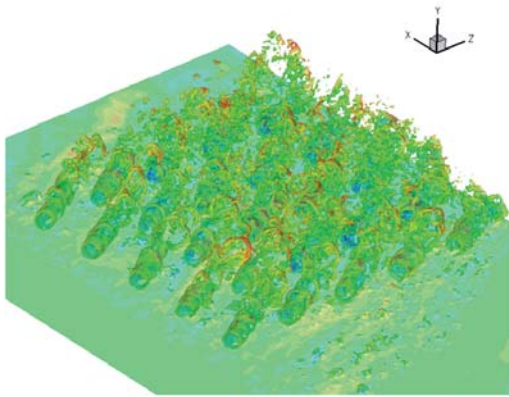
Fig. 3.6. Numerical simulation of the flow structures inside an offshore wind farm computed using the Ellip-Sys3D code. Such computation is important for determining overall power generation, since the interaction between the turbines and their wakes determines the local flow in the wind farm

Concluding remarks

The wind energy research at DTU described here has focused on two of the traditional areas in which DTU has been active for more than three decades. However, this only constitutes part of the total wind energy research DTU conducts. The research on the future grid for energy distribution, the ability to forecast days in advance the availability of wind energy, materials research, the wind design conditions and economics are all visible on the global scale and contribute to fulfilling the targets for sustainable energy.

The authors

Erik Lundtang Petersen: MSc and PhD (1973), DTU. Research on atmospheric turbulence, Pennsylvania State University. Head, Meteorology Section, Risø National Laboratory, 1977; Head, Wind Energy Department, 1985–. Coordinator, *European wind atlas* for the European Commission, 1982–1989. Editor, *Wind Energy*, 1998–2008. DONG Grant of Honor 2002; Poul la Cour Prize, European Wind Energy Association, 2007. Currently Steering Committee member and Chair, Working Group on Wind Conditions, European Technology Platform for Wind Energy and the European Wind Initiative.

Jens Nørkær Sørensen: MSc and PhD (1986), DTU. Research experience, French Aeronautics and Space Research Center. Faculty member, DTU, 1989–. Head, Fluid Mechanics Section, 1996; Professor, 2001–. Gorm-Petersen Memorial Grant Award, 1987; Alexander Foss Gold Medal Award, 1999; Statoil Prize, 2002; Member, Danish Academy of Technical Sciences, 2005. Member, Steering Committees, European Academy of Wind Energy, Danish National Committee for IUTAM, Danish Center for Applied Mathematics and Mechanics and Fluid.DTU. Editorial board member, *Wind Energy*. Research: fluid mechanics and wind energy.

More to explore

Sørensen JN, Hansen MOL, Hansen KS, ed. The Science of Making Torque from Wind, 28–31 August 2007, Technical University of Denmark. *Journal of Physics: Conference Series* 2007: 75. Bristol, IOP Publishing.

Hansen MOL, Sørensen JN, Voutsinas S, Sørensen N, Madsen HA. State of the art in wind turbine aerodynamics and aeroelasticity. *Progress in Aerospace Sciences* 2006: 42: 285–330.

Vermeer LJ, Sørensen JN, Crespo A. Wind turbine wake aerodynamics. *Progress in Aerospace Sciences* 2003: 39: 467–510.

Thor S-E, Weiss-Taylor P. Long-term research and development needs for wind energy for the time frame 2000–2020. *Wind Energy* 2002: 5: 73–75.

Snel H. Review of the present status of rotor aerodynamics. *Wind Energy* 1998: 1: 46–69.

Troen I, Petersen EL. *European wind atlas*. Roskilde, Risø National Laboratory, 1989.

Petersen EL, Troen I, Frandsen S, Hedegaard K. Danish windatlas. *A rational method of wind energy siting*. Roskilde, Risø National Laboratory, 1981.

Petersen EL, Mortensen NG, Landberg L, Højstrup J, Frank HP. Wind power meteorology. 1. Climate and turbulence. *Wind Energy* 1998: 1: 2–22.

Petersen EL, Mortensen NG, Landberg L, Højstrup J, Frank HP. Wind power meteorology. 2. Siting and models. *Wind Energy* 1998: 1: 55–72.

Wind Atlas Analysis and Application Program (WAsP, www.wasp.dk)
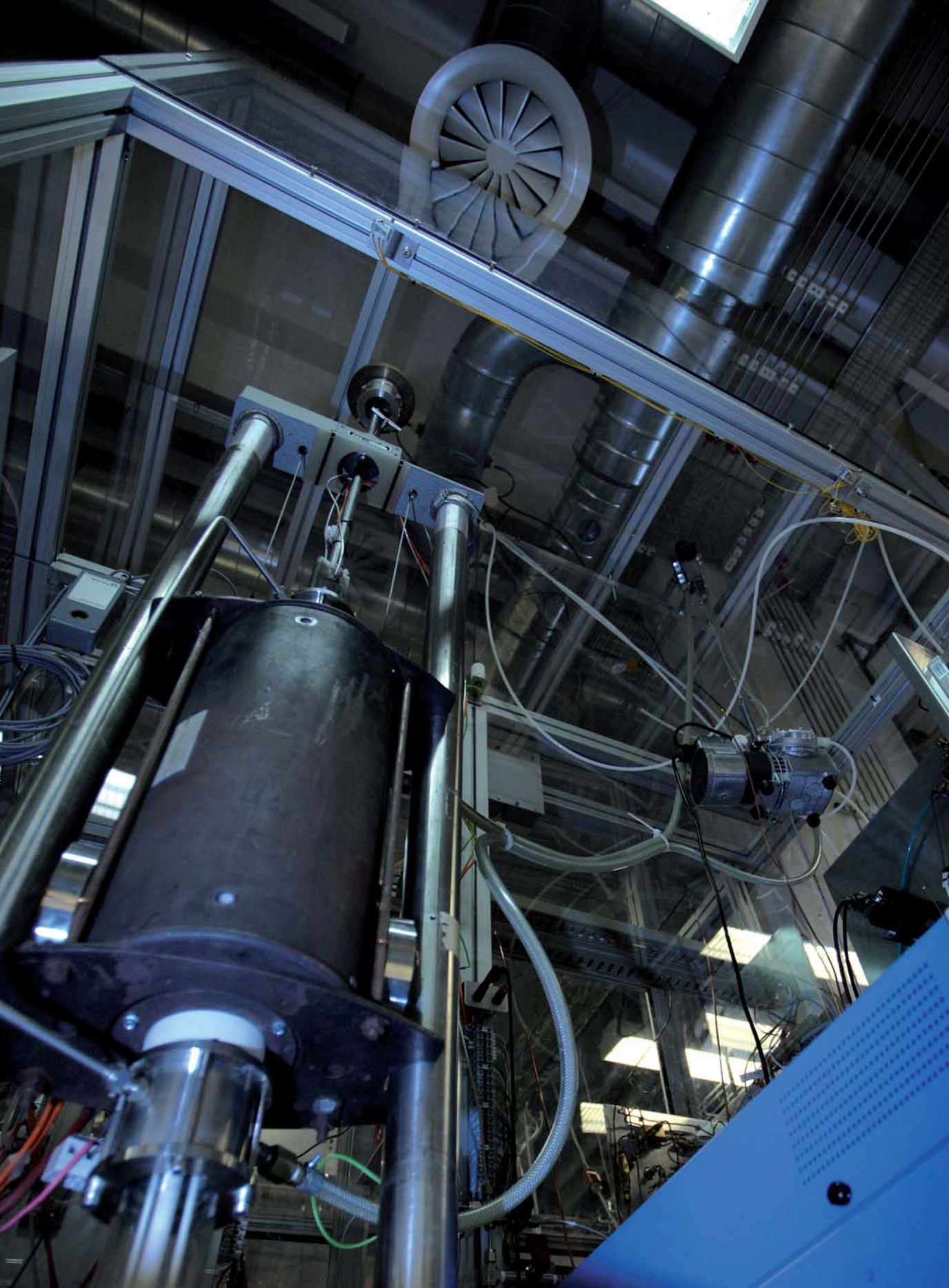
Fuel cells and electrolysis

for converting renewable energy

Mogens Mogensen & Søren Linderoth

Fuels cells and electrolyzers may help to achieve 100% sustainable energy supply – including transport. Electricity from renewable resources such as wind, solar and wave can be converted by means of electrolyzers into synthetic fuels for transport and for storing energy for later use when needed. Thus, large offshore wind farms in the North Sea may produce methane (natural gas), which can be transported to customers using Denmark's existing gas pipeline network. Fuel cells can then convert the natural gas into electric power and heat efficiently when this is needed and with little or no pollution.

A fuel cell is an electrochemical cell that can convert the chemical energy of hydrogen and oxygen directly into electricity. Several types of cells exist. Fig. 4.1 shows the principle of solid oxide cells. The cell is reversible, as Fig. 4.1 also explains.
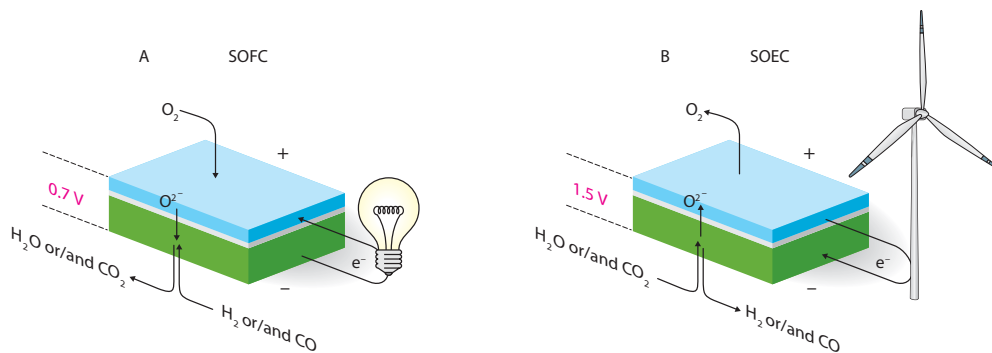
Fig. 4.1. Principle of a reversible solid oxide cell. Such an electrochemical cell consists of a thin electrolyte layer (white) with electrodes on both sides. The cell can be operated as a solid oxide fuel cell (SOFC) (**A**), generating electricity while consuming H_2 or CO at the negative electrode and O_2 at the positive electrode. If operated in the reverse mode as a solid oxide electrolyzer cell (SOEC) (**B**), then electricity is consumed with H_2 + CO being produced from H_2O + CO_2 at the negative electrode and O_2 at the positive electrode. The overall reversible reaction is:

$$H_2 + CO + O_2 \xrightleftharpoons[SOEC]{SOFC} H_2O + CO_2 + \text{electric energy } (\Delta G) + \text{heat } (T\Delta S)$$
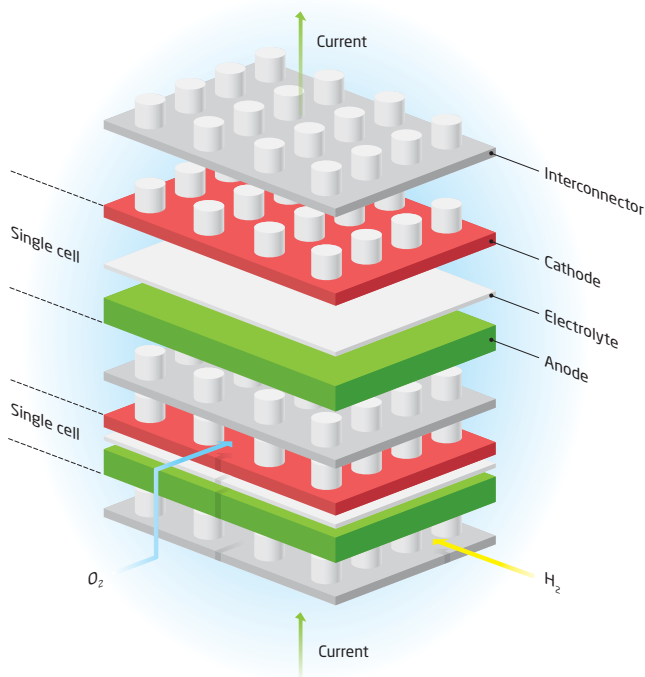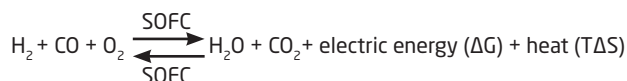
Fig. 4.2. Sketch of a cell stack. The cells must be stacked in series, as the voltage across a single cell is only about 1 volt. Many designs of cell stacks (assemblies) have been reported. The simplest is the bipolar flat plate design of the type shown here and used by the Risø DTU National Laboratory for Sustainable Energy and Topsoe Fuel Cell A/S.

Research and development of solid oxide cells at Risø DTU

The arguments for the choice of fuel cell type for Denmark's fuel cell program in 1988 were as follows. 1) Solid oxide fuel cells are high-temperature cells that can convert natural gas and other fuels without separate fuel-processing units that operate at a higher temperature than the cell stack. 2) They were considered to have fewer problems with materials than other types of high-temperature fuel cells. 3) They are reversible: the same cell can be used for electrolysis. 4) It was the least-explored type of fuel cell, and thus Denmark's consortium had a better chance to get to the international forefront. The actual work of Denmark's solid oxide fuel cell consortium started gradually in 1989, and by mid-1990 about 20 people were employed within this project, of which 15 were at Risø National Laboratory. Between 20 and 30 employees worked on the solid oxide fuel cell projects until 2001. In 2001, a new consortium between Risø National Laboratory and Haldor Topsøe A/S was created. The number of employees within the consortium has been growing since then and is now about 200. The purpose of the company, Topsoe Fuel Cell A/S, is to commercialize the solid oxide fuel cell technology.

It was recognized from the very beginning that the development of solid oxide fuel cells had to be based on scientific understanding. An integrated effort involving studies from the nano-sized structures of the surfaces and composite electrodes to the development of expertise on cell and stack fabrication was therefore initiated. The solid oxide fuel cell group did not have access to the techniques for nano-studies from the very start, but as these opportunities gradually became available, not surprisingly, the nanoscale studies were very important for understanding the electrochemistry. Research and development on solid oxide fuel cells is an interdisciplinary undertaking involving several disciplines:

- ceramic and powder metallurgy processing for cell fabrication;
- materials physics and chemistry, including metallurgy, defect chemistry of ceramics and oxidation of metals;
- electrochemistry, theory and especially advanced characterization methods;
- electron microscopy (scanning, transmission and high-resolution transmission electron microscopy), including advanced methods of sample preparation to make good specimens from the composite porous structures;
- surface science methods such as X-ray photoelectron spectroscopy, time-of-flight secondary

ion mass spectrometry, atomic force microscopy and controlled-atmosphere high-temperature scanning probe microscopy; and
- modelling of the physics, chemistry and economics involved in the technology.

Examples of the results of the two decades of work are given below.

Cell fabrication

Denmark's consortium had a clear strategy from the very beginning of focusing on inexpensive cell production methods that could be scaled up for mass production by industry. Many groups were already deeply engaged in solid oxide fuel cells and with several different designs. Denmark's consortium decided for a flat-plate design, especially due to the high power density achievable. The cells consist of thin ceramic or cermet (ceramic-metal composite) layers (Fig. 4.3). Inexpensive methods for fabricating such layers include such processes as tape-casting slurries, spray-painting slurries or screen-printing pastes. The slurries and pastes contain the ceramic and/or metal powders that eventually will constitute the single layer of the cell. Using such methods, Risø successfully developed methods that have been scaled up to a pre-production line (capacity of 20,000 12×12 cm cells per
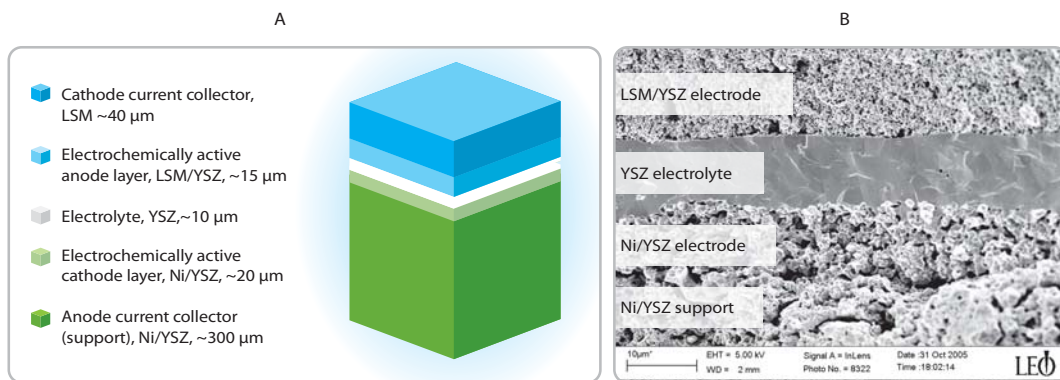
Fig. 4.3. **A**. Elements of the second-generation Risø DTU solid oxide fuel cell. LSM: lanthanum strontium manganate ($(La_{0.75}Sr_{025})_{0.95}MnO_3$), an electron-conducting ceramic. YSZ: yttria-stabilized zirconia ($Zr_{0.84}Y_{0.16}O_{1.92}$), an oxide ion conduction ceramic electrolyte. Both electrodes are porous composites containing about an equal volume fraction of electrode material and YSZ. **B**. A scanning electron microscope micrograph of a cross-section (fracture) of a real solid oxide cell produced at Risø DTU. The cell, which has a thick porous support, consists of two porous composite electrodes, one on each side of the dense (gas-tight) YSZ electrolyte.

year as of 2008), and these cells have been delivered to Topsoe Fuel Cell A/S for fabricating stacks since 2002. Topsoe Fuel Cell A/S has now built a pilot plant with a capacity of 5 MW per year, and the cell fabrication technology is the one developed by the Risø DTU National Laboratory for Sustainable Energy. The purpose is to demonstrate the technology on a near-commercial scale.

The cell fabrication development takes place stepwise. Massive research and development efforts on electrodes, electrolytes and support have taken place continually. When the progress of single components seems sufficient to make it worthwhile to integrate these results into a new cell, a new cell may be developed. As all the layers have to be sintered together into one solid cell in a few steps, preferably one, such work requires considerable labor. The structure and properties of each layer must be such that all layers in the unsintered cell will shrink with the same rate at all temperatures, and finally the electrolyte layer must be totally dense and gas-tight, and the electrode and support layers must be porous with different pore sizes and distributions. Thus, only a few cell types have been attempted over the years. Fig. 4.4 depicts the different types (generations) of cells fabricated at Risø DTU. Only the production of second-generation cells has been scaled up so far. Fig. 4.5 shows how this cell is high-performing and totally reversible (working in both fuel cell and electrolysis mode). The other cell types have only been fabricated on a laboratory scale. The improved electrochemical performance has been used to decrease the operating temperature, because the cost of the balance of plant of a fuel cell power plant becomes higher for higher operating temperatures, and the cost of the balance of plant is foreseen to be the major part of a fuel cell electricity generator.
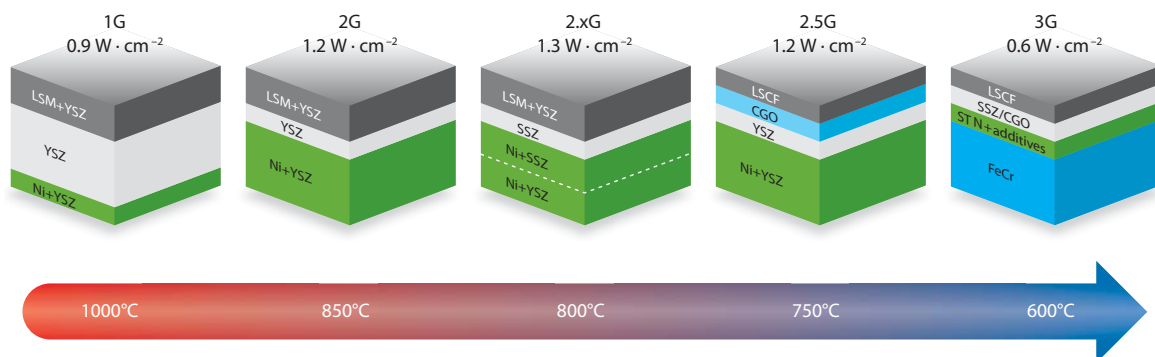
Fig. 4.4. Generations of cells with optimization over time (the generations) in operating temperature and electrochemical performance. The first-generation (1G) cell has a thick electrolyte (YSZ) as the mechanical backbone and thin electrodes of Ni-YSZ and LSM-YSZ. The second-generation (2G) cell has a thick Ni-YSZ support and a thin electrolyte. The first generation was ready in 1994 and the second generation in 2002. Many of these cell types have been produced and demonstrated in stacks. The 2G cell is the basis for the ongoing pre-commercial demonstration activities carried out by Topsoe Fuel Cell A/S. The 2.xG, 2.5G and 3G cells are all in the laboratory stage, and relatively few have been fabricated. They represent different lines of research and development. The 3G cell with metal support is especially directed towards a much less expensive cell and is the latest development. The combinations of materials shown are only examples, as the selection of materials is still under consideration and experimentation. The power density for a given cell at a given cell voltage is an exponential function of temperature. The reaction rates are thermally activated and follow an Arrhenius expression with apparent activation energy of about 0.7 eV. Despite this, the power density (W · cm^{-2}) at 0.6 V per cell has been increased, while the operating temperature has been lowered significantly. SSZ: scandia-stabilized zirconia. LSCF: lanthanum strontium cobalt ferrite. CGO: cerium gadolinium oxide. STN: strontium titanate with niobium oxide.
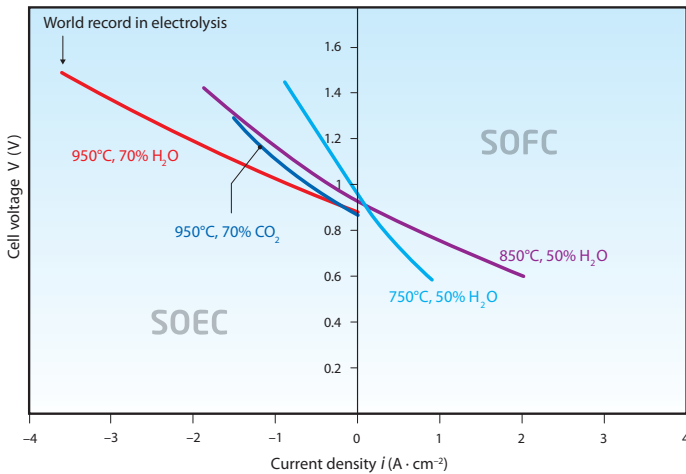
Fig. 4.5. Cell voltage versus current density for 2G Risø DTU cells in fuel cells (SOFC) as well as in electrolysis (SOEC) mode at a few selected conditions. The current density of 3.7 A · cm^{-2} at 950°C at 1.5 V is the highest known value reported. Current density of 3.7 A · cm^{-2} produces 15.5 normal m^3 of H$_2$ per m^2 of cell per hour. The smooth passage of the curves across zero current proves the reversibility of the cells. H$_2$ comprises the remainder up to 100% for H$_2$O as does CO for CO$_2$.

Fig. 4.5 shows that the solid oxide cells are fully reversible, and the performance (the internal resistance) in electrolyzer mode is almost as good as in fuel cell mode. Further, Fig. 4.5 shows that CO$_2$ can also be electrolyzed at a rate that is somewhat lower but still similar to that of steam electrolysis.

Degradation of cells

An important parameter is the lifetime of the cells and stacks. An interesting aspect here is that carrying out useful and high-quality studies of lifetime and degradation mechanisms has required produc-

ing many high-quality and highly reproducible cells in the pre-pilot facility. A large matrix of cell tests was carried out for 1500 h or more and, in one case, a cell was tested for 2 years under rather harsh conditions. This has naturally created knowledge about the problems related to durability and degradation mechanisms, which in turn reveal the potential for improving the lifetime under given operating conditions. The development of electrochemical impedance spectroscopy enabled the contribution of each of the electrodes and the electrolyte to the overall degradation of the cell to be quantified. Fig. 4.6 shows an example.
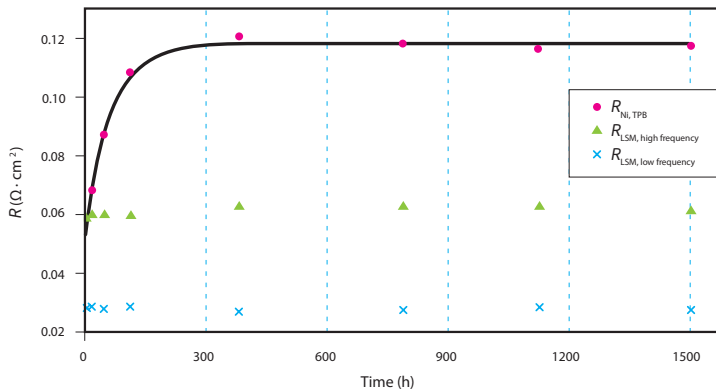
Fig. 4.6. Durability of the 2G cells. The area-specific cell resistance R (Ω · cm^2) is the sum of the resistance of the cell components, which are plotted versus time. $R_{Ni, TPB}$ is the H$_2$ electrode resistance; $R_{LSM, high frequency}$ + $R_{LSM, low frequency}$ is the O$_2$ electrode resistance. They are obtained by mathematical analysis of electrochemical impedance spectroscopy recorded during testing at 0.75 A · cm^{-2} and 750°C as a function of time. After 300 hours, all resistance values are constant with time: no further degradation occurs within the accuracy of the measurement at the test conditions. TPB: triple-phase boundary.

The line shows the fit for the development over time of $R_{Ni, TPB}$ – the polarization resistance of the Ni-YSZ hydrogen electrode, which is dominated by the resistance of the electrochemical reaction at the three-phase boundary between Ni,YSZ and gas given by an expression of the type

$R_{Ni, TPB} = k_1 + k_2 \cdot (1 - \exp(-t/\tau))$.

k_1 and k_2 seem to be constants, which are characteristics of the Ni-YSZ composite structure, and the time constant, τ, is a function of the steam partial pressure: the higher the steam partial pressure is, the shorter the time constant.

Materials science and electrochemistry studies

The Risø DTU National Laboratory for Sustainable Energy has been performing many diverse materials science and electrochemical investigations. Such subjects as the mechanical strength of cells, metal oxidation of interconnects and 3G cell supports, glass chemistry and physics of the sealing materials and modeling of all processes from the atomic level to integral cell stack have been and are studied in multi-person-year projects.

Other major research fields have been electrode studies using electrochemical impedance spectroscopy, electron microscopy (scanning and transmission), atomic force microscopy and controlled-atmosphere high-temperature scanning probe microscopy (a novel technique developed as part of the Risø DTU solid oxide cell projects), X-ray photo electron spectroscopy and time-of-flight secondary ion mass spectrometry. These methods have been used in studying possibly detrimental chemical reactions between the cell materials during fabrication, which involves sintering at high temperatures up to 1350°C. Further, Risø DTU has been extensively studying the electrode reaction mechanisms, including the presence and effect of segregation of impurities and chemical constituents of the cell to the interfaces, and in particular to the three-phase boundary.

Fig. 4.7 illustrates the results of an electrochemical impedance spectroscopy study on how nanoparticles may improve electrode performance.

Fig. 4.8 shows two atom force microscopy images of the segregation of SiO_2 containing glassy impurities to the rim of Ni model electrodes, often called point electrodes, which are single contact areas of Ni electrodes and YSZ electrolytes. The contact areas are elliptic or circular 100–300 μm in diameter. The rim of the electrode – the three-phase boundary – is where the H_2/H_2O electrode reaction would be fastest for a clean electrode.
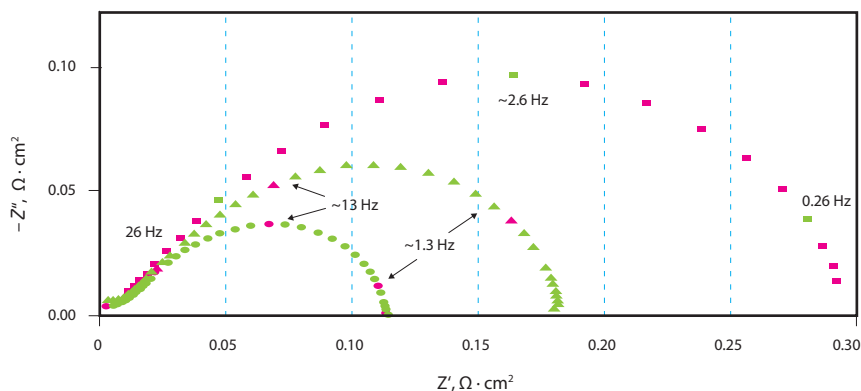
Fig. 4.7. Electrochemical impedance spectroscopy plot of simple composite electrodes of $(La_{0.6}Sr_{0.4})_{0.99}CoO_3$ and $Ce_{0.9}Gd_{0.1}O_{1.95}$ (■) and the same electrodes impregnated with nanoparticles of $Ce_{0.9}Sm_{0.1}O_{1.95}$ (△) and CeO_2 (○) at 600°C and zero current in air. The especially surprising aspect is the great improvement: the decrease in electrode polarization resistance (Z'), by nanoparticles of pure ceria (CeO_2), because pure ceria is not supposed to have any electron or ion conductivity in air. The interpretation is that the nanoparticles change the surface chemistry of $(La_{0.6}Sr_{0.4})_{0.99}CoO_3$ and $Ce_{0.9}Gd_{0.1}O_{1.95}$ and, through this, accelerate the oxygen electrode reaction.
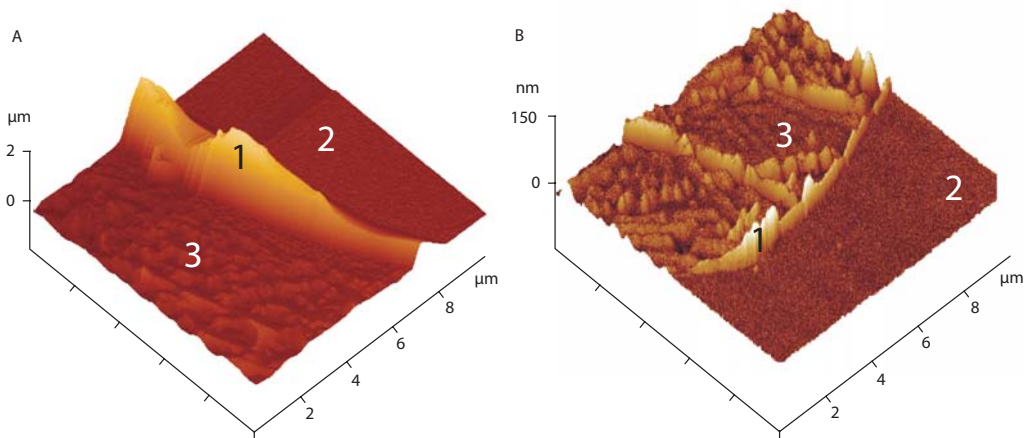
Fig. 4.8. Atomic force microscope pictures of rim ridge (**1**) of non-conducting impurity at the three-phase boundary of a Ni point electrode on a YSZ electrolyte after cool down and removal of the Ni electrode. **A**. From "impure" Ni = 99.8% pure Ni. **B**. From "pure" Ni = 99.99% Ni. The flat regions (**2**) are the YSZ surface exposed to the H_2 gas, and the uneven structures (**3**) are the YSZ side of the YSZ-Ni interface during the electrode experiment.

Fig. 4.8 explains why the degree of impurity of the electrode and electrolyte material largely determines the electrode polarization resistance.

Fig. 4.8 also illustrates that solid material is rather mobile at the operating temperature. Thus, the need for in situ characterization of the detailed structures of the cell surfaces and especially the composite structures was recognized. This initiated the development of controlled-atmosphere high-

temperature scanning probe microscopy in cooperation with DME – Danish Micro Engineering A/S. Fig. 4.9 shows that the construction of this instrument was successful such that lateral resolution of less than 500 nm of the variation of the electrical properties can be obtained. Controlled-atmosphere high-temperature scanning probe microscopy is continually being refined.

The details of the structure and composition of
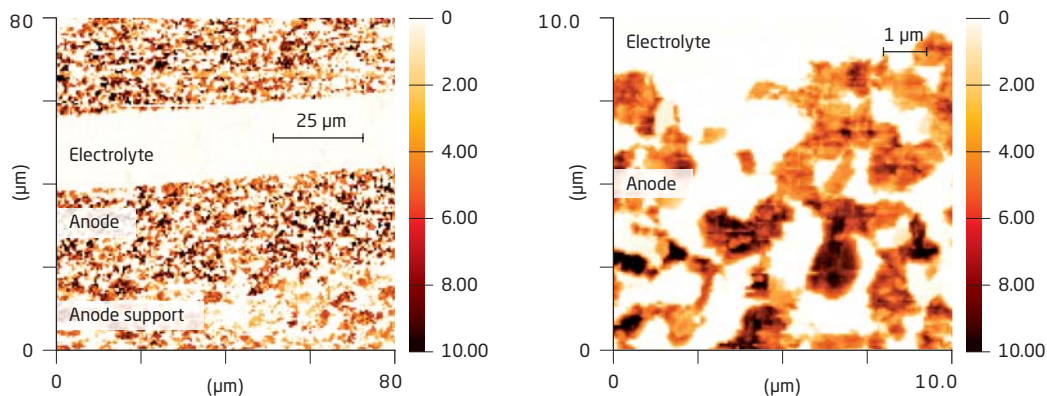
Fig. 4.9. **Left**. Controlled-atmosphere high-temperature scanning probe microscope micrograph of a cross-section of an SOFC anode-electrolyte precursor structure (of a symmetric cell) taken in air at 650°C. The property represented is the conductivity of the surface. The best conducting phase (dark) is NiO and the more resistive phase (white) is YSZ. The technique enables detailed changes to be followed over time in situ, such as during the reduction of NiO to Ni when shifting from air to an $H_2.N_2$ atmosphere. **Right**. The higher magnification shows that the lateral resolution of the controlled-atmosphere high-temperature scanning probe microscopy conductivity mapping at this temperature is well below 500 nm.
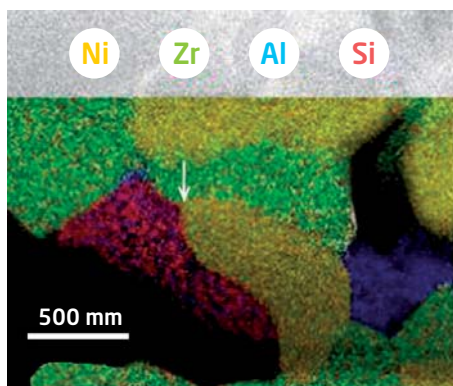
Fig. 4.10. Micrograph showing the chemical composition as measured using energy-dispersive spectroscopy mapping of an SOFC anode using scanning transmission electron microscopy. The colors represent the elements given in colors at the top. An electrical blocking alumina-silicate impurity particle is observed at a point, which probably originally was a three-phase boundary in a Ni-YSZ cermet electrode. The arrow marks the original three-phase boundary, which became blocked by the impurity during operation.

the inside of the composite electrodes have had to be studied ex situ so far. Transmission electron microscopy and scanning electron microscopy with high resolution have been our main workhorses. The capabilities of these techniques were expanded recently by the availability of the focused-ion-beam cutting technique and scanning electron microscopy cross-beam machines, which enables thin slices of porous composite materials to be cut at the desired positions with high accuracy. Fig. 4.10 and 4.11 show examples of results that have been obtained by transmission electron microscopy.

The possibility of cutting an electrode into slices using the focused-ion-beam scanning electron cross-beam microscope reveals the possibility of reconstructing a three-dimensional (3D) image of an actual composite electrode. Combining such 3D structures with electrochemical impedance spectroscopy results for the electrode performance and micromodeling of the electrode structure facilitates correlation of the details of the structure with the electrode performance. Fig. 4.12 illustrates this idea. This enables researchers to predict how the optimal structure should look and thus is a tool for the scientists responsible for optimizing ceramic processing for fabricating improved electrodes.
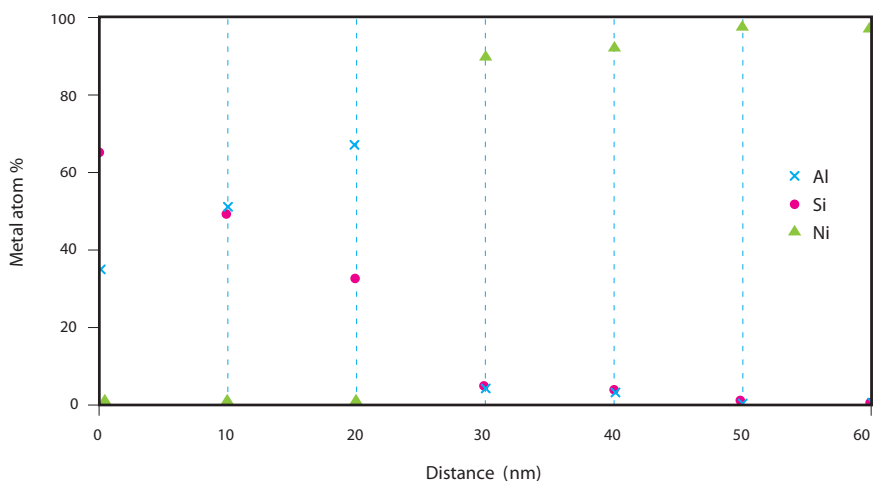
Fig. 4.11. The metal atom percentages of Ni, Al and Si as a function of distance across the interface between an Ni particle and a glassy impurity in a real SOEC hydrogen electrode of Ni-YSZ composite similar to that in Fig. 4.10 as measured by energy-dispersive spectroscopy using transmission electron microscopy. Al (in the form of Al^{3+}) appears to segregate to the Ni, but this may be due to random inhomogeneity in the glassy impurity phase. The resolution is in the nm range.
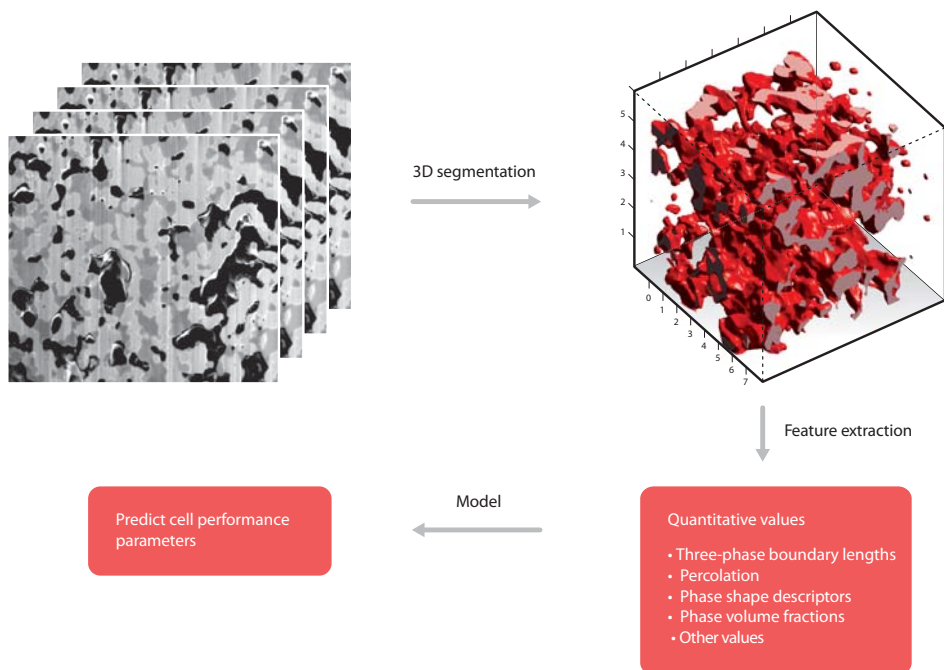
Fig. 4.12. Construction and analysis of electrode structures using a focused-ion-beam cutting and field emission gun scanning electron cross-beam microscope and advanced quantitative image analysis. As the electrochemical performance of the electrode was measured before the microscopy, parameters can be calculated such as the length of the three-phase boundary, specific polarization resistance and, from this, electrode and cell performance may be predicted based on structural parameters.

The authors

Mogens Mogensen:
Research Professor, Fuel Cells and Solid State Chemistry Division, Risø DTU National Laboratory for Sustainable Energy; 36 years in electrochemistry. MSc in 1973 and PhD in 1976 from the Department of Metallurgy, DTU. After a post-doctoral period at the Department of Chemistry A, DTU, and a short period in the battery industry (Hellesens A/S), he was employed at Risø National Laboratory in 1980. Manager for numerous projects and programs and now the Strategic Electrochemistry Research Center (SERC). Has published 250 scientific papers and has 16 patents.

Søren Linderoth:
Professor and Head, Fuel Cells and Solid State Chemistry Division, Risø DTU National Laboratory for Sustainable Energy. MSc in 1984 and PhD in 1987 from DTU. Has published 200 scientific papers and has 20 patents. Has performed research in several fields related to materials science, such as atomic defects studied by positron annihilation techniques, magnetic nanostructures studied by Mössbauer spectroscopy, high-temperature corrosion, electrolytes for solid oxide cells and thin films for functional ceramics.
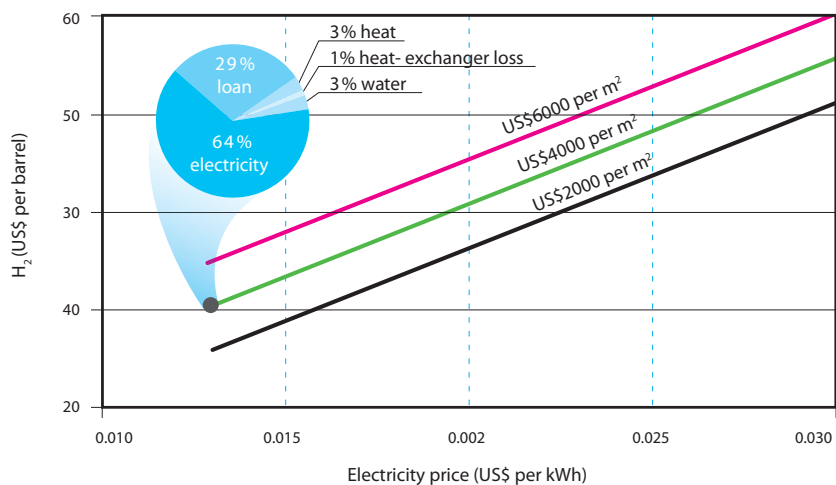
Fig. 4.13. Calculated cost of hydrogen (converted to the equivalent price of US dollars per barrel of crude oil) produced by SOEC as a function of the cost of electricity. The lines represent the stack production costs per cell area. A pie of the distribution of the cost is shown for an electricity cost of US$0.013 per kWh and electrolyzer cost of US$4000 per m².

Economic estimates

Even though predicting the costs of a new technology in a relatively early phase of development is difficult, we think that estimating the order of magnitude of the costs is important. Fig. 4.13 illustrates the case of calculated hydrogen production costs, which has been converted to the equivalent price of US dollars per barrel of crude oil for simple comparison. As both the thermodynamics and the kinetics of electrolysis of CO_2 are very similar to the electrolysis of H_2O, the calculated price of hydrogen given here may indicate the price of synthesis gas. This means that when suitable durability of the cells and stacks has been demonstrated, competitive electrolyzer-based systems should be able to be produced for the production of CO_2-neutral liquid hydrocarbon fuel if the pair of (electricity price and crude oil price) is above the lines in Fig. 4.13. The three lines represent three estimated production costs of given cell areas. The three numbers are meant to be slightly pessimistic (US$6000 per m²), realistic (US$4000 per m²) and optimistic (US$2000 per m²). The situation of an electricity cost of US$0.013 per kWh and electrolyzer cost of US$4000 per m² (the black dot) has been broken down into cost contributions in the pie. This shows that, even if electricity becomes

very inexpensive, it is the main contributor to the cost of hydrogen (synthetic fuel). In other words, the investment costs will be relatively marginal for electricity costs above US$0.01–0.02 per kWh.

Outlook

As mentioned, Topsoe Fuel Cell A/S is starting a pilot plant for cell and stack production in 2009. The electric power generation capacity of the stacks produced each year is 5 MW. These stacks are intended for a demonstration and pre-commercial phase. If this phase is successful, an industrial-scale facility is expected in 2011. If this succeeds, this also means that getting the electrolyzer capability utilized become easier and more realistic. Our vision is that co-electrolysis of H_2O and CO_2 using renewable electricity could supply the ordinary cars of today with CO_2-neutral gasoline in the future due to the possibility of producing synthetic liquid hydrocarbons from synthesis gas, which can be derived from CO_2 and steam by electrolysis using renewable energy sources such as wind and hydroelectric power. Nuclear power companies have expressed great interest in this technology, because the efficiency of nuclear power plants improves if they operate at full power all the time.

Concluding remarks

The Prime Minister of Denmark has recently declared with full support from Denmark's Folketing (parliament) that energy consumption in Denmark must move towards 100% sustainable energy supply free of fossil fuels. This is a wonderful and great challenge for the whole society and especially for us at DTU. The Fuel Cells and Solid State Chemistry Division is working towards this goal with great enthusiasm. It is a special pleasure to carry out research and development when society very clearly expresses its appreciation.

More to explore

Bossel U. *The birth of the fuel cell* 1835–1845. Oberrohrdorf, Switzerland, European Fuel Cell Forum, 2000.

Vielstich W, Lamm A, Gasteiger HA, eds. *Handbook of fuel cells: fundamentals, technology and applications*, Vol. 1–4. New York, Wiley, 2003.

Vielstich W, Gasteiger HA, Yokkokawa H, eds. *Handbook of fuel cells: advances in electrocatalysis, materials, diagnostics and durability*, Vol. 5–6. New York, Wiley, 2009.

Singhal SC, Kendall K. *High temperature solid oxide fuel cells: fundamentals, design and applications*. New York, Elsevier, 2003.

Photonics engineering in a new light

Lars Dittmann & Paul Michael Petersen
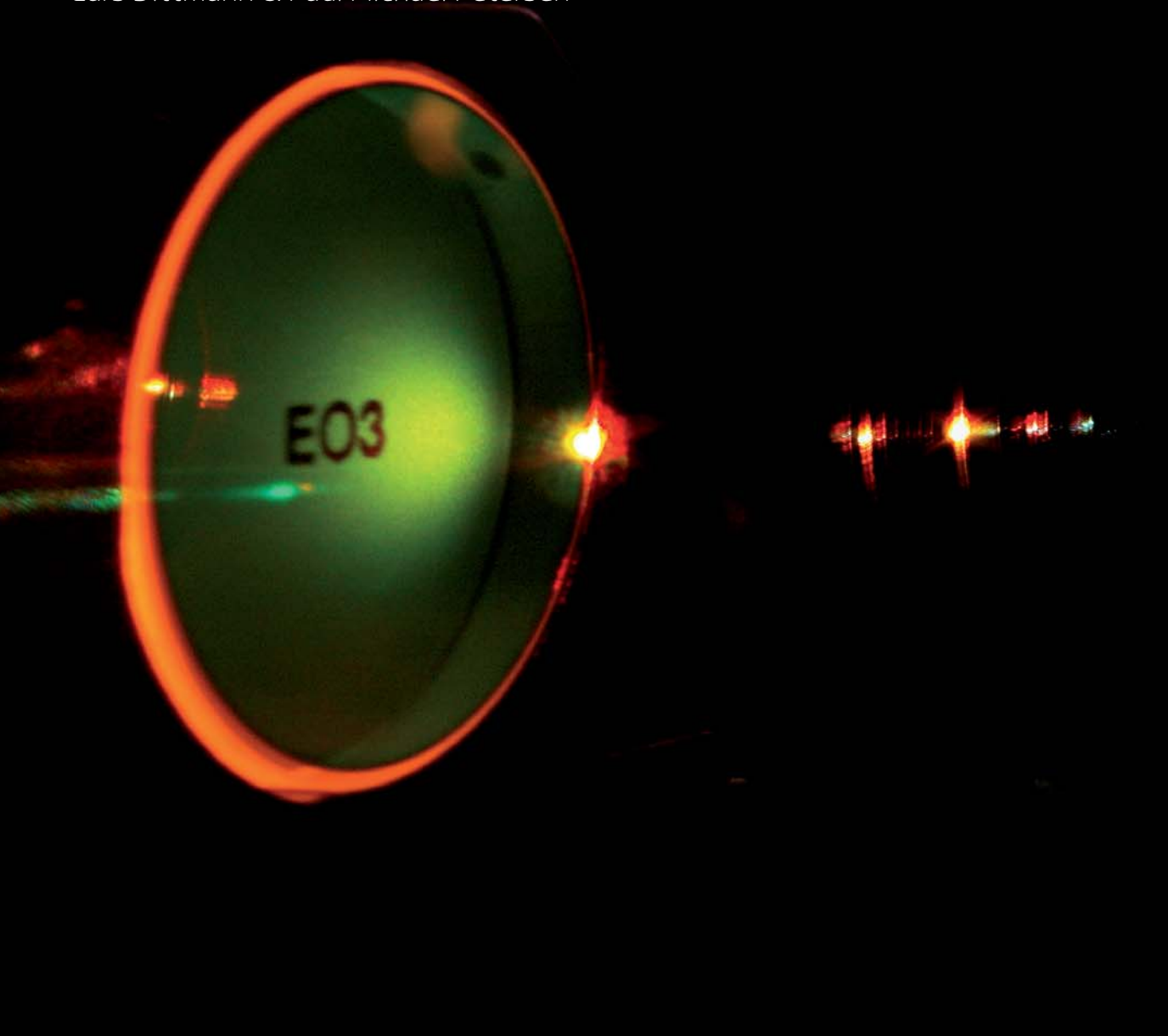
Photonics engineering is an exciting technology that increasingly influences our daily lives. Developing new light-emitting diode (LED) light sources considerably reduces the electricity used in lighting. In medicine, optical technology is enabling new therapies that improve health, and lasers have been one of the key enablers for developing modern information technology and telecommunication.

Further use of photonic technology is seen as a way to limit or even reduce electricity consumption in the continually expanding Internet.

The DTU Department of Photonics Engineering has built strong competencies in photonics. This article reviews how the Department has developed exciting new applications in lighting, medicine and communication.

Energy-saving LED lights may replace incandescent light bulbs

The European Union recently decided to ban incandescent light bulbs to reduce energy consumption. LED can become the perfect energy-efficient substitute for attractive but inefficient incandescent and halogen lighting. No other lighting technology can match the quality, efficiency and lifetime of LED lighting. In contrast to incandescent bulbs that create light and heat from filaments, LEDs convert electricity with high efficiency directly into visible light. However, several technical problems must be solved before LED lighting can break through. The DTU Department of Photonics Engineering is developing future LED light sources with high color rendering that save significant energy. LED lighting has developed tremendously in recent years. The efficiency of LEDs has doubled every 3 years, and LEDs are expected to become more energy efficient than the most efficient conventional lighting sources within the near future. In addition to the energy savings, LEDs have several other advantages: small and compact emitters with high flux, robust, no ultraviolet or infrared radiation when used for visible illumination and long lifetimes (20,000–100,000 hours) provided proper thermal management. In addition to energy savings, money will be saved due to fewer replacements of light sources.

Ongoing projects at the DTU Department of

Photonics Engineering aim at developing high-quality LED light sources with high color rendering for replacing incandescent lamps. These light sources are based on spectral design: mixing colored LEDs and using holographic optical elements for efficient mixing. Novel micro- and nanostructured optical elements are being developed for efficient color mixing and light control. The purpose is to develop high-quality and energy-efficient LED light sources and illumination solutions for numerous specific applications such as work lighting, showcase lighting, greenhouse lighting, path lighting and LED-based optical sensors.

Nanotechnology is essential in developing new enhanced LEDs with higher energy efficiency. Photonic crystals and photonic quasicrystals can help to efficiently extract light from LEDs and to form a desired emission profile.

New LED lamps for the Royal Danish Collections at Rosenborg Castle

An interdisciplinary project between the DTU Department of Photonics Engineering, the Royal Danish Collections and the Danish company Lumodan has developed new LED sources for display cases at Rosenborg Castle. The project, headed by Carsten Dam-Hansen of the DTU Department of Photonics Engineering, has demonstrated new LED light sources with variable color temperature down to a correlated color temperature of 2200 K that demonstrate very high color-rendering properties suitable for illuminating display cases in the Royal Treasury.

Rosenborg Castle used 5-W incandescent bulbs with a correlated color temperature of 2200 K in their display cases for many years. The aims of developing a new LED light source were to achieve the same high level of color rendering as for the 5-W incandescent bulbs, to reduce the power consumption and to remove radiant heat from the light source.

The LED technology used in our light sources is based on spectral mixing of LEDs and holographic diffusers that led to a uniform white light distribution in the light beam. Fig. 5.1 shows the spectral distributions of the light from the 5-W incandescent bulb and the new LED source operating at a correlated color temperature at 2200 K. The spec-
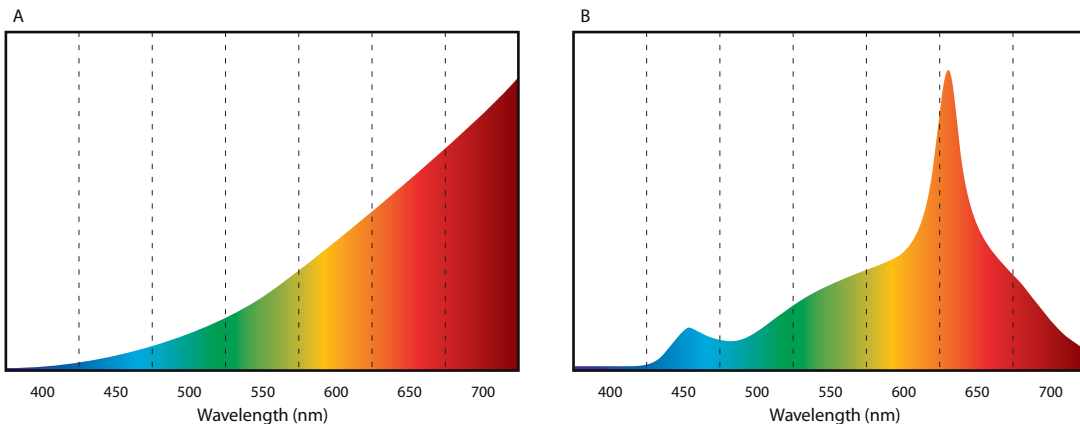
Fig. 5.1. **A**. Measured spectral distribution of a 5-W incandescent bulb, which has a correlated color temperature of 2200 K and a color-rendering index of 99. **B**. Measured spectral distribution of the new LED light source adjusted to a correlated color temperature of 2200 K and a color-rendering index of 94.7.

tral distribution of the light is used for calculating the correlated color temperature and color-rendering index of the light sources. The wavelength distribution of the new LED source in Fig. 5.1B shows a strong red content and little blue and green light, which is almost similar to the spectral distribution of the incandescent light bulb in Fig. 5.1A. The new LED light source has a color-rendering index of 94.7, meaning that colored objects are rendered as naturally as under the incandescent light. A color-rendering index of 100 yields color rendering 100% identical to the incandescent light.

Fig. 5.2 shows photos of gold objects under LED illumination with decreasing color temperature in the range from about 3000 K to 2200 K. In Fig. 5.2a, a commercial warm-white LED is used. The research shows that the appearance of the gold objects differs widely under the different types of illumination. Even a small variation in color temperature is clearly visible. Illumination with the new LED source in Fig. 5.2 b–f with color temperature from 2760 K to 2200 K provides illumination with a high color-rendering index up to 97.1. At 2200 K, the new LED source produces illumination that is visually identical to that produced by the 5-W incandescent bulbs.

The results of our research show that new LED light sources can be spectrally designed to replace 5-W incandescent bulbs with the desired color qualities and without the heat radiation that is an inherent part of the light from incandescent bulbs. In addition, the LED light source saves significant

energy compared with the incandescent light bulbs. Using the new LED light source in a display case reduces energy consumption by about 70% compared with the 5-W incandescent bulbs.

New optical techniques for medical applications

Within medicine, optical techniques are used increasingly to develop new therapeutic and diagnostic methods. Optics is providing new tools to physicians for noninvasive diagnostics and for early detection of diseases.

Developing state-of-the-art laser systems is of utmost importance for deriving new clinical procedures, diagnostics and therapy and for biochemical quantification. The goal of our research activities at the DTU Department of Photonics Engineering is to develop new lasers and systems with improved characteristics in terms of beam quality, wavelength, tunability of laser wavelength, output power, noise and coherence properties.

The Department is developing high-brightness diode lasers. These high-brightness, external-cavity laser diodes emerge as the next generation of compact lasers that have the potential of replacing conventional high-power laser systems, such as neodymium:yttrium aluminum garnet (Nd:YAG) lasers and CO_2 lasers in many existing applications. The laser diodes have high output power and unique beam quality, which enables the output beam to be focused on a diffraction-limited spot.

Fig. 5.2. Photos of gold samples under LED illumination at different correlated color temperatures. **a**. Illumination of gold samples using a commercial Lumileds LED. **b–f**. Illumination of gold samples when the correlated color temperature of the new LED light source is reduced from 2760 K to achieve the desired 2200 K (F). CRI: color-rendering index.

Further, the laser diodes may be realized as tunable laser systems at almost any wavelength in the blue, green, red and near-infrared regions. One example of such a compact high-brightness laser system from the Department is a single-mode infrared laser. The source is tunable in the range from 788 to 815 nm with an almost diffraction-limited beam over the entire tuning range and with an output power of up to 2 W and output power that is four times the previous state of the art. The research at the Department also focuses on the important development of compact green and blue lasers based on second-harmonic generation of infrared and red high-brightness laser diodes. We recently developed a frequency-doubled 531-nm laser with output power up to more than 1.5 W. This is a world record for green frequency-doubled diode lasers.

The following section describes a new state-of-the art blue diode laser system that can be used for diagnosing cancer. The new laser system has been used for feasibility tests in fluorescence diagnostics during photodynamic therapy. Photodynamic therapy is a cancer treatment that uses a light-sensitive drug and a laser of a particular wavelength. The photosensitive material accumulates selectively in the cancer cells and, when illuminated, singlet oxygen is formed that leads to necrosis of the closest available cancer cells.

A new blue laser system for fluorescence diagnosis of cancer

The development of blue lasers is highly interesting because these lasers allow physicians to diagnose the presence of cancer cells but also because physicians can provide better cancer treatment. We recently developed a new blue 404-nm frequency-doubled laser system based on a tapered 808-nm pump laser diode with external feedback. Using a new very compact cavity design, shown in Fig. 5.3 (right), we have demonstrated highly efficient frequency doubling. The external cavity consists of two reflecting concave mirrors and a nonlinear crystal (periodically poled potassium titanyl phosphate) placed between the mirrors. One of the mir-
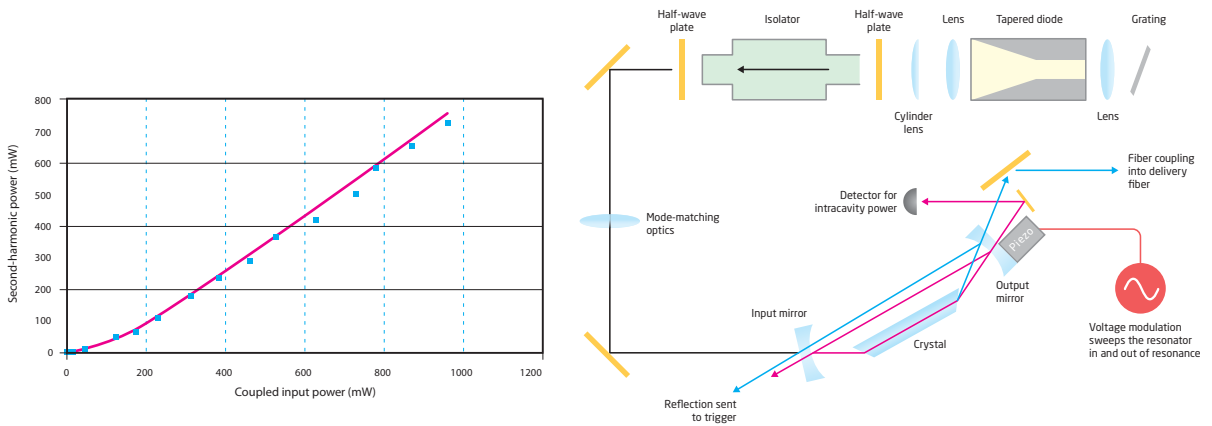
Fig. 5.3. **Left**. The blue output peak power at 404 nm (blue light) versus the infrared power in the external cavity. The blue laser light may be used for diagnosing cancer using fluorescence. **Right**. The experimental set-up with second-harmonic generation in an external cavity. The output laser beam from an infrared tapered diode laser is sent through an optical insulator and into an external cavity consisting of two mirrors and a nonlinear crystal that converts the infrared radiation into blue light.

rors in the cavity is piezo-controlled and thus can be used to sweep the cavity in and out of resonance, which allows for pulsed second-harmonic generation.

The pulsed laser system is attractive because time-gated detection allows suppression of background light during fluorescence diagnosis. Fig. 5.3 (left) shows the peak powers of the pulses versus the coupled input power in the external cavity. The pulses produced are up to 720 mW in peak power. This is the highest second-harmonic generation power obtained in periodically poled potassium titanyl phosphate at 404 nm.

Our 404-nm laser system, operating in continuous-wave mode, has been used in the clinic for fluorescence diagnostics on patients that were treated with aminolevulinic acid photodynamic therapy for skin cancer. The blue laser output illuminates

the cancer lesion, and the induced fluorescence is picked up after a bandpass filter that transmits specific wavelength regions. The laser system was moved to Lund Laser Centre in Sweden to perform clinical experiments on a patient with cancer lesions between the left eye and the left ear. Fig. 5.4A and B show different wavelength regions before and after the photodynamic therapy, displaying the fluorescence peaks at 635 nm and 700 nm before the photodynamic therapy. Fig. 5.4B shows that these fluorescence peaks disappeared after the treatment. The fluorescence signal before and after the treatment can be measured to determine whether the treatment is completed. The future research challenges in laser-induced fluorescence include developing new lasers with high peak power in the blue and ultraviolet part of the spectrum. In particular, for diagnostic autofluorescence imag-
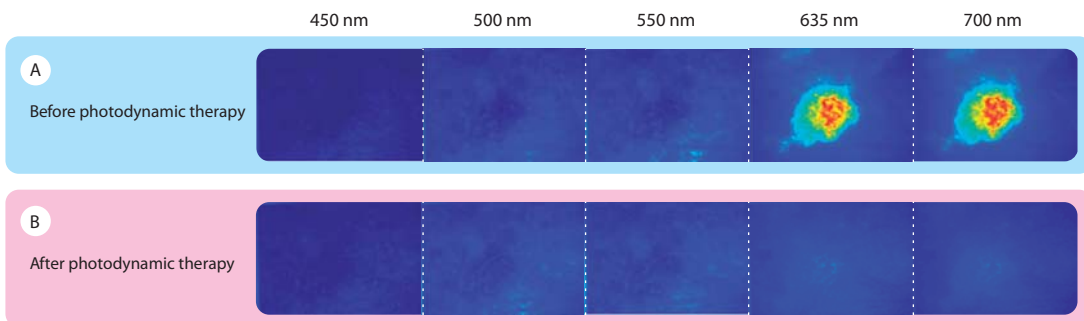
Fig. 5.4. Fluorescence imaging in a clinic with the blue 405 nm laser. **A**. Before the aminolevulinic acid photodynamic therapy. **B**. After photodynamic therapy.
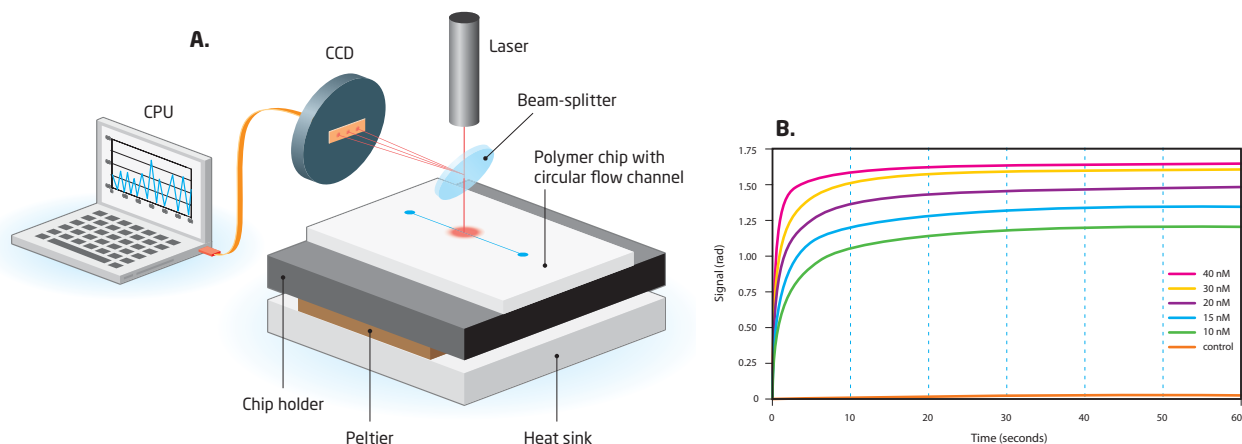
Fig. 5.5. **A**. Schematic of the backscattering interferometry principle. **B**. Real-time association plots are shown for protein A–binding IgG at various nanomolar concentrations. CCD: charge-coupled device. CPU: central processing unit. rad: radiation absorbed dose.

ing, a 340-nm source is required for screening applications in which avoiding administering any exogenous fluorescence marker to a large patient population would be beneficial. This clinical application will minimize any risk of side effects of the examination.

Backscattering interferometry – label-free biosensing

At the DTU Department of Photonics Engineering, we have developed new optical biosensors that can be used for measuring small concentrations of proteins. A promising biosensor is backscattering interferometry (Fig. 5.5.A), a refractive index measuring technique based on light interacting with a microfluidic channel. In this method, the physical variable being measured is the change in refractive index with time, which can be caused by numerous bulk properties or solute interactions. Backscattering interferometry has the highest figure of merit among candidate detection schemes, considering the combination of sample volume, time, sensitivity and system simplicity for doing label-free, in-solution protein–protein interaction studies. Fig. 5.5.B shows the real-time detection of protein A–binding immunoglobulin G (IgG) at various nanomolar concentrations. The capacity to measure any circulating disease marker or mediator in blood samples can develop numerous applications in clinical medicine. In addition, backscattering interferometry may be applied to other areas where analyzing small volumes of liquid samples is desir-

able, including drug discovery and environmental monitoring.

Monitoring this shift is the basis of this sensor. The capillary tube optical train has been modeled with high accuracy. One major step forward in developing backscattering interferometry was to transition from capillary tubes and into the microfluidic network. We recently showed that backscattering interferometry is capable of detecting label-free and free-solution binding with unprecedented detection limits in a microfluidic device. The unique optical train used in backscattering interferometry allows near real-time quantification of solutes at zeptomole levels, corresponding to changes in refractive index at the 10^{-6} level, within detection volumes of tens of picoliters and binding constants spanning six orders of magnitude from pM to μM. Such a wide dynamic range is relevant because most circulating molecules of pathophysiological interest in human disease can be found at levels from fM to μM.

Towards the next generations of Internet – with higher capacity and lower power consumption

We are living in a globalized world or a global village. Distances are no longer as important as before, and many people visit remote places all over the world, even for relatively short visits.

This trend also applies to the communication infrastructure known as the Internet. Today users retrieve information from remote corners of the

world just as commonly as the latest news from the local supermarket. Distance is no longer an issue, and users do not care where the information they retrieve is stored.

The key invention that has enabled this revolution is the optical fiber, which has allowed signals to be transmitted with an extremely high information rate over very long distances.

Researchers at DTU recently broke the terabit per second wall. For the first time in history, they demonstrated that a single wavelength can carry an information rate at 1.2 Tbit/s – or 1,200,000,000,000 bits per second.

The record high bitrate was achieved by using an all-optical synchronous time domain multiplexing technique (Fig. 5.6).

At a time when 2.5 Gbit/s is the most common bitrate for optical networks and 10 Gbit/s is about to become the most cost-efficient solution for new installations, the need for Tbit/s operation might seem in the distant future.

However, after a century with an access network based on copper wires, the access network is gradually being replaced with optical fibers (complemented by wireless mobile access for less demanding applications). Initially, the bitrate provided on the fiber to the home is 10–100 Mbit/s. But just as the bitrate on the copper cable went from kbit/s to Mbit/s in less than a decade, the fiber to the home system can easily be upgraded from Mbit/s to Gbit/s – when demand arises and the backbone is capable of handling it.

With respect to demand, video applications finally are about to become the killer application (although many people might start to consider that we are crying wolf). While broadcasters are working on providing high-definition TV, it has already become a commodity for home video applications, as the price of camcorders has declined and home videos in high definition are shared on YouTube and similar services. While broadcasters are working hard on providing airborne TV in high definition, wired TV over the Internet in cinema quality – 4k video – is evolving fast in parallel and might surpass traditional high definition.

So the question is not when Tbit/s optical systems will be needed but rather how to integrate them efficiently into the network. Even system bitrates of 100 Gbit/s are a challenge to integrate into

1280 Gbit/s

6.25 ps
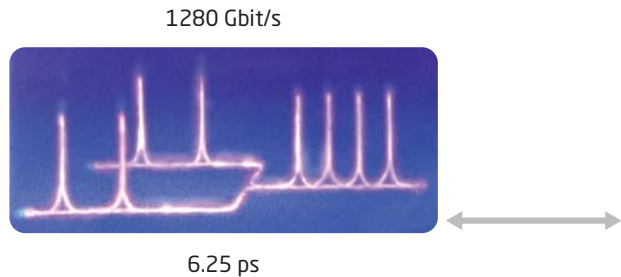
Fig. 5.6. Optical time domain multiplexing of short optical pulses used to achieve 1.2 Tbit/s capacity in a single wavelength

the networks, as the internal speed inside the network nodes is much higher due to the need to speed up to handle several inputs and outputs at the same time.

Moore's law for integrated electronics is often referred to as the explanation for the continual growth in the capabilities of computers and electronic devices. But transmission capacity is growing even faster, and the span between the capacity of a single fiber or wavelength and the capability of processing Internet-related protocols is continually increasing, and new paradigms for core network implementation are required.

An additional impact of pushing electronics to its very limit to cope – as well as possible – with the advantages of optical transmission is an extreme increase in power consumption. The Internet is based on protocols with intensive use of memory and complex algorithm processing. The Internet has become a global success as it enabled global connectivity in a simple manner at a time when computer networks were in their infancy and being built on an ad hoc basis. But the current implementation of the Internet is based on conditions and assumptions that are 35 years old and are today more important for service integration rather than for network integration. The use of classical Internet technology in the global infrastructure is too expensive both in terms of equipment cost but even more important in terms of electricity consumption.

The next generation of Internet(s) must be simplified to enable proper integration of electronics and photonics and to reduce electricity consumption (Fig. 5.7). Today large network operators consume more than 1 TWh annually.
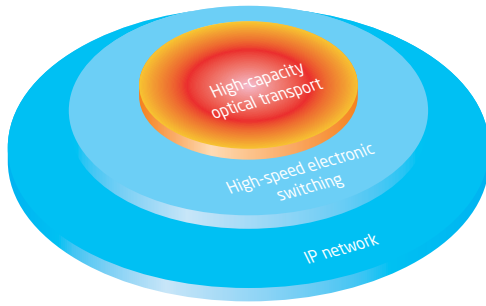
Fig. 5.7. Future use of Internet protocol (IP) as application or service integrating technology: edge technology

Rethinking the network structure

Globally, communication infrastructure is responsible for more than 2% of all CO_2 emissions, and typical equipment costs more to power during its lifetime than the actual cost of purchasing it. With the continual growth, this proportion is rising if the current strategy is maintained and could easily reach 10% within the next 15–20 years.

So many factors motivate rethinking the whole network structure. Simplifying the functionality in the core network can limit the memory and processing power required. But minimizing continuous translation from optical to electrical signal – and back again – can save considerable electricity. Ideally, the signal should remain in the same domain as long as possible. As only optical fiber can

support the requirements for high-capacity, long-distance transmission, the goal should be to design network nodes where transit traffic remains in optical format while passing the node.

Several attempts have been made to replace functions in electronic domain with optical counterparts. However, simple 1:1 substitution cannot be done. Optical components cannot be integrated as electronic components, and optical memory is still not a reality. The network must be designed according to existing components under the assumption that demand can optimize the cost of the components.

Optical technology is hard to integrate efficiently into the Internet because it is best suited for circuit switching or synchronous time division, whereas the Internet and its protocols are based on packet switching and asynchronous computer architectures.

Some years ago, we built an all-optical packet switch network together with partners in the DAVID project. But component technology is still not sufficiently mature for further development in this direction, and the focus today is on intermediate steps towards all optical networks based on time-domain multiplexing.

Recent research into network design has primarily focused on new network node design and methods of combining resources with different capacity granularity to offload the power-consuming electrical nodes and provide power and cost-efficient end-to-end services (Fig. 5.8).

This has the clear advantage of being capable of exploiting new developments in optical component technology and optical multiplexing techniques. Circuits provided as either wavelength or sub-wavelength can be complemented with optical burst and packet switching when the technology matures.

Fig. 5.8. New network architecture that enables the integration of different transport network technologies. PBT: provider backbone transport. GMPLS: generalized multi-protocol label switching. TMPLS: transport multi-protocol label switching. NRPS: network resource provisioning system.
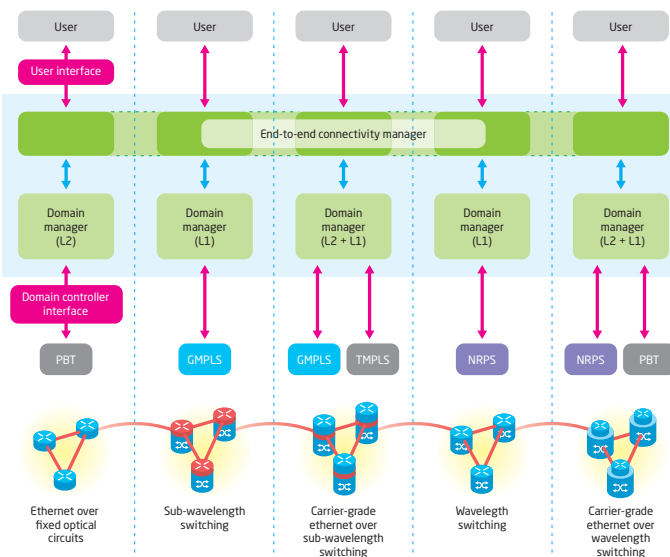
More to explore

Chi M, Jensen O, Holm J, Pedersen C, Andersen P, Erbert G, Sump B, Petersen PM. Tunable high-power narrow-linewidth semiconductor laser based on an external cavity tapered amplifier. *Optics Express* 2005: 13: 10589–10596.

Jensen OB, Andersen PE, Sumpf B, Hasler K-H, Erbert G, Petersen PM. 1.5 W green light generation by single-pass second harmonic generation of a single-frequency tapered diode laser. *Optics Express* 2009: 17: 6532–6539.

Lundeman J, Jensen O, Andersen P, Andersson-Engels S, Sumpf B, Erbert G, Petersen PM. High power 404 nm source based on second harmonic generation in PPKTP of a tapered external feedback diode laser. *Optics Express* 2008: 16: 2486–2493.

Lundeman J. *Blue laser source for laser-induced fluorescence*. PhD dissertation. Lund, Lund University, 2008 (Lund Reports on Atomic Physics, LRAP-396).

Sørensen HS, Larsen NB, Latham JC, Bornhop DJ, Andersen PE. Highly sensitive biosensing based on interference from light scattering in capillary tubes. *Applied Physics Letters* 2006: 89: 151108.

Bornhop DJ, Latham JC, Kussrow A, Markov DA, Jones RD, Sørensen HS. Free-solution, label-free molecular interactions studied by backscattering interferometry. *Science* 2007: 317: 1732–1736.

The authors

Lars Dittmann:
MSc, 1988 and PhD, 1994, DTU. Head, network competence (both optical and electrical networks), Research Center COM, DTU, 1999–2008. Before that, responsible for electronic switching and ATM networks at the Center for Broadband Telecommunication. Currently Professor and group leader, DTU Department of Photonics Engineering, responsible for network technology and service platforms and coordinator for the DTU communication cluster. Involved in several EU and national projects on packet switching, contributing traffic and protocol analysis and design of electronic integrated circuits and systems.

Paul Michael Petersen:
MSc in engineering and PhD in physics, DTU. Currently Head of Research, DTU Department of Photonics Engineering and Adjunct Professor of Optics, Niels Bohr Institute, University of Copenhagen. Head, Center for Biomedical Optics and New Laser Systems, DTU. Has 20 years of research experience in laser physics, nonlinear optics and optical measurement techniques and has headed several collaborative research projects within laser physics. Has authored or coauthored more than 100 scientific publications and 9 patents.
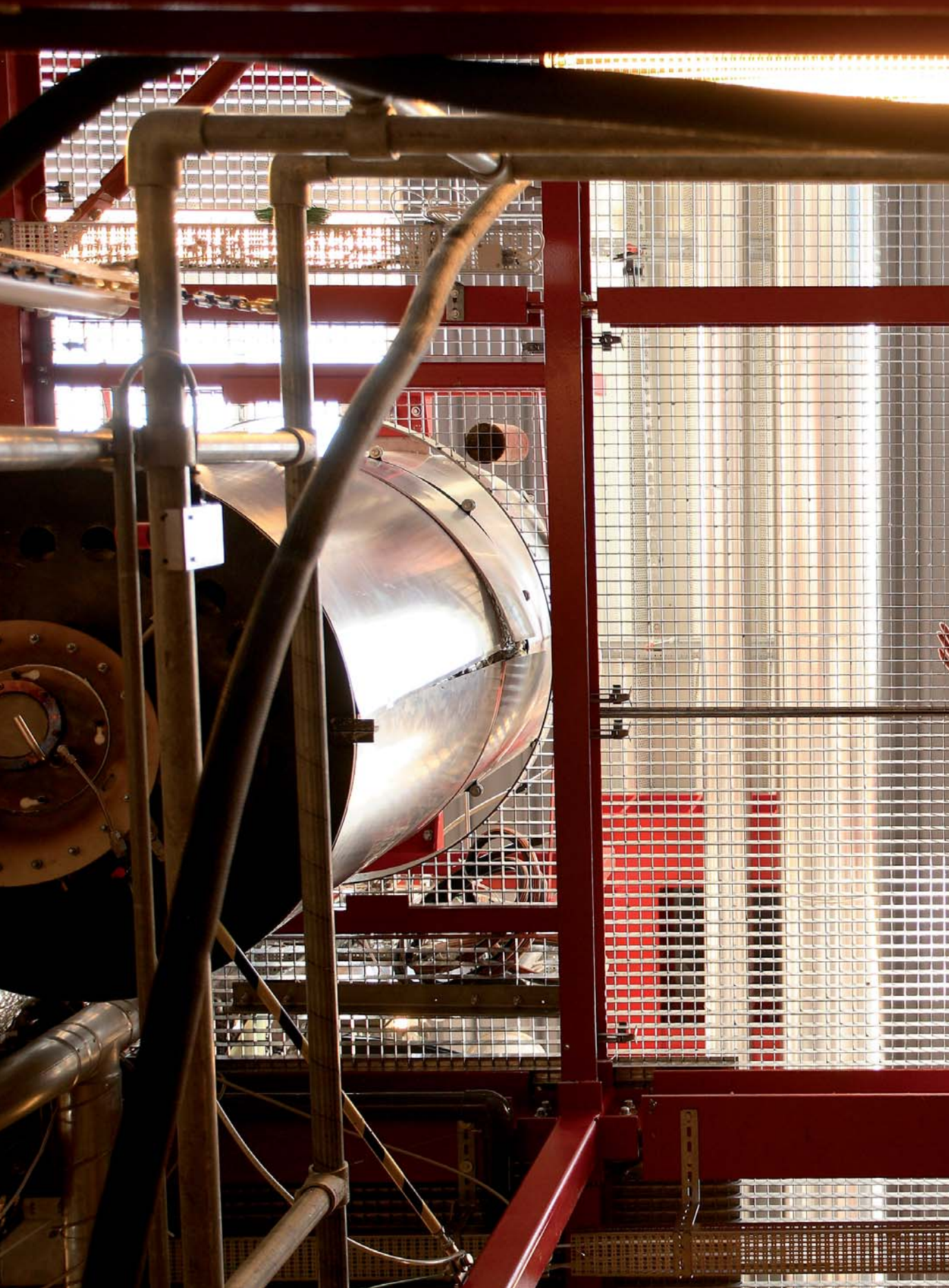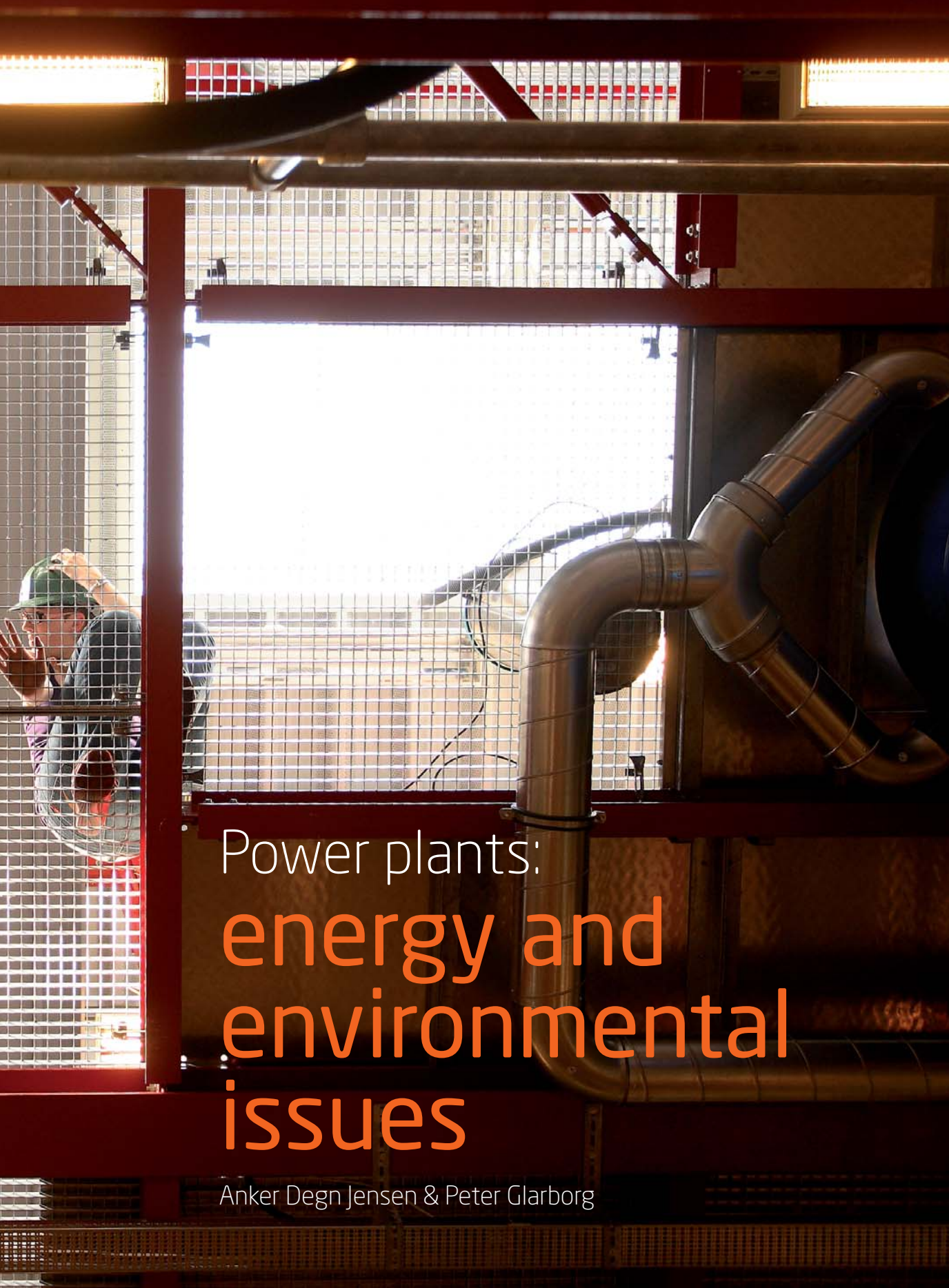
Power plants:
energy and environmental issues

Anker Degn Jensen & Peter Glarborg

King Edward I of England banned the use of coal in London as early as 1272 due to the heavy formation of smog. Nevertheless, coal continued to be used out of necessity even though breaking the law would lead to execution. Today, the supply of energy must be sustainable and minimize harmful emissions to the environment. Modern society increasingly depends on a continuous supply of energy; now (besides heat) in a refined form as electricity and high-energy-density liquid fuels for transport purposes.

The timeline in Fig. 6.1 shows the challenges the power sector in Denmark (and most high-income countries) has encountered during the past 50 years. In the 1970s, oil was the most common fuel for generating electricity, but the oil crises led to rapid conversion to coal, which was less expensive, more abundant and geographically spread.

The CHEC (Combustion and Harmful Emission Control) Research Center at the DTU Department of Chemical and Biochemical Engineering was founded in 1987 to establish fundamental science-based knowledge to advance power plants and energy-intensive industrial processes with the highest efficiency and lowest emission of harmful pollutants (Box 6.1).

The CHEC Research Center program originates from the concern about forest death in central Europe and the acidification debate, and the initial research projects were related to processes for reducing emissions of nitrogen oxides (NO_x) and sulfur dioxide (SO_2) – the main contributors to acid rain. Since then, additional harmful components from thermal processes have been identified, and flue gas cleaning equipment constitutes a major part of a modern power plant (Fig. 6.2).

The stack in Fig. 6.3 illustrates the typical concentration levels of several relevant species in the flue gas from combustion processes before any cleaning. The emission of a species is given as flow times concentration. The flue gas flow is related to the efficiency of the process, and the concentration in the flue gas depends on the fuel and process conditions. The electrical efficiency η_{el} (the ratio of the power output to the input of chemical energy in the fuel) is limited by the temperature and pressure of the steam put into the turbine. Modern coal-fired power plants in Denmark are among the most efficient in the world, with electrical efficiency close to 50% versus the European average of about 35%. The overall energy efficiency can be improved further by using the waste heat for district heating. This is common in Denmark but not very common in many other European countries. The high efficiency thus leads to reduced emissions due to the lower flue gas flow. Nevertheless, the content of nitrogen, sulfur and chlorine in the common fuels (Table 6.1) results in considerable concentrations of NO_x, SO_x and HCl in the flue gas, which must generally be handled by post-combustion cleaning processes. In the 1990s, decentralized production of electric power and heat was promoted and there was a desire for greater fuel flexibility and the use of renewable energy sources such as waste and agricultural residue. The use of biomass introduced considerable technical challenges, since its content of inorganic elements such as chlorine, sulfur and alkali metals led to operating problems with in-furnace deposition and corrosion. Consequently, research in inorganic chemistry and ash transformation became an important part of

Fig. 6.1. Timeline showing the challenges of the electric power sector over the past 50 years. The CHEC Research Center was founded in 1987.
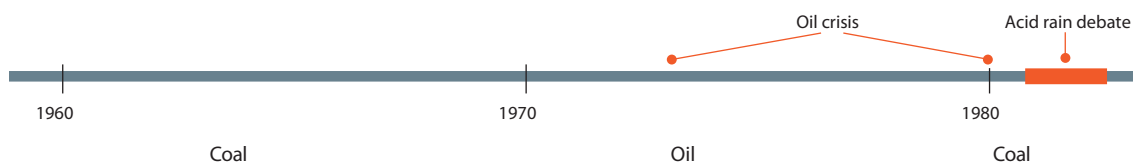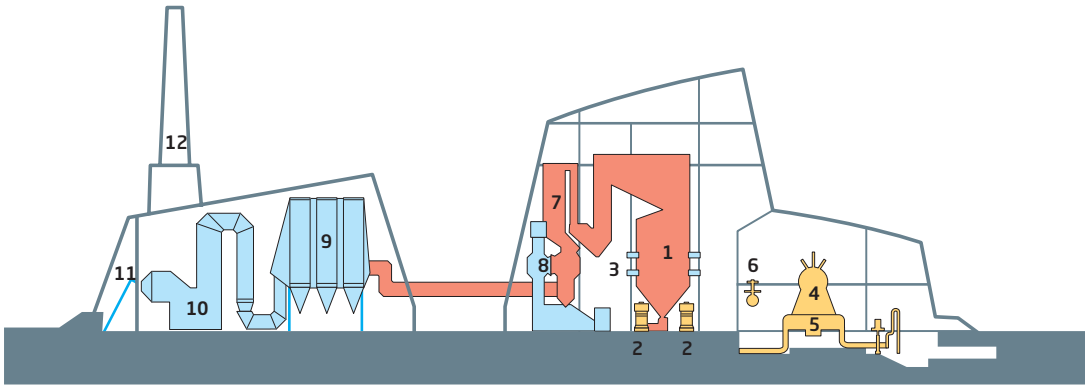
Fig. 6.2. Layout of a typical pulverized fuel power plant (Avedøre unit 1). **1**. Combustion chamber. **2**. Coal mills. **3**. Burners. **4**. Turbine. **5**. Condenser. **6**. Feed water pump. **7**. Removal of NO_x from flue gas. **8**. Air preheater. **9**. Electrostatic precipitator for capturing particles. **10**. Removal of SO_2 from flue gas. **11**. Gypsum storage. **12**. Chimney. Courtesy of DONG Energy A/S.

Concentration (in molar fraction) of typical species in the flue gas from combustion processes

1	N_2
10^{-1}	H_2O, CO_2, O_2
10^{-2}	
10^{-3}	SO_2, NO Particulate
10^{-4}	HCl, CO
10^{-5}	SO_3, N_2O, NO_2
10^{-6}	Na, K
10^{-7}	Cd, Hg, particulate organic matter,
10^{-8}	polyaromatic hydrocarbons

Emission = $\underset{f(\eta)}{\text{flow}} \times \underset{f(\text{process, fuel})}{\text{concentration}}$
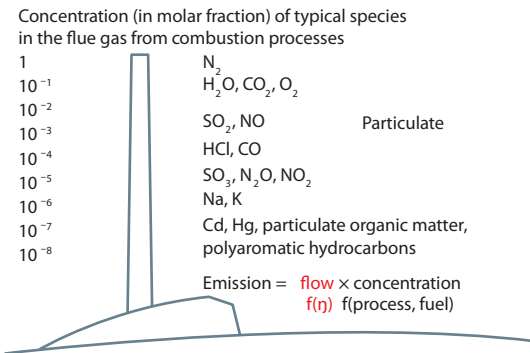
Fig. 6.3. Concentration (in molar fraction) of typical species in the flue gas from combustion processes

the CHEC Research Center program. The research on the combustion of renewable fuels has become even more important, since CO_2 has been added to the list of undesired emissions, being a greenhouse gas leading to global warming. Biomass and waste are considered CO_2-neutral fuels, and major efforts are being made to facilitate the use of these fuels for generating electricity. Already in the mid-1990s, the CHEC Research Center assisted Denmark's electric power industry in conducting the first full-scale global tests of coal and straw

co-firing at centralized power plants; use of CO_2-neutral fuels is now seen as a major path toward more sustainable generation of electric power and heat.

The rest of this chapter presents results from past and ongoing research at the CHEC Research Center on controlling pollutant emissions and discusses some future challenges.

Fuels	N	S	Cl
Fuel oil	0.2–0.9	0.1–4.5	<0.1
Straw	0.3–1.5	0.10–0.24	0.1–1.7
Other annual biomass	0.1–3.5	0.03–0.6	0.01–0.6
Wood	0.03–1.0	<0.10	<0.10
Coal	0.5–2.5	0.3–4.3	0.01–0.10
Plastics[a]	0.00–0.01	0.0	0.0; 50
Paper	0.1–0.2	0.05–0.3	0.03–0.40
Residential solid waste	0.2–1.0	0.10–0.4	0.1–0.9

[a]Plastics such as high-density or low-density polyethylene contain no chlorine, whereas polyvinyl chloride contains about 50%.

Table 6.1. Concentration (mass percentage) of nitrogen, sulfur and chlorine in selected fuels

Global CO_2 alert

1987   1990   2000

CHEC

Biofuels and waste

New processes: gasification, oxyfuel, CO_2 absorption etc.

Box 6.1. The CHEC (Combustion and Harmful Emission Control) Research Center

The CHEC Research Center has about 70 researchers, students and technicians. A vital part of the research is conducted in close collaboration with industry and international research organizations. The industrial relations cover close collaboration on joint projects with mutual exchange of staff and collaboration on experimental research. This ranges from microscale and pilot plants (Fig. 6.4) to full-scale industrial production plants combined with mathematical models of processes at scales from the molecular level to full-scale plants in operation. This approach ensures highly relevant research and efficient exchange of technology, expertise and know-why. The Center's research fields include:

- inorganic metal species in high-temperature processes;
- high-temperature heterogeneous reactions;
- high-temperature homogeneous reactions;
- low-temperature gas cleaning processes;
- fluid dynamics in chemically reacting systems; and
- laboratory facilities and measuring techniques.

Fig. 6.4. The CHEC semi-industrial scale facilities for combustion, gasification and flue gas cleaning

Emission of sulfur oxides

Most fossil fuels, as well as biofuels and household waste, contain sulfur (Table 6.1). The sulfur is largely released to the gas phase during combustion, either as simple species such as H_2S or as complex organic compounds. Following release, the gaseous sulfur is oxidized rapidly to sulfur oxides, mainly sulfur dioxide (SO_2). A minor part of the SO_2 may be oxidized further to sulfur trioxide (SO_3). Sulfur trioxide is thermodynamically favored at lower temperatures, but kinetic limitations often prevent an SO_3/SO_2 partial equilibrium from being attained. The presence of SO_3 is undesirable in the combustion process, since it enhances corrosion problems, mostly by promoting chlorine release in the superheater deposits ($2KCl + SO_3 \rightarrow K_2SO_4 + HCl$).

Because of the acidification problems, legislation
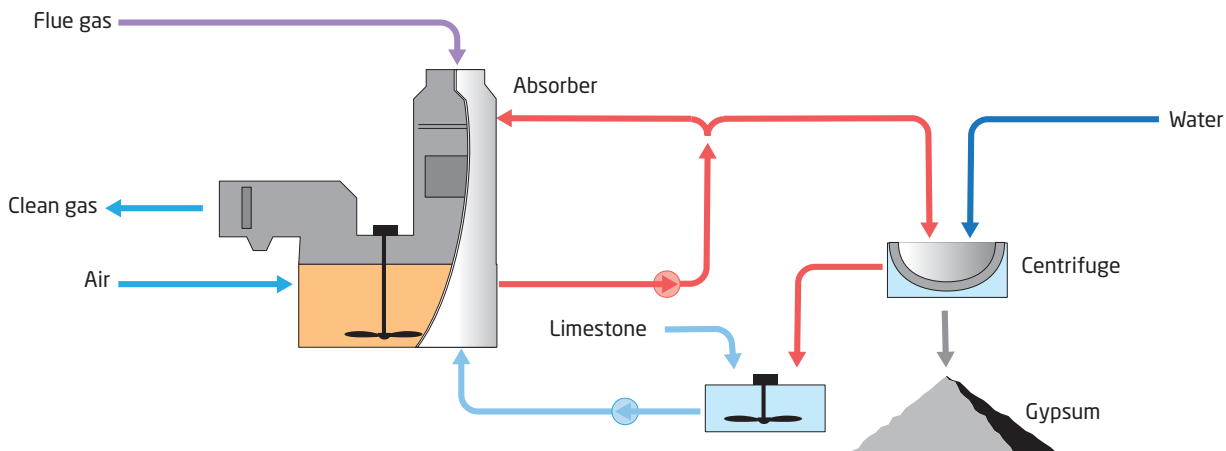
Fig. 6.5. Process for wet flue gas desulphurization (stage 10 in Fig. 6.2). Flue gas containing SO_2/SO_3 is led into a large reactor, where it reacts with a slurry of finely ground limestone ($CaCO_3$) in water. The sulfur species are absorbed into the liquid phase, where the reaction $SO_2 + CaCO_3 + 0.5O_2 + 2H_2O \rightarrow CaSO_4, 2H_2O + CO_2$ takes place. The formed gypsum ($CaSO_4, 2H_2O$) is filtered out of the slurry and sold to the plasterboard industry. CHEC has contributed to developing and optimizing this technology through pilot and full-scale experiments and mathematical modeling. Pilot-scale experiments showed the potential of adding of trace amounts of organic acid buffers. This method is now used at several wet flue gas desulfurization plants, pushing the degree of desulfurization from about 92–95% up to close to 100%. Current research deals with foaming behavior, gypsum crystal formation and use of flue gas desulfurization under oxyfuel combustion conditions.

on the emission of sulfur oxides was introduced in the mid-1980s and increasingly tightened over the years. In 2000, a tax of €1.30 per kg of SO_2 emitted was imposed on large power plants in Denmark (exceeding 25 MW of electricity). These steps are the main reason for the decrease in Denmark's annual SO_2 emissions from 173,000 tons in 1987 to about 9000 tons today.

The main technology that has enabled this emission reduction is wet flue gas desulfurization (Fig. 6.5). This technology is very efficient and converts an undesired pollutant to a useful byproduct: gypsum. Recent advances in the process have allowed power plants to push the degree of desulfurization from about 92–95% up to close to 100%, and the problem of emitting sulfur species from large power plants has largely been solved.

Emission of nitrogen oxides

Emission of NO_x from combustion and high-tem-

perature industrial processes continues to be a major environmental concern. NO_x is formed either from the fixation of N_2 in the combustion air at high temperatures or from the oxidation of nitrogen chemically bound in the fuel. Nearly all combustion processes lead to formation of NO_x, emitted mostly as NO, with smaller amounts of NO_2 and even N_2O. NO is subsequently oxidized to NO_2 in the atmosphere. NO and NO_2 are acid-rain precursors and participate in generating photochemical smog, whereas N_2O is a greenhouse gas.

In large-scale power plants, NO_x is formed mainly from oxidation of the fuel-N. Unlike other emissions that arise from fuel impurities, oxidation of nitrogen species may yield a harmful product (NO) or a harmless component (N_2), depending on combustion conditions. NO_x may be abated by modifying the combustion process to limit the oxygen availability in the flame. Unfortunately, attempts to control NO_x through delayed mixing often result in problems with incomplete fuel oxida-

tion and enhanced deposition and corrosion. Large-scale power plants often prefer downstream flue gas cleaning, preferably selective catalytic reduction of NO, a very efficient NO_x control technology.

Following SO_x, legislation imposing an overall NOx emission limit for electric power plants was introduced in the 1980s. This caused the power plants to install selective catalytic reduction on the newer plants and low-NO_x burners on the older units. This has reduced the annual NO_x emissions from about 170,000 tons in 1987 to about 40,000 tons today (Fig. 6.6). In 2010, a tax of €0.67 per kg of NO_2 emitted will be imposed on large units, facilitating further reduction.

The concern about NO_x emissions and the need to comply with increasingly stringent regulations have motivated vast research. Restrictions that used to apply only to large combustion systems such as electric power plants are now a concern on ever-smaller scales and continue to be a driving force for research and development in many industrial high-temperature processes. Further, a wide range of solid fuels can be expected to contribute to the energy supply in the future. In addition to coal, biomass fuels and household and industrial waste become more important. The use of novel fuels and fuel mixtures introduces new challenges in pollution control.

Emission of CO_2

There is general consensus that the human combustion of fossil fuels is significantly responsible for the rise in the CO_2 concentration in the atmosphere that has led to global warming. The International Panel on Climate Change (IPCC) has recently concluded that avoiding an increase in temperature of more than 2°C, which could have dramatic adverse effects, requires keeping the CO_2 concentration in the atmosphere below 450 parts per million (ppm) by volume. This is a major challenge, since the CO_2 concentration is about 390 ppm by volume and is rising by about 2 ppm by volume annually.

CO_2 emission from heat, power and industrial production can be reduced by several measures. Rapid conversion from fossil fuels to renewable energy is attractive in fulfilling the goal of stabilizing the CO_2 concentration. However, fossil fuels account for about 80% of the world's primary energy demand, with coal comprising about 25%. Geothermal, solar and wind energy meet only 0.4% of the global energy demand and biomass and waste 11%. Further, China and India are expanding their use of coal very rapidly. China is currently constructing the equivalent of two 500-MW coal-fired power plants per week. This indicates that the use of coal will expand in the future because it is plentiful (enough to last several hundred years at the current rate of use), inexpensive and geographically spread globally.

NH$_3$/urea
Burnout air
1000°C   800°C

Selective
catalytic
reduction

Reburn fuel   1200°C

370°C

Low-NO$_x$ burner

370°C

Air preheater
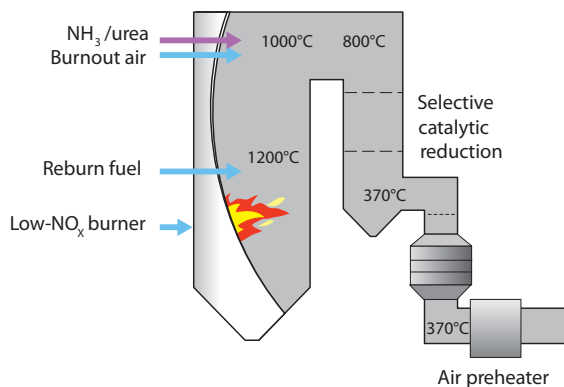
Fig. 6.6. Significant expertise has been achieved in reducing nitrogen oxide emissions from pulverized-coal fired boilers, either by primary means (in particular low-NO$_x$ burners or over-fire air injection), by injection of reducing agents (such as ammonia (NH$_3$) or reburn fuels (natural gas, biomass or coal) or by catalytic flue gas cleaning with ammonia (selective catalytic reduction). Some of these methods may not be viable for fuels such as biomass and waste. These fuels are commonly burned on a grate, a technology that precludes the use of low-NO$_x$ burners. Further, catalytic flue gas cleaning is often difficult due to catalyst deactivation caused by trace components such as alkali metals in the flue gas from biomass or waste – a problem CHEC is also working on solving.
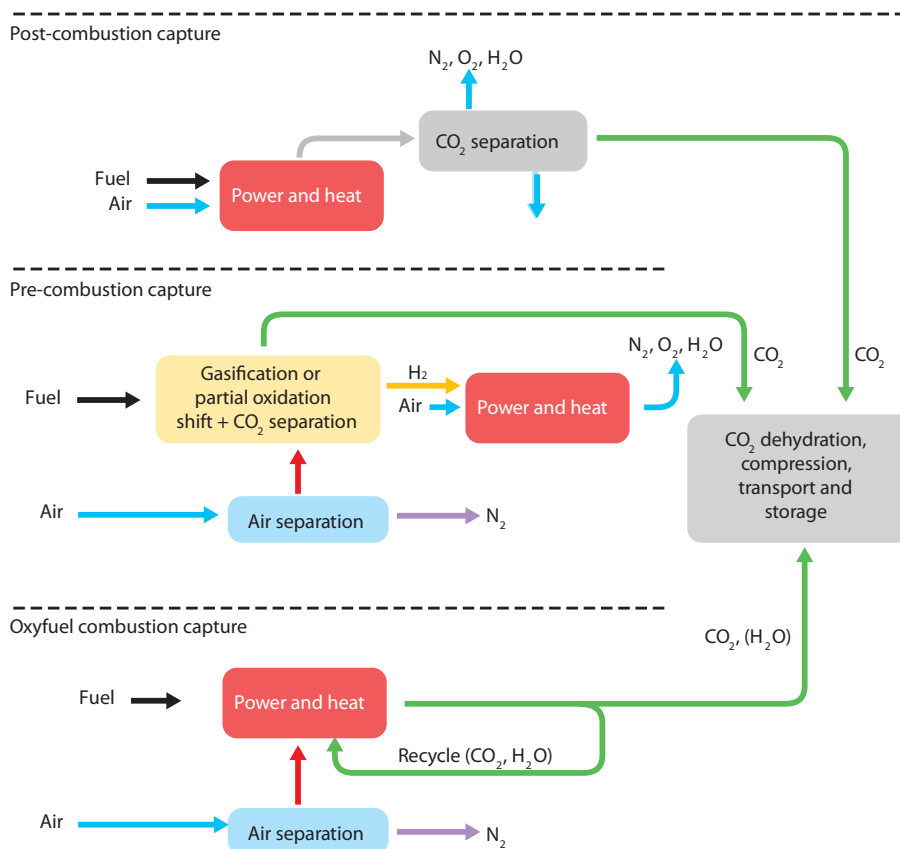
Fig. 6.7. Possible plant configurations for the three main categories of carbon capture technology: heat, power and industrial production

The continued use of fossil fuels implies that carbon capture and sequestration is necessary to reduce the emissions of CO_2. Introducing this worldwide will be a formidable task: about 1.5 million tons of CO_2 per year needs to be captured and stored for a 500-MW coal-fired plant. The global emissions from coal combustion in 2000 were about 10 billion tons: about 10 km^3 of liquid CO_2. The goal of the European Commission is almost zero emission of CO_2 from coal-fired power plants from 2020, requiring that many large demonstration plants be built and tested on a short time scale.

Fig. 6.7 outlines the three most frequently suggested processes for coal combustion with carbon capture and sequestration. In all schemes, the final product is a stream of almost pure CO_2 that can be compressed, liquefied and transported to an underground storage site. In the post-combustion

capture of CO_2, the power plant is basically identical to the plant shown in Fig. 6.2, with an additional process added before the chimney to remove CO_2 from the flue gas. Processes currently being investigated involve scrubbing the flue gas with alkanolamine solvents such as monoethanolamine or a cold aqueous solution of ammonia. An advantage of this process is that it can be used to retrofit existing power plants for carbon capture and sequestration. DONG Energy, Vattenfall and others are testing this technology at a pilot plant at Esbjergværket, and Vattenfall is currently planning to demonstrate this technology in full scale at unit 3 at Nordjyllandsværket (a 900-MW pulverized coal–fired power plant). The captured CO_2 is intended to be stored underground at a site near the power plant.

Pre-combustion capture involves gasifying the

fuel to a gas consisting mainly of CO, CO_2, H_2 and H_2O, which is converted to form CO_2 and H_2 in a catalytic reactor and finally CO_2 is removed from the flue gas as in the post-combustion capture process. The stream of pure hydrogen (H_2) is burned in a gas turbine. The pre-combustion capture process applies pure oxygen in the pressurized (30–50 bar) gasification process and thus requires an air separation plant upstream of the gasifier. Despite the high plant efficiency promised by this technology, few electricity-producing units exist today (without the carbon capture and sequestration capability). The availability of these plants is relatively low because of the short operating experience compared with the mature, conventional pulverized coal–fired power plants. This technology is not an option for retrofitting existing pulverized fuel plants and requires developing a hydrogen gas turbine.

Oxyfuel combustion involves combusting the fuel in pure oxygen obtained from an air separation plant. Flue gas is recycled to the combustion chamber to control the flame temperature. This combustion process leads to a flue gas consisting mostly of CO_2 and water. Water is easily removed by condensation, leaving an almost pure stream of CO_2 ready for sequestration. Compared with the post-combustion capture process, implementing the oxyfuel combustion technology in existing pulverized coal–fired power plants will require greater changes to the plant configuration. This is mainly because substituting recycled flue gas (mainly CO_2 and water) for nitrogen in the oxidizer alters the combustion chemistry and physics.

With the present technology, these processes are very costly. Estimates indicate that each will result in an efficiency loss of 8–12 percentage points, corresponding to an electrical efficiency loss of 25–30% at the power plant. None of the processes has yet been demonstrated in full scale and will require significant research and development efforts before a winner can be found.

Power plants in Denmark have been testing biomass as fuel in co-combustion with coal since the mid-1990s, and plants have been in commercial operation since about 2000. With Denmark's highly efficient power plants, with total efficiency of about 90%, this is probably the most efficient way of using biomass for reducing CO_2 emissions. Combined with carbon capture and sequestration, this is a particularly attractive option where the net CO_2 emission may even be negative, with CO_2 being drawn out of the atmosphere.

The authors

Anker Degn Jensen: MSc and PhD, DTU, 1996. Postdoctoral Fellow, Advanced Fuel Research Inc., USA, 1996/1997. DTU Department of Chemical and Biochemical Engineering: Assistant Professor, 1997; Associate Professor, 2000; Professor, 2007–. Research: catalysis, particle technology and combustion and gasification processes.

Peter Glarborg: MSc (1984) and PhD (1987), DTU. Professor, Department of Chemical and Biochemical Engineering, DTU Clean Power Chair. Associate Editor, *Combustion and Flame*. Chair, Scandinavian-Nordic Section and Member, Board of Directors Combustion Institute. Research: combustion and harmful emission control, high-temperature chemistry, chemical kinetics.

Present and future challenges

Fossil fuel resources provide 85% of Denmark's energy consumption, with more than 90% based on thermal processes. The Government of Denmark has recently stated its vision to lower CO_2 emissions and to make Denmark independent of fossil resources. Today, the only way to fulfill this vision is to increase strongly the generation of wind energy and the generation of power and heat from biomass and waste while simultaneously removing CO_2 from the flue gas from fossil fuel–fired power plants.

In the years to come, the energy sector will still be based mostly on thermal processes and depend significantly on fossil fuels, especially coal. Significantly reducing CO_2 emissions will require developing technology for increased fuel flexibility, thermal efficiency, carbon capture and sequestration, control of harmful emissions and integrated energy systems to optimally use available resources.

The share of biomass and waste in power and heat generation needs to be increased significantly, and power efficiency needs to be increased when using alternative fuels. Using alternative fuels such as biomass and waste is an efficient way to reduce CO_2 emissions. Biomass and waste are currently used mainly on decentralized, grate-fired units. However, the biomass share for co-combustion with fossil fuels in the centralized, pulverized-fuel power plants is increasing, and co-combustion of industrial waste in these units is being tested. The ambition is to replace coal as the main fuel with biomass, using coal only as back-up fuel.

Compared with fossil fuels, biomass and waste are difficult to handle in terms of pretreatment, combustion and solid residue. In particular, the content of inorganic species (such as sulfur, chlorine, potassium and sodium) is a concern due to enhanced propensity for deposition and corrosion. Currently, this severely constrains the electrical efficiency of the decentralized, grate-fired units, which seldom exceeds 26–27%, and the share of alternative fuels in the centralized pulverized-fuel power plants, which seldom exceeds 20%. Research and technology development must ensure enhanced conversion efficiency on the grate-fired units, based on novel furnace designs and/or use of additives or other means. For the pulverized-fuel power plants, the aim will be to develop a 100% fuel-flexible unit, which will involve pretreating alternative fuels, developing novel burner designs and developing methods to improve ash quality and control emissions.

No commercial technology is available yet for capturing CO_2 on existing power plants, and implementation within a short time frame will require considerable technology development. Capturing and storing CO_2 from power plants is estimated to double the cost of generating electricity. Activities are desirable on oxyfuel combustion, CO_2-capture technologies and technologies based on gasification (integrated gasification combined cycle). The emphasis on each technology should be reviewed regularly, however, to ensure that the activities support present carbon capture and sequestration strategies in Denmark's industry.

More to explore

Publications of the CHEC Research Center: www.chec.kt.dtu.dk/Publications.

Jaccard M. *Sustainable fossil fuels*. Cambridge, Cambridge University Press, 2005.

Pachauri RK, Reisinger A, eds. *Climate change 2007: synthesis report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva, Intergovernmental Panel on Climate Change, 2007.

The future of coal: options for a carbon-constrained world. An interdisciplinary MIT study. Cambridge, MA, Massachusetts Institute of Technology, 2007.
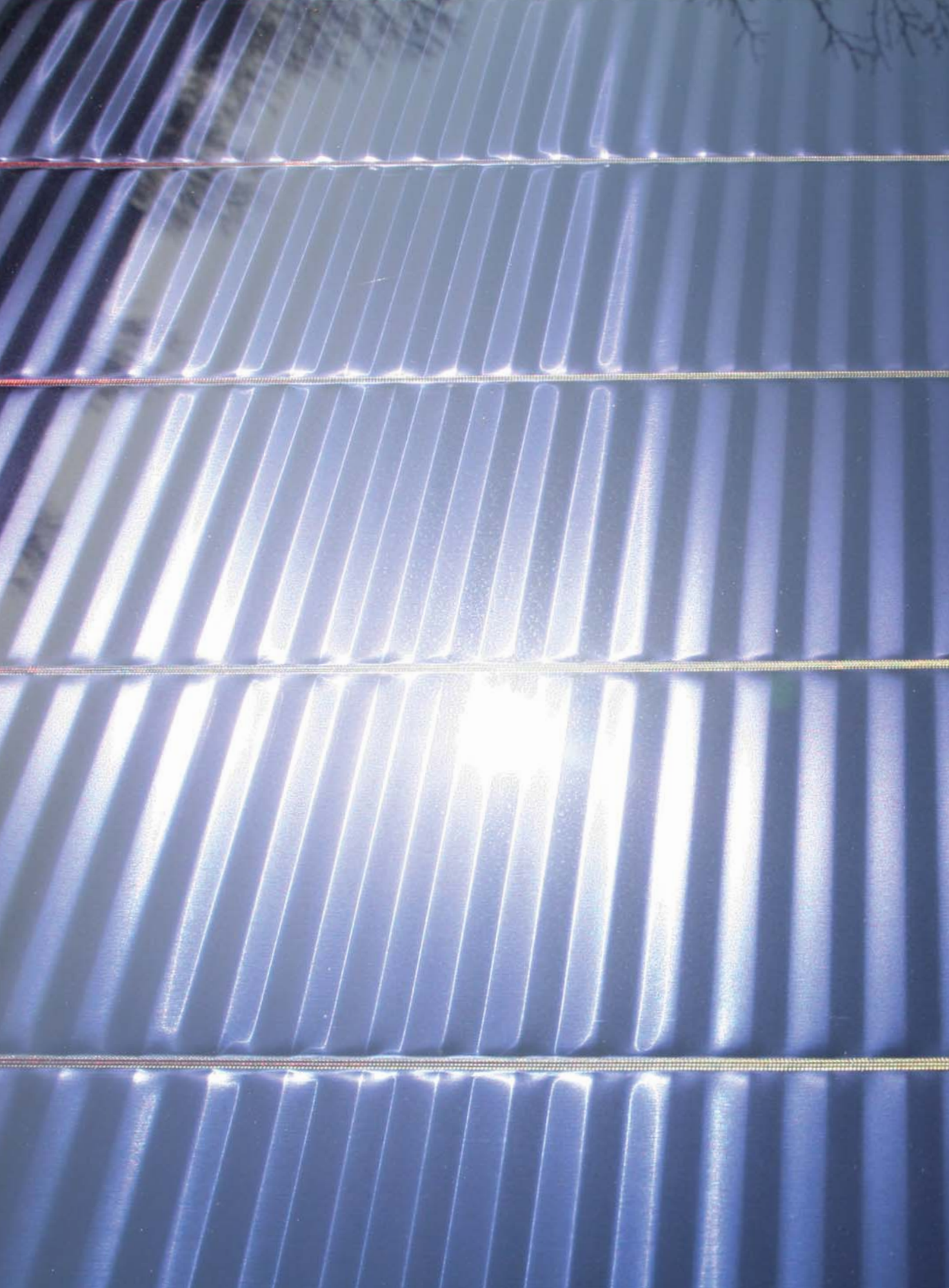
Buildings:

consuming and conserving energy

Bjarne W. Olesen & Carsten Rode

Introduction

Buildings nowadays must be energy efficient. High-income countries use 40% of total energy to provide healthy and comfortable environments in buildings. Thus, since energy resources are scarce, energy optimization of buildings and the related heating, ventilation and cooling systems are required.

Reducing energy consumption is one goal, and another is achieving sustainable sources of energy such as renewable energy sources with low or no emissions of greenhouse gases (mainly CO_2).

Buildings must provide a healthy and comfortable indoor environment that enhances the performance of the occupants. The indoor environment should not be too cold or too warm, should have no unacceptable odour, should have an acceptable relative humidity and should be free of harmful pollutants. Buildings should have good daylight conditions without blinding and no disturbing noise from systems, neighbors or the outside. In a normal office, the energy used costs only 1% of the cost of human labor (including compensation, insurance and other costs). A slight decrease in performance or increase in sick leave therefore easily exceeds the whole energy budget. Thus, maintaining indoor environmental quality when reducing the energy demand of a building is of utmost importance.

Is there a potential conflict between an acceptable indoor environment and energy conservation? Ensuring a good indoor climate requires ensuring that buildings be sufficiently well ventilated with fresh air and have sufficient heating and cooling to keep the room temperature comfortable. Nevertheless, enhancing the ventilation rates, heating and cooling also increases energy consumption.

Energy performance of buildings

Buildings are made more airtight to increase energy efficiency, and exterior walls and windows are highly thermally insulated. The challenge is how to make the buildings more energy efficient without compromising the indoor environment.

Energy-efficient buildings can be built that have a healthy indoor environment and are attractive, function well, can be maintained and are durable.

These functional requirements must be fulfilled simultaneously. This may not be as easy as it used to be – or at least new methods are needed to design and realize such buildings.

Highly energy-efficient new buildings can be built today that cost only about 10% more than conventional buildings. However, the price is a relative measure. Many components in advanced buildings may be rather expensive today, but as soon demand for these components is sufficient, the price will also come close to the price of the traditional products they replace. This has been the case with windows, which insulate much better today than they did 15 years ago.

However, such buildings must be very finely tuned to optimize the indoor climate, durability of the building components and energy efficiency. Although some demonstration buildings have been constructed, it still needs to be determined how such finely tuned buildings can become the standard for all people, companies and institutions with their different needs. Further, retrofitting existing buildings to similar good standards should also be realistic and economical, and this may be a significant challenge.

Design of buildings

The answer is integrated design. Today buildings cannot be optimized for one parameter only. All components and functions of a building have to be designed in an integrated way such that the design can be optimized for all design issues – even when they may initially seem to contradict each other.

Integrated design can only be realized when engineers work closely with other designers of buildings:

- the architects who produce fantastic designs that show how a building should function and how it should look;
- the construction managers, who make plans that cleverly translate the architectural and engineering design into realizable projects, including the costs; and
- the contractors, who know how it should be built in practice.

Engineers are important contributors to the integrated design process. The engineers translate the laws of the physics into what can be realized in

terms of important functional parameters, such as load-bearing capacity and deflection, temperature and energy flows and indoor environment conditions.

Building physics is the engineering discipline that focuses on many of the physical parameters that constitute the interface between buildings, the heating, cooling and ventilation systems installed and the building occupants. Building physics focuses on:

- the thermal environment within and around buildings;
- moisture conditions in these environments;
- the condition of air and its movement;
- the atmospheric condition;
- light; and
- acoustics.

These parameters are mutually dependent: for instance, the humidity depends highly on the temperature. And the condition in one part of a building depend on the conditions in other parts; thus, indoor air conditions depend on the heat and

moisture conditions in the surrounding walls. Buildings can be regarded as melting pots of physical phenomena and building components that interact with one another. The condition of the cocktail is also strongly influenced by the users of buildings, and these users are critical assessors in appraising the quality of the indoor environment.

Building physics and tools

DTU had a leading role in a recent international project in which researchers from 40 institutions from all over the world collaborated to formulate and test computational models to predict the hygrothermal state of buildings: the heat, air and moisture conditions of the indoor and outdoor environments and the structure that separates the two environments (Fig. 7.1). The project was carried out under the auspices of the International Energy Agency's research program on energy conservation in buildings and community systems. Researchers in the project represented 19 different countries.

Diffusion

Air leakage

Water vapor in indoor air
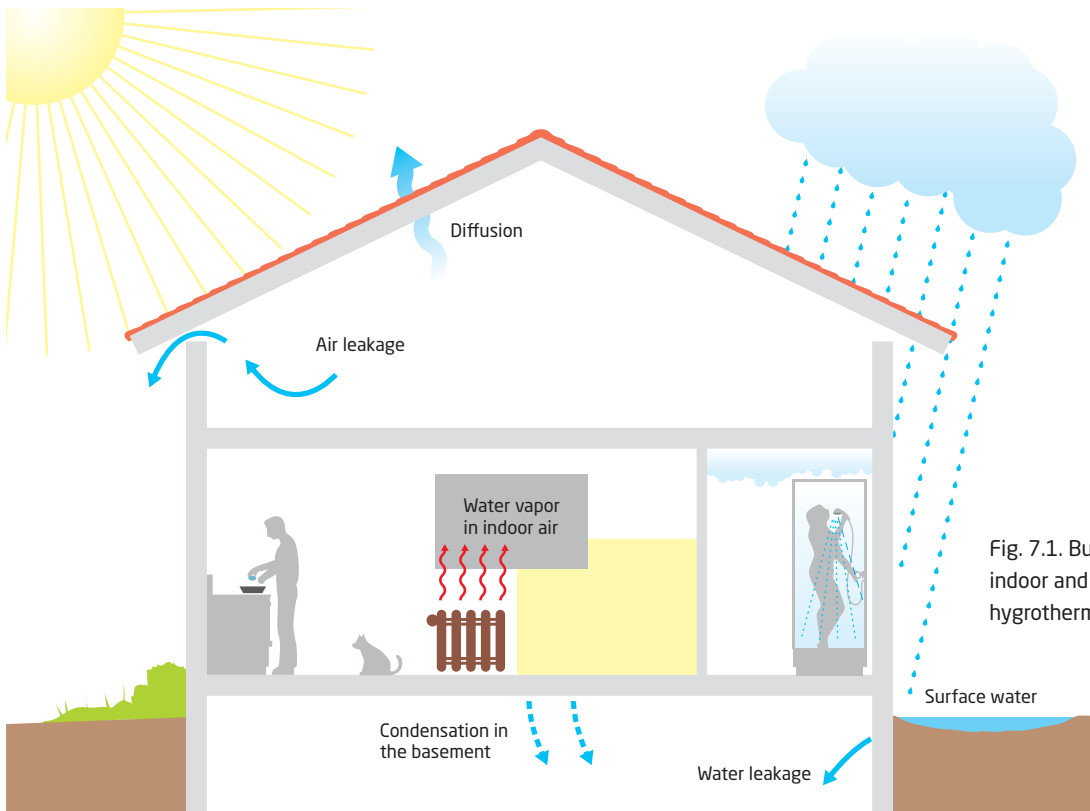
Fig. 7.1. Building with indoor and outdoor hygrothermal loads

Surface water

Condensation in the basement

Water leakage

The research combines theoretical models for individual and well-known processes to simulate the combined effects of heat and mass flow in rooms and components of buildings. Many of the processes are stochastic, and the material properties vary non-linearly with the driving potential. Back-up activities are therefore important in determining material properties and empirically investigating the heat and mass flow conditions in test rooms and real buildings.

The research involves many modelers and experimentalists, but the whole scope of the research has local angles of relevance, since building traditions and climate differ from region to region. Although the physics is theoretically always the same and many analytical paradigms can be reused, the problems are not the same from country to country. It is therefore natural to carry out the research in coordinated international projects.

A coordinated effort may involve the activity of 50 researchers who work together for 3–4 years. Participating in such work normally requires at least 3–4 researchers at each participating institution, such as DTU.

The tools developed in this project are important cornerstones for developing principles for designing buildings that are energy efficient, durable and have a good indoor environment.

For instance, one topic investigated was analyzing demand-controlled ventilation. Buildings should be ventilated sufficiently when occupied, but the ventilation rate may be reduced at other times. This may reduce energy consumption by 15% without affecting the quality of indoor air.

Building physics also analyzes the conditions ensuring that processes that can harm the indoor environment do not take place. Building physics analysis enables buildings to be designed such that the thermal conditions are comfortable and avoids conditions risking mold growth. Models such as those emerging from this project may help to ensure this.

Energy renovation of buildings

Building physics is important in developing new buildings and in energy renovation of the existing building stock. The new buildings constructed every year correspond to about 1% of the existing number of buildings. Buildings from around 1980 or older are considerably less energy efficient than new ones. For this reason, buildings that are considered old today will probably be the main contributors to the building sector's energy consumption and greenhouse-gas emissions for at least the next 50–100 years (Fig. 7.2). Finding new ways to renovate the existing buildings will therefore be very interesting. Several projects at the DTU Department of Civil Engineering focus on energy renovation of existing buildings: how to do this in a way that is economically attractive, saves significant energy and enhances the indoor environment. This is the only way to substantially reduce energy consumption and greenhouse-gas emissions from the building sector within our lifetimes given the slow pace of renewal of the building stock.

Fig. 7.2. Old buildings will probably be the main contributors to the building sector's energy consumption and greenhouse-gas emissions for at least the next 50-100 years
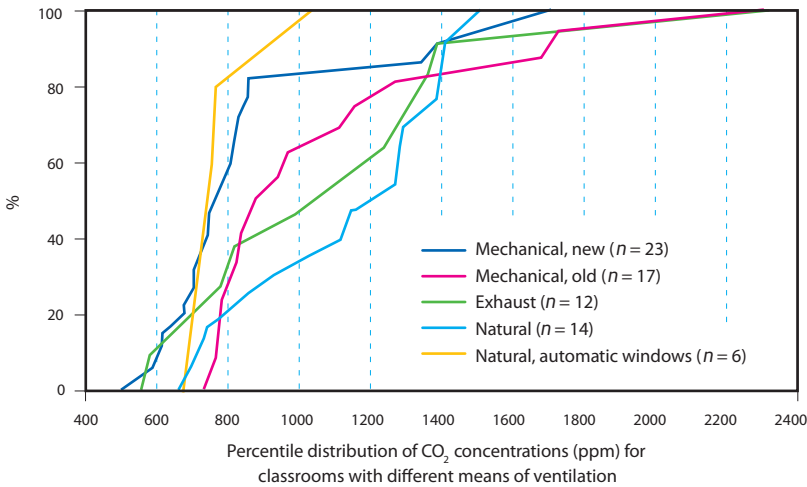
Fig. 7.3. Studies by the DTU Department of Civil Engineering show that high concentrations of CO_2 in schools indicate insufficient ventilation. About 50% of the classrooms with natural ventilation had CO_2 concentrations exceeding 1200 ppm. The numbers (n) indicate the number of classrooms measured. The steeper the curve, the higher the ventilation rate. Thus, natural ventilation with automatic windows produced CO_2 concentrations of less than 800 ppm in 80% of the classrooms versus 20% of classrooms with natural ventilation without automatic windows.

The indoor environment may be acceptable in old buildings, but this may result from a relatively air-permeable building envelope. If energy renovation is not carried out correctly, then indoor climate problems may arise, such as in many buildings renovated for energy conservation after the energy crises in the 1970s.

Today considerable money is being invested in developing new energy sources. Investing in reducing demand is a win-win situation, however. It saves energy, reduces costs and makes using renewable energy sources easier. One goal of the research at the DTU Department of Civil Engineering is to develop concepts and components so that, by 2025, we can make net zero-energy buildings or even plus-energy buildings: buildings that produce energy and can be used for charging electric cars, which are thus independent of fossil fuels.

One trend is heat recovery by natural ventilation. This requires heat recovery units with very small pressure drops and little use of auxiliary energy by pumps and/or fans.

Indoor environment

Building physics is necessary, but buildings also need to be heated, cooled and ventilated more efficiently. People have high demands for comfort (such as air-conditioned cars), and global warming will increase the demand for cooling. Further, several field studies from the DTU Department of Civil Engineering show that many homes, kindergartens, schools and offices are underventilated and that temperatures in summer are often unacceptably high (Fig. 7.3).

The International Centre for Indoor Environment and Energy at the DTU Department of Civil Engineering has performed extensive field studies examining a possible relationship between asthma and allergy and the indoor environment. In large studies in Sweden and in Bulgaria, the Centre found that many homes are underventilated and have unacceptable concentrations of phthalates. For both factors, a lower ventilation rate and higher concentrations of phthalates increased the odds ratio for asthma and allergy cases among children. A similar study is now being done in Denmark. Here the researchers are also studying the indoor

environment in kindergartens. Many of the kindergartens both with and without mechanical ventilation are underventilated, similar to schools.

Evaluating the quality of the indoor environment is often very difficult. There are numerous factors, such as temperature, humidity, emission of pollutants from materials, bioeffluents from people, lighting conditions and noise. Detailed evaluation based on measurements therefore requires many resources, expensive instruments and analysis. The most common measurements are room temperature and humidity combined with measuring the CO_2 concentration to indicate the ventilation rate. Nevertheless, detailed analysis often requires taking air or dust samples and expensive analysis of the samples. Measuring the subjective reaction of people exposed to the environment being studied is also very important. Studying the human reactions has been the core of the research conducted at the International Centre for Indoor Environment and Energy. Both in field studies and in laboratory studies this requires data from many samples (people and buildings) to statistically evaluate the results. It also requires multidisciplinary research involving psychology and medical expertise. Engineers may take measurements and record people's subjective reactions. Medical experts are required, however, to explain why people react as they do, such as what happens inside the human body and how the body absorbs air pollutants (through the skin, orally or through the lungs).

Energy performance of systems

How can buildings be heated and cooled more efficiently?

Approaches for reducing energy consumption to maintain comfortable temperatures in buildings include using water-based radiant cooling systems, which may use the building's thermal mass in combination with night cooling or heating, or cooling by pipes embedded in floors, walls or ceilings (slab heating or cooling).

The energy advantages of such systems are the high water temperature for cooling and low water temperature for heating (low-exergy systems). Low-temperature heating means that the temperature of the water used to transfer heat is slightly higher than room temperature. Similarly, high-

temperature cooling means that the temperature of the water used for cooling is slightly lower than room temperature This increases the efficiency of cooling systems (chillers and heat pumps) and the potential for using renewable energy sources such as ground heat exchangers, evaporative cooling and adsorption heat pumps. Further, using the storage capacity of the building can reduce the peak loads, and much of the energy use can be transferred to nighttime.

The Centre has been involved in several building projects using this technology, such as Bangkok International Suvarnabhumi Airport (Fig. 7.4) and the Copenhagen Opera House. Through a joint program of the European Commission and the Association of Southeast Asia Nations (ASEAN), the Centre was involved in a demonstration project in Malaysia in which the concept will be used in a net zero-energy office building. The Centre is also involved in a demonstration project in Denmark.

How can buildings be ventilated more efficiently?

Concepts such as reducing internal pollution loads (low-emitting building materials and furnishings), demand-controlled ventilation (only ventilating when needed), air cleaning and more efficient distribution of the ventilation air in the building (complete mixing, displacement or personalized ventilation) will all reduce energy demand and may enhance the indoor environment.

The DTU Department of Civil Engineering is studying all these concepts. The concept of personalized ventilation has especially been a significant part of the research during the past decade. This research requires combining many skills. First, it requires knowledge of air flow and such tools as computational fluid dynamics to optimize the air outlets. Normally an air outlet would mix the air very rapidly with room air, but for personal ventilation, an air jet is desired without introducing room air in the jet stream. The research also requires many studies with human subjects to be able to evaluate the acceptance of such systems. People differ vastly, so data from many people are needed. The aim is to generate personalized flow with two zones: the first zone of clean personalized air covering the area in front of the PC monitor a person

Fig. 7.4. Bangkok International Suvarnabhumi Airport. The International Centre for Indoor Environment and Energy at the DTU Department of Civil Engineering has participated in a demonstration project here.
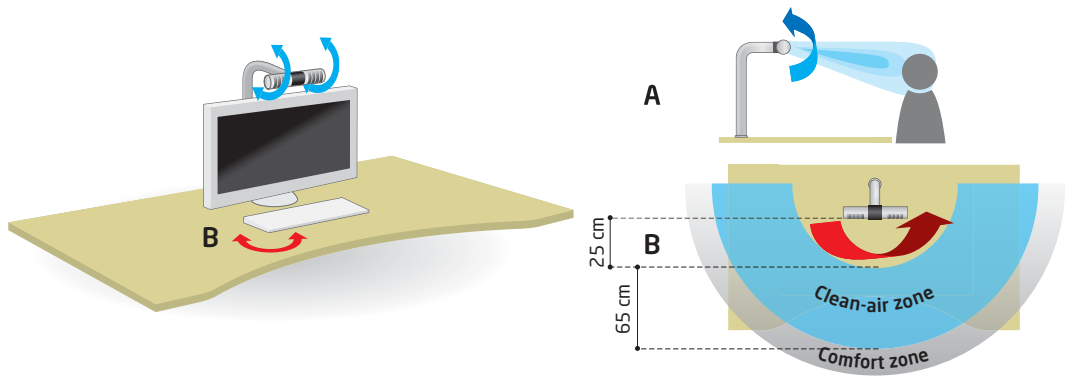
Fig. 7.5. The first prototype of the personalized ventilation system has been manufactured and been tested in a full-scale air movement room

occupies most of the time and the second zone, the comfort zone, covering a much larger area at the desk within which the person can inhale air cleaner than the background air in the room and can substantially improve his or her thermal sensation by convection cooling of the body (Fig. 7.5). The first prototype of the personalized ventilation has been manufactured and been tested in a full-scale air movement room using a breathing thermal manikin resembling a person combined with many tests with real subjects.

The personalized ventilation system has now been commercialized and installed in a couple of buildings. It is being installed in several other buildings. A field survey in one building located in downtown Copenhagen was initiated to identify occupants' response to the personalized ventilation. This shows one example of transferring many years of research to industry. It also shows a way of improving the indoor environment, providing the occupants with some personal control and reducing energy consumption.

More to explore

Woloszyn M, Rode C. *Whole building heat, air and moisture response.* Volume 1. *Modelling principles and common exercises.* Paris, Executive Committee on Energy Conservation in Buildings and Community Systems, International Energy Agency, 2008.

Olesen BW, Blutssen P, Roulet C-A. Ventilation and indoor environmental quality. In: Awbi H, ed. *Ventilation systems: design and performance.* Milton Park, Taylor & Francis, 2008.

Proceedings of the 11th International Conference on Indoor Air Quality and Climate: Indoor Air 2008, Copenhagen, Denmark, August 17–22, 2008. Santa Cruz, CA, International Society of Indoor Air Quality and Climate, 2008.

Babiak J, Olesen BW, Petrás D. *Low temperature heating and high temperature cooling.* Brussels, Federation of European HVAC Associations (REHVA), 2007 (REHVA Guidebook 7).

The authors

Bjarne W. Olesen:

MSc, Civil Engineering, 1972. PhD, Laboratory of Heating and Air Conditioning, DTU, 1975. Research scientist, Laboratory of Heating and Air Conditioning, DTU, 1972–1990. Part-time Product Manager, Brüel & Kjær, 1978–1992. Senior Research Scientist, College of Architecture, Virginia Tech, 1992–1993. Head, Research & Development, UPONOR-VELTA GmbH KG & Co., Norderstedt, Germany, 1993–2004. Professor of Indoor Environment and Energy and Director, International Centre for Indoor Environment and Energy, DTU, 2004–. Awarded the Ralph Nevins Award (1982), Distinguished Service Award (1997), Fellow Award (2001) and Exceptional Service Award (2006) from the American Society of Heating, Refrigerating and Air-Conditioning Engineers, where he is on the Board of Directors. Medal of Honor, German Engineering Society (VDI-TGA), 2005. International Honorary Member, Society of Heating, Air-Conditioning and Sanitary Engineers of Japan. Chair, CEN TC228 Heating Systems in Buildings. Convener, ISO TC159/SC5/WG1 Ergonomics of the Thermal Environment and CEN TC156/WG12, Ventilation of Buildings – Indoor Environment. Active in several other ASHRAE-CEN-ISO-DIN standards committees on indoor environment and energy performance of heating, ventilation and air-conditioning systems.

Carsten Rode:

MSc in Civil Engineering, DTU, 1987, PhD, DTU, 1990. Research Visit, Roof Research Center, Oak Ridge National Laboratory, 1989. Research assistant, Thermal Insulation Laboratory, DTU, 1990. Consultant on market preparation of the MATCH-program – Moisture and Temperature Calculations for Constructions of Hygroscopic Materials, 1990. Research Fellow, Thermal Insulation Laboratory, DTU, 1990–1992. Senior Researcher, Danish Building Research Institute, 1992–1996. Associate Professor, Department of Civil Engineering, DTU, 1996–. Head, Section for Building Physics and Services, Department of Civil Engineering, DTU, 2008–. Chair, International Society for Building Physics, 2009–. Expertise in building physics: combined heat and moisture transfer processes in structures, indoor climate, hygrothermal properties of building materials and numerical modeling.

Biorefineries:
converting biomass into valuable products

Irini Angelidaki & Kim Pilegaard

Biotechnology was the first technology to produce chemicals, fuels and other industrial products. Goods were produced on a large scale based on biomass as early as the 19th century. Early industrial products from biomass include pulp and paper production from wood, nitration of cellulose from guncotton and viscose silk, soluble cellulose for fibers, production of furfural for nylon and production of acetate, lactic acid, citric acid and ethanol. Later, fossil fuels set the progress of biomass-based production of industrial products on standby.

There is general concern about the current fossil-fuel system, which is largely based on finite resources that are not sustainable. In addition, the political volatility of many oil-producing countries and the rapid fluctuation of the fuel market are encouraging governments to plan long term to decouple from dependence on fossil fuels. Concern about the instability of fossil-fuel supply, limits on fossil-fuel reserves and especially environmental concerns have brought new focus on white biotechnology. White (or green) biotechnology uses biomass as feedstock instead of fossil fuels for production by biological conversion processes, bio-based fuels, chemicals, solvents etc. Using biomass as a raw material instead of fossil fuel has the advantage of working in a closed sustainable carbon cycle, in contrast to the open cycle of using fossil fuel with net release of greenhouse gases to the atmosphere (Fig. 8.1).

Numerous compounds can be produced from biomass, although only a few can be produced economically compared with present fossil fuel–based technology. Besides the interest in new chemicals, a strong interest in producing biofuels and bioenergy has brought biotechnology into focus.

Biomass

Globally, biomass resources are mainly from wood and agricultural products and waste. Agricultural residue mainly comprises lignocellulosic biomass. Lignocellulose is the term for the structural parts of plants. It consists of cellulose, hemicellulose and lignin. Cellulose is an organic polysaccharide of glucose and can be broken down by enzymes (cellulases and glucosidases), although the process is slow. Hemicellulose is a heteropolysaccharide containing mainly C-5 sugars such as xylose and arabinose and the C-6 sugar mannose. The composition of hemicellulose varies between plant species. Lignin is a term for amorphous, three-dimensional polymers that have a phenylpropane structure. Lignin is very resistant to degradation and can be used for combustion if it can be separated in dried form since it has a high heat value.

Many biofuels are derived from sugar cane, corn and wheat. Cereal straw represents the largest biomass resource from agriculture in Denmark (5.2 million tons in 2006): 26% is directly burned for household heating and in power plants, 19% for fodder and 12% for bedding; the remainder (43%) is plowed in. Other major resources in Denmark
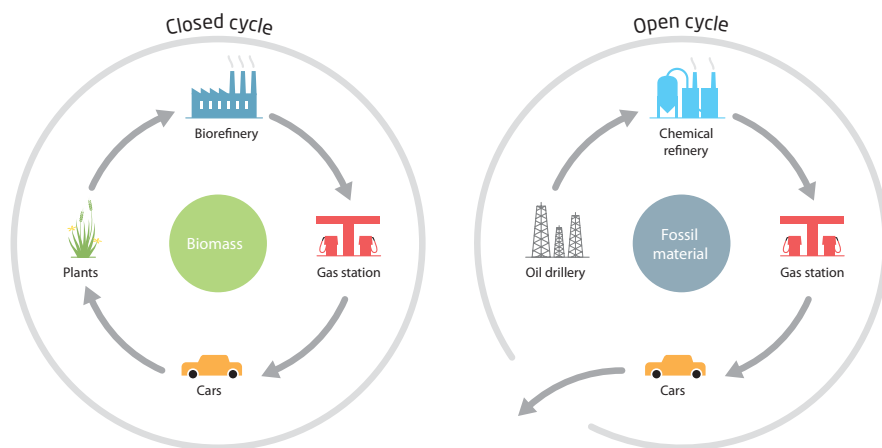
Fig. 8.1. The closed carbon cycle using biomass as raw material for fuels, chemicals and energy in contrast to using fossil fuel as raw material and releasing net carbon
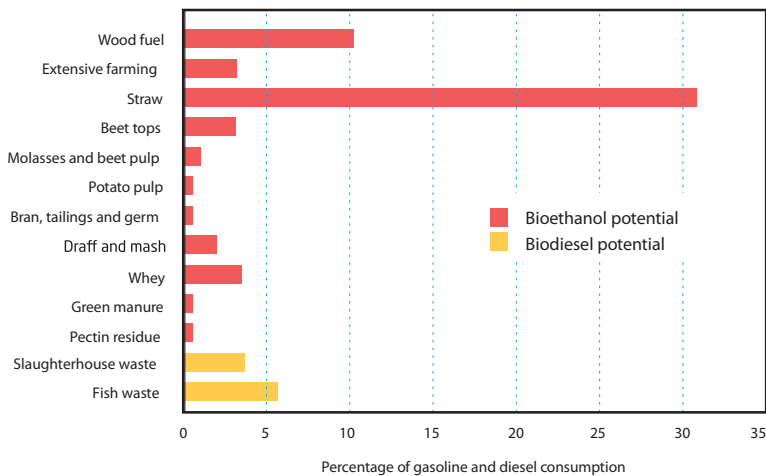
Fig. 8.2. Potential for converting biomass to transport fuels in Denmark. Both C-6 and C-5 sugars are presumed to be converted to bioethanol (source: Blume et al., 2008).

include wood, manure and organic waste from industry and households. In addition, extensive grasslands and dedicated energy crops could be considered if they can serve other additional ecosystem purposes such as landscape management, groundwater protection, biodiversity and carbon sequestration and preventing nutrient leaching. A recent report shows that converting straw from Denmark's agricultural sector into bioethanol can cover up to 30% of the existing fossil fuel consumed in transport (Fig. 8.2). Wood resources can contribute 10% and various types of waste biomass an additional 6%. Fish and slaughterhouse waste can cover up to 9% of current diesel consumption. However, if such extensive resources are used, this might have other effects on agricultural ecosystems, such as reducing soil carbon stocks and soil fertility and influencing biodiversity.

Biorefineries

A biorefinery integrates biomass conversion processes to produce fuels, electrical power and chemicals from biomass. By producing multiple products, a biorefinery can take advantage of the differences in biomass components and intermediates and maximize the value derived from the biomass feedstock according to the market situation and biomass availability. The bulk of the products are biofuels and bioenergy, and chemicals are produced in smaller amounts. Although the amounts of other products are small, they often have much higher specific value.

Biorefinery systems and design

Biomass is complex. Plant biomass consists of the basic products carbohydrate, lignin, protein and fat and a variety of substances such as vitamins, dyes, flavors and aromatic compounds.

Many biorefinery concepts have emerged in recent years based on different feedstocks or/and different processes and products.

Four main types of biorefinery systems have been defined recently:
- lignocellulosic biorefineries, based on wood and straw;
- whole-crop biorefineries, based on such raw materials as grain and maize (whole crop);
- green biorefineries, based on grasses;
- two-platform biorefineries, with sugar and syngas (synthesis gas) platforms; and
- oily-crop biorefineries based on whole-crop utilization of oily crops.

Many of the proposed biorefinery systems focus on producing fuels for transport. However, new ideas are emerging continually, such as biorefineries based on cultivating algae, especially after the ethical quandaries of using agricultural soil for producing biofuels have emerged.

Achieving a high degree of advanced processing is theoretically possible technically. Technical, socioeconomic, political and environmental interaction plays an important role in developing biorefineries (Fig. 8.3).

Economics is often the most important factor determining the application of the technology. However, political decisions and priorities can
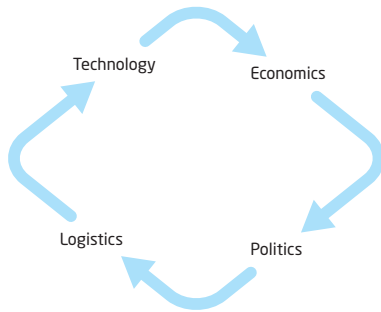
Fig. 8.3

often motivate development, which results in technologies that are more advanced and price reductions. Logistical issues and infrastructure are also important factors and depend on medium- to long-term political strategic planning.

Biorefineries based on oily crops

Significant investment has been made in the biodiesel sector in recent years. The European Union has become the world leader in biodiesel production, and demand for biodiesel fuel for cars is increasing. However, producing biodiesel competitively and sustainably is difficult. The Rapeseed Biorefinery (a project coordinated by DTU with the participation of the University of Southern Denmark, Faculty of Life Sciences of the University of Copenhagen, Aarhus University, Novozymes A/S and Emmelev Mølle A/S) will utilize the whole crop of rapeseed biomass (in contrast to the seeds only, which is the practice today) by combining seed and straw processing. Besides food and animal feed, the rapeseed biorefinery can produce a multitude of biofuels, bioenergy, fertilizers and high-value chemicals (Fig. 8.4).

Electricity

Heat

High-value products
(such as chemicals)

Fertilizer

Biodiesel

Bioethanol

Biohydrogen
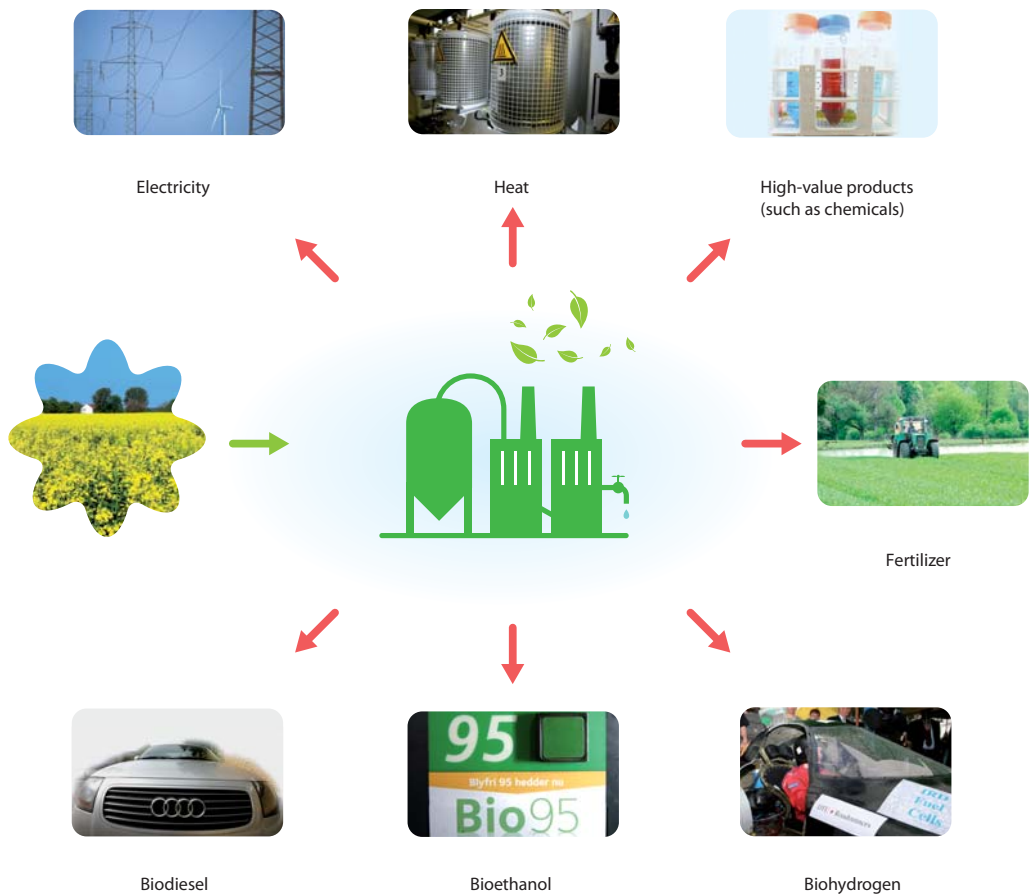
Fig. 8.4. The DTU rapeseed biorefinery

The rapeseed plant is divided into two streams: the seeds and the straw.

The seeds are treated by an innovative enzymatic process (hemicellulase), resulting in four fractions (hulls, oil, sugar and protein). Biodiesel is produced from the oil fraction by transesterification with methanol. High-value chemicals (phospholipids, tocopherols, sterols, dicarboxylic acids and epoxidized oleochemicals) are also derived. The glycerol released in biodiesel production as a byproduct can also be used for producing high-value-added products. Alternatively, extremely thermophilic microorganisms can be used to convert glycerol to ethanol or butanol.

High-value products such as antioxidants, glucosinolates, anticancer pharmaceuticals and high-quality protein rich in lysine and methionine can be recovered from the other parts (hulls, syrup and protein).

From straw, hexoses can be converted to bioethanol by yeast and pentoses can be converted to biohydrogen. Alternatively, pentoses can be converted to bioethanol by extremely thermophilic bacteria. The effluents from different processes will be treated anaerobically to stabilize them and to produce methane. Finally, the treated effluents will be used as biofertilizer.

By using the whole rapeseed crop (seeds and straw), energy production will increase from 28% of the total plant energy content (by using the seeds only) to 49% (by using the whole crop producing second-generation biofuels) along with the production of high-value-added products and biofertilizer.

Converting lignocellulosic matter to bioethanol

Diesel and gasoline constitute the main fuels used for transport. The world's main oil reserves are found in a small part of the world, mainly in Middle Eastern countries, which reduces the security of energy supply for many other countries. Domestic production of fuels, such as bioethanol, reduces dependence on oil-producing countries. In addition, oil reserves are limited and alternative renewable energy sources are therefore required eventually. Environmental awareness and the threatening climate change have resulted in extreme interest in biofuels. Finally, bioethanol is a renewable en-
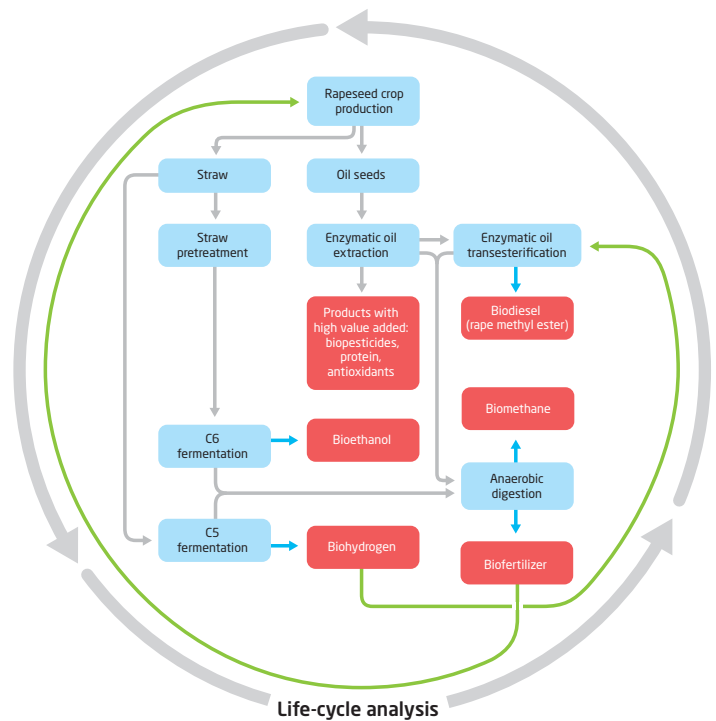
Fig. 8.5. The DTU rapeseed biorefinery concept

ergy source that can be directly implemented in the established transport systems as an additive to gasoline.

The use of ethanol for road vehicles is not new. Already in 1908, Henry Ford used ethanol to run his motor vehicle, believing it would be the fuel of the future. Ethanol later proved not to be economically competitive with fossil oil as this sector matured and more-abundant resources were identified.

Mature technologies for bioethanol production are based on using substrates such as sugar cane juice or cornstarch. These are also called first-generation technologies. Since the cost of raw materials can exceed 50% of the cost of bioethanol production, and because of the recent competition between producing food or biofuel on scarce land, recent efforts have focused on using lignocellulosic biomass.

Lignocellulosic biomass is the most abundant type of biomass on earth. Large amounts of lignocellulosic biomass are wasted today as agricultural residue, such as corn or rice stover, biofiber,

Fig. 8.6. Process steps for producing bioethanol from lignocellulosic biomass

woodchips, waste etc. However, in contrast to the established first-generation technology, technologies for lignocellulosic ethanol production (also called second-generation technologies) have not yet been fully developed. The first-generation technologies mainly include fermentation and distillation, whereas new process steps are needed for the second generation (Fig. 8.6).

Fig. 8.7. summarizes the main differences between first- and second-generation ethanol production.

Although using lignocellulosic biomass for producing biofuels has obvious advantages, the process is also facing significant challenges that need to be addressed to enable second-generation bioethanol production:

- biomass pretreatment;
- new effective enzymes are needed;
- utilization of the hemicellulose part of the sugar (mainly consisting of pentoses); and
- disposal of effluents.

Pretreatment and enzymatic hydrolysis

Since the cellulose and hemicellulose are embedded in lignin, a pretreatment step is necessary before the polymers can be broken down to simple sugar by enzymes for subsequent fermentation. Several techniques have been developed for this purpose, including acid and alkaline hydrolysis and elevated temperature and pressure. One of these techniques, wet oxidation (high temperature and pressure with added oxygen) was originally developed within the Risø DTU National Laboratory for Sustainable Energy for extracting uranium from ore from Kvanefjeldet in Greenland. This technique also turned out to be able to break down complex organic compounds and was therefore exploited for pretreating straw. The technique is, however, quite energy-intensive and thus expensive. Other techniques such as wet explosion and treatment with hydrogen peroxide are more aggressive and more expensive. In practice, a simpler hydrothermal solution seems to be more feasible,

- The substrate is
  storage polysaccharides:
  sucrose from sugar cane
  and starch from corn and wheat.

- Biomass is not pretreated before
  enzymatic hydrolysis.

- Optimized commercial enzymes
  are available.

- The substrate is
  structural polysaccharides:
  lignocellulosic material
  (straw, corn stover, wood and waste).

- Biomass needs to be pretreated to
  facilitate enzymatic hydrolysis.

- Expensive, noncommercial
  enzymes are available.

Fig. 8.7. Main differences between first- and second-generation ethanol production

such as that DONG Energy developed in the EU-funded IBUS (Integrated Biomass Utilization System) project. The biomass is generally heated to 150–200°C during the pretreatment step. Another important part of the pretreatment is to separate lignin from cellulose, because lignin cannot be converted to ethanol.

After the pretreatment, the cellulose and hemicellulose are hydrolyzed to monosaccharides by means of enzymes. Cellulose is broken down by cellulase followed by glucosidase. Breaking down hemicellulose requires another set of enzymes including xylanase. This enzymatic hydrolysis is normally carried out at a temperature of 50°C.

Utilization of hemicellulose

Sugar is released after pretreating lignocellulosic material. Two types are released: hexoses (the main constituent of cellulose) and pentoses. Hexose can effectively be converted to bioethanol, and the process is carried out with high yield and productivity by *Saccharomyces cerevisiae* or recombinant *S. cerevisiae*. *S. cerevisiae* is by far the best-known ethanol producer today but cannot convert pentose. No effective microorganisms for the industrial conversion of pentose (the main constituent in hemicellulose) to bioethanol have been found yet, although several promising recombinant candidates for pentose fermentation have been described and presented as future solutions. Meanwhile, these organisms have not yet proven their applicability on a large scale. These organisms often have relatively low productivity, low ethanol tolerance and high sensitivity to the inhibitors present in the hydroly-

sate (the liquid stream of thermal pretreatment of lignocellulose biomass) from the pretreatment step.

DTU is working in several directions to find cost-effective methods of utilizing pentose. Due to the limitations in the conversion of pentose into bioethanol, an obvious solution would be to investigate alternative methods of utilization.

Conversion of pentose to ethanol

An ideal microorganism to be used as an industrial ethanol producer for second-generation ethanol production should fulfill several requirements, such as:

- fermenting essentially all the carbohydrate present in lignocellulose;
- ethanol tolerant; and
- substrate tolerant.

Many fermentative extremely thermophilic microorganisms have the capacity to produce ethanol from pentose and hexose. DTU has screened in hot springs, anaerobic digesters, sediments and other places. Several candidates have been enriched or isolated from the screening.

One very promising microorganism has been isolated from an extremely thermophilic process (70°C) operated as a continuously mixed reactor with household waste at a retention time of 1–2 days (Fig. 8.8). This organism can be directed to produce ethanol with a high yield (>70% at low pH, about 5); at higher pH, it produces more hydrogen. This possibility for manipulating the metabolic pathway of the microorganism enables products to be altered according to the market situation and the demand for specific products.
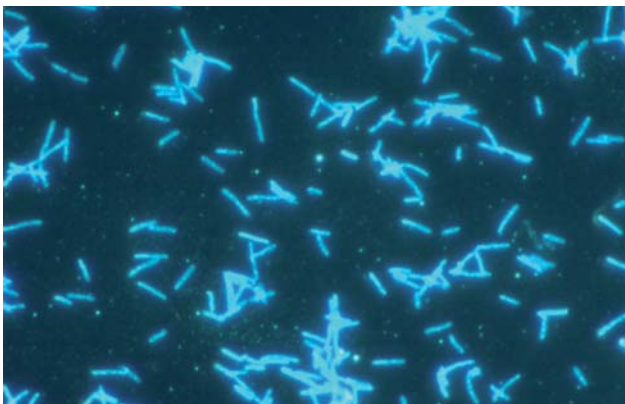
Fig. 8.8. Newly isolated organism that can convert xylose to ethanol

Converting pentose to biohydrogen

Fermentative biohydrogen production is an emerging technology and has received increasing interest in recent years as a sustainable energy source for fuel cells. The dark fermentative hydrogen process is environmentally friendly, cost-effective and sustainable. Moreover, this process is considered a promising treatment technology for organic waste and/or residue with simultaneous clean, highly efficient energy production. During the dark fermentation process, hydrogen is produced together with CO_2 in the gas phase and organic acids and solvents in the liquid phase as the end-products. Substrates that have been used for hydrogen dark fermentation are mainly carbohydrate-containing feedstock such as glucose, sucrose and starch. We have used pentose for producing biohydrogen. Cultures of extremely thermophilic bacteria have been enriched for biohydrogen production and adapted to convert hydrolysate. Hydrolysate contains most of the pentose and is a harsh environment for microbial growth because of several toxic substances formed during the thermal pretreatment process. Compounds such as furfural, hydroxymethyl-furfural and organic acids are examples. Long-term adaptation of the enriched microbial culture enabled the organic compounds in hydrolysate to be converted to hydrogen and entirely detoxified.

We have developed a two-step process in which biohydrogen is produced in a first step and methane in a subsequent step (Fig. 8.9). The process can be optimized by recycling the methane produced through the hydrogen reactor and thus reducing the hydrogen partial pressure, resulting in thermodynamically increased efficiency.

The gas mixture produced comprises CH_4 and H_2. Using this in internal combustion engines leads to many advantages in terms of combustion efficiency and engine performance due to the specific physical and chemical properties of the two fuels.

A new process was developed recently that applies a slight voltage potential in the reactor to convert the organic matter into hydrogen: electrohydrogenesis.

Biogas

The biogas process is an established technology and considered the most efficient way to convert a broad range of biomass to energy. Although biogas has mainly been used for producing electricity and heat, biogas can be upgraded for use in transport. However, infrastructure is required to use it generally in transport. Nevertheless, the biogas process is very versatile and non-selective with regard to substrate and is therefore an excellent way to remove organic matter and polish effluent streams. Codigestion of waste streams has been shown to be a way of optimizing the bio-gas process to increase substrate utilization and to decrease process inhibition. Biogas is a complex microbiological process requiring different groups of bacteria to collaborate in a balanced way for successful digestion.

Fig. 8.10 shows the anaerobic digestion process schematically.

Several groups of microorganisms are involved in the conversion process, such as hydrolytic, acidogenic and acetogenic bacteria and methanogenic Archae. Archae are distinctive from bacteria and are supposedly older evolutionary than bacteria. We examined the distribution of Archae and

Fig. 8.9. Process for producing hythane (H_2 + CH_4 mixture)
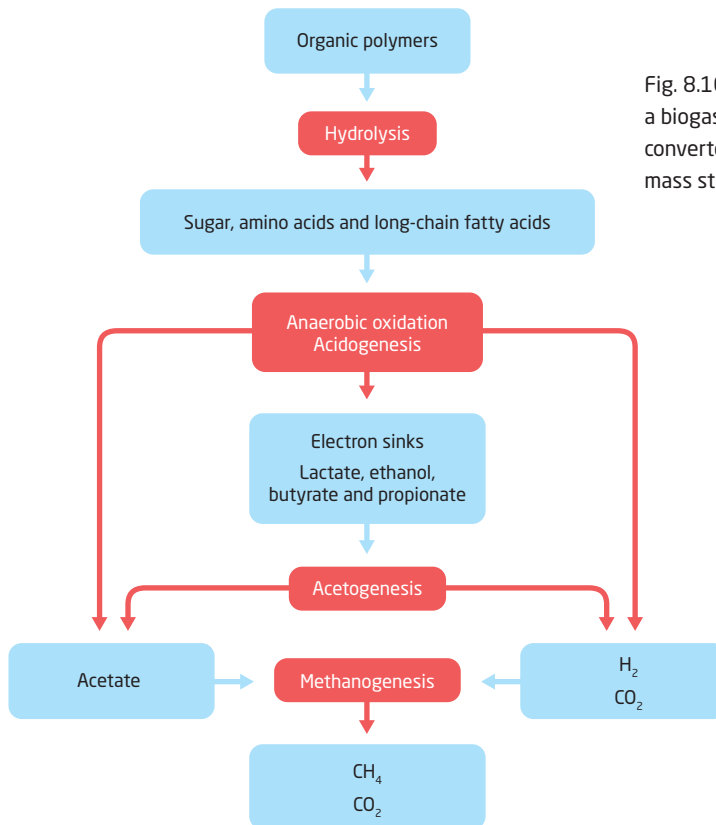
Fig. 8.10. The primary conversion processes in a biogas reactor, in which organic matter is converted to biogas, with typical relative mass streams

bacteria in biogas plants in Denmark by using specific probes targeting 16S RNA, which can produce different colors for bacteria and Archae. Fig. 8.11 shows the distribution of bacteria and Archae in the Fangel biogas plant. Understanding the factors determining the establishment of specific methanogens may enable manipulation of the microbial composition of a biogas reactor and thereby increase the efficiency of the reactors.

Sustainability of biofuels

For biofuels, the focus should be on the potential of biofuels to reduce global warming: reducing emissions of greenhouse gases (most importantly CO_2, N_2O and CH_4). Several other issues are also highly relevant such as air pollution with soot, aerosol particles, nitric oxide, carbon monoxide and ozone, which affects human health. In addition, ozone is a greenhouse gas and negatively
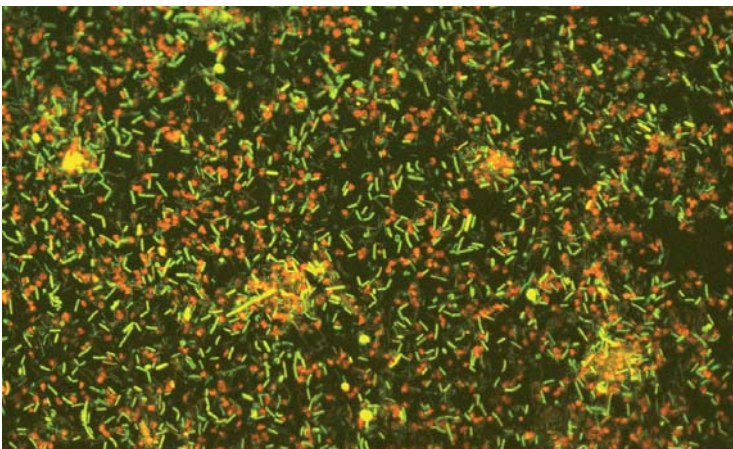
Fig. 8.11. The red microorganisms are methanogens, and the green ones are bacteria (hydrolytic, acidogenic and acetogenic)

Global warming potential

	0%	20%	40%	60%	80%	100%	120%

Methane manure, optimized
Methane manure + cosubstrate, optimized
100% recycled vegetable oil methyl ester (France)
Ethanol, whey
100% recycled vegetable oil methyl ester (Switzerland)
Methanol, fixed bed
Methane, wood
Methanol, fluidized bed
Ethanol, sugar cane
Ethanol, grass
Ethanol, wood
Ethanol, sweet sorghum
Ethanol, sugar beets
Methane, sewage sludge
Methane, grass biorefinery
100% soy methyl ester (USA)
Methane, biowaste
100% palm oil methyl ester
100% rape methyl ester (Switzerland)
Methane, manure + co-substrate
Methane, manure
100% rape methyl ester (Europe)
Ethanol, corn
Ethanol, rye
Ethanol, potatoes
100% soy methyl ester (Brazil)
Natural gas, Euro 3 emission standard
Diesel, low sulfur, Euro 3 emission standard
Gasoline, low sulfur, Euro 3 emission standard

Fig. 8.12. Relative global warming potential of biofuels (source: Zah et al. 2007)

affects plant growth. The energy used to produce, handle and process the feedstock should be considered as well as the change of land use for growing fuel crops and their potential influence on food production and food prices.

Sustainability analysis is quite complicated, and comparing analyses is often difficult. An activity has therefore been started to certify biofuels using a common set of criteria.

As shown previously, a wide variety of feedstock can be used for biofuels. The task of calculating the sustainability of biofuels is not easy, and there is some controversy about exactly how much

greenhouse-gas emissions are reduced depending on how many factors the analysis includes. Fig. 8.12. gives an overview of the global warming potential of various biofuels relative to gasoline. Biogas from manure has the lowest global warming potential, but lignocellulosic ethanol, which typically saves 50–80% compared with fossil fuel, is an attractive technology. Corn ethanol and biodiesel from rapeseed oil save much less, often only about 20%. However, for biodiesel only seeds were used, whereas using the whole plant can substantially change the sustainability of oil seed plants for energy production. Another benefit of bioethanol is

that replacing as little as 6% of the gasoline avoids the need to add the toxic methyl tertiary-butyl ether (MBTE) to increase the octane rating.

When land-use change is considered, the greenhouse-gas balance might become negative. A recent study comparing energy solutions for transport concluded that the highest-ranking solutions were wind-powered battery-electric or hydrogen fuel cell vehicles. The lowest-ranking solutions were corn and lignocellulosic ethanol. It was even concluded that they may actually worsen climate and air pollution. The main reasons for this are that, despite the relatively high overall greenhouse-gas savings, there are other environmental issues, especially for lignocellulosic ethanol, which requires a large land footprint and results in high air pollution, increasing mortality rates.

The use of biomass for biofuels should therefore be considered carefully, and biorefineries should be justified not solely on their biofuel production but also on the production of high-value products that can substitute for fossil fuel.

More to explore

Blume SB, Hauggaard-Nielsen H, Jensen ES. *Opgørelse af den danske biomasseressource til brug for fremstilling af biobrændstoffer til transportsektoren frem mod 2020* [Calculation of Denmark's biomass resources available for producing biofuel for transport until 2020]. Risø, National Laboratory for Sustainable Energy, Technical University of Denmark, 2008 (Risø-R-1665(DA)).

Jacobson MZ. Review of solution to global warming, air pollution and energy security. *Energy and Environmental Sciences*, 2009:2:148–173.

Zah R et al. *Life cycle assessment of energy products: environmental assessment of biofuels.* St. Gallen, Empa (Swiss Federal Institute for Materials Science and Technology) Technology and Society Lab, 2007.

The authors

Irini Angelidaki:
MSc in Chemical Engineering, 1989 and PhD, 1992, DTU. Professor, Department of Environmental Engineering, DTU. Specializes in anaerobic microbiology and processes; development of mathematical models for simulation of the anaerobic process and investigation of methods for monitoring and control of the biogas process; development of physical, chemical, enzymatic and microbiological methods for degradation of lignocellulosic material; isolation and characterization of anaerobic bacteria; molecular methods for characterization of bacteria; biological production of hydrogen; microbial fuel cells; bioethanol production; and biorefineries.

Kim Pilegaard:
MSc, 1975, PhD, 1978, and Postdoctoral Fellow, 1978–1981, University of Copenhagen. Risø DTU National Laboratory for Sustainable Energy, 1981–. Head, Biosystems Department, 2005–. Research: bioenergy and effects of energy use on the environment; microbial techniques for energy, bioprocessing, physical and chemical processes, resources, environment and sustainability.

Computer simulation:

dynamical systems and multiscale modeling

Per Christian Hansen, Jan S. Hesthaven & Jens Juul Rasmussen

Computer simulation plays a central and ever-increasing role in everyday life. Daily weather forecasts result from intensive computation, and search engines on the Internet rely on rapid computation on extremely large data sets. Likewise, train schedules are produced by computer solutions, and hearing aids that can adapt to noisy environments run highly sophisticated computation in real time. In the near future, advanced simulation tools will help physicians in diagnosis and even patient-specific treatment and will improve the control of electricity generation and distribution.

The developments in computer technology and mathematics are leading to a new and exciting era in which complete systems can be modeled and simulated (and not just their individual components) and simulation tools are being developed that are powerful enough to analyze and synthesize highly complex systems. This chapter surveys the use of computer simulation to study complex mul-

tiscale systems and illustrates some of our efforts with a case study from fusion energy research.

Simulation science
Problems in engineering are often highly complex, and the computer is an indispensable tool in modern engineering, such as for simulating machines and processes and for improving their design and functionality. Computing how much energy a wind turbine generates requires understanding the detailed flow around the wings, the mechanics of the wings, the tower and the weather at the site. The role of computer simulation is even more dominant in engineering science for development, design and validation purposes. Devising and implementing computer simulation tools are becoming as important as the two traditional approaches in engineering and science: developing theoretical models and conducting laboratory experiments (Fig. 9.1). Computer simulation is the third leg of

Fig. 9.1. The triangle of modern engineering science

Theoretical model development

Computer simulation

Laboratory experiments
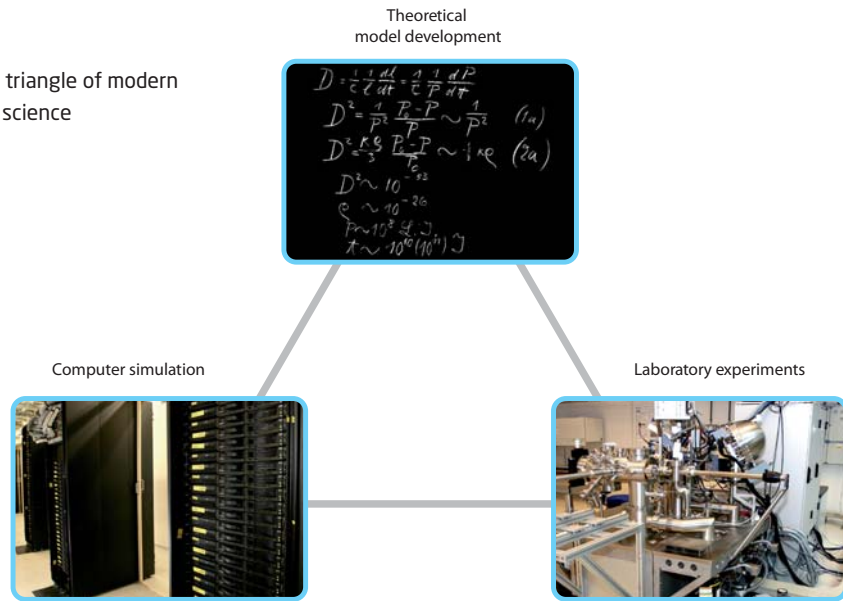
modern science, in addition to theory and experiment.

There is every reason to expect this importance to grow in the future. To illustrate this, Henry F. Schaefer III of the University of Georgia predicted that computational modeling will expand its share of modeling by one percentage point annually until it reaches 50% in 2030. At that point, experimental and computational modeling will be of equal importance for analyzing, synthesizing and designing complex systems across the natural sciences and engineering (Fig. 9.2).

Traditionally, computational efforts have carefully modeled the individual components of the complex system and used experience and application insight to draw conclusions about the performance of the complete system. Although this approach has been highly successful in the past, the increasing demand for improved performance of new technology pushes this approach to its limit. We are beginning to experience a paradigm shift toward full system focus to enable the modeling, design and synthesis of revolutionary new technologies and devices (Fig. 9.3).

Complete systems need to be simulated in a wide range of different applications, such as patient-specific drug design, hydrological studies and optimal energy planning. In the near future, such advanced computer simulation will also be required for validation purposes. These and similar types of sophis-
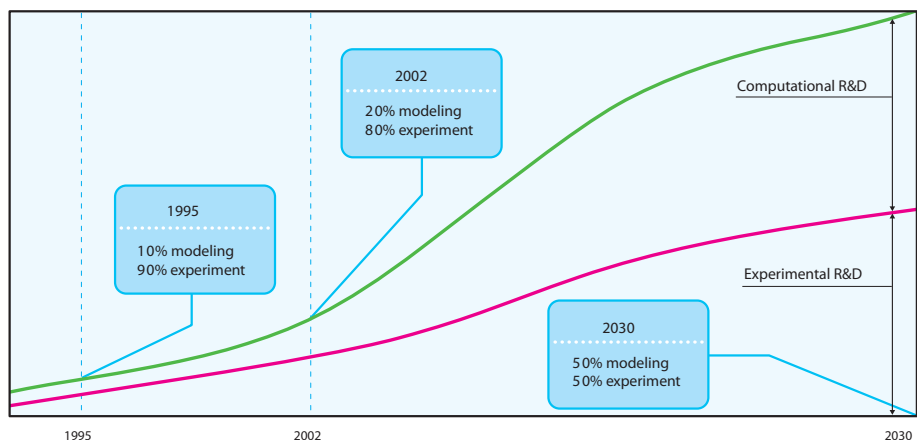
Fig. 9.2. Trends in modeling of complex systems, 1995–2030
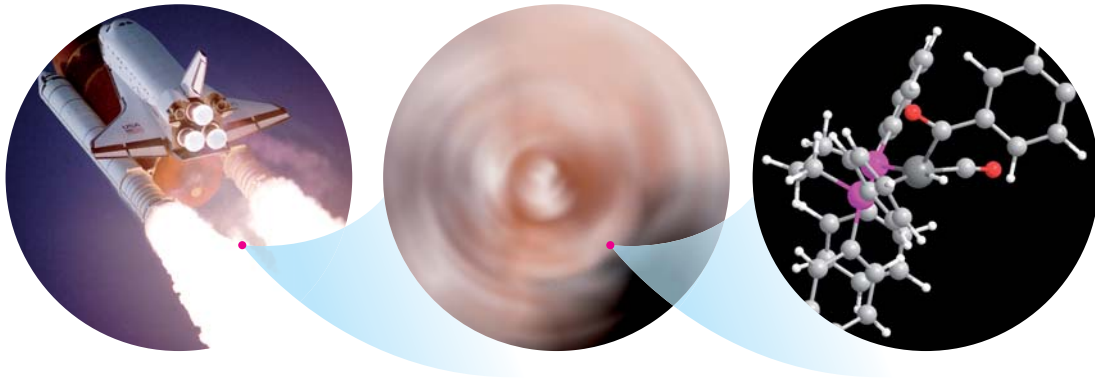
Fig. 9.3. Given the increasing demand for reliability, modern computer simulation efforts must account for all parts and scales of a problem and the important coupling between these scales. This leads to the introduction of multi-scale science and the simulation of complex systems.

ticated simulation can only be achieved by using multiscale techniques.

Major advances in both computational technology and mathematical techniques have initiated this shift in the approach towards computational modeling. The main research challenge lies in the interdisciplinary collaboration between application scientists, computational scientists and applied mathematicians needed to enable this effort. In particular, ideas and methods need to be developed to enable interscale coupling in multiscale and multiphysics systems, based on models of the individual components.

Computers today and tomorrow

In 1965, Gordon E. Moore (co-founder of Intel) predicted that the number of transistors on a computer chip would double every 1–2 years. He was correct: over the past 40 years, the performance of central processing units (CPUs) has doubled every 18 months – a prediction now known as Moore's law. To fully appreciate the magnitude of this growth, imagine buying a car today for €20,000. If the development of this car had followed a similar path, 5 years ago it would have cost €200,000 and there would just be a few cars around. Twenty years ago, the price would have been €200 million and it would have had the status of the space shuttle. Five years from now it would cost only €200 and auto mechanics would be unnecessary as the car would simply be disposed of after use.

There is no other known example of such growth in history. It is unique and is transforming all aspects of society and how society addresses scientific questions and seek to develop innovative new technologies. If this development continues, in 2023 a €1000 chip will have the complexity and capacity of the human brain. In 2036, this same chip will cost a cent. In 2049, a €1000 chip will equal the brain complexity of the entire human race and only a decade later this chip will cost a cent.

Nevertheless, the need to model increasingly complex problems continues to outpace this remarkable performance growth. All predictions for the future indicate that the gap between what can be accomplished by simply relying on the continued growth of computer performance and what will be required to simulate central societal problems will continue to grow. This increasing performance gap prevents attempts to address key problems related to the environment, new energy forms, national security and the health of the population through individualized medicine if they rely only on computational power.

As more and more complex problems are studied, the quantity of data increases dramatically. At the same time, users produce data and need to process them and perform simulation in shorter and shorter time frames. The obvious solution to these conflicting demands is to distribute the computation among several processing units, resulting in a parallel computing model. This idea was introduced more than 50 years ago, and since the early

1990s, it has been the dominant path towards larger and faster supercomputers. However, even this approach, leading to the clustering of hundreds, thousands or even tens of thousands of individual computer units into large networks, is beginning to see its limits in two main ways: the chips need more and more electric power (most of which is dissipated into heat, which again requires extensive cooling of the systems), and the gap between the CPU performance and the rest of the computer – in particular, the memory – gets wider and wider.

In 2004, this picture changed when multi-core CPUs appeared on the market. Here, the required increase in computational power is obtained by concurrent execution of the programs on several units – instead of increasing the clock cycle. Four- and eight-core CPUs are currently available, and in the near future, this number will increase dramatically. Moore's law is still valid, but instead of doubling the clock-speed every 18 months, cores will double in that period.

On a larger scale, researchers continue to use parallel computer systems for simulation. Such systems consist of clusters of smaller computers connected through a fast network (distributed computing) or larger machines with many processing units connected to a shared memory through a crossbar switch (shared memory computing). DTU uses several computers of both types, such as Niflheim at the Department of Physics, Alfheim at the Department of Mechanical Engineering, the installation at the Department of Systems Biology and the High-Performance Computing Installation (Fig. 9.4).

A very interesting recent alternative to increasing the computational power by replicating CPU cores – in which each core is quite complex – is executing the numerical computation on a graphics-processing unit (GPU), which consists of many small and specialized cores. In many cases, these GPUs are simply high-end graphics cards, driven down in price by a worldwide demand for extreme graphics for computer games. If this is done correctly, the performance of such cards can be used to accelerate large-scale simulation by one to two orders of magnitude for many types of calculation. Already today, GPU-based mini-parallel computers with four specialized graphics cards are available, taking up no more space than a standard personal computer yet reaching performance levels reserved for large-scale supercomputers just a few years ago.

Fig. 9.4. DTU's High-Performance Computing Installation currently consists of two symmetric multiprocessing computers with 96 dual-core processors in total and more than 600 GB of memory
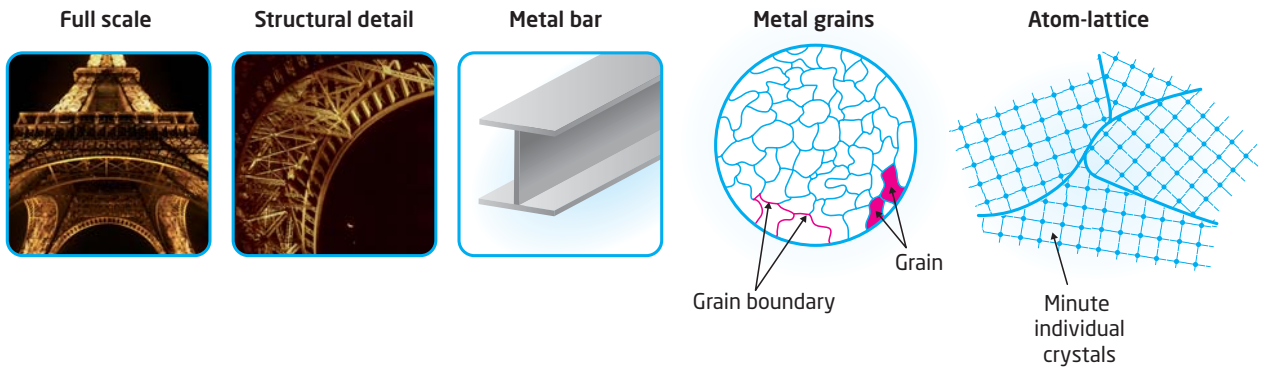
Fig. 9.5. A multiscale structure – from full scale, over structural detail to the metal bar, metal grains and atom lattice

In the future, and in a sense already today, there are several trends. The number of cores per CPU will increase, while the increase in clock speed is likely to stagnate to limit power consumption and heat production. An emerging model for the efficient use of novel technology is a hybrid model in which the algorithms are designed to switch the workload between the CPU and the GPU, depending on what is the most appropriate technology. In fact, the first computer on the top 500 list of June 2009, the IBM Roadrunner, is such a machine – this was the first computer that could execute more than 1 petaflop per second (10^{15} floating-point operations per second).

The modeling challange

The remarkable increase in computational capacity during recent decades makes one wonder why the gap between what can be modeled and what should be modeled continues to grow. As already mentioned above, the answer to this can be found in the interest and the need to simulate systems that are more complex and to require a higher degree of fidelity of the results. This requires not merely considering the individual components of larger systems but doing so in a fully coupled fashion. Consider, as an illustration, the modeling of a large structure such as the Eiffel Tower (Fig. 9.5).

If one is merely interested in the structural stability of the structure, then simple force considerations may well suffice. However, modeling the more delicate features requires complementing this with detailed analysis at levels of increasing detail,

including the molecular scale, to understand crack-tip development and the strength of the welding, among many other things. Further, these microscopic effects affect the dynamics and structural integrity of the whole structure and therefore need to be reflected in the modeling of the full structure.

Modeling the full structure using a molecular approach may be tempting. However, since the molecules are about 10 nm and the full structure about 100 m, this indicates that about 10 orders of magnitude in spatial scales would need to be covered in each direction. Even if one could encode one degree of freedom in a single hydrogen atom, storing the information required to represent the whole structure would weigh 1 million tons – and we have not even begun to compute yet. We can expect a similar scale separation in the temporal dimension, rendering the simulation even more costly. This is clearly not feasible.

The need to model such systems introduces the need for new formulations to model the problem hierarchically, often with different models and computational techniques used at the different levels (Fig. 9.6). For the large scales, a macroscopic model is often used, typically developed using basic principles of conservation of mass, momentum and energy, leading to models based on differential equations and often known as continuum models. Methods for computationally solving such problems are well developed and include standard techniques such as finite difference and finite element methods. These models typically allow the solution to be computed relatively rapidly, even for large and geometrically complex systems undergoing
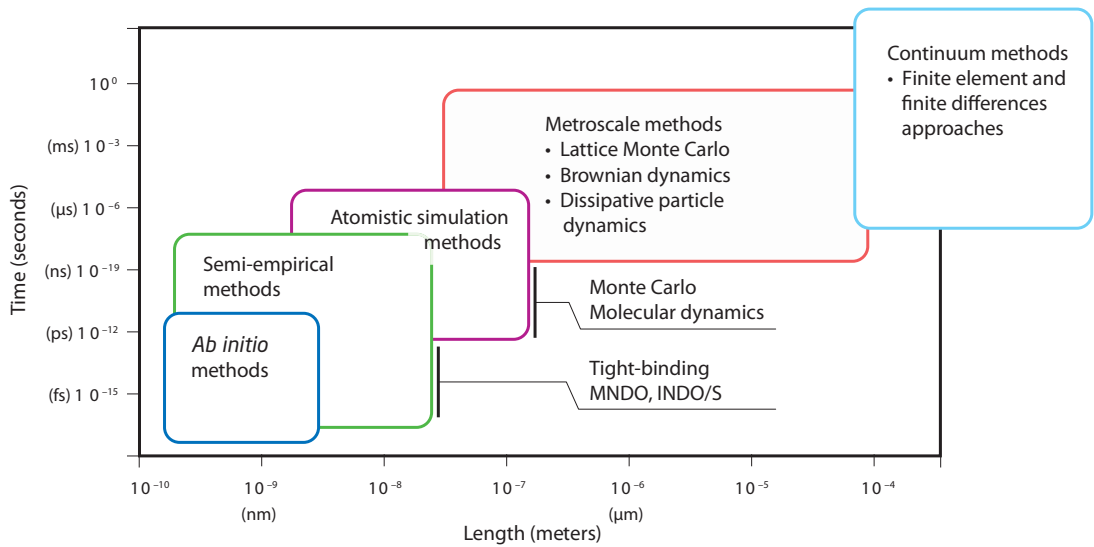
Fig. 9.6. Techniques used for modeling as a function of time and length scales

time-dependent changes. In the past decade, methods for solving such problems have improved dramatically through the development of geometrically flexible, robust and accurate high-order methods that are essential to enable the efficient modeling of dynamic problems over long time scales.

When phenomena with a scale length close to the mean free path of the molecules in the problem at hand become important, such as for crack-tip propagation in a mechanical construction, such simple macroscopic models break down and a more detailed model is required.

At the other end of the scale spectrum are the purely atomistic *ab initio* models, based on quantum mechanics and many-body systems. As already mentioned, these cannot realistically be used to model the whole system. The challenge is how to accurately and efficiently couple these types of models to enable the full system to be modeled.

Returning to the Eiffel Tower example, this means that we should not (and cannot) choose between, say, a classical structural model and a model based on molecular structures. Instead, we must link the different models at the different scales into a complete model of the Tower, in which the most appropriate model is used at each scale. For example, molecular models are used to describe the development of cracks at the microscopic level, and

the effects of these cracks are propagated to the structural model that describes the stability of the Tower at the macroscopic level.

Multiscale modeling
Such modeling techniques, known as multiscale modeling, have become central in recent years to enable the simulation of complex engineering systems. The main challenge lies in identifying where the fine scales enter into the macroscopic models and in exploiting the understanding of the scale separation in the physical system. For instance, one can use a microscopic model, obtaining the state variables from the macroscopic model through a process called downscaling, to compute the local material properties around a crack tip and, through upscaling, use these microscopic properties to compute a macroscopic quantity. Such techniques often lead to equation-free modeling, since the actual equations being solved are not known explicitly but the impact of the equations can be obtained through a hierarchy of models.

Another interesting and recent development is the development of novel computational techniques that can model phenomena effectively across many scales. Such techniques, most often found in the modeling of fluid phenomena, typi-

cally rely on a Lagrangian description of the physical system in which the continuum medium is described by a large collection of particles with various properties. Such mesoscale methods, exemplified by smooth-particle hydrodynamics, lattice-Boltzmann models and dissipative particle dynamics, hold great promise for modeling and coupling between the extreme scales or even as stand-alone multiscale simulation tools. Nevertheless, such mesoscale techniques are considerably more computationally demanding than more traditional continuum models.

Even with the development of efficient and accurate multiscale techniques, many challenges lie ahead since these developments remain in their infancy and are largely application-specific. As the granularity of the physical model increases, more and more parameters are introduced to describe local material properties, concentrations, reaction rates and local thermodynamic properties. The vast majority of these quantities are unknown or only known with substantial uncertainty.

Although straightforward techniques such as Monte Carlo and quasi–Monte Carlo methods are appealing, they are often impractical for quantifying uncertainty in complex physical systems. Several alternative methods, based on either statistical or probabilistic ideas, are beginning to enter into large-scale simulation and certification efforts. However, they all lead to a substantial increase in the computational workloads because the sensitivity of the solutions needs to be quantified with respect to the high-dimensional parameter space. Problems are no longer just three-dimensional and time-dependent but can have a parameter space with hundreds of dimensions.

Modeling of turbulence and transport in magnetized plasmas

Numerical simulation has become an indispensable tool in contemporary fusion research, which aims at developing a sustainable energy source for the future (Box 9.1). Simulation is now widely applied to model various aspects of the dynamics of the hot plasma confined by a magnetic field. Further, computer simulation tools are increasingly applied in developing new materials used in constructing the device and, in particular, for model-

ing the interaction of these materials with the immense neutron flux originating from the fusion processes.

The dynamics of magnetically confined plasmas is a true multiscale problem, with spatial scales ranging from the electron gyro radius (sub-millimeter) to the size of the machine (several meters) and temporal scales from the electron gyro period (10^{-10} second) to the energy confinement time of seconds. Even if the fundamental equations that govern the plasma dynamics were known, simulating plasma dynamics would be a formidable problem. Thus, despite the considerable progress in modeling and the advances in high-performance computing, we are still far from full-scale numerical modeling of a fusion device with the ultimate goal of developing predictive models.

One of the most challenging and difficult areas in fusion research is understanding and controlling the transport of plasma particles and energy across the confining magnetic field. It is necessary to bridge the separate aspects and describe the turbulence and the slow, large-scale evolution of the profiles self-consistently for progressing towards predictive capability. Some of the first results have been obtained using simplified equations describing the turbulence and transport in the edge region of magnetized plasmas by means of the ESEL (Edge-SOL-Electrostatic) model developed by the Risø DTU National Laboratory for Sustainable Energy. The model includes the self-consistent evolution of the pressure and flow profiles – in the two-dimensional plane perpendicular to the magnetic field. The geometry comprises distinct production and loss regions, corresponding to the edge region and the region where the magnetic field ends on material surfaces designed to take up the power load from the escaping plasma, the scrape-off layer. The simulations are resolving the smallest spatial scales of the dynamics and are run over time spans from the ion gyro period to several energy confinement times, more than 10^8 time steps to obtain time series containing the slow as well as the fast dynamics and allowing advanced statistical treatment. The results revealed the possibility to reproduce, using a first-principle model without free parameters, the statistical features of the plasma fluctuation and transport and to predict the density and temperature profiles within limitations (Box 9.2).

Box 9.1. Fusion energy research

Fusion energy, which powers the Sun and the stars, is released when light elements, such as deuterium and tritium, fuse together. This occurs at very high temperature, where all matter is in the plasma state. Plasma is a collection of free ions and electrons that is usually macroscopically charge neutral but contains enough charged particles that the dynamical behavior is governed by collective interactions. In neutral gases, pressure forces and collisions determine the dynamics. In plasmas, the charge separation between ions and electrons gives rise to electric fields, and the flow of charged particles sets up currents and in turn generates magnetic fields. These self-consistent fields provide long-range interactions and result in a wealth of complex phenomena. In addition, plasmas are influenced by externally applied electric and magnetic fields.

Open and worldwide coordinated fusion energy research started 50 years ago. It was realized early on that building a reliable fusion power plant would be extremely challenging. However, the prospect of fusion power is very attractive. Fusion offers a safe, clean, CO_2-free energy source with a fuel that is abundantly available everywhere. The challenges are multiple; they do not lie primarily in heating the plasma to the necessary 150 million K (Kelvin) but in confining the plasma particles and energy effectively and for sufficient time for the fusion processes to produce excess energy. It is important to isolate the hot plasma from the material walls of the container, and the walls have to be able to sus-

tain the energy fluxes produced by the escaping fusion products. Aside from plasma physics issues, materials science faces immense challenges in developing new materials that are stable and resistant to large energetic neutron fluxes and intermittently high power loads.

The most successful strategy for confining the hot plasma is magnetic confinement, with the tokamak (toroidal chamber with magnetic coils) configuration now being close to producing net energy. The basic principle is that the plasma is confined in a torus by an imposed strong toroidal magnetic field in combination with a weaker poloidal field created by a current in the plasma. Fig. 9.7 shows the largest and most advanced fusion experiment in the world, the Joint European Torus (JET) at Culham Science Centre in the United Kingdom, the central facility in the European Fusion Development Agreement.

The success of JET and several other experiments has led to the next step: the construction of ITER (International Thermonuclear Experimental Reactor), a large-scale research and technology development tokamak, in Cadarache, France. The construction phase is expected to last for the next 10 years. With ITER, fusion research has turned into a true worldwide collaborative effort: one of the major scientific endeavors being undertaken globally. The purpose of the project is to demonstrate physically and technically viable net fusion energy production (500 MW) on a large scale.
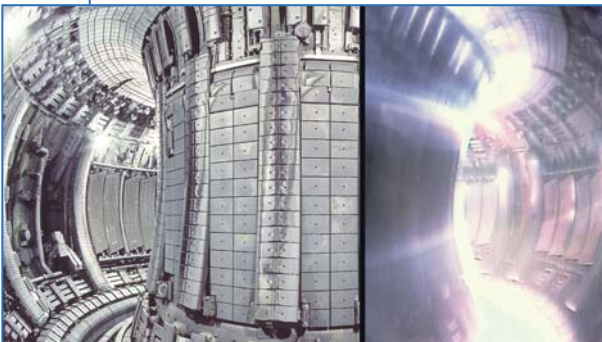
Fig. 9.7. The JET tokamak experiment. The toroidal vacuum chamber is seen from the inside. The height inside the chamber is about 4 m. The left frame shows the empty chamber, and the right frame shows the plasma during a discharge. Note that the light emission is concentrated near the wall with the highest neutral background. ©EFDA-JET

Box 9.2. Transport in the plasma-edge region

Fig. 9.8 illustrates the simulations of plasma-edge dynamics, showing the evolution of the particle density and the vorticity (the rotation of the velocity) during a burst in the turbulence intensity. The particle density takes the form of a localized blob of enhanced density that is generated in the edge region and propagates outwards through the scrape-off layer in a similar fashion to the experimental observations. The propaga-tion velocity of the blob structures corresponds to around 5% of the ion sound speed, close to the experimental observed blob velocity. Associated with the blob structure in the particle density is a dipolar vorticity field, which provides the mechanism for advecting the whole structure. Since vorticity is not directly measured in experiments, the simulation has added to the understanding of the blob dynamics.
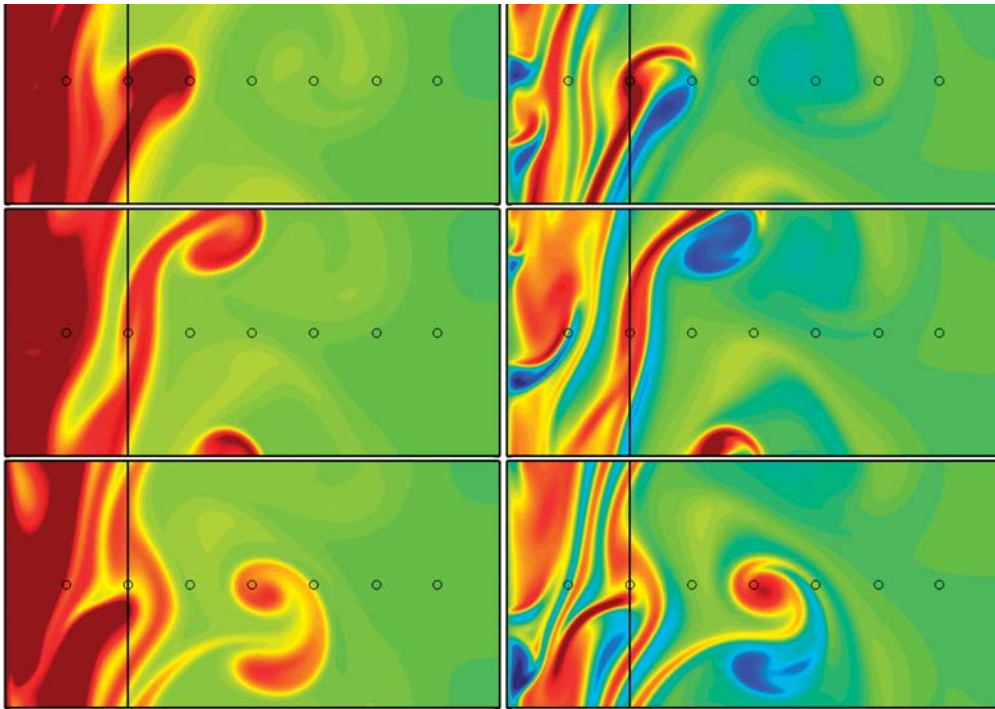
Fig. 9.8. Evolution of plasma blobs from simulations. **Left column**. Density evolution (red is high density). **Right column**. Vorticity evolution (red is positive and blue negative vorticity). The vertical line designates the border between the edge region and the scrape-off layer. The time interval between the frames is 500 ion gyro periods, corresponding to 32 μs for deuterium ions and a magnetic field of 2 T

Conclusion

The combination of powerful computer technology, innovative and novel simulation techniques and a close interface between the simulation and experimental results provides unrealized opportunities to harvest the power of simulation science. We are just at the very beginning of this transformation in the way we do science, the way we develop insight into new complex systems and the way we develop new technology through innovation.

One challenge is educating and equipping the new generation of scientists and engineers for this paradigm shift. This change will require more strongly emphasizing computational and mathematical tools than those to which most past engineers or scientists have been exposed and emphasizing the development of core competencies in these fields as simulation science matures as an everyday tool.

According to Horst Simon of Lawrence Berkeley National Laboratory, "Computing is changing more rapidly than ever before, and scientists have the unprecedented opportunity to change computing directions". DTU has a great challenge and opportunity to properly educate the next generation of engineers in developing efficient algorithms and physical models, to ensure that the next-generation models can take full advantage of the unique opportunities to harvest the power of emerging computational technologies.

More to explore

Garcia OE, Horacek J, Pitts RA, Nielsen AH, Fundamenski W, Graves JP, Naulin V, Rasmussen JJ. Interchange turbulence in the TCV scrape-off layer. *Plasma Physics and Controlled Fusion* 2006: 48: L1–L10.

Garcia OE, Naulin V, Nielsen AH, Rasmussen JJ. Turbulence and intermittent transport at the boundary of magnetized plasmas. *Physics of Plasmas* 2005: 12: 062309.

Hesthaven JS, Warburton T. *Nodal discontinuous Galerkin methods: algorithms, analysis, and applications*. New York, Springer, 2008 (Springer Texts in Applied Mathematics 54).

Kritz A, Keyes D. Fusion simulation project workshop report. *Journal of Fusion Energy* 2008: 28: 1–59.

Naulin V. Turbulent transport and the plasma edge. *Journal of Nuclear Materials* 2007: 363–365: 24–31.

Oden JT et al. *Simulation based engineering science*. Arlington, VA, National Science Foundation, 2006 (http://www.nsf.gov/pubs/reports/sbes_final_report.pdf).

Sørensen HHB, Hansen PC, Petersen DE, Skelboe S, Stokbro K. Efficient wave function matching approach for quantum transport calculations. *Physical Reviews B* 2009: 79: 205322–205331.

The authors

Per Christian Hansen:
MSc in electrical engineering, PhD and DrTechn in numerical analysis, DTU. BIT Prize 1990, Statoil Prize 1994; highly cited mathematician (ISI Web of Knowledge).
Currently Professor of Scientific Computing, DTU. Research: computational methods for inverse problems.

Jan S. Hesthaven:
MSc and PhD in numerical analysis, DTU. Alfred P. Sloan Research Fellow 2000, National Science Foundation Career award 2001. Professor of Applied Mathematics and Director, Center for Computation and Visualization, Brown University. Research: computational methods for time-dependent partial differential equations.

Jens Juul Rasmussen:
MSc in electrical engineering; PhD in plasma physics. Risø National Laboratory: 1978 –. Adjunct Professor in Nonlinear Sciences DTU 2001–2007. Presently Professor in Plasma Physics and Head of Plasma Physics and Technology Program, Risø DTU. Research: turbulence and transport in magnetized plasmas, nonlinear dynamics.

Waste and climate

– energy recovery and greenhouse gases

Thomas H. Christensen

Introduction

Society is politically, economically and technically facing the increasing challenge of climate change and struggling to reach targets for reducing greenhouse-gas emissions. No single solution is available, and every sector of society must analyze its current contribution and consider how to improve performance.

This focus on climate change and greenhouse gases has also influenced waste management. The main purpose of waste management is to remove the waste from industrial and residential areas without adversely affecting public health and the environment. Nevertheless, society's need to find renewable energy sources and reduce greenhouse-gas emissions has given a new dimension to waste management: how and how much can waste management contribute to this societal challenge of reducing greenhouse-gas emissions?

Waste management primarily emits greenhouse gases because of fuel use in collection trucks and facilities and waste treatment: for example, incinerating plastic or releasing methane (CH_4) from an anaerobic digestion process. However, waste management can also produce energy that may replace the energy from fossil fuel and recover materials, thereby conserving energy in producing new resources and products. The challenge is to find a balance between emitting greenhouse gases and contributing to reducing the emissions of greenhouse gases at the societal level that lead to the best overall contribution to reaching the goal of reducing these emissions. A life-cycle perspective is needed to balance between aspects because the actual emissions and the reductions take place in different sectors of society and sometimes in different countries. The greenhouse-gas issues have been instrumental in changing the focus from managing waste as a problematic residue to managing waste as a potential residual resource.

During the past decade, the DTU Department of Environmental Engineering has performed laboratory research, pilot-scale experiments and full-scale testing with the aim of establishing mass balances, energy budgets and emission accounts for waste management technologies. The technological insight and data form the basis for modeling waste management technologies and systems with respect to resource and energy recovery. The EASE-WASTE (Environmental Assessment of Solid Waste Systems and Technologies) model integrates this in a life-cycle perspective and has been used for calculating the greenhouse-gas accounts presented here.

Greenhouse gases in waste management
The main greenhouse gases in waste management are carbon dioxide (CO_2) of fossil origin, CH_4 and nitrous oxide (N_2O). Occasionally, chlorofluorohydrocarbons (CFCs) from old refrigerators and insulation materials and sulfur hexafluoride (SF_6) from, for example, thermal glass may locally be a source of greenhouse gases.

CO_2 of fossil origin primarily originates from the use of electricity and fuel and from the combustion of plastic materials. Energy use is associated with all technical processes within waste management, while the contribution from plastic is from waste incineration or waste-to-energy plants. CO_2 of recent biogenic origin is considered neutral with respect to global warming. Most organic material in the waste originates from short-rotation crops and thus was recently synthesized by taking up CO_2 from the atmosphere. Forestry has a longer rotation period, but paper waste is only a small contributor to the stock of biological carbon bound by forestry, and this justifies the assumption that biogenic CO_2 from waste management can be considered neutral with respect to global warming. However, this implies that biogenic carbon not released to the atmosphere as CO_2, such as that stored in the soil after applying compost or buried in a landfill, avoids emission and should thus be ascribed a negative global warming potential of -3.67 kg of CO_2 equivalent per kg of carbon stored. The time horizon associated with the consideration of stored carbon is discussed later. Fig. 10.1 illustrates in an idealized world, using paper as a specific case, how

carbon is circulated and used within and between the waste sector, the paper industry, the energy sector and the forestry sector. Modeling of the contributions to increased CO_2 in the atmosphere for a range of waste management scenarios involving recycling paper, incinerating paper with and without energy recovery, landfilling paper and using forestry biomass for energy production showed that this global warming potential is internally consistent and thus useful for assessing how waste management best contributes to reducing greenhouse-gas emissions.

CH_4 has a global warming potential 25 times higher than that of CO_2. It originates from degrading organic matter under anaerobic conditions. The main sources of CH_4 within waste management are landfills and anaerobic digesters. Microniches in composting processes may be anaerobic, and composting processes may thus also release small amounts of CH_4, but rarely does more than a few percent of the degraded carbon become CH_4. Given the high global warming potential of CH_4, strict control of release is important. The means include collection and flaring or using the gas or using biofilters in which biological processes oxidize CH_4 into CO_2.

N_2O has a very high global warming potential of 298 kg of CO_2-equivalent per kg of N_2O. N_2O is primarily formed during the biological conversion of nitrogen: during oxidation of ammonium into nitrite and nitrate and during reduction of nitrate into nitrite and free N_2. Very little of the nitrogen is released as N_2O, and its formation is hard to control. Composting processes and the use of compost in soil are likely to be the main sources within waste management.

Waste collection and transport
Waste collection and transport are primarily linked to diesel combustion and do not differ from any other bulk transport by truck. Much of the cost of waste management is associated with collecting and transporting waste, and much is done to minimize the costs. This usually involves optimizing collection schemes and transport routes, implicitly reducing fuel consumption. However, as illustrated later, the energy and global warming aspects of waste collection and transport are minor compared
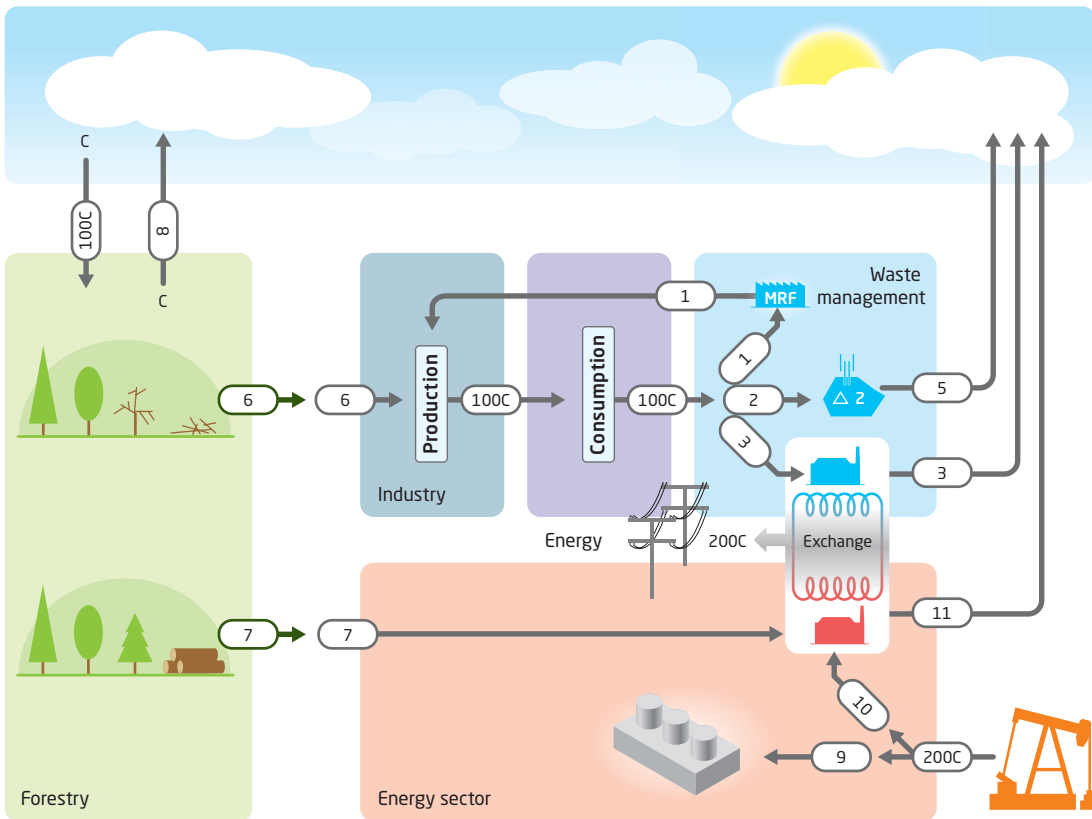
Fig. 10.1. Simplified model for calculating carbon (C) balances in waste management including allied industries (energy, paper and forestry). MRF: materials recovery facility.

with the recovery and recycling aspects of the collected waste.

Recycling materials

Recycling of materials in the waste often saves substantial energy in the industries using the recovered materials instead of virgin materials. These savings are accounted for in the involved industrial sectors, but since waste management provides the materials, these savings are considered indirect savings downstream to waste management.

Metals such as aluminum, steel, iron and copper require substantial energy to produce from ore, and scrap metal is a valuable material for smelters. Recovering the metals from scrap is less energy intensive, and the overall recycling process saves energy. The greenhouse-gas savings are substantial:

for example, 5000–19,000 kg of CO_2 equivalent per ton of aluminum and 500–2400 kg per ton of steel. The actual value depends on the technology used and the location of the smelter. Smelters running on hydroelectric power may reduce greenhouse-gas emissions less.

Glass recycling involves cleaning and reusing bottles and remelting crushed glass (cullet) for producing new glass containers and bottles. Depending on purity, cullet can be added in significant quantities to the production of new glass from virgin materials (quartz sand, soda, lime, etc.), reducing the melting temperature and hence the energy needed. About 400 to 1400 kg of CO_2 equivalent is saved per ton of glass waste.

Recycling is simple for clean plastic of a single type, such as pure low-density polyethylene. The mechanical process involves melting, extrusion

and granulate production. For dirty and mixed plastic, the process is more complicated and demanding. Virgin production of plastic uses about twice as much energy as the energy content bound in the plastic itself. Plastic recycling saves 0–1500 kg of CO_2 equivalent per ton.

Paper and cardboard recycling saves from minus 300 to plus 1200 kg of CO_2 equivalent per ton. The actual savings depend on the product quality and the pulping technologies. Most recycling processes are primarily based on fossil energy, whereas some virgin technologies use energy from biomass. If the biomass saved by recycling paper is assumed to substitute for fossil fuel, the overall saving is probably 1800–4400 kg of CO_2 equivalent per ton. This aspect nicely demonstrates that reprocessing the waste material requires energy but the savings depend on what it replaces, perhaps not directly but at the system boundary. The latter is not always easy to define, because it depends on so many other factors in society.

The stipulated greenhouse-gas savings from recycling waste are substantial, but the quantity and quality of materials are important from a waste management viewpoint. Thus, based on 1 ton of waste, paper recycling is by far the dominant recycling activity reducing greenhouse-gas emissions.

Incineration

Incinerating waste releases significant energy. The lower heating value of waste is typically around 10 GJ/ton depending on what the waste contains. Modern incinerators produce electricity and heat, with the heat often used for district heating. The electricity produced amounts to 20–28% of the lower heating value, and with flue gas condensation the overall energy recovery may be close to 100%. The electricity is sold to the grid, where it replaces electricity that elsewhere would have been produced at high marginal cost. The source of this marginal electricity determines the potential savings. In Denmark, this is usually electricity produced from coal without co-production of district heating. The savings are therefore likely to be high, probably about 0.9 kg of CO_2 equivalent per kWh of electricity produced. Elsewhere, the marginal production of electricity may differ. In the scenario calculations shown later, European mixed electric-

ity was used, which is close to the greenhouse-gas emissions from electricity produced using natural gas: about 0.5 kg of CO_2 equivalent per kWh.

The substitution value of the heat produced largely depends on the local conditions, because hot water and steam cannot be transported very far without substantial losses. If the local district heating system is small and hosts a power plant, then producing heat in the waste incinerator, even if fed to the district heating system, may not substantially reduce greenhouse-gas emissions if the power plant already has excess heat. Thus, estimating the substitution value of the heat produced by incinerating waste may require more detailed analysis of the local conditions and heat market.

The incinerator itself also emits greenhouse gases from combustion of plastic materials and any use of fossil fuels for start-up operations, whereas the CO_2 from organic waste is considered neutral. Depending on the waste incinerated, about 30–50% of the CO_2 emitted is of fossil origin. The current information is based on calculations considering the expected composition of the waste and the observed energy content, but methods for direct measurement in the flue gas are underway. A precise estimate is important because waste incinerators, in contrast to power plants, are not subject to CO_2 quotas but part of the national reporting on greenhouse-gas emissions.

An additional option is producing fuel from the waste that can be used in power plants and industrial kilns (such as cement kilns) to replace fossil fuels, often coal. Power plants and industrial kilns usually have less extensive flue-gas cleaning systems, so using a refuse-derived fuel requires a clean waste fraction. In particular, the mercury and the chloride concentrations must be closely controlled. Refuse-derived fuel can be used flexibly and directly replaces coal based on the energy content. As shown later, this may provide some advantage compared with electricity production in an incinerator if this electricity replaces marginal electricity based on natural gas.

Balancing the direct emissions from the incinerator against the savings obtained from the electricity and the heat produced that otherwise would have been generated from fossil fuels, 100–800 kg of CO_2 equivalent per ton of waste may be saved overall.

Biological treatment

Composting organic waste involves aerobic degradation and generates CO_2 and compost. The composting process emits few greenhouse gases, primarily from the use of machinery and emissions of CH_4 from anaerobic niches in the waste and small amounts of N_2O from the mineralization of nitrogen. The overall emissions are believed to be about minus 10 to plus 300 kg of CO_2 equivalent per ton of organic waste.

Digestion of organic waste is an anaerobic degradation of the organic waste in a closed reactor. The main product is biogas, constituting about equal quantities of CH_4 and CO_2. The CH_4 is often used to produce electricity and sometimes heat for district heating. The greenhouse-gas emissions are related to the use of fuels and unintended leakage of CH_4. Savings are obtained from the electricity and heat delivered to society. The overall emissions are believed to be between minus 300 to 0 kg of CO_2 equivalent per ton of organic waste.

The stabilized organic products from biotreatment used on land for crop production supply the soil with nutrients and organic matter. If the organic residue is used rationally according to a fertilization scheme, it may substitute for the production and use of the mineral fertilizers N, P and K (N and P are the most important nutrients in this context). This may save 4–80 kg of CO_2 equivalent per ton of incoming organic waste. The carbon in the stabilized material will slowly decompose in the soil. This process may take decades; in particular, the compost has a high fraction of humus-like material that degrades very slowly. Model calculations show that, after 100 years in the soil, the compost may still have about 10% of the carbon left. This should be considered a saving in greenhouse-gas emissions of 2–70 kg of CO_2 equivalent per ton of organic waste.

Landfilling

Landfilling waste generates CH_4 and CO_2 from the anaerobic degradation of the organic matter. The degradation process continues for decades after disposal. Most of the CH_4 is generated within 30 years after disposal, but even after 100 years about half the biogenic carbon is left in the landfill.

In engineered landfills, the landfill gas is extract-

ed by vacuum and burned to convert CH_4 to CO_2. If the amount of gas is limited or highly variable, the gas is burned in a flare, but substantial gas is often used to generate electricity by burning the gas in an engine with a generator. Sometimes the heat is also used, at least for internal purposes at the landfill.

Much of the gas is not collected unless the landfill is covered by a synthetic liner. The uncollected gas may escape through cracks or coarse soil and emits substantial greenhouse gases. If the gas flow is low and distributed over large areas of the landfill, the CH_4 may get oxidized passing through the soil cover. CH_4-oxidizing bacteria are present in most landfill cover soils and use oxygen diffusing in from the atmosphere to oxidize CH_4 rising from the waste below. The CH_4 is oxidized into CO_2, thus reducing the global warming potential from 25 kg of CO_2 equivalent per kg of CH_4 to 0 kg of CO_2 equivalent per kg of biogenic CO_2. However, if the flow is high or the soil too compact, CH_4 oxidization is limited.

The biogenic carbon left in the landfill after 100 years, the time period modeled, is a greenhouse-gas saving and ascribed a negative global warming potential as described earlier. For 1 ton of municipal waste with a substantial content of organic waste and paper (without separate recycling schemes), the biogenic carbon left equals as much as a saving of 130–180 kg CO_2 per ton of waste landfilled.

The overall greenhouse-gas contribution from landfilling may be minus 70 to plus 170 kg CO_2 equivalent per ton.

System considerations

The input to the waste management system is the waste in terms of quantity and composition. Municipal solid waste is often one third paper and paper products by weight, one third kitchen organic waste and one third other waste. Fig. 10.2 shows model estimations of the energy that can be recovered from 1 ton of waste by a range of waste management scenarios. For all the scenarios, collecting and transporting the waste is a very small part of total energy. Using 10–15 liters of diesel for collecting and transporting 1 ton of waste is not important when the waste contains the equivalent of
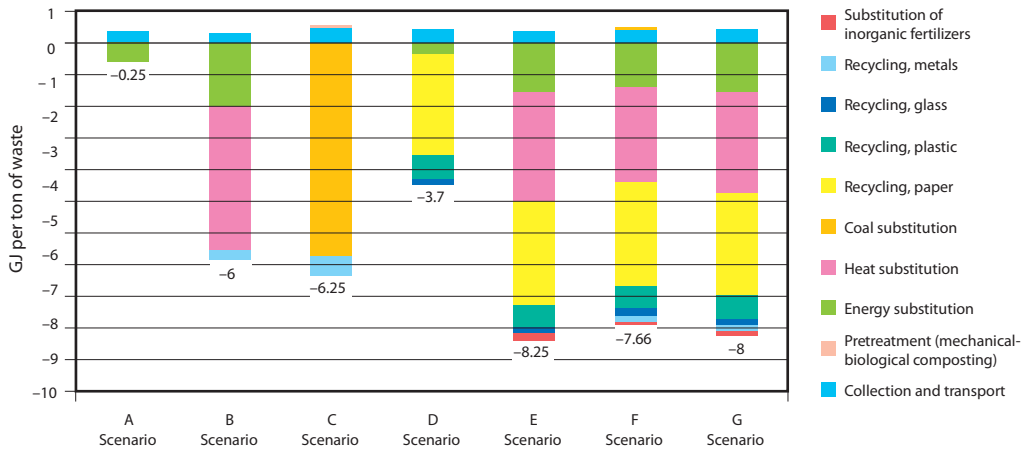
Fig. 10.2. Energy balance (negative values are recovered energy) for seven waste management scenarios. **A**. Landfilling. **B**. Incineration. **C**. Production of fuel for coal-fired power plants. **D**. Materials recycling and landfilling. **E**. Materials recycling and incineration. **F**. Materials recycling, organic waste composting and incineration. **G**. Materials recycling, organic waste digestion and incineration.

200–250 liters of diesel per ton. Seven scenarios have been modeled.

A: A modern landfill with a high level of gas collection using the collected energy for generating electricity recovers 0.25 GJ net per ton of waste. This scenario is included as a reference since landfilling organic waste is prohibited in many countries.

B: Incineration with 20% electricity generation and 40% heat recovery for district heating saves 6 GJ net per ton of waste. The heat recovery rate is typical for Europe but low for Denmark, with 65–80% heat recovery given the extensive district heating.

C: A refuse-derived fuel fraction is produced and used in a power plant instead of coal. The remaining waste is stabilized by composting and landfilling. This mechanical-biological treatment technology is common in central Europe. This saves about 6 GJ per ton, equivalent to the amount of coal saved without considering how the coal is converted into energy at this point. This emphasizes the fact that energy recovery not only depends on the amount recovered but also on the form. For the other scenarios, the energy is in the form of electricity and heat.

D: This scenario also focuses on landfilling, but much waste is recycled before landfilling: paper, plastic and metal. The overall savings are 3.7 GJ per ton, of which paper recycling is the major contributor.

E: This scenario has the same recycling scheme as in D, but the residual waste is now incinerated, with energy recovery similar to B. The overall energy saved is now about 8 GJ per ton: paper recycling, electricity generation and heat production are the three main contributors.

F and G. These two scenarios are identical to E except that an organic fraction is collected separately at the source: in F for composting and in G for anaerobic digestion. Digesting organic waste generates some electricity from the biogas produced, but the overall energy recovery is nearly identical, about 8 GJ per ton. Paper recycling and energy recovery from incineration are predominant.

Fig. 10.3 presents the same scenarios expressed in net CO_2 equivalent per ton of waste. The picture is somewhat similar to that in Fig. 10.2, since all waste-management scenarios emit negative greenhouse gases or reduce global warming. However, two issues are different.

The first difference between energy recovery and reductions in greenhouse-gas emissions is that the scenarios with significant landfilling (A and D) contribute relatively more to reducing greenhouse gases than they do to energy recovery. This is due to the biogenic carbon left in the landfill after the 100-year period considered for the modeling. About half the organic carbon is still present in the landfill and thus constitutes a saved emission since biogenic carbon, if released as CO_2, is considered neutral.

The second difference between the energy recovery and the reduction in greenhouse-gas emissions is that scenario C, which produces refuse-derived fuel for use in power plants, reduces greenhouse gases the most. This is because this scenario is credited the full greenhouse-gas emissions avoided from burning coal, whereas many of the other scenarios that recover the same amount of energy are credited the greenhouse-gas emissions from producing average European energy, resembling using natural gas, which produces less greenhouse gases than coal.

The scenario modeling shows that waste management can contribute to reducing greenhouse-gas emissions, primarily through energy recovery and paper recycling. Storing organic carbon in landfills is also an option but less attractive than recovering the energy in the waste. Waste management can probably recover net energy of about 8 GJ per ton of municipal waste, and the net reduction in greenhouse gases could be about 500 kg of CO_2 equivalent per ton of waste. The actual value depends strongly on the type of energy replaced by the energy recovered from incineration. This reduction in greenhouse-gas emissions from waste management corresponds to about 2% of the current load caused by an average European person.

Conclusion

Waste management can contribute to society's efforts to reduce greenhouse-gas emissions by minimizing its own use of fossil fuels, by increasing the recycling of materials and energy recovery and by binding biogenic carbon in soil and landfills. The
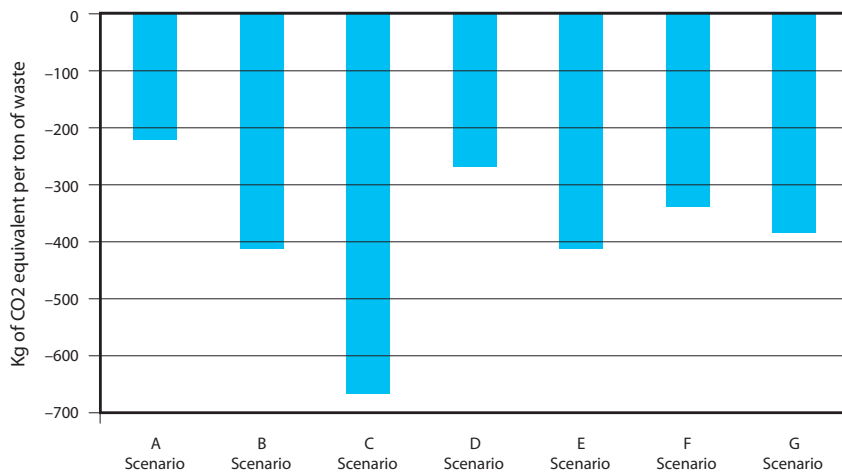
Fig. 10.3. Global warming contribution (negative values are reductions) for seven waste management scenarios with a 100-year time perspective. **A**. Landfilling. **B**. Incineration. **C**. Production of fuel for coal-fired power plants. **D**. Materials recycling and landfilling. **E**. Materials recycling and incineration. **F**. Materials recycling, organic waste composting and incineration. **G**. Materials recycling, organic waste digestion and incineration.

scenario modeling suggests that energy recovery and paper recycling are the most important factors in reducing greenhouse-gas emissions. Waste management may reduce current net greenhouse-gas emissions from an average European person by about 2%.

For waste management, the challenge is to develop a system that uses the least energy, minimizes greenhouse gases and maximizes energy recovery and material recycling. The challenge is to balance material recycling and energy recovery and to maintain the overall perspective, although much of the reduction in greenhouse gases is outside the waste management system. The solution is to develop transparent data on all levels and to establish models synthesizing the complex data into communicable terms. Climate-friendly waste management is not the solution but one of many necessary contributions to meeting targets for reducing greenhouse-gas emissions. The waste needs to be managed, so why not do it the climate-friendly way?

More to explore

Bruun S, Hansen TL, Christensen TH, Magid J, Jensen LS. Application of processed organic municipal solid waste on agricultural land: a scenario analysis. *Environmental Modeling and Assessment* 2006: 11: 251–265.

Christensen TH, Gentil E, Boldrin A, Larsen A, Weidema B, Hauschild M. C balance, carbon dioxide emissions and global warming potentials. *Waste Management and Research* 2009: doi:10.1177/0734242X08096304.

Gentil E, Aoustin E, Christensen TH. Waste management: accounting of greenhouse gases and global warming contributions. *Waste Management and Research* (in press).

Manfredi S, Christensen TH. Environmental assessment of solid waste landfilling technologies by means of LCA-modeling (EASE-WASTE). *Waste Management* 2009: 29: 32–43.

The author

Thomas Højlund Christensen: MSc in Environmental Engineering, 1974, DTU; PhD in Water Chemistry, 1980, University of Washington, Seattle; Dr.Agro. in Soil Chemistry, 1990, Royal Veterinary and Agricultural University, Copenhagen. Professor, DTU Department of Environmental Engineering. Professor at DTU in environmental technology since 1990, specializing in waste and resource management. Recent focus: integrated modeling of waste management systems.

Forecasting and optimizing transport

Oli B.G. Madsen & Otto Anker Nielsen

Introduction

Transport is about quality of life and prosperity – about connections between people, families and businesses. Transport infrastructure helps ensure that people can get to work and that products and goods can be transported to their destinations. This makes transport and infrastructure a vital cornerstone of social welfare.

The average Dane spends more time on transport than on completing primary education, and households in Denmark spend more money on transport than on food on average.

Denmark's shipping fleet is among the largest in the world and earns considerable foreign currency for Denmark: even more, for example, than agriculture.

During the past 20 years, road traffic in Denmark has increased by 50%, leading to considerable congestion problems. Measurements by the DTU Department of Transport show that road congestion in Copenhagen alone cost €1.34 billion in 2006. At the same time, the political and public focus on energy consumption, CO_2 emissions, noise and road safety has increased.

Improving decision support in the transport sector

Public authorities invest large sums in transport networks to improve performance. Using the funds efficiently requires being able to predict and evaluate the effects of the investment. This includes the effects on road traffic, energy use and CO_2 emissions, safety and local emissions and noise.

The effects of transport investment are usually very complex. Travelers from a given origin may change the frequency of trips, the destination, the mode of transport, when they make their trips and the routes they follow. Each route comprises a sequence of links and nodes in the network. Travelers' decisions depend on their socioeconomic characteristics and the quality of the transport options. The more traffic on a given transport network, the more congestion and the longer the travel times. This may make some travelers change their journey compared with their preferred option. Traffic models predict the behavior of travelers' complex decisions while they fulfill the equilibrium conditions that describe the congestion in the network.

After the travelers' behavioral responses to policies have been predicted, impact models can be used to forecast the effects derived on energy use and CO_2 emissions, local emissions and noise and road safety. Socioeconomic evaluation can then quantify these effects into monetary values, such as in cost–benefit analysis.

Transport companies can obtain great benefits by optimizing their operations. Examples are the design of an overall shipping network and locations of terminals, designing the schedules for the next year and reoptimizing transport for short-term occurrences. The decision problems are often considerable. The bus timetable in Copenhagen, for instance, is defined by more than 250 billion decision variables. Developing heuristics that can solve transport problems within reasonable calculation times is therefore a challenge, but the magnitude of the problems is such that manual approaches usually fail to find near-optimal solutions.

Transport science

In recent decades, transport science has developed into its own field, drawing on developments in mathematical modeling, statistics, software engineering, economics and behavioral research. Applied research has benefited from the availability of computers that are more powerful, better mathematical methods and general advantages in the field. This means that larger and more realistically described problems can now be solved than previously.

Transport forecast and optimization problems tend to be very large. Many of the applied problems described here would have taken years of computer power to solve by standard methods on a powerful PC. Brute force methods – such as using a cluster of computers – are therefore seldom the path forward compared with developing more advanced mathematical solution algorithms and clever ways of simulation.

Traffic models

Traffic models predict passenger choices by describing these as mathematical models. The traditional four-step model describes the choices as 1) the number of trips originated from a given

zone, 2) the choice of destination, 3) the choice of mode and 4) the route choice. If route choice models consider congestion, the fourth step finds equilibrium between route choice, traffic loads on the network and the resulting congestion.

It has long been recognized that the traditional four-step model is too simple to describe choice processes in transport. More advanced models include decisions on when to travel, choices of mode chains – such as bicycle to the station, train to the city and bus for the last segment of the journey. Activity-based models describe individual or even household activity patterns during the day, and derived from this, the transport decisions.

Research has developed discrete choice models to describe travel demand. The models are usually based on stochastic utility theory to describe both explainable and unexplainable aspects of behavior – the latter by an error term. Thus, some parts of the models describe a deterministic explainable part of travel behavior, and other parts describe unexplained or stochastic parts of behavior by statistical distributions. Newer models also include statistical distributions of passengers' preferences: taste heterogeneity. Closed-form mathematical models can be derived for some fairly simple cases.

Many models, however, cannot be solved analytically and need to be solved by simulation. More advanced statistical estimation methods and model structures have therefore been developed to refine the stochastic component of the models, by which the correlation between choices can be incorporated into the models, such as describing how buses and trains are often thought of differently from cars when choosing modes.

Research also focuses on the theoretical foundation for formulating the deterministic part of the models, drawing on microeconomics and behavioral theory. Advanced models can also capture nonlinearity in choice models today.

Transport models are thus formulated based on a mix of mathematical theory to fulfill logical constraints, economic methods to fulfill budget and time constraints and statistical methods to fit data. Models are typical estimated empirically based on transport surveys (various forms of interviews), experiments in which respondents answer hypothetical questions and aggregate observations such as traffic counts.

In recent years, new data have become available such as Global Positioning System (GPS)–based tracking of cars, Internet-based surveys and automatic counting equipment in public transport. This has enabled much larger samples to be obtained at reasonable cost and information on travel behavior to be indirectly derived from these new data sources. Most models are estimated on a cross-section of time, although some models are based on time-series and observations over time, such as panel data. The Danish Transport Survey has been carried out for a number of years, which enables changes in behavior to be analyzed over time. Applying models and using them for forecasts requires assuming – or forecasting – how the explanatory (exogenous) variables change over time. Some traffic models interact with economic models and models on the development of socioeconomic variables to refine the forecast input to the traffic models themselves.

Congestion

One of the major challenges in Denmark's transport system is increasing road traffic congestion. Fig. 11.1 illustrates an often assumed relationship between flow and speed at a road link, and Fig. 11.2 is an example of measurements on a real road.

The two figures illustrate how the speed decreases marginally with the flow at low levels. Speeds are even quite high near the capacity limit; in Fig. 11.2, speeds have decreased from an average
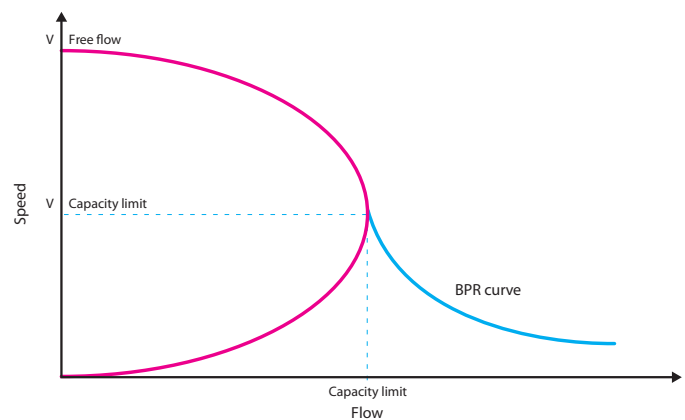

Fig. 11.1. The normally assumed relationship between speed and flow. BPR curve: curve based on a formula devised by the United States Bureau of Public Roads in 1964.

of 110 km/h to about 95 km/h near the capacity level of about 4000 cars per hour (distributed on two lanes). When the capacity is reached, however, speed and flow rates become unstable, and the speeds vary considerably. When the traffic flow switches from a fairly free-flow condition to a congested situation, both speed and flow rates reduce significantly. The road measured in Fig. 11.2 only manages to service 500 cars per hour and with an average speed as slow as 15 km/h in the worse case.

The figures illustrate that a road performs pretty well up to about 90% of the capacity, after which it becomes more unstable. A small increase in flow – and demand – at this stage may lead to a very steep deterioration of road performance. Or vice versa, if demand could be reduced marginally, performance would improve considerably.

Examples of policies to reduce congestion are introducing road pricing or congestion charging, using intelligent transport systems to control flows by traffic signals and information or increasing capacity by enlarging roads or building new roads. Predicting the effects of such initiatives and assessing the effects of transport policies are challenging, since the travel times depend on the traffic flows, but travelers' choices and hence the traffic flow depend on the travel time.
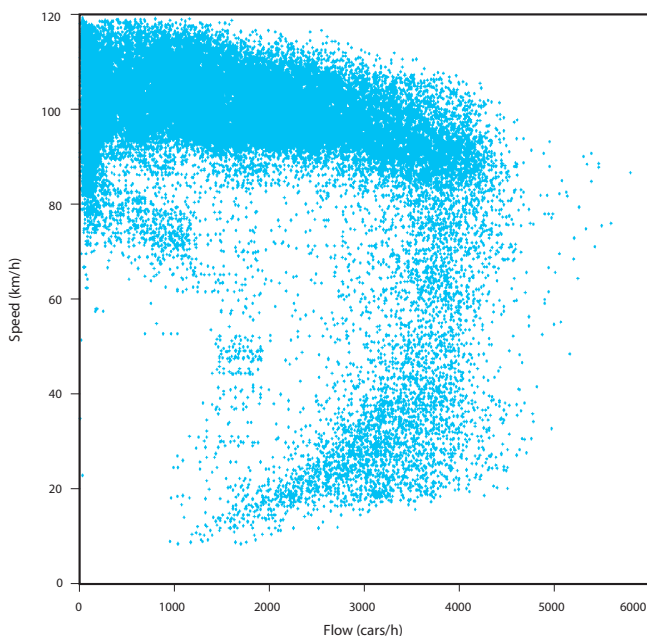
Traditional equilibrium models of traffic assignment assume a continuous speed–flow curve at each link and a static description of time. The models assign traffic onto the traffic network and find the equilibrium between the speed-flow curve and the route choice, where no traveler can save time by unitarily changing route. The speed–flow curve in Fig. 11.1 may continue above the capacity limit as indicated by the dotted line. An example is the widely used BPR formula that often provides a good fit to the upper part of the curve but ignores what happens after congestion develops (lower part of the curve).

In real networks, queues at one link may spill back on other links. So the congestion at each link cannot be modeled independently of other links, and each time period cannot be modeled independently from the other time periods. The queues start to grow at the beginning of the rush hour, and they increase over time until they peak. At the end of the rush hour, the queues start to decrease, but the network might still be congested even though the inflow of additional cars is lower than the capacity of the network. Finally, after the rush hour, the network is back to a free-flow state.

Travel times may vary significantly over time, which means that route choices may also vary over time. If a traveler leaves an outer suburb at 06:30, the network might be uncongested at that time, but when he or she reaches the city at 07:30, there might be considerable congestion.

This means that the two traditional assumptions in traffic assignment models – independence between links and independence between time periods – are no longer correct. The instability at the capacity limit also means that modeling variability in travel time is more important here than in uncongested networks. Comparing Fig. 11.1 and 11.2 shows this, where the real observations in Fig. 11.2 show much variation around the assumed curve in Fig. 11.1.

New generations of models therefore try to model the dynamics of traffic flows (dynamic traf-

Fig. 11.2. Example of a measured relationship between speed and flow (Ring Road 3 in Copenhagen)

fic assignment models) or to simulate the flows (macro, meso or micro simulation models, which model flows, packages of cars or each car separately). This makes it possible to model lines and travel time variability and how this influences travelers' choices.

Public transport models
Mathematically, public transport models are quite different from road transport models, since public transport is operated using timetables with discrete departures over time, whereas flows in road networks are often assumed to be continuous. Examples of timetable-based networks are urban public transport networks, national rail networks, air transport networks and many freight transport networks, such as international container shipping routes.

Timetable-based models describe the supply by the exact timetables, and the transport models calculate the passenger flows through these networks. Only in recent years has modeling large public transport networks become possible because of combining more efficient mathematical solution methods, larger computers and the availability of data in digital formats so that coding data in models has become less cumbersome.

Congestion and delays are also key in public transport. A recent research trend is to link railway simulation models that simulate railway operations with traffic models that simulate passenger flows through the system. This combination can be used to predict the delay distributions of new timetables. The DTU Department of Transport has worked on several projects for DSB S-tog a/s (suburban railway in Copenhagen) to develop methods of modeling and forecasting passenger delays and to gather information on passenger flows and estimate passenger flows from counting trains. This information can be used to improve future timetables.

Optimizing transport
The general public and governments have increasing requirements for energy consumption, noise and emissions. For both passengers and goods, the answer is flexible and multimodal transport. Making multimodal transport attractive and competitive requires good transport planning systems and using information technology equipment such as rapid communication systems, Global Positioning System (GPS), radio frequency identification (RFID) and on-board computers. The planning systems are based on mathematical models, optimization software and fast computers. The systems can be used at all levels from designing a transport network to handling disruptions and dynamic planning of the daily route for a lorry.

Vehicle routing
The simplest vehicle routing problem is known as the traveling salesman problem, originating more than 50 years ago, when traveling saleswomen were a very small minority. Assume that a salesman has a list of customers to visit on a given day. The salesman must start from home and return home at the end of the day. Plan a route such that all the customers are visited and the total distance traveled is minimized. It seems easy to solve, but finding the optimal solution might be difficult manually even with only 20 customers. Try to find the solution to the problem shown in Fig. 11.3 assuming that you can travel via straight lines between the customers.
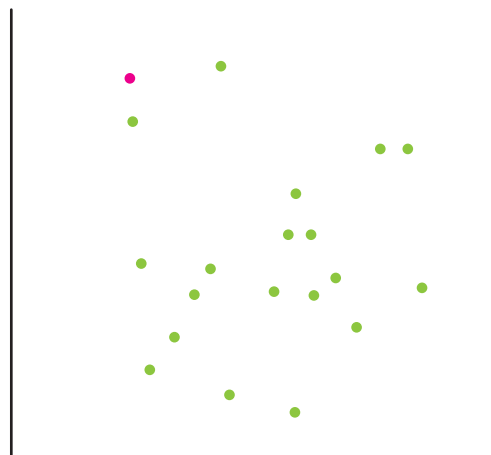
Fig. 11.3. Customers involved in the traveling salesman problem. Fig. 11.5 shows the optimal solution

Following streets instead of straight lines makes this much more difficult. The traveling salesman problem is very easy to formulate mathematically

using binary variables $x(i,j)$, which are equal to 1 if the salesman travels directly from customer i to customer j and equal to 0 otherwise. This is combined with two sets of constraints. One set ensures that each customer is visited once and another set ensures only one trip instead of several smaller separate trips. However, finding the optimal solution is quite difficult mathematically. The first optimal method appeared in 1970. The method was slow and could usually only handle a few customers. Today problems with 100 customers can be solved routinely, and with sufficient computer time, some problems with thousands of customers can be solved. The most recent solution methods for the traveling salesman problem are based on successive solutions of a linear programming problem (having a linear objective and a set of linear constraints, meaning that the assumption about binary variables is relaxed). After a linear programming problem is solved, several separation algorithms are applied to identify new special constraints (valid inequalities) that make the linear programming solution more integral and connected. The new constraints are included in the linear programming problem, which is solved again. The procedure continues until all original constraints are respected.

The traveling salesman problem is the simplest vehicle routing problem, and you can prove that the route must not cross itself. However, extending the problem to include several vehicles with capacity constraints and to include customer demands can obtain optimal solutions in which the specific route for each vehicle crosses itself and crosses the other routes. Under these circumstances, planning the routes manually might be very difficult. Fig. 11.4 shows the optimal solution to a vehicle routing problem with 100 customers and a specified time interval for each customer within which the vehicle is supposed to start servicing the customer. The depot is located in the middle and, to avoid confusion, the first and last legs of the 15 tours are not shown. Note that one customer just southwest of the depot is served individually. The figure demonstrates that this solution would be very hard to find manually. The optimal solution to most problems with 100 customers can be found in a few seconds using computers and modern solution methods. However, problems with 100 customers still exist that have not yet been solved optimally.

The vehicle routing problems mentioned above are still rather simple problems from an application viewpoint. To transport goods and people efficiently, dozens of mathematical models have been formulated and both optimality-finding methods and heuristic methods have been developed. Apart from minimizing the total distance traveled, the objectives might also consider environmental and energy aspects. Heuristic methods are based on an intelligent search, where an optimal solution is not guaranteed.
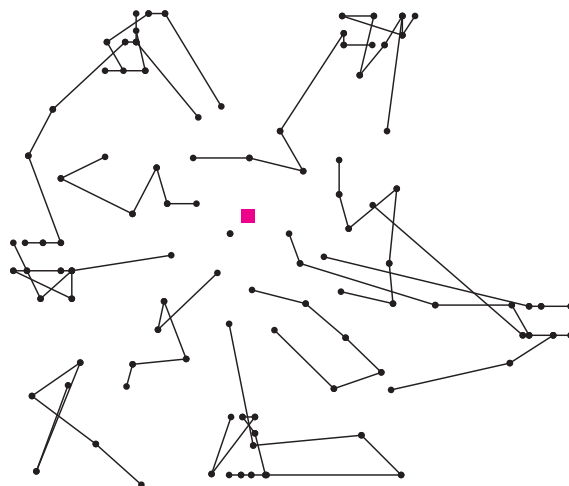
Fig. 11.4. Optimal routes for visiting 100 customers with time window requirements
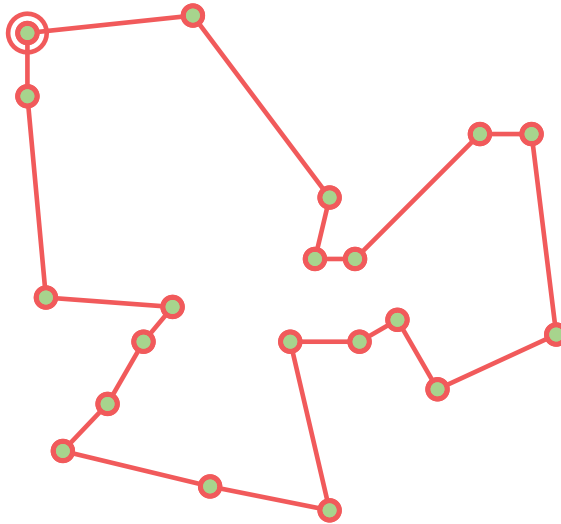
Fig. 11.5. Optimal solution to the travelling salesman problem shown in Fig. 11.3

Ship scheduling and network design

Denmark's shipping fleet is among the largest in the world and earns considerable foreign currency for Denmark. However, competition is hard, the vessels are costly to buy, they use substantial fuel and they emit considerable CO_2 and other pollutants. Planning the operation of the fleet as effectively as possible is therefore important. The planning at the strategic level may include what vessels to buy, what network to operate and where to lo-

cate terminals. The planning at the tactical level may concern how to schedule the vessels and whether to accept a transport order.

Imagine, for example, that you are running a container line and then you buy another container line serving its own network (Fig. 11.6). Now you have two networks and have to combine them. You have two fleets of vessels each with certain characteristics, forecasts of the demand for container transport and several potential harbours and ter-
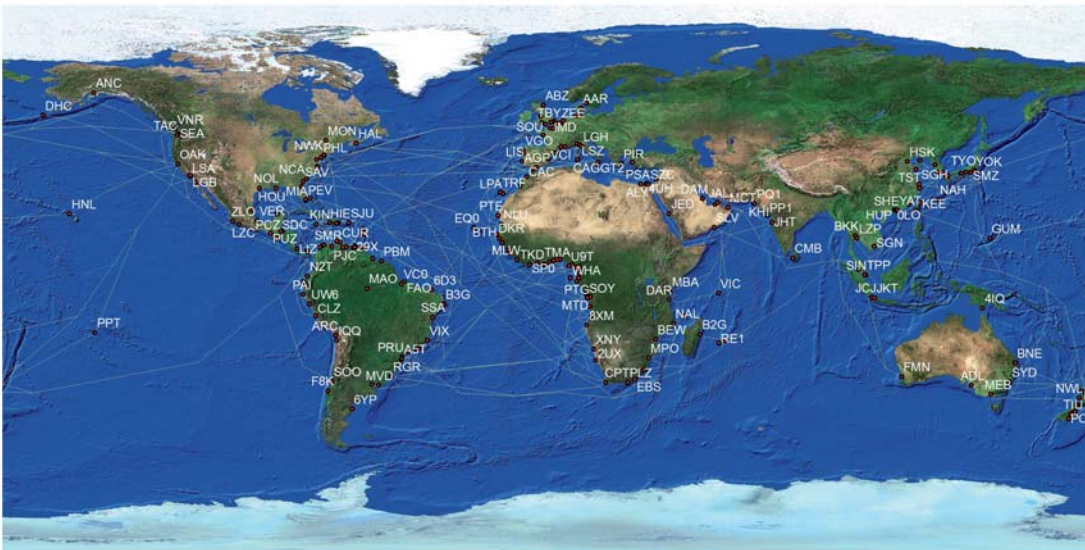
Fig. 11.6. Example of a container shipping network

minals to consider visiting. What should the new network look like and how should the vessels move in the new network? This can be formulated as an optimization problem of large dimensions and solved afterwards.

Combining traffic models and transport optimization

Many attempts have been made to combine real-size traffic models with transport optimization models. The following section presents an example.

Coordinated timetabling

Creating an efficient timetable is a challenge in a modern timetable-based transport system, such as a public transport system in a metropolitan area. Passengers often experience waiting times between two scheduled services. The DTU Department of Transport estimates that passengers in Greater Copenhagen spend about 200,000 hours every day waiting for connecting buses or trains. Planning or modifying the timetables to optimize or at least balance well the two objectives – passenger service and operating costs – is therefore important. The problem can be called the simultaneous vehicle-scheduling and passenger-service problem. You can imagine that if the timetable is adjusted to fit at one transfer point, it might fit much worse in other parts of the transport system. The transfer points must therefore be processed at the same time to find the optimal balance. The problem can be modeled as an integer programming problem and can be solved either optimally (for smaller problem dimensions) or by a heuristic method. Both approaches have been applied to the Greater Copenhagen bus network.

An example of a heuristic method is the tabu-search method, where the algorithm at a given point is trying to improve the best solution (or several candidate solutions) by testing variations. If a tested solution proves worse, it might still be chosen. The previous solution is then marked as tabu for a certain period of time. If the solution proves better then it is added to a list of solutions to be examined and improved further. When the list cannot be improved – after some further attempts – the best solution in the list is chosen.

The solution is typically examined through an object function, which is the criterion to optimize. Examples include minimizing travel time, maximizing the bus companies' revenue or a combined function. The government may want to maximize the benefits for society, which combines the benefits for the passengers, the cost of operation and the environmental effects, such as CO_2 emissions. The object functions are often nonlinear. A small change in the timetable of a bus route – such as reducing travel time by 1% by closing one bus stop – may enable the timetable to be served with 9 buses and 9 bus drivers instead of 10 – a saving of 10%.

One of the challenges of optimizing timetables with regard to passenger flows is the relationship between the timetable and the passenger flow (Fig. 11.7). The figure shows that two arrivals ($m + 1$ and $m + 2$) correspond with the same departure ($n + 1$). If passengers know the timetable – typically in low-frequency systems – more passengers will arrive with run ($m + 2$) than with ($m + 1$), since ($m + 2$) corresponds best with ($n + 1$). However, if the schedule of line n changes, this may influence the arrival pattern from line m. In a complex bus network, with many alternative routes, changes in flows are more complex, since passengers may completely reroute. One fundamental challenge of optimizing timetables is therefore to link the optimization model and the passenger flow prediction model – a bi-level problem.

The results of the tests in the Copenhagen network are encouraging and indicate a potential for reducing passenger waiting times by 10–12% with no additional vehicle scheduling costs. The best configuration of the model uses "synthetic" passenger flows that are generated by assigning passengers onto the network as if there were no transfer time. This resulted in a better solution than using the flows in the existing timetable. This means that the existing timetable is so far from optimum that the heuristics for timetable optimization are trapped in a local non-optimal solution from which the iteration between flow prediction and the optimization model cannot escape.

A more effective and user-friendly public transport system will also make more people switch from private car transport to public transport and thereby save energy and CO_2 emissions.

Time at stop X

$n+1$

$m+3$

$m+2$

n

$m+1$

$n-1$

m

Arrivals, route A   Departures, route B
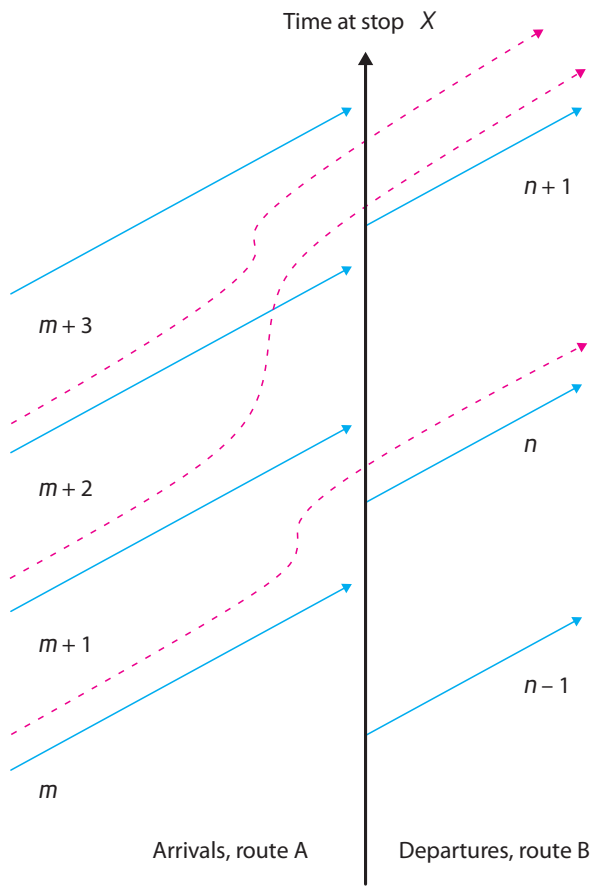
Fig. 11.7. Example of correspondence between two bus routes at a station

Conclusions and research challenges

Integrating transport forecasting models and optimization models is a promising research field that can contribute much to society, since the benefits of better planning and operation in the transport sector largely exceed the costs of the research and development of decision support tools. The examples illustrate the considerable benefits of using such methods. The trend is in the direction of models that can forecast and optimize even larger problems more realistically and closer to the optimal solutions, using less calculation time.

The combination of more efficient methods, fast-er computers and more standardized availability of data makes solving real-life problems easier and thus facilitates the use of research results in practice. Some problems are very large, and algorithmic refinement is therefore an important research topic. One example is a recent air transport optimization model for Greenland developed by the DTU Department of Transport. The first attempt to solve the problem by standard optimization methods would have required an estimated calculation time with a powerful PC of 12 years, but the final method used only 4 days to solve the same problem. The solution would be of great benefit for Greenland in reducing travel time and operating costs in its air transport system.

Behavioral data are required to estimate transport models, and such data are often costly to collect. One interesting research theme is therefore developing new methods of collecting data, such as by using GPS, and new, more efficient methods to design and carry out surveys – such as intelligent efficient statistical designs and adaptive surveys based on portable computers or the Internet. An ongoing research theme is developing methods to estimate transport models and developing new, more advanced models and mathematical formulations that reflect behavior.

Many public authorities and transport firms use transport models for forecasting and optimization models for improving operations. Applied projects are often carried out in collaboration between universities, consulting firms, information technology firms and the end-users of models. Some studies are very large, such as the new European TRANS-TOOLS Transport Model, which covers all of Europe and all modes, and the new Danish National Transport Model that cost many millions of euros to develop. The costs must be viewed based on the possible gains of choosing the right projects and the right design of each project. The TRANS-TOOLS model, for example, is used to set priorities for the European Union's support to national investments in the main trans-European transport network, and the new Danish National Transport Model has been used to refine decisions in the recent €12.5 billion infrastructure plan. Optimization models will definitely reduce the operating costs of the transport companies by more than the cost of developing the models.

More to explore

Applegate DL, Bixby RE, Chvátal V, Cook WJ. *The traveling salesman problem – a computational study*. Princeton, NJ, Princeton University Press, 2006.

Golden B, Raghaven S, Wasil E. *The vehicle routing problem – latest advances and new challenges*. Berlin, Springer, 2008.

Sheffi Y. *Urban transportation networks – equilibrium analysis with mathematical programming methods*. Englewood Cliffs, NJ, Prentice-Hall, 1985 (web.mit.edu/sheffi/www/urbanTransportation.html).

Train, K. *Discrete choice methods with simulation*. Cambridge, Cambridge University Press, 2003 (elsa.berkeley.edu/books/choice2.html).

Wilson NHM, Nuzzolo A, eds. *Schedule-based modeling of transportation networks: theory and applications*. Berlin, Springer, 2009 (Operations Research/Computer Science Interfaces Series, Volume 46).

The authors

Oli B.G. Madsen:
MSc (Electrical Engineering), PhD (Operations Research) and dr.techn. (Transport Optimization), DTU. Professor, Department of Transport, DTU. Director, Center for Traffic and Transport, DTU, 2000–2007. Visiting professorships, Massachusetts Institute of Technology and University of Auckland, New Zealand. Research: transport optimization, multimodal transport, intelligent freight transport and vehicle routing.

Otto Anker Nielsen:
MSc (Department of Roads, Transport and Urban Planning) and PhD (Transport Modeling), DTU. Professor, Department of Transport, DTU. Visiting professorships, Delft University of Technology, Netherlands and University of Montreal, Canada. Has also worked as Manager of Research and Development at ScanRail. Research: transport modeling, route choice models and traffic assignment, public transport planning, railway planning and operations, road pricing and congestion.
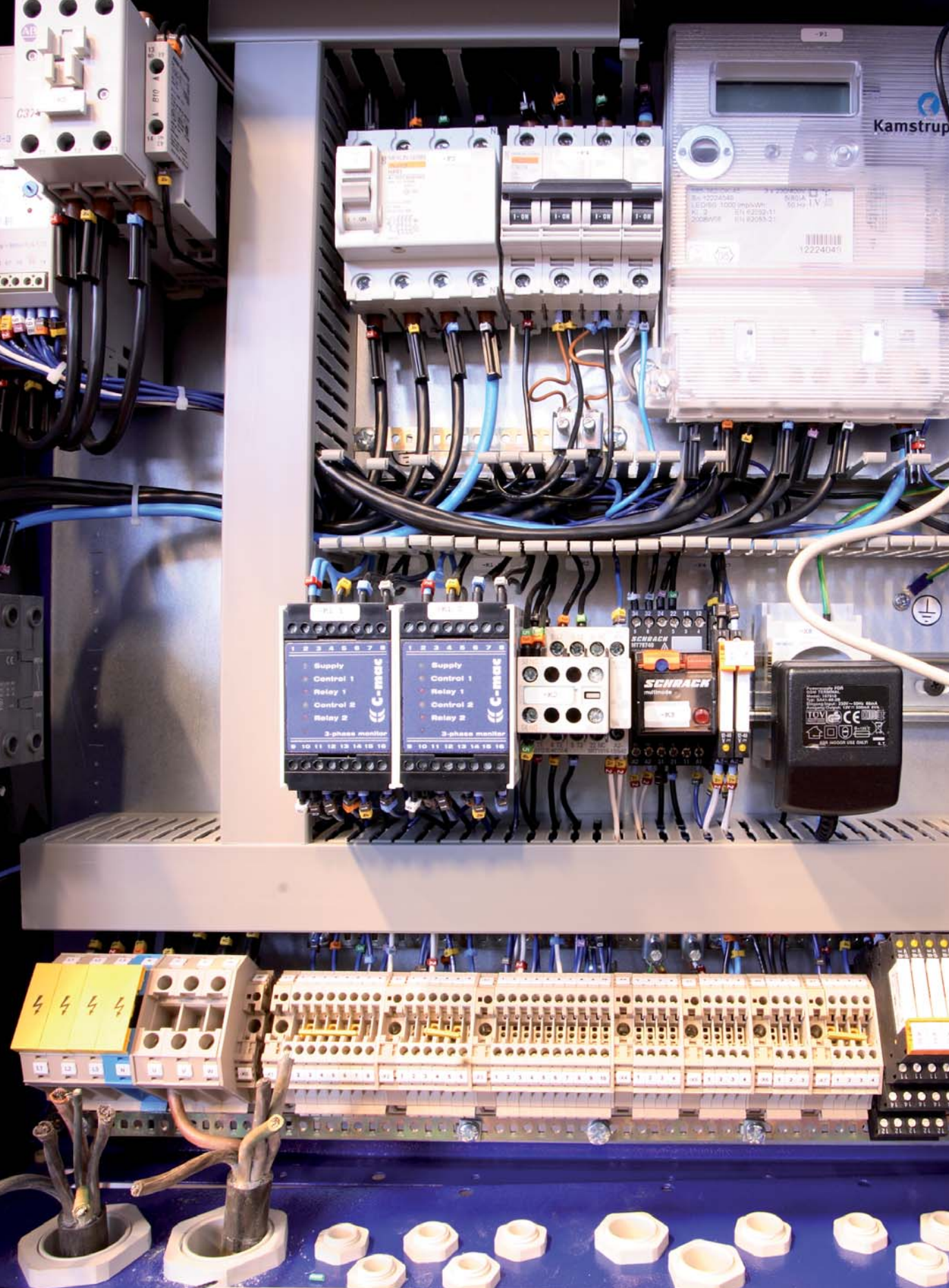
Towards a new grid infrastructure:

Denmark as a full-scale laboratory

Jacob Østergaard

A windy morning in January 2025, Denmark's nearly 5000 wind turbines are running at full speed and generate 6000 MW – enough to cover the electricity demand of your house and the rest of Denmark. Together with 1000 others, your house is a member of a virtual power plant in your region. This system automatically trades the electricity you need on the spot market, where the wind power and generation from other CO_2-neutral sources are sold.

This morning, the market price is relatively low due to the extensive wind power generation, so the electric car in your garage has decided to fully charge its battery, even though it knows that you will only commute the usual 40 km today. Despite the weather forecast, the weather is changing. At 07:00 the wind speed starts to slow, and after 30 minutes wind power generation has dropped to 4000 MW. The reduction in expected generation is reflected in a real-time price signal the system operator broadcasts to all end-users and virtual power plants in the system. The signal indicates that reducing demand to ensure the balance will provide economic benefits. Your electric car, which is now 90% charged, decides to stop charging, and the difference between the real-time price signal and the previously traded spot-market price results in you earning (saving) €3.

At 07:30, you open the refrigerator and are about make your brown-bag lunch. In a flash, as you hardly notice, the buzzing of your electric pantry stops before you close the door. Your refrigerator has just made a small but valuable contribution to preventing a large electricity blackout. A sudden failure in a large transformer station outside the city has resulted in tripping of a 400-MW offshore wind farm. The fault is very severe, because 5 minutes earlier an important transmission cable incidentally was taken out of operation. Back in 2009 this would definitely have led to a major blackout, which would have taken many hours to restore manually and tremendously affected the entire society. But here in 2025, your refrigerator and other intelligent appliances together with some online biomass-fired thermal power plants quickly re-establish the power balance. The sudden lack of generation results in a rapid drop in the frequency of the system, and the electronic controller of your refrigerator detects this. In a split second, the controller decides to disconnect from the grid. So do thousands of other appliances. The drop in frequency is stopped, but the frequency is still low.

To re-establish normal frequency and restart your refrigerator, the system operator adjusts the real-time price signal to activate more responsive demand or generation. Your car in the garage, which has advanced built-in vehicle-to-grid functionality, now starts selling power to the grid by discharging its battery. You have pre-programmed the car, so unless you plan for a long-distance drive, the car is allowed to discharge its battery to 75% if this will save you money. The real-time price is twice the spot-market price now, so you earn another €3. The extra generation in the system brings the frequency back to the normal 50 Hz, and your power system–friendly appliance automatically returns to normal operation after 5 minutes.

The tripping of the large wind farm has resulted in unintentional power flows in the grid, and a sneaking voltage collapse is threatening the system. A new early-warning system based on phasor measurement units has been installed that can automatically detect the approaching voltage collapse. Further, this system evaluates potential actions to avoid the collapse and warns the system operator or, if time is limited, directly activates the most op-timal action. This specific morning, the phasor measurement unit–based system decides to isolate three local subgrids in islanding mode, relieving the pressure on the system. Local controls in the three subgrids take over and ensure uninterrupted operation without any customers noticing (Fig. 12.1).

Intelligent energy solutions

This story could be a simplified picture of the future electric power system in 2025, which is designed and operated to accommodate large shares of renewable energy. New intelligent energy solutions based on information and communication technology and automation will have been introduced. The system will differ substantially from the present grid, which is far more passive and has limited capability to absorb intermittent generation.

The described future paradigm shift of the power system is needed to achieve the political goals for renewable energy. In 2007, the Government of Denmark adopted an ambitious energy policy requiring that renewable energy comprise at least 30% of demand in 2025 and that fossil fuel use be reduced by 15%. This is expected to require that wind energy comprise half of electricity generation in 2025. The political objective is 100% independence of fossil fuels in the long term.

Climate change is not solely on the political agenda in Denmark. The European Union (EU) has agreed on binding targets for reducing CO_2 emissions. Nevertheless, Denmark has chosen to be in the forefront. The 20% share of wind power in Denmark's electric power system and the United Nations Climate Change Conference in 2009 in Copenhagen are visible results of this.

There is consensus in the international energy community that the electrical grid needs to be transformed into an intelligent user-interactive energy infrastructure. The EU has established a European technology platform for the electricity system of the future, called SmartGrids, gathering leading European industry and academia within electric energy. The vision is to enable the future smart grid: an electricity network that can intelligently integrate the actions of all users connected to it – generators, consumers and those that do both – to
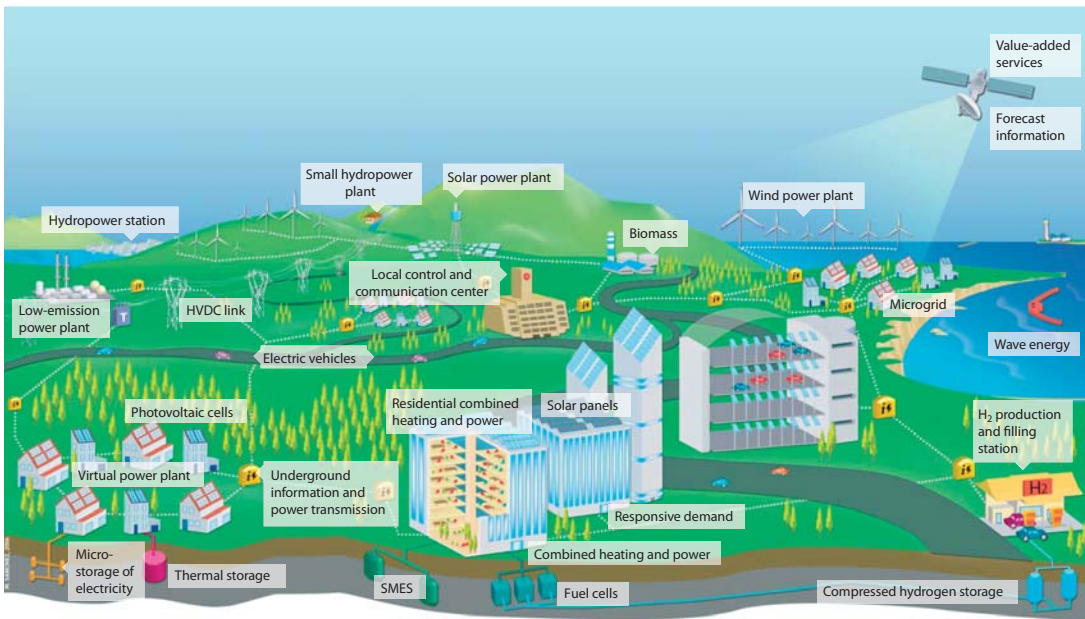
Fig. 12.1. The smart electrical energy system in 2025: wind power and other renewable energy sources cover 50% of the average electricity demand, and intelligent energy technologies based on information and communication technology assist in maintaining the power balance and stability. HVDC: high-voltage direct current. SMES: superconducting magnetic energy storage.

efficiently deliver sustainable, economical and secure electricity supplies (Fig. 12.2).

Denmark's electrical transmission system operator, Energinet.dk, and DTU have taken steps to bring Denmark into a global leading position within intelligent power systems. In October 2004, Eltra (now Energinet.dk) and DTU agreed to establish the Centre for Electric Technology at the DTU Department of Electrical Engineering. Since then, the Centre has been developed into a vital research group with about 40 researchers and a large portfolio of research that is highly relevant for an environmentally sound future. In 2008, the partnerships of the Centre for Electric Technology were extended to include DONG Energy Generation and Siemens Wind Power.

The Centre for Electric Technology aims at developing a more intelligent, flexible and automated electric infrastructure to handle the future expansion of renewable energy on a liberalized energy market maintaining a very high reliability of supply. The Centre has developed a research plan and has decided to focus the research within intelligent

electric power systems, with the three focus areas:
- system integration and grid connection of distributed generation and renewable energy sources;
- new network and control architectures; and
- increased flexibility through end-use demand response and integration with other energy systems.

Some of this ongoing research is introduced here.

Intelligent appliances can improve system stability

In interconnected electricity systems, electricity demand and generation must be balanced in real time. An imbalance is directly reflected in the frequency, which deviates from its normal value of 50 Hz (in Europe). After faults and during changes in generation or demand, the balance must quickly be re-established by activating reserves, typically large thermal power plants that are prepared to change

Fig. 12.2. Global electricity consumption is steadily increasing; here is artificial light seen from space in 2000

their generation. In the Nordic power system, Nordel, the frequency-controlled disturbance reserve must be 1000 MW and must be activated linearly between 49.9 and 49.5 Hz. This reserve is distributed to the different countries according to the size of the dimensioning fault: for example, eastern Denmark must have 100 MW. These reserves are costly (up to €100,000 per MW per year). Further, this type of frequency-activated reserve maintains the normal continuous frequency control. The potential and economy of loads compatible with demand for frequency-controlled reserve have been investigated, several types of control logic for demand for frequency-controlled reserve have been developed, power system impact has been analyzed, potential business models have been designed and an implementation strategy has been suggested. The results show that the demand for frequency-controlled reserve technology is promising from several perspectives.

Technically, demand for frequency-controlled reserve can be used to provide reserves and enhance the control of power system frequency while fulfilling power system requirements such as linear activation according to frequency deviation. Envi-

ronmentally, the demand for frequency-controlled reserve technology is pollution free, in contrast to traditional generation reserves. The cost of such reserves can be low and an attractive business model exists for both society and the involved parties. Given that fluctuating renewable energy is continually being increased in power systems, frequency control will become critical in the future. The demand for frequency-controlled reserve is a novel technology that can facilitate such trends by providing high-quality service at low cost and with zero pollution. If implemented, unique advantages in market competition can be gained to realize the business potential for manufacturers in Denmark. The Centre for Electric Technology has started experimental research involving testing and demonstrating 200 appliances with demand for frequency-controlled reserve together with Vestfrost A/S, Danfoss A/S, Electronic Housekeeper ApS and Ea Energianalyse A/S (Fig. 12.3). Denmark's Energy Technology Development and Demonstration Programme (EUDP) is funding the project.

Early-warning systems and system awareness tools

A phasor measurement unit is a device that provides synchronized measurements of real-time phasors of voltage and current and measures the frequency. A time signal from GPS (Global Positioning System) satellites synchronizes the individual phasor measurement units. Synchronizing the sampling process of the voltage and current, which may be hundreds of kilometers apart, enables measured phasors from different locations in the power system to be put on the same phasor diagram. This could not be done in the past without very expensive atomic clocks.

The accuracy of phasor measurement units is defined in terms of total vector error, which should be within 1%. This implies a phase error within ± 0.01 rad ($\pm 0.57°$) or a maximum time error of ± 26 µs. This high accuracy of the phasor measurement units elevates the standards of power system monitoring, control and protection to a new level.

Researchers and engineers at the DTU Centre for Electric Technology have developed phasor measurement units by combining scientific competencies within electric power engineering, signal processing and computer science. Phasor measurment units have been manufactured and together with Energinet.dk installed in seven locations in Denmark's 132–400 kV transmission systems.

Phasor measurement units most affect power systems operation through applications that use their unique capabilities to provide synchronized measurements at dispersed locations in the grid. Such applications require phasor measurement
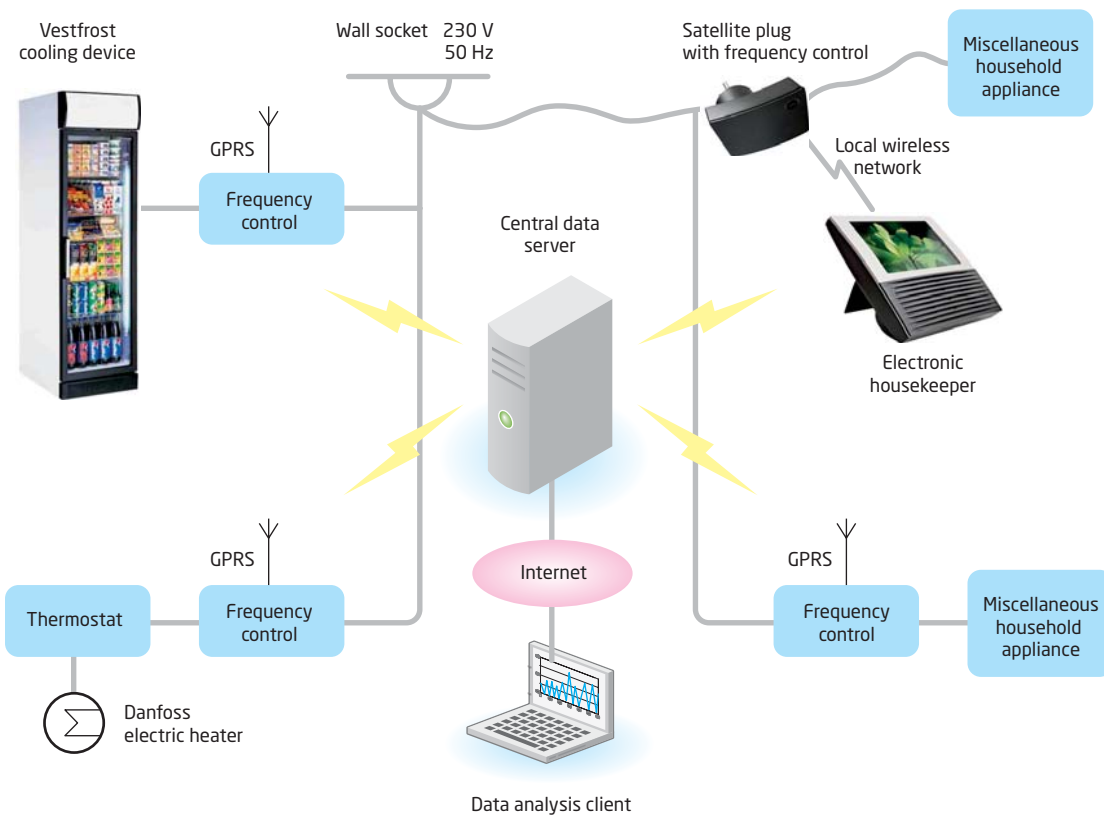
Fig. 12.3 Set-up for demonstrating 200 intelligent appliances supplying a frequency controlled reserve. GPRS: general packet radio service

unit data to be collected at a control center where the data can be analyzed. For research purposes, a server at DTU collects the data from Denmark's phasor measurement units (Fig. 12.4)

In electric power systems in which the proportion of renewable generation has increased significantly, the existing grid is not always designed to cope with the new transmission requirements. In these cases, phasor measurement unit applications can provide intelligent cost-effective monitoring and control solutions that reduce the need for new transmission lines. Ongoing research is taking place at the DTU Centre for Electric Technology in collaboration with major partners regarding online assessment of the power system stability, system awareness tools and blackout prevention systems. This research involves fundamental theory development, algorithm development, advanced simulation, processing of large quantities of data, developing software and experimental verification.

Electric vehicles can help integrate wind power into the grid

An electric vehicle results in less than half the CO_2 emission of a conventional car with an internal combustion engine. Intelligent charging of the car based on the availability of wind power can further reduce or eliminate transport emissions. Electric vehicles also have the potential to play a major role in the economical and reliable operation of the grid with high penetration of renewable energy and they will be an important measure for balancing Denmark's energy system (Fig. 12.5). Analysis by Energinet.dk shows that introducing electric vehicles together with heat pumps can account for about 40% of Denmark's obligations under the EU's 2020 targets regarding: 1) renewable energy penetration, 2) CO_2 emission in sectors not subject to quotas, 3) the share of renewable energy in transport and 4) energy efficiency.

An electric vehicle will be a storage device for
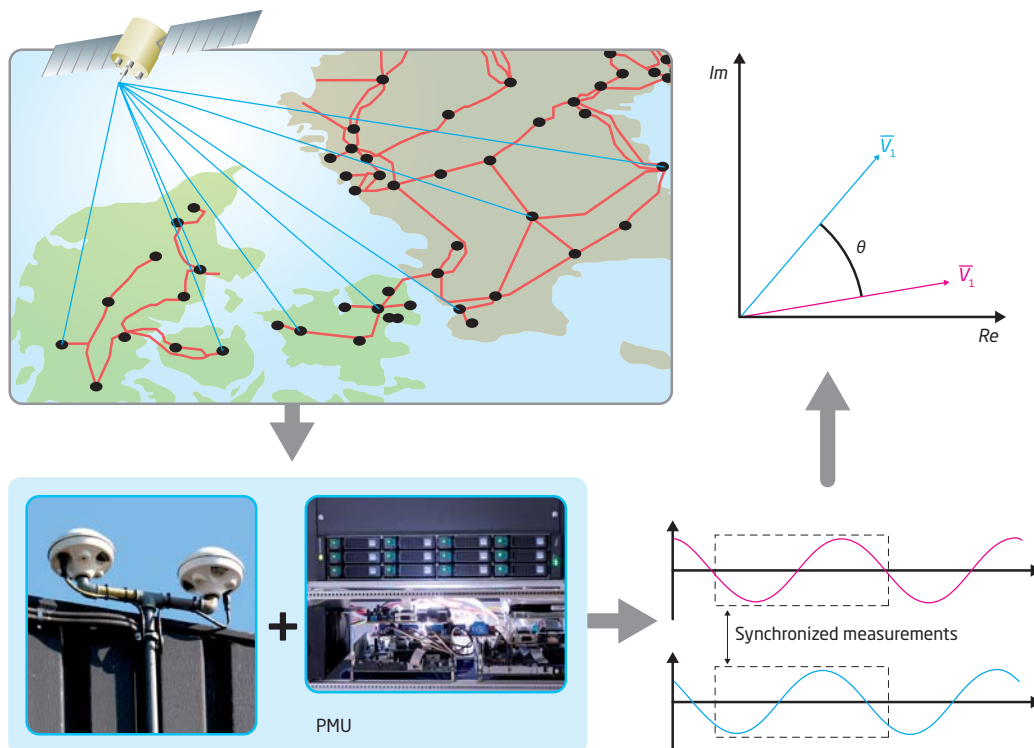
Fig. 12.4. A novel measurement system based on phasor measurement units (PMU) providing measurements in the power system using GPS signals for time synchronization. Differences in the phase angle of the measured 50 Hz voltage signals at various measuring points (illustrated with two measurements – a red one and a blue one) can be determined by the very accurate time stamping and used for advanced supervision, protection and control applications.

Fig. 12.5. Electric vehicles with vehicle-to-grid functionality can dynamically communicate with and exchange power with the electric power system and thereby ensure real-time balancing of the fluctuating wind power. TSO: transmission system operator.

smoothing electric power fluctuation from renewable resources, especially wind power, and provide valuable system services for reliable grid operation. With the proper technology, the cars can run on wind power and enable an increased share of renewable energy for supplying conventional electricity demand and thereby provide an overall economical, reliable and sustainable energy system.

Denmark does not have a car industry, and Denmark's background for developing electric vehicles themselves is limited. Nevertheless, Danish companies and research institutions have very strong knowledge and competencies regarding designing, developing and operating power systems with high penetration of distributed generation. Further, Danish industry is involved in technologies critical to the widespread use of electric vehicles such as information and communication technology and control systems for charging and discharging batteries. This forms an ideal base for developing systems and integration solutions for electric vehicles (Fig. 12.6).

Denmark's competencies can be used to develop optimal system solutions for integrating electric vehicles with the electrical system, including network issues, market solutions and optimal interaction between different energy technologies. Further, Denmark's electric power system provides an excellent platform for demonstrating developed solutions, thereby providing the commercial basis for exporting Denmark's technology. Further, the advantage of being a first mover constitutes a business advantage and enhances the opportunity for strongly influencing future standards for system integration of electric vehicles to achieve optimal

utilization of the electric vehicles in the power system.

In spring 2009, the Edison project was started. The project is the world's largest electric vehicle project, focusing on intelligent integration of electric vehicles with a power system based on renewable energy. The consortium partners in the project are the Danish Energy Association, DONG Energy, DTU Department of Electrical Engineering, DTU Department of Transport, DTU Department of Informatics and Mathematical Modelling, EURISCO ApS, IBM, Risø DTU National Laboratory for Sustainable Energy, Siemens and Østkraft (Bornholm's distribution system operator).

Fig. 12.6. Students and employees from the DTU Centre for Electric Technology examine the Tesla Roadster (pure electric, 0–100 km/h in 4 s, 250 hp, >200 km/h, 360 km per charge). Today electric cars can be compared with conventional cars based on internal combustion engines.

Bornholm as a full-scale laboratory for intelligent energy

From an international perspective, Denmark's electric energy system is unique, with 20% penetration of wind energy, and during selected hours, wind power alone exceeds the demand. Further, Denmark's faces steps no country has taken: developing a system that can manage wind power and distributed energy resources of the planned future proportions. Denmark is probably facing some of the greatest technology developments in its power system so far, in particular because the system changes are expected to involve major implementation of modern information and communication technology that has not been used in the power sector before.

Denmark's unique power system provides very good opportunities for developing, testing and demonstrating new innovative solutions. Combining advanced experimental facilities at the universities with a full-scale true test-bed can provide a quantum leap for developing future low-carbon electric energy systems, thereby strengthening the energy sector in Denmark in general and poten-

tially putting Denmark in the forefront of the global market for intelligent energy solutions.

The ongoing research at the DTU Centre for Electric Technology benefits from Denmark's unique electrical system. Research is being carried out in collaboration with Østkraft, Bornholm's distribution system operator. Bornholm corresponds to about 1% of Denmark in land area, load and population. The island has a representative distribution grid with 30% wind power penetration, more than 27,000 customers and a peak load of 55 MW. The system can be isolated from the interconnected Nordic grid, and a power system with truly high wind power penetration can be studied. This makes Bornholm the perfect case for research, development and demonstration of new technology.

When the Bornholm grid is electrically isolated from the larger grid, frequency stability can become problematic, and today most larger wind turbines have to be shut down, even during such well-planned operations. The current research focuses on some of the problems seen in Bornholm related to high wind power penetration and includes developing islanding capability, using phasor measurement units, demonstrating electric vehicles, developing coordinated frequency control by wind turbines and experimenting with demand as frequency-controlled reserve. The ambition is to carry out research, development and testing within about 5–8 years leading to full-scale demonstration of the future intelligent electric energy systems of 2025 described previously.

Conclusion

A huge research task is ahead in the coming years. The future intelligent user-interactive electricity system has to be developed to efficiently deliver sustainable, economical and secure energy and fulfill the political ambitions regarding CO_2 emission. New technologies have to be introduced and new

28,000	Customers
55 MW	Peak load
16	60/10 kV substations
	Power plants:
39 MV	Diesel (oil)
37 MV	Steam (oil)
37 MV	Steam (oil or coal, combined heat and power)
30 MV	Wind (wind)
2 MV	Biogas
4	District heating systems
6	Heat generators
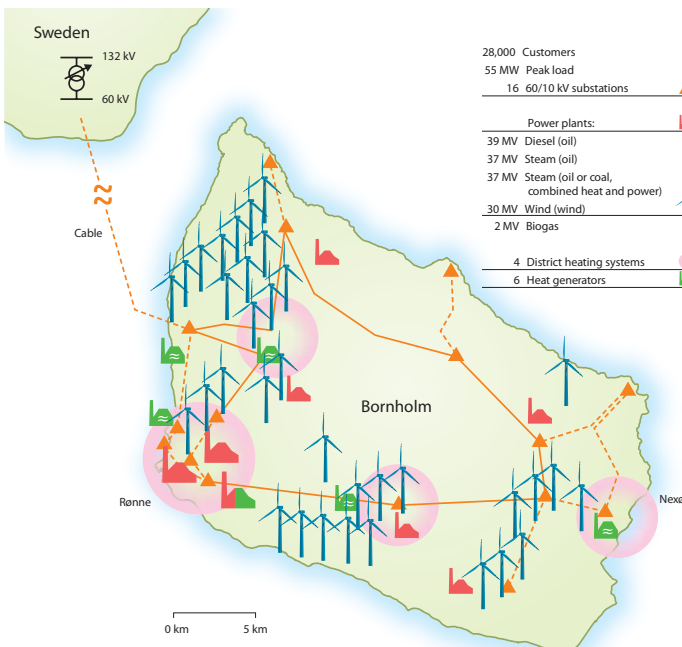
Fig. 12.7. Bornholm's power system, with more than 30% wind power penetration and unique potential for isolated operation in island mode, makes it a terrific full-scale laboratory for future intelligent power systems.

system architectures need to be developed. A new architecture will be based on subgrids with increased integrated control of the distribution and transmission systems and active network functions at the distribution level, enabling participation in the electricity market and delivery of services from distributed demand, generation and storage resources. Developing coherent and consistent control architectures will be the prerequisite for releasing the potential of many of the resources in the distribution networks and for an overall economically optimal system design.

Denmark has the opportunity to be a global front-runner in this development. Several new technologies and solutions for the future intelligent grid are already on their way based on the ongoing research. Denmark's unique grid with a high share of wind power constitutes a unique full-scale laboratory facilitating the future technology development – it provides a unique opportunity for Denmark to show the way and provide the global solutions for the long-term goal of 100% independence from fossil fuels.

The author

Jacob Østergaard: MSc, DTU, 2005. Professor, Head, Centre for Electric Technology, DTU Department of Electrical Engineering, 2005–. Researcher, Danish Energy Association, 1995–2005. Research fields: electric power systems, wind power integration, active demand.

More to explore

Vision and strategy for Europe's electricity networks of the future. Brussels, European Commission, 2006.

Xu Z, Gordon M, Lind M, Østergaard J. Toward a Danish power system with 50% wind – SmartGrids activities in Denmark. In: *Proceedings of IEEE PES General Meeting, Calgary, Alberta, Canada, 2009*. Paris, IEEE, 2006.

Chen Y, Xu Z, Østergaard J. Frequency analysis for planned islanding operation in the Danish distribution system – Bornholm. In: *Proceedings of the 43rd International Universities Power Engineering Conference*. Padua, UPEC2008 Secretariat, 2008 (www.upec2008. org/files/summaries/1_271_summary.pdf).

Saleem A, Lind M. Multiagent intelligent control system for subgrid operation of distributed networks with large amount of distributed generation. In: *Proceedings of the IEEE PowerTech 2009, Bucharest, Romania, 2009*. Paris, IEEE, 2009.

Xu Z, Togeby M, Østergaard J. *Demand as frequency controlled reserve: final report of the PSO project*. Kgs. Lyngby, DTU, 2008.

Steps toward a power system with 50% wind energy: EcoGrid phase 1. Kgs. Lyngby, DTU, 2009.
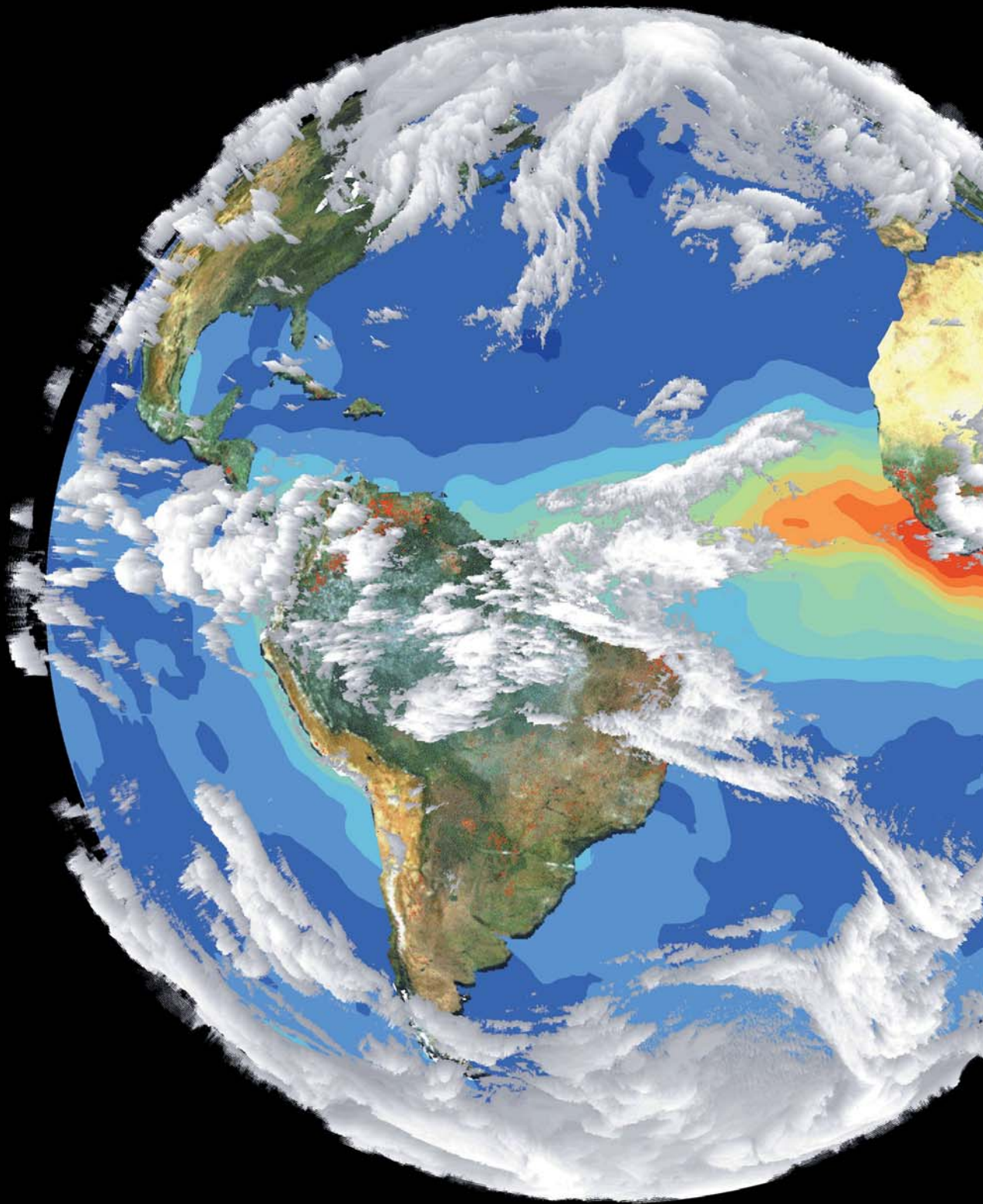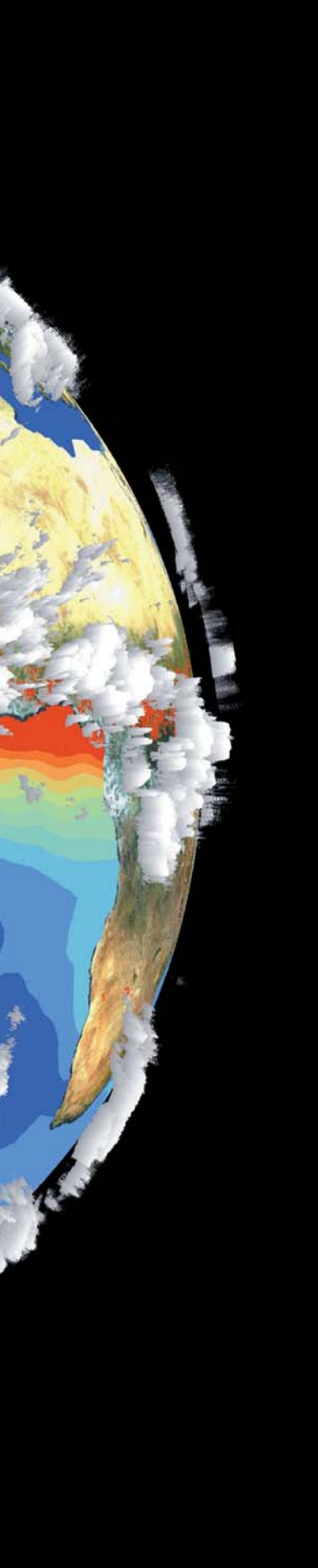
Changing
energy
systems
to prevent climate change

Kirsten Halsnæs & Hans Larsen

This chapter discusses how future energy systems can be changed rapidly and economically efficiently as needed to prevent global temperature increases from exceeding 2°C. We assess the magnitude of the effort required to reduce the emissions of greenhouse gases for different parts of the energy sector and indicate that meeting such a temperature stabilization target requires more rapid and ambitious development and deployment of low-carbon energy technology than the current level. Accordingly, we recommend which technologies could potentially contribute to meeting the challenge of reducing global greenhouse-gas emissions in the short term and long term: 2050, when annual greenhouse-gas emissions should be reduced significantly. We also assess the costs and other requirements of enforcing rapid deployment of low-carbon technologies.

Background

The Fourth Assessment Report (Synthesis Report) of the Intergovernmental Panel on Climate Change (IPCC) concluded that "Warming of the climate system is unequivocal, as is now evident from observations of increases in global average air and ocean temperatures, widespread melting of snow and ice and rising global average sea level." Based on this conclusion, the IPCC suggested that avoiding dangerous anthropogenic influence on climate change requires stabilizing the global atmospheric concentrations of greenhouse gases over the next centuries. The IPCC discussed the climatic consequences of alternative stabilization levels, and observations of current changes in ecosystems and human livelihoods together with integrated assessment of the Earth system and the biosphere have led to the conclusion that global average temperature increases above 2°C to 3°C in this century could pose high risk in terms of increased water stress, coral bleaching, loss of biodiversity, health effects and flooding. This chapter focuses on the challenge of limiting global temperature changes to below this boundary risk condition.

Several international energy economics models have assessed mitigation potential, technical options and costs corresponding to various climate stabilization scenarios. We review some of these results as a background for more detailed discussion about how greenhouse-gas emissions can be sub-stantially reduced during the next 30 to 50 years by using new low-carbon technologies. We especially emphasize energy supply technologies, transport and the building sector.

The stabilization challenge

Limiting global temperature changes to 2–3°C will require very large and rapid reductions in green-house-gas emissions (Fig. 13.1).

Fig. 13.1 shows (in colored banners) the ranges of global scenarios for greenhouse-gas emissions that give a 50% probability of meeting a given temperature stabilization target. Meeting 3°C cor-responds to what is called category III, and this target requires that global emissions decease after 2020 and that emissions in 2050 be 30–60% below the 2000 level. Correspondingly, meeting a 2°C target, as shown in Fig. 13.1, requires that global emissions decrease by 2015 and that emissions in 2050 be 50–85% lower than in 2000.

Meeting these targets is a huge challenge for the energy sector, and even more so since global emissions are currently rising rapidly and there is no indication that this will be reversed soon in terms of reaching a tipping point. Many initiatives to mitigate climate change are being taken and plans are being developed in individual countries and regions, but the results in significantly reduc-ing greenhouse-gas emissions have yet to be seen.
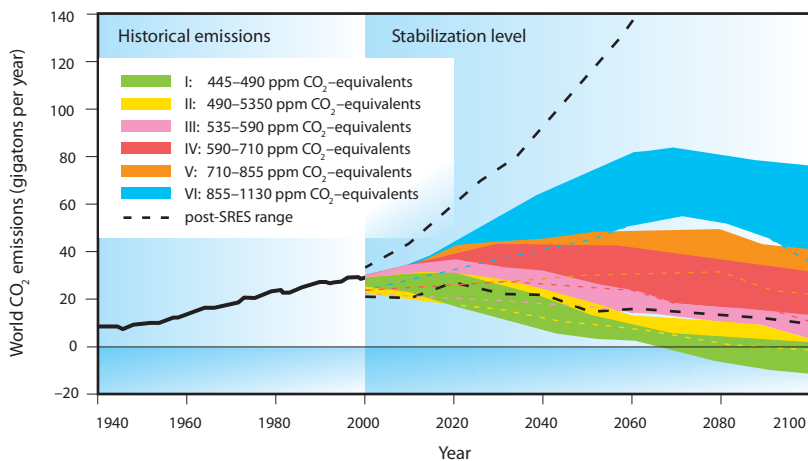
Fig. 13.1. Stabilization pathways and global greenhouse-gas emissions corresponding to alternative global warming levels, 2000–2100. SRES: based on the IPCC *Special report on emissions scenarios.*
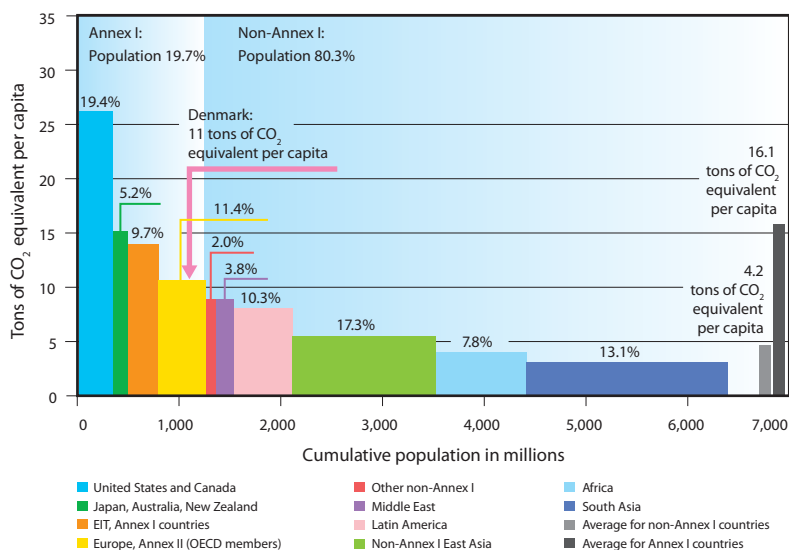
Fig. 13.2. Global CO$_2$-equivalent emissions per capita in global regions and percentage of total emissions accounted for by each region. EIT: countries with economies in transition, including the Russian Federation, the Baltic countries and several other countries in central and eastern Europe. Source: IPCC.

Transforming the energy systems of the world into systems that are rapidly reducing greenhouse-gas emissions is a tremendous challenge and even more so, as this has to go hand in hand with securing energy supply for future economic development and giving affordable energy to the 2 billion people without access to modern energy today.

Greenhouse-gas emissions vary between global regions due to differences in economic development and the carbon intensity of current energy systems. Fig. 13.2 shows the greenhouse-gas emissions per capita in different parts of the world. It is striking that 80% of the world population lives in the non-Annex I countries (mostly low-income countries), but the greenhouse-gas emissions from this group only comprise about 50% of the global emissions. Global climate change policies therefore need to include a strong element of burden sharing which, for example, in relation to energy systems, can imply transferring new efficient technologies to low- and medium-income countries and supporting the implementation of climate-friendly technologies by financial mechanisms.

Mitigation potential

The *Report on mitigation policies* of the IPCC Fourth Assessment Report assessed in great detail mitigation studies for energy supply, transport, buildings, industry, agriculture, forestry and waste. The studies were based on sectoral technology-oriented modeling and assessment. The IPCC used these studies to generate an overview of mitigation potential and costs covering a period until 2030 with a cut-off cost level of US$100 per ton of CO$_2$ equivalent (Fig. 13.3).

IPCC concludes that energy supply and the transport sector will contribute greatly to future global greenhouse-gas emissions, and this will be a very strong obstacle to achieving stabilization scenarios at low temperature levels. Meeting a global stabilization scenario of 2°C, for example, could imply that total global emissions in 2030 should not exceed about 18 million tons of CO$_2$ equivalent, and meeting a 3°C scenario could imply that emissions should below about 36 million tons of CO$_2$ equivalent. In this way, the 3°C scenario could be met with the IPCC mitigation potential assessed in the Fourth Assessment Report that brings the level down to about 33 million tons of CO$_2$ equiva-

Fig. 13.3. Mitigation potential for 2030 assessed by the IPCC Fourth Assessment Report, Working Group III for a cost below US$100 per ton of CO_2 equivalents. The light green part at the bottom of the figure shows total greenhouse-gas emissions remaining after the full mitigation potential identified in the sectors is implemented. The upper part of the figure then shows the mitigation potential as shaded areas. Adding the remaining greenhouse-gas emissions and the mitigation potential shows that the total emissions in the baseline case are projected to increase by about 22% from 2004 to 2030. Implementing all sectoral mitigation options that cost less than US$100 per ton of CO_2 equivalent would reduce greenhouse-gas emissions by about 45% in 2030. Relatively large reductions in greenhouse-gas emissions have been assessed for the building sector and industry, and electricity savings are included in the end-use sectors buildings and industry, which expands these sectors' mitigation potential. Energy supply and transport, in contrast, have very small reductions.

lent in 2030, but the 2°C scenario requires greater efforts. The actual mitigation potential that was assessed only covers some of the options that have been assessed in the literature; more options exist for a higher cost than the US$100 cut-off level the Fourth Assessment Report used as a benchmark for the period until 2030.

Technological options for reducing greenhouse-gas emissions in energy supply and end-use in buildings and transport

Large-scale reductions in greenhouse-gas emissions can be pursued through a variety of options. One option is to change the energy system and base it on renewable energy technologies such as wind, solar and biofuels with no or very low emissions and to combine this with improving the efficiency of the fossil fuel–based supply technologies used today. Another option is to improve con-

version efficiency and to develop new low-energy end-use technologies for private households, transport and industry. A third option is to upgrade the energy transmission and distribution systems with smart grids and distributed production to improve overall efficiency significantly. As *Risø Energy Report 7: future low carbon energy systems* from the Risø DTU National Laboratory for Sustainable Energy states, no single solution is deemed sufficient to meet this huge challenge, so all these climate-friendly solutions must be implemented jointly.

A major challenge is to ensure that total global greenhouse-gas emissions peak within 10–15 years from now and then subsequently decline rapidly (Fig. 13.1 further illustrates scenarios for greenhouse-gas emissions). This rapid turnaround has to be tackled using today's technologies or technologies that are almost ready to enter the market. This has to be followed up by developing energy systems that offer sustained low greenhouse-gas emis-

sions in the future. Some options have very low costs of less than US$20 per ton of CO_2 equivalent, and some options even have negative costs because they improve efficiency and thereby save more in fuel cost than the capital investment. Other options are currently expected to cost more than US$100 per ton of CO_2 equivalent, but this cost might be reduced in the future by large-scale penetration of new technologies, research and development and innovation.

Fossil fuels still supply 80% of global energy supply, and changing this significantly will take decades. Various renewable energy-supply technologies can contribute significantly: wind, solar, bioenergy, geothermal energy and ocean energy, including wave technologies.

Wind energy comprises about 20% of Denmark's electricity production, and this share is expected to grow to 50% in the coming decades. Wind energy, however, is only a few percent of global energy supply but is growing about 15% per year. According to IPCC estimates, the cost per ton of CO_2-equivalent reduction for wind power is about <US$0 to US$50 per ton, spanning from fully competitive to negative costs, depending on local conditions. Hence, the prospects are good for wind energy contributing large scale to world electricity production.

Photovoltaics is a mature technology for electricity production but contributes very little to global electricity supply today due to relatively high costs. The IPCC estimates the cost per ton of CO_2 reduction is about US$50–100. The technology is developing rapidly, and the annual growth in deployment of the technology is about 40%.

Large-scale hydroelectric energy and geothermal energy are mature technologies already used today. The cost per ton of CO_2 reduction varies with local conditions, but the technology has the potential for wider use, as some projects have low CO_2-reduction costs.

Ocean energy, including wave energy, is still at the research and development and demonstration stages, and the technologies are estimated to have limited global potential.

Nuclear fission energy is a mature competitive technology with very low CO_2 emissions; the limiting factor for further deployment is political in terms of lack of public acceptance in many countries.

New sustainable energy supply technologies with low greenhouse-gas emissions that could enter the market in the middle of the century are hydrogen-fueled technologies and fusion. Hydrogen-fueled technologies are at the research and development and demonstration stages today, and the major challenge is developing technologies for sustainable hydrogen production by means of renewable energy sources. Fusion technology is being pursued through international cooperation on the design and construction of the large fusion reactor ITER, which is projected to start operation in 2016. The first commercial fusion reactor is estimated to be in operation in 2050.

Two end-use sectors attract great attention in relation to reducing greenhouse-gas emissions: the building and transport sectors.

The building sector has huge potential for reducing greenhouse-gas emissions at very low cost – in some cases even at negative costs. The technologies needed to harvest this potential are largely already available. An integrated part of future low-energy houses might be technologies to produce part or all of the electricity needed for the house, such as solar collectors or small fuel cells.

The transport sector also attracts great attention; the challenge, however, is much greater than in other parts of the energy system. Transport today depends on fossil fuel almost 100%, and the technologies to change this are not available at competitive costs. One possibility is to introduce biofuels, but the technologies, including non-food second-generation technologies, are still at the research and development and demonstration stages. Biofuels might be used in cars, ships and airplanes. The other possibility is to electrify transport. This option is relevant for cars and railways but not for air and sea transport. The limiting factors for electrifying road transport – cars – are costs and range: developing inexpensive, light and efficient batteries. Further, electric cars only become sustainable when electricity production is sustainable.

The future energy system will link the various distribution systems such as the electricity, gas and heating networks. The electricity system will comprise centralized production units and numerous decentralized units that will largely be based on renewable energy technologies.

The grids will be developed into smart grids op-

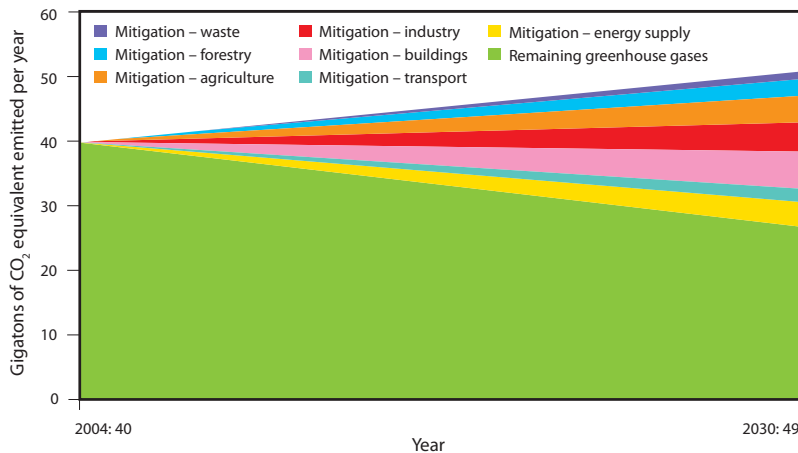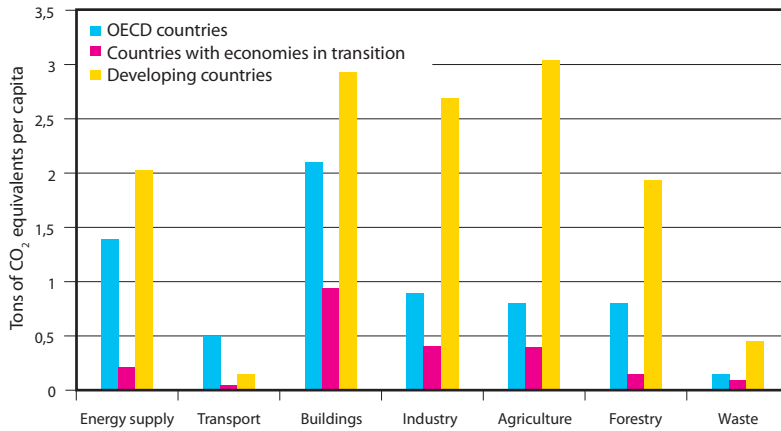Fig. 13.4. Mitigation potential by sector and region in 2030 as assessed by the IPCC Fourth Assessment Report, Working Group III for a cost below US$100 per ton of CO_2 equivalent

erated using advanced information technology and communication equipment, which will make increasing the proportion of intermittent renewable energy sources in the system easier and thereby contribute to reducing greenhouse-gas emissions.

An electrified transport sector, including electric vehicles, may become an integrated part of such an energy system, where the car batteries can be charged when inexpensive surplus electricity is available, and the batteries are then used as reserve capacity in high-demand situations. The expected future trend in the building sector towards low-energy houses or even energy-plus houses will similarly make them suitable for flexible energy supply and use managed by smart communication systems.

The future energy system could thus emit much less greenhouse gas than today's system, and this could be achieved by combining numerous advanced solutions, including supply-side options and end-use efficiency, and by managing the total system through a smart grid with advanced communication systems (Box 13.1).

Developing countries

An important complexity in meeting the global stabilization challenges will be to facilitate a process of technology transfer to developing countries that can integrate them into global mitigation efforts. Fig. 13.4 shows the mitigation potential in different global regions of the world in accordance with the results shown in Fig. 13.3.

As Fig. 13.4 illustrates, the mitigation potential is very large in developing countries compared with other regions for all sectors except transport. Relative to OECD countries, industry, agriculture and forestry offer especially large low-cost mitigation potential in developing countries. The mitigation potential of the transport sector in developing countries is relatively small largely because of the low level of development of this sector. An example is India, whose 1.1 billion people only have about 10 million private cars. This development level certainly is far from saturation, so a large mitigation potential must be expected to emerge over time in developing countries.

Implementing options for reducing greenhouse-gas emissions in developing countries requires that mitigation policies and national sustainable development goals be aligned. The Development and Climate Project studied the potential for such combined policies. The approach was to use several indicators to measure how specific energy policies from 2000 to 2030 simultaneously could meet energy policy goals, sustainable development and low greenhouse-gas emission trajectories. The studies included an assessment for India that used

Box 13.1. Prospects for cleaner energy technologies

Wind energy has good prospects to provide a large share of global electricity generation. Worldwide, 27 GW of new wind energy capacity was installed in 2008, bringing the total installed capacity connected to the grid to 121 GW.

Photovoltaic technology is developing rapidly, but increasing its share of global electricity generation requires reducing costs. Only 5.4 GW of new photovoltaic capacity was installed globally in 2008, increasing the total installed photovoltaic capacity connected to the grid to 13 GW.

Large- and small-scale hydroelectric energy are mature technologies already used today. Globally, 25–30 GW of new large-scale hydroelectric capacity was installed in 2008, bringing the total to 860 GW. The equivalent figures for small-scale hydroelectric energy are 6–8 GW in new capacity and 85 GW total.

Geothermal energy is a mature technology used for electricity production but at a low level. Only 10 GW was installed at the end of 2008. However, the installed capacity for heat production was 50 GW.

Nuclear fission energy is a mature and competitive technology. The limiting factor for further deployment is political in terms of public acceptance in many countries.

Ocean energy (including wave energy), hydrogen-fueled technologies and fusion energy are still at the research and development level.

the energy and gross domestic product intensities of greenhouse-gas emissions and local pollutants, investment, electricity access, efficiency of electricity generation and renewable energy as indicators of joint policy goals that could be considered as important representative indicators of India's sustainable development policy goals. Fig. 13.5 shows the development of these indicators for India.

Fig. 13.5 illustrates that the energy system of India is expected to offer decreasing CO_2 and energy intensity of the economy as well as decreasing local pollution in terms of SO_2 emissions, but this does not imply that greenhouse-gas emissions are expected to decrease absolutely during this period. Per capita electricity consumption and investment in new power plants increase rapidly during this
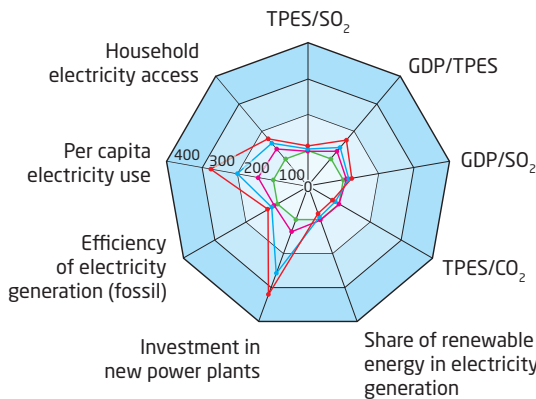
Fig. 13.5. Indicators for sustainable development, energy and climate change in India for 2000–2030. The diagram is structured such that outward movement in the indicators over time represents improvement in the sustainable development of India's energy system.

period in accordance with national development goals, and it is expected that as much as US$1.2 trillion will be invested in energy supply infrastructure from 2006 to 2030. More climate-focused policies are needed to integrate India better into international climate policies.

The mitigation options in developing countries such as India are largely the same as those in industrialized countries. But the time frame for policy-making and priorities as well as the opportunities differ somewhat. India is expected to use its huge coal reserves in new coal-fired power plants – these could be developed into clean coal projects including carbon capture and storage in Clean Development Mechanism projects. Renewable energy could likewise be pursued in international collaboration – such as collaboration on wind energy between Denmark and India. With regard to end-use technologies, leapfrogging to low-energy solutions would be attractive in accordance with what India can produce and export to industrialized countries. In general, a country with rapid growth in the economy and energy sector development has huge opportunities to change the systems to modern standards without insurmountable additional costs.

Conclusions

Future global energy systems must be changed rapidly to limit global temperature changes. Limiting global temperature changes to 2–3°C will require very large and rapid reductions in greenhouse-gas emissions. If a 3°C global temperature increase is considered acceptable, there is some time to make the necessary changes. However, if only a 2°C global temperature increase is acceptable, time is short and action is needed now and not only tomorrow. Many mitigation options are available, and no single option can solve the problem. On the contrary, all options need to be implemented, both new energy-supply options and major changes in the future energy system and in end-use technologies, especially in transport and building. Further, new ambitious policies and measures are needed to facilitate the transfer and funding of clean energy technology deployment in developing countries, and the establishment of such mechanisms is one of the key issues being negotiated for the United Nations Climate Change Conference (COP15) in Copenhagen in December 2009.

More to explore

Core Writing Team, Pachauri RK, Reisinger A, eds. *Climate change 2007: synthesis report. Contribution of Working Groups I, II and II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Geneva, Intergovernmental Panel on Climate Change, 2008 (www.ipcc.ch/ipccreports/ar4-syr.htm).

Metz B, Davidson O, Bosch P, Dave R, Meyer L, eds. *Mitigation of climate change. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, Cambridge University Press, 2008 (www.ipcc.ch/ipccreports/ar4-wg3.htm).

Larsen H, Petersen LS, eds. *Risø energy report 7: future low carbon energy systems*. Risø, Risø DTU National Laboratory for Sustainable Energy, 2008 (www.risoe.dtu.dk/Risoe_dk/Home/Knowledge_base/publications/Reports/ris-r-1651.aspx).

Renewables global status report: 2009 update. Paris, Renewable Energy Policy Network for the 21st Century (REN Secretariat), 2009 (http://www.ren21.net/pdf/RE_GSR_2009_Update.pdf).

The authors

Kirsten Halsnæs:
MA, Economics, University of Copenhagen, 1983; PhD, Economics, Roskilde University, 1991. Head, DTU Climate Center. More than 15 years of experience in studies of climate change mitigation and adaptation. Convening lead author, Intergovernmental Panel on Climate Change. Author of several international papers and books about climate change, energy, sustainable development and international policies.

Hans Larsen:
PhD, DTU, 1973. Head, Systems Analysis Division, Risø DTU National Laboratory for Sustainable Energy. Many years of experience in energy systems analysis and climate change technologies. Member, Danish Academy of Technical Sciences, World Energy Council, Programme Committee 2004–. Bureau Member, Intergovernmental Panel on Climate Change, 2003–2007.

Characterization of the fish communities present in waters near Denmark (eastern-northern North Sea, Skagerrak, Kattegat, Belt Sea and western Baltic) during cold (left side) and warm (right side) conditions. All species shown were caught in waters near Denmark in the late 1990s and early 2000s, but the abundance of the warm-adapted species has increased and is expected to increase further as climate change progresses in the 21st century, while the abundance of the colder-adapted species is expected to decline. The species shown are herring, cod, saithe, sprat, plaice, red mullet, anchovy, sardine, gilthead seabream, European seabass, swordfish, bullet tuna, common sole and bogue. Content: Brander et al. (*ICES Marine Science Symposium* 2003: 219: 260–273) and MacKenzie et al. (*Global Change Biology* 2007: 13: 1348-1367).

Can we sustain fisheries as the climate changes?

Brian R. MacKenzie & Keith Brander

lobal production of fish and other marine products provides a vital part of human food supply and is essential to the livelihoods of many people. Total annual world landings from marine capture fisheries (based on "wild" populations of fish) rose steadily until about 1990 and have since remained at around 70 million tons (Fig. 14.1). Aquaculture production (freshwater and marine) has been rising steeply and is expected to equal capture production by about 2030. There are serious concerns that global fish production will be difficult to sustain in the future, and the additional threat posed by climate change is being assessed to anticipate effects and to develop adaptive strategies.

Questions and applications

How will changes in temperature, pH, salinity, oxygen and nutrient supply affect marine primary production? Will marine ecosystems alter their composition and properties to different states? How does biodiversity affect the resilience and productivity of marine systems? Will the geographical distributions of fish and other marketable taxa shift? How should the management of fisheries and of

marine ecosystems respond to changes due to climate? The processes and interactions between fisheries and climate change are far from completely understood. Nevertheless, urgent action is needed to avert further serious harm. These complex issues transcend disciplinary and international boundaries. Extensive international collaboration in science and policy-making is therefore fundamental to much of our work. A large proportion of the DTU work related to fisheries is carried out for organizations such as the International Council for the Exploration of the Sea and the Intergovernmental Panel on Climate Change. Our work is applied in European fisheries management (such as how to design and implement the new European Marine Strategy) and supports the development of agreements on sustainable fisheries and policy on climate adaptation and mitigation.

Denmark's location in the Baltic region provides a natural focus for research in a region that is particularly sensitive to certain aspects of climate change and has particularly rich records for reconstructing the effects of past climate change. The Baltic Sea is only a few thousand years old and has far fewer fish species than the adjacent North Sea; only species that can tolerate low salinity survive
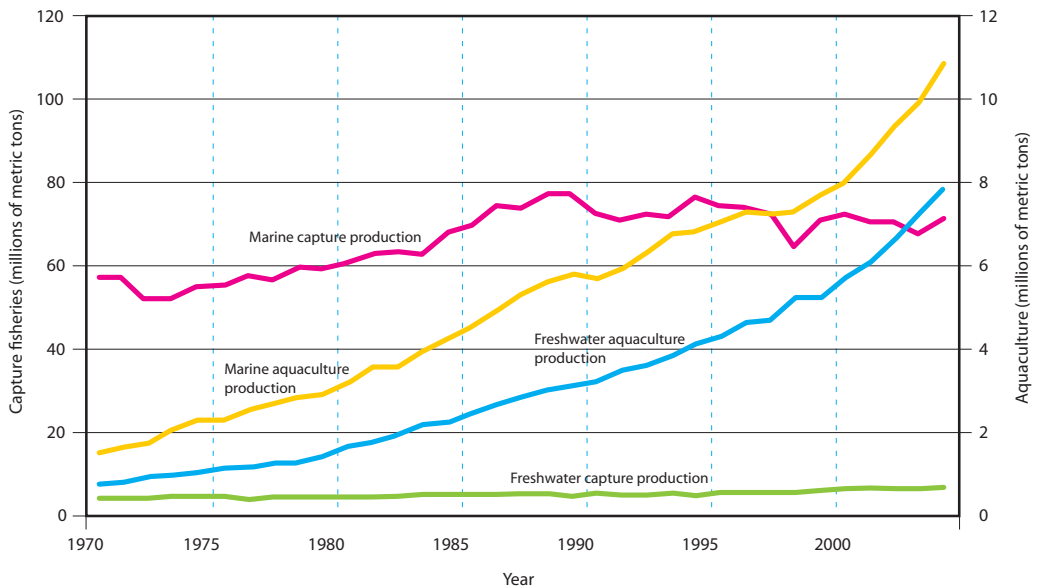

Fig. 14.1. Trends in world capture fisheries and aquaculture production, not including problematic data from China. Source: data from Food and Agriculture Organization of the United Nations (2007).

and thrive there. Climate change will probably increase the input of fresh water and reduce the inflow of saline water from the North Sea, causing the salinity to decline. The lower number of species and the probability of rapid decline in salinity make the Baltic a valuable test case for studying the effects of climate change on fisheries and how to respond to them.

People harvest fish species from wild populations but do not have the same control over their growth and reproduction that they do with farmed terrestrial animals. The main scientific method for understanding processes and measuring the effects of factors such as climate is to carry out experiments, but the ability to carry out experiments on marine species and ecosystems is very limited. The main alternative methods iterate progressively between developing an understanding of key processes, retrospective analysis, modeling and testing hypotheses with existing datasets. The following sections describe what can be learned from the past, the present understanding of ecological processes governing fisheries production, particularly in the Baltic Sea, and how the future effects of climate can be predicted for purposes of management and policy.

Historic changes in climate and their effects

Predicting the future of marine ecosystems under climate change is a daunting task, given their complexity and the interacting effects of many drivers, including changes in temperature, salinity, oxygen, pH, vertical mixing and nutrient supply. Although such predictions are being made, their credibility is rarely evaluated and is generally likely to be low. Studies of periods in the past when conditions resembled those we are trying to predict provide an alternative form of prediction by analogy. However, their utility as predictors of the future is restricted, especially because the choices society makes concerning pollution, agriculture, land use, fisheries, marine protected areas and so on will also alter the future composition, productivity and resilience of marine ecosystems.

What can be learned from studies of the past? A joint fisheries and archaeological study with a colleague from the Natural History Museum of Denmark investigated the species composition of fishbone remains in human settlements in Denmark

from the Stone Age Atlantic Warm Period (about 7000–3900 BC). This period was characterized by sea temperatures 2–3°C higher than late 20th-century temperatures and therefore similar to the temperatures forecast for northern Europe in 2100. The remains included bones of anchovy, red mullets, swordfish, European seabass and black seabream, which are all species typical of warmer waters than those observed near Denmark until recently, but many are now returning to Denmark's waters (see illustration on pages 148–149).

The fishbone remains included thousands of cod bones. This is surprising, because recruitment of North Sea cod declines when sea temperature increases. However, fishing pressure was low in the Stone Age, and the negative effect of higher temperature on cod was partly offset by the lower fishing mortality. Cod fisheries in the North Sea might be able to continue throughout the 21st century despite increasing temperatures, but this requires achieving lower fishing mortality for cod.

The history of the Baltic marine ecosystem during the 20th century can be regarded as a large-scale experiment in ecological engineering as human activity eradicated marine mammals and added nutrients, which increased production and increased fishing pressure (Fig. 14.2). In addition, major variation in hydrographic conditions (sea ice, temperature and salinity) can be associated with large-scale climatic variability over the northern Atlantic region. The long time series of data for Atlantic cod (*Gadus morhua*) in the Baltic Sea and the large variability in the stock size and in forcing factors provide a unique opportunity to investigate the relative effects of human and natural forces on the dynamics of this fish population. We show, using a variety of models, how human activities have interacted with the cod population and demonstrate that a combination of top-down (fishing and predation) and bottom-up (nutrient loading and climate forcing) processes determined the dynamics of the eastern Baltic cod stock in the 20th century.

Recent changes in climate and the effects on distribution of marine species

Waters near Denmark such as the Baltic Sea and the North Sea have been warming since the late
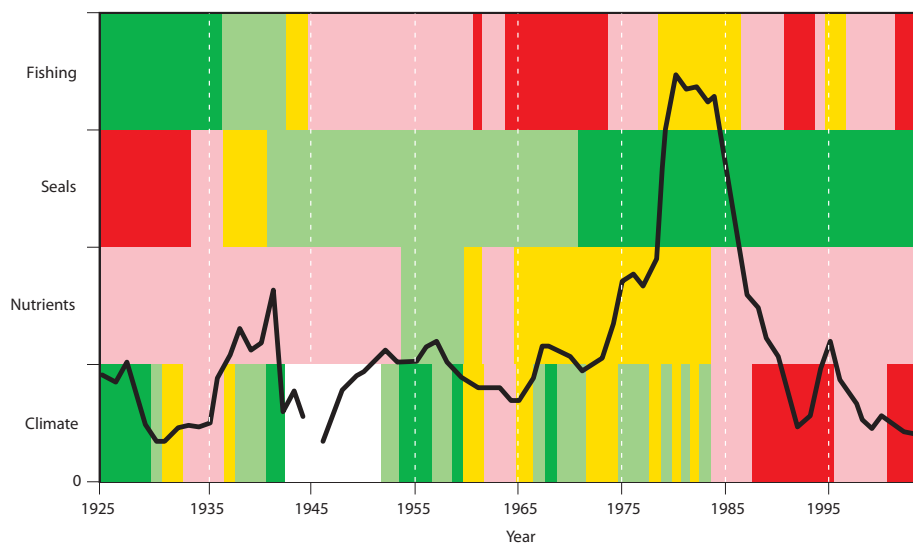
Fig. 14.2. Positive and negative effects of fishing, seals, nutrients and climate on the biomass of cod (black line) in the Baltic Sea during the 20th century. The large increase in cod biomass (and catch) during the 1970s is believed to be due to the simultaneous occurrence of positive (or less negative) effects of all four factors. The color scale from green to dark red represents levels of the forcing factor ranging from beneficial to detrimental and is based on quintiles of the ranges for each factor. Source: Eero et al. (*Proceedings of the ICES Annual Science Conference* 2008/J:08).

1980s. Daily sea temperature monitoring at lightships and harbours and a variety of other temperature measurements show that surface temperatures in the early 2000s were warmer than at any other time since instrumented measurements began (Fig. 14.3).

The rate of warming of the North Sea and Baltic Sea during the past 20 years has been much more rapid than the global trend and much more rapid than the rate predicted by the Intergovernmental Panel on Climate Change for the coming century. This is largely explained by multi-decadal climate variability over the northern Atlantic, but the processes causing this regional warming and cooling are not well known. The distribution of marine species in northern European waters has also changed remarkably rapidly, partly because of the high rate of warming they have experienced. Such high rates of warming and distribution shift are unlikely to continue, and there may even be periods of cooling lasting for a few years, as the temperature reverts closer to the regional and global means.

The research surveys of the DTU National Institute of Aquatic Resources and the catches of commercial fisheries show that "new" species such as

anchovy, red mullet, thicklip mullet and sardine are appearing in the North Sea and Baltic Sea with increasing frequency (see illustration on pages 148–149). These species are typically found in warmer waters further to the south (such as the Bay of Biscay), but the recent warm conditions have expanded their thermal habitats farther to the north. Anchovies may once again be reproducing near Denmark, as they did during a warm period in the 1940s, when fish egg sampling studies showed that the species was reproducing in the Kattegat. Even swordfish have been caught or found washed ashore in Denmark's waters (Belt Sea and Bornholm) during 2006–2008. Several of the new warmer-water species are already being fished commercially and will be followed with interest by sport fishers. Whether they become a significant proportion of future catches depends on the continuing rate of climate change, the establishment of new breeding populations and ensuring that they are not overexploited.

We recently carried out a meta-analysis of marine biota to detect the effects of climate change on distribution, abundance and other characteristics (particularly seasonality). The data were assembled
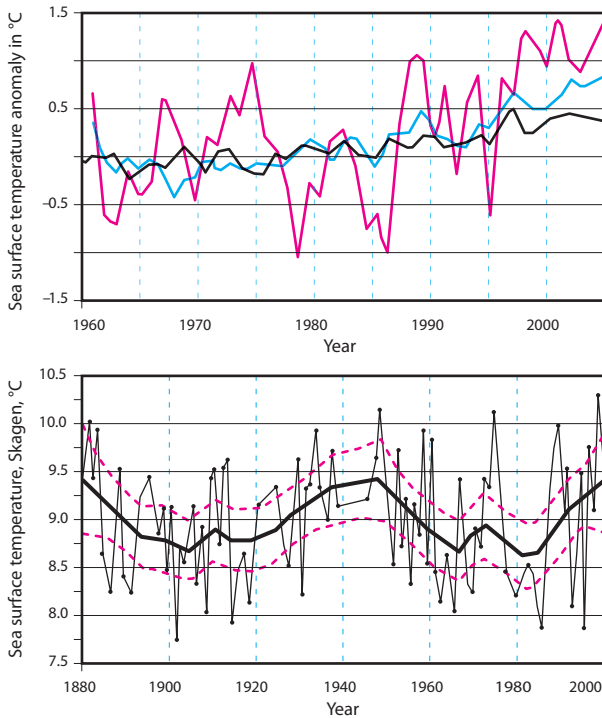
Fig. 14.3. **Top left**. Average annual anomaly of sea surface temperature (relative to the average for the 30-year period from 1961 to 1990). As the area covered by the average increases from local (Denmark's waters in red) to regional (northern Atlantic in blue) to global (in black), the geographical variability is smoothed out. Since 1980, the northern Atlantic has warmed at roughly twice the global rate, and Denmark's waters have warmed at roughly twice the northern Atlantic rate. The rates are 0.016°C, 0.029°C and 0.064°C per year respectively. The rapid warming of the northern Atlantic is partly due to a positive phase in multidecadal variability, which is likely to reverse. **Lower left**. A time series of annually averaged sea surface temperatures measured at the Danish lightship at Skagens Rev during 1880–1979 and updated with data from the International Council for the Exploration of the Sea. This time series is based on daily recorded measurements using standardized, calibrated instruments; each plotted value is an average of about 360 measurements. The thick dark line is a statistical fit from a general additive model (GAM), and the red lines are 95% confidence limits. A compilation of data sets, including this one and others from the North Sea and Baltic Sea, shows that sea surface temperatures in the early 2000s exceeded those seen since the 1860s–1880s (source: MacKenzie & Schiedek, *Global Change Biology* 2007: 13: 1335–1347). **Right panel**: An example of a lightship similar to the one at Skagens Reef on which the temperatures in the lower left panel were recorded. The vessel in the photograph was anchored near Gedser Reef, Denmark in the western Baltic Sea until 1970 and is now moored in Nyhavn, central Copenhagen.

from routine trawl surveys for fish (100 time series) and from a variety of sampling gears and survey methods for zooplankton (83), benthos (85) and birds (20). The analysis was part of a report produced by the International Council for the Exploration of the Sea at the request of the OSPAR Commission, which is responsible for protecting the marine environment in the northeastern Atlan-

tic from the adverse effects of human activities. It showed that the observed changes were consistent with expected climate-change effects in three quarters of the 288 cases examined. Climate is not the only cause of changes in distribution, but it is clearly having a detectable effect on many marine species. For seabirds and marine mammals that have life-history strategies that can adapt to vary-
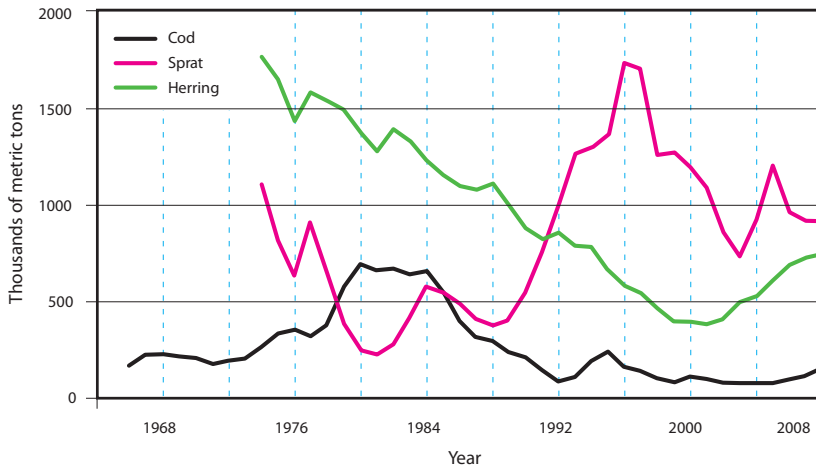
Fig. 14.4. The biomass of adult cod, sprat and herring in the Baltic Sea east of Bornholm, 1966–2008. Source: *Proceedings of the ICES Annual Science Conference* 2008/ ACOM:06.

ing environmental conditions, the evidence of climate-change effects was weaker.

Baltic fish populations and their sensitivity to climate

Much of our research in the Baltic Sea has focused on three main fish species, cod (*Gadus morhua*), sprat (*Sprattus sprattus*) and herring (*Clupea harengus*), which dominate both the total fish biomass (Fig. 14.4) and the commercial catch. Moreover, they interact ecologically with each other: adult cod are predators of both sprat and herring, but sprat and herring can be important predators of cod eggs and larvae; sprat and herring themselves compete for the same food during much of their lives and compete with cod larvae for the same food (Fig. 14.5).
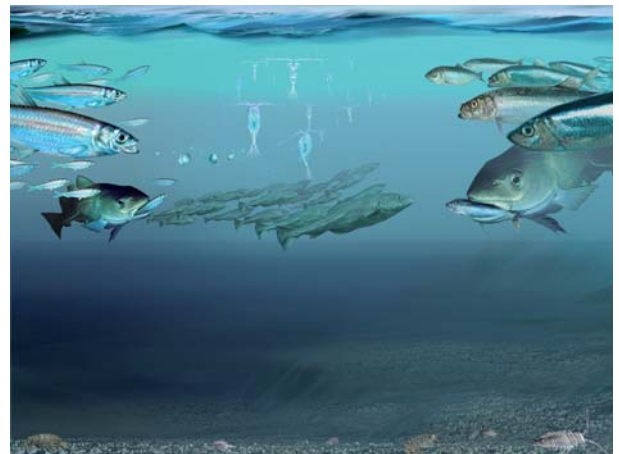
Fig. 14.5. Predator-prey and competitive interactions among the three most abundant fish species in the central Baltic Sea (east of Bornholm). **Left panel:** In years when the deep layer of the water column is oxic, large cod are abundant and prey on herring and sprat, but sprat and herring prey on cod eggs and herring prey on cod larvae; large cod also prey on benthic animals such as crustaceans. Sprat and herring compete with each other, and with cod larvae and young juveniles for zooplankton prey. **Right panel:** The deep layer of the Baltic Sea is anoxic. Species interactions in the water column are similar to those in the left panel, but benthic animals are absent, and cod must feed in the water column. Cod become less abundant because their eggs cannot survive in the low oxygen concentrations of the deep layer. Herring and sprat become more abundant because biomass of their main predator has declined. Content: MacKenzie et al. (*Global Change Biology* 2007: 13: 1335-1347).
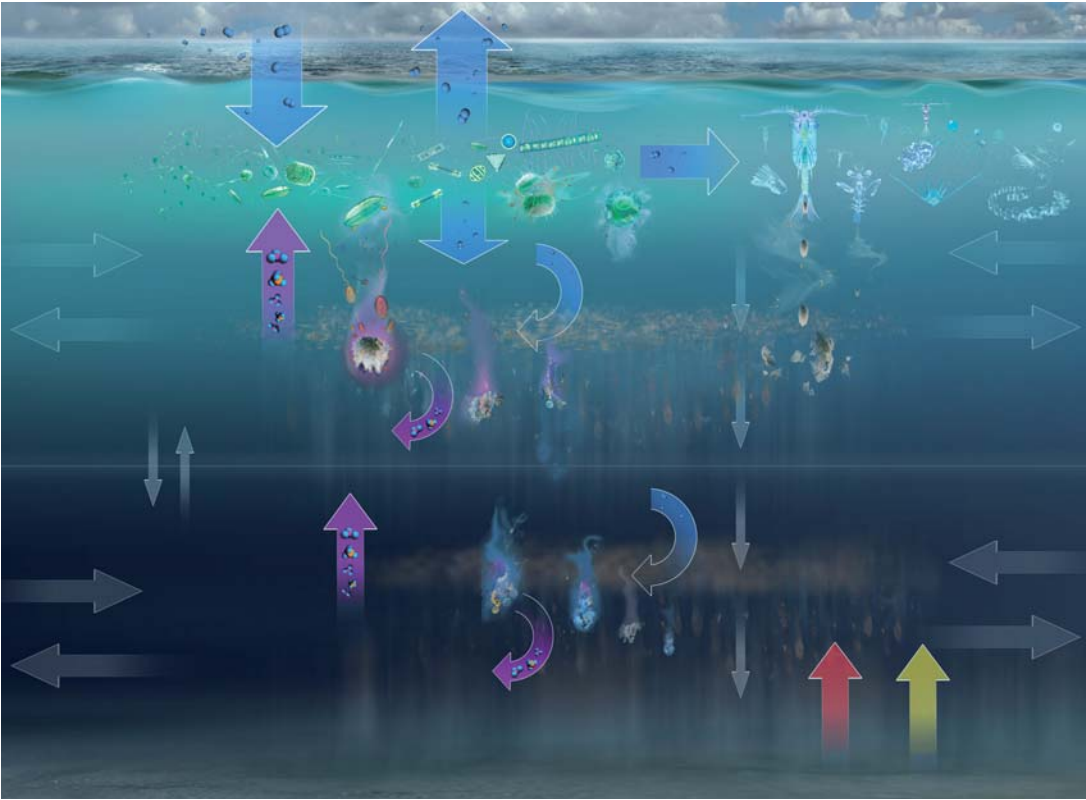
Cod is particularly sensitive to the low salinity of the Baltic Sea. It cannot reproduce in waters below 11 parts per thousand (psu), and unfortunately for cod, this means that most of the Baltic is off-limits for reproduction. In fact, the only areas where salinity exceeds this threshold are parts of the three deep basins (Bornholm, Gdansk and Gotland) where the relatively dense, high-salinity water is near the bottom.

However, the same deep-water masses that occasionally have sufficient salinity are frequently also anoxic. The oxygen concentration is often low due to the natural topographical and hydrographical situation of the Baltic Sea and eutrophication-related production and decomposition of organic matter (Fig. 14.6). The water column in the Baltic does not become mixed because, unlike the North Sea, tidal mixing is very weak and wind mixing is not sufficient to overcome the strong density gradient due to relatively fresh water on top of relatively salty water. Even the hurricane of December 1999 was not sufficient to completely mix the water column in the Bornholm Basin.

Fig. 14.6. Biotic and abiotic processes affecting suitable habitat for successful cod reproduction in the eastern Baltic Sea. The processes include the production of organic matter, including fecal pellets, by phytoplankton and zooplankton, its subsequent decomposition by single-celled heterotrophs and bacteria, and the exchange of chemical compounds across the air-sea and water-sediment interfaces and across the permanent density gradient ("pycnocline"; depicted by a thin horizontal line across the figure) between upper fresh and lower saline water. The decomposition process consumes oxygen and releases nitrogen- and phosphorous-containing compounds back to the water column, which can subsequently stimulate new primary production. Other aerobic biota consume oxygen as part of respiration. Dissolved oxygen concentrations in the deep layer can increase after major inflows of saline water from the Kattegat and North Sea. Between inflows the deep layer can become anoxic with high concentrations of hydrogen sulfide (H_2S). Content: Neumann and Schemewski (*Journal of Marine Systems* 2005: 56: 195–206). See also Fig. 14.7.
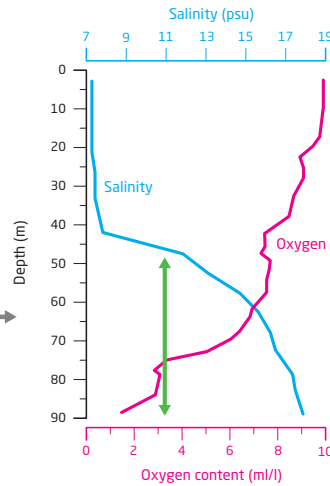
Fig. 14.7. Vertical profiles of salinity and oxygen concentration from a central location in the Bornholm Basin, Baltic Sea. The green arrow delimits the depth layer whose salinity (11 practical salinity units) and oxygen concentration (2 ml/l) allow cod eggs to be fertilized and hatch.

Laboratory experiments on the physiological limits for successful fertilization and hatching of cod eggs can be used to determine the limits of the cod reproductive habitat (water with sufficiently high levels of oxygen and salinity). Then the depths in the sea at which these tolerance limits occur are used to define a vertical layer of water with minimum physiological requirements. Measurements of salinity and oxygen concentrations from vertical profiles taken during cruises (Fig. 14.7) are plotted and integrated horizontally to calculate the volumes of suitable cod reproductive habitat in the three Baltic basins. The time series of these reproductive volumes can be compared with the rate of production of new recruits. Such comparisons show that the Bornholm Basin has always had some reproductive volume but that its size has varied substantially since the 1950s. The other two basins have essentially been reproductive dead zones during most of the last 20 years, contributing to a significant decline in the recruitment of cod. The existence of sufficient reproductive volume affects how quickly the population might recover from a reduced stock level.

Why have the volumes varied so much over time? The Baltic Sea is relatively isolated from the North Sea; the shallow sill formed by Denmark's straits limits the horizontal exchange of saline water masses; and large-scale weather and atmospheric processes determine the frequency of inflows from the North Sea. Major inflows of saline, oxygen-rich water from the North Sea are infrequent, and the oxygen in the deep saline layer of the three basins is gradually consumed by aerobic metabolism. Measurements of salinity from Dan-

ish lightships since the 1890s show that the frequency of these events declined after about 1980 from 1 per 2–3 years to 1 per 10 years; there have only been two major inflows (January 1993 and January 2003) since the early 1980s. These inflows are important both for the survival of cod eggs and for the abundance of prey for cod larvae. Superimposed on the decline in inflow frequency is the continuing eutrophication of the Baltic Sea, which stimulates the production and decomposition of organic matter and reduces oxygen levels. Thus, both natural and human-induced environmental changes affect cod population dynamics.

While the environmental conditions deteriorated, the fishing pressure increased and remained high. Landings peaked in the early 1980s and remained high during most of the decade. However, the combination of increasing fishing mortality and deteriorating reproductive volume proved unsustainable, and the overall biomass of the population declined to a record low. New restrictive fishing regulations were implemented to promote recovery, such as lower quotas, longer closed season, larger mesh in nets and larger minimum size of captured cod.

Together with colleagues in European Union projects, we are now developing coupled models of the physical oceanography (such as major Baltic inflows from the North Sea) and the biological processes (such as rates of photosynthesis by phytoplankton, rates of oxygen consumption and decomposition of organic matter) that control the oxygen concentration in the water column (Fig. 14.6). Our objective is to be able to use process knowledge to quantify, describe and predict the fluctuation in cod reproductive volume and cod popula-

tion dynamics, in response to both climate-change and nutrient-loading scenarios. Such a tool would contribute to implementing new ecosystem-oriented policies for the Baltic Sea on eutrophication, climate change and fisheries management.

Sprat reproduction is less sensitive to salinity and oxygen stresses than cod reproduction. The difference in sensitivity between the species is due primarily to differences in egg buoyancy; sprat eggs are more buoyant than cod eggs and float higher in the water column, where oxygen concentrations are higher. However, sprat eggs require higher temperatures to hatch than cod, and hatch success in the laboratory declines when temperatures fall below 3-4°C. Sprat recruitment in the Baltic Sea increases with temperature. Warm temperatures have several beneficial effects on sprat recruitment: eggs are more likely to hatch, growth rates of larvae are higher and the supply of zooplankton food for the larvae is higher in warm years to meet the increased metabolic and energy demands of the growing larvae.

Temperatures in sprat-spawning areas rose during the late 1980s and have generally remained above average since then. As a result, the conditions for sprat reproduction have been favorable and recruitment has been high. Over the same period cod, which is the main predator of sprat, suffered due to both high fishing pressure and poor environmental conditions. Sprat biomass has therefore increased due to a combination of improved reproductive environment and reduction in predation pressure due to the decline of cod biomass (Fig. 14.5).

Effects of future climate change and interaction with fisheries management
Predicting the effects of future climate change on marine ecosystems and fisheries requires coupling climate models and biological models. The product of this coupling is not very reliable, but strategic decisions about fisheries and marine ecosystems require guidance on the effects of alternative management options under different scenarios of climate change. These decisions can be better informed if based on thorough understanding of processes affecting ecosystems and populations and how fishing interacts with these processes to

increase vulnerability to population collapses. As we have shown here, a variety of research approaches are achieving this understanding.

For example, we conduct direct experimental fishing surveys to monitor changes in the abundance and distributions of various stages of fish life-history (eggs, larvae and adults) and the abiotic conditions in marine ecosystems (illustration on pages 148–149 and Fig. 14.5–14.7). Laboratory experiments are needed to measure key physiological processes (such as growth rates relative to food supply), behavior and species interactions (such as dietary preferences) under controlled conditions. This knowledge can then be incorporated quantitatively into population and ecosystem models. The model outputs are then compared with field data for reliability and accuracy and are updated as conditions in nature change (such as due to sustained changes in temperature or nutrient loading or the accidental introduction of a non-native species that preys on or competes with a native species).

The expected output includes ecosystem-based, interdisciplinary approaches to assist policy-makers and inform stakeholders about the effects of various human activities; another output could be spatially and temporally resolved three-dimensional models and animations of interactions between and within species and with the abiotic environment. New research questions (such as how climate change will affect the spatial distributions of predators and prey and thereby affect their dynamics and yields to fisheries) should be addressed and would form a powerful public outreach vehicle for communicating the complexities of fish population and ecosystem dynamics to wider nonspecialist audiences.

The limited number of species present in the Baltic Sea and the environmental conditions, especially salinity, being close to the tolerance limits of marine species mean that relatively simple ecological models can capture much of the dynamics and show strong responses to climate-induced changes. Only three species – cod, herring and sprat – produce most of the fisheries yields, but fully explicit interacting life-history models of all three would be too complex to handle if coupled to explicit models of phytoplankton and zooplankton production. Instead, we have developed simpler ecosystem-based models that integrate across trophic levels, include multiple drivers (especially climate and fishing) and
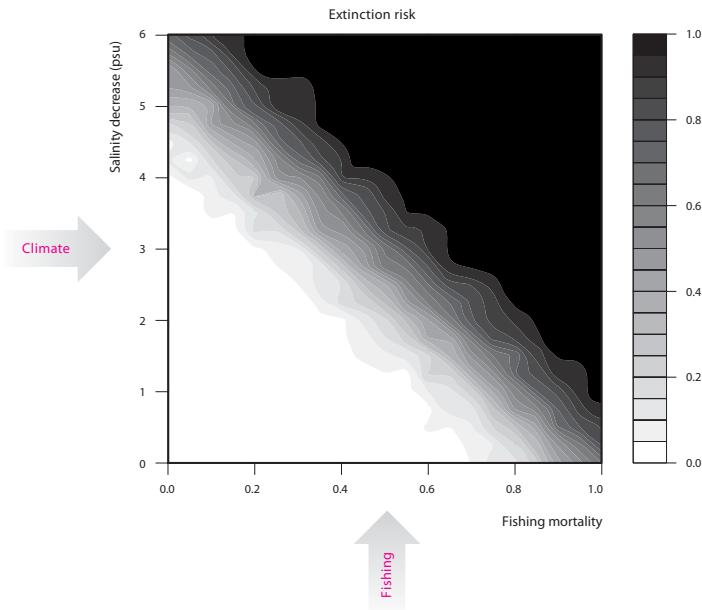
Fig. 14.8. The risk that cod will become extinct in the Baltic Sea during the 21st century depends on both change in climate (particularly salinity) and controlling the level of fishing. Source: Lindegren et al. (ICES CM 2008/OCC:08).

estimate the uncertainty of the forecasts from all sources. The models have been used to explore the dynamics of cod in the 21st century as climate and fishing mortality change. The results indicate what level of fishing will be needed to remain within acceptable levels of the risk of stock collapse, depending on how the climate changes (Fig. 14.8).

Other human-induced pressures such as habitat degradation, pollution and eutrophication are also stressing fish populations and causing loss of biodiversity at all levels from ecosystem biodiversity to genetic change. Again, our Baltic test case is a useful example. Here we can expect that increases in temperature will reduce oxygen concentrations (interacting with eutrophication), and a reduction in salinity will further stress resident marine species such as cod, sprat, plaice and turbot. Their spatial distributions will contract if the negative effect of salinity reduction outweigh the positive effects of warmer temperature and will probably shift further south and west within the Baltic Sea. However, new species that might immigrate to the Baltic because of warmer temperatures may not be able to reproduce in its lower salinity. As a result, the over-

The authors

Brian R. MacKenzie: MSc, Dalhousie University, 1985; PhD, McGill University, 1992. Professor with Special Responsibilities in Fisheries Oceanography, National Institute of Aquatic Resources, DTU. Has worked in Denmark since 1992. Investigates how climate and fishing affect the dynamics of marine fish populations and ecosystems. Member and former chair, several expert committees, Census of Marine Life and the International Council for Exploration of the Sea. Active in large international (such as EU) research projects and networks of excellence.

Keith Brander: PhD in Fish Population Dynamics, University of East Anglia. Senior Researcher, National Institute of Aquatic Resources, DTU. Coordinator, ICES (International Council for the Exploration of the Sea)/GLOBEC (Global Ocean Ecosystem Dynamics) Cod and Climate Change program. Has worked in fisheries management and carries out research on climate change and marine ecosystems. President, Sir Alister Hardy Foundation for Ocean Science, 1999–2005. Lead author on fisheries, Fourth Assessment Report, Intergovernmental Panel on Climate Change, 2007.

all fish biodiversity of the Baltic fish community could decrease. Ecological theory predicts that the resilience and stability of ecosystems is generally higher when more species are present; the vulnerability of the Baltic ecosystem to perturbations, including those due to fishing and climate change, could be increased by a reduction in biodiversity. Conversely, we can expect that action that alleviates stress due to any of these types of pressure may have beneficial effects in relation to effects due to the other types of stress, because, for a variety of reasons, less-stressed populations and ecosystems are likely to be more resilient. We can make this case by looking at the consequences of fishing mortality on food webs, fish communities and fish populations.

When a fish population is exploited, the biomass is reduced and the population becomes more sensitive to other types of pressure, including climate. Many species of fish have a natural lifespan of decades, which enables them to withstand long periods of poor environmental conditions. Fishing reduces life expectancy and truncates the age structure of the population, with adverse consequences for the natural buffering capacity. Removal of older fish may adversely affect larval survival and recruitment in species in which older fish produce more viable offspring and spawn over a wider geographical and seasonal span. Fishing and habitat degradation may reduce the spatial heterogeneity or geographical structure of fish stocks, and genetic variability can be affected.

Yield models and empirical studies show that reducing fishing mortality will reverse downward trends in biomass and age structure, thus enhancing the resilience of the population and making it better able to withstand additional stress due to climate change. In addition, reducing fishing mortality produces other benefits, including higher catch rates and high, stable yields, and these are likely to accrue regardless of the actual rate and magnitude of climate change. This is therefore a no-regret, win-win strategy. The other benefits are increasing catch-rate and increasing total yield if the stock has been overfished. The increase in catch-rate means that the same catch can be taken with lower fishing effort and therefore less use of fuel, which is a climate mitigation benefit. Thus, despite the difficulty of predicting future fisheries production and distri-

bution changes, we can already recommend management actions that will be robust to the uncertainty in both climate and biological predictions. The main way to sustain fish populations and fisheries even though climate is changing is to intensify efforts to deal with the old familiar problems of overfishing, loss of biodiversity, habitat degradation and pollution.

More to explore

Barange M, Field J, Harris R, Hofmann E, Perry I, Werner F. *Marine ecosystems and climate change*. Oxford, Oxford University Press (in press).

Brander KM. Global fish production and climate change. *Proceedings of the National Academy of Sciences of the United States of America* 2007: 104: 19709–19714.

Cury PM, Shin Y-J, Planque B, Durant JM, Fromentin J-M, Kramer-Schadt S, Stenseth NC, Travers M, Grimm V. Ecosystem oceanography for global change in fisheries. *Trends in Ecology and Evolution* 2008: 23: 338–346.

International Council for the Exploration of the Sea. The effect of climate change on the distribution and abundance of marine species in the OSPAR Maritime Area. *ICES Cooperative Research Reports* 2008: 293: 1–45.

MacKenzie BR, Gislason H, Möllmann C, Köster FW. Impact of 21st century climate change on the Baltic Sea fish community and fisheries. *Global Change Biology* 2007: 13: 1348–1367.

Stenseth NC, Ottersen G, Hurrell JW, Belgrano A. *Marine ecosystems and climate variation*. Oxford, Oxford University Press, 2004.

The adventurous journey of Spaceship Earth

Henrik Svensmark

During its worst climatic episodes, about 2300 million years ago and about 700 million years ago, the Earth was in deep freeze. Even tropical coastlines would have looked like this scene from Antarctica today, to judge from evidence of stones dropped by icebergs near the equator. The explanation is that our planet suffered from the effects of close encounters between other galaxies and the Milky Way, which provoked enhanced star formation rates and intense cosmic radiation.

The most profound questions with the most surprising answers are often the simplest to ask. One is: why is the climate always changing? Historical and archaeological evidence of global warming and cooling that occurred long before the Industrial Revolution and the geological traces of far greater variations before human beings even existed require natural explanations. Regrettably, the global sense of urgency about understanding current climate change as well as possible does not extend to profound investigations of natural influences. As a result, a small group of physicists at the DTU National Space Institute has enjoyed a near monopoly in one unfashionable but highly productive line of research for more than a decade.

Here the surprising answer about those never-ending natural changes of climate is that galactic cosmic rays, atomic particles coming from the supernova remnants left by exploded stars, appear to play a major part (Fig. 15.1). By ionizing the air, cosmic rays help to form the aerosols, cloud condensation nuclei, required for water droplets to condense and create low-altitude clouds. As these exert a strong cooling effect, increases or decreases in the cosmic ray influx and in cloudiness can significantly lower or raise the world's mean temperature. This is our central hypothesis, but the earth sciences have an unfortunate history of spurning external inputs.

Telescopes on interplanetary spacecraft can look back at the Earth and see a small blue dot. It makes a pretty color contrast with red Mars and white Venus, but it is still only a dot in the vast scenery of cosmic space. Although 2009 is the 400th anniversary of the astronomical telescope, many people have been slow to catch up with what astronomy has taught us. They do not like to admit that our planet is itself a kind of spaceship, not very large by cosmic standards, and exposed to great hazards as it travels through a busy universe.

For example, three centuries had to pass before most scientists were ready to acknowledge the scale of environmental disasters caused by traffic accidents in the solar system. The danger was obvious to Edmond Halley (famous for his reappearing comet) as soon as his friend Isaac Newton showed that comets follow capricious orbits around the Sun. In those days, the greatest disaster imaginable was Noah's flood, and in 1694, Halley proposed to
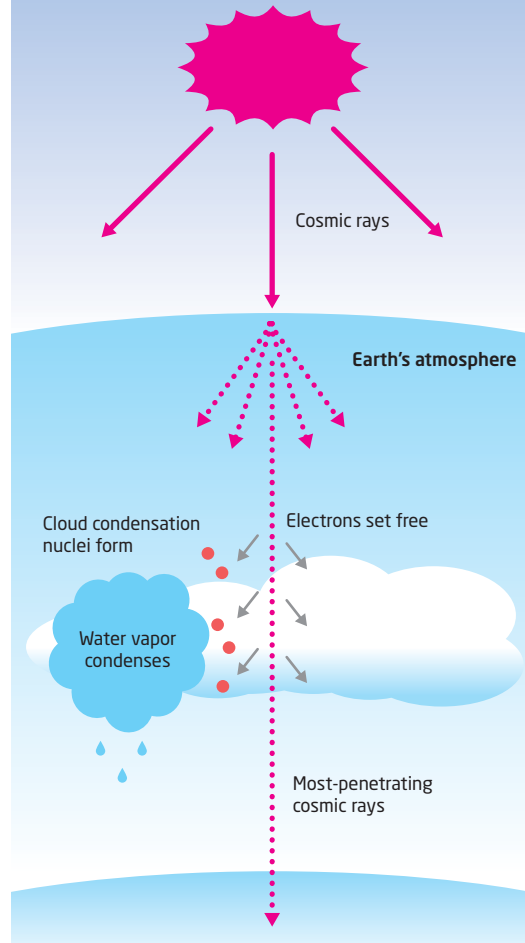
Fig. 15.1. The hypothesis that attributes climate change to variation in the influx of cosmic rays from exploded stars in the galaxy focuses on the seeding of clouds at low altitudes, where cosmic rays are scarcest and the variation has the greatest effect.

the Royal Society of London that a comet hit the Earth. The Caspian Sea might be the impact crater, he suggested, and the ocean rushing in to fill the hole could have caused the biblical flood.

How ridiculous! That was the opinion of the surveyors of strata and collectors of fossils who founded the modern earth sciences. They wanted no outside interference from natural influences beyond the Earth. Their pioneering treks across mountains and deserts and oceans taught them, correctly, how magnificent and active our planet is. Incorrectly, they decided it was far too grand to be troubled very much by anything in the sky. Long after Halley, astronomers repeatedly pointed out the risks from comets and the later-discovered asteroids. A buried crater identified under the coast of Mexico in 1991 settled the issue. The Chicxulub Crater is

170 kilometers wide and was created by a collision with a comet or asteroid. Dated at 65 million years ago, it coincided exactly with the environmental catastrophe that wiped out the dinosaurs and terminated the Mesozoic Era. After 300 years of neglect, Halley's hypothesis is now intergovernmental business, and Spaceguard projects aim to protect the Earth from the next big traffic accident. Just since 1991, the count of known near-Earth objects has increased from about 200 to more than 6000.

This historical analogy does not prove the correctness of our own offering from astronomy, the cosmic-ray theory of climate change. The fate of Halley's idea simply warns us that, even if we are right, we have to be very patient about persuading the colleagues in the earth sciences who still prefer traditional home cooking – terrestrial explanations for terrestrial events. Meanwhile, we need the geologists' reports on past changes of climate, because they provide some of the very best evidence for our controversial theory (Fig. 15.2).

When ice drifts across the ocean and then melts, gritty material carried by the ice drops onto the sea floor. Marine geologists recover cores from the seabed and find layers of exotic ice-rafted debris that denote cool periods periodically interrupting warmer periods with little or no debris. The leading investigator of recent climate history in the North Atlantic was the late Gerard Bond of Columbia University, and in 2001, he and his colleagues reported on ice-rafting events since the end of the most recent ice age. The last of them was the Little Ice Age, when all the Viking settlers in Greenland perished and Sweden took Denmark by surprise with an invasion across the frozen sea. The Medieval Warm Period preceded the Little Ice Age, and alternating cold and warm spells extend back 11,500 years in Bond's seafloor record. Long before the modern industrial age, our ancestors in the Epipaleolithic Period, Neolithic Period, Bronze Age and Iron Age endured an ever-changing climate.
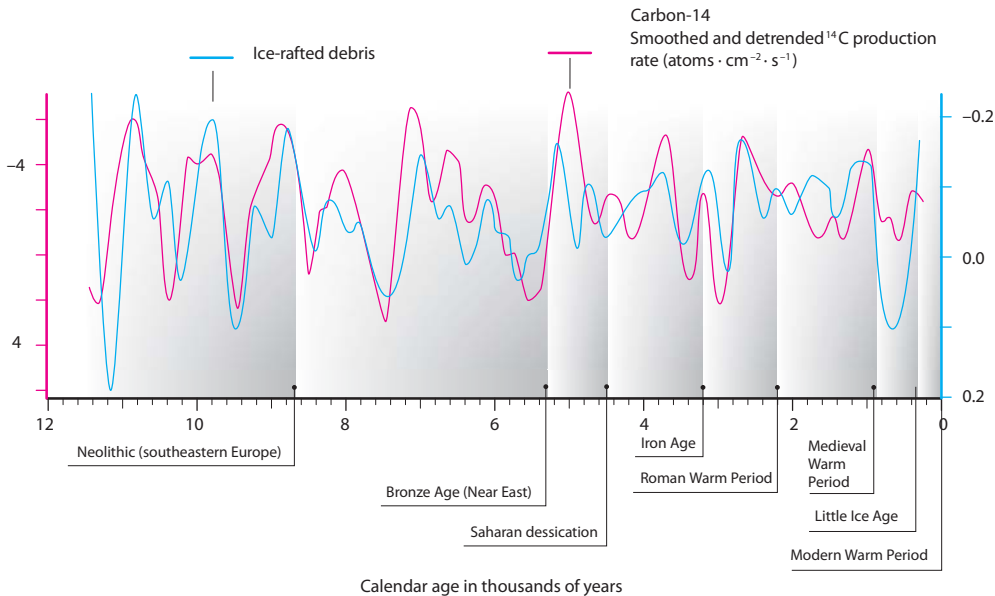
Fig. 15.2. Cool and warm periods during the past 11,500 years demonstrated by peaks and troughs in the red curve showing ice-rafted debris on the North Atlantic seabed. They correspond with high and low influxes of cosmic rays, indicated here by the blue curve of production rates of carbon-14, smoothed and detrended. An important source of error is an imperfect match between ages in the ocean-core and radionuclide series (2σ typically ~200 years). Even so, the correlation in the case illustrated is 44%, and 56% was obtained with beryllium-10 for the period 11,500 to 3000 years ago.

Cosmic rays and climate change

The repeated cold spells coincided with times of high cosmic-ray influx, recorded by high counts of radioisotopes, carbon-14 and beryllium-10, made when cosmic rays react with terrestrial atoms. Bond's team imagined that the high cosmic rays were merely a symptom of a magnetically quiet Sun, which would also be less bright. Nevertheless, variation in solar irradiance measured by satellites suggests that they would be too small to explain this cooling and warming. An amplifier is needed, and in our hypothesis, the cosmic rays are not just a proxy for solar activity but a causal agent, helping to seed clouds that cool the world. Bond's ice-rafting events occurred when a magnetically weak Sun failed to repel as many cosmic rays as it does when it is most active. We should therefore count the behavior of our parent star as a cosmic hazard. At the present stage of the Sun's middle age, the maxima and minima in cosmic rays seen over 12,000 years probably represent the typical range of solar magnetic variability.

How can we check that cosmic rays, rather than irradiance or any other solar effect, were mainly responsible for these climate changes? One way is to look much farther back in time. The natural parti-cle accelerators that fire off the cosmic rays from supernova remnants evolve slowly over 100,000 years or more. During an interval as short as 12,000 years, the background of cosmic rays scarcely changes. But now change the time scale drastically. Over millions of years, as Spaceship Earth travels through our galaxy, the Milky Way, in company with the Sun and the rest of the Solar System, cosmic rays vary much more, which has nothing to do with the Sun. Instead, the frequency of supernova events depends on where we are in the galaxy.

An astrophysicist at the Racah Institute in Jerusalem, Nir Shaviv, heard about our theory that cosmic rays can cool the world, and he came up with a fascinating story about climate change in the remote past. Geologists tell us that over hundreds of millions of years the Earth has sometimes been a hothouse, with virtually no ice at all, and sometimes an icehouse with large ice sheets spreading out from the poles. Shaviv said that the cold periods came whenever the Solar System visited one of the spiral arms of our galaxy, where the hot, blue and most explosive stars are concentrated and cosmic rays are particularly intense. In the darker spaces between spiral arms, cosmic rays were

Mass density	
ρ_{local} (now)	$0.115 \pm 0.1\ M_\odot \cdot pc^{-3}$
$\bar{\rho}$	$0.145 \pm 0.1\ M_\odot \cdot pc^{-3}$
$\rho_{arm}/\rho_{interarm}$	$\approx 1.5 - 1.8$

Timing of spiral-arm crossings	
Sagittarius-Carina	34 ± 6 million years
Scutum-Crux	142 ± 8 million years

Pattern speed	
Ω_p	$13.6 \pm 1.4\ km \cdot s^{-1} \cdot kpc^{-1}$
φ_1	$25 \pm 10°$
φ_2	$100 \pm 10°$
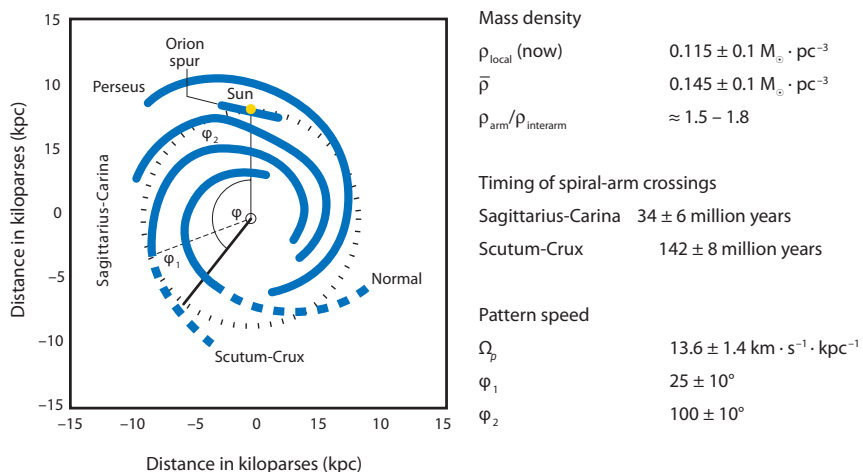
Fig. 15.3. The Sun's motion relative to the Milky Way Galaxy's spiral arms over the past 200 million years is defined by changes in the angle φ. The most recent spiral-arm crossings φ_1 and φ_2 were at ~100° and ~25°. Climatic data show rhythmic cooling of the Earth whenever the Sun crossed the galactic midplane, where cosmic rays are locally most intense, at intervals of ~34 million years. From these geophysical data, important astrophysical data can be deduced, as shown at right.

scarcer and the Earth was a hothouse. Shaviv's first papers on this theme caused a stir, and *Discover* magazine called it one of the top science stories of 2002. Our group in Copenhagen followed up Shaviv's propositions with astrophysics of our own. In 2006, for instance, I turned the reasoning around and used the climate data to give us a better measure of some key features of the galaxy. The trick was to use as a yardstick a well-dated 34-million-year climate cycle associated with the Solar System's oscillations through the mid-plane of the galaxy, where the cosmic rays are most intense (Fig. 15.3).

Star clusters and cosmic rays

About 4 million years ago, as I also realized, Spaceship Earth blundered into special danger when it entered a low-pressure region called the Local Bubble, blown by intense winds from hot, massive stars. Some of those stars were exploding as supernovae, and when we calculate the intensifying cosmic ray background, the fastest rate of increase was about 2.75 million years ago. That result is exciting because the North Atlantic was turning icy at just

that time, and reduced rainfall in Africa changed some of the jungle into the open savanna where the first human beings evolved.

Although the map of the Milky Way's spiral arms may be the key to many episodes of climate change in the past, we see the galaxy edge-on, from our viewpoint in the suburbs, and nearby spiral arms spoil the view of distant ones. The map is so uncertain that astronomers even argue about how many major spiral arms there are. In our latest astrophysics, we sidestep that problem by making no assumptions about the spiral arms and instead focus attention on the Solar System's most important escorts on its orbit around the center of the galaxy (Fig. 15.4).

Fig. 15.4. The Pleiades, the nearest open star cluster, are 490 light-years away and were formed together as a large group. Their most massive siblings blew up as supernovae and contributed to the cosmic rays at the Earth towards the end of the Mesozoic Era.

They are the open star clusters, of which the Pleiades, or Freja's hens as the Vikings called them, are just the most visible example. Together with many similar clusters that have long since scattered beyond recognition, the surviving clusters provided the nearest sources of cosmic rays over hundreds of millions of years. Each cluster was born as a huge family of stars, of which the most massive ones blew up as supernovae between 3 and 40 million years ago after the birth of the cluster (Fig. 15.5).

These nearby supernovae become a guide to the cosmic-ray influx to the Earth. From the University of Vienna's WEBDA catalogue, which gives their ages, I selected 283 open clusters lying within 1 kpc, or 3300 light-years. The Space Telescope Science Institute's Starburst99 program then enabled me to compute the average rates of nearby supernovae in intervals of 8 million years, going back to the early days of animal life 500 million years ago. The highest supernova rates were twice the lowest rates and corresponded neatly with the icehouse times identified by geologists. Low rates went with the hothouses.

During the dinosaurs' reign, for example, severe global warming of 6.5°C, dated by geologists at about 183 million years ago, coincided with a particularly low supernova rate 184–176 million years ago. Our peak supernova rate 296–288 million years ago fits very well with the acme of the Late Paleozoic icehouse 294–289 million years ago, when the dominant egg-laying tetrapods, the sphenacodonts, took refuge in the tropics. In the whole 500 million years period of our study, there are no gross mismatches between supernova rates and climate. By 450 million years ago, the timing is off by a few percent, but that is within the growing uncertainty of the astrophysical calculations so far back in time.

If Noah's flood was the worst terrestrial catastrophe that Edmond Halley could imagine, geologists speak nowadays of episodes in the Proterozoic Eon about 2300 and 700 million years ago when there were glaciers at the equator, even at sea level. Whether or not their tag Snowball Earth is a valid description, these were evidently the coldest times ever known, and they seem to tell of yet another hazard for our little blue dot, arising this time from intergalactic traffic accidents. A close encounter between the Milky Way Galaxy and another galaxy
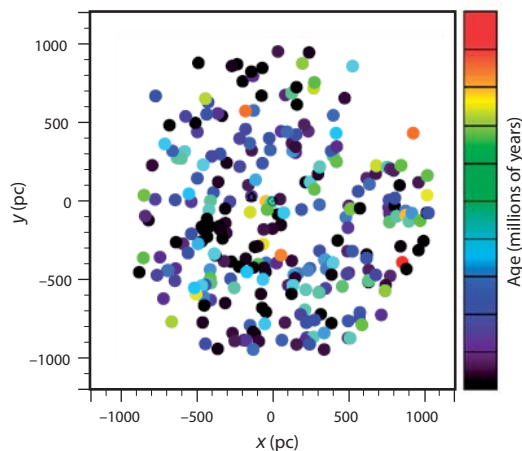
Fig. 15.5. For stellar clusters within 1 kpc (~3300 light-years), the colors denote the ages of the clusters. Most of them are relatively young survivors of a process of dispersal that has led to the disappearance of many old clusters. Although the analysis can consider this, the disadvantage is increasing uncertainty as one goes back in time. The symbol in the center denotes the position of the Solar System.

can provoke a starburst event – a stellar baby boom. Nir Shaviv as well as Raul and Carlos de la Fuente Marcos in Madrid have linked those Proterozoic deep-freeze events to exceptionally high star formation rates, which implied very high doses of cosmic radiation for Spaceship Earth. Less dramatic, but equally telling, is an ice-free world that lasted for a billion years in the mid-Proterozoic Eon, when the star formation rate was persistently low.

From the Little Ice Age a few hundred years ago, via the Late Paleozoic glacial peak 290 million years ago, to the Snowball Earth episodes much farther back in time – could all those cold climates have coincided with cosmic-ray peaks just by chance? I may not think so, but Earth scientists still have completely different theories of their own, usually invoking carbon dioxide as the most important driver of climate change, even in the remote past. So far, there is little meeting of minds about the cosmic influence.

Search for evidence

Our long-range explorations through space and time would be less meaningful if there were not

clear evidence from observations and experiments that cosmic rays really do affect the clouds and the climate, here and now. Temperatures both in the upper air and in subsurface ocean water continue to follow the reductions and increases in cosmic rays during the solar cycles. Since the 1990s, the temperature changes also reflect a recent moderation of the Sun's magnetic activity after its intensification during the 20th century.

Meanwhile, our experimental work has revealed the microphysical and chemical mechanism by which cosmic rays help to seed liquid-water clouds in the lower air. It is a matter of ion-induced nucleation of micro-aerosols, small clusters of sulfuric acid and water molecules, which can then grow into the cloud condensation nuclei on which water droplets form. In 2005, our team completed the world's first laboratory experiment on the role of cosmic rays in aerosol formation in the basement of the Rockefeller Complex on Juliane Maries Vej in Copenhagen.

We called the experiment SKY (*cloud* in Danish) and used natural cosmic-ray muons coming through the ceiling, supplemented by gamma-ray sources when we wanted to check the effect of increased ionization of the air. A 7-m^3 reaction chamber contained purified air and the trace gases that occur naturally in unpolluted air over the ocean. Ultraviolet lamps mimicked the Sun's rays, and instruments traced the chemical events. We were surprised by how quickly the ionizing radiation worked, in a split-second, suggesting that electrons released in the air act as efficient catalysts for building the microaerosols.

To verify that what we saw really was the effect of cosmic rays and not some other chemical reaction, we made a small-scale version of the experiment called SKY@Boulby. It first confirmed the SKY results at sea level and is now operating 1.1 km underground in partnership with the Boulby Dark Matter Laboratory in a mine in England, where overlying rocks block nearly all the cosmic rays. The ionization can be reduced by three orders of magnitude. Next, we want to build an experiment of intermediate size, where we can vary temperature and pressure to determine the influence of altitude on the aerosol reaction. In connection with these Proterozoic deep-freeze events, we will test the effect of putting exotic gas mixtures in the reaction chamber, to simulate the very different atmospheres that the Earth possessed long ago.

In 2008, two groups of investigators independently claimed to have falsified our hypothesis about cosmic rays and clouds. They announced that they had failed to detect a necessary consequence of our mechanism: an emphatic reduction in cloud cover that should follow a large decline in the influx of cosmic rays. The Sun arranges this convenient test of our hypothesis by coronal mass ejections – blasts of gas from the Sun, which can act like a magnetic umbrella and quickly reduce the cosmic-ray supply for a few days.

To be fair to these critics, the effect on clouds in the aftermath of such Forbush decreases (as they are called) is not easy to find. Changes in cloudiness from day to day that occur for quasi-random weather reasons tend to mask the influence of cosmic rays, and the full impact is delayed for more than a week. Nevertheless, when looked for with sufficient care, the impact on clouds turns out to be not just detectable but huge.

Crucial to our own investigations of the Forbush decreases is thorough understanding of the importance of penetrating muons, which are the chief source of ionization in the lower atmosphere. We pool data from 146 cosmic-ray monitors around the world to ascertain the heliospheric character of each Forbush decrease and compute the consequential loss of muons (Fig. 15.6). In this way, we can rank the events according to their effect at low altitudes. Then we review the satellite data on clouds, including those from the Special Sensor Microwave/Imager, carried on spacecraft of the United States Defense Meteorological Satellite Program, which looks down on the clouds and measures their content of liquid water. In the aftermath of the most effective Forbush decreases, the data show billions of tons of liquid water disappearing from the sky as if by magic.

In rebuttal of the negative reports on this subject, I was able to preview our Forbush results at the 33rd International Geological Congress in Oslo in 2008. While there, I was pleasantly surprised by how open-minded quite a few of the Earth scientists seem to have become about alternative theories of climate change.
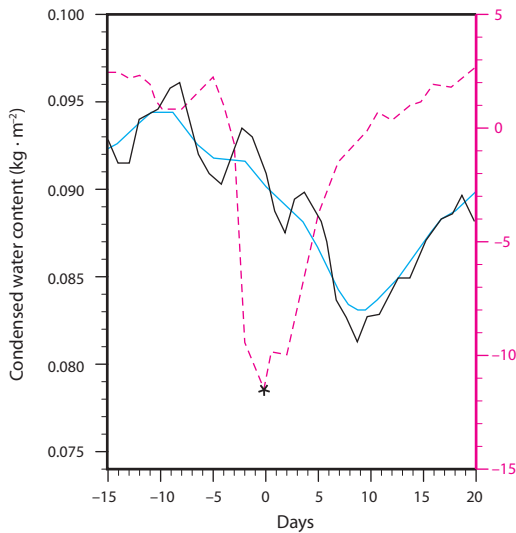
Fig. 15.6. A large drop in cosmic rays recorded at the Climax station in Colorado (red broken line, red % scale) on September 13, 2005 is an example of a Forbush decrease due to a coronal mass ejection from the Sun. The black curve shows the liquid-water content of clouds over the oceans, as observed by the satellite-borne Special Sensor Microwave/Imager instrument, dropping to a minimum about 10 days later. That interval is a guide to how long it takes for small aerosols made by cosmic-ray action to grow into cloud condensation nuclei. The blue curve (Gaussian smoothing with a width of 2 days) reduces the cloud fluctuations due to quasi-random weather factors.

Future perspectives

Opportunities for further investigation in cosmo-climatology, as we call it, remain so exciting and demanding that our small group can scarcely handle them all. Apart from plans for new experiments already mentioned, an overall aim is to make cosmoclimatology an exact science. For example, cosmic-ray variation has greater climatic effects in some regions than in others. By considering these regional aspects, we should be able to interpret historical, recent and current climate change rather precisely and offer new input to climate models that still struggle to describe cloud behavior convincingly.

Perhaps we should also take a tentative step towards regional climate forecasting. If so, southern Asia is an obvious place to start, as Indian physicists report an apparent influence of cosmic rays on the vigor of the summer monsoon. The well-being of billions of people depends on coping year by year, sometimes with too little rain, sometimes too much, so predictions that are more reliable would be a very useful application of our style of climate science.

Falsifiability is the supreme test of competing hypotheses, illustrated here by the Forbush decreases, and in this case, our hypothesis survives. Attempts to falsify the anthropogenic-warming hypothesis run into the problem that the computer models appear to be infinitely adjustable. Perhaps the issue will be settled not by reasoned debate but by the unusual state of the Sun, which has shown remarkably little magnetic activity in the past 2 years.

We have come to a parting of the ways. Until recently, the arguments were about alternative explanations for the rise in temperature during the 20th century. Now the expectations are completely different. In keeping with the experience of the past 12 millennia, if the cosmic ray hypothesis is right we should now see definite cooling instead of anthropogenic warming. Again, patience is called for, as we wait for measurements by satellites, weather balloons and ocean buoys to deliver their verdict in the years ahead.

More to explore

Scientific papers by Henrik Svensmark and his colleagues are listed at: http://www.space.dtu.dk/English/Research/Research_divisions/Sun_Climate/Publikationer.aspx

Svensmark H. Cosmoclimatology: a new theory emerges. *Astronomy & Geophysics* 2007: 48(1): 1.18–1.24.

Svensmark H, Friis-Christensen E. *Reply to Lockwood and Fröhlich – the persistent role of the Sun in climate forcing.* Copenhagen, DTU National Space Institute, 2007 (http://www.space.dtu.dk/upload/institutter/space/research/reports/scientific%20reports/dnsc-scientific_report_3_2007.pdf).

Svensmark H, Calder N. *The chilling stars: a new theory of climate change.* Thriplow, Icon Books, and New York, Totem Books, 2007 (updated 2008). Also translated into Danish, Dutch, German, Hebrew and Swedish.

Information on *The cloud mystery*, Lars Oxfeldt Mortensen's TV documentary about Henrik Svensmark's work (http://www.sciencebits.com/thecloudmystery).

General online information from the Center for Sun-Climate Research of the DTU National Space Institute, with links to news releases: http://www.space.dtu.dk/English/Research/Research_divisions/Sun_Climate.aspx

The author

Henrik Svensmark: Professor with Special Responsibilities in Atmospheric Effects of Cosmic Rays and Head, Sun-Climate Research Center, DTU National Space Institute. PhD in superconductivity, DTU, 1987. Postdoctoral fellow, superconductivity, University of California at Berkeley, 1987–1990. Worked on fluctuation in nonlinear systems at the Niels Bohr Institute before joining the Danish Meteorological Institute in 1993. While there, he embarked on solar-terrestrial physics in 1996 and continued this upon joining the Danish Space Research Institute (now the DTU National Space Institute) in 1998.
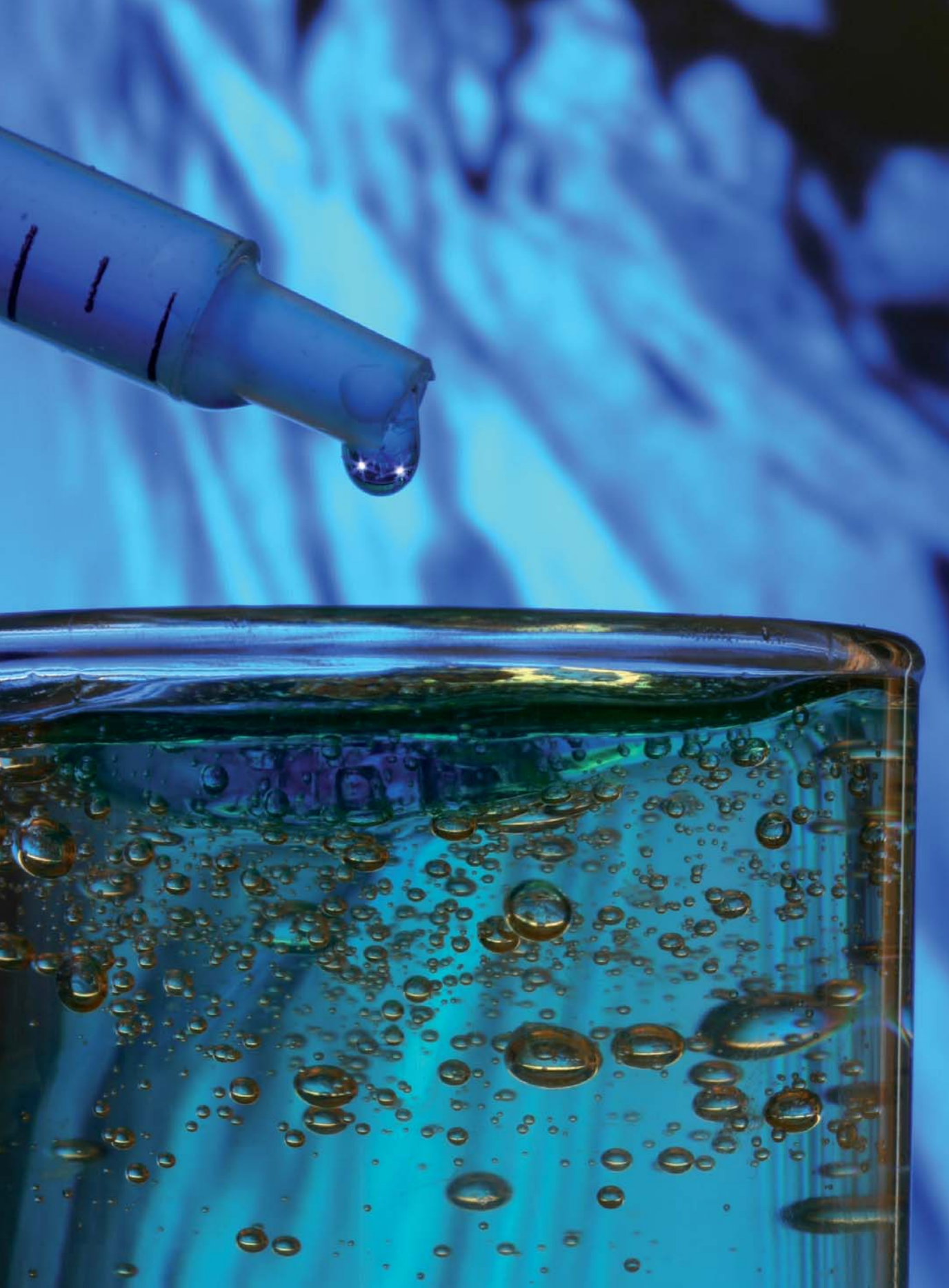
Process design and production of chemicals, food ingredients, fuels and pharmaceuticals:

multidisciplinary chemical engineering

Anne S. Meyer, John M. Woodley & Rafiqul Gani

Introduction

In the past century, chemical engineering has evolved from a means to design processes on a case-by-case basis to a mature field of engineering. Modern chemical engineering rests on established fundamental concepts such as engineering thermodynamics, unit operations, reaction engineering and systems engineering and now includes biochemical engineering and polymer technology. Chemical engineering uses fundamental tools such as mass and energy balances and process simulation to design and improve processes. Today chemical engineering stands at a hugely exciting point as it embraces the biological, nano- and polymer sciences as the new driving forces (alongside chemistry) that create new processes, materials and products. Chemical engineering is the epitome of a subject for a technical university since it applies the science base to create wealth and to help to solve some of the world's major problems in energy supply, health, environment and climate change. However, what makes chemical engineering so exciting

today is that the newest developments occur right at the interface of several disciplines and that application-oriented work promotes the development of new knowledge. Indeed, a hallmark of the modern profession is that the research integrates product and process requirements. This chapter presents our view of modern chemical and biochemical engineering with a few selected examples.

Chiral pharmaceutical intermediates

Pharmaceuticals are particularly difficult molecules to manufacture, and several challenges need to be addressed for effectively developing and designing processes. First, the target molecule needs to be made to specifications using a process developed in a very short timeline, brought about by the limited patent life of the product and the need to recover high discovery costs. The second challenge arises from the need to integrate new technologies, especially biotechnology. Most pharmaceuticals manufactured today (and those in the pipeline) are

small molecules, although an increasing number have more than one chiral center and are complex multifunctional molecules. Unsurprisingly therefore, biotechnology has a role in synthesizing such molecules via fermentation (to supply starting materials), microbial catalysis (to carry out complex conversions involving enzymes that cannot be isolated or requiring multiple steps) and immobilized enzymes (to create chiral center(s) using reusable catalysts). In all these cases, the biotechnological production methods need to be integrated with multi-step chemical synthesis and purification technology.

For a process to be economically viable, certain process metrics must be achieved. For biocatalytic processes, two are particularly important. The usual requirement is to achieve product concentrations comparable to chemical processes of at least 100 g/l. Given that, in nature, enzymes work at millimolar levels of reactant, enzymes will always operate away from their natural conditions. Process solutions as well as protein engineering are available to achieve such concentrations while maintaining sufficient enzyme activity. The other metric depends on the cost of the enzyme and is best expressed as the grams of product per gram or activity unit of biocatalyst. For commercial processes, this metric needs to be at least 2000 for an enzyme and 15 for a whole-cell process (reflecting the different costs of each). Techniques such as immobilization become important to provide the ability to recycle (especially for enzymes). At the DTU Department of Chemical and Biochemical Engineering, we are examining new ways to test catalysts that enable them to be exposed to defined reactor conditions. Laboratory tests in reactors that mimic industrial-scale operation allow parameters to be obtained that can be used directly in mathematical models linking catalyst properties with process performance. In this way, hypotheses can be tested on the likely process implications of modifying catalysts (for example, by protein engineering or changing an immobilization support) or changing the reactor or process design. The research involves developing suitable process models and devising a suitable method for testing biocatalysts that accurately mimics industrial-scale operation.

Platform chemicals from biomass

The increasing cost of oil is driving specific interest in producing new fuels from biomass, but producing chemicals from biomass, carbohydrate and sugar is undoubtedly equally important. Indeed, there are many potential future fuel sources aside from biomass. However, in a world with limited (or very expensive) oil, the origin of the chemicals of the future is less clear. Existing infrastructure is based on the use of the seven established platform chemicals: toluene; benzene; xylene; 1,3-butadiene; propylene; ethene; and methane. In the short term, we could consider whether we can use the same infrastructure and just create the seven chemicals from alternative sources. However, in the longer term new processes will need to be devised based on a different set of platform chemicals. One group will be based around glucose: the hydrolytic product of starch and cellulose and therefore readily available from biomass. Biorefineries will need to develop a structure that can manage a range of types of feedstock, a range of technologies and a range of products. This presents a considerable challenge for design and optimization and for process integration. For example, at the DTU Department of Chemical and Biochemical Engineering we are examining suitable process technology to convert glucose or fructose to the building blocks 5-hydroxymethylfurfural (HMF) or 2,5-furandicarboxylic acid (FDA) (Fig. 16.1). The greatest value is obtained by going all the way from glucose to FDA. Even in this small reaction pathway there are many alternative technologies. Some can be integrated together, some provide the required yield and selectivity, some are difficult to implement and some are untested at large scale. This illustrates very well the challenge that design engineers face. The major use of glucose is in food applications as a feedstock for sorbitol and high-fructose corn syrup, and it is also a major starting-point for technical as well as potable alcohol production. All the potential technologies to use glucose (whether approved for food or non-food production) need to be able to overcome the pH and temperature instability and limited solubility in organic solvents. Thus, the nature of glucose therefore means that one obvious starting-point is to use enzymatic catalysis (water-based and under mild conditions).
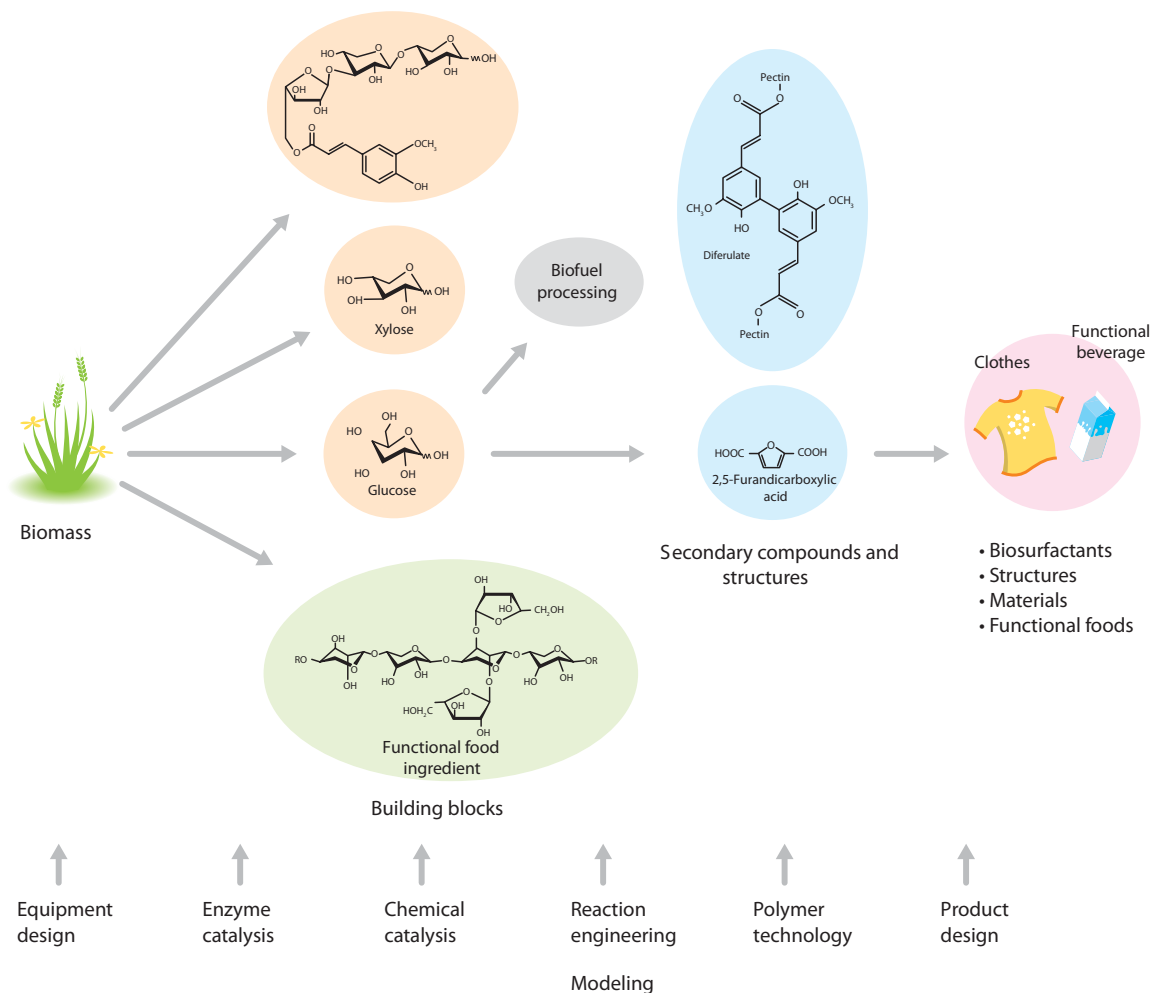
Fig. 16.1. Integration of disciplines in product and process development: developing chemicals, biofuel, materials and food ingredients from biomass

Nature produces vast amounts of biomass driven by sunlight via photosynthesis:

$$nCO_2 + nH_2O \rightarrow (CH_2O)_n + nO_2.$$
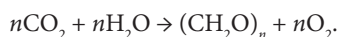
However, using biomass for producing chemicals and fuels is still in its infancy, with only 3.5% being used for food or non-food purposes. Plant biomass consists mainly of carbohydrate, lignin, protein and fat. Of an estimated 170 billion metric tons of biomass produced every year, roughly 75% is carbohydrate, which makes biomass carbohydrate the most abundant renewable resource. Together with their amenability to enzymatic processes, this makes carbohydrates the center of attention for new and greener feedstocks to replace petroleum for producing commodity chemicals as well as fuel.

Challenges in upgrading plant biomass

In plant biomass, most of the carbohydrate is stored as sugar polymers such as starch, cellulose or hemicellulose. Starch is the second largest type of biomass produced and is commonly found in corn, wheat, rice and potatoes. The total world production in 2004 was 60 million tons, of which more than 70% came from corn. Starch is industrially hydrolyzed to glucose by the three enzymes: α-amylase, glucoamylase and pullulanase.

Cellulose is a glucose polymer consisting of linear chains of glucopyranose units. Unlike starch, cellulose is a crystalline material in which inter- and intramolecular hydrogen bonding gives rise to the very stable cellulose fiber. Hemicellulose is a common name used for different types of polysaccharides consisting of short, highly branched

chains of carbohydrate units, including five- as well as six-carbon units such as xylose, galactose, glucose, mannose and arabinose. Hemicellulose is much easier to hydrolyze than cellulose. The structured portion of biomass, such as straw, corn stover, grass and wood, is made of lignocellulose, composed mainly of cellulose (30–60%), hemicellulose (20–40%) and lignin (10–30%). Both cellulose and hemicellulose consist of carbohydrate components, whereas lignin is a highly branched aromatic polymer.

Intensive research is being conducted on the use of lignocellulosic raw material as a biomass source for producing chemicals and fuel. However, this research still faces considerable challenges because lignocellulose is remarkably resistant to hydrolytic and enzymatic attack. Current technologies involve energy-demanding thermal pretreatment of lignocellulose to break down the extremely stable cellulose–hemicellulose–lignin composites before adding cellulose-hydrolyzing enzymes. The energy required for this pretreatment is a barrier that presents one of the most expensive processing steps in converting cellulosic biomass to ethanol, thus requiring significant improvement. At the DTU Department of Chemical and Biochemical Engineering, we are examining novel pretreatment strategies, including both novel equipment design and novel approaches to the pretreatment and enzymatic hydrolysis reaction designs to convert the cellulose to glucose. A high yield of glucose from the biomass without excessive levels of undesirable byproducts is a major prerequisite in the whole biomass-valorization chain (Fig. 16.1). The work requires 1) advanced understanding of the physicochemical behavior of the lignocellulosic feedstock, 2) detailed insight into the structural features of the biomass that retard liquid penetration and challenge the enzymatic attack on the insoluble cellulosic substrate and 3) practical experiments employing new innovative equipment designs. Hence, developing novel technology at the interface between classical chemical engineering and biochemical engineering is necessary to enable the efficient use of lignocellulosic materials. Classical chemical engineering can provide solutions related to thermal processing concomitant with physical treatment and large-scale equipment design, and biochemical engineering can provide

novel enzyme solutions and bioreactor design (Fig. 16.1). Given the great abundance of lignocellulose, upgrading lignocellulose will undoubtedly become an attractive option in the future. It is therefore important that we continue to develop processes that can economically convert lignocellulose into alternative fuel and chemicals.

Health-promoting food ingredients from plant biomass

Valorizing byproduct streams resulting from large-scale agroindustrial processes such as starch, sugar, plant oils, wine, beer and fruit juice processing is becoming vital for ensuring the competitiveness of agriculture-based industry. Upgrading these low-value byproducts can also assist in offsetting the prospective increasing costs and environmental concerns of disposing of the residue and can provide an important contribution to improving bioresource utilization. Dietary fiber and prebiotics are indigestible dietary carbohydrate that can promote health by supporting the growth of beneficial bacteria in the digestive tract and via other mechanisms. Byproduct streams such as those mentioned are rich sources of indigestible plant fiber and thus prospective starting materials for obtaining bioactive carbohydrate and, in certain cases, even antioxidants that can be upgraded for use as food ingredients with potential health benefits (Fig. 16.1). Target biofunctional structures can be produced by using selective enzymatic reactions, including custom-made mixtures of monoactive plant polysaccharide–modifying enzymes (available thanks to the progress in genomics). New oligomers can be obtained either by selective hydrolysis of plant fiber material or built by the enzyme-catalyzed conjunction of mono- and oligomers. The provision of well-defined plant fiber structures may even provide an improved foundation for understanding and predicting the putative health potential of different types of dietary fiber and prebiotics. From an engineering viewpoint, the work involves a holistic biorefining approach including detailed insight into the raw material composition and the prevalent bonds present (requiring detailed analysis), innovative, selective deconstruction of the material to "harvest" the desired structures (we modify enzymes for this) and separating and purifying

the target substances. The research thus involves developing targeted biocatalysis technologies, including applying designed mixtures of monoactive enzymes and new applied enzyme technology principles such as sequential enzyme dosing and using alternative bioreactor designs to handle highly viscous processing streams. The research thus requires detailed knowledge about the molecular features of plant materials and advanced analytical methods and equipment, in addition to using state-of-the art reactor and separation technologies and equipment. This type of research represents a demanding and costly, interdisciplinary endeavor but is a modern example of exploiting life sciences through technology to benefit society. The chain approach is simply a prerequisite for innovative developments leading to improved bioresource utilization.

Managing complexity through a systems approach

Process systems engineering promotes solving a problem systematically. In this way, although the chemical engineering community has traditionally applied process systems engineering to solve problems for the oil and petrochemical industries, its potential application range is much wider. This is because the word "process" also implies the process of solving a problem; design of a biochemical and biological process for converting biomaterial to specific chemicals; and the process of finding and designing chemicals with desired properties. Product-process design and development in the life sciences and pharmaceutical, food and related industries, in contrast to the oil and petrochemical industries, mainly depends on experiment-based trial-and-error approaches. Further, unlike the oil and petrochemical industries, in the life sciences, pharmaceutical, food and related industries, problems associated with product-process design and development involve the following distinct features.

- *Multiple scales*. Important data related to the chemicals come from different sources at different scales of time and size. For example, the properties that define the product characteristics are based on the microstructure of the molecule or material, whereas the process behavior that

needs to be monitored and controlled during operation is defined by the macroscopic (end-use) properties of the chemical system.
- *Multidisciplinary*. Converting the biomaterial through biocatalysis requires knowledge of organic synthesis, enzymes, reaction catalysis, bioreactor design and operation. Information about these topics comes from different disciplines.
- *Computer-aided techniques*. A lack of models to predict the behavior of the chemicals at different scales, of enzymes during organic synthesis and of reaction kinetics means that appropriate model-based computer-aided techniques have not been developed, and using experiment-based techniques is the only option.

Advances have been made on each of these issues in the specific field of chemical and biochemical engineering. For example, multiscale polymerization reactors have been developed to investigate the operation of reactors; technical and economic assessment related to the sustainability of biofuel has been conducted using data from engineers, economists and scientists; and computer-aided systems have been developed to perform routine mass and energy balances in chemical and biochemical processes. The demand for improved chemical-based products made from more sustainable raw material resources and for using more efficient processes to make them, however, requires tackling these issues and others in an integrated manner. This means that methods and tools suitable for current and future product-process development need to manage complex situations that require processing data and knowledge from different sources and at different time and size scales. That is, the scope of the problems that need to be solved has increased. A systems approach that can efficiently manage the complexity is therefore an absolute prerequisite for achieving real progress in several fields within modern chemical and biochemical engineering.

Fig. 16.2 highlights the multidimensional and multiscale nature of problems. At the micro and meso scales, the related problems are dealing with the microstructure of the molecules or materials and their properties; at the macro scale (the traditional field of application of chemical engineering), the related problems are mainly dealing with the
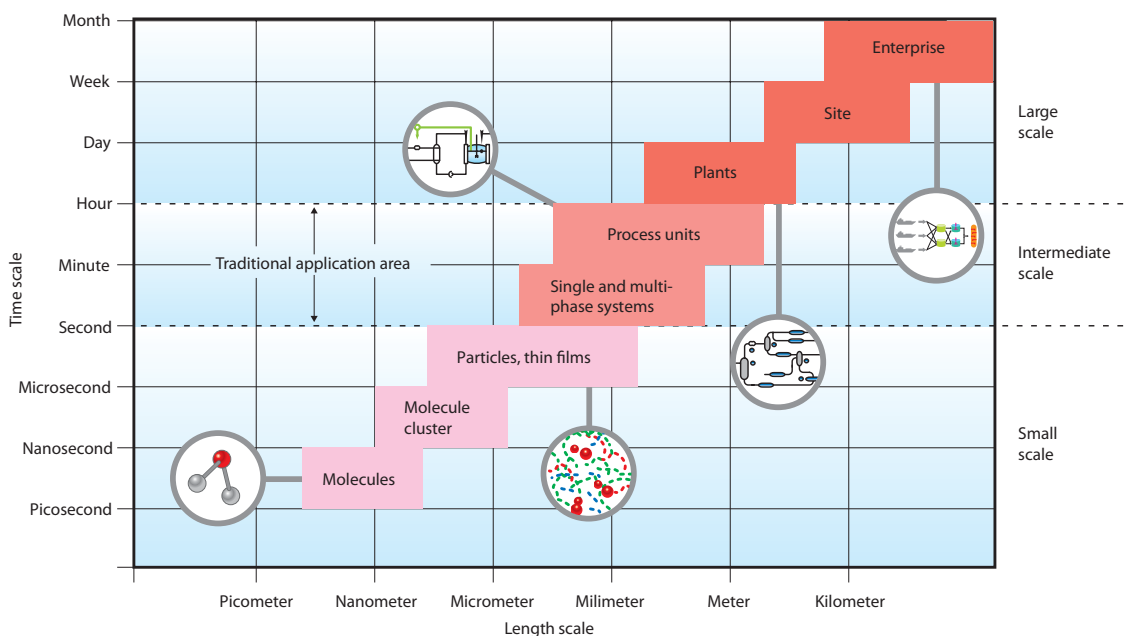
Fig. 16.2. Multiscale nature of product-process design problems

process and its operation to produce a desired chemical; at the mega scale, the related problems include dealing with company-wide optimization and supply chain issues.

To manage the complexity, a systems approach would develop a framework (the architecture of the software) for handling the diverse set of methods and tools needed to solve a wide range of problems for a potential computer-aided system. Such systems need to have:

- a knowledge base of data: for example, of the active ingredients, solvents, polymers, etc.;
- a library of models: for example, models to predict the properties – in case data are not available – of active ingredients, solvents, polymers, etc.; models to predict the controlled release from the microcapsule; and models to predict the behavior of the mixing process;
- a design method: for example, guiding the engineer or scientist through the sequence of steps needed to identify the best solution; and
- other associated methods or tools: such as a tool to analyze data; a tool to create the missing model; and a tool to screen feasible alternatives.

The main idea is to break down a complex problem into a set of subproblems that are easier to solve and to identify those that can be solved through model-based solution approaches. Solving these subproblems according to a predetermined sequence helps to reduce the search space through each subsequent subproblem solution, until a subproblem cannot be solved by using models anymore. At this point, the experiment-based trial-and-error approach takes over to determine the final solution. The advantage of this combined hybrid (systems approach) is that, during the early stages, when enough data and models are available (or could be easily generated), the search space is rapidly reduced. In the later stages, when quantitative values become important and data and models become more unreliable, the experimental resources are used, sometimes only to evaluate a few feasible alternatives to identify the truly innovative and best solution. An example of a computer-aided system that has been developed at the DTU Department of Chemical and Biochemical Engineering to manage this complexity is the Virtual Product-Process Design Laboratory.

The Virtual Product-Process Design Laboratory

The idea is the following: instead of performing the experiments needed to search for a product and its process to manufacture it, the engineer or scientist performs virtual experiments through the Virtual Product-Process Design Laboratory software. The software therefore contains:

- a large knowledge database of chemicals, solvents, plants, microcapsule devices, etc.;
- a large collection of models: for predicting properties, for controlled release, for mixing, etc.;
- design algorithms: methods for designing formulations, methods for designing molecules, methods for designing polymers, methods for synthesizing process flowsheets etc.; and
- other tools: software for predicting properties; software for generating models; software for designing equipment; and software for designing experiments, etc.

All of these are organized through a framework for efficiently managing the complexity. Fig. 16.3 gives an overview of the main features of the Virtual Product-Process Design Laboratory software, which has been used in designing and evaluating the controlled release of an active ingredient (codeine) through a polymeric microcapsule. In the first step, the problem is defined (the "documentation" box of the Virtual Product-Process Design Laboratory provides the identity of the active ingredient, the desired controlled release parameters etc.). In the second step, the application source (codeine released into the body) and the primary properties of the solvent and the polymer (needed by the controlled release model) are selected. If the user cannot provide this information, methods for designing solvents and polymers are used to generate a list of candidates from which to select. In the next step, the functional properties needed to evaluate the controlled-release design are selected and
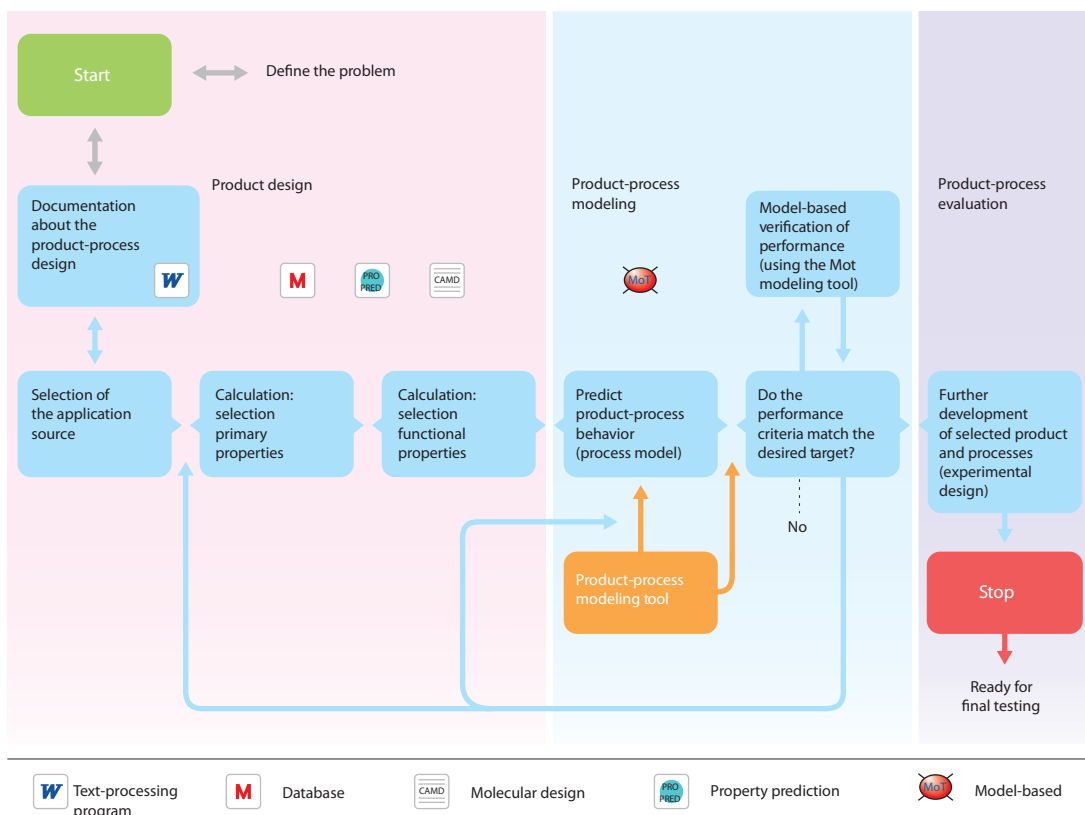

Fig. 16.3. Virtual Product-Process Design Laboratory

calculated. If models are not available, the modeling software helps to generate new models. In the next steps, the product performance model is used to predict the product behavior. If the desired (target) performance is matched, then the last step of verifying the product performance through experiments is performed. If the target is not matched, the process can be repeated from any of the earlier steps with a new design alternative.

Note that multiscale models have been used, data and knowledge from different disciplines have been used and design and evaluation problems have been effectively used by solving a collection of subproblems according to a predetermined sequence. The final step (not shown) would be to select a few of the alternatives and perform the experiments necessary to validate the selection. The experiments are therefore done not to design the product but to verify the product. This approach has the potential to save time and money in bringing a chemically based product to the market. The accuracy and range of application of the Virtual Product-Process Design Laboratory software depends on the available data and models in the software.

Conclusions

Chemical engineering is an established field with a century of development. For much of that century it has been taught at DTU, from its origins in technical chemistry to the modern department with a Virtual Product-Process Design Laboratory and controllable and automated pilot-scale equipment. Although computer-based models can reduce the cost and time of such experimentation, the final validation is necessary and requires significant resources for pilot-scale experimentation and testing. Much work is still needed, however, to solve the many challenging problems in product-process design from the life sciences and pharmaceutical, food and related industries. The future of chemical engineering requires an interdisciplinary approach to provide the base for developing new solutions. Indeed, the integration of several disciplines is a hallmark of modern chemical and biochemical engineering that will continue to foster new and exciting research platforms that match the challenges of modern society and can foster new opportunities.

More to explore

Pollard DJ, Woodley JM. Biocatalysis for pharmaceutical intermediates: the future is now. *Trends in Biotechnology* 2007: 25: 66–72.

Dale BE. Greening the chemical industry: research and development priorities for biobased industrial products. *Journal of Chemical Technology and Biotechnology* 2003: 78: 1093–1103.

Meyer AS, Dam BP, Lærke HN. Enzymatic solubilization of a pectinaceous dietary fiber fraction from potato pulp: optimization of the fiber extraction process. *Biochemical Engineering Journal* 2009: 43: 106–112.

Gani R. Chemical product design: challenges and opportunities. *Computers and Chemical Engineering* 2004: 28: 2441–2457.

Rodriguez RM. *Computer-aided multiscale modeling for chemical product-process design*. PhD dissertation. Lyngby, DTU Department of Chemical and Biochemical Engineering, 2009.

Anne S. Meyer:

PhD, DTU, 1993. Professor, Head, Center for BioProcess Engineering, Department of Chemical and Biochemical Engineering, DTU. Research: applied enzyme technology and kinetics and plant biomass upgrading processes. Member, Danish Academy of Technical Sciences and received the Danisco Award in 2005 and the Grundfos Award in 2008.

John Woodley:

PhD, Biochemical Engineering, University College London,1988. He worked for ICI Ltd and was a lecturer, reader and professor at University College London in Biochemical Engineering until 2007. Professor, Department of Chemical and Biochemical Engineering, DTU, 2007–. About 120 publications and 160 conference presentations. Awarded a DEng from the University of Manchester. Fellow, Institute of Chemical Engineers, UK.

Rafiqul Gani:

PhD, Imperial College, London. Department of Chemical and Biochemical Engineering, DTU, 1985–. Currently Professor and Director, Computer Aided Process Engineering Center. More than 150 publications in peer-reviewed journals and numerous invited conference presentations. Editor-in-Chief, *Computers & Chemical Engineering*. Member, Danish Academy of Technical Sciences and Fellow, Institute of Chemical Engineers, UK and American Institute of Chemical Engineers. Member, Executive Board, European Federation of Chemical Engineering.

Nutrition, genetics and health

Inge Tetens & Ulla Vogel

A healthy diet can be described as foods or nutrients. A nutrient is defined as a chemical substance that is needed by any organism to grow or to maintain life. Nutrition recommendations were previously generally given as foods known to prevent diet-related diseases, such as scurvy. The increase in knowledge and the mounting evidence of the role of important nutrients and other food components has led to new possibilities for developing evidence-based dietary guidelines for preventing disease and promoting health.

Nutrient recommendations are derived by measuring the average requirements of a nutrient plus a safety margin ensuring that 97–98% of the healthy population is covered. Nutrition recommendations are translated into food-based dietary guidelines that are developed by integrating the scientific knowledge about the relationships between individual nutrients and foods and health and about nutrient intake recommendations compared with the observed dietary intake of nutrients and foods in populations. The food-based dietary guidelines are conceptually easier to understand

for the final consumers and can be used in developing public health practice guidelines (Box 17.1).

The immediate goal of food-based dietary guidelines is to promote health by providing consumers with advice on food choices and eating patterns. The long-term goal is to provide advice on optimal food choices and eating patterns to maintain health, to improve the quality of life, to reduce the risk of disease and potentially to increase longevity. Awareness is increasing among scientists and politicians that such advice should be evidence-based.

Our approach in providing the most updated optimal and evidence-based food-based dietary guidelines starts in some of the basic disciplines of human nutrition, including measuring dietary intake and developing and using biomarkers of food and nutrient intake and metabolism and exposure to food contaminants. A new and upcoming approach is to study the interplay between nutrition and genetics using molecular epidemiology in relation to the risk of disease to obtain more detailed knowledge about the interaction between diet and genetics.

Box 17.1. The 6-a-day campaign

The 6-a-day campaign was launched in Denmark in 1999 by the 6-a-Day Partnership, a public-private partnership with representatives from government agencies, nongovernmental health organizations and the fruit and vegetable industry. The major objective of the campaign was to create awareness and encourage all Danes to eat and enjoy eating 6 or more servings of fruit and vegetables every day (Fig. 17.1). The long-term goal of the campaign was to reduce the risk of certain lifestyle-related diseases, especially cardiovascular disease and certain types of cancer.

Fig. 17.1. Posters used as part of the 6-a-Day campaign (courtesy of the 6-a-Day partnership and Grønttorvet København A/S)

The background for the 6-a-day campaign was the results of the two nationally representative dietary surveys conducted in Denmark by the National Food Agency (now the DTU National Food Institute) in 1985 and 1995, respectively. The surveys showed that an average Dane consumed much less fruit and vegetables than recommended by the scientific community.

Has the 6-a-day campaign been successful? According to our data from the National Survey of Dietary Habits and Physical Activity, the trends are clear. The average consumption of fruit and vegetables among adults increased about 100 g per day from 1995 to 2000–2001. However, the latest survey suggests that this increase has stagnated for both adults and children (4–15 years old) (Fig. 17.2) and that only 16% of the adult population meets the recommendation of 600 g per day.

The 6-a-day case highlights the need for continual surveillance of food intake and for knowledge on the relationships between food, nutrition and health.

Fig. 17.2. Only 16% of the adult population meets the 6-a-Day recommendation of 600 g of vegetables and fruit per day

Measurements of food intake

From a public health viewpoint, the major reason for measuring food intake is to evaluate the adequacy and safety of the food people eat. The objective of virtually all dietary assessment is to estimate the habitual or average long-term intake for the group or individuals of interest. Habitual represents what is usual in the long term. However, measuring habitual intake is extremely difficult because food intake varies widely from day to day and over a longer period. Measuring food intake poses several inherent problems that require different methods depending on the objectives of the study. Measuring food intake is not a simple measure of only one variable but requires data on the intake of many different food items.

Many different approaches can be taken to measuring food intake at different levels. A quantitative record of food intake (such as a record of foods and beverages consumed during several days by individuals) is suitable for assessing differences between individuals and demographic subgroups and seasonal and/or geographical differences. We have

adapted a 7-day dietary record as the most suitable method for evaluating the adequacy of the diet in Denmark. A representative population sample 4–75 years old fills out a questionnaire (Fig. 17.3) on 7 consecutive days and reports all foods and beverages consumed. The main objective of this survey is to identify the proportion of the population that may have inadequate intake of energy and nutrients. Another objective of the survey is to provide valid data for food safety assessment. The dietary intake information must be interpreted in the light of appropriate biological measures of nutritional status. For the food safety assessment, it is of great importance that the dietary data obtained reflect the habitual level of intake of the foods.

The measurements and the evaluation of food intake are likely to become more complex in the future, given the increasing diversity of the food supply and the increasing recognition of the need to be able to assess accurately the intake of food and nutrients and of non-nutrient constituents of foods and dietary supplements. The ever-more-complex food supply requires that the consumers

Fig. 17.3. A DTU National Food Institute survey aimed at identifying the proportion of the population who may have inadequate intakes of energy and nutrients. A representative sample of Denmark's population 4–75 years old fills out a questionnaire on 7 consecutive days and reports all foods and beverages consumed.

be able to describe the food in adequate details for the study purposes. New technological developments will therefore be required for developing future dietary assessment methods.

New measures of dietary intake, such as indexes or nutrient profiling systems, will be developed as a new tool to evaluate this intake. With the increasing number of foods and food constituents of in-

terest in relation to health, appropriate biomarkers for food intake and health or preventing the risk of disease need to be developed.

Food composition tables

Compiling reliable and high-quality food composition tables is an important instrument for measuring and evaluating dietary intake. At the DTU National Food Institute, we compile the Danish Food Composition Table (http://www.foodcomp.dk), freely available for users at the international, national, household and individual levels. The Food Composition Table comprises about 1050 selected foods with data on the content of energy and 111 nutrients. On average, each food has 72 nutrient values altogether. The table includes all foods that form most of the food supply in Denmark and that constitute major contributors to the diet in the forms most commonly obtained. The data are based on original analytical values, imputed values derived from analytical values from a similar food, calculated values derived from recipes and from ingredients and borrowed values from other tables and databases.

No food database system fulfills the needs of all compilers and users of food databases. Recent international collaboration has considerably improved the development and compatibility of food composition data. However, the future requires greater consideration and expansion of the existing food database as new innovative food products are introduced, including new functional foods, fortified foods and novel foods expected on the future food market.

We have developed a unique program at the DTU National Food Institute. The General Intake Estimation System (GIES) links data on food intake, food recipes and food composition tables (Fig. 17.4). Along with the expanding demands for the use of data collected in dietary intake research, we experience increasing demand for flexible programs such as GIES to fulfill new research needs. This includes especially the dietary intake and adverse effects of food and nutrients in which the relationship between diet and health is studied. Such studies investigate both health-promoting and adverse effects of the diet or dietary constituents and contaminants.

Toxicological studies at the DTU National Food Institute are used to identify toxic substances in foods or food contact materials using, for example, rodent models and cell lines. In these studies, we identify food contaminants in specific food items such as acrylamide in carbohydrate-rich foods (such as French fries and potato chips). The rodent models are often used to establish dose–response relationships that are useful for regulatory authorities in regulating. The recent development of biochemical and genetic biomarkers that can be studied in human populations makes it possible to link knowledge about the specific intake of food items in a given population and the knowledge about the content of harmful substances in the food items with the occurrence of diseases in population-based epidemiological studies.

The study of nutrition and disease

Molecular epidemiological studies have emerged as a new research discipline in recent decades and constitute a useful tool in studying the complex interplay between nutrition, lifestyle, genetics and disease. These studies investigate the distribution of environmental and nutritional biomarkers and genetic variation between diseased and healthy people to identify potential risk factors for the disease in question. Molecular epidemiology as a research field has increased enormously since the human genome was sequenced in 2001. The advantage of the field is that the association between food and human health is studied directly. The major disadvantage is that epidemiology only identifies associations and thus no causal relationships. The associations found therefore have to be confirmed in controlled settings such as human intervention studies.

Definition of genetic variations and polymorphisms

The human genome comprises 3 billion base pairs distributed on 23 chromosomes. Humans have two copies of the genome, inheriting one copy from each parent. On average, one of every 1000 base pairs in the genome varies between individuals, resulting in 3 million potential differences between two people (Fig. 17.5). The genetic variation in-
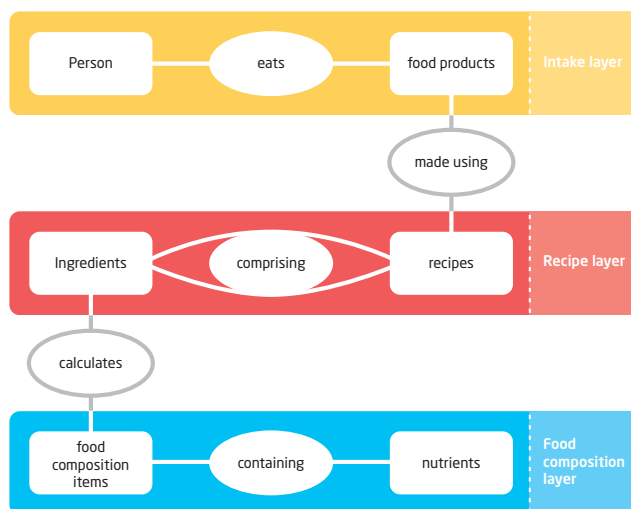

Fig. 17.4. The General Intake Estimation System (GIES) links data on food intake, food recipes and food composition tables

cludes either single nucleotide polymorphisms, which constitute most the genetic variation, or deletions or insertions of more than one base pair. Deletions or insertions may cover large DNA fragments, deleting whole genes. By definition, polymorphisms have an allele frequency exceeding 1%; genetic variation with allele frequencies lower than that is termed mutations.

A functional effect of the polymorphism is a prerequisite for a biological effect. Consequently, much of the current molecular epidemiological research aims at identifying the functional polymorphisms and their interaction with environmental factors such as food intake.

Most of the polymorphisms that have functional effects are either promoter polymorphisms or missense polymorphisms causing an amino acid substitution in the coding region of a gene. The resulting protein will therefore have another amino acid in the given position, and this may cause changes in the catalytic property of the enzyme. A further class of biologically active polymorphisms changes the stability or the splicing of the messenger RNA (mRNA). This usually results in lowered amounts of intact mRNA, which in turn gives lower expression levels of the encoded protein.
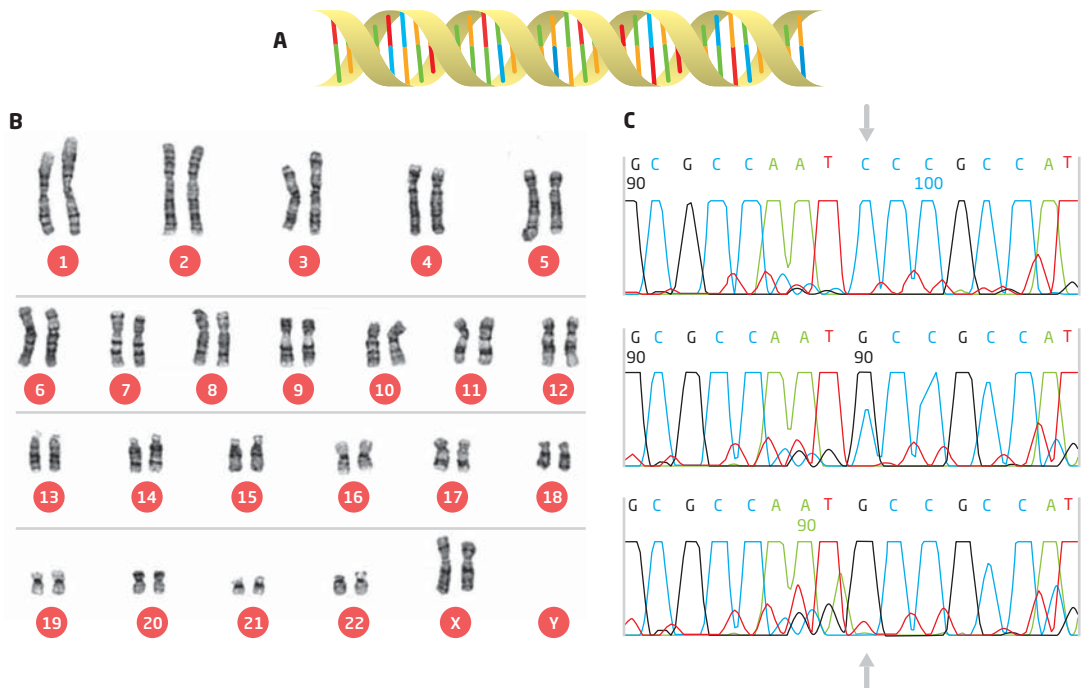
Fig. 17.5. Human DNA. **A**. The human genome consists of 3 billion nucleotides. **B**. Each person has two copies of the genome distributed on 23 chromosomes, resulting in 46 chromosomes. **C**. When the DNA from one individual is sequenced, both copies are sequenced simultaneously. The person in the middle has two different nucleotides at the position marked by the arrows, indicating the presence of a single nucleotide polymorphism.

Interactions between nutrients and genetics

We have worked with nutrition and genetics using the large, prospective Diet, Cancer and Health cohort of 57,053 Danes established by the Danish Cancer Society in 1993–1997. At enrollment, detailed information on diet, smoking habits, lifestyle, weight, height, reproduction status, health care and other socioeconomic characteristics and environmental exposure were collected. Blood, urine and fat tissue were sampled and stored at –150°C. The participants had no cancer diagnosis when entering the cohort, thus ensuring that all information on lifestyle was given before any cancer diagnosis and therefore unbiased. Moreover, the biological samples were collected from healthy individuals who would later develop cancer.

Interaction between nutrition and genetics in cancer risk: alcohol-related breast cancer

Alcohol is a well-known risk factor for breast can-
cer. Alcohol intake is associated with a 7% increased risk of breast cancer per 10 grams consumed per day, and breast cancer constitutes an estimated 60% of the alcohol-induced cancer worldwide. Women in Denmark have a high alcohol intake compared with women in other European countries, and breast cancer is the most common cancer form among women in high-income countries, including Denmark. One of nine Danish women develops breast cancer before the age of 75 years.

The mechanism behind alcohol-induced breast cancer is not clear. Both ethanol and the primary metabolite acetaldehyde are classified as human carcinogens. The mechanism behind alcohol-induced breast cancer has not been fully elucidated. Ethanol has been proposed to induce breast cancer by several different mechanisms: 1) by increasing the level of the female sex hormone estrogen, 2) by disturbing folate metabolism and 3) by the metabolite acetaldehyde acting as a mutagen.

We searched for gene–environment interactions between genetic variation and alcohol intake in relation to breast cancer risk to try to understand how alcohol causes breast cancer. We found new evidence that the peroxisome proliferator-activated receptor-γ2 (PPARγ2) is implicated in alcohol-mediated breast cancer. The gene *PPARγ* encodes the peroxisome profilerator-activated receptor γ, a transcription factor involved in differentiating fat cells and controlling the energy balance. The *PPARγ* Pro12Ala polymorphism in the *PPARγ2* isoform results in a proline-to-alanine amino acid substitution. In vitro studies have shown that the variant *PPARγ2* [12]Ala transcription factor gives 30% less transcriptional activation of target genes.

We found that variant allele carriers were at lower risk of breast cancer (incidence rate ratio = 0.67, 95% confidence interval (CI): 0.46–0.97) in a case–control study of 423 postmenopausal women with breast cancer cases and matched controls nested within the prospective Diet, Cancer and Health Study. The lowered risk was caused by an interaction between genotype and alcohol intake. Alcohol consumption was associated with a 1.21-fold increased risk of breast cancer per 10 grams of alcohol per day (95% CI: 1.06–1.35) among homozygous wild-type carriers, whereas alcohol was not associated with breast cancer risk among variant allele carriers (*P* for interaction = 0.005) (Fig. 17.6). This means that, for homozygous carriers of the wild-type pro-allele, the intake of 10 grams of alcohol per day (less than one drink a day) is associated with a 20% increased risk of breast cancer. Thus, the result indicated that high *PPARγ* activity promoted alcohol-related breast cancer, whereas low *PPARγ* activity prevented alcohol-related breast cancer. Commonly used over-the-counter headache pills such as ibuprofen are known to activate *PPARγ*. They are classified as nonsteroidal anti-inflammatory drugs (NSAID). We therefore tested whether the use of NSAID changed the interaction between *PPARγ* Pro12Ala and alcohol intake in relation to breast cancer risk (Fig. 17.7). NSAID use led to increased risk of alcohol-related breast cancer, primarily among variant allele carriers. Thus, NSAID use increased the carcinogenic effect of alcohol in relation to breast cancer, and especially among variant allele carriers, who were otherwise not at risk of alcohol-related breast cancer.
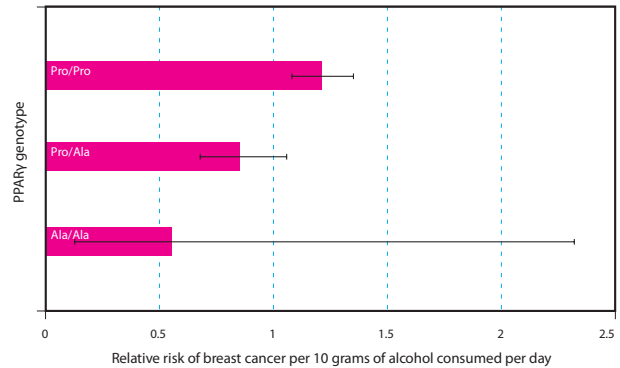

Fig. 17.6. Relative risk of breast cancer per 10 grams of alcohol consumed per day in relation to the polymorphism *PPARγ* Pro12Ala among 361 postmenopausal women with breast cancer and 361 matched controls. For Pro/Pro carriers, the intake of 10 grams of alcohol per day (almost one standard drink) is associated with a 20% increase in the risk of breast cancer. For carriers of the other genotypes, alcohol intake is not associated with breast cancer risk, as the risk estimates are below 1. The difference in the rate of alcohol-related breast cancer between genotypes was statistically significant (*P* = 0.01). Seventy percent of women in Denmark are Pro/Pro carriers; 28% carry one Ala allele, and 2% carry two Ala alleles.

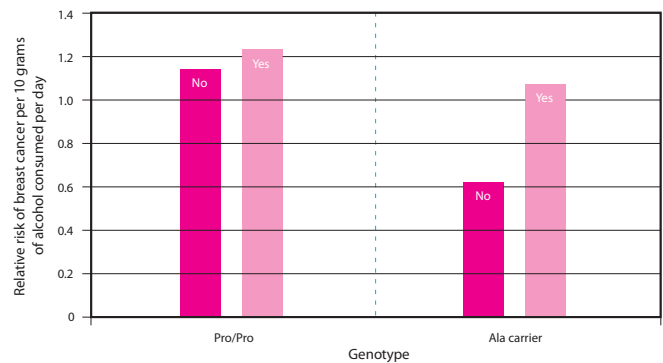Fig. 17.7. Interaction between *PPARγ* genotype and use of NSAID in relation to alcohol-related breast cancer. The rate of breast cancer per 10 grams of alcohol consumed per day is shown for the combination of NSAID use and *PPARγ* genotype. Variant allele carriers have lower *PPARγ* activity, whereas NSAID use increases *PPARγ* activity. The higher the *PPARγ* activity, the higher risk of alcohol-related breast cancer.

We tested this association in the whole Diet, Cancer and Health cohort, where we had data on both alcohol consumption and use of medication. We found that NSAID use among women with a moderate alcohol intake of 5–13 g/day was associated with a 1.75-fold increased risk (95% CI: 1.31–2.36) of breast cancer compared with non-users of NSAID with a similar alcohol intake. Among women with a low alcohol intake, NSAID use was not associated with an increased risk of breast cancer. Thus, we may be able to explain why NSAID use is associated with increased breast cancer risk in Denmark, whereas in countries in which women drink less alcohol, NSAID use protects against breast cancer. Many nutrients, foods, environmental pollutants and medicines activate *PPARγ*, and it is hoped that future research will elucidate whether these agents also increase the risk of alcohol-related breast cancer. The example illustrates that molecular epidemiology is a powerful tool in the study of the complex interplay between lifestyle and genetics in relation to human disease.

Polymorphisms as biomarkers

We used the prospective design of the Diet, Cancer and Health cohort to answer fundamental questions about risk factors for cancer and for the important development of genetic and biochemical biomarkers of exposure, metabolism and disease. Below is an example in which a polymorphism is used as a biomarker of exposure.

Glutathione peroxidase 1 is encoded by the gene *GPX1*. It is part of the antioxidant defense and neutralizes hydrogen peroxide, which is formed when electrons escape from the electron transport chain in the mitochondria. Hydrogen peroxide is also formed from water whenever reactive oxygen species are formed. This is termed oxidative stress. Oxidative stress has been proposed to be a risk factor for cancer, especially breast, lung and colon cancer.

The GPX1 Pro[198]Leu polymorphism results in an enzyme in which a proline residue at position 198 in the enzyme is replaced by a leucine residue, resulting in a 30% reduction in enzymatic activity. The allele frequency of GPX1 Pro[198]Leu is about 0.3 among Caucasians; 10% of a Caucasian population carry two copies of the variant allele and 40% one copy of the variant allele and one copy of the wild-

type allele. The polymorphism is thus an example of a commonly occurring missense polymorphism resulting in a less efficient enzyme. Glutathione peroxidase activity in red blood cells measured in 754 postmenopausal women correlated with GPX1 Pro[198]Leu, such that glutathione peroxidase activity decreased 5% per [198]Leu copy. This indicates that the GPX1 [198]Leu enzyme is also less catalytically active than the wild-type [198]Pro enzyme *in vivo* and that this results in lowered glutathione peroxidase activity level in red blood cells even though two other glutathione peroxidase enzymes are present in red blood cells. Carriers of the variant *T*-allele of *GPX1* Pro198Leu had a 43% higher risk of breast cancer than homozygous wild-type allele carriers (95% CI = 1.07–1.92). The association between glutathione peroxidase activity and breast cancer risk was not statistically significant. This supports the notion that oxidative stress is a risk factor for breast cancer. Moreover, the polymorphism turned out to be a stronger risk factor for breast cancer than the glutathione peroxidase activity in the single blood sample that was available. This means that the glutathione peroxidase polymorphism better indicated a person's overall glutathione peroxidase activity than the single measurement of glutathione peroxidase activity. This may be because several factors influence the glutathione peroxidase activity, including alcohol intake, sex, intake of fruit and vegetables and smoking.

The DNA required for the polymorphism is much more readily available (for example, from urine, hair or mouth swab) than red blood cells, making the glutathione peroxidase polymorphism an attractive biomarker of the antioxidant defense in future studies. With about 3 million genetic variations in the human genome and an estimated 50,000 of the polymorphisms having a biological effect, establishing one genetic biomarker is a very small step forward to understanding the complex interplay between food, nutrition and genetics in relation to health and disease.

Concluding remarks

The multidisciplinary nature of the science of food and health is a research challenge that should be overcome to effectively influence the food and health of individuals and populations. Modern ad-

vances in technology and innovative tools to further improve the quality of food intake data together with advances in molecular epidemiology will enhance the translation of information on food intake into associations between intake or exposure and health or disease risk. New investment and expansion of the national food databases and further development of high-throughput techniques for the polymorphism and gene analysis combined with advances in the statistical handling of larger amounts of data are much needed. It is very important that epidemiological studies that identify associations be followed by human intervention trials to prove causal relationships. Such procedures will ensure the basis for evidence-based dietary guidelines for the future.

The magnitude and complexity of the future food and health challenges require concerted multidisciplinary and multisectoral efforts from everyone involved. Specific needs for the future include:

- conducting cross-sectional surveys to understand the most important determinants of dietary habits in the overall population and specific groups;
- establishing associations between nutrients or food and health or disease risk in large prospective population-based studies;
- developing statistical methods and technology to handle large and complicated data sets;
- complementing association studies with mechanistic studies providing evidence of the biological basis of the observed associations; and
- conducting human intervention studies to ensure that cause–effect relationships between nutrients or foods and health are substantiated.

The authors

Inge Tetens (front right): MSc, PhD. Head, Nutrition Department, DTU National Food Institute, 2005–. Associate Professor, Royal Veterinary and Agricultural University, Copenhagen, 1996–2005. Research fields: human nutrition, applied nutrition, human nutrition intervention studies.

Ulla Vogel (behind left): MSc, 1991, PhD, 1995, University of Copenhagen. Professor with Special Responsibilities in Diet and Cancer, Department of Toxicology and Risk Assessment, DTU National Food Institute, 2008–. Senior Scientist, National Research Centre for the Working Environment, 1999–2008. Research fields: molecular epidemiology and identification of carcinogens.

More to explore

Tetens I, Oberdörfer R, Madsen C, Vries J. Nutritional characterisation of foods: science-based approach to nutrient profiling. *European Journal of Nutrition* 2007: 46(S2): 4–14.

Tjonneland A, Olsen A, Boll K, Stripp C, Christensen J, Engholm G, Overvad K. Study design, exposure variables, and socioeconomic determinants of participation in Diet, Cancer and Health: a population-based prospective cohort study of 57,053 men and women in Denmark. *Scandinavian Journal of Public Health* 2007: 35: 432–441.

Vogel U, Christensen J, Nexo BA, Wallin H, Friis S, Tjonneland A. Peroxisome profilerator-activated receptor[1]2 Pro12Ala, interaction with alcohol intake and NSAID use, in relation to risk of breast cancer in a prospective study of Danes. *Carcinogenesis* 2007: 28: 427–434.

5 µm

A
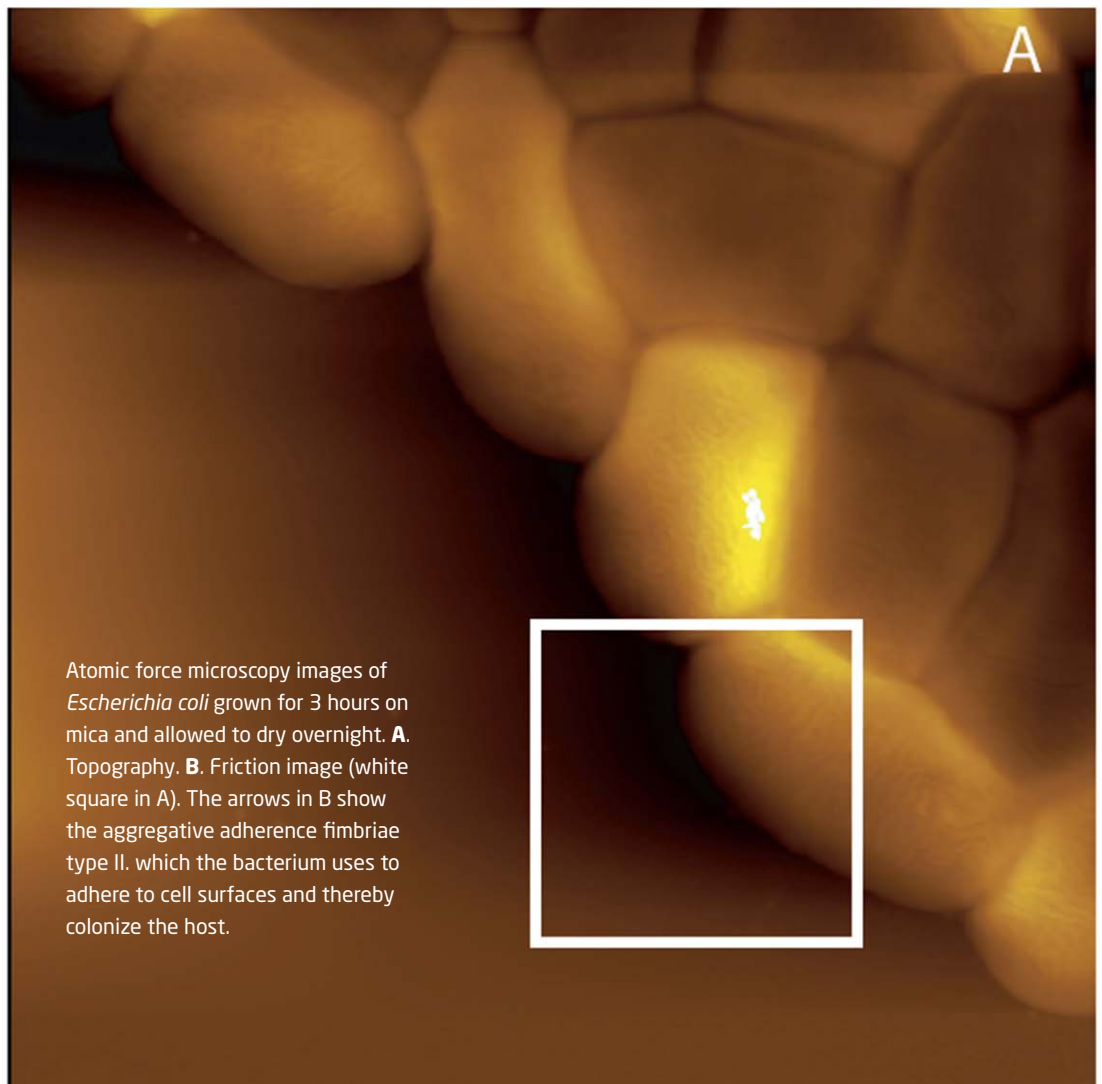
Atomic force microscopy images of *Escherichia coli* grown for 3 hours on mica and allowed to dry overnight. **A**. Topography. **B**. Friction image (white square in A). The arrows in B show the aggregative adherence fimbriae type II. which the bacterium uses to adhere to cell surfaces and thereby colonize the host.

0
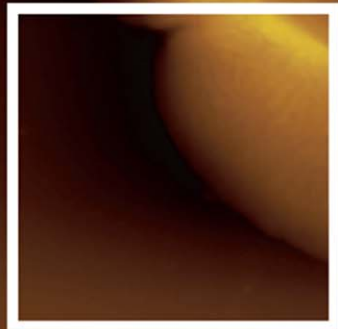
5 µm

Microorganisms
– the good, the bad and
the indispensable

Lone Gram & Frank Aarestrup

Introduction

Microorganisms, such as bacteria, can both benefit and threaten the health and well-being of humans and animals. Microorganisms cause disease in all eukaryotic organisms, and people are constantly battling infectious or intoxicating agents. Humans and animals master numerous defense strategies again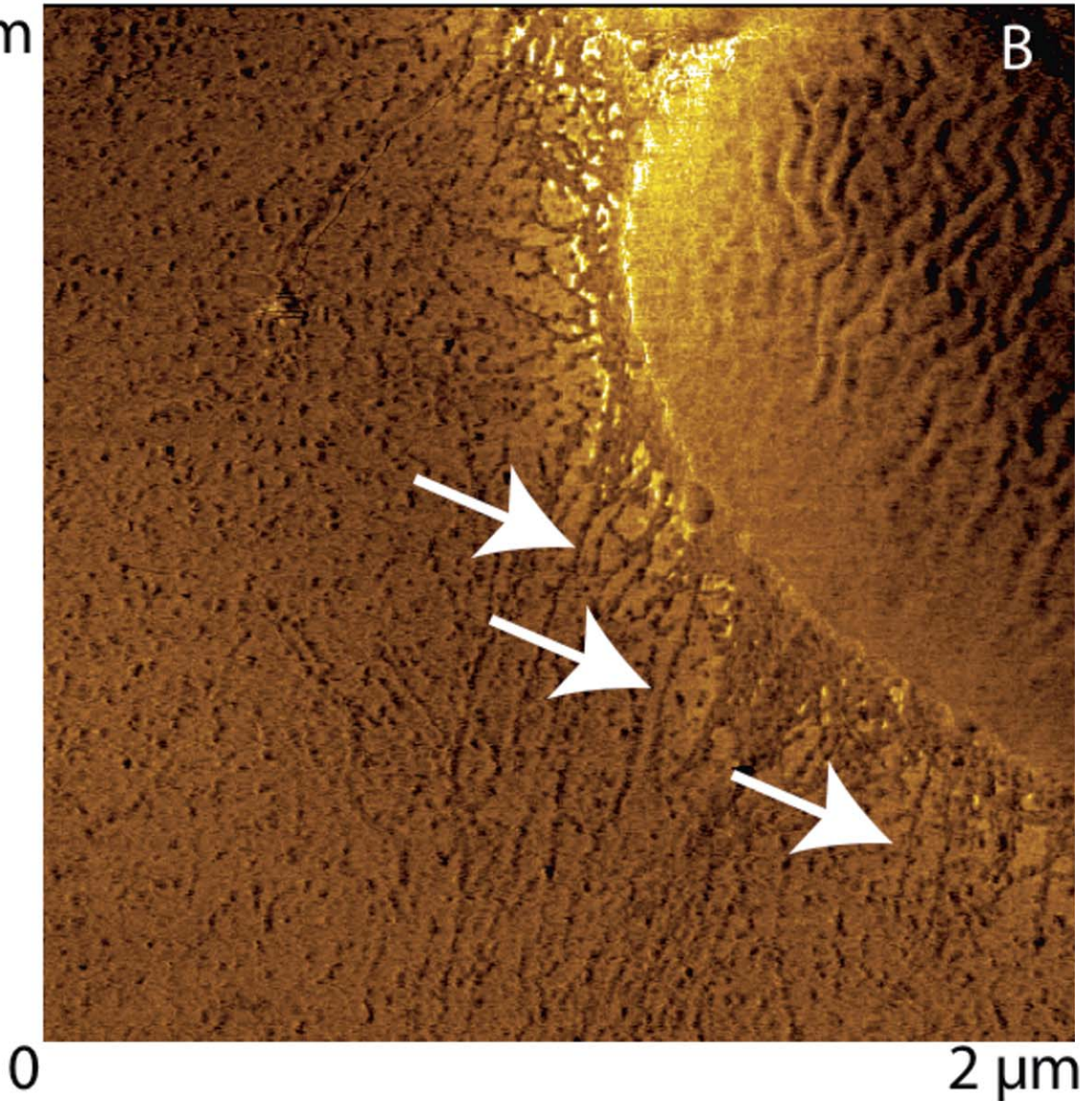st microbial infection. Improved hygienic control, the use of vaccines as disease prophylaxis and the use of antibiotics in treatment have brought about major progress in controlling infectious diseases. However, several types of bacteria are rapidly becoming resistant to the antibiotics used, and the emergence of methicillin-resistant *Staphylococcus aureus* in both hospitals and communities is alarming, as is the recent trend in antibiotic resistance in gram-negative bacteria due to extended-spectrum beta-lactamase activity. The use of antimicrobial agents may also change the balance of the normal bacterial flora and provide opportunities for naturally resistant bacteria such as enterococci and *Clostridium difficile* to emerge.

Microorganisms produce primary and secondary metabolites such as enzymes or antibiotics that are useful to humans, and biotechnological production exploits microorganisms. Further, mixtures of microorganisms are deliberately used in biological sewage systems. The commensal microbiota contributes to human health, and environmental microorganisms are key players in our environment. They are responsible for converting organic matter and hence form the foundation for all life. Microbial metabolism, such as in the marine environment, affects global climate.

The adaptability of microorganisms, their continued evolution and the global emergence of old and newly emerged pathogens and importance in coexistence as well as beneficial microorganisms with potential in biotechnology and disease control are unsurpassed by any higher organisms (animals and plants). This chapter gives examples of ongoing microbiological research at DTU and the approaches and techniques used to survey and understand new and emerging pathogens and to exploit new microbial metabolites as tools to control pathogenic microorganisms.

Pathogenic bacteria and antibiotic resistance – changes in a globalized world

Infectious pathogens and toxic organisms constitute continual threats to humans and animals and the emergence and spread of bacterial pathogens is becoming a topic of great scientific and practical importance. This has and will influence human and animal evolution and how society is organized.

Understanding the complexity of infectious diseases requires studying the bacteria in their ecological niches and the environmental circumstances under which they evolve into pathogenic organisms. Knowledge of the global trends and changes in society therefore needs to be combined with basic research into the molecular regulation and evolution of the bacteria and their hosts. Infections spreading between animals and humans especially constitute a challenge, because the bacteria must be studied not only in humans but also in the animal reservoirs.
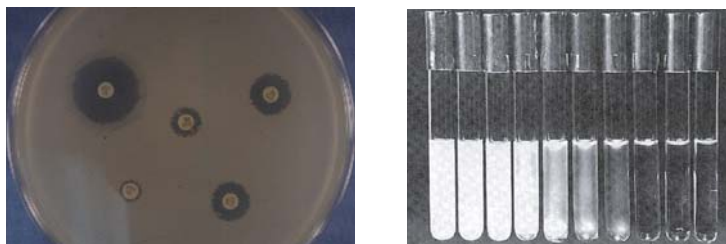
Fig. 18.1. Susceptibility testing by disk diffusion and determination of the minimum inhibitory concentration (at which no growth can be observed). **Left**. In disk diffusion, the bacterium is spread on an agar plate and disks containing different types of antibiotics are applied. During overnight incubation, the antibiotic diffuses from the disks and inhibits the growth of susceptible but not resistant bacteria. **Right**. To determine the minimum inhibitory concentration, a defined quantity of bacteria is inoculated in broth tubes or on agar plates containing increasing amounts of antibiotic (typically two-fold dilutions)

Bacteria resistant to antimicrobial drugs are rapidly emerging, and this problem has continued to follow the introduction of all new antimicrobial compounds. The antimicrobial agents used for domesticated animals are often the same or belong to the same classes as those used in human medicine. Modern livestock production uses huge amounts of these antimicrobial agents. This provides favorable conditions for selection, persistence and spread of antimicrobial-resistant bacteria capable of causing infections in both animals and people. DTU has been in the international forefront of understanding the events leading to resistance and in determining the importance of antimicrobial resistance, especially in foodborne and animal-borne pathogenic agents (Fig. 18.1).

Antibiotic resistance – magnitude and nature
Denmark was the first country to establish integrated monitoring of the incidence and trends in antimicrobial resistance among bacteria isolated from livestock, food and humans. This has become the international standard for monitoring bacterial antibiotic resistance in livestock and food products. Further, Denmark has surveyed how banning a glycopeptide antibiotic growth promoter (avoparcin) markedly reduced antibiotic-resistant bacteria in livestock (Fig. 18.2). Of enterococci isolated from broilers, resistance dropped from more than 70% to less than 10% in 3 years. However, resistant enterococci kept being isolated from pigs at around 20% for the first 3 years.

We used molecular methods to characterize the *Enterococcus faecium* isolates collected during the monitoring and found that all isolates from pigs, in contrast to isolates from broilers, had the same clonal lineage. Further, we showed that all isolates had genes encoding resistance to glycopeptides and another antibiotic class, macrolides, located closely together on the same plasmid. This suggested that resistance to glycopeptides persisted among pigs because macrolides continued to be used for pigs. Thus, molecular characterization of the bacteria provided suggestions for an unexpected event happening in real life. Fortunately, the use of macrolides decreased in 1998, and the monitoring program demonstrated the effect in real life, which confirmed the predictions based on the molecular characterization. This example also shows the importance of combining data and isolates from monitoring with basic information obtained in the laboratory.

Denmark's monitoring programs have enabled new types of pathogens to be monitored in farms in Denmark. Studies combining statistical analysis of monitoring data and data on antimicrobial use and molecular characterization of the isolates have shown that the emerging resistance among *Salmonella* does not seem to be associated with emerging antimicrobial resistance in already established clones but is mainly due to novel resistant clones that spread and replace the susceptible clones. This happens even though the resistance is located on transferable plasmids. This has forced us to reconsider how antibiotic resistance in bacteria emerges
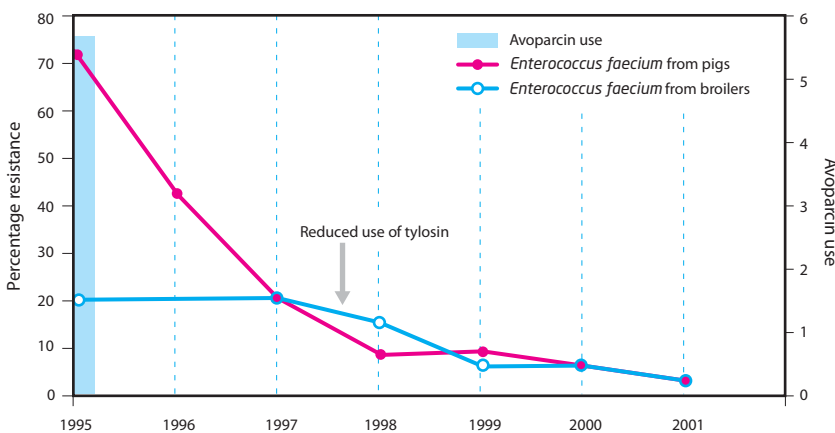

Fig. 18.2. Reduction in prevalence of bacterial resistance to glycopeptides following the ban of avoparcin in 1995 and reduced use of macrolides after 1998

and spreads and has provided potential new ways of limiting resistance.

Molecular characterization of tetracycline resistance genes from bacteria combined with phylogenetic analysis has shown that specific genes seem rather confined within specific clonal lineages. Together with the previous study and others, this suggests that horizontal gene transfer is not as frequent as initially believed but instead happens occasionally and then resistant clones subsequently spread, sometimes globally. This has provided a new focus for our research in which we aim to identify the global hotspots for the evolution of resistance and understand the relative importance between acquiring resistance and the spread of the resistant clone. Thus, besides combining basic research with field studies, DTU has also increased research activities globally in recent years.

From home to the globe

As mentioned pathogenic bacteria need not only to be studied and understood in the human or animal niche they occupy, they also need to be studied in a global perspective, as people, animals and goods are shipped globally. DTU has developed a mathematical simulation model for estimating the attribution of different sources, including foods of national or imported origin, for *Salmonella* infections in humans. This model is based on the following formula:

$$\lambda_{ij} = M_{j} * p_{ij} * q_{i} * a_{j},$$

where

λ_{ij} = the expected number of cases per year of type *i* from source *j*;

M_{j} = the amount of source *j* available for consumption per year;

p_{ij} = the prevalence of type *i* in source *j*;

q_{i} = the subtype-dependent factor for type *i*; and

a_{j} = the food-source dependent factor for source *j*.

It is assumed that the $\{q_{i}\}$ for subtypes with the same level of resistance (susceptible, resistant, multiresistant or nalidixic acid–resistant) and belonging to the same serotype were equal. The model has formed the basis for targeted interventions for controlling *Salmonella* nationally since the late 1990s and played a major part in reducing the incidence of salmonellosis in Denmark (Fig. 18.3). The model requires accurate monitoring data from (almost) all potential reservoirs. The model has been devel-
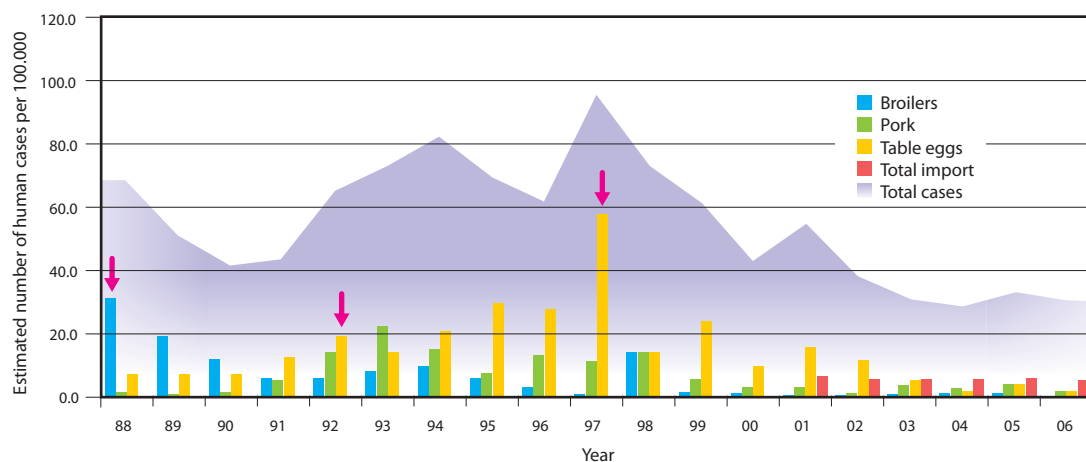

Fig. 18.3. Trends in human salmonellosis attributed to different sources and indicated targeted actions, 1988–2006. The arrows indicate the implementation of national control programs for broilers in 1992, for pigs in 1992 and layers in 1997.
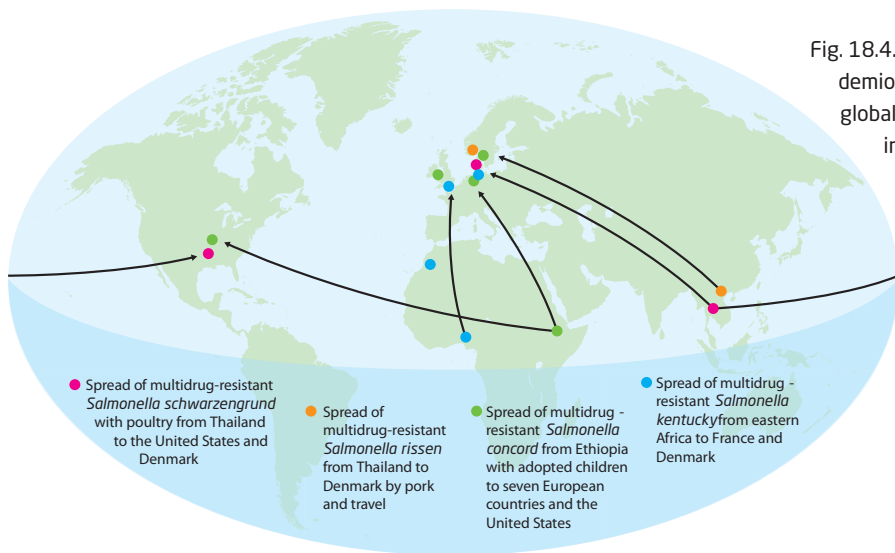
Fig. 18.4. Global molecular epidemiology: examples of the global spread of *Salmonella* in which DTU has been involved in the elucidation

● Spread of multidrug-resistant *Salmonella schwarzengrund* with poultry from Thailand to the United States and Denmark

● Spread of multidrug-resistant *Salmonella rissen* from Thailand to Denmark by pork and travel

● Spread of multidrug-resistant *Salmonella concord* from Ethiopia with adopted children to seven European countries and the United States

● Spread of multidrug-resistant *Salmonella kentucky* from eastern Africa to France and Denmark

oped over time based on several years of data. It does not require extended computer power but can be run on a normal PC. Today the model is set up in WinBUGS software and normally run with five independent Markov chains of 30,000 iterations each.

This exemplifies a very short distance between the science of developing detection of and methods for characterizing bacteria and a mathematical simulation model and political action being taken almost immediately based on the results.

This source attribution calculation has shown that imported food products and traveling cause an increasing proportion of the human infections in Denmark. Since antimicrobial resistance is higher in most countries from which food is imported than in Denmark, bacteria originating in other countries cause most infections with antimicrobial-resistant *Salmonella*. Hence, improving food safety in Denmark requires improving food safety on a global scale. In recent years, DTU has participated in studies confirming the global spread of bacteria, especially *Salmonella* clones, resistant to multiple antimicrobial agents. This has included spread with poultry from Thailand to the United States and Denmark, with pork and travel to Denmark, with adopted children from Ethiopia to several countries and spread by an unknown mechanism from eastern Africa to Europe (Fig. 18.4).

Beneficial bacteria

Denmark has traditionally used microorganisms to ferment food and feeds. Control of large-scale food and beverage fermentation requires detailed understanding and prediction of microbial metabolism. Microorganisms serve as a constant source of novel enzymes and other bioactive compounds. More recently, Denmark has also deliberately used beneficial microorganisms as disease control modulators in humans, animals and plants: probiotic or biocontrol microorganisms. Several crucial processes in society are microbe based, ranging from biological sewage treatment to solid-state fermentation.

Microorganisms produce compounds that either kill or inhibit the growth of other microbes. A well-known example is penicillin, produced by the fungus *Penicillium*, discovered by Alexander Fleming in 1928. However, bacterial resistance to known antibiotics has led to searches for novel antimicrobial agents.

The search for microorganisms with useful properties has until recently been mainly a land-based endeavor. There are numerous examples of terrestrial fungi or bacteria as a source of novel bioactive compounds such as antibacterial agents or hydrolytic enzymes. The next section describes recent work at DTU focusing on the marine environment for novel bioactive compounds.

Antimicrobial peptides from bacteria

Almost all living organisms produce cationic peptides with antimicrobial properties. In multicellular organisms such as animals and plants, they are an important part of the innate defense, and epithelial cells and the white circulating blood cells produce various antimicrobial peptides. Most have a broad antimicrobial spectrum and execute their killing effect by acting in a detergent-like manner; others have a more specific target. Not only eukaryotic organisms produce cationic peptides. Many bacteria also produce antibacterial peptides or proteins: bacteriocins, which are presumably used in bacterial competition. One such bacteriocin, colistin, is a polypeptide produced by *Bacillus polymyxa*. This has been used for several decades as an antibiotic. Polypeptides can be toxic for the host species, especially when injected into the bloodstream. Thus, until recently, colistin and other polymyxins had limited usage in treating *Escherichia coli* diarrhea in animals. However, due to the emergence of resistance to virtually all other antibiotics, colistin has recently been used in combination with more traditional antibiotics to combat *Pseudomonas aeruginosa* infection among people with cystic fibrosis and as a last resort in treating *E. coli* septicemia. This is a good example of reviving an old product after several decades, and its use is based on combining laboratory experiments and clinical practice.

Lactic acid bacteria produce bacteriocins that selectively kill or inhibit closely related bacteria. Since lactic acid bacteria are generally recognized as safe and have a long history as fermenting organisms contributing to preserving food, their bacteriocins and the cultures producing lactic acid bacteria have being investigated as preservatives. Several bacteriocins from lactic acid bacteria are effective against the foodborne pathogen *Listeria monocytogenes*, which can cause listeriosis. This serious infection primarily affects people with a immune system suppressed because of age, medication or pregnancy. Although listeriosis is very rare – Denmark had 51 cases in 2008 – the fatality rate is typically 20–30%. *L. monocytogenes* is ubiquitous and is present in most raw materials for food production. Its presence in foods is not inherently hazardous, but its growth in foods must be prevented to prevent disease. At DTU we have isolated *Carnobacterium piscicola*, a lactic acid bacterium from a fish product. *C. piscicola* produces a bacteriocin called carnocin that effectively kills or inhibits *L. monocytogenes* in ready-to-eat fish products. The gene encoding the bacteriocin is encoded on a plasmid, and removing this plasmid, and hence the bacteriocin, reduces the *Listeria*-inhibiting capability (Fig. 18.5).

These studies demonstrated that several naturally occurring bacteria in food actually serve a protective and preserving function, and several companies market lactic acid bacteria as bioprotective cultures. This and other similar observations have led the food industry to search for and develop novel protective probiotic strains for use in food products.
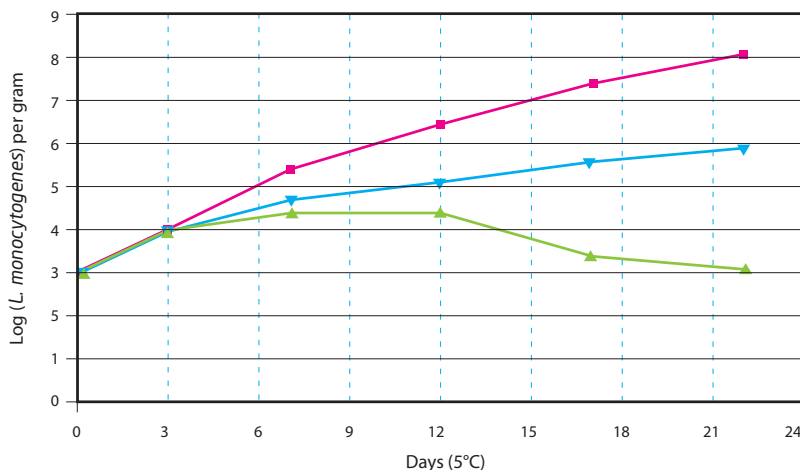

Fig. 18.5. Growth of *Listeria monocytogenes* in vacuum-packed cold-smoked salmon. *L. monocytogenes* as mono-culture (■) or co-inoculated with a carnobacteriocin-producing bacteria (▼) or a non-bacteriocin-producing mutant (▲).

Novel antibacterial compounds from marine bacteria

Oceans cover 70% of the Earth and constitute 95% of the biosphere. The biological richness of the oceans is unsurpassed, and 34 of 36 known phyla are in the marine environment. The ocean microbiota plays a central role in degrading organic matter and providing nutrients to algae and hence to completing the food chain. They are among the most abundant marine organisms.

It has been known for decades that marine microorganisms produce compounds that inhibit other microorganisms, but only during the past decade has this potential source of novel compounds been investigated more systematically. DTU is participating in this exploration of the marine environment. In 2006–2007, the DTU National Institute of Aquatic Resources and Department of Systems Biology joined the Galathea 3 Expedition to establish a global collection of bioactive, antibacterial culturable bacteria from almost all marine environments. The long-term goal is to determine the potential use of bioactive secondary metabolites from culturable marine bacteria. Specifically, we searched for marine bacteria that could inhibit pathogenic bacteria or interfere with their virulence.

The Galathea 3 data confirmed that ocean surface water contains about 10^6 cells/ml (Fig. 18.6) and that only a fraction of these can be cultured under standard laboratory conditions. The culturable bacterial count varied but averaged 10^3 colony-forming units per ml. Internationally, several large projects, including one led by Greg Venter, have set out to describe the 99.9% of unculturable ocean microbiota by use of shotgun sequencing and metagenomic analyses. These studies will provide unsurpassed insight into the organisms and the genes present in this large unculturable fraction. Being able to culture the microorganisms known to produce bioactive compounds is a major benefit, however, as the compounds can be produced, extracted and tested in a range of biological systems. We are currently pursuing this using 500 types of bioactive marine bacteria collected during the Galathea 3 Expedition.
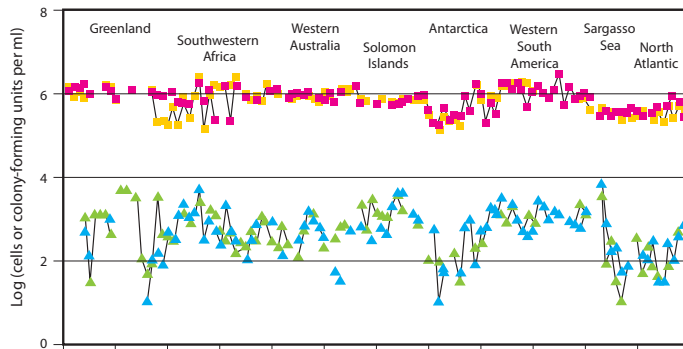

Fig. 18.6. Bacterial density in ocean surface waters along the Galathea 3 route. The squares depict the total numbers of cells counted by microscopy on fluorescence-stained (SYBR Gold) cells. The triangles depict the numbers of culture cells cultured on marine agar.

Bioactive culturable bacteria

Of the culturable microbiota, 4–13% of the colonies were capable of inhibiting the growth of other bacteria. Most of the bioactive culturable bacteria belong to one of three groups: the *Roseobacter* clade, the *Pseudoalteromonas* genus or the *Vibrionaceae* family. The DTU Department of Systems Biology is currently extracting and elucidating the structure of bacterial compounds with biological activity.

The *Roseobacter* clade is fascinating. They predominate the upper ocean waters and may comprise as much as 30% of the prokaryote population. They are typically associated with algae and metabolize the algal sulfur compound dimethylsulfoniopropionate, producing dimethylsulfide, an important component of cloud formation. Several species will grow as single cells under aerated conditions, and no antibacterial compound is formed under these conditions. However, if the bacteria are allowed to form a biofilm at the air–liquid interface, the bacteria grow as rosettes and produce a highly antibacterial compound, tropodithietic acid (Fig. 18.7). The mechanism by which the compound exerts its antibacterial activity is not known, but resistant mutants have not been isolated. This finding is promising if future use of this antibacterial compound is envisioned.
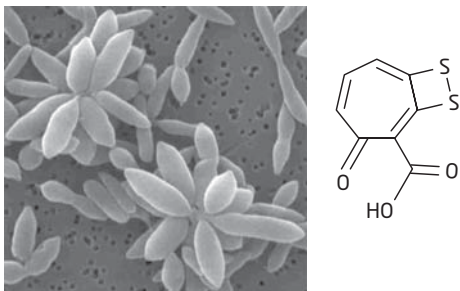
Fig. 18.7. **Left**. Rosettes of a *Roseobacter* clade strain as visualized by scanning electron microscopy. **Right**. Tropodithietic acid, an antibacterial compound produced by the rosettes.

Pseudoalteromonas consists of two major groups; one produces intense pigments of red, purple or yellow color. The pigmented strains almost always inhibit other bacteria and typically produce a range of antibacterial compounds (Fig. 18.8). A few of these are known, but we are currently isolating novel antibacterial agents from this group. The pigmented *Pseudoalteromonas* are primarily associated with the surfaces of eukaryotic marine organisms (seaweed, fish and Mollusca) and are predominantly isolated from tropical and subtropical waters. The role of their intense pigments is not known, but we hypothesize that they serve as protection from ultraviolet light and will therefore investigate their antioxidant properties. Although they are not related to controlling pathogenic bacteria, antioxidant compounds are used in various products and processes.
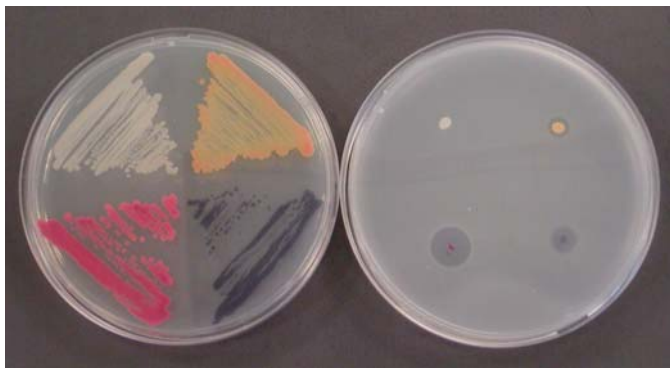

Fig. 18.8. Four *Pseudoalteromonas* strains and the antibacterial activity of the three pigmented strains

The last major bioactive group, the *Vibrionaceae*, is mostly known due to the many species that are pathogenic to fish and humans, including *Vibrio cholerae*. Many *Vibrio* species produce an array of extracellular compounds (enzymes and toxins) that function as virulence factors, and several may have antibacterial activity. Even if the producing strain or culture is pathogenic, one could hypothesize that a purified compound, expressed in a nonpathogenic organism, could have biotechnological potential.

The isolation of bioactive marine bacteria focused on their ability to inhibit or kill other (pathogenic) bacteria. However, the marine bacteria also produce compounds that do not directly kill the target pathogens but interfere with their virulence. Recent results from the University of Copenhagen indicate that some of the marine bacteria produce compounds that specifically interfere with the regulation of virulence in *Pseudomonas aeruginosa* or *Staphylococcus aureus*.

Future perspectives – microbiology at DTU
Understanding the global changes and emergence of bacterial pathogens, the complex interactions between multiple bacterial species in natural communities or selecting and engineering bacteria for production purposes requires thoroughly understanding the basic physiology and regulation of single bacterial isolates. Understanding the behavior of single isolates requires knowing their role, behavior and real-life response to changes in their natural environment. Finally, robust specially designed informatics technologies (databases, computer science and bioinformatics) are required to handle the vast amount of information and required analysis.

Our current and future research builds on an attempt to integrate real-life (*in vivo*), laboratory (*in vitro*) and computer (*in silico*) sciences to study pathogenic and beneficial bacteria. We want to take the world into the laboratory and the laboratory into the world under the umbrella of information sciences and statistical modeling.

The new DTU has an impressive number of microbiologists and masters a range of both classical and very sophisticated techniques for studying, controlling and using microorganisms. We have

only given a few examples of the microbiological research conducted at DTU. We are currently linking these groups in a network, very much inspired by a similar collaborative network structure at Harvard University (the Microbial Sciences Initiative). We expect that this network will organize postgraduate teaching, PhD supervision, seminars and colloquiums and that this will give DTU students access to some of the best-qualified research environments in this field. We envision this network as combining strong research groups in relevant fields from different departments. A small and efficient core secretariat organizing symposia, lectures, journal clubs and other activities is essential for efficient flow of information and to ensure that the network functions.

More to explore

Diamond JM. *Guns, germs and steel: the fate of human societies*. New York, Norton, 1996.

de Kruif P. *Microbe hunters*. Boston, Mariner Books, 2002.

Aarestrup FM, ed. *Antimicrobial resistance in bacteria of animal origin*. Washington, DC, ASM Press, 2006.

Fenical W, Jensen PR. Developing a new resource for drug discovery: marine actinomycete bacteria. *Nature Chemical Biology* 2006: 2: 666–673.

Bowman JP. Bioactive compound biosynthetic capacity and ecological significance of the marine bacterial genus, *Pseudoalteromonas*. Marine Drugs 2007: 5: 220–241.

Zasloff M. Antimicrobial peptides of multicellular organisms. *Nature* 2002: 415: 389–395.

The authors

Frank Møller Aarestrup:
DVM, 1992, PhD, 1996, Dr. Vet. Sci., 2000, Royal Veterinary and Agricultural University, Copenhagen. Research Professor, DTU National Food Institute, 2002–. Head, WHO Collaborating Centre for Antimicrobial Resistance in Foodborne Pathogens and the Community Reference Laboratory for Antimicrobial Resistance of the EU. Has many years of experience in controlling antimicrobial resistance and in international collaboration.

### Lone Gram:
PhD, 1989, Royal Veterinary and Agricultural University, Copenhagen. Research Professor in Microbial Ecology and Physiology of Aquatic Organisms, DTU National Institute for Aquatic Resources, 2000– and in collaboration with the DTU Department of Systems Biology, 2005–. Research fields: bacterial ecology and physiology, bacterial interactions, novel antibacterial compounds.
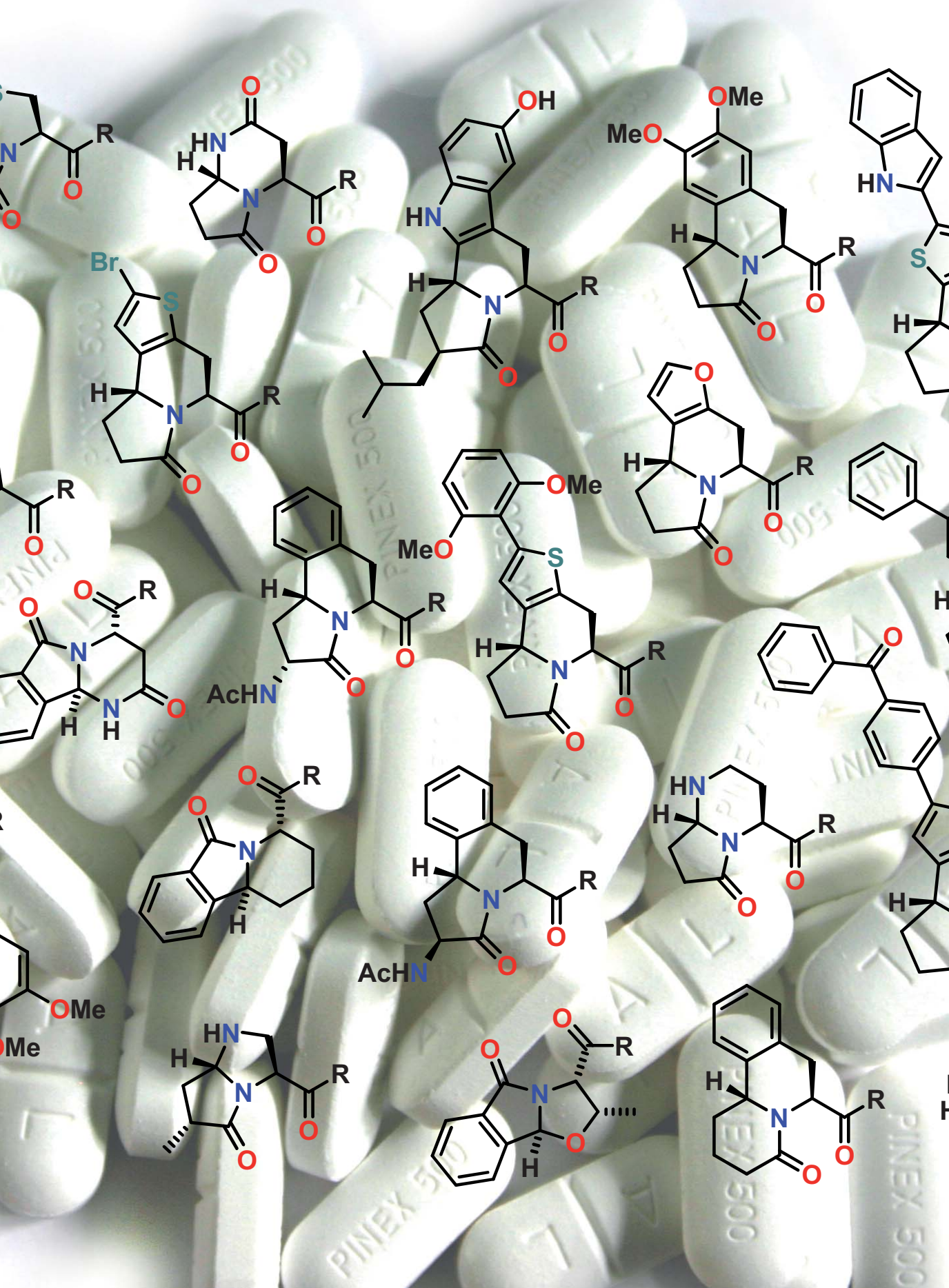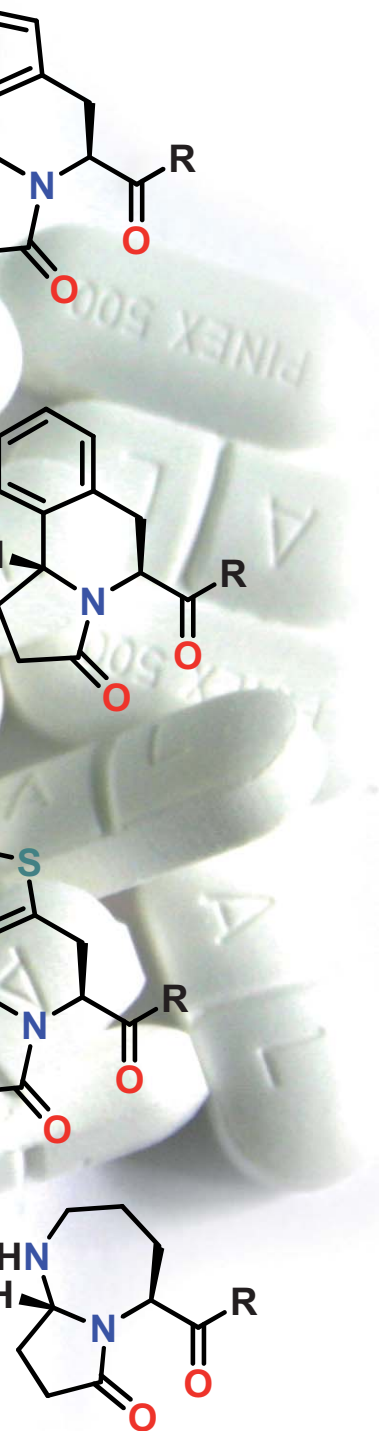
Small molecules and medicine

Thomas E. Nielsen

Small organic molecules are valuable for treating diseases and constitute most medicines marketed today (Fig. 19.1). Such molecules are also highly useful as chemical probes to study the functions of genes, cells and biochemical pathways. Their effect on recent life sciences research has been dramatic, providing both new tools for understanding living systems and a smoother transition from biology to medicine. After a period of widely recognized advances in genetics and molecular and disease biology, chemistry is returning to a position of prominence in the consciousness of the larger scientific community.
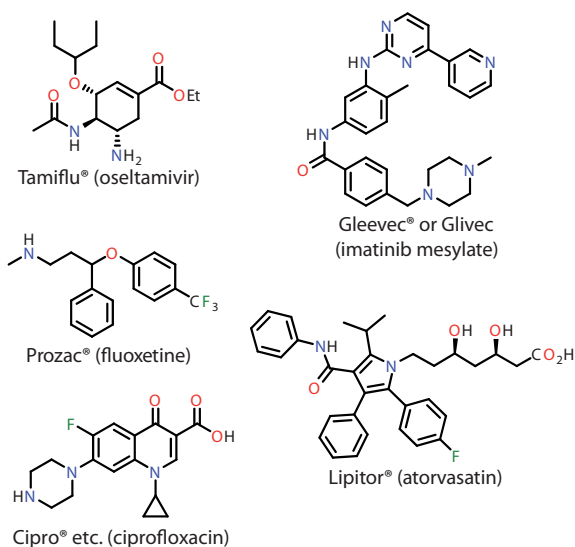
Fig. 19.1. Synthetic small molecules with diverse, biomedically important properties produced by major pharmaceutical companies. Gleevec® or Glivec® (imatinib mesylate, Novartis) is used to treat a blood cancer called chronic myeloid leukemia. Cipro® and other trade names (ciprofloxacin, Bayer and other manufacturers) has broad-spectrum antibiotic activities. Tamiflu® (oseltamivir, Roche) displays antiviral activity against influenza viruses. Lipitor® (atorvastatin, Pfizer) is used to lower serum cholesterol and the risk of heart disease. Prozac® (fluoxetine, Eli Lilly and other manufacturers) is an antidepressant used for mood disorders.

Perturbations of suitable model systems and molecular analysis of the consequences constitute the paradigm for studying complex biological phenomena (Fig. 19.2). Along these lines, advances in DNA sequencing (genomics), protein expression and approaches addressing interactions between proteins (proteomics) have provided detailed knowledge about the macromolecules of higher eukaryotes and the networks underlying their interactions. Importantly, these techniques have uncovered a wide range of proteins whose functions remain unknown.

Complementary to systematic DNA mutagenesis and RNA interference studies, the use of small molecules (chemical genetics) to modulate gene expression or protein function is emerging as an increasingly important and promising approach

for elucidating the roles of these proteins, thus posing a new direction for research in the postgenomic era. What are small molecules? They are not the macromolecules of living systems (DNA, RNA and proteins) consisting of thousands to millions of atoms. Small molecules comprise far fewer atoms, typically less than 100. They represent the kind of molecules nature uses to regulate cellular processes and information flow and are therefore, not surprisingly, valuable for treating diseases. For example, the antidepressant fluoxetine (Fig. 19.1) is a small molecule consisting of only 40 atoms. Such molecules are also highly useful as probes to study, for example, the individual functions of multifunctional proteins, cell circuitry and the physiology of living systems. Recent technological developments (high-throughput screening) have elevated the small-molecule approach to biology such that it poses a truly revolutionizing impact on modern life sciences research. Small molecules have the potential to rapidly diffuse in and out of cellular systems. Here they can interact selectively and reversibly with proteins and other macromolecules. In this way, small molecules offer rapid, conditional, dose-dependent and often reversible control of cellular functions, which cannot be achieved with traditional genetic methods (Fig. 19.3).

Small-molecule challenges for the pharmaceutical industry

In the hunt for new drugs, the pharmaceutical industry has faced the challenges noted above for more than a decade. Intense screening campaigns using large compound collections prepared via traditional combinatorial chemistry and parallel synthesis techniques have been launched. However, such molecular libraries are not meeting the demands and constraints of these new potential targets for drug discovery, as seen, for example, in the failure of traditional medicinal chemistry compounds to disrupt protein–protein interactions. Despite advances in the technology behind analytical instrumentation and robotics that have enabled ultra-high-throughput synthesis and the generation of an ever-increasing number of compounds per year, the number of new molecular entities making their way to market as drugs has been halved during the past decade (Fig. 19.4).

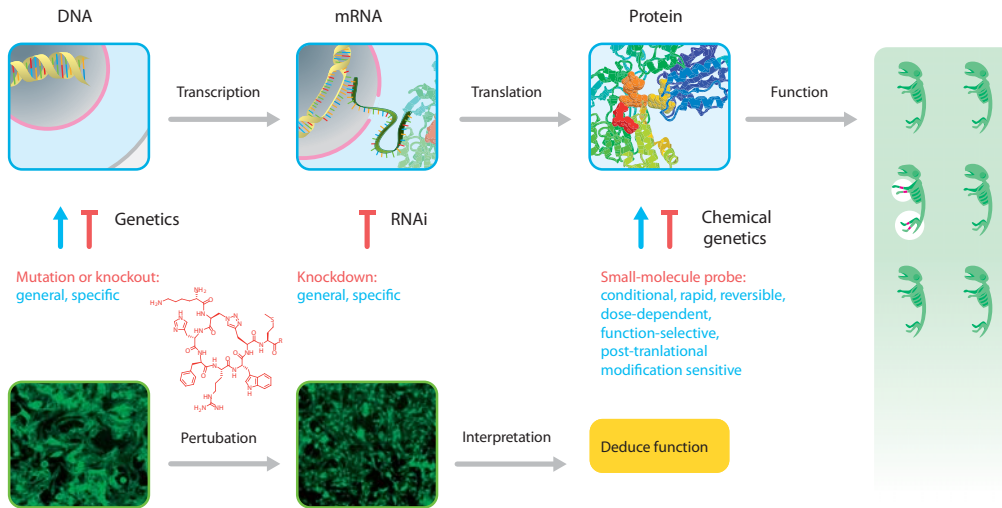Fig. 19.2. Opportunities for small-molecule perturbagens and probes in the study of biological phenomena and medicine. mRNA: messenger RNA. RNAi: RNA interference.
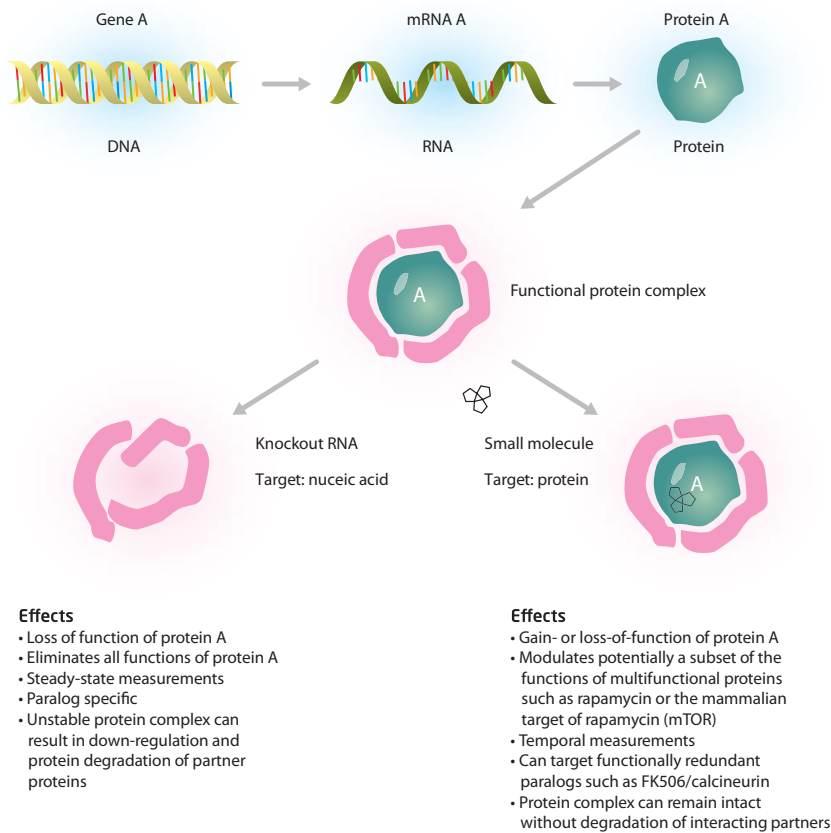

Fig. 19.3. Differences between nucleic acid-based and small molecule–based modulation of protein function, emphasizing the reasons small molecules are being used with increased frequency. mRNA: messenger RNA. RNAi: RNA interference.
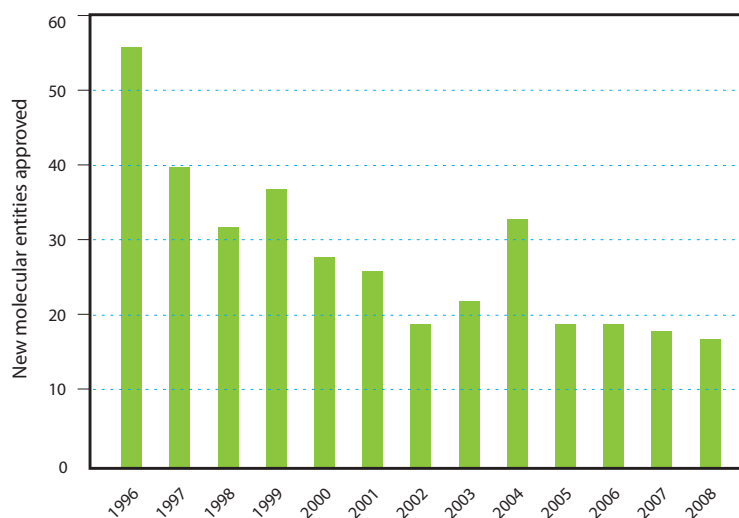
Fig. 19.4. Number of new molecular entities approved by the United States Food and Drug Administration, 1996-2008

The decline in drug discovery successes and high clinical development costs have caused the global productivity of the pharmaceutical industry to fall dramatically. As the major pharmaceutical companies have reached a size that requires generating several billions of dollars in revenue every year to satisfy investor expectations, the focus on R&D expenditure has never been more intense. The rapidly growing R&D costs during the last decade have not been paralleled by a similar market productivity trend. Similarly, R&D expenditure in the pharmaceutical industry roughly increased from US$10 billion in 1993 to US$30 billion in 2004. The cost of bringing a new prescription drug on the market is now estimated to easily exceed US$600 million (some very recent estimates even exceed US$1 billion!), implying that certain drugs might never generate the revenue needed to cover initial investment. The upcoming expiration of several major blockbuster drug patents and the related rise of less-expensive generic alternatives are further worsening the situation. The regulatory demands of drug approval are stricter than ever, and the low-hanging fruits have all been picked. Realizing that the days of the blockbuster drugs are over, pharmaceutical companies are now increasingly turning their attention to rare and complex disease indications. In this context, discussing the perfor-

mance of small molecules in disease-relevant screens becomes relevant. There is widespread agreement that current compound collections for screening are lacking, and chemistry strategies that yield advances in higher-performing small molecules are therefore in great demand. Based on the chemistry performed during the past 20 years, larger compound libraries would be expected to yield many new hit and lead structures, but these expectations have clearly not been met. Deficiencies in the performance of current compounds are further evident in colloquialisms such as "undruggable targets" and "crowded intellectual property space" limiting "freedom to operate".

The extent to which the decline in drug-discovery successes is due to shortcomings in synthetic chemistry remains controversial and inherently difficult to address. However, small-molecule synthesis combined with small-molecule screens in open data-sharing environments (such as Chem-Bank and PubChem) are now beginning to illuminate the structural properties of small molecules most likely to affect assay performance. Research in this area is revealing the value of using compounds that are poised for optimization during follow-up studies or for modification during, for example, target identification studies. However, an overall successful outcome will require manufacturing optimized compounds for broad distribution or for preclinical or clinical investigations, and synthetic pathways must thus be short and efficient. Collectively, organic synthesis faces substan-

tial challenges. These trends require advances in synthetic planning and methods. Synthetic routes are needed to create optimized small molecules better able to interact specifically and potently with disease-relevant proteins and for improved versions of candidates identified in biological discovery efforts. What are the structural features of small, organic molecules most likely to yield specific modulation of disease-relevant protein functions? How can candidate structures poised for optimization be effectively synthesized? It remains difficult to predict which small molecules will best modulate a given biological process or disease state, especially with limited prior knowledge, such as protein crystal structures, of the macromolecular targets. Researchers must therefore systematically screen tens or hundreds of thousands of small molecules to find a successful match between a chemical and its target.

Small molecules come in a variety of different shapes: they may be flat, rod-like or spherical and may contain a variety of atoms with many functions. Natural products are small molecules that tend to be complex, highly three-dimensional in structure and very diverse, whereas compounds made by traditional medicinal chemistry tend to be simple, flat and alike. Natural products represent a prime source of "chemical diversity" and, for example, the hit rates of screening natural product libraries for novel antibiotics by far exceed those of the prevalent compound libraries in the pharmaceutical industry. Much of the medicinal chemistry practiced in industry fails to provide compounds suitable for such purposes as developing antibiotic drugs, imparted by the strategic choice of focusing

on quantity rather than quality (or chemical diversity). Given their track record, such as in discovery of anticancer drugs and antibiotics, natural products may therefore conveniently inspire the design of novel small molecules. Nevertheless, from an experimental viewpoint, the overall synthesis of such natural products is extremely challenging and laborious, and the structural complexity clearly limits the scope for pharmaceutical applications. A key question therefore is how synthetic small molecules can be created that mimic the structural features of potent natural products but are accessible in a few synthetic steps.

Diversity-oriented synthesis

Planning and performing multistep synthesis of specific natural products in the past has resulted in the recognition, and often, resolution of limitations in synthetic methods. The synergistic relationship between chemistry and biology is even more profound as synthetic organic chemists approach the challenges noted above by launching new strategies and methods. One such approach emphasizes diversity as a key parameter in designing molecular libraries. Belief has been growing in the hypothesis that identifying small molecules that can modulate a range of disease-relevant protein functions specifically requires collections of candidate small molecules having diverse structures. Creating compound collections that are maximally diverse increases the probability that different library members may perturb different proteins. Diversity-oriented synthesis offers the potential to meet these demands. Unlike traditional target-oriented syn-

Diazepam
0 stereocenters
Synthesis: 1-2 steps

Diversity-oriented
synthesis compound
5 stereocenters
Synthesis: 3-4 steps

Ginkgolide B
11 stereocenters
Synthesis: >25 steps
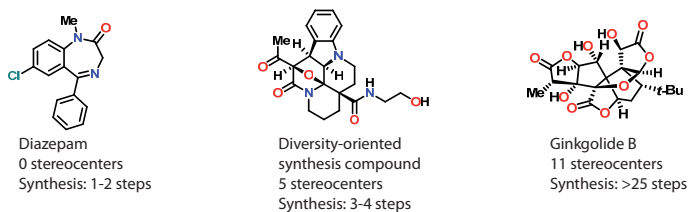
Fig. 19.5. Categories of small molecules. Drug: diazepam, a typical medicinal and combinatorial chemistry compound resulting from analog collections having known or predicted properties. Diversity-oriented synthesis compound: typically resulting from a collection of complex and diverse structures with unknown properties. Natural product: ginkgolide B (a constituent of *Ginkgo biloba*), a typical natural product with single target structure having known or predicted properties.

thesis strategies, the diversity-oriented synthesis approach enables chemists to efficiently synthesize libraries of complex and structurally diverse small molecules in a small number of synthetic steps. Rapidly optimizing the properties of these small molecules by structural modification in follow-up studies requires synthesis that is short and modular, with structures of the candidates that possess orthogonal chemical functionality facilitating the appending of substituents onto their core skeletons. Diversifying structures of small molecules by altering stereochemistry and skeletal arrays rather than by altering appendages has been a central tenet of the most successful approaches to diversity synthesis. The compounds derived from such pathways must have the potential to incorporate appendages regio- and stereoselectively at a variety of attachment sites during postscreening maturation stages. When a target has been identified, the later stages are feedback loops that yield focused libraries with tuned activities. However, systematically creating structurally diverse and complex small molecules in a few synthetic steps is far from trivial. The picture is further complicated if the invoked synthesis methods must be amenable to the solid support, which could be highly beneficial for generating larger compound libraries (such as more than 200–500 compounds).

The synergistic relationship between organic synthesis planning and methods is even more significant as synthetic organic chemists tackle the new challenges noted above. The objects of synthesis planning, no longer limited by the biochemical transformations used by cells in synthesizing naturally occurring small molecules, require radically new strategies and methods. Several efforts to identify planning concepts for synthesizing small molecules having at least some of the features described above have been reported in recent years. These strategies include biology-oriented synthesis (Waldmann), molecular editing (Danishefsky) and libraries from libraries (Houghten). The next section focuses on a striking example of diversity-oriented synthesis conceived when I was at the Center for Solid-Phase Organic Combinatorial Chemistry at the Carlsberg Laboratory, with Morten Meldal. The DTU Department of Chemistry is exploring this further. Pathways of this kind, which in a few synthetic steps lead to structurally complex and

skeletally and stereochemically diverse small molecules, preferably poised for optimization, may have the potential to achieve the goals noted above in the future.

Solid-phase synthesis of pharmaceutically relevant heterocycles via *N*-acyliminium chemistry

Protected natural and non-natural α-amino acids are readily available from synthetic and commercial sources. Benefiting from decades of optimization of peptide synthesis methods, such building blocks readily fulfill a desirable criterion of full stereochemical control in their mutual coupling. Nielsen & Meldal connected building blocks by using standard peptide coupling procedures to yield masked peptide aldehydes (Fig. 19.6). Treatment with acid liberates the corresponding aldehyde, which immediately condenses with the amide backbone to generate an *N*-acyliminium intermediate. By changing nucleophilic moieties positioned in the side chain (R^2) of a strategically positioned amino acid residue, new ring systems were formed by cyclization to the *N*-acyliminium intermediates. In these reactions, for example, the intramolecular *N*-acyliminium Pictet-Spengler cyclization leads to molecular structures, where the hydrogen in the newly formed stereogenic site always bears a *cis* relationship to R^2. As the relative stereochemical orientations of the R^1 and R^3 substituents do not interfere with the *N*-acyliminium cyclization, 8 of 16 possible stereoisomers of the resulting products are accessible using this approach.

Complete stereochemical diversification is thwarted by the current inability to overcome the substrate-controlled face selectivity posed by the *N*-acyliminium intermediates, thus suggesting a challenge for future synthetic efforts as incomplete collections of stereoisomers may impair efforts to extract powerful stereochemistry-based structure–activity relationships from primary screening data. Stereochemistry-based structure–activity relationships can provide important clues that facilitate optimization and modification studies following the discovery of a small-molecule lead.
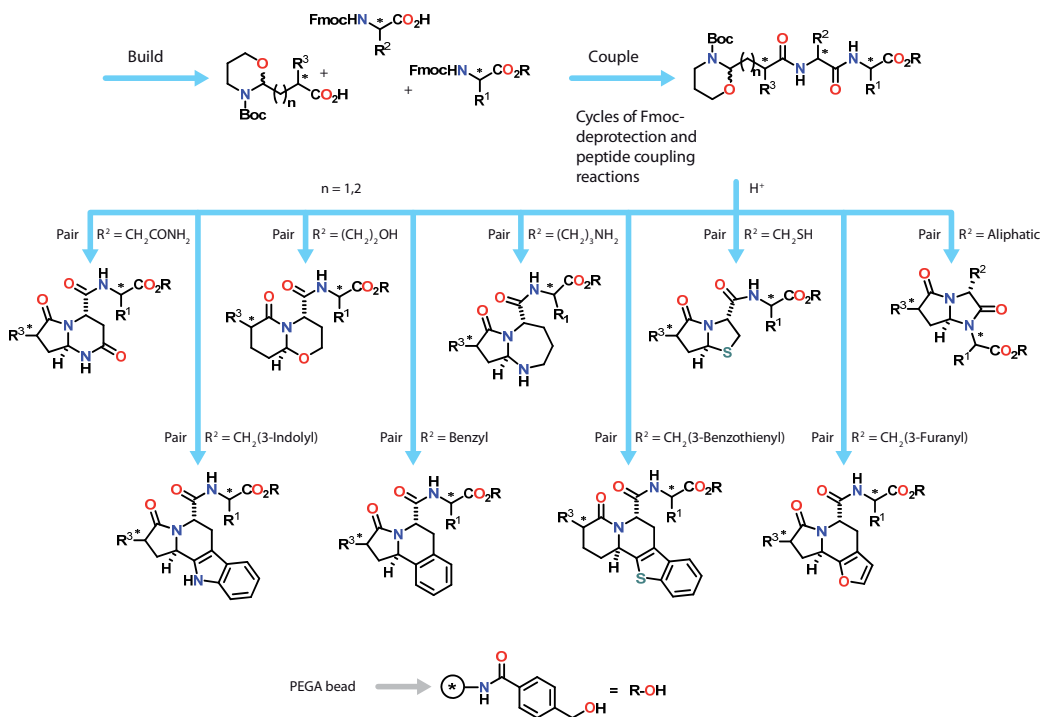
Fig. 19.6. Solid-phase peptide deprotection and amide bond formations in the couple phase (coupling functional groups: amine or activated carboxylic acid) were followed by aldehyde-amide condensation and subsequent addition of a nucleophile to an iminium intermediate in the pair phase (pairing functional groups: *N'*-acyliminium ion or heteroaromatic ring, aromatic ring, amine, carbamyl, amide, alcohol and thiol). Fmoc: 9-fluorenylmethyloxycarbonyl. PEGA: polyethylene glycol monoacrylate.

Outlook

By establishing the Molecular Libraries Screening Centers Network in 2005, the United States National Institutes of Health (NIH) Roadmap for Medical Research clearly illustrated the growing impact of synthetic small molecule probes on biology. This major strategic investment now allows academic centers across the United States to operate as national, high-throughput screening centers to create biologically relevant chemical tools on an unprecedented scale. The aims are to create a public collection of chemically diverse small molecules, allow high-throughput screening to identify compounds active in target- and phenotype-based assays, perform medicinal chemistry to transform hits into chemical probes, implement novel technologies and deposit screening data into a freely accessible public database. This unique initiative has empowered public-sector biomedical researchers to use small molecules as probes to systematically explore genes and cell circuitry on a scale previously available only to scientists in the pharmaceutical industry. In a second phase of the program, now more than ever realizing the potential and impact of small molecules in medicine, the NIH funded a network of nine centers in 2008; the Molecular Libraries Probe Production Centers Network, at about US$70 million annually over 4 years. The goal is to increase the pace of development and use of small-molecule probes for exploring biological processes and for developing new therapies for disease.

An underlying challenge for these efforts is developing methods for creating diverse and structurally complex small molecules in a few synthetic steps – a formidable and difficult task that only synthetic organic chemists can solve. No research group in a Danish university setting has truly undertaken this approach for the development of molecular libraries. However, this is urgently needed if
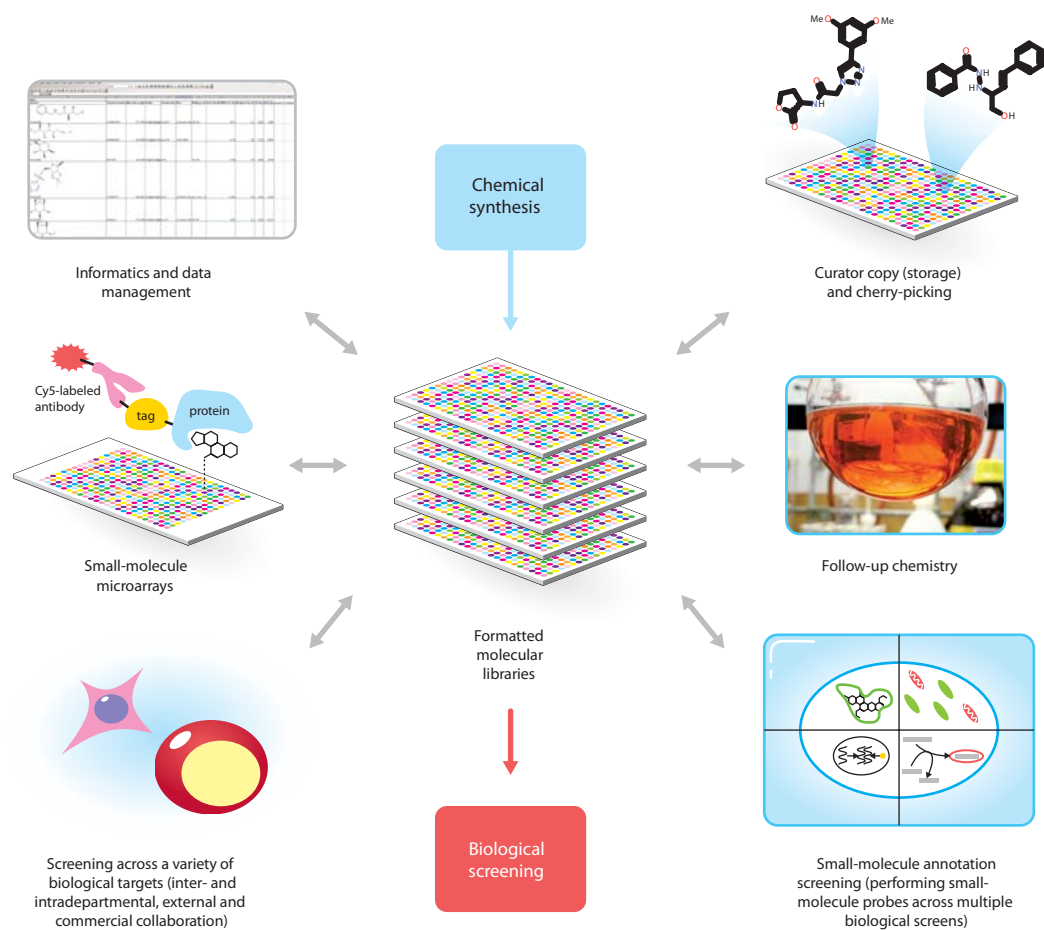
Fig. 19.7. A chemical biology platform: an integrated approach to generating and identifying biologically active small molecules

Denmark wants to establish a position at the forefront of chemical biology. Current research at the DTU Department of Chemistry aims at exploring this most important challenge of modern organic chemistry: the development of diversity-oriented synthesis techniques for creating molecular libraries. It is the hope to lay the first bricks in what may evolve into a national research center for molecular library development. In the future, I envision this center to be engaged in a European network of centers for high-throughput screening. As these activities are now emerging in Europe, as exemplified by the establishment of ChemBioNet in Germany, major strategic investment within the European Union is expected within the near future. We need to be prepared for such collaborative research activities as active players in molecular library de-

velopment. The DTU Department of Chemistry is currently laying the foundation for a chemical biology platform at the international level (Fig. 19.7).

As the very nature of the diversity-oriented synthesis approach is to create molecular libraries capable of modulating the functions of many different proteins, the access to many diverse high-throughput screening assays is highly desirable to justify these hypotheses. Diversity-oriented synthesis molecules have the potential to perturb many biological pathways specifically, and to best verify this, collaboration with national and international screeners of expertise in distinct areas of disease biology has been initiated. These efforts mainly focus on identifying compounds with antibacterial and anticancer properties. With competencies in diversity-oriented synthesis, combinatorial chemis-

try and molecular library development, research groups at the DTU Department of Chemistry are now developing efficient, general, state-of-the-art methods for designing, synthesizing, analyzing and handling chemically diverse libraries of small molecules. As a result, the number of molecular entities produced in-house is expected to increase dramatically in the years to come. All compounds are being submitted to an intradepartmental compound collection, and with systematic storage in 96- and 384-well plates (as dimethyl sulfoxide stock solutions) at low temperature, the basis of a compound pipeline feeding assay developers and screeners across DTU is in the making. To this ever-growing compound collection, chemists can deposit small molecules for systematic distribution and screening across a variety of assays relating to many facets of human disease biology. In this way, biomedical researchers at DTU will have unparalleled opportunities for discovery that previously almost exclusively resided in the private sector. In addition to creating exciting science, these initiatives will also educate a new, strong generation of young scientists prepared for the multidisciplinary research efforts increasingly demanded by both academia and industry. The scientific discoveries to emerge, such as new biochemical probes (small molecules) for the study of bacterial infections and cancer, will illustrate the effects of the planned research activities – with direct implications for human health.

More to explore

ChemBank: chembank.broad.harvard.edu

PubChem: pubchem.ncbi.nlm.nih.gov

Burke MD, Schreiber SL. A planning strategy for diversity-oriented synthesis. *Angewandte Chemie International Edition* 2004: 43: 46–58.

Tolliday N, Clemons, PA, Ferraiolo P, Koehler AN, Lewis TA, Li X, Schreiber SL, Gerhard DS, Eliasof S. Small molecules, big players: the National Cancer Institute's Initiative for Chemical Genetics. *Cancer Research* 2006: 66: 8935–8942.

Baxendale IR, Hayward JJ, Ley SV, Tranmer GK. Pharmaceutical strategy and innovation: an academics perspective. *ChemMedChem* 2007: 2: 768–788.

Schreiber SL, Kapoor TM, Wess G. *Chemical biology: from small molecules to systems biology and drug design*. Weinheim, Wiley-VCH, 2007.

Nielsen TE, Schreiber SL. Towards the optimal screening collection: a synthesis strategy. *Angewandte Chemie International Edition* 2008: 47: 48–56.

The author

Thomas E. Nielsen: MSc (1999) and PhD (2002), DTU. Postdoctoral studies with Morten Meldal, Center for Solid-Phase Organic Combinatorial Chemistry, Carlsberg Laboratory (2003–2005) and Stuart L. Schreiber, Department of Chemistry and Chemical Biology, Harvard University and Broad Institute of Harvard and MIT (2006–2007). Associate Professor, Department of Chemistry, DTU (2007–). Deputy Director, Center for Antimicrobial Research (2009–).
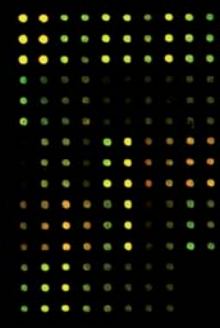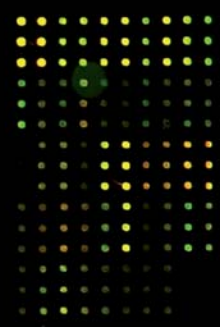
New-generation vaccines

– combining immunology with organic chemistry and bioinformatics
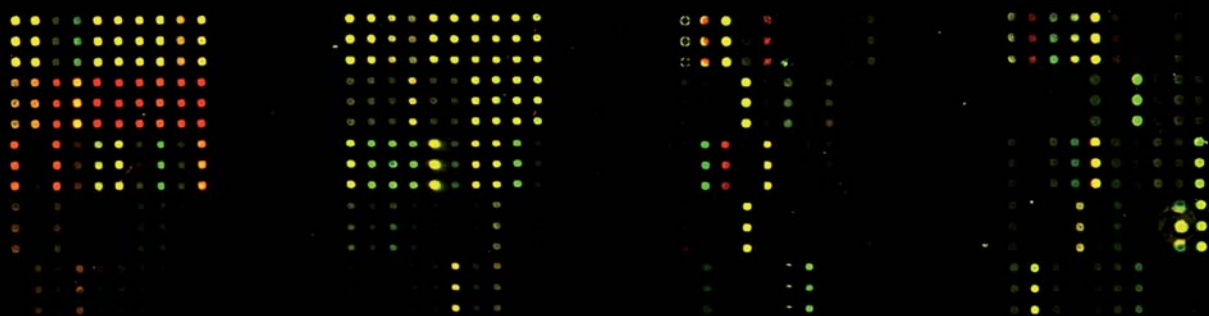
Peter M. H. Heegaard & Ole Lund
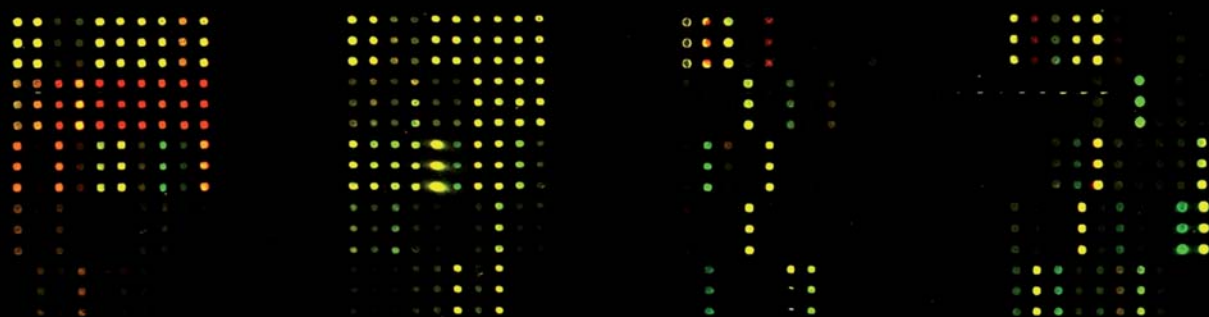
The challenge of infectious diseases

Infection and the response to infection have been long known in human and animal medical history. Inflammation is the immediate host response to infection, although it is not recognized as such, and Celsus quite precisely described inflammation in the medical literature as early as 100 AD. The somewhat delicate practice of protecting against infectious disease by injecting infectious biological material into an individual, known as variolation for smallpox virus, dates back to the early 1700s. This was later developed into the art of vaccination, which is the process of giving a vaccine to an individual. A vaccine traditionally contains modified pathogens (microbial agents that can cause disease) that will not cause disease but resemble the pathogen sufficiently to provoke the vaccinated host to become immune to the disease. Vaccination exploits the unique ability of the immune system of the host to adapt to and thereby remember challenges to an infectious agent after being exposed.

Many successes have been experienced in combating infectious diseases using vaccines, such as the eradication of smallpox, which was officially declared in 1980, and the dramatic decrease in formerly widespread diseases such as polio, measles and rinderpest (cattle disease). Nevertheless, infectious diseases are still abundantly present around the globe and account for some of the most fatal (Ebola virus), most globally widespread (malaria, tuberculosis and bacterial and viral diarrhea) and most enigmatic diseases (prion diseases) known.

Modern lifestyles give rise to additional complicating factors, including the increased mobility of humans (tourism) and animals (food trade), leading to a greatly increased rate of spread of pathogens, and climate change that leads to the redistribution of vectors (carriers of pathogens) for especially viruses to geographical areas in which human and animal populations are naive to the pathogen (without pre-existing immunity). Some of these factors contribute to the danger of new pathogens, as they can now potentially be distributed rapidly and become exposed to a large and diverse population. Examples include the new severe acute respiratory syndrome (SARS) virus in 2003, West Nile fever virus, first reported in about 1999 in Western countries (both human pathogens)

and bluetongue virus (infecting mostly cattle and sheep) on the rise in northern Europe, with the first case in Denmark in late 2007.

The incidence (number of cases per year) of antibiotic-resistant bacterial pathogens has increased dramatically. This has been caused by excessive use of antibiotics, which creates a selection pressure for the development of antibiotic-resistant bacteria. This is especially a problem in the hospital sector and in the intensive animal production worldwide. This renders the widespread use of antibiotics for treating bacterial infections problematic. Moreover, antibiotics cannot be used for prophylactic treatment (treatment before the onset of disease) of animals and humans, and no antiviral agents exist for general use against viruses.

Thus, vaccination still stands out as the best means of controlling or even eradicating infectious agents. However, the methods and strategies that have very successfully been applied to develop vaccines for a range of infectious diseases have proven inadequate for controlling a whole range of "difficult" old infectious diseases (malaria, dengue fever, *Shigella* spp.) and new ones (HIV). According to Stefan Kaufmann (*Nature Reviews Microbiology* 2007: 5: 491–504), "All the vaccines that can be developed by trial and error have now been developed." Some of these difficult infectious agents attack the immune system itself or are able to change rapidly to avoid the immune system.

New-generation vaccines that can target such infectious agents are thus needed. Although knowledge of the immune system has grown tremendously in recent years, these new insights have not generally led to the development of new generations of vaccines. In other words, immunological knowledge has not been applied fully for developing vaccines. The main reason for this is probably the success of the old-fashioned vaccines based on empirical knowledge about how to inactivate or kill microorganisms, how to formulate them with good adjuvants (vaccine helper molecules) and how to administer them in helpful doses to individuals. These vaccines have worked very effectively, but no one really knows how and why they worked at the molecular level. It has therefore not been possible to derive general principles based on old-school knowledge of vaccinology to design new vaccines using molecular components of the

immune system and of the microorganisms. It has therefore not been possible to develop new-generation vaccines with increased efficacy against the difficult pathogens, with minimized side effects and with generic applicability against new emerging pathogens.

This is now expected to improve since:

- new biological networks controlling immune responses have been discovered;
- systems biology wet-laboratory approaches that allow relevant and complete biological interactions to be monitored have become available;
- bioinformatics and systems biology approaches enabled by good computers and good computational methods, including neural networks, have been developed; and
- new methods for producing pure biomolecules with precise biological activities have been developed.

All of this is present within the DTU National Veterinary Institute and Department of Systems Biology. The National Veterinary Institute has ongoing research on several aspects of the immediate ("innate", see below) responses to infection, mainly in cattle and pigs, based on well-described experimental infection models, and microarray, quantitative polymerase chain reaction and antibody-based assays for host expression patterns, combined with advanced microscope equipment for in situ labeling of pathogens, host factors and cellular markers (laser confocal microscopy) and for precise sampling of minute amounts of tissue (laser capture microdissection).

The National Veterinary Institute also has ongoing research on the organic synthesis of replicas of microbial biomolecules that, when presented in multiple copies on polymers, are intended to work as very specific and chemically fully defined immunopotentiators for vaccine use. Finally, the DTU Department of Systems Biology has worked for many years on computational prediction of the binding of peptides to major histocompatibility complex (MHC) molecules from different species, based on binding data from analysis of interactions between the molecules in question and extending these to new molecules.

Here we describe this in more detail and emphasize DTU's current research in these areas and how this can be integrated to further the development of new and efficient vaccines, based on nature's own defense systems for the benefit of human and animal populations (Fig. 20.1).
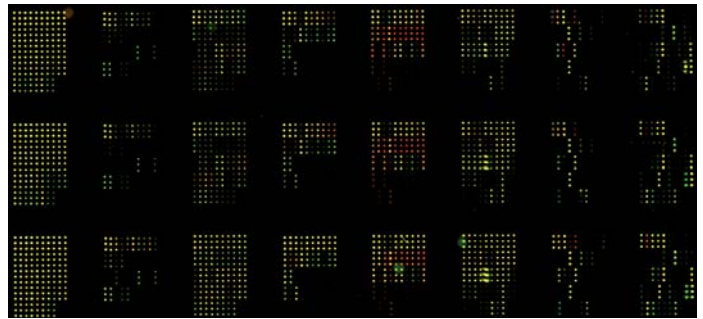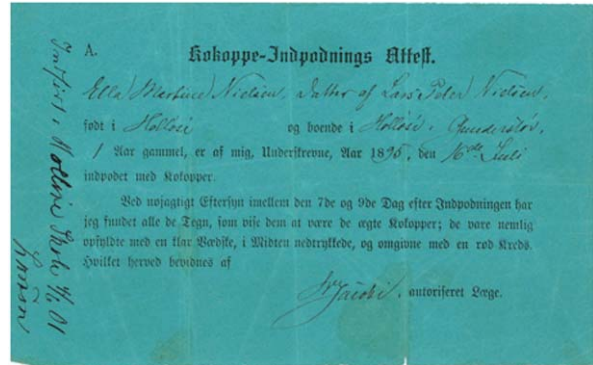

Fig. 20.1. Monitoring host responses now — and then. **Top**. A Danish cowpox vaccination certificate (1862) reporting a successful inoculation of cowpox in a 1-year-old girl, reporting that "By careful examination, between the seventh and the ninth day after inoculation" the doctor found "real cowpox blisters being whole and undamaged, filled with a clear liquid, slightly depressed in the middle and surrounded by a red circle". This is an example of a host response that was correlated with successful vaccination. **Bottom**. An immune-focused microarray representing more than 200 porcine genes relevant for the host response against infection. The microarray was used to analyze the expression of these genes in an uninfected pig (green) compared with a pig infected with the lung pathogen *Actinobacillus pleuropneumoniae* (red). If the expression of a certain gene is increased in the infected pig, the spot will turn out red; if it is decreased, it will be green; and if there is no change, it will be yellow. This gives a very detailed picture of the host response to infection. By correlating such data with data on the development of disease (clinical information), a "good" response correlating with good resistance to disease can be defined and used in developing optimal adjuvants and vaccines that induce similar "good" host responses.

The innate and the adaptive immune systems: new biological networks and mapping them by using new analytical methods

The immune system in higher vertebrates consists of an innate, nonadaptive part and an adaptive part that evolutionarily originated in jawed fish 600 million years ago. The adaptive part comprises effector molecules (antibodies or immunoglobulins) and cells (T and B lymphocytes, antigen-presenting cells and others).

The whole system is aimed towards optimizing the efficiency of immunity (rapidity, magnitude, adequacy and durability) while minimizing its activation by harmless substances, including self-molecules and commensal microorganisms that are part of the normal flora, such as in the gastrointestinal tract. To act most efficiently, these decisions have to be made as early as possible after the body is exposed to a potential pathogen. This is the job of the innate immune system, which initiates the response without being specific to the pathogen and without building up long-term immunity specific to the pathogen (immune memory). The adaptive immune system carries out both these activities. The innate immune system is always on the alert, reacts rapidly and is pivotal in setting off the right kind of adaptive immune responses. It has become increasingly clear that the seemingly uniform and invariable reactions of the innate defense response are in fact quite diverse at the molecular level and consequently lead to different adaptive immune responses.

The innate immune system comprises cellular factors such as the important antigen-presenting dendritric cells (see below) and macrophages and granulocytes that can immediately kill microbes. The innate immune system recognizes conserved molecular structures specific to microbial agents such as bacteria, viruses and parasites. These structures have been named pathogen-associated molecular patterns (PAMPs) and are recognized by corresponding innate host-cell pattern-recognition receptors. Occupation of pattern-recognition receptors by PAMPs leads to molecular signals that report on the nature of the microbe and control the ensuing early defense mechanisms and the balance of the subsequent adaptive immune responses. Interest in these host molecules reacting during the host response to infection has surged in recent years, and although many details have not been elucidated, the understanding of the nature and interactions of such factors will be pivotal in the rational development of new vaccines.

In the next step, the T and B cells of the adaptive immune system react to molecules from the invading microorganism (antigens). B cells produce antibodies that can neutralize microbes, and T cells can kill infected cells. These responses are long lasting and thus create an immune memory that is the basis of immunity. T cells do not recognize microbes directly but only when parts of its proteins called peptides are bound to the MHC molecules.

Fig. 20.2 gives a very simplified picture of the extremely complicated processes of the adaptive immune response. Binding the foreign antigen, a B-cell receptor leads to cellular uptake and degradation of the molecule into protein fragments (peptides) that are then bound by the cell surface MHC molecules. If an antigen-experienced ("primed") T cell then recognizes the peptide–MHC complex, the B cell will be licensed to produce its antibodies. Thus, in addition to being restricted by the binding to the MHC, the adaptive immune response also needs primed T cells to set off.

Importantly, the innate immune system controls the priming process. The innate cells responsible for priming are the dendritic cells that form networks throughout the body along the skin and mucosal surfaces where they continually take up antigens and become activated. Activated dendritic cells produce stimulating molecules (cytokines) that control the development of different effector T-helper (Th) cell types: cellular immunity. During infection, large numbers of activated dendritic cells with MHC-bound microbial peptides on their surface will migrate to the nearest draining lymph node, where they prime only the T cells that are specific for the presented peptides (Fig. 20.2). Depending on the types of T-helper cells being induced, the cellular immunity may cover the spectrum from immunity slanted towards dominant cellular reactivity being particularly efficient against, for example, viruses and intracellular bacteria (Th1 responses) to immunity that neutralizes predominantly extracellular pathogens and toxins and comprising a mixture of cellular and antibody-based immunity (Th2 responses). Importantly, after clearing the microbes, a subset of effector T cells
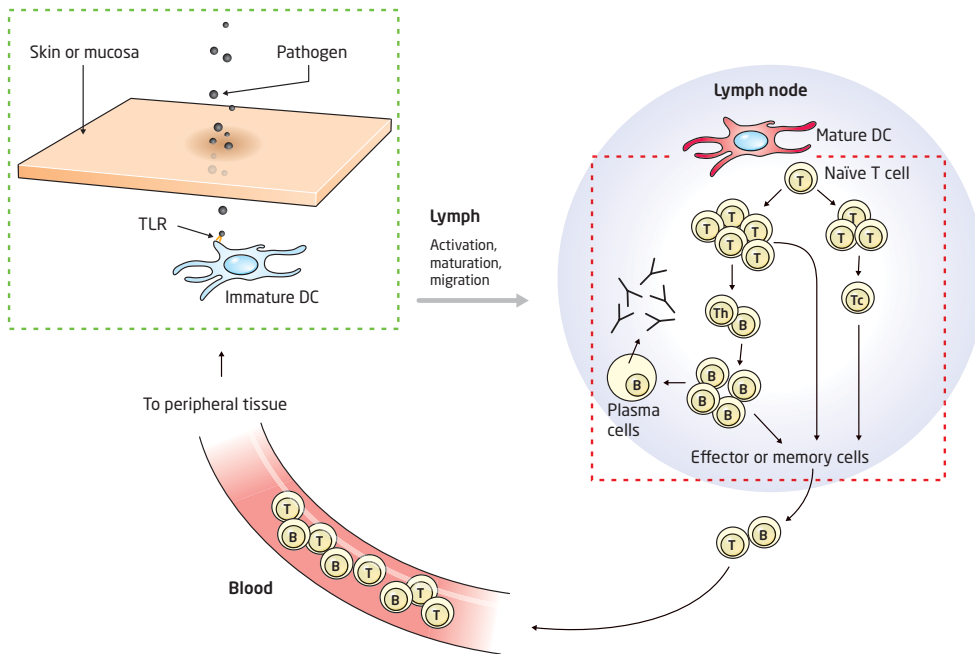
Fig. 20.2. Cell types and basic physiology of the immune system being activated by an intruding pathogen. TLR: toll-like receptors. DC: dendritic cells.

survives and differentiates into long-lasting memory T cells that respond rapidly to antigen even in the absence of co-stimulation by MHC-bound peptides. Likewise, memory B-cell responses are more rapid and generate higher-affinity antibodies of subclasses efficiently linking up with cellular immunity. This immune memory is a very important characteristic of the adaptive immune system and the reason that vaccination works.

Thus, MHC binding is a prerequisite for the induction of efficient full-spectrum immunity (cellular and antibody-based) as well as immune memory (long-lasting immunity). Another prerequisite is the efficient and relevant priming of T cells, an important activity of the innate immune response.

Host and pathogen variability: using computers to predict interaction between MHC and peptides

As described above, MHC molecules on the surface of antigen-presenting cells bind short peptides and present them to T cells. The peptides can be

regarded as fingerprints of the proteins from which they are derived. The T cells have been trained to distinguish between fingerprints of their own proteins (self-proteins) and foreign proteins, such as those stemming from an infectious agent. There are two major types of MHC molecules. MHC class I molecules sample from inside the host's own cells, and MHC class II molecules sample from the outside of these cells, thus complementing one another to give the immune system a total overview. Fig. 20.3 shows a MHC class I molecule displaying a bound peptide, seen from the perspective of a T cell looking down on a cell to check whether it is infected. If the T-cell receptor of the T cell can bind strongly to the peptide–MHC complex, it will be activated to kill the infected cell.

There are many slightly different MHC molecules with variable peptide-binding specificity, each variant of MHC molecule binding a "set" of peptides conforming to the specific binding pattern recognized by the specific MHC molecule variant. In humans, more than 3000 MHC variants have been discovered so far. Each MHC molecule variant has a different binding specificity and pres-
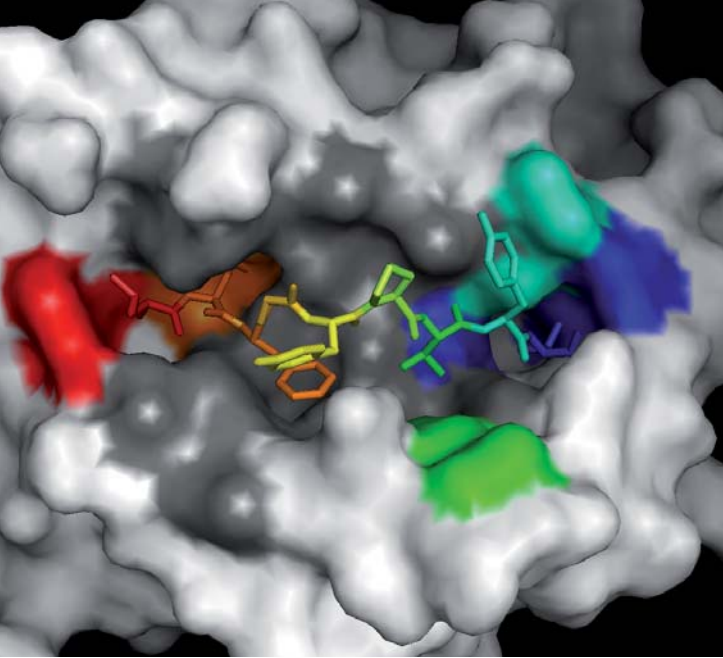
Fig. 20.3. MHC molecule in complex with a peptide. The MHC molecule is shown using surface representation, and the bound peptide is shown using sticks. The figure is based on the protein structure determined by Khan (Khan AR et al. *Journal of Immunology* 2000: 164: 6398-6405). The residues in the peptide are shown in different colors, and the surface of the MHC molecule is shown in the same color as the residue in the peptide with which it is reacting. Regions reacting with more than one residue in the peptide are shown in gray.

ents a distinct set of peptides to the immune system. The grooves of MHC variants have different shapes, which in turn is the reason why they bind different sets of peptides.

The diversity of MHC molecules means that any two people are likely to carry different sets of MHC molecules. If the same virus infects two people, they will thus display different fractions of its proteins to the immune system. The genetic mechanism behind the generation of this diversity of MHC molecules is not fully understood. It has been shown that individuals who carry different versions of each gene (heterozygous) rather than two identical copies (homozygous) are less likely to be killed by an infection. Another hypothesis is that people carrying rare variants of MHC molecules are more likely to survive an infection since there is greater selective pressure on microbes to adapt to the most common variants of MHC molecules.

In any case, a prerequisite for a peptide antigen to function as a vaccine is that it can bind to the MHC molecules of the individuals subjected to the vaccine. The vast genomic variation of the infectious pathogens and the diversity of the host cellular immune system are major challenges in attempting to identify such T-cell epitopes to be used for vaccine purposes. We and other researchers have developed computer tools during the past decade based on available sequence and binding data to predict which sets of peptides different MHC molecules can bind. The most successful methods have been data driven: using available experimental data about which peptides can bind as input and using this to predict which other peptides can bind. The simplest method is to look for a consensus motif. Fig. 20.4 shows examples of peptides that have been found experimentally to bind the human MHC variant called HLA A*0201. Eight different peptides, each with nine amino acid residues (shown in one-letter code) are shown. The letters in all positions seem to be fairly random, except for position 2, where only L (and one M) are found, and on position 9, where only four different amino acids are found (L, V, T and S). A simple rule (consensus motif) for finding binders is then to scan a protein for positions in its amino acid sequence where an L or an M is followed seven positions later by an L, V, T or S. Such algorithms are easy to implement in software but have limited accuracy. A more advanced method would be to consider the frequency of the amino acids on the different positions rather than just the binary information of whether an amino acid is found or not (peptide motif method). L, for example, has a much higher frequency than M in position 2. A prediction algorithm can therefore be constructed for each position in the peptide that looks up in a table the probability of having that amino acid in that position and then multiplies these probabilities to get a prediction score. Such matrix methods have been shown to have higher accuracy than the motif-based methods. These methods do not take into account possible correlations between different positions in a peptide, which determine whether a peptide will bind. For example, as seen in Fig. 20.3, the binding groove may not be able to accommodate two adjacent amino acids that both have large side chains. Methods that can

consider such correlations such as neural networks are currently among the most successful tools for predicting MHC-binding peptides. *Immunological Bioinformatics* describes these prediction methods in more detail (see More to Explore).

The methods described above can be used to develop specific prediction tools for a given MHC variant given that binding and sequence data are available for it. Thus, we have characterized the binding specificities for about 100 different human MHC variants. Even if this covers most of the most common variants, there is a long way to go to be able to cover all variants. We and other researchers have therefore recently developed pan-specific methods that can interpolate not only between peptides but also between MHC molecules.

All these methods use algorithms implemented on computers to produce lists of predicted MHC-binding peptides from the antigen and/or translated genome in question. Prediction algorithms are improved and new algorithms developed on computers as described above using available sequence and binding data.

In contrast to the empirical methods conventionally used for developing and producing vaccine, modern vaccine design relies on carefully selecting one or a few subunits of the pathogen (antigens) or even fragments of antigens (peptides). Modern vaccine design aims to achieve the difficult combination of stability, efficiency, minimal side effects, selective immunostimulation, low production price and high safety. For some vaccines, especially veterinary vaccines, the vaccine also often needs to induce antibodies that can be discriminated from the antibodies induced after natural infection with the pathogen in question (discriminating infected from vaccinated animals).

The MHC-binding prediction methods described here are a major help in avoiding as vaccine antigens peptides that do not bind to a high proportion of MHC molecules in a relevant human or animal population, thus securing the first prerequisite of a vaccine peptide antigen: that the MHC will present the antigen to the primed T cells during the immune response.

However, many such highly specific low-mo-

Fig. 20.4. Example of peptides binding HLA A*0201, showing the concept of binding patterns

lecular-weight peptide antigens, even if foreign to the host and good MHC binders, are not immunogenic by themselves because they cannot stimulate dendritic cells: the innate part of the immune response. Such otherwise ideal vaccine candidates need adjuvants to become adequately immunogenic.

Adjuvants: old methods and new problems and the potential of using synthetic replicas of biomolecules

An adjuvant may be defined as a substance enhancing the immune response to the molecules mixed with it. It will not change the antigenic specificity of the immune response, just its magnitude. In this way, an adjuvant may be compared with an enzyme speeding up a chemical reaction without changing the products of the reaction. Adjuvants, however, work on a much more complex system, and this may affect the type of immune response produced to a certain degree. Since adjuvants work by stimulating the innate (initiating) part of the immune system, innate immunology has been called "the science of adjuvants".

An ideal adjuvant–antigen mixture combines the precise minimal molecular antigenic structure necessary for conferring the desired immune specificity with a powerful and selective immuno-stimulation activity efficiently inducing the most optimal type of immunity (cellular, humoral, Th1, Th2 etc.) to the highest possible level. A specific challenge to modern vaccine development is that a Th1 response is needed to obtain protection against a range of important and widespread infections, but safe and efficient Th1-inducing adjuvants are generally lacking or unsuitable for human use. The development of adjuvants has taken place concomitantly with the development of vaccines and has historically been empirically based. Such traditional adjuvants are complex and somewhat undefined mixtures of surface-active compounds, microbial components and/or various polymers and lipids, which are very efficient in increasing the immunogenicity of co-injected molecules but are also largely not approvable for human use. Next to alum (various aluminum salts), water-in-oil and oil-in-water emulsions stabilized by surfactants are the most frequently used traditional adjuvants.

Alum is approved for human use but is not a good Th1 inducer.

Adjuvants can be classified as either delivery systems or immunostimulants. Some delivery systems function by aggregating and/or polymerizing antigens, minimizing their spreading from the injection site (depot effect), increasing their half-life and enhancing their uptake and processing by local antigen-presenting cells that are known to take up particles within specific size ranges by specific processing pathways. Immunostimulants used for adjuvant purposes are often microbial products (PAMPs; see above – these are among the most powerful adjuvants known) or their synthetic replicas or endogenous immunoactive compounds, such as cytokines, chemokines or co-stimulatory molecules.

We have studied the use of dendrimers for delivering and immunopotentiating antigens with low immunogenicity with the overall aim of developing generic all-synthetic and molecular fully defined adjuvants for vaccine purposes. Dendrimers (Fig. 20.5) are monodisperse, hyperbranched synthetic polymers first described in the late 1970s and available commercially in a range of types and sizes. Dendrimers consist of a core and a well-defined branching pattern defining the dendrimer generation. An increase in dendrimer generation leads to a doubling of the number of functional groups on the multivalent surface of the dendrimer. Molecules linked to the highly multivalent dendrimer surface are presented very efficiently to the surroundings and are freely accessible for interactions with surfaces and soluble molecules. The chemistry of dendrimer surfaces is very versatile, allowing a wide range of coupling chemistries.

Dendrimers show controllable cytotoxicity and interaction with cell membranes, biodistribution and biological half-lives, all of this mainly depending on the surface groups of the dendrimer. As the dendrimer surface is highly derivatizable, there is considerable room for creating dendrimers with desired specific biological properties. For example, lipids (for example, a tripalmitate PAMP (Fig. 20.5) or simple fatty acids) can endow dendrimers with membrane-interacting and -immunostimulating properties, and high-generation amino-surfaced dendrimers can result in inflammation and activation of the complement system. Good induc-
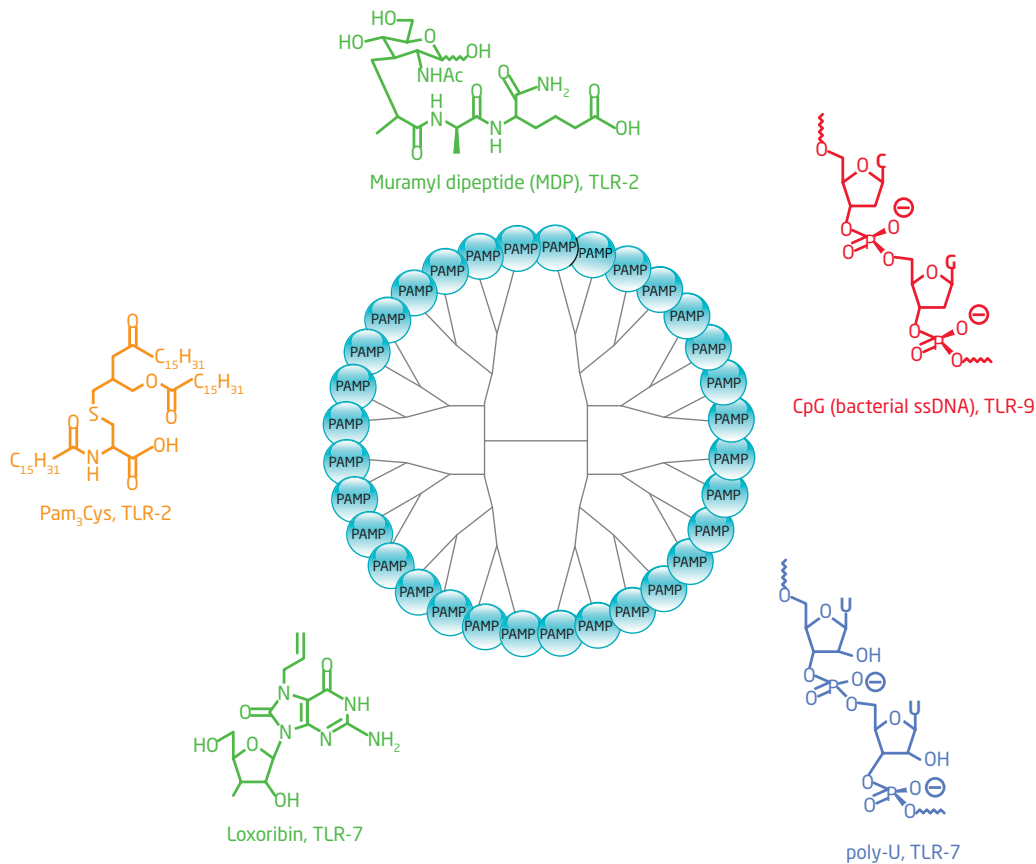
Fig. 20.5. Dendrimer pathogen-associated molecular pattern (PAMP) concept showing the basic dendrimer scaffold decorated with surface-bound PAMP molecules and examples of relevant PAMP molecules. TLR: toll-like receptor.

tion of immunity against a peptide can be obtained by including a peptide antigen of interest in the structure.

The simplest explanation for the effect of such lipid moieties is that the lipids promote contact with the cell surface membranes of antigen-presenting cells and thereby optimize exposure of an attached peptide antigen to these cells and act as immunostimulants activating the cells. Further, MHC restriction of T-cell activation can be circumvented by introducing broadly reactive (universal or promiscuous) T-cell peptide epitopes into the dendrimer. T- and B-cell epitopic peptides may also be combined by dimerizing two dendrimers, each containing its own type of peptide. Dendrimers may also contain targeting moieties in addition to immunostimulating ligands and antigenic moieties to facilitate interaction with highly

reactive cells of the immune system. Mannose moieties, for example, target the construct to dendritic cells carrying a mannose binding receptor on their surface.

We are studying the possibilities of coupling various PAMP-related molecules to dendrimer surfaces to obtain specific immunostimulation profiles for combination with antigens to obtain a desired type of immune response with the right antigenic specificity. As a first step, carefully synthesized and characterized PAMP-coupled dendrimers are biologically characterized with respect to their ability to activate white blood cells (from pig blood). This is measured by the induction of specific host factors related to different aspects of the immune response in these cells, using enzyme-linked immunosorbent assays for detection at the protein level and microarrays and quantitative reverse-

transcribed polymerase chain reaction methods for detecting specific messenger RNA molecules. Preliminary findings indicate that coupling to dendrimers potentiates the immunostimulating property of a range of PAMP-derived structures and that different constructs can induce different sets of host factors.

Conclusion and outlook

It is not easy to predict whether an immune response and the ensuing immune memory directed toward a given antigen will actually protect against disease caused by the infectious agent. As explained above, the first prerequisite is the ability to obtain a high level of immune reactivity against the agent, and this depends on selecting good T-helper epitopes and efficient adjuvants, which may additionally be tailored to bias the immune response in the most optimal direction. The ability of the immunized host to show early and topically precise responses at the initial site of infection is highly important, and the outcome of the race between the memory-based immune reactivation and the pathogen-driven disease development is decisive in the outcome of the infection. In the optimal situation, pre-existing antibodies are present to neutralize a pathogen or its toxins immediately during the establishment of the infection, and vaccine-induced immunity boosted at the same time

by the infection will cooperate with the pre-existing antibodies to eradicate the pathogen (sterilizing immunity). Systemic and mucosal immunity can interact in this and constitute the most efficient immunity achievable to completely prevent the establishment of infection.

Future vaccines for protecting against "difficult" infections need to include efficient Th1-inducing adjuvants and should preferably have the ability to induce immunity at mucosal surfaces. To achieve this with minimal adverse effects and maximum reproducibility, such new adjuvant systems should preferably be molecularly defined and may comprise delivery systems combined with immunostimulants to induce Th1-biased immune responses. Further, such vaccines when employing minimal antigens will need to consider that obtaining the T-cell stimulation needed for an optimal adaptive immune response and for the efficient induction of immune memory (the hallmark of a good vaccine) requires defining peptides with good MHC binding. Dendrimers may be useful generic platforms for developing such defined and safe vaccines with new properties and application potential.

The authors

Peter M. H. Heegaard:
MSc in Biochemistry, 1985 and PhD, 1992, University of Copenhagen. Senior Research Fellow. Protein Laboratory, 1990–1992. Senior Scientist, National Veterinary Institute, DTU, 1993–2008. Professor with special responsibilities in the innate immunology of pigs and cattle, DTU National Veterinary Institute, 2008–. Leads several research projects on immunology, especially infection biology and innate immunology of pigs and cattle, bioorganic chemistry for vaccine and diagnostic purposes and prion disease biology.

Ole Lund:
MSc in Engineering, 1991 and PhD, 1995, DTU. Senior Scientist, Bioinformatics Advanced Technologies, Hørsholm, Denmark, 1998–2001. Associate Professor, DTU, 2002–2007. Professor, Center for Biological Sequence Analysis, DTU, 2007–. Coordinator and principal investigator on the projects Discovery of Epitopes of NIAID Category A–C Pathogens Using Bioinformatics and Immunology and Genome- and HLA-wide Scanning and Validation of Cytotoxic CD8 T-cell Responses against *Mycobacterium tuberculosis*.

More to explore

Lund O, Nielsen M, Lundegaard C, Kesmir C, Brunak S. *Immunological bioinformatics*. Cambridge, MIT Press, 2005.

Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* 2009: 61: 1–13.

Boas U, Cristensen JB, Heegaard PMH. *Dendrimers in medicine and biotechnology*. London, RSC Publishing, 2006.

Skovgaard K, Mortensen S, Boye M, Poulsen KT, Campbell FM, Eckersall PD, Heegaard PMH. Rapid and widely disseminated acute phase protein response after experimental bacterial infection of pigs. *Veterinary Research* 2009: 40: 23.

Kaufmann SHE. The contribution of immunology to the rational design of novel antibacterial vaccines. *Nature Reviews Microbiology* 2007: 5: 491–504.

Pulendran B, Ahmed R. Translating innate immunity into immunological memory: implications for vaccine development. *Cell* 2006: 124: 849–863.
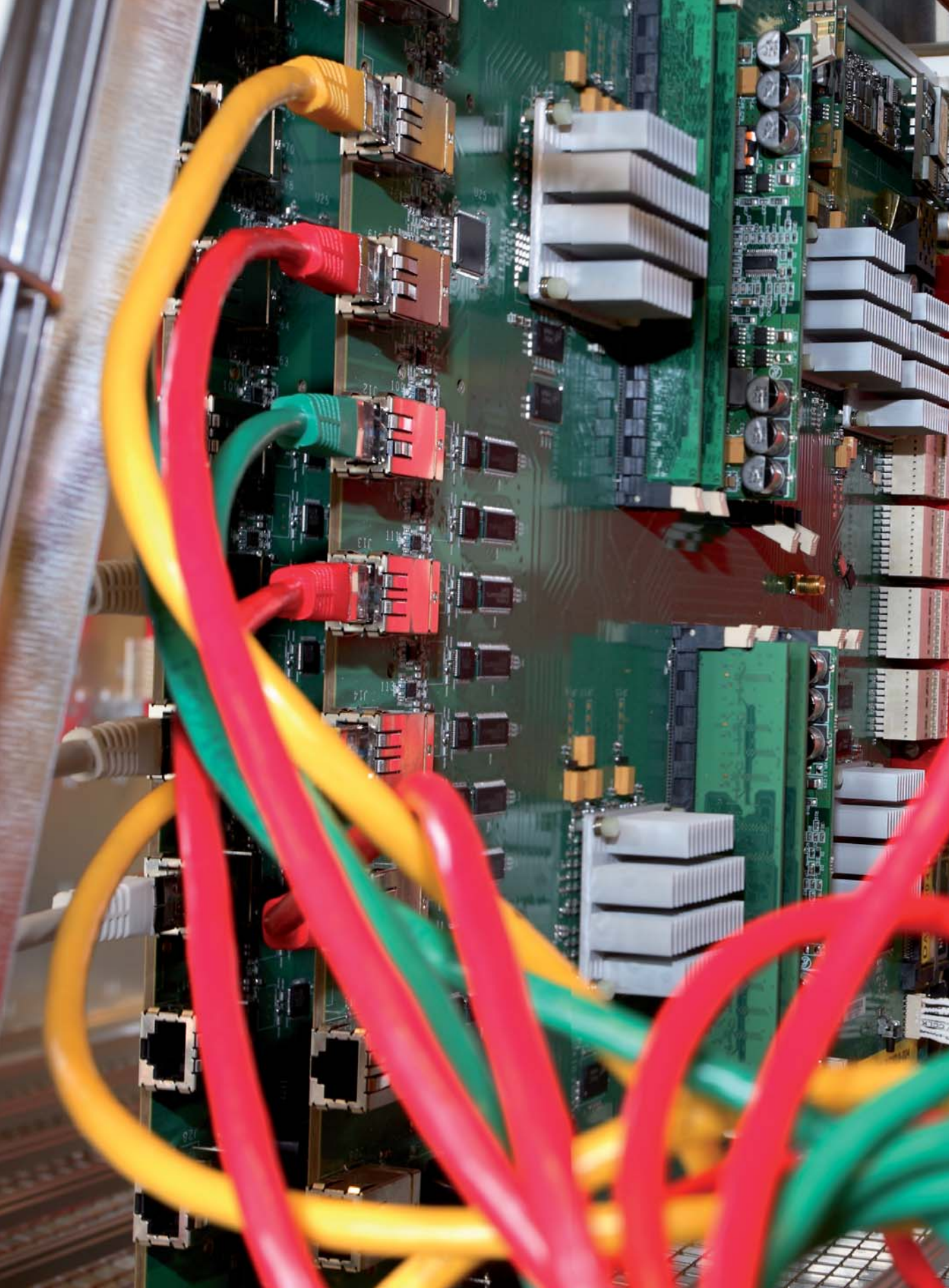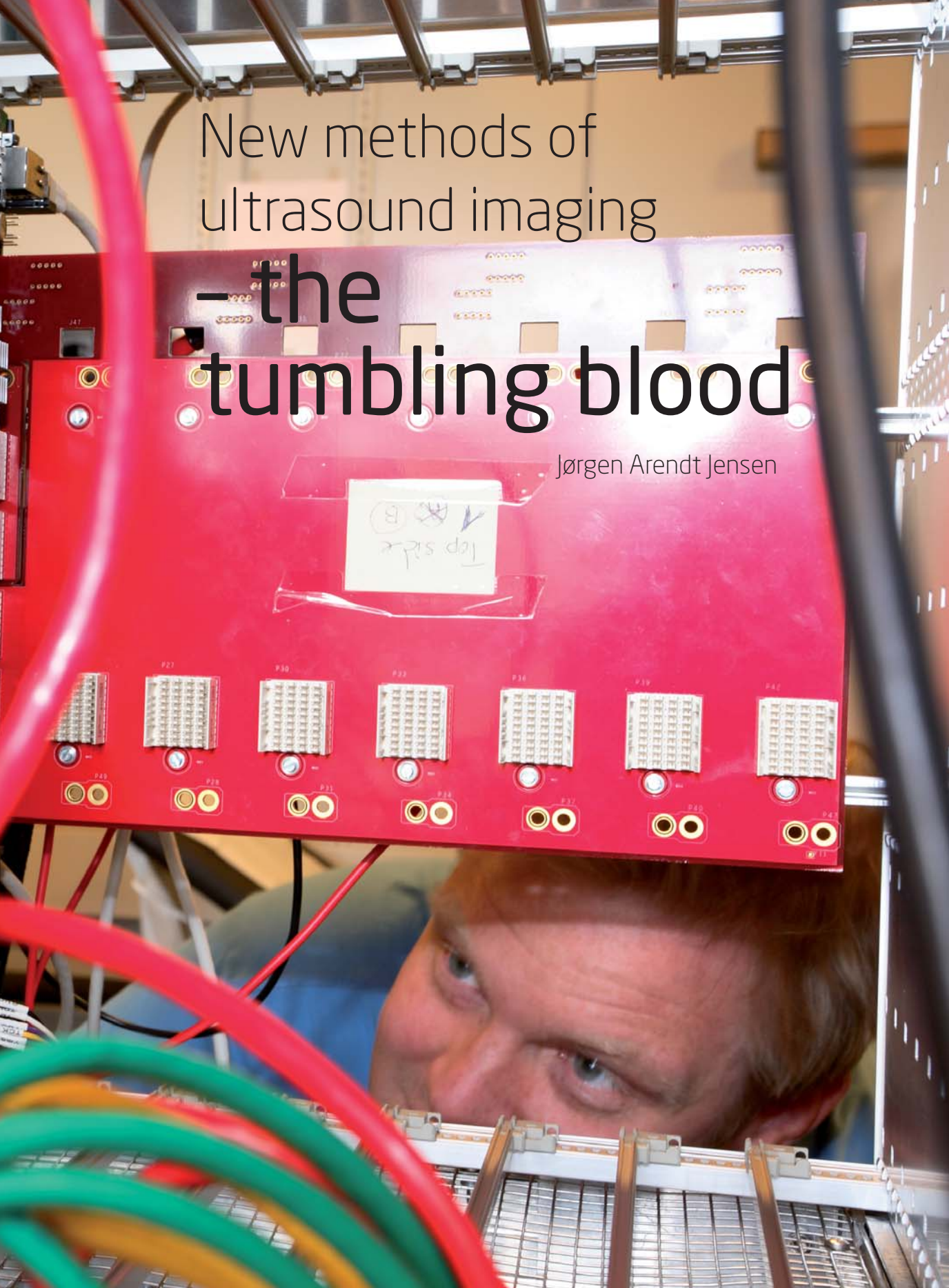
New methods of
ultrasound imaging
– the
tumbling blood

Jørgen Arendt Jensen

Millions of ultrasound examinations are conducted each year to reveal the development of fetuses and the functioning of the heart and to diagnose nearly any soft tissue structure in the human body. Ultrasound is frequently used in modern hospitals because it is convenient and safe. It uses nonionizing radiation, which is harmless to the patient and painless, and ultrasound examinations can be conducted without contrast agents and positioning the patient into narrow tubes as in computed tomography (CT) and magnetic resonance imaging (MRI).

The major advantage of ultrasound compared with other methods of imaging is its interactivity. It is conducted in real time, and the images are shown like a video stream of up to 60 frames per second. It is thus possible to study the dynamics of the body and manipulate the tissue to see whether it is hard or soft and to easily take out tissue samples for later histological studies. It is even possible to study the pulsating blood flow in the human body and thereby evaluate, for example, the functioning of the umbilical cord or the flow around heart valves.

Ultrasound, however, also has problems. The grainy image is different from other methods, creating difficulty in interpreting the image, and physicians often have to spend significant time to learn how to use ultrasound. The sharpness of the image depends on the depth in tissue, and penetration can be a problem for obese people, a problem that is increasing rapidly. Further, visualizing the blood flow depends on the direction of the ultrasound beam, making flow images difficult to interpret for complex types of flow such as those at branching vessels or valves. There is thus a need to investigate whether modern ultrasound can be further developed to increase its clinical efficacy. This has been the focus of the DTU Center for Fast Ultrasound Imaging since it was established in early 1998 as a collaboration between the Technical University of Denmark, BK Medical, Herlev University Hospital and Rigshospitalet – Copenhagen University Hospital by a grant from the Danish National Research Foundation's THOR program. The mission of the Center is to develop novel ultrasound methods and systems that can significantly alter the clinical potential of ultrasound imaging.

What is wrong with traditional ultrasound imaging?

Conventional ultrasound images are acquired by emitting a short ultrasound pulse of about 3–10 MHz into the body. Sound at these frequencies can be focused like light and can be directed in one image direction. The sound propagates in the tissue at about 1540 m/s for soft tissue. Depth in tissue and time are proportional, and the arrival time of the reflected signal indicates the depth from which the signal originates. The tissue structures reflect and scatter the ultrasound, which is received by the same transducer. In the 1970s and 1980s, the sound was emitted by concave single elements with a fixed focus, and imaging was performed by moving or rotating this element. Modern scanners use multi-element probes using 32 or 64 elements at the same time to emit and receive the ultrasound. The beam is then focused by delaying the signal to each element relative to the other elements to form a concave wave front. In receive mode it can be dynamic, and the focus can be made to track the beam down through the tissue. In transmit mode, however, only a single focus is possible, as the sound cannot be manipulated when it has left the transducer. Modern ultrasound images are therefore only optimally focused at one depth, and the image quality degrades before and after the transmit focus depth.

Another problem with modern ultrasound is the sequential data acquisition: one image line is acquired at a time. For a depth of D, one line takes

$$T_l = \frac{2D}{c}$$

seconds to arrive, when the speed of sound is c and the sound has to travel both forth and back to the ultrasound probe. This restricts how quickly images can be displayed, as a limited amount of lines can be acquired per second. This is often a problem if deeper-lying structures are to be imaged rapidly, such as the heart. The problem is significantly aggravated for three-dimensional ultrasound, where volumes of data have to be acquired, and this very significantly limits the volume rate.

Commercial scanners can also estimate the velocity of the blood. This is performed by repeatedly acquiring ultrasound lines from the same direction. Comparing the signals can then reveal the

motion from line to line and yield the velocity. This corresponds to taking snapshots of a moving object. Measuring the motion between the two images and dividing by the time between images gives the object's velocity. The basic drawback with this technique is that it can only measure the velocity towards or away from the transducer. The blood vessels often run parallel to the skin surface and the systems, and thus the velocity component transverse to the ultrasound beam cannot be measured. Further, the blood flow in the body is often not laminar and does not move in just one direction. Around valves and in bifurcations, the pulsating flow is much more complex.

A possible solution – synthetic aperture imaging

As indicated above, there are many good reasons for trying to improve ultrasound imaging, such as the focusing, frame rate or velocity estimation. The research field is very multidisciplinary, encompassing acoustics, data acquisition systems, signal and

image processing as well as clinical studies. The breakthroughs often come from combining novel acoustic fields with new signal-processing schemes, and this can only be tested realistically by constructing new systems that can acquire such new field signals.

One possible method is to use synthetic aperture imaging. The basic principle is to acquire a complete set of data and then synthesize or reconstruct the ultrasound image from this data. Fig. 21.1 shows the acquisition of data.

From these data, a whole image can be created with low resolution, since there is only focusing during the reception stage. The next element is used in transmit mode, and the data are stored. This is done for all elements in the transducer, and a complete data set for all combinations of transmit and receive elements is acquired. Combining all these low-resolution images will yield a high-resolution image, which is dynamically focused in both transmit and receive modes. Synthetic aperture imaging has several advantages.
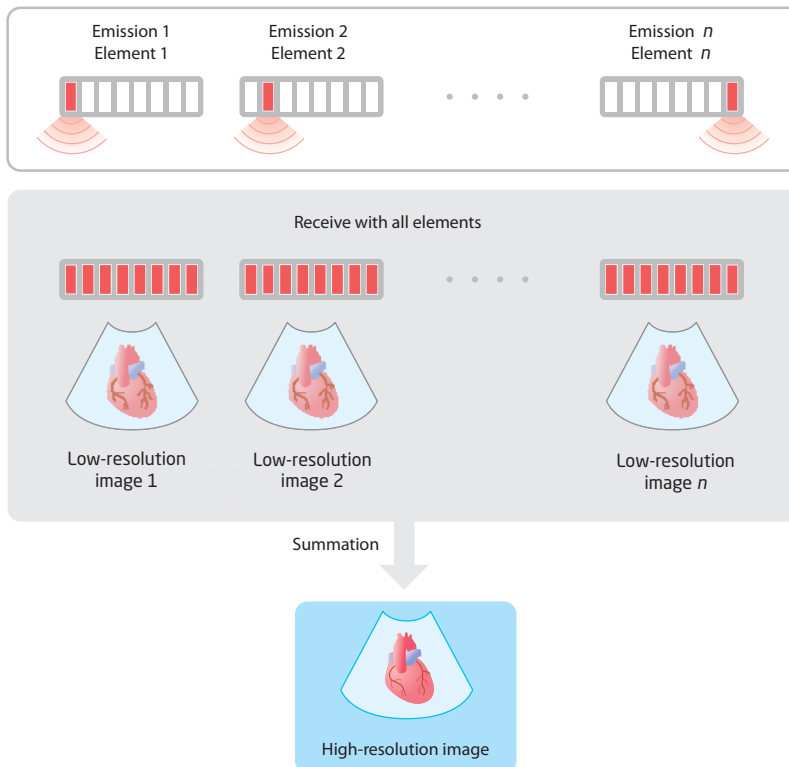

Fig. 21.1. Acquiring data using a synthetic aperture

- The precise origin of the ultrasound and its reception point are known. This enables precise travel times to be calculated from transmit to receive and thereby optimize focusing in both transmit and receive modes. The images therefore have the best possible sharpness for all depths. This also increases the contrast.
- A new image can be made after each emission, and very rapid frame rates can be achieved, making it an ideal candidate for three-dimensional imaging.

Synthetic aperture imaging, however, also has very significant drawbacks that have prevented the method from being used commercially. The penetration depth is very limited, since only a single element is used in transmit mode. Motion is a potential problem, as data used at one point in the image are acquired over several emissions. This also precludes estimating velocity, which is thought to be impossible for synthetic aperture data. The last problem is the very high number of calculations needed. Complete data sets are nice, but all the data have to be processed. This requires 100 to 200 times the processing power of current ultrasound scanners, and no systems exist for synthetic aperture imaging. The challenge was therefore to address all these issues and to try to solve them.

Making it work in the clinic

The first step is to actually get synthetic aperture data. We followed two approaches: simulation and measurements. The first entailed further developing our simulation tool to make it rapidly simulate synthetic aperture data. The program can model the transducer, its emission of sound, tissue interaction and the reception and processing of the data. It has been instrumental in developing the new techniques and revealing their performance in Monte Carlo simulations. This has often involved very large simulations, and this has been performed on the 100-CPU Linux cluster of the DTU Center for Fast Ultrasound Imaging. The Center put the program on the Web so that other research groups and companies can use it free of charge, and this has resulted in its acceptance as a gold standard for simulating ultrasound systems.

The second approach for getting synthetic aperture data is through real-time measurements, which is considerably more cumbersome as no systems existed for this. Commercial systems do not have the necessary memory capacity, and research systems cannot be bought. In collaboration with I-O Technologies, we therefore developed the research scanner RASMUS from 1998 to 2002, especially constructed for acquiring and storing synthetic aperture data. The scanner is capable of sending out any kind of signal in 128 simultaneous channels and of sampling 64 channels simultaneously at 40 MHz and 12 bits precision. It generates 5.1 GB of data per second and stores them in its 16 GB of memory. It then takes 8 hours to store on disk and up to 1 week of processing before the 3-second real-time synthetic aperture video is ready for viewing. The system is very flexible and can be programmed in a special language for doing all kinds of either traditional or synthetic aperture imaging or a combination of the two.

The first problem in synthetic aperture imaging is the poor penetration due to using just one element in transmission. This can be solved by sending out more energy by using a longer pulse, but this reduces the resolution in the image. The solution is to use frequency coding, where the frequency changes as a function of time (Fig. 21.2).

Processing the received data by using a specially matched filter will compress the pulse again, as the early frequencies are delayed more than the late frequencies, enabling more energy to be transmitted and still maintain resolution. A second approach uses a combination of transducer elements that are defocused to transmit energy in the whole region. Combining the two actually makes it possible to penetrate further into the tissue with synthetic aperture imaging than conventional imaging (Fig. 21.3). Here a tissue-mimicking phantom has been scanned using both the conventional approach and the synthetic aperture principle. The bright white targets are nylon wires, which can be seen down to 6 cm for the conventional image and down to 10 cm in the synthetic aperture image. This improvement of 50% shows that synthetic aperture imaging should be ideal for obese patients. The wires are also smaller, indicating that the resolution of the system is higher.
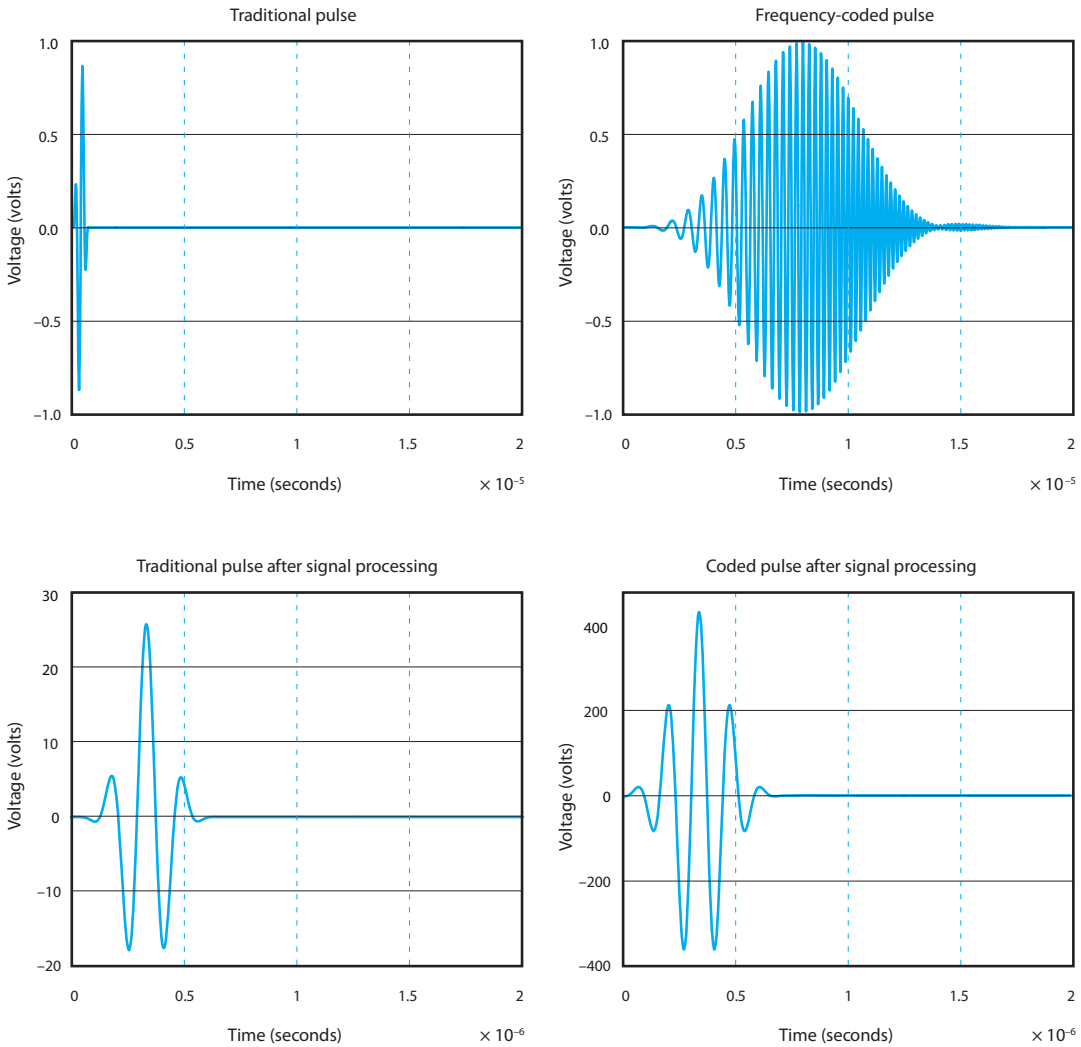
Fig. 21.2. Coded ultrasound pulse. The traditional pulse is in the left column and the coded pulse in the right column. The emitted pulse is in the top row and the processed pulse in the bottom row. They nearly have the same length, but the amplitude is higher for the coded pulse. This gives rise to an increase in signal-to-noise ratio and thereby imaging depth.

The next step is to see whether these advantages actually translate into a better image in the clinic and that the motion of the tissue does not distort the images. We used the RASMUS scanner to acquire both a conventional image and a synthetic aperture image at the same time. We can then compare the actual imaging, as the patient, scan time, transducer and measurement systems remain the same. Three seconds of video were acquired for seven people and three different scans. The data were processed and displayed in a double-blinded trial to three experienced physicians. Penetration depth increased using synthetic aperture. Image quality also increased ($P < 0.001$). The results show that in vivo ultrasound imaging using a synthetic aperture is feasible for abdominal imaging without severe motion artifacts.
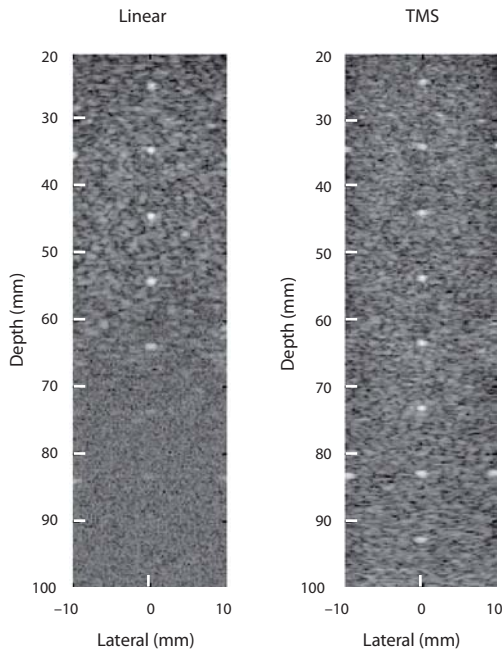
Fig. 21.3. Comparison of a conventional ultrasound image (**left**) and synthetic aperture image (**right**) of an artificial tissue phantom. The bright white targets are nylon wires. TMS: temporally encoded multi-element synthetic transit aperture.

The tumbling blood

One of the main obstacles to introducing a synthetic aperture system on the market is that blood flow imaging is thought to be impossible. Data are acquired over several ultrasound emissions, and the blood cells move between emissions. The key idea to solving this problem was to ignore the motion. It is possible to compare two motion-blurred images and still determine the motion and, thus, velocity. In synthetic aperture imaging, this is performed by only measuring a few emissions and then repeating the process. The method has several advantages. Unlike conventional imaging, in which only one direction is probed at a time, this approach probes the whole imaging field at the same time. Data for flow measurements are thus available for all points in the image at all times. This has enabled very rapid flow imaging. Fig. 21.4 shows the first synthetic aperture flow image. The red color shows velocities away from the transducer and blue towards it. The color intensity indicates the

magnitude of the velocity. The image has been generated from only 24 pulse emissions, enabling more than 1000 of these images to be acquired and displayed per second. A conventional scanner would use more than 400 emissions for one image like this. The new approach thus enables the flow dynamics to be followed very closely with good temporal resolution.

The complete data set also makes it possible to focus data in all directions, not only along a certain direction from the transducer. Data can be focused along the flow direction and can be tracked precisely along this direction so that any velocity vector can be detected including transverse to the normal ultrasound directions. This paves the way for truly quantitative velocity imaging. The synthetic aperture data set is also continuous in time for all positions in the image. The velocity estimators can therefore use as much data as needed for obtaining high precision. This is shown in Fig. 21.5, which displays several velocity profiles from a flow rig. Here the velocity is estimated at 90° or transverse to the ultrasound beam, where a normal scanner would not be capable of estimating any velocity. The method estimates the correct velocity with a standard deviation of 1.2% using 128 emissions, enough to make more than 50 images per second for the full heart.
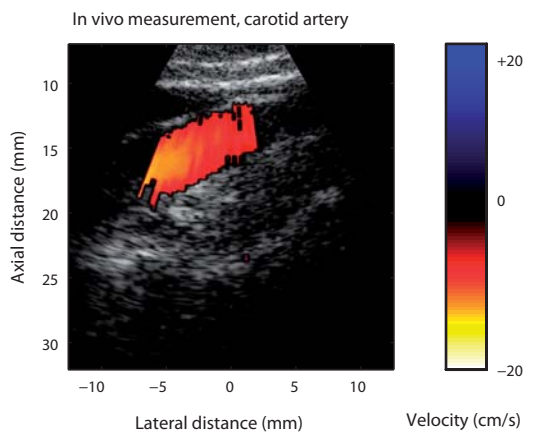

Fig. 21.4. In vivo synthetic aperture color flow image of the carotid artery scanned with a 7-MHz linear array transducer on a healthy 29-year-old man
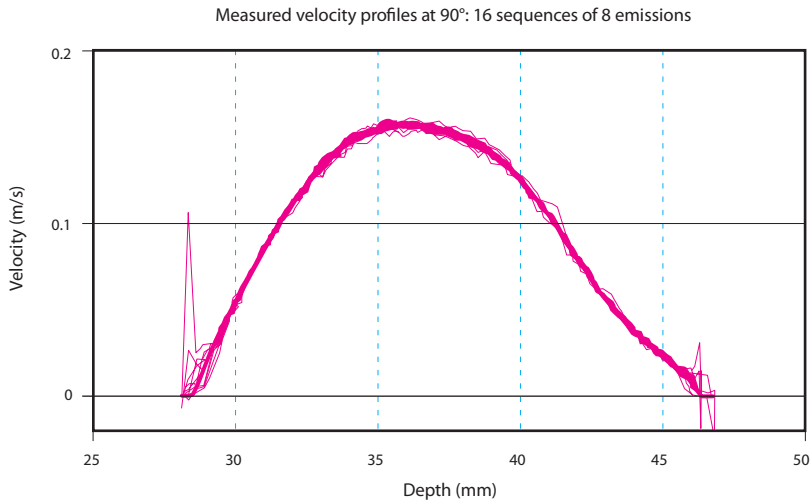
Fig. 21.5. Twenty independent estimated profiles from the flow rig at a 90° flow angle

The Center for Fast Ultrasound Imaging has developed several other methods for estimating vector velocity that are being studied extensively in preclinical trials, revealing new information about the human vasculature. Fig. 21.6 shows an example of this, displaying the vector velocity in the carotid bifurcation supplying the brain with blood. The arrows give the velocity vectors and the colors give the velocity magnitude. The flow is close to transverse to the ultrasound beam at many positions and far from parallel to the vessel surfaces at many positions. There is a marked change in velocity direction between the two images, and the flow swirls at the carotid bulb in the lower part of the image. These changes happen within less than 50 ms, which shows the importance of both rapid imaging and being able to detect the full velocity vector.
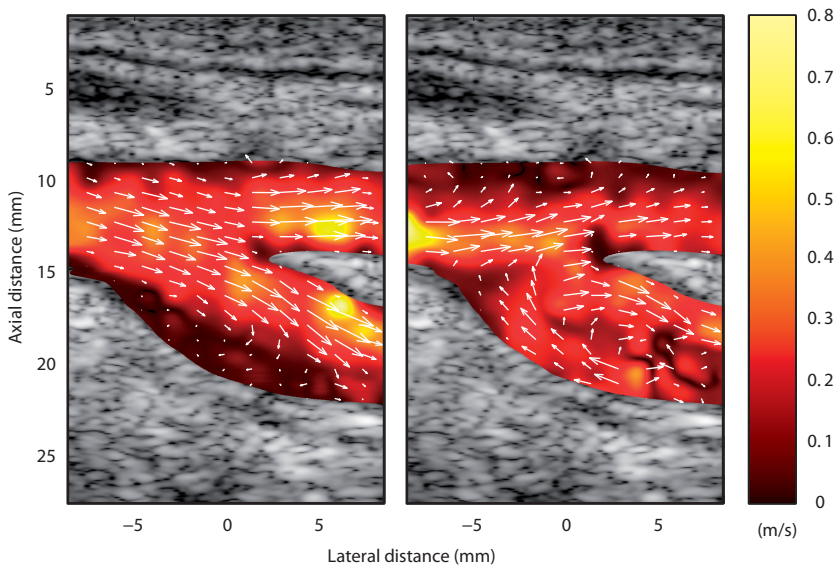

Fig. 21.6. The vector flow in the bifurcation of the carotid artery at the time of the peak systole (**left**) and shortly after the peak systole (**right**)
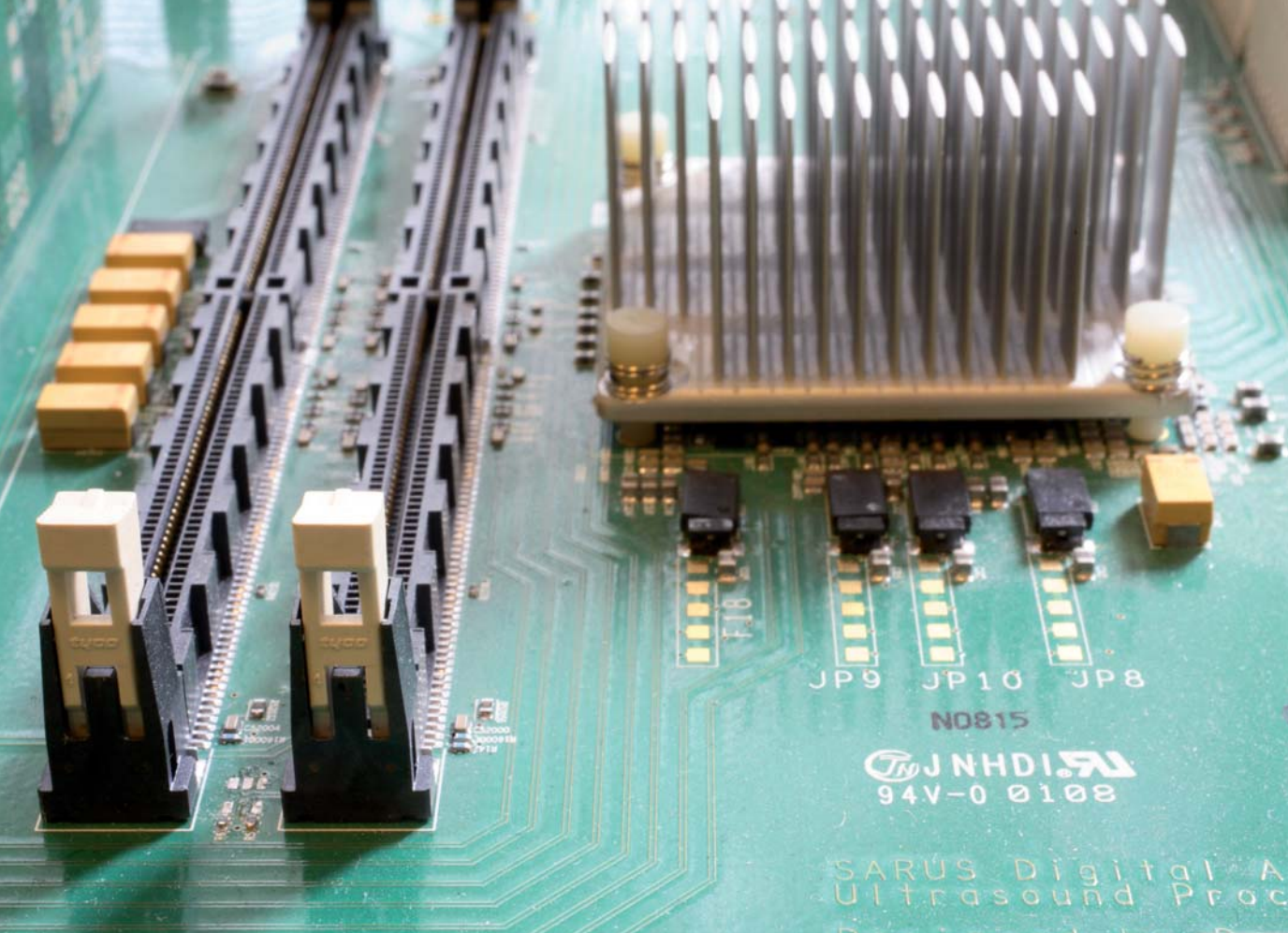
Fig. 21.7. Part of the newly developed central board in the SARUS (Synthetic Aperture Real-time Ultrasound System) research platform for faster data acquisition and real-time processing

What's next?

The images shown have all been acquired with the RASMUS system in real time and processed offline by our Linux cluster. It can take several days and even weeks to process the 16 GB of data, and this makes it practically impossible to perform larger clinical studies. There is thus a real need for real-time performance to fully exploit the interactive aspects of ultrasound. In 2005, we therefore joined forces with Prevas A/S to start making the next research platform SARUS (Synthetic Aperture Real-time Ultrasound System) for faster data acquisition and real-time processing. Fig. 21.7 shows the central board in the system. It is capable of sending out 16 signals in parallel and sampling 16 signals at 70 MHz and 12 bits. Most importantly, the system houses five large field-programmable gate arrays

that each contains 160 signal-processing units working at 500 MHz. The complete system houses 64 of these boards and is capable of making 25,600 billion multiplications per second. The system can easily be loaded with new code for implementing any kind of conventional and synthetic aperture imaging. The system has 500 GB of RAM and can generate 140 GB of data per second in its 1024 channels, which can be processed in real time. All these resources enable interactive imaging and further advancing synthetic aperture ultrasound imaging. The system will be operational in 2009.

The SARUS system can be combined with new two-dimensional matrix transducers having $32 \times 32 = 1024$ elements. Using these, the ultrasound beam can be focused in a volume and three-dimensional data can be acquired. Combining this

with three-dimensional estimation of vector velocity gives data sets with seven dimensions when time is included. The new system is capable of acquiring such data, and our challenge is now to develop the methods and estimators capable of extracting this information. A challenge is then to present this to the user and visualize how complex flow around, for example, a heart valve can best be presented.

Conclusion

The advancement of medical ultrasound images entails combining a large array of disciplines. The new methods for anatomic imaging and the display of vector velocities all rely on advances in the generation of acoustic fields, the acquisition of complete data sets, the development of new signal-processing methods and estimators, efficient implementation that can function in clinical settings and the demonstration of its usefulness and efficiency in the clinic. It inevitably requires a large array of people, collaboration between institutions, good experimental facilities and generous funding from outside sources. We have been fortunate to have this during the past 11 years.

More to explore

Hansen KL, Udesen J, Gran F, Jensen JA, Nielsen MB. In vivo examples of flow patterns with a fast vector velocity ultrasound method. *Ultraschall in der Medizin – European Journal of Ultrasound* (in press).

Jensen JA, Nikolov S, Gammelmark KL, Pedersen MH. Synthetic aperture ultrasound imaging. *Ultrasonics* 2006: 44: e5–e15.

Jensen JA, Hansen M, Tomov BG, Nikolov S, Holten-Lund H. System architecture of an experimental synthetic aperture real-time ultrasound system. In: *IEEE Ultrasonics Symposium 2007*. Paris, IEEE, 2007: 636–640.

The author

Jørgen Arendt Jensen: Professor of Biomedical Signal Processing, DTU Department of Electrical Engineering and Adjunct Professor, University of Copenhagen. Founder and head, Center for Fast Ultrasound Imaging, DTU. He founded the Biomedical Engineering Group at the DTU Department of Electrical Engineering and initiated the degree program in Medicine and Technology jointly offered by DTU and the University of Copenhagen.

Microfluidic devices for
health applications

Jörg P. Kutter

May 4, 2020. It is early morning in a one-family house in suburban Copenhagen. Christian Hansen wakes up to the chime of the alarm in his personal digital assistant. The device wishes him good morning and outlines his daily schedule, including a visit to a physician for his semi-annual health check. The biosensors built into the toilet bowl quickly assess Christian's state of health by performing a few rapid screening tests. No direct feedback is given, since no alarming results are found. However, a note is beamed to the diet log in Christian's personal digital assistant advising taking it easy on spicy food. Further, the summaries of the biosensor readouts are directly forwarded to his physician's office to be merged with his medical entry in the central database of the Health Care System Denmark.

Christian showers, dresses and goes to the kitchen to meet the other family members. The house computer suggests a breakfast tailored to Christian's recent biomonitoring results, medical history and the current contents of the fridge and the cupboards. It also suggests supplements based on the weather forecast and Christian's schedule.

Christian cycles to work. At a main intersection, displays underneath the traffic lights summarize the current air quality for the most common pollutants and the pollen counts. Later, Christian leaves for his physician's office. After a general examination, checking blood pressure, heart and lungs, the assistant pricks Christian's finger to sample his blood. A few droplets are directly collected from the fingertip onto the sample patch of a flat piece of polymer the size of a credit card. The assistant puts it in a shoebox-sized machine on the lab bench and asks Christian to wait in the reception area for a few minutes. After 15 minutes, Christian and his physician discuss the results of blood screening tests, including several protein biomarker fingerprints and single nucleotide polymorphism tests for most common types of cancer and neurodegenerative diseases. The test results are automatically compared with previous results to discover trends or acute changes and with blinded results from Christian's parents and siblings in case of genetic predisposition for diseases. One condition was flagged as "yellow", and the physician suggests medication. The physician sends the relevant medical data, including genomic and proteomic data, to a pharmaceutical company of Christian's choice, which prepares a tailor-made, personalized cocktail of drugs with just the right dosages and ships it directly to Christian. The physician explains that he will receive an applicator with pain-free microneedle arrays to introduce a colloid solution of drug-loaded nanoparticles into the bloodstream to ensure proper dosage and delivery solely to the targeted tissue.

Background

These examples may sound like science fiction. Nevertheless, most of these examples are the subjects of intense research efforts, and applications of the principles and devices developed may appear in everyday life within the next decade. At the DTU Department of Micro- and Nanotechnology, we use micro- and nanotechnological approaches to work towards these goals. In particular, I describe some challenges of trying to tackle issues related to health care, health monitoring, drug delivery and drug development using micro- and nanotechnology. However, because all relevant biochemical processes take place in (aqueous) solutions, I limit the rest of this chapter to describing microfluidic devices: devices used to manipulate tiny amounts of liquids (in the nanoliter range) and perform chemical and biochemical processes and protocols on the molecules contained in these small volumes. Other names are lab-on-a-chip devices, micro-total analysis systems or biomicro-electromechanical systems (Fig 22.1).

Microfluidic devices allow the manipulation of tiny amounts of liquids within the confined geometry of a microchannel network in a very controlled manner. This technology enables chemical and biochemical processes known from chemical, biochemical, medical and forensic laboratories to be implemented on a device the size of a common microscope slide or CD, thus allowing work with minute sample volumes while saving chemicals, energy and time and reducing the waste produced. However, more importantly, entirely new ways of performing chemical processes become available because liquids confined in very small geometries have unique physical behavior (Box 22.1).
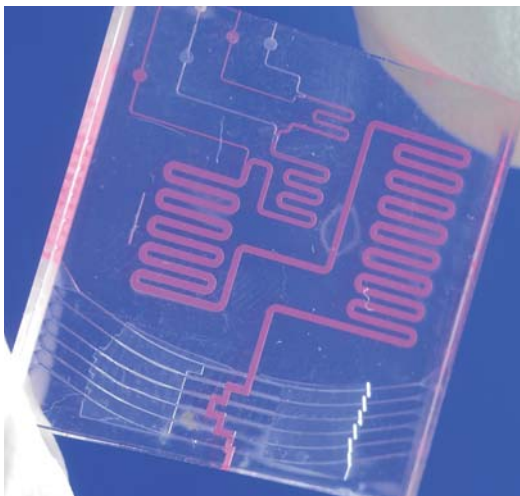
Fig. 22.1. A microfluidic device can be fabricated from different materials. Here, a sandwich structure of polymethyl methacrylate (PMMA) and the patternable photoresist Su-8 are used, and the features were realized using photolithography. Several chemicals can be mixed together sequentially to facilitate a chemical reaction that produces a colored product if the analyte of interest is present. Towards the end of the channel (lower part of the chip), optical cuvettes are used to probe the solution with external light and measure the light absorbance, which is proportional to the concentration of the analyte to be measured. It is envisioned that microfluidic devices will be able to directly communicate with personal digital assistants to exchange data or update working protocols.

Developing microfluidic devices for applications in medical diagnosis and therapeutics is a very interdisciplinary endeavor. Success requires integrating input and contributions from the engineering disciplines (fluid mechanics, electrics, optics, materials and fabrication) and from chemistry, physics, biology and medicine. The DTU Department of Micro- and Nanotechnology has established fruitful collaboration between focused specialists and big-picture creative generalists. Similarly, hiring highly skilled experimentalists as well as theoreticians with a strong foundation in the underlying principles and intuition for new phenomena that have not yet been exploited is essential. Great leaps in scientific and technological development will only occur in the interaction between theory and experiment.

Possibilities and challenges

There are many ways to use microfluidic devices for health care applications. Many efforts so far have been in diagnostics, in which microfluidic devices can work with small samples while providing answers more rapidly with less reagent. Thus, immediate answers and reactions to these answers in terms of treatment become possible either in the physician's office or at bedside for hospitalized patients. Further, screening much of the population at regular intervals becomes feasible, both because procedures are less invasive and labor-intensive and because the costs for the public health care

system are drastically reduced. A second important field for using microfluidic systems within health-related issues is drug delivery: substantially improving the administration of drugs to specific loci in the body where their action is required without losing functional material on the way and having to overdose to account for this. This requires working with nanocontainers or nanocarriers that have biochemical identification "kits" incorporated to allow them to locate the correct cell or tissue type. In addition, external triggers (such as electromagnetic pulses) or triggers based on recognizing a particular chemical environment might be used to initiate release of the transported cargo.

Microfluidic devices also constitute tools for basic research activities that allow better understanding of the underlying processes involved in disease development and progression and thus give medical experts pointers on how to design strategies to fight these diseases.

The following section gives examples of ongoing research at the DTU Department of Micro- and Nanotechnology using microfluidics for basic investigations and direct health-related applications. Many of these examples have been realized thanks to grants from national and international research funds. Research within micro- and nanotechnology and especially for applications in the life sciences always needs resources, including hardware, personnel and consumables. The continued output of scientific results at the highest levels largely depends on adequate funding.

Box 22.1. The basics of microfluidics

The two fundamental principles that govern microfluidics are laminar flow and diffusion. These two phenomena can be exploited favorably to realize working schemes only possible in the microdomain, but sometimes they need to be counteracted to achieve goals.

An aqueous solution confined in a small channel of a few micrometers in width and depth behaves very similarly to how honey or toothpaste "flows" in everyday life experience. This is because viscosity is the dominant force for small length scales, whereas inertial forces (which give rise to turbulent flows) are almost negligible. Thus, flows are very predictable and well behaved, and merging flows run alongside each other rather than into each other (see the merging flows at the junction in the left side of Fig. 22.2). This is laminar flow. If you imagine more than two flows merging, you can "squeeze" a certain lamina between two adjacent laminae and thus achieve liquid focusing. This can be used to control the position of cells with respect to the width of the channel to make sure they all move through a focused laser beam for detection.

In the case depicted in Fig. 22.2, two streams are laminated: one contains a colored dye and the other only contains pure buffer solution. The two streams only mix because of chemical diffusion across the interface between the two laminae. No convective transport happens in this case across the interface, only diffusive transport.

However, because of the small dimensions, diffusive transport is a much more efficient means of transport than on everyday scales, in which sugar molecules dissolving from a cube of sugar at the bottom of a cup of coffee take a very long time to reach the upper layers of coffee by diffusion only – instead we stir with a spoon to induce convective transport. On the microscale, a "spoon" is typically not needed. Decreasing the diffusion distance (by reducing the channel dimension) by a factor of 10 reduces the time a molecule takes to diffuse across this distance 100-fold! Thus, choosing appropriate geometries and operating conditions enables the mass transport of molecules between adjacent fluidic laminae to be controlled very precisely. This is exploited in many fluidic designs, such as selectively filtering out molecules with a higher diffusion coefficient from a solution containing impurities of larger molecules with a lower diffusion coefficient.

In Fig. 22.2, the situation assumes complete mixing at the end of the channel. However, this is only achieved if the length and width of the channel and the flow velocity are carefully chosen and matched to the diffusion coefficient of the molecules to be mixed.
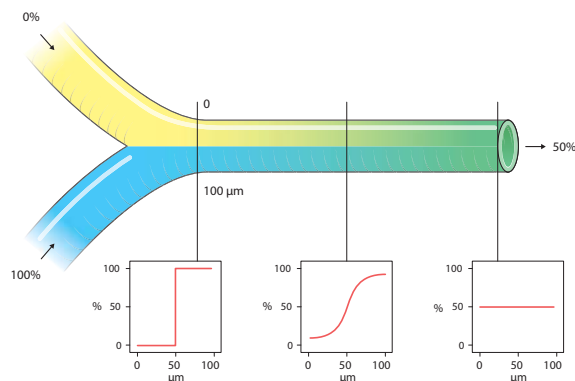

Fig. 22.2. Two microchannels are merged into one. However, because of laminar flow, the two streams flow side by side undisturbed, much like the red and white stripes emerging from the nozzle of a toothpaste tube, even though these are aqueous solutions in the microchannels. Mass transport transverse to the flow direction occurs only by diffusion. The plots below the channel depict the concentration profile across the width of the channel, as it develops from a distinct step function to a flat line, indicating complete mixing of the two solutions.
(Source: Jörg P. Kutter, DTU Nanotech)

Examples: basic investigations into health-related issues

Cellular dynamics at the nanoscale
The important length scales of microfluidic devices are of the same order of magnitude as the size of cells. Thus, these devices are intrinsically suited for working with cells, manipulating them and addressing and monitoring even individual cells. Fluidics are mainly used to move cells around, select them, bring them to specific locations on the chip and perfuse them with nutrients and/or specific chemicals (drugs, inhibitors and toxins), but the fabrication technologies used to make these chips enable the integration of other functional elements – microelectrodes and waveguides – and provide nanostructured surfaces. The EU-funded project EXCELL (Exploring Cellular Dynamics at Nanoscale) is combining these possibilities to explore the functions of cells and the molecular machinery inside cells (Fig. 22.3).
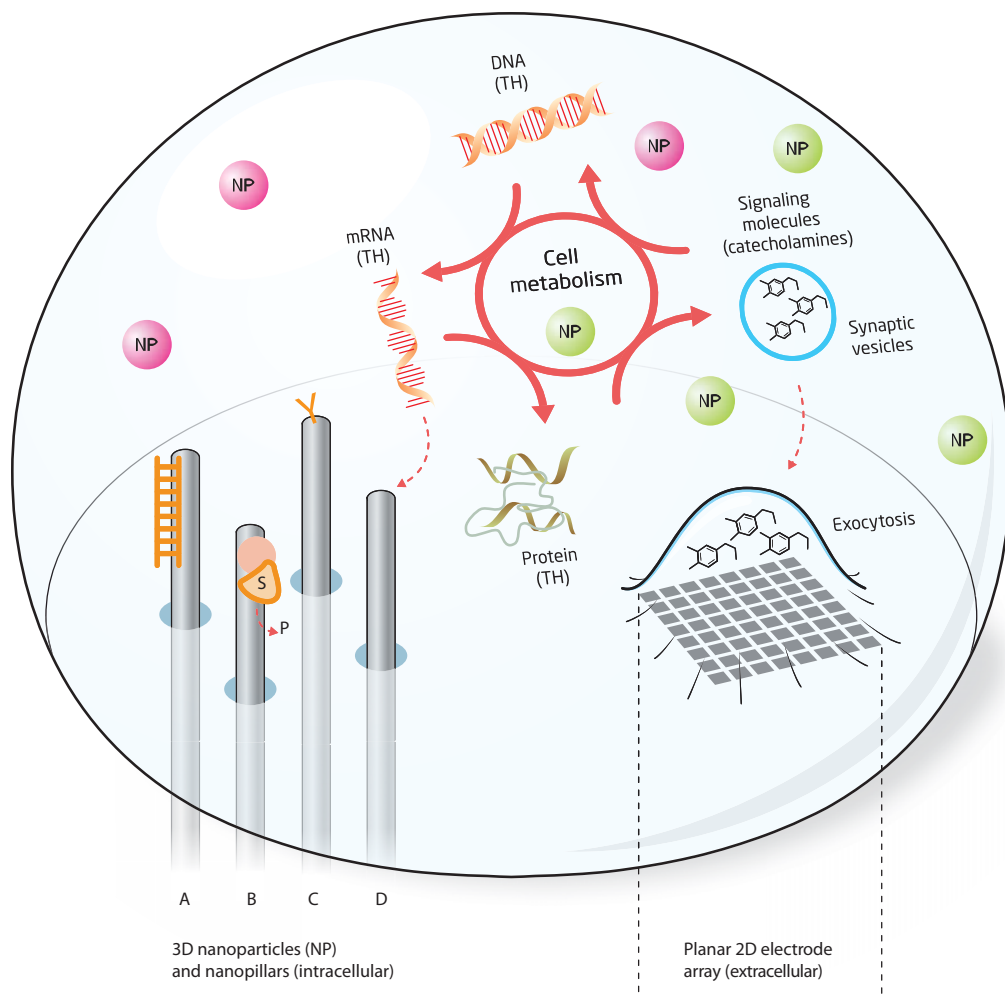
Fig. 22.3. A cell is nature's own microfluidic device, and micro- and nanotechnological tools need to be developed to explore the inner workings of cells. Here, nanoparticles and 3D nanoelectrodes probe the inside of the cell and investigate, for example, processes involved in cell metabolism. External planar electrode arrays can help to examine chemicals released by the cell (exocytosis), either spontaneous or during stimulation. NP: nanoparticles. mRNA: messenger RNA. TH: tyrosine hydroxylase. (Source: Jenny Emnéus, DTU Nanotech)

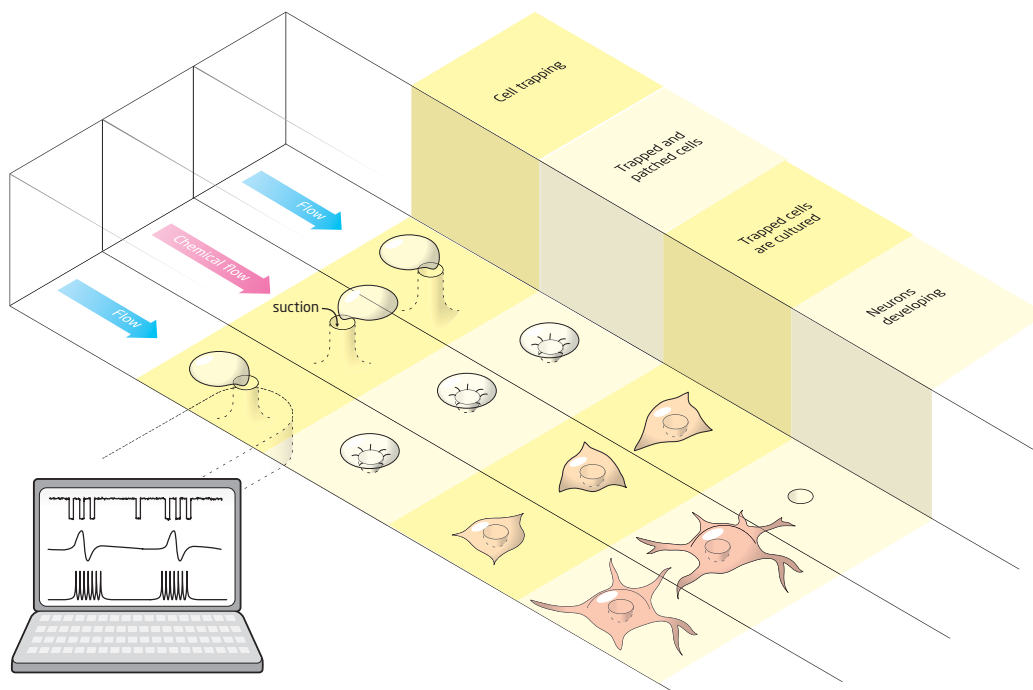Fig. 22.4. A fluidic microchip with an array of microholes on the bottom of the channel. Cells are transported from the left and trapped on the holes as suction is applied to these holes from underneath. If the suction is strong enough, the cell membrane is locally ruptured and allows electrical access to the cell interior. These trapped cells can now be exposed to different solutions containing, for example, drugs or toxic substances and the cells' reaction monitored electrically or optically. Alternatively, the trapped cells can be cultured, establishing, for example, a neuronal network, as shown on the right. Now, one specific neuron can be exposed to a particular solution using laminar flow and the cell's reaction and communication to its neighbors can then be studied. (Source: Massimo Alberti, DTU Nanotech)

In a related project, microholes provide anchor points at which cells can be trapped and then exposed to chemicals while their immediate reaction can be monitored electrically and optically. Fig. 22.4 shows a perspective view of the device and how it is designed to work. It is machined from various polymeric materials, and the central chip with an array of microholes is made from oxidized silicon. Trapping cells on tiny apertures allows the cell membrane to be ruptured locally to get electrical access to the cell interior, thus enabling electrical measurements to be used to monitor the response of the remainder of the cell to stimulating chemicals, especially the transmembrane transport of ions. This technique is known as patch clamping and is a gold standard in cell investigation. Again, the goal is to improve understanding of basic cell functions and subsequently exploit this accrued knowledge for fighting diseases or developing new drugs.

Cell culturing and programming
Traditional formats, such as cell-culturing flasks, can certainly provide the necessary environment for cells to propagate. However, performing this in traditional formats screening a range of parameters for optimum cell growth conditions is cumbersome, labor-intensive and resource-intensive. Microfluidic handling (laminar flow and fast mixing) can be used to easily prepare and supply many different conditions for cell culturing to an array of cell-culturing chambers, thus allowing parallel, simultaneous screening instead of serial screening. In devices such as that depicted in Fig. 22.5, more realistic *in vivo*–like conditions can be established, allowing the investigation of cells in an almost natural environment.
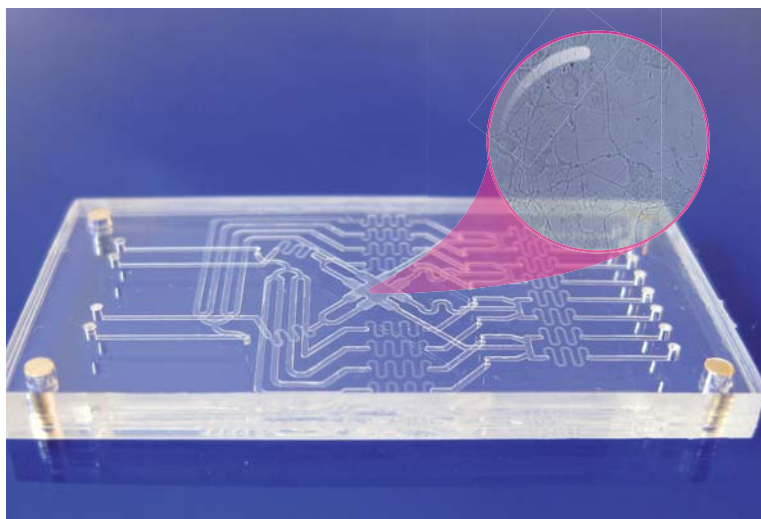
Fig. 22.5. A device made in polymethyl methacrylate using micromilling featuring a number of microchannels and other fluidic elements to provide defined and stable chemical conditions to the cells being cultured in the center of the device. The insert shows cells grown in such a device for several days. Conditions can quickly be changed to create chemical stimuli, and the cells' responses can be monitored optically through the transparent material.

Understanding the relationship between protein structure and function

Deeper insight into the relationship between a protein's structure and function is of great interest for protein-related research within pharmaceuticals, drug development and disease progression. One good method of investigating structure and structural changes of proteins in solutions is small-angle X-ray scattering, which requires an X-ray source to illuminate the protein solution. Scattering the X-ray photons off the proteins and protein complexes provides insight into their three-dimensional shape. Although the resolution is not as high as X-ray crystallography, small-angle X-ray scattering provides information on conformational changes in proteins resulting from changed solution conditions. However, changing these conditions is still time- and sample-consuming and labor-intensive, and microfluidic devices can provide options not previously available to increase sample throughput and automation at the often very expensive beamlines where beam time is very limited.

The DTU Department of Micro- and Nanotechnology is involved in the BioXTAS project, which is developing a microfluidic front-end that allows rapid and automated sample preparation before measuring small-angle X-ray scattering, with drastically reduced consumption of precious protein and a measurement volume of only 600 nl (see photo introducing this chapter). Setting up feedback between the data analysis and the sample

preparation allows automated search in the buffer parameter space, thus minimizing the number of experiments needed and, accordingly, protein consumed. The results from such experimental campaigns are useful for developing new drugs, understanding membrane protein function, elucidating fibrillation pathways in such diseases as Alzheimer's disease and optimizing formulation conditions for insulin.

Examples: applying microfluidic devices to health-related issues

Cancer diagnosis

Many diseases (especially cancer) can be traced back to single-point variations in the DNA base sequence: point mutations and single-nucleotide polymorphisms (SNPs). Rapidly and easily detecting these single-point variations will considerably improve the potential for diagnosing cancer and thus early and successful treatment options. Analyzing point mutations and single-nucleotide polymorphisms on a microfluidic device requires integrating several steps: DNA purification, DNA amplification by polymerase chain reaction, enzymatic clean-up, mini-sequencing, hybridization and detection. Microfluidics and integrated lab-on-a-chip devices enable all functions to be combined on a single device, using small amounts of sample, while providing answers within minutes instead of
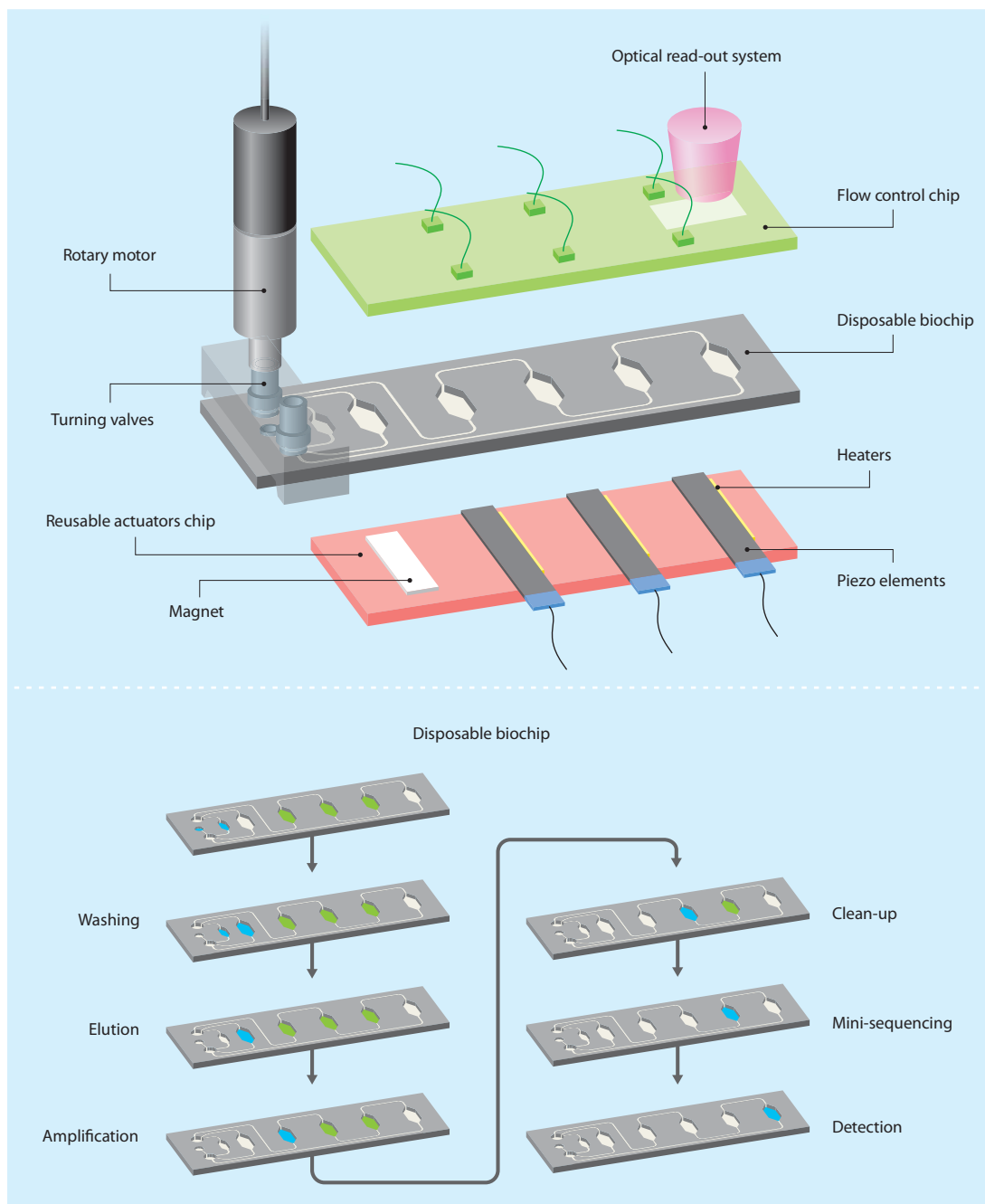
Fig. 22.6. A highly integrated lab-on-a-chip device containing a disposable microfluidic chip and several reusable functional element layers. Upper panel. The disposable central chip, where the chemical and biochemical processes occur, is sandwiched by a fluidic control unit on top and an actuator unit from below. Here, functional elements for controlling the flows, setting and controlling temperatures, manipulating magnetic particles and optical monitoring and read-out are included. Lower panel. The various process steps necessary for analyzing single-nucleotide poly-morphisms are executed sequentially by moving the solution to different chambers on the disposable chip until the final detection step. (Source: Monica Brivio, DTU Nanotech)

several hours to days with more traditional approaches. The EU-funded project SMART-BioMEMS (Development of an Integrated MEMS-based DNA Analysis Chip with Active Flow Control Components), in which the DTU Department of Micro- and Nanotechnology is a partner, attempted to realize such devices (Fig. 22.6). The central microfluidic element is made from polymer to be disposable, and a pneumatic manifold, turning mechanisms for microvalves, heating elements and magnetic elements are integrated in the reusable set-up surrounding the polymer chip.

Early diagnosis of neurodegenerative diseases
Neurodegenerative diseases (such as Alzheimer's and Parkinson's diseases) are becoming a serious burden for the health systems in many high-income countries where the population enjoys longer and longer lifespans. Once one of these diseases becomes obvious, with clear changes in normal functioning and behavior, treatment is no longer possible. Screening methods therefore urgently need to be developed that enable the entire older population to be regularly examined for signs indicating onset. Treatment can only be effective if disease onset is discovered early on. Diagnosis requires determining the occurrence of certain proteins, plaques and fibrils (which are accepted biomarkers for these diseases) by taking cerebrospinal fluid samples. This is painful, risky, rather expensive and not suitable for routine screening. In the EU-funded project NeuroTAS, the DTU Department of Micro- and Nanotechnology is involved in developing a microfluidics device for highly sensitive detection of certain protein biomarkers for neurodegenerative diseases in (easily available) blood samples. Microfluidic techniques involve immunocapturing and enriching the biomarker proteins using nanoparticles coated with appropriate antibodies and exploiting the short diffusion distances in microfluidic channels for enhanced mass transport. The microchip system includes other functionalities, such as microelectrophoresis, immune-based detection, electrochemical detection and fluorescence and absorbance detection to allow different, complementary analysis options. Physicians are planning to evaluate the systems developed with blood samples from real patients previously examined with conventional methods. Chip systems designed to work with patient samples cannot be reused. Consequently, cost considerations come into play in designing and producing these systems. The disposable part of the system should therefore contain as few expensive materials as possible and is usually made from polymer materials. However, the biocompatibility of polymer materials must be assessed before starting large-scale production.

Diagnosing HIV infection
Diagnosing HIV infection is a challenge, especially in low-income countries where sophisticated equipment is not readily available, especially not in remote locations. Nevertheless, inexpensive, reliable and easy-to-use diagnostic equipment is crucial to help to control the spread of the infection and to get health care to people in need. In the EU-funded project DetectHIV (Sensitive Nanoparticle Assay for the Detection of HIV), the DTU Department of Micro- and Nanotechnology is a - partner in developing such a device (Fig. 22.7). Again, a tiny blood sample is sufficient and, again, antibody-coated magnetic nanoparticles are used inside microchannels to catch a specific protein indicating the presence of the virus. The assay is set up such that a positive event will have two nanoparticles bound to the virus, thus forming a doublet. After the incubation, all particles (doublets and singlets) are then pushed onwards in the microchannel and additional flows are brought in to provide hydrodynamic three-dimensional focusing. This positions all particles in a very defined horizontal and vertical position in the channel cross-section and ensures that all pass through the light beam emanating from an integrated waveguide at the same velocity. An integrated waveguide collects the scattered light from these particles and passes them to a detector. Doublets produce higher scattered signal intensity than singlets and can thus be distinguished. Here, again, the central part of the system, the microfluidic chip, is made from inexpensive and disposable polymer materials, whereas the more expensive parts are collected in a housing, which can accept a new disposable chip for the next analysis or diagnosis.
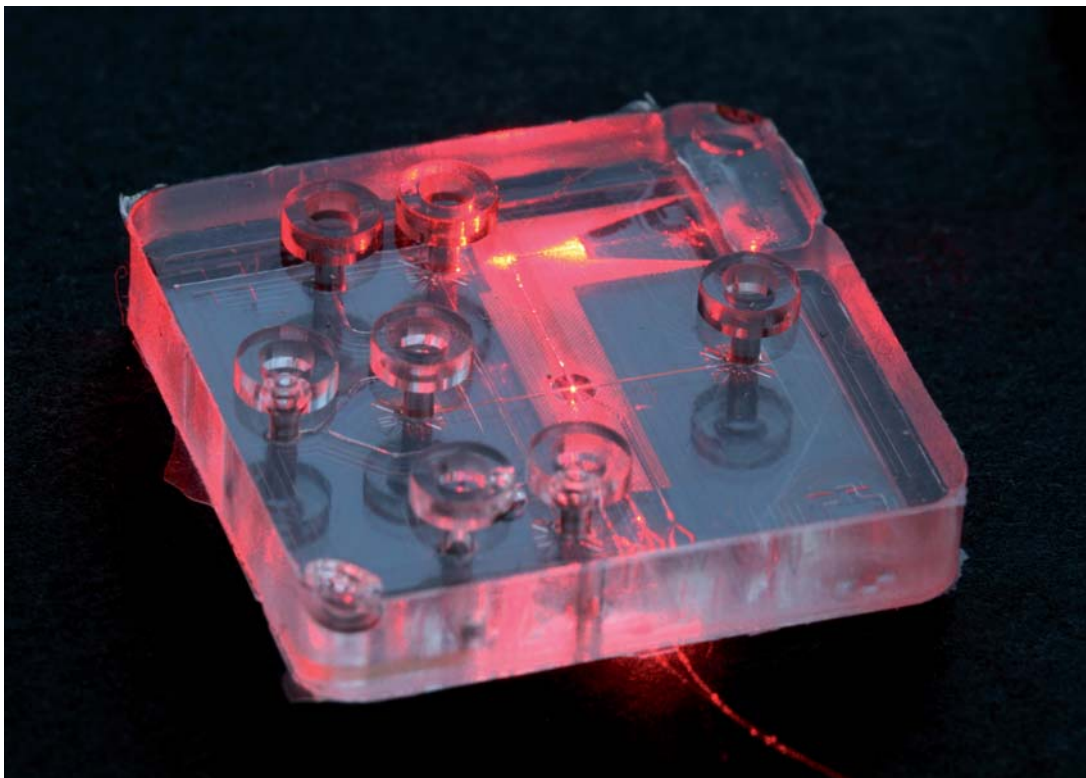
Fig. 22.7. An all-polymer microfluidic device intended for distinguishing singlets and doublets of micron-sized particles by their light-scattering signature. To facilitate this, the particles in the microchannel need to be hydrodynamically focused using laminar flow to "force" all particles through a defined detection window, where waveguides bring in laser light and pick up the light scattered off of the passing particles. In the picture, the main channels run from left to right, whereas the central measurement region is in the center of the chip. Laser light is coupled into the waveguides via a lens structure integrated in the chip (upper right corner). Scattered light is picked up by waveguides and then coupled into an optical fiber (lower edge of the chip), sending the light to a detector.

Outlook

Micro- and nanotechnology will become major enabling factors in improved health monitoring and health care in the future. This is true for high-income countries, where the diseases of civilization can be detected earlier and much more easily (in particular, diseases that do not have a major impact until old age) and new therapeutic options can be provided for the general public. Both early and rapid diagnosis and advanced treatment options will significantly reduce the health care costs to society, including treatment, care, pharmaceuticals, rehabilitation and follow-up care, thus alleviating one of the major cost factors in the national budget of many high-income countries. Similarly, the devices developed using micro- and nanotechnology will help to control the onset, outbreak and spread of diseases in the low- and middle-income countries, thus drastically improving the quality of life. Finally, systems like the microfluidic devices described above are already planned to be used for other important health-related tasks. These include food safety and food quality control, environmental monitoring, water management, indoor climate and even detecting airborne pathogens or explosives. Miniaturized devices with functionalities far beyond those of a simple sensor will become ubiquitous in everyday life.

The author

Jörg P. Kutter:
MSc, 1991 and PhD, 1995, University of Ulm, Germany. Postdoctoral Fellow, Oak Ridge National Laboratory, 1996–1998. Assistant Professor (1998–2000), Associate Professor (2000–2006) and Professor with Special Responsibilities (2006–), Department of Micro- and Nanotechnology, DTU. Research interests: lab-on-a-chip systems, microfluidics, application of these technologies in the life sciences.

More to explore

Geschke O, Klank H, Telleman P, eds. *Microsystem engineering of lab-on-a-chip devices.* 2nd edn. Weinheim, Wiley-VCH, 2008.

Bruus H. *Theoretical microfluidics.* Oxford, Oxford University Press, 2008.

Kutter JP, Fintschenko Y, eds. *Separation methods in microanalytical systems.* Boca Raton, FL, Taylor & Francis, 2006.

Hardt S, Schönfeld F, eds. *Microfluidic technologies for miniaturized analysis systems.* New York, Springer, 2007.

Schattner E. A chip against cancer: microfluidics spots circulating tumor cells. *Scientific American* 2009 (www.scientificamerican.com/article.cfm?id=a-chip-against-cancer).
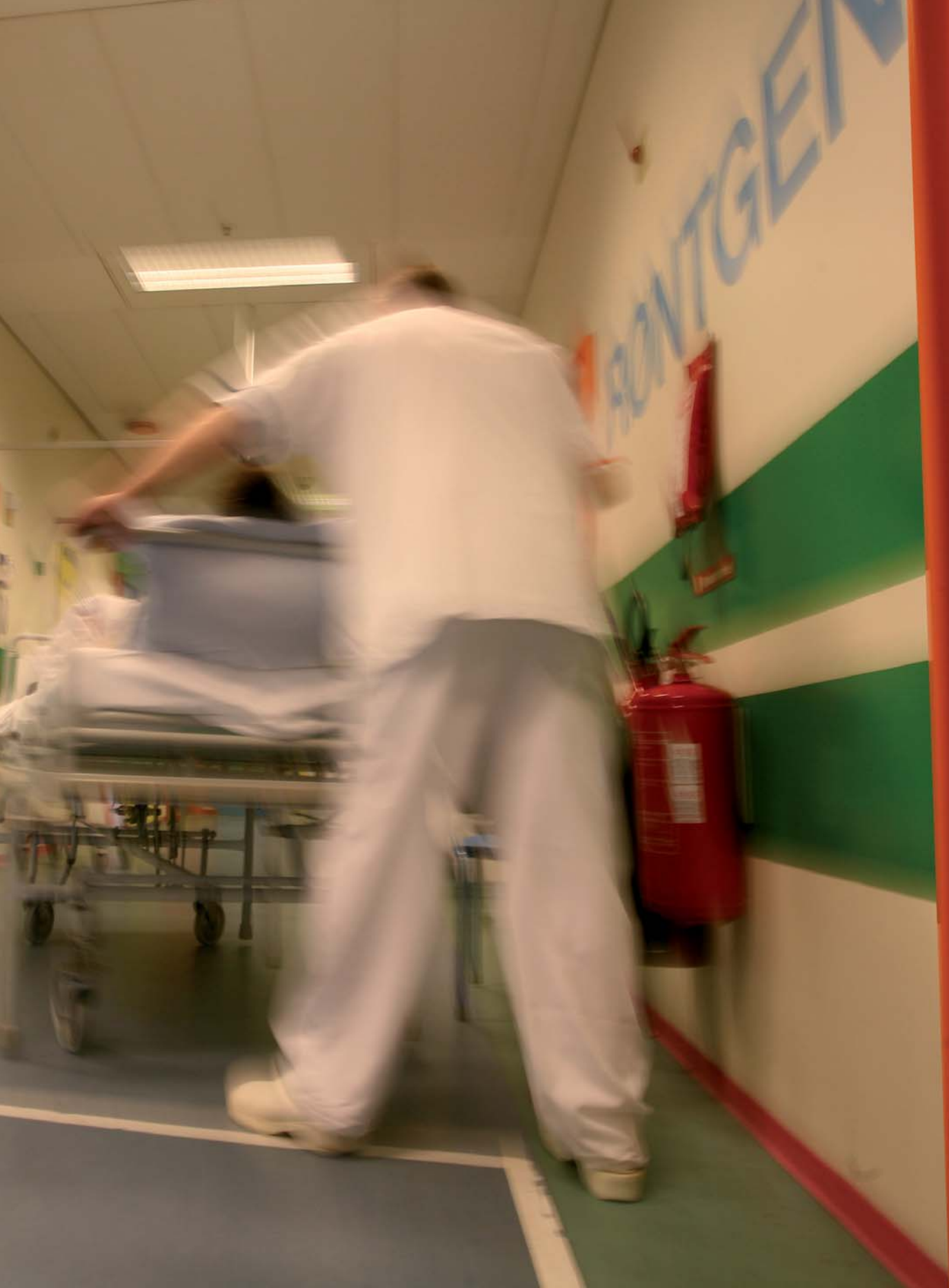
An engineering approach to
health care

Peter Jacobsen, Martin Grunow
& Henning Boje Andersen

Health care systems around the world are facing major challenges. In Europe, the proportion of people older than 65 years is expected to almost double by 2050. Chronic diseases are on the rise. One of the greatest problems already today is the lack of human resources in health care. Many European countries have too few physicians and nurses to tend to the aging population. New medical and information technologies for health care are evolving rapidly, and well-informed patients expect to receive such treatments. However, the costs for introducing and using the latest medical technologies are exploding while health care budgets are not likely expand significantly.

Calls for greater efficiency in health care have therefore become stronger, while compromising on quality is not an option. In fact, within the past decade, there has been a marked rise around the world in the attention to the quality of care and thus to the fact that avoidable human and system errors in health care make patients ill or injured every day.

Engineering methods and experiences are increasingly contributing to solving some considerable challenges facing health care in high-income countries. Several great problems in health care may only be solved by advanced technical methods typically mastered by engineers and rarely by health care professionals. Others may be solved only by applying socio-technical engineering approaches capable of analyzing and modeling the complex interplay among technical, organizational and human factors.

Engineering methods have been used in industrial research for years and helped in optimizing efficiency and quality. The challenge is now to integrate this experience with health care. This chapter describes engineering methods – discrete event simulation and causal safety models – that can contribute significantly to solving many problems in health care. The methods and techniques described do not inherently solve anything – except perhaps some equations – but contribute substantially to the challenge of optimizing the flow of patients and resources needed to provide care.

Discrete event simulation is primarily used in health care for optimizing patient flow, work flow and thus efficiency. However, this approach may address quality issues, including patient safety. This requires applying even wider engineering methods and techniques, considering models of medical errors that transcend the delayed delivery of services. The collaboration of research groups in the newly formed DTU Department of Management Engineering offers an opportunity for addressing health care efficiency and health care quality in a unified approach. This contribution sketches some of the previous work in the groups and the research challenges that are addressed in the future collaboration.

We first review discrete event simulation methods and some tools to model the domain and run simulations. Then we present a case study of patient flow to, within and from an emergency department. The emergency department environment receives patients who have suddenly fallen ill or become injured. The emergency department staff must therefore act from case to case, and specialists from other departments are frequently called to assess or diagnose patients. Thus, the emergency department environment has many care transitions: situations in which one health care worker transfers responsibility for a patient to another health care worker. Care transitions, however, are known to be risky: vital information may be lost or not even addressed when a busy nurse or junior physician presents a patient to a senior physician. Thus, developing and testing models for optimizing patient flow and the use of human and other resources requires including information about the risk of disrupting care in transferring responsibility for a patient between staff members. Accordingly, in the last section we describe the structure of a classification (taxonomy) of adverse events related to care transitions and outline how the using the taxonomy may generate information about care transition risks, which in turn may be used to optimize the deployment of available health care staff resources to achieve efficient and high-quality care.

Discrete event simulation

In general terms, simulation is the process of designing a model of a real or imaginary system and then conducting numerical experiments. Single modelers or groups of modelers create simulation

analysis by using a simulation tool, such as simulation software. The modelers observe the real system, define the area of concern and identify the key parameters of interest (Fig. 23.1). With these in mind, they design a computer model that captures the crucial aspects of the real system. The art is striking the right balance between the level of detail required to correctly model the system behavior and the effort required to build the model and run the simulation analysis. The model is then used to conduct numerical analysis in the form of scenarios testing the system response to changes in some of its components or inputs from the environment. The results often lead to rebuilding or refining the model.

 The term model is equivalent to a computer model in this case. An advantage of using a computer is that complex scenarios can be performed, and they may be performed in compressed time. A discrete event simulation is one in which the state of the simulation model changes at discrete points in time as triggered by events. Simulation software is usually divided into four groups:

- general purpose programming languages;
- simulation languages;
- simulators; and
- customized simulation systems.

The groups allow for different levels of modeling flexibility and different levels of ease for model building. They also represent the evolution of simulation software generations. With the help of the first group of tools, computer simulation was performed using general-purpose languages. Examples are C, C++ and Basic. These tools require great effort from the modelers, as they have to build up the simulation program themselves, programming it line by line. Despite the effort required, this approach is still used because it offers great flexibility. In the 1960s, programming expressions that were more specific were introduced in an effort to simplify the model-building process. These tools are called simulation languages and regarded as higher-level programming environments, although they still require specialized programming skills. Examples include GPSS (General Purpose Simulation System), SLAM II and SIMAN. Today's software simulation packages, called simulators, facilitate model building in specific simulation pro-
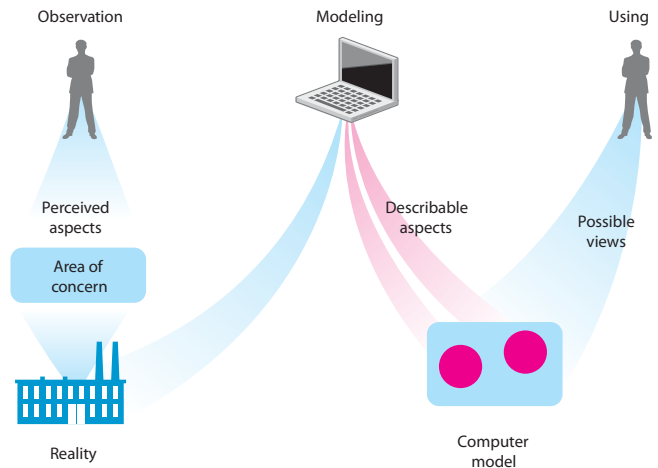

Fig. 23.1. From reality to simulation model

gramming environments. Models are built by using built-in templates. Further, the introduction of graphical modeling environments with the latest generation of simulators provides the ability to build a simulation model even more rapidly by using graphical symbols and defining the relationships between them through a pointing device and short logical statements. These templates represent predefined programming macros that, built together, form the simulation model. Examples are Pro-Model, Arena and Enterprise Dynamics. The fourth group, customized simulation systems, covers tools that are highly specialized for a specific purpose. An example is Siemens Plant Simulation. The templates provided in this software tool cover typical resources used in manufacturing environments. The use of these predefined templates allows for very efficient model building, but these simulation systems run the risk of constraining the modeler to some standard building-blocks that may not match the modeled objects. This is why the programming macros for the templates in some simulation systems are accessible and can thus be customized. Moreover, some simulation systems allow for interaction with other software packages such that, for example, programs generated by general programming languages can be included.

 Selecting a suitable simulation tool requires profound knowledge of the available software options and understanding of the requirements related to the simulation project. When the simulation software has been selected, all required input data have to be identified. With the increased use of comput-

ers in traditional simulations domains such as manufacturing, more and more data become available in databases from which they need to be extracted and analyzed. If this is not the case, data need to be measured. The observed data are matched to a suitable probability function, which is then used to generate events in the simulation analysis. Other data may also be obtained from interviews, but such information must be handled with care and validated.

Applying discrete event simulation in health care systems: analysis of an emergency department

Simulation in health care

Traditional methods for configuring production or service systems are often static. Norm times are used in calculating capacity and costs. Such an approach is not suitable in health care, which has no rigid time ranges for the duration of even standard activities such as appendectomy. Most activities in health care are related to people and thus depend strongly on the individuals involved. Here, simulation is a particularly useful tool as it can include stochastic events based on different probability distributions. However, the data required to describe the stochastic processes are often not available in the units providing the health care services. Hence, they must be deduced from other information such as corrected shift plans or collected through measurements.

Specialized simulation system packages include templates, designed mostly for industry applications such as conveyors or automatically guided vehicles. The central building blocks for health care simulation models need to be identified and made available as templates to make creating simulation models easier in this sector.

Case study in a emergency department in Greater Copenhagen

A current trend in health care is gathering activities in larger units. One such area of concentration is emergency departments. Emergency departments are difficult to schedule, as treatment cannot be planned in advance. They are also expensive to run, as personnel with different qualifications need to be available at all times. Health care authorities in Denmark and in other countries have therefore recognized that centralized emergency departments allow more resource efficiency. Hence, some departments are closed, while others are being redesigned and expanded. Moreover, using resources more efficiently requires developing new patient flows. The following case illustrates how discrete simulation can be used to analyze configuration options.

The purpose of the project was to verify the proposed concept for a new emergency department in Hillerød Hospital, a large hospital in Greater Copenhagen. The interdependence between the design options can only be modeled through complex stochastic distribution functions. Simulation was therefore chosen as the most suitable analysis tool. Together with physicians and nurses from the department, a triage point and method were defined. The reason for doing triage is to optimize resource use. Triage means that the emergency department assigns the patient to a group on entry. This group is associated with a sequence of activities, and the appropriate resources (such as personnel and equipment) are allocated.

Three patient flows emerge from the triage point. Together with a proposed layout, these data formed the basis for the simulation analysis (Fig. 23.2).

As indicated in Fig. 23.2, 60% of the patients walk in and 40% arrive by ambulance. The triage clusters the patients in three groups: white, red and blue. The dotted line indicates a patient requiring diagnostic imaging other than X-rays (mostly computed tomography) and a stippled line for patients needing X-rays. These arcs simultaneously represent transitions in which the patient is transferred to another person and a risk for misinformation and impaired health care service quality. The next section deals with this.

A patient in the red group often has a common condition, which means that one or more vital parameters are affected. Such a patient has injuries or diseases that impair their condition if they are not treated immediately. A red patient hence requires acute action. Blue patients, in contrast, can wait up to 3 hours, and the waiting time for white patients who have minor symptoms is not limited.
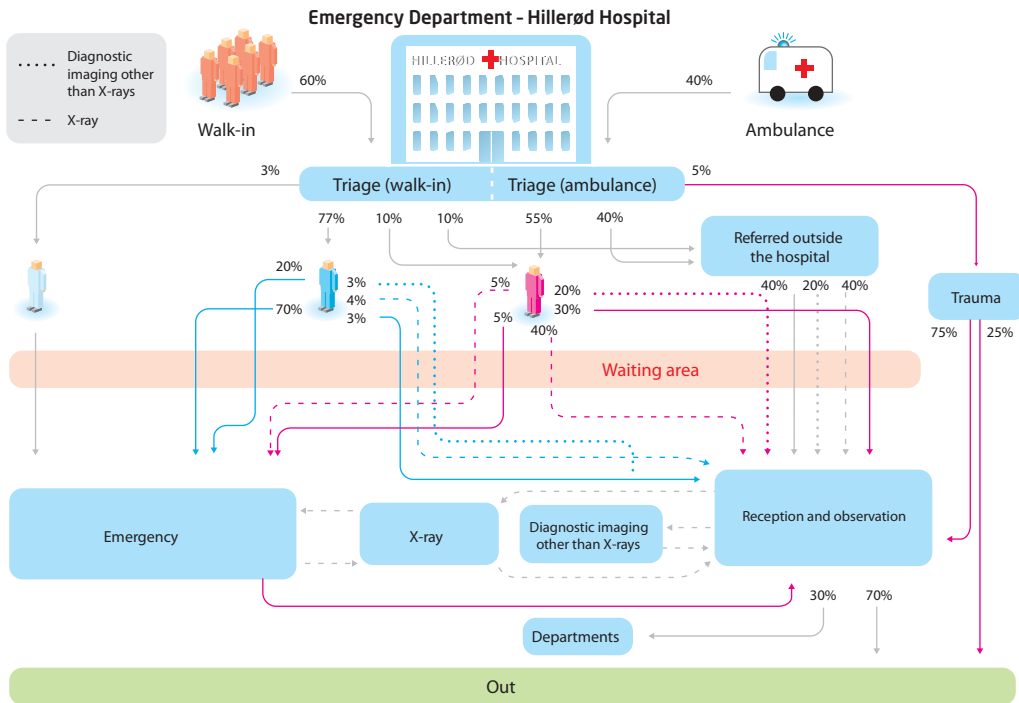
Fig. 23.2. Patient flows at the Emergency Department of Hillerød Hospital, Denmark

Numerical analysis
All input data were gathered in spreadsheets. This enabled the scope of the numerical analysis to be easily defined in collaboration with the stakeholders. Due to the concentration of activities, the numerical experiments focused on the total number of patients arriving at the emergency department.

Fig. 23.3 shows the lead time before X-rays or diagnostic imaging other than X-rays as a function of an increase in the number of patients. Some of the patient flows react more sensitively than others. As red patients have first priority, their lead time is less affected than the lead time for the blue patients.
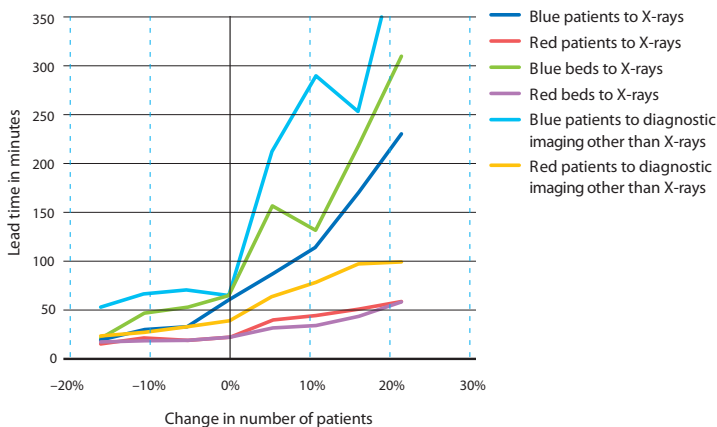

Fig. 23.3. Lead times before X-rays or diagnostic imaging other than X-rays as a function of an increase in the number of patients
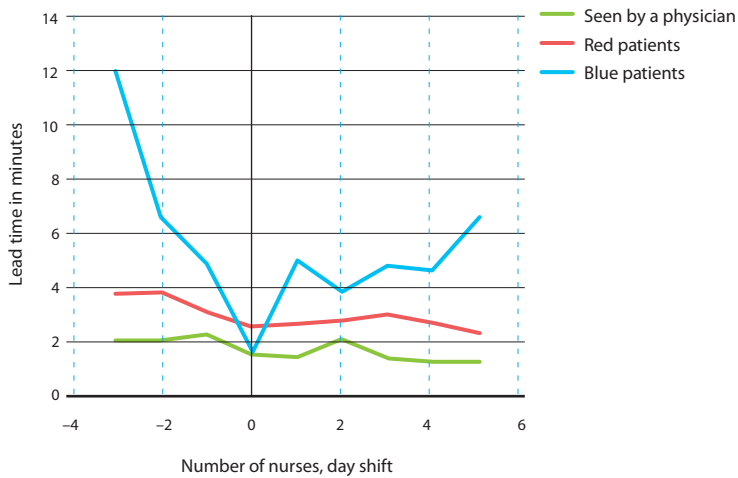
Fig. 23.4. Lead times in the waiting room as a function of number of nurses on a day shift

Further scenarios were conducted in which the patient lead time was related to the resources available. The example is shown in Fig. 23.4, which illustrates the lead time in minutes as a function of the number of nurses on the day shift. Again, as red patients have priority, the influence of staffing on these patients is minor, whereas the impact on the blue patients is severe. For the blue patients, reducing the number of nurses by two increases waiting time by 280%. Full parametric analysis was carried out including other shifts and other employee groups.

The main results of this simulation analysis were as follows.

- The lead times show no reaction to an increase of patients of at most 10% – as long as X-rays and diagnostic imaging other than X-rays are not required. But the lead times increase sharply when the requirements for X-rays and other types of diagnostic imaging follow the known distributions. Hence, X-rays and other types of diagnostic imaging are the bottlenecks in an expanded emergency department. Investment should be focused here.

- At current patient load levels, a bed is occupied for 75% of the simulation time. If the additional cleaning activities are also considered, this is a high but acceptable load. For expansion, additional investment is also required here.

- The employees spend 3–4% of their working

time walking from one workplace to another. This was perceived as a good result for the proposed layout.

- The employees show a high degree of utilization, such as 90% for the nurses. This is defined as the time used on assigned work compared with the total available time minus breaks. Additional staff is required for expansion.

- At current load levels, reducing the number of nurses under the current level mostly influences patients requiring X-rays or other diagnostic imaging. But at night, the number of nurses can be reduced without significantly increasing lead times.

- Reducing the number of physicians under the current level affects waiting times more than changing the number of nurses. Increasing the number of physicians yields minor improvements.

This study focuses on efficiency only. The analysis does not include the impact of the centralized configuration on the quality of the health care provided. Including quality requires developing novel approaches for integrating potential hazards in the simulation approaches. The following section therefore focuses on the failures of transition and how they affect the quality of care (Box 23.1).

Case 1. A male in his sixties enters the Department of Gastrointestinal Surgery in the early afternoon with signs of possible ulcer. A gastroscopy carried out in the evening shows a small non-bleeding ulcer. At 01:40, the patient has vomited and calls the nurse. The nurse observes large quantities of fresh blood and coagels in vomit and calls the duty officer immediately. The patient has low blood pressure, and the physician asks the nurse to send for four bags of blood immediately. The nurse transmits the instruction to the assistant nurse, who goes to the nurse station office and relays the request to the blood bank at 01:50.

At the blood bank, the call is received by a bioanalyst who is alone on duty. She does not get the impression that the request is acute and therefore follows normal procedures, instructing the hospital porters via the electronic patient administration system that blood bags are to be taken to the ward.

The porters are occupied with other tasks and their normal round is therefore delayed by 15 minutes. In the ward, they are busy attending to the patient, and 30 minutes elapses before they

press for the delivery from the blood bank and explain that the patient's life is at danger. The bioanalyst now calls the porter, who rushes the blood bags to the operating room to which the patient has now been transferred. The patient is operated on and survives.

Case 2. A patient has arrived at the Emergency Department and is transferred to the observation section for observation of lipothymia (gradual loss of consciousness). On arrival in the Emergency Department, the patient's electrocardiogram is recorded and several blood tests are initiated. On the following day, at the patient round after noon, the physician doing the round discovers that the electrocardiogram has never been examined and that the patient has had acute myocardial infarction. The patient is rushed to the Department of Cardiology.

These cases are drawn from the Danish Patient Safety Database established in 2004 when Denmark introduced the world's first confidential, non-punitive, mandatory and nationwide system for reporting and learning from adverse events.

Care transitions – delays and breakdowns

As mentioned above, care transitions are risky but unavoidable. Transition occurs whenever the responsibility for a patient passes from one health care professional to another. At a small scale, this happens when a physician in the emergency department asks for X-rays for a patient. The physician delegates part of the care responsibility to professionals in the Department of Radiology, or a nurse asks a porter to take the patient to the Department of Radiology. At a larger scale, it happens in a transition between two physicians or when a new shift arrives and assumes responsibility for all patients.

Delays and breakdowns in care transitions: a causal safety model

Some characteristics are common to most transitions in an emergency department. Thus, written documentation is rarely used, and the patient chart is rarely available – since it has not yet been transcribed after dictation by the physician(s). Further, transitions also vary considerably: some are face-to-face encounters, and others occur when a nurse calls a physician or a junior physician calls a more senior physician and asks for advice or calls for assistance. Here, the urgency with which the case is presented may be crucial.

During a care transition, information may be forgotten or suppressed – often due to interruptions from other workers who might be dealing

with patients with more acute needs. Similarly, worries and concerns about the patient's condition may not be presented such that the receiving physician or nurse grasps the patient's situation adequately. Finally, although trauma patients are always tended to immediately by a team with several specialists, most patients arriving in an emergency department are not usually urgent cases, although their condition is unplanned and typically unforeseen. Thus, the staff members tend to treat the most urgent cases first. But sometimes the apparent urgency is either misjudged or not conveyed adequately when a nurse or physician calls for advice or assistance. For example, in case 1 in Box 23.1, the staff in the blood bank misunderstands an urgent request for blood transfusion bags. Fig. 23.2 depicts care transition points in emergency departments, involving the transfer of responsibility of care as defined above. The transition points include situations in which responsibility for providing a specialized service is transferred to another organizational unit such as X-ray, other diagnostic imaging or a clinical laboratory.

Care transition failures happen when actions or omissions by health care staff result in actual or potential harm to the patient. Examples of failed transitions are acts of miscommunication and failure to – perhaps forgetting to – transmit, respond to or request information.

Capturing adverse events during care transitions

Several studies have sought to identify the causes and circumstances behind medical errors, or more generally, adverse events. An adverse event comprises actions or omissions by health care staff resulting in actual or potential unintended harm to a patient, including death in the most extreme cases. Industry uses causal models extensively to aid in structuring and capturing information from accidents and incidents for the purpose of identifying factors that need to be changed to improve safety. Fig. 23.5 depicts a general model of causal factors involved in adverse events, showing only the major groups of factors.

A causal model of accident and incident factors may be associated with a taxonomy (a classification system) of the unsafe acts and the underlying factors behind accidents. This section describes a taxonomy developed in a current study of adverse events focusing on care transition failures and their causes in hospital settings and including patient flow through an emergency department and wards and departments that receive patients from or perform analysis for the emergency department, such as clinical laboratories, X-ray and diagnostic imaging.

A taxonomy is a hierarchical classification system for ordering either natural or human-made objects or phenomena. Several taxonomies have been developed for industrial and transport accidents and
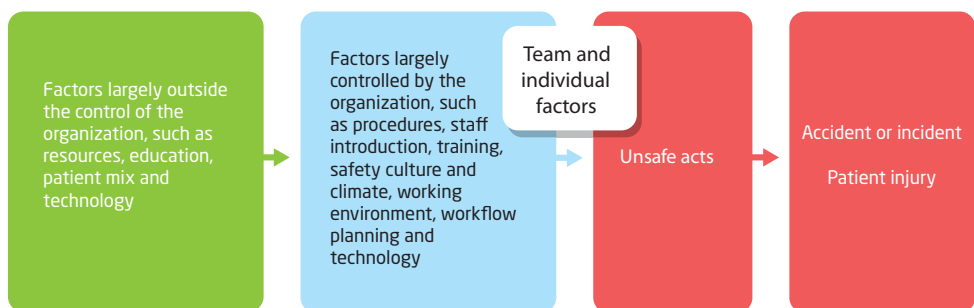

Fig. 23.5 A general model of the causal factors involved in adverse events, showing only the major groups of factors

incidents, and several taxonomic systems have been developed for adverse events in hospitals in recent years. The taxonomy described here targets transition failures and not all types of adverse health care events. The taxonomy was developed to structure and classify narratives elicited from health care staff during interviews, reports submitted by health care staff to the Danish Patient Safety Database and accident analysis reports involving transition failures.

The taxonomy distinguishes between active failures (encounters and contacts between health care professionals) and background factors causally linked to the active failures or to the specific case of (potential) patient harm. Active failures are acts of failed communication among health care staff or the failure of an attempt to transfer responsibility for patient care. To be useful, the categories of failed communication must include not only obvious omissions and acts of mishearing utterances but also omitting to ask relevant questions or to address aspects about the patient that, according to accepted standards of care, should have been explored. An example is explicitly communicating the degree of urgency of a request (case 1 in Box 23.1). Besides communication failure, failure to actually transfer responsibility arises not because the communication is unclear or misinterpreted but typically because the recipient is too busy with other patients.

Background factors comprise organizational and individual factors to the extent they can be distinguished. Organizational factors are largely within the control of the organizational unit – the department or the hospital (Fig. 23.5). For instance, clinical procedures or guidelines on introducing new personnel to their tasks. In contrast, cuts in overall budgets may be an underlying cause of busy and crowded wards that, in turn, may make a given adverse event more likely. But since the overall budget is largely outside the control of the individual department or hospital, we do not categorize this type of factor as organizational.

If a care transition fails because a physician or nurse misjudges the situation, a given adverse event report may not contain sufficient information to decide whether this is caused by "organizational failure" or "individual failure". In general, if a department fails to deliver a care service that it is supposed to be able to deliver, the failure is normally categorized as organizational. For instance, if staff members are not properly introduced to their tasks (such as not being adequately trained in using information and communication technology) or if junior staff members have to take on tasks that require competencies and experience they lack, the failure is organizational, according to the definitions of the taxonomy categories. If experienced staff members fail to apply the level of competence that may be expected from them (in relation to their education, training and job position), the failure is classified as individual.

Prospects and conclusions

Although discrete event simulation models have increasingly been used to optimize patient flow and the use of resources in health care, the potential of such modeling has not been extended to capture transition failures. Current models are focusing on optimizing efficiency and thus capture temporal and resource dependence very well. However, quality issues are equally important in modern health care settings, so while current discrete event simulation models are capable of generating a wide range of results to improve patient flow running through more or less optimized transition points, the models have not been designed to capture transition failures compromising patient care because of incorrect or inadequate communication.

In the last two sections, we described a framework for structuring and capturing information about care transitions. The framework was originally and primarily used to apply results to the design of work flow and information flow mechanisms as well as team training to improve coordination and reduce the risk of transition failures. However, conventional simulation may be expanded to include risk information and thus results about the frequency of types of failure and how they are distributed among causal factors. Discrete event simulation is particularly suited, because the adverse events, including the transitions, can be added to efficiency-focused models. Simulation can then be used to identify the root causes of quality problems. Investment and personnel education measures can be then focused to obtain the largest gains in quality and efficiency.

The prospect of integrating adverse events and

their implications into conventional discrete simulation represents not current or completed work but a research agenda. We regard extending the discrete event simulation approach to include risk data as a promising step towards integrating the two major dimensions in optimizing health care management – efficiency and quality.

More to explore

Akkerman R, Knip M. Reallocation of beds to reduce waiting time for cardiac surgery. *Health Care Management Science* 2004: 7: 119–126.

Andersen HB, Hermann N, Madsen MD. Hospital staff attitudes to models of reporting adverse events: implications for legislation. In: Spitzer C, Schmocker U, Dang VN, eds. *Proceedings of International Conference on Probabilistic Safety Assessment and Management (PSAM 7 – ESREL'04), Berlin, 2004.* Berlin, Springer, 2004: 2720–2725.

Grunow M, Günther H-O, Yang G. Development of a decision support model for scheduling clinical studies and assigning medical personnel. *Health Care Management Science* 2004: 7: 305–317.

Itoh K, Omata N, Andersen HB. Human error taxonomy for analysing health care incident reports: assessing reporting culture and its effects on safety performance. *Journal of Risk Research* (in press).

Jun JB, Jacobson SH, Swisher JR. Application of discrete-event simulation in health care clinics: a survey. *Journal of the Operational Research Society* 1999: 50: 109–123.

Lowery JC. Getting started in simulation in healthcare. In: *Proceedings of the 1998 Winter Simulation Conference.* Paris, IEEE, 31–35.

Rasmussen J. *Information processing and human–machine interaction.* Amsterdam, Elsevier, 1986.

Reason JT. *Managing the risks of organizational accidents.* Aldershot, Ashgate, 1997.

Standridge CR. A tutorial on simulation in health care: applications and issues. In: *Proceedings of the 1999 Winter Simulation Conference.* Paris, IEEE, 49–55.

Saltman RB. Decentralization, re-centralization and future European health policy. *European Journal of Public Health* 2008: 18: 104–106.

VanBerkel PT, Blake JT. A comprehensive simulation for wait time reduction and capacity planning applied in general surgery. *Health Care Management Science* 2007: 10: 373–385.

The authors

Peter Jacobsen:
MSc and PhD (1983), DTU. Research experience, Brigham Young University, 1984. Consultant, 1985–1995. Faculty member, DTU, 1995–. Currently Associate Professor, DTU Department of Management Engineering. Research: planning, optimization and control of production systems, design and control of health care systems and simulation modeling.

Martin Grunow:
MSc in Industrial Engineering, PhD and Dr. habil., Technical University of Berlin. Researcher, Degussa AG, Hanau, Germany. Assistant and Associate Professor, Technical University of Berlin. Guest professor, University of Vienna. Currently Professor in Operations Management and head, Operations Management section, DTU Department of Management Engineering, 2006–. Research: manufacturing and service operations, supply chain management, IT-based decision support.

Henning Boje Andersen:
Mag.art., University of Copenhagen (philosophy; dissertation: epistemic logic, gold medal 1976). Oxford University Danish scholarship. External Lecturer, University of Copenhagen, 1979–1984. Researcher and subsequently Senior Researcher, DTU Risø National Laboratory for Sustainable Energy, 1984–2009. Currently Senior Researcher, DTU Department of Management Engineering. Research: human factors in safety critical domains; risk and accident analysis; design and evaluation of human–machine interaction systems; and organizational learning from accidents and incidents.