Springer Series Complexity"

EMERGENCE, ANALYSIS AND OPTIMIZATION OF STRUCTURES

Concepts and Strategies across Disciplines

K. Lucas, P. Roosen (Eds.)

Es werden folgende Farbcodierungen verwendet, um jeweils neue Beiträge in übergreifenden Kapiteln einzelnen Autoren zuzuordnen. Hierbei werden jedoch üblicherweise die Passagen nicht wortgenau eingefärbt, sondern nur im Rahmen gewisser Formulierungsverschleifungen:

- Grötschel
- Hartmann
- Költzsch
- Strübing
- Roosen
- Lucas
- Mehnen
- Preuß
- Schuster

Nach Akzeptanz der entsprechenden Passagen wir die farbliche Kennzeichnung wieder entfernt.

Inhaltsverzeichnis

Preface

(Lucas/Roosen, 3 pages)

Discussion of Origin of the Book: The BBAW study group. Short historic reflection what typical questions arise in transdisciplinary discussions, and how different opinions of their solutions are. Is there such a thing as an optimum in more than purely technical cases?

1

Introduction

We live in a world of structures. In the form of plants and animals, as well as of landscapes and clouds we are surrounded by the beautiful structures of nature (Fig. 1). Human civilization has added artifacts of particular beauty and geometry, from the

Abb. 1. Structures in nature: Plants and animals

egyptian pyramids to the structures of modern urban settlements (Fig. 2). Modern industrial development has brought to appearance artificial structures in the form of engines and products of various types, such as airplanes, automobiles, power stations,

Abb. 2. Artificial structures: Complex Buildings

chemical production sites and many more. (Figures) We start learning to design molecular structures, such as proteines or nucleid acids, for medical applications by experimental methods. We organize our social and economical life in societies, (Figures) companies and logistical decisions. We invent medical treatment procedures, mathematical algorithms and theories explaining the phenomena of the world and we create structures of art in music, poetry and paintry (Figures). All these structures, be they of a material or a immaterial nature, are in a process of permanent development, accompanied by decay and resurrection, and, in particular, characterized by a trend to progress and complexity.

What do we know about the principles of the processes creating structures? Since Darwin we know that the structures of the living nature emerge from ongoing stochastic mutation and selection, leading to an increasing survival and reproduction. The same principles can frequently also be observed to operate in processes leading to our artificial structures, no matter at which area we look, i.e. technical structures, structures in the social sciences as well as structures of art and intellect. Yet, at the same time we note that especially in the technical domain highly functional structures are frequently brought to appearance in one targeted throw. An example is the conventional design of a modern building or a bridge as a monumental architecture using appropriate structural analysis computer codes, i.e. without the necessity of any evolutive trial and error approach. Nevertheless, in modern approaches the numeric structural optimization is increasingly based on evolutionary algorithms which compute properties of competing structural candidates to figure out the “best solution´´. Finally, every once in a while a revolutionary invention or intellectual concept, apparently without any evolutionary roots, strikes and alters the world.

Given the fact that structure generating processes are studied in almost any scientific field the question arises to what extent they can be reduced to common foundations. In particular, it appears fruitful to investigate whether progress can be made in analyzing and, where applicable, optimizing them by transferring the knowledge accumulated in the various disciplines across the frontiers.This book does not contain final answers. In chapter 3 a selection of examples illustrates the issues involved. Problems and solution procedures are presented that are typical for the disciplines contributing to this book, ranging from engineering, over natural sciences up to the social and economic sciences. They are not written as in-depth treatments for the specialist. On the contrary, they are presented with the aim to explain area-specific questions and methods of scientific analysis in a way digestible for non-specialists.

As can be expected, there are differences between the fields with respect to approach, analysis, and solution strategies, but there is much more in common than scientists focussing on disciplinary work normally anticipate. Therefore in Chapter 2 some transdisciplinary fundamentals are discussed. We start from a first look at the structures and structure building processes in simple systems, then proceed to the problem of modeling reality. Next, the question of optimality is discussed, and general optimization methodologies are scrutinized as depending on the various types of modeling.

Finally, in Chapter 4, we summarize the results and further challenges of this interdisciplinary approach to understanding structures and illustrate the perspectives of a new research field.

2

Transdisciplinary Fundamentals

The analysis of structures and structure generating processes in the various disciplines rests on the basis of some transdisciplinary fundamentals, more or less common to all fields of application. These fundamentals will be discussed in the present chapter. First we define our understanding of structures and their properties as treated in this book. After that we discuss the various aspects of modeling real world structures in terms of artificial mental concepts. Here we reflect the necessity of introducing simplifications to the real world structures, in order to make them accessible to analysis.

On the basis of models further progress in analyzing structures and structure generating processes can be made. A deep rooting question concerning structure generating processes is that of the circumstances under which they develop towards a predetermined goal. In the technical as well as the economical sciences we frequently analyze structures with the aim of optimizing them. When a single goal can indeed by specified we can apply the algorithmic methods that offer themselves for this process. However, this is too simplified a view to be realistic. A closer look usually reveals the existence of not one but of concurring and potentially conflicting goals. Furtheron, we are faced with various uncertainties in modeling and model evaluation. Finally, since the structures to be discussed here are created to serve the needs of people, the assessment of quality measures is influenced by subjective apprehension. In elucidating all these realistic facts, we find that in the general view there seems to be no predetermined goal, not even in technology and economy, in particular in a long term perspective. On the other side there are evidently predeterminable goals in a partial sense, e.g. a restricted short time view, in a restricted area of consideration, or in a restricted system of values. This then seems to close the gap to the biological subsystems, which frequently are partially optimized to a remarkable extent, although the full organism is not.

We encounter structures in many different appearances. They may be material ones, like ordered matter, or immaterial ones, like social interrelations or concepts of thought. Scientific problem solving reveals structures that may be characterized as distillates arising from data analysis, general problem knowledge, experience, experiments, contemplation, and in many other appearances. Which structures are recognized and considered important for the specific niche of the real world dealt with and which structures finally emerge as key objects of study strongly depends, though, on the scientific and cultural background of the investigators and on the viewpoints chosen.

Material objects, i.e. structures, as well as concepts and ideas, are interlinked by more or less complex networks. We refer to this phenomenon of a structure being part of a network generally as its interaction with an environment. To illustrate this view, let us consider a simple object, e.g. a window in a house. A window has a certain internal structure as it consists of translucent elements, a frame, an operating handle, and some inner mechanics. On the other hand the window is embedded in the house and, together with the house, in the neighboring environment that may define varying interactions with it. Taking account of these interactions other properties of the window become important, like its direction with respect by sunshine, its relative position in the room, its size, its operability etc. . Last but not least, in a sociological or psychological context, the window in the house may be seen as a purely abstract functionality connecting the adjacent sides for visible information while at the same time separating them. This last aspect reflects the window as a purely ideational structure. Each view defines a particular net of interactions, that makes the window analyzable and optimizable in a systematic way. So structuring is part of our problem-solving processes. Even our material world is not just structured in itself but mere materiality resisting our acting towards it. Thus, we come to distinguish entities and qualities related to what our actual problem and purpose is.

1 Theories, Models, Structures

Three notions will frequently appear throughout the book: structure, model, and theory. They are — not only here — utilized in a wide range of contexts. None of the words has a universally accepted definition. Their meaning depends on the disciplinary environment in which they are used and on the educational, sociological, or scientific background of the persons employing them. This book cannot and will not make an attempt to provide precise definitions, but rather will explore the meaning of these notions in different contexts, explain their usage by examples, point at differences, and work out common features with the aim to help improve interdisciplinary communication.

Theories. To scientists, a phrase "…theory" signals a particularly well-tested set of scientific ideas for which there is some range of phenomena where the theory gives correct predictions every time it is applied. A theory represents the well-established understanding of a system of objects, mechanisms, or processes. It is often formulated in terms of mathematical formulas. These make the intellectual concepts precise that have led to formulating a theory. A theory can never be proved to be complete and final. New discoveries may result in generalizations, like the special theory of relativity or quantum mechanics extending Newton's laws of mechanics to treat processes with velocities approaching the light velocity or phenomena on an atomistic scale. What system of ideas is called a theory depends on the history and development of a scientific field. While nobody doubts that there is a mature theory of electromagnetism, some may question that psychoanalysis, a family of psychological theories, has a similar scientific status.

Models. Models are employed to describe certain aspects of objects or phenomena without claiming to represent reality. Typical physical models are the models architects construct to show the outside appearance of a building. If this building is an opera house, a sound engineer responsible for the acoustics may build a very different model. He would ignore the outside and focus on the acoustically relevant aspects of the interior, such that experiments can be made to simulate the sound propagation.

A conceptual or abstract model is often phrased in mathematical terms using variables, equations, inequalities, or other logical and quantitative relations. The aim of such mathematical models is to enable reasoning about the modeled object or process within the underlying mathematical theory. Just as in physical modeling, one and the same object can be investigated by employing different mathematical models. Each of these mathematical models will ignore certain aspects and focus on others. The reason for using a multitude of models is that we are frequently unable to understand the process or object on the whole. Instead, we try to capture certain aspects by partial models.

At present, for instance, nobody is able to set up a mathematical model of an opera house representing all aspects important for the construction of such a building at the same time. One mathematical model may reflect the statics of the building, another air conditioning and heating, a third sound propagation, etc. All these models depend on each other, but to reduce complexity, special features are singled out while others are kept fixed.

The relation between model and theory is not clear-cut. The word theory is usually employed for “general aspects”. Quantum theory and thermodynamics in physics, algebra and probability theory in mathematics, and equilibrium and game theory in economics are well established theories addressing broad ranges of phenomena. One would, however, not speak of a theory of motor vehicles. Cars are modeled, and there is not only one model. There are mathematical and physical models for the aerodynamics, models for stability and crash simulations, models for the engine, one for the catalytic converter, etc. Many theories are employed to model a vehicle properly.

Theories and models are both judged by their ability to predict consequences. Gravitational theory can be employed to compute the trajectory of a space craft flying to Mars very exactly, and various mathematical models for cars have significantly helped to make vehicles safer, to reduce energy consumption, and produce cars more efficiently.

Structures. A usual definition says that the structure of a thing is how the parts of it relate to each other. This is not really helpful. It seems, though, that we have an intuitive feeling or maybe even an inherent understanding for what structure is. We may even assume that the way our brain operates on its sensory input strongly determines the generation of (abstract) structures. Originating from the ancient impulse of making the material environment managable by ordering and introducing causality, facilitated by the existence of our cerebrum, we extend this method to purely theoretical constructs as well. Only this ordering process makes them manageable with respect to understanding and, sometimes, melioration. Of course, there is history, tradition, experience, there is process in general: We do not determine structures anew in every act of problem-solving. Embedded in our collective memory, in our language, our rules of conduct, social or technical norms, or our scientific knowledge and last but not least in our artefacts there are given results of former acts of structuring, resulting from problem-solving in other, sometimes similar contexts. We use to start working on the base of these given structurations, we recognize a window when we see ’it’. In our everyday practices working with given structures prevails compared to inventing new structures. Therefore we tend to assume that these given structures are part of ’the nature of objects’. However, this essentialist view reveals itself as insufficient as soon as we come across new problems as it is often the case in sciences and humanities. Solving these new problems requires us to establish new views on the case and its context.

The interactions of a structure with its environment have to be carefully observed in any analysis. Usually, we tend to isolate an object from the surrounding world for analytical and abstracting purposes in the sense of defining a system, i.e. the object under consideration, and the environment. Without this isolation an abundance of influences may have to be taken into account when an object is to be modeled and investigated. Let us return to our window example that we might like to optimize with respect to its energetic behaviour. In a first try, we limit our point of view to the interior structuring, considering everything else as a non-interacting environment. Accordingly, we model its inner structure only and subject the available parameters defining it to an optimization procedure with respect to the minimum loss of heat. Then it is quite obvious that the window will turn into something small (because the surrounding walls are better isolated), thick (as more material reduces the overall heat conduction), not very translucent (since heat-reflecting panes tend to be dim), and directed southward (in order to accumulate energy gains by direct sunlight). Clearly the ’optimized’ object is not the technical solution we had in mind as human beings when we started out optimizing it. Had we, instead, isolated our perspective to just one of the other aspect we would have come to other, but similarly unsatisfying results. So, any one concerned with the assessment of optimality of a window in a house, such as the carpenter, the psychologist, the police and, finally, the housewife cleaning it, will arive at different results. We conclude that an ’optimal’ window can thus only be achieved if the model includes the relevant interactions with its environment. Recognition of these interactions will by itself lead to a multitude of mostly conflicting and subjective property demands as will be treated later in this book. The impossibility of treating the window independent of its interactions with the environment reflects the common observation that there is a wide variety of window concepts in buildings. They arise from the rather individual interactions with the environment along with a plentitude of conflicting and subjective preference settings. So, an analysis of structures and structure generating processes requires understanding of the objects under consideration as well as their interactions among themselves as well as with their environment.

2 A simple example: Structures in Fluids

In order to arrive at fundamental insight and classifications, a first look on structures and structure generating processes is reasonably focused to highly simplified systems, where the objects and the interactions between them are well understood. An example of such systems are fluids. The objects are the molecules, the internal interactions are those between the molecules, and the external interactions those between the fluid and its environment. Due to the simplicity of these interactions important and formal insights can be attained that will be generalized to much more complex, macroscopic structures in other fields of interest. The science providing a universal theory in such systems is thermodynamics. Thermodynamics tells us that there are two fundamentally different types of structures, referred to as static and dynamic.

2.1 Static structures

Let us first look at static structures. A prototype of a static structure generated by nature in a fluid is a snowflake (fig. 3). Such structures are also referred to as structures in equilibrium. Here, equilibrium is that state of a system attained when it is isolated from its environment and when all relaxation processes have come to an end. So, all interactions with the environment are eliminated and we have a final, i.e. not any more changing state.

The global principles leading to ’static structures’ are well understood. In an isolated system, the second law of thermodynamics requires, that the entropy of the system under consideration must increase in any real process. The state of maximum entropy, consistent with the constraints of the isolated system, is the equilibrium state and plays the part of a predetermined goal. In that state the rate of entropy production is zero, because all processes have come to an end. Under suitable constraints (but not always) maximum entropy will require the generation of macroscopic static structures, such as that of a snowflake. When we follow the structure generating process by looking at the modeled interactions of the molecules on a computer over time, we see that the evolutive relaxation process to maximum entropy indeed leads to the emergence of a static structure, when the constraints of the system are appropriate. As a very simple example, when we cool a pure gas, say water vapor, we observe that the molecules will at a certain temperature generate a structure, in particular a spatial separation of two

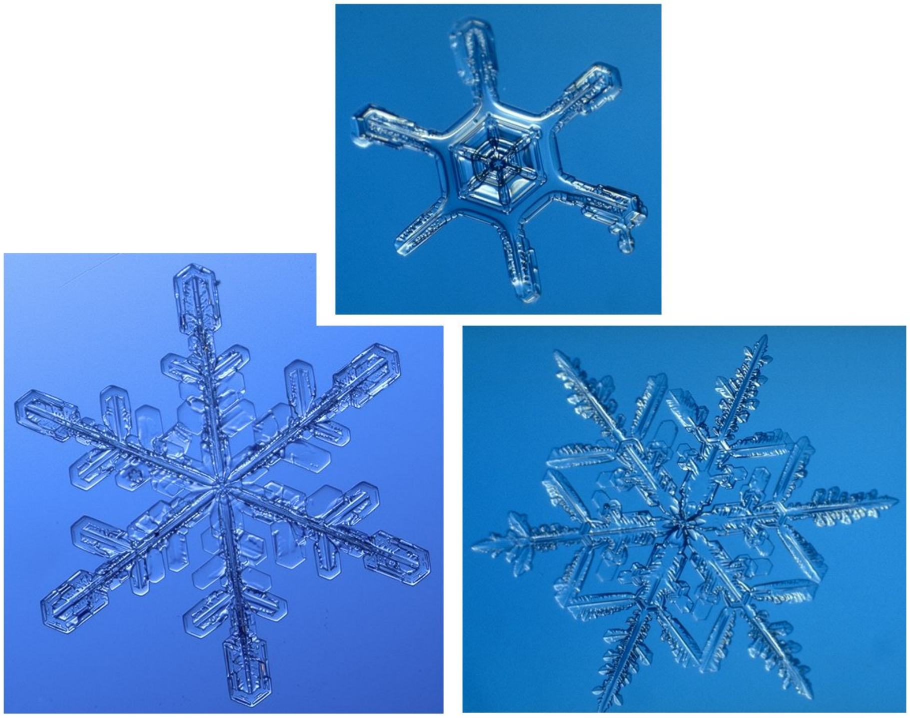

Abb. 3. The always varying structures of snowflake crystals. Although every crystal differs from all others there a common structuring principles, like the six-fold geometry.

phases, gas and liquid. This structure will persist after isolation of the system, i.e. after eliminating all interactions with the environment. So it is a static structure.

Much more complicated static structures are known to appear in fluid mixtures as a result of interactions between the molecules by cooling processes from the disordered gaseous state, like the phase and reaction equilibria exploited in an industrial scale in chemical processes. The particular appearance of the structures depends on the particular intermolecular interactions. The plentitude of internal interactions between the objects, is the reason for the enormous plentitude of static structures in fluids, that we observe. The entropy maximum principle explains that the existence of a structure is subject to particular constraints. It will disappear when the constraints are violated, e.g. at lower and higher temperatures and pressures, since the entropy maximum will then require a structureless state. We note that for purely mechanical systems, i.e. systems at zero temperature, the maximum entropy definition for isolated systems transforms into one of minimum free energy. [81]. So, starting from the geometry of a simple molecule, as calculated today in a standard way from quantum mechanics by looking for the minimum energy, up to that of a folded protein, as studied in biochemistry, molecular equilibria can be considered as static structures (fig. 4). We note that conditions can occur in which the minimum of (free) energy is not attained by a system on its way to equilibrium since the required reaction energy is not available. We then find frozen states, which when the hindrance is eliminated, invariably move to the state of minimum free energy with the emergence of the associated static structures.

We summarize that static structures in simple fluids originate from the particular interactions between their molecules and are generated by relaxation processes towards a predetermined goal, which here is maximum entropy, unless hindered by particular constraints. These relaxation processes can be reproduced on a computer, although the final result can in such simple cases more easily and reliably be calculated in one rational step by a direct mathematical optimization routine. Static structures are

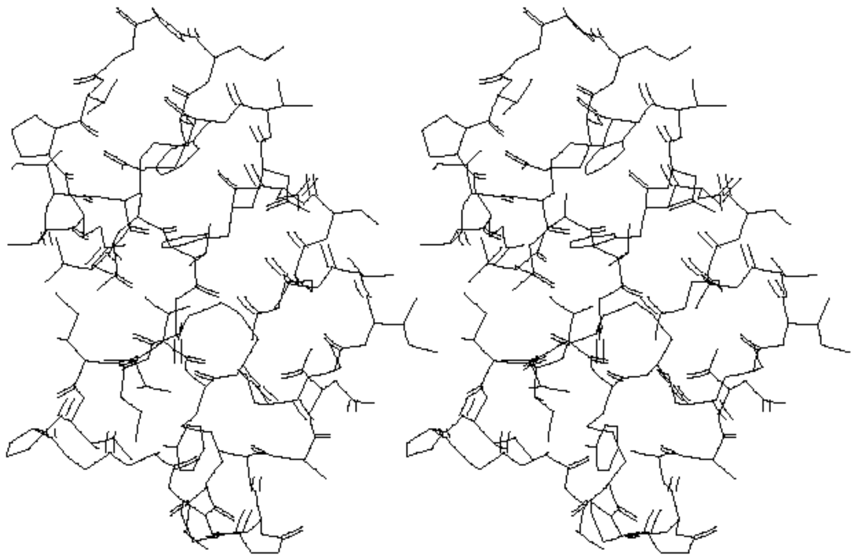

Abb. 4. Folded three-dimensional structure of a proteine molecule, representing the state of minimal free energy. The structure is displayed in a stereographic representation. If observed with a sqinting view the spatial structure appears after a while of eye adaption.

predictable and reproducible. So, given the full definition of the molecular interactions in a thermodynamic system, the equilibrium state can be formally predicted, at least in principle, and the same static structure will be reached independently from a starting condition of the development. Clearly, static structures in fluids are no more than a simple example for much more complicated static structures placed into the society as technological artifacts, such as engines, buildings etc. We shall see, however, that those artifacts share important fundamental properties with those of the static structures of fluids.

2.2 Dynamic structures

The interpretation of structures as static, i.e. as arising from relaxation processes towards a predetermined goal and remaining constant after eliminating all interactions with the environment, is adequate only in a restricted view. Evidently there are other types of stable structures which are generated and kept in existence by interactions with the environment through a continuous transfer of energy and matter in open systems. This input of energy and matter stabilizes a state away from equlibrium, and so, these structures are referred to as nonequilibrium or dynamical structures. They appear to us in the beauty of living nature, as well as in form of processes, social networks and in practically all operational appearances of our man-made artifacts (Fig. 5).

Again the study of fluids, on the basis of interactions between simple molecules, is a convenient basis for getting insight into the generating and sustaining processes of such dynamical structures. Two simple and well-known examples may serve to illustrate the point and will then be generalized. One is the Bénard-instability, studied in hydrodynamics. (Fig. 6). A horizontal fluid layer between two plates of different temperatures in a constant gravitational field will show a structureless heat conduction phenomenon at small temperature differences. However, at a sufficiently large temperature difference the structureless state becomes unstable. Convection occurs, and the heat flow and along with it the entropy production increases. Although accompanied by a production

Abb. 5. A bridge as representative of a man-made dynamic structure. Even if it appears static at first glance, the continuous care in terms of inspection and maintenance should be kept in mind that is needed to retain its functionality over a long period of time.

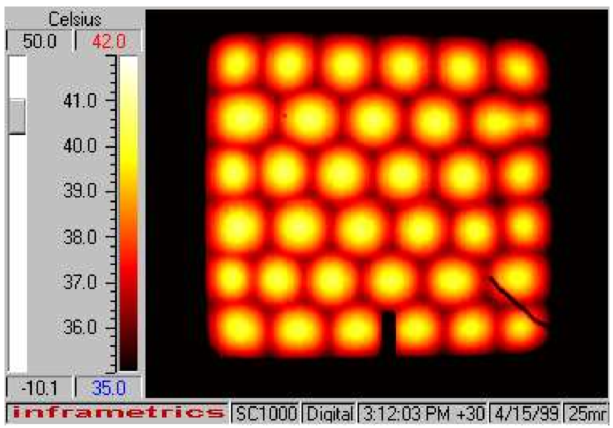

Abb. 6. Infrared camera view of the free surface temperature field above the onset of the Marangoni-Bénard instability in a 5 mm layer of silicone oil heated from below. The pattern is hexagonal, apart from the influence of lateral walls. (Curtesy of Physical-Chemistry Department of the Faculty of Applied Sciences — ULB)

of molecular chaos, this increase of entropy generates regular hydrodynamic patterns, a stable dynamical structure. So, although the entropy evidently is increased as a whole there are areas in the system which are obviously characterized by a local reduction of entropy, the structured Bénard cells. A non-zero rate of entropy production, kept up by an input of energy, is the basic source of this structure generating process.

The second example refers to dynamical structures generated by a certain kind of chemical reactions, the Belousov-Zhabotinskii reactions [1]. When specified chemicals are transferred to a reaction apparatus under suitable circumstances, a regular change of colours and also a migrating spatially coloured pattern can be produced. A similar moving pattern structured by the same principles of an oszillating biochemical process, although with much lower frequency, has been observed in a special mutant form of mice (fig. 7). Again, there is a considerable rate of entropy production associated with such a reaction process. And again this rate of entropy production is a source of structure generation.

So, in both of the above nonequilibrium structure generating processes, we find that a non zero entropy production rate is associated with the structure generation and

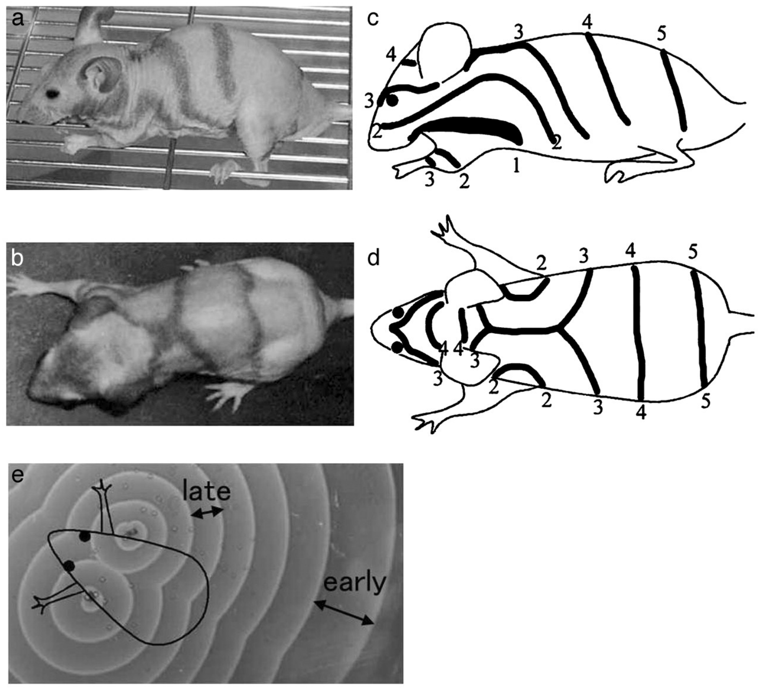

Abb. 7. Traveling stripes on the skin of a mutant Foxn1^{tw} mouse, with (a and b) showing side and top views of an adult mouse, respectively. (c and d) Schematic drawing of a typical wave movement in an adult mouse. The numbers by each line represent the time course of a wave. The time interval between the numbers is $\approx$30 days. (e) Traveling waves formed by a Belousov-Zhabotinskii (BZ) reaction. The shape of a mouse is superimposed. The outermost wave is the first wave. Intervals between waves are relatively wider for the first a few waves, then become shorter and constant. Experimental BZ waves are made by the standard method described in ref. [1]. To trigger two separate centers of concentric waves, two iron grains had been put on the dish. (Courtesy of Noboru Suzuki, Masashi Hirata and Shigeru Kondo)

preservation. This entropy production is made possible by specific interactions with the environment. The particular type of structure depends on both, the interactions with the environment as well as the internal interactions of the molecules. When looking at the processes of dynamic structure generating in such comparably simple systems in detail some interesting features become visible that can be generalized to more complex situations. So, we observe that in the course of the still rather simple-structured nearest-neighbor interactions between the molecules long-range, macroscopic structures are created. In the case of the Bénard effect, i.e. in a hydrodynamic system far from equilibrium, small fluctuations, which are damped close to equilibrium, are augmented as a coherent effect and create a new type of macroscopic structure, stabilized by the exchange of energy with the environment. In the Bénard cells the global temperature difference between lower and upper interface plate is so great that the *local* environment of the interacting molecules is modified in such a way that in a first step of structure evolution small-scale deviations from the global temperature distribution arise by stochastic effects. They modify the environment and the associated interactions. These cumulated effects, taking place in the immediate surrounding, add up to stabilize and extend the self-organizing and stabilizing long-range structures.

Similar effects become apparent in the chemical or biological structures referred to above. In particular, we find an unexpected element of randomness, non-reproducibility and non-predictability in such dynamical fluid structures. The precise and detailed form of the Bénard cells is not predictable and not reproducible. They may have different sizes and rotation directions. The same is true for all dynamical structures of thermodynamic systems, notably those arising in chemical reactions. The theoretical foundation is the fact that the appearance of dynamical structures is a stability problem. Increasing the deviation from the equilibrium state leads to a point where this state becomes instable. The point of instability is referred to as a bifurcation point, at which fluctuations of an intrinsic nondeterministic character lead to an impredictable new state of the system, which is usually chosen out of various possibilities. Once this state is attained the system will further develop in a deterministic way on increasing its distance from equilibrium until a new bifurcation occurs. This new bifurcation is, of course, predetermined by the outcome of the earlier bifurcation, i.e. there is an element of history in structure generating processes far from equilibrium. They thus obtain an evolutionary character. There is no predetermined goal for the structure being obtained in detail. The particular appearance of the structure is determined by the particular interactions effective within the system and with its environment.

It should be noted, that increasing the driving forces beyond strenghts sufficient for (long-range) structure creation will usually lead to increasingly finer substructures by further bifurcations until at some point deterministic chaos production sets in: The dynamically created structures change so fast that they are no more identifiable as such. Accordingly a certain local predictability can be maintained only for short periods, with decreasing exactness in structuring forecast due to accumulating nascence of new structure nuclei. The ordering principle leading to such states remains the same, however.

We note the fundamental difference of the dynamic vs. static structures in simple fluids. The latter are characterized both by a maximum of entropy in an isolated system and a zero entropy production, while the former are stabilized by a non-zero entropy production, accompanied by an inflow and outflow of mass and energy. Also, contrary to static structures, dynamic structures cannot persist after isolation. They need the interactions with the environment. As soon as the energy inflow and outflow, along with the resulting entropy production, are reduced under a critical value, the dynamic structures will collapse to a structureless or staticly structured state.

3 Real-world structures and structure-generating processes

Let us now generalize the fundamental properties of structures and structure generating processes, as elucidated by a consideration of simple fluids, to the material and immaterial structures of our society. Are such material structures as an automobile or a bridge static or dynamic structures? The answer evidently depends on the type of analysis. An artificial structure may well by classified as static, when we consider its creation. Although the internal objects interacting may be quite specific for each object under consideration, they will be identifiable. Let us consider an automobile as an example. Here we have wheels, the engine, carriage and a number of electronic and mechanical devices as interacting objects. The predetermined goal for generating the static structure of an automobile may be minimum cost, minimum fuel consumption or others. However, if we consider the operation of an automobile or its performance on the market, we have to consider it as a dynamical structure. It is determined by the interactions with the environment, such as acceptance by the customer and maintenance. Without such interactions it will eventually be destroyed and turn into structureless heap of rust and plastic. The external interactions with the customers make it depend on a plentitude of objectives and the plentitude of automobiles on the market reflects such interactions with the environment.

The structure generating processes leading to technical structures can be considered as evolutive relaxation processes towards a predetermined goal changing over time by interactions with the environment. So, they have elements of static as well as dynamic structures. Value assessment for automobiles has changed over time the predetermined goals, such as design, safety, fuel consumption, all being summarized as success on the market. A similar example is a power plant process with a varying number of feedwater preheaters. Subjected to an optimization towards maximum efficiency, such a system will eventually approach an idealized Carnot process by adapting an infinite number of feedwater preheaters. However, when minimization of investment and fuel costs is considered simultaneously as predetermined goals, quite different power plant processes will arise from the optimization process. Also, when analyzed over a long time period, there will not be a predeterminable goal for a power plant, since assessments and values change in the society. The actual power plant processes will then adapt themselves to the varying short time goals in an evolutionary process. So, we see, not only the intrinsic interactions of the objects but also external interactions in terms of objective functions and value assessments are responsible for the plentitude of structures that we observe. While the view of a static structure is adequate in a restricted sense, we generally have to consider all the technical and sociological artifacts as dynamical ones.

In the social sciences a most common structural issue of vast importance is what we call social stratification. That is, the as we know unequal distribution of life chances among the members of a society. Income, education, job opportunities, status attributed on the base of recognized social capital — all this adds up to a view of society in the large as structured in various ways. While social stratification is a type of macro structure effected only indirectly by individual acts, we have likewise to deal with meso-level and micro-level structures where the actors' influence on structure-building processes are easier to see: work flow and hierarchies in organizations, sets of relations in peer groups or the microstructures of interaction most often invisible to actors — at least as long as interactions proceed without crisis and on routine grounds.

We conclude that most of the real world material and immaterial structures and structure generating processes that we wish to study in this book belong to the class of dynamical structures. All structure generating processes, technical, sociological, biological, are associated with interactions with the environment, accompanied by entropy production. Based on simple balances of energy and entropy, this requires an input of energy with a low entropy content and the capability of the system to export the entropy produced by the structure generating process. This can easily be verified for the earth as a whole. Here, we profit from structure generation by influx of solar energy, which has a low entropy content due to the high temperature of the sun, and the export of low temperature heat by radiation from the earth to the surrounding space, which carries all the entropy with it. Looking at technical systems confirms the general conclusions. Low entropy energy must be transferred to the system, such as fuel or electricity, to generate and preserve structures such as a running engine or an illumination during night time. For living organisms the low entropy input is contained in the food. High entropy energy must be exported in the form of waste heat or waste material.

The details of structure generation depend on the specific intrinsic interactions of the system as well as on the interactions with the environment. Generally we have to consider much more complicated interactions in analyzing the structures of our civilized world. For example, generating the structure of a bridge or a power plant requires the consideration of interactions between material properties, technological components, natural laws, and, in particular, man made values of efficiency and desirability. Sociological structures are determined by interactions between human beings. We shall return to the analysis of these interactions in chapter 4, when we discuss structures and structure generating processes in various disciplines. No matter how complicated the considered structure is: As soon as the inflow of energy and the export of entropy are blocked, all structures will collapse. A bridge without maintenance, i.e. after isolation from all artificial energy and mass transfer, will eventually rust away in a structure destroying process. A structure in logistics will cease to exist when nobody will survey it and keep up its existence. An infrastructure will eventually disappear when it is not continuously used. A structure of thought will be destroyed as soon as the human beings adhering to it will stop thinking it.

The analogous fate is bound to happen for any natural and artificial structure, including human life.

Structures are not timeless, but at the same time they are not necessarily progressive. If we think of the Inuit and their knowledge about snow and ice, we can easily see both that the ability to differentiate varies with the tasks to perform and that established fine grained structurations can get lost as soon as they lose their relevance for problem-solving: For driving a car on a road in winter it is pointless to distinguish 30 or 40 different types of snow, as the Inuit do. Instead, we just diffentiate by a binary distinction: snow or no snow, frozen or not frozen. In domains where the properties of 'snow' are more important, like in winter sports, at least some additional differentiations (powder snow, crusted snow, . . .) are maintained even in our modern civilization.

In sociology the most important and at the same time the most problematic relation is the link between structure and acting: While in mechanics it can be said that structures have direct causal impact (like the molecular structure of a certain metal that is directly responsible for the ability of the metal to carry a certain weight), in society things are different: processes in the social domain consist of acting and they necessarily involve actors. The often urged layman view, that structural factors cause certain actions, is problematic since in society people act on the base not of structures, objects, problems or other people. They rather act on their interpretation of structures, objects, problems and so on. Though by and large we might say that social structures have an enormous impact on educational success (as documented in various transnational learning and teaching effectivity studies), this is not necessarily true for every single case. Every detailed study of, say, decision making in educational processes would show how actors continuously (though not always consciously) interpret their context, including aspects of status, economic resources, family traditions in order to come to a conclusion about, say, going for another degree or not.

It is clear that all structure generating processes in principle share the basic properties of dynamic structures, such as non-reproducibility, non-predictibility in detail and lack of a predetermined goal. This is immediately evident, e.g. in sociological structures, in those of diseases and in those associated with weather and climate. The associated structure generating processes are difficult to control and impossible to predict in detail. However, the laws of their generation can be analyzed by considering the relevant intrinsic interactions and fruitful generalizations, as well as average predictions can be made by statistical methods. Furthermore we will see that even dynamic structures *can* be optimized by deliberately blinding out their dynamical nature in representing them by simplified, quasi-static models. This is a reasonable strategy as long as the human optimizer is aware of its limitations. Accordingly, the intrinsic interactions in such system models as well as those with the environment are to be carefully modeled, with the dynamically dominated effects of bifurcations and determinstic chaos etc. being as precisely accounted for as possible.

4 Modeling Reality

It is an established fact that the vast majority of systems or processes in the real world are so complicated that there is no hope and even no sense in trying to analyze them in full detail. Instead, scientific analysis has to be liberated from the confusing plentitude of phenomena in order to reduce the complexity of seemingly unsurveyable problems to something that is amenable to analysis. The method of analysis may be mathematics in the technology oriented applications or observation and thought along with creating notions and their operational interactions in the social sciences and in the humanities. The very process of modeling even a small part of reality is naturally accompanied by a loss of realism, in the sense that some aspects are deliberately eliminated from further consideration. So, there is always the danger of an unacceptable disparity between model and reality, which has to be taken into account when conclusions about model behavior are transferred to conclusions about reality. Models cannot be justified or evaluated within the categories right or wrong. Rather, they are either useful or not useful. The judgement about this classification is subject to a comparison with reality, e.g. by experiment in the natural sciences.

4.1 A simple example: Models of fuids

We wish to illustrate the problem associated with modeling reality on the basis of a simple example, once again chosen from the thermodynamics of fluids, treated at varying levels of complexity, each level being adequate to describe certain physical phenomena in different state domains of a substance. We know, that almost any material can be encountered in a liquid or a gaeous state, both being summarized as the fluid state. How can the properties of this technically very important state adequately be modeled, e.g. in order to derive additional information about its behaviour in certain artificial boundary conditions, like chemical apparatuses?

Let us first consider gas of monatomic molecules at normal temperatures and pressures. An adequate model for this reality is that of small billiard balls. This implies a particular conception of the nature of their interactions. In the billiard ball model of a fluid the members are hard balls with a negligible volume, and the nature of the interactions is such that, apart from occasional elastic bounces, there is no interaction at all. So, the billiard balls move independently of each other and their motion is described by the laws of classical mechanics. On the basis of this model one finds formal expressions for the energy, the temperature, the pressure and the entropy of such a gas which are in perfect agreement with the properties of a real gas like Argon over a significant temperature range at normal pressures. To illustrate, we plot the molar isobaric heat capacity of Argon divided by the universal gas constant against temperature in Fig. 8. It is a constant of value 5/2, in perfect agreement with precise measurements.

Abb. 8. Fluid phase behaviour of Ar, shown by its molar isobaric heat capacity divided by the universal gas constant against temperature, for a limited range of pressure and temperature.

This quantity determines the temperature dependence of thermodynamic behavior and plays a significant role in various engineering applications.

It is remarkable that such an ultrasimplified model can be useful in describing the properties of a fluid. It claims that the molecules of Argon do not have any type of interaction with each other, neither that of attraction nor that or repulsion. Quite evidently this is wrong since it is in conflict with our basic knowledge of the interactions between the molecules in fluids. The contradiction is only apparent, however, and its resolution is that the simple billiard ball model is only probed in a very limited region of the phase diagram, i.e. only at normal temperatures and pressures. In this region, indeed, the interactions between the molecules, while basicly effective, are not essential for the macroscopic properties of the fluid, due to the relatively large average distance between them. Events, in which two molecules approach each other so closely that the interactions between them become noticeable are simply too rare to be statistically significant and therefore can be neglected in calculations of the macroscopic properties. So, while the billiard ball model is a useful model for the properties of a real mono-atomic gas at normal temperatures and pressures, it does definitely not represent the general situation in an arbitrary fluid.

When the properties of a fluid like Argon are to be analyzed over a larger region of states, it becomes necessary to change the model. Going to higher temperatures, say 10,000 K, real Argon starts to ionize, an effect which cannot properly be described by the billiard ball model since it does not allow for a disintegration of the balls into nuclei and electrons as happens in reality. Going to higher pressures and lower temperatures the phase behavior of Argon becomes entirely different from the prediction of the billiard ball model. Fig. 9 shows the fluid phase behavior of Argon over a large region of states. The region adequately described by the billiard ball model is shown as a shaded region in the phase diagram.

Abb. 9. Fluid phase behaviour of Ar for a large region of states.

A most significant deviation from the prediction of the billiard ball model is the effect of condensation. This is a simple example for a structure generating process in a fluid. Increasing the pressure at a sufficiently low temperature from the low pressure billiard ball region eventually leads to a situation where the average distance between the molecules decreases so strongly that the short-range interactions cannot be overlooked any more. Condensation takes place, and in a gravitational field the liquid and the gaseous region are separated by a meniscus, clearly pointing to the appearance of a spatial structure. Hence, a model for a fluid over the full region of states requires the interactions between the molecules to be taken into account. These can be modeled on the basis of quantum mechanics. With these interactions the properties of fluid Argon can then be predicted over a large region of states by the methods of statistical mechanics along with mathematical algorithms. We note that even this more demanding problem of predicting the fluid phase behavior of Argon over a large region of states, including the structure of the vapor-liquid equilibrium, is usually based on a model which significantly reduces the real complexity. This particularly holds for more complicated molecules, where the interactions depend on their internal structures, and so, on many coordinates. The models for the intermolecular interactions currently used for predicting structures of fluid phase behavior are strong simplifications of the real world. However, when used in the combination with an adequate model for the connection between the macroscopic fluid phase behavior and the intermolecular interactions we are able to make predictions of great practical value. This is shown in Fig. 10 where a prediction of the vapor-liquid

Abb. 10. Vapour-liquid equilibrium modeling estimate compared to measured data (left), showing good conformance. XYZ calculated with the same model of molecular interaction exhibits significant discrepancies between model prediction and measurements.

equilibrium in the system is compared to experimental data. Evidently, the model, while being far away from the true situation in the fluid, is a useful representation of reality in the context of vapor-liquid equilibrium prediction. It is by no means guaranteed, however, that this model predicts equally well other aspects of fluid phase behavior of the same system, as can indeed be verified.

Various aspects of modeling reality, which hold quite generally in all branches of science, can be studied from the simple example of fluid phase behavior. First, a model has to be tailored to the problem to be discussed, e.g. predicting the structure of the vapor-liquid equilibrium in a fluid mixture. There is no sense in extending the model to more and more complexity if only a segment of reality is supposed to be analyzed. So, when the gas phase properties of a fluid such as Argon are to be analyzed, there is no sense in setting up a model for the intermolecular interactions between its molecules. Further, when only the vapor-liquid equilibrium structure in a fluid mixture is to be considered, a satisfying model can be based on a rather crude model for the intermolecular interactions. The demands on model complexity become more and more complex as further aspects of fluid phase behavior are to be studied, such as liquid-liquid equilibrium structures and heat effects. A model representing a large segment of reality will necessarily be much more complex and thus require more effort on applying it to the prediction of fluid phase behavior by statistical mechanical and mathematical methods.

Similar layers of increasing detailedness occur in many modeling efforts when coping with the need of increased fidelity of reality. As an illustration, let us have a closer look on the energetic supply of a residential area with district heating and electrical power supply. A first order of rather crude detailedness is the consideration of the added-up peak needs in heating and electrical power for the intended set of houses. If the existing electrical power generators and heating stations provide these respective maximum values the demand will be satisfied. If the district is to expand by building additional houses and a sensible statement is required on how many of them should be allowed, the power requirement model has to be refined. The coincidence that all customers are drawing their peak requirements at the same time will never occur. Hence it would be much more realistic to model the requirements by a time-resolved demand structure on known peak demand days separately for electrical energy and heat. A further increase in the number of supportable houses in the residential area requires a correlation of electrical power and heat demands, opening up the possibility to utilize low-temperature waste heat of the electrical power generation for domestic heating purposes. Here the coincidence of respective demands will have to be mapped, requiring again a more complicated model.

4.2 General principles of modeling

In the context of analyzing structure generating processes we explicity base our considerations on process models with a certain self-organizing intelligence. Such a model is particularly fruitful in our interdisciplinary context since in many applications it can be implemented in the form of a computer code with an ab-initio unfathomable abundance of possible system reactions upon changes of boundary conditions. Such models are usually called simulation systems, as they try to mimick the causalities of the real world.

In the humanities a somewhat different notion of modeling prevails, as the primary goal of those disciplines is mostly description, not targeted change. Here a descriptive social network is one of the primary ways of modeling. Through the use of formal representations — like directed graphs showing how each unit is linked to certain other units in a network — the model-based analysis of social networks has the aim of uncovering structural patterns and studying their emergence, their consequences, and their temporal transformations. Various techniques are employed to develop respective models as well as to analyze them using mathematical, statistical, and computational methods (e.g. simulation of network processes). In sociology, archives of network data and related computer programs to analyze them are typical resources available to social network analysts. In this way, the area of social network modeling relates to the general sociological goal of understanding how social processes generate, maintain and change social structures, and how social structures both serve to constrain and enable social action. Typical modeling target issues are structured social inequalities, including social demography, socio-historical research, discourse analysis, ethnographies, and policy analysis.

In the causality-emphasizing models a given set of starting conditions is fed into the implemented simulating rule set of mutual dependencies, and the computer calculates the evolutive development of related output or monitoring variables. Although only a limited number of causal dependencies can be implemented as computer code the variation span of possible outcomes can be vast, due to the effect of combinatorial explosion. In short, it describes the fact, that possible outcomes of modeling variable state changes tend to be multiplicative. A nice episode in the musical domain, taken from [73], pinpoints this phenomenon:

„As a boy, John Stuart Mill was alarmed to deduce that the finite number of musical notes, together with the maximum practical length of a musical piece, meant that the world would soon run out of melodies. At the time he sank into this melancholy, Brahms, Tchaikovsky, and Rachmaninoff had not yet been born, to say nothing of the entire genres of ragtime, jazz, Broadway musicals, blues, country and western, rock and roll, samba, reggae, and punk. We are unlikely to have a melody shortage anytime soon because music is a combinatorial system. If each note of a melody can be selected from, say, eight notes on average, there are 64 pairs of notes, 512 motifs of three notes, 4,096 phrases of four notes, and so on, multiplying out to trillions and trillions of musical pieces.“

To conclude, even if the amount of modeled variations in an evolutive simulation system are few and easily manageable, the potentially reachable variation space need not necessarily be so: *Simple rules can lead to complex and seemingly purposeful behavior.*

Statements of this kind are often made and supported by more or less abstract formalisms, but an illustration is more easily remembered. For this purpose we choose a two-dimensional cellular automaton, known as John Conway's game of life, running on a rectangular grid of infinite or finite size. Grid nodes (mostly depicted as cells) are occupied or empty, and there are only four rules determining the pattern in the next generation:

i) Occupied cells with no or one occupied cell in the neighborhood become empty,
ii) occupied cells with two or three occupied cells in the neighborhood stay occupied,
iii) occupied cells with four or more occupied cells in the neighborhood are emptied, and
iv) empty cells with three occupied neighboring cells become occupied.

Given these rules the unfolding of the cellular automaton in time is completely determined by its initial conditions, but nevertheless, an incredible richness of different dynamical behaviors results from the 'game of life' rules. Some of these patterns are even suggestive of purposeful design. For example, from one neither random nor fully ordered initial pattern a forever living periodical structure develops that emits small gliding motifs in one precisely defined direction. (figure 11) As to be anticipated, larger structures, although adhering to the same simple rules, show even much more diverse and astonishing features up to the point of reacting to 'environmental' influences in almost intelligent manners [2].

A well-known evolutive simulation engine, the computer game 'Civilization' [79] (fig. 12), may serve as an example that combines a large rule set and a mixture of continuous and discrete variables. In a vast network of mutual dependencies human activities like exploration of unknown territory or technology, war and diplomacy are simulated in their effect on an endeavoring society. The player, thought as a ruler of his people, has to make decisions about which improvements or units to build in each city, where to build new cities, and how to transform the land surrounding the cities for maximum benefit. The number of possible simulation engine reactions is so large that it can be interpreted as almost limitless, even though there are boundary conditions for every variable. The actual state of the game is represented by a larger amount of system variable states. Adding to the variety of simulation answers is the additional inclusion of stochastic elements, like natural catastrophies. These set of features causes each game to develop different from any preceding, even if the player tries to repeat a former scenario.

Similar simulation systems, although less popular, are used in scientific modeling of development processes, e.g. to understand the development of social groups depending on political boundary condition settings over time. As soon as the behaviour of human beings is modeled the simulations tend to become unrealistic, though: Representing the

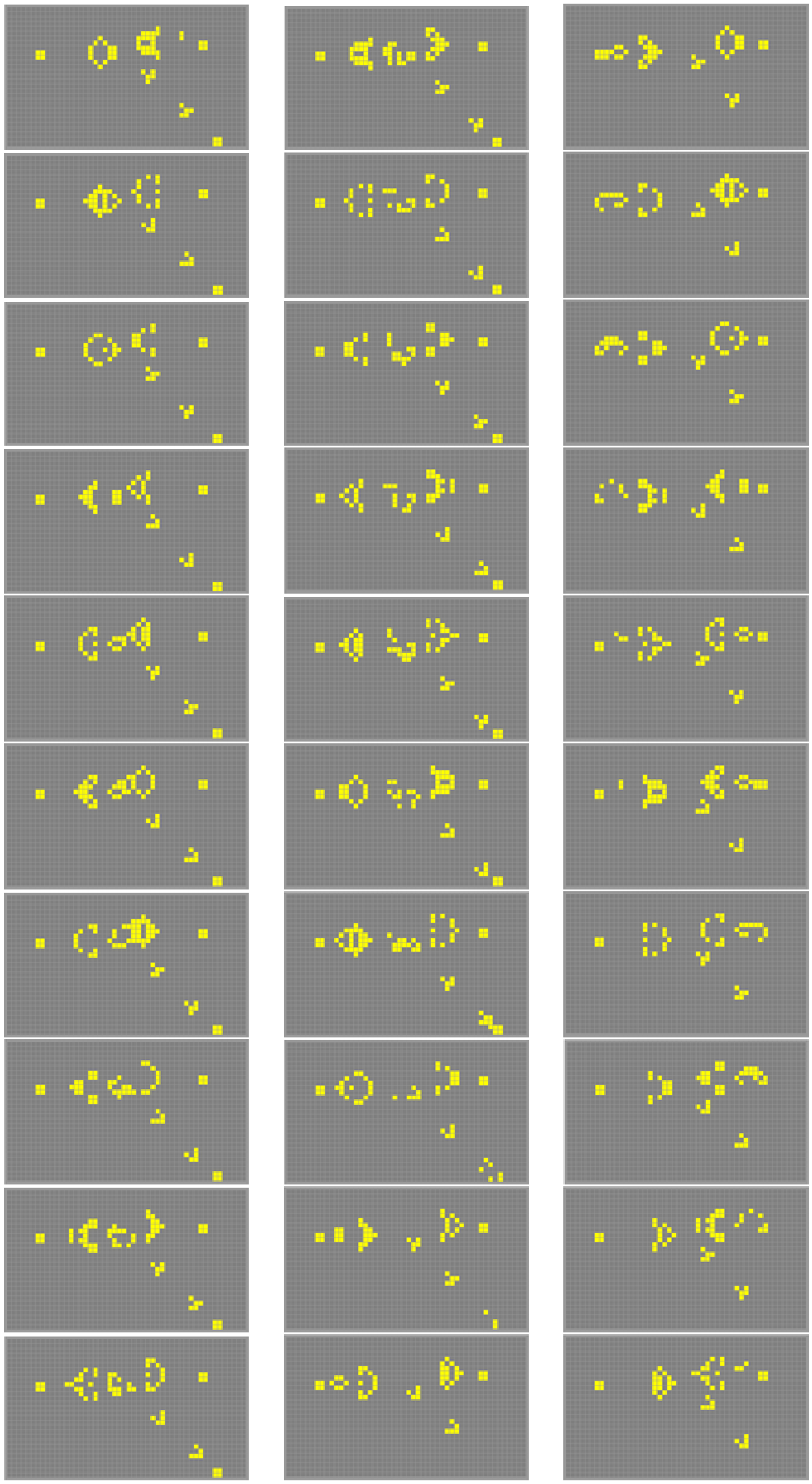

Abb. 11. Example sequence of a Game of Life configuration, repeating its core structure with a period of 30 interations and continuously producing 'gliders'. Images are sorted column-wise.

Abb. 12. Example screen of the society simulation game 'Civilization', taken from [79]. Some of the state-determining variables are depicted or listed in the corners of the screen, others are represented by respective parts of the drawing.

twists of the human mind in its reaction to more or less comprehendable boundary conditions is still a wide field of future simulation development.

A final remark appears adequate in relation to such process simulation in systems. Mathematical analysis strongly favors steady and differentiable dependencies, since they allow the application of rigorous and well-assessed optimization algorithms for system melioration. But this represents a significant simplification. So, frequently these simulation systems also contain large numbers of internal decision variables to reflect real-world conditions: If, for example, the real features of a hot water piping in a house, or the set of interconnections of chemical engineering apparatuses, is to be realistically represented, the discrete diameters of tubes available for construction purposes must be respected. If any calculation in the respective model is dependent on tube diameters its outcome will accordingly 'jump'. To represent tube diameters in such systems with a continuous variation usually impedes realistic calculations — especially so, if optimality points are sought in a respective framework.

4.3 The Economy of Modeling

Connected to the problem of an appropriate model complexity, discussed from the perspective of real-world reproduction fidelity, is the consideration on the effort that should go into the conception of a model from a practical point of view. Even if more detailedness would be desirable in many cases of optimization efforts with respect to a noticeable reproduction fidelity improvement, it may prove rather impractical to do so. Two main reasons to restrict modeling depth exist: i) the time requirements for evaluating a complex model, and ii) the required labour to set up the complex model, compared to the expected improvement the model promises to cause in an optimization effort.

An example for the first case is the weather forecast. Even if there were a model to describe a complex phenomenon like the atmospheric development, but the time needed to converge the modeling equations is too large, the real phenomena that should have been forecasted are on a smaller timescale than the forecast itself. This effect of extended computing power requirements is in most cases a transient one, caused by always lacking contemporary availabilities of adequate calculational power. Similar effects show up if rather fast phenomena, like real-time control units and similar applications, are to be scrutinized. The operation of an unmanned vehicle, striving to cover a given distance in a sensible amount of time, is such an application. Equipped with a set of sensors, like stereographic cameras, tire slip and bearing controls etc. the steering control unit must interpret many signals in combination, as no single input channel provides satisfying information on its own. Each channel input is ambiguous to a certain extent, so the optimizing interpretation of all available inputs needs to be balanced. If the model underlying this optimizing interpretation is too complex and thus too slow the vehicle is stuck in the ditch before the control unit gives the signal to circumvent the obstacle. Here a timely circumvention of a non-existing ditch, assumed by a crude but conservative interpretation of the acquired sensors data, is better than a too late one of a correctly recognized ditch.

An example for the second case, is the economization of the required human power input into the modeling of a given custom case. As an example a typical problem of chemical engineering shall be discussed. A general task in this field is the anticipatory performance evaluation of a coupling of several basic apparatuses into a so-called flowsheet (fig. 13). The model may be optimized for improved performance with respect to different target criteria, like throughput, product payoff etc. .

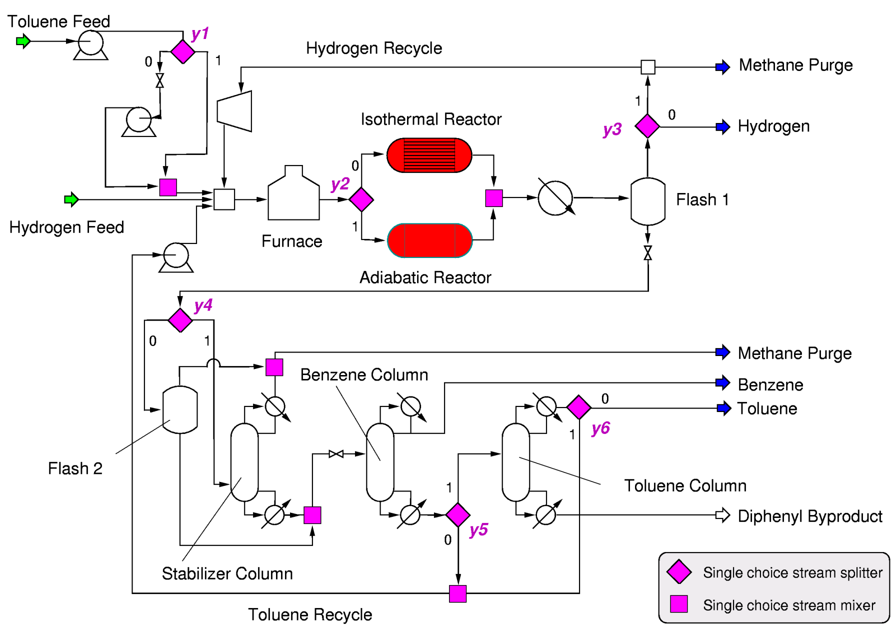

Abb. 13. Flowsheet of a benzene production plant, taken from [33]. Two educts are fed into a (mostly) serial connection of several basic apparatuses (reactors, pumps, distillation columns, . . .), producing the desired output of benzene, but possibly an additional by-product (diphenyl), depending on the concrete choice of the chosen path.

The boundary conditions the real plant has to obey are manifold, being mostly of thermodynamic nature. First of all, conservation laws for matter and energy have to be fulfilled. Second, the properties of the reacting substances, like separation efficiencies in distillation columns, impose limits on what a single apparatus can achieve. Third, at several points in the flowsheet there are decision points on the principal layout of the future plant. In the flow sheet this is represented by alternative paths the materials can take, leading to different product strategies, (human) reaction possibilities upon changes in market prices, apparatus developments etc.. Fourth, there are up to two recycle branches, recirculating some intermediately created substances back to the plant inlet. These potential recycle streams must be considered in the balances.

Basically there are two differing modeling approaches to obtain an optimal path choice and setting of respective operative values — the equation-based and the sequential modular aproach. In an equation based approach all simultaneously valid equations and inequality conditions are formulated for the complete system, creating quite a vast mathematically described collection of dependencies. Please note that in this approach values and settings expressing conditions in the educt section of the plant are frequently connected to values at the end or in the center of the system. Accordingly it is a very complex and time-consuming task to model a chemical reaction system on this basis. If done so, one is rewarded by a 'closed model' that can be subjected to a purely mathematical, rigorous and usually very effective and fast treatment. If small parts of the setup are changed, a lot of reorganizing work has to be put into the model, though, due to the strong interconnections even of spatially far separated units.

There is a caveat in the depicted flowsheet, though. It contains six decision points where a binary selection of one path or the other has to be set. Depending on the chosen path, the mathematical system of equations and inequalities changes significantly, barring a combined modeling of the complete plant. Instead, separate models must be created for every sensible combination of binary decisions. Even if some of the $2^5 = 32$ purely combinatorically defined ones can be dismissed by engineering reasoning quite a number will have to be elaborated to cover the interesting range. Accordingly such a modeling treatment is very expensive with respect to the involved human effort.

An alternative method is the setup of a sequentially modular model where mathematical representations of the individual apparatuses are coupled in an manner strongly resembling their physical interconnection: Substance and energy streams are coupled mostly on a local basis, splitting up the very complex interrelationships (like: the sum of discrete mass flows summarized over every parallel stream of the flow sheet must remain constant) into local subsystems called unit operations. These unit operations, being internally represented by smaller sets of equations and inequalities, may relatively simply be switched on and off for decision variable resetting, requiring rather few changes in global bilance adaptations. The modeling effort on the whole is less demanding than the equation-based one, but produces, by its very nature, a computational problem. Due to the localized, hierarchical nature of the modeled interactions the unit operations need to converge in themselves before respective stream redefinitions are to be forwarded to the hierarchically higher level of the unit-combining equation sets. Especially in the case of recycling streams such models have a tendency to oscillate as changed output streams of one unit influence the next in row, but eventually affect their own inputs again due to the circular loops. Such computational problems *may* be overcome by some increased calculational efforts, like decreased time steps, or upper bounds on the changes of important variables with auto-adapted time step decrement.

To summarize, the setup of an equation-based simulation mode requires an elevated amount of human effort that will be honored both by fast optimization times and, even more important, more stable system answers in case of mathematically difficult conditions. Whether this effort is useful in a practical sense depends on the expected return relative to an existing design. Similar decisions on the sensibility of respective modeling efforts are commonplace in most meliorating problems, up to the point that an elaborated treatment may not be adequate.

4.4 Uncertainty

The only thing that we can be sure of is the fact that we cannot be sure of anything.

In many cases the real world shows numerous imponderabilities which pinpoint different categories of uncertainty. Appropriately embodying theses uncertainties into engineering models provides ways and means to express the limited rationality of a plentitude of real world phenomena computationally. In harmony to the definition of uncertainty introduced by BOTHE [BOTHE93], uncertainty can be understood as a gradual assessment of the truth content of a proposition, related to the appearance of a specified event.

The following diagram (Fig. 14) demonstrates the potential subcategories of uncertainty and assigns them to corresponding theories that have been elaborated in the last two decades. Three different subcategories can be distinguished:

- stochastic uncertainty, which describes the random results of multiply repeated experiments where the boundary conditions have always to remain unaltered,
- informal uncertainty, which describes information deficits due to the limited information sources and an only small number of observations,
- lexical information, which quantifies a relevant real world fact in terms of a linguistic variable by ranking the membership of that fact to a defined uncertain set.

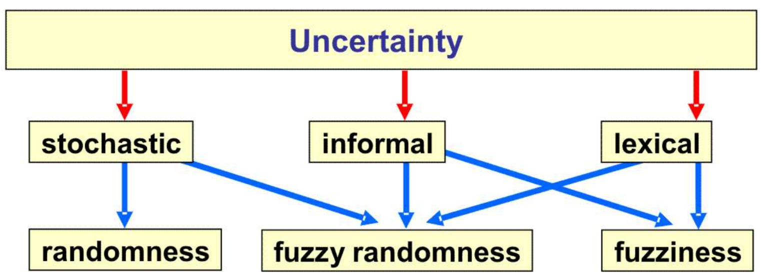

Abb. 14. Subcategories of uncertainty

There is a rather traditional treatment framework to calculate the effect of uncertain input or measurement values (Fig. 14, left) on the outcome of a functionally dependent output value: the error propagation method. This kind of uncertainty effect is almost omnipresent in the modeling of real world systems. It poses no major problem on interpretation of results as long as the model is not a self-evolutive one, meaning that the modeled responses of the system are directly depending on a set of input parameters. As an illustration, let us consider a very simple model for determining the required power output of a vehicle engine, driving at constant speed on a motor way. The modeling is done with a simple formula:

$$P = \frac{\rho}{2}(v + v_0)^2 A v c_w$$

(with $\rho =$ density of the air, $v =$ velocity of the vehicle, $v_0 =$ headwind velocity, c_w air resistance coefficient, and $A =$ head face of the vehicle)

The rule of error propagation then directly tells us how uncertainties in a given variable, say, the air resistance coefficient of the vehicle c_w, lead to a respective uncertainty in the calculated required locomotion power:

$$\Delta P = \frac{\rho}{2}(v + v_0)^2 A v \cdot \Delta c_w$$

Or, with other words, the *relative* uncertainty of required power, $\Delta P/P$, is identical to the relative uncertainty of the vehicle's air resistance coefficient, and the *absolute* uncertainty in P is constant for an assumed constant uncertainty in c_w. This is a dependency the common mind usually expects. But what about uncertainties in the velocity? If we apply calculus to the modeling equation we find for the absolute error in P

$$\Delta P = \frac{\rho}{2} c_w A (3v^2 + 4vv_0 + v_0^2) \cdot \Delta v$$

or, if we again look at relative uncertainties,

$$\frac{\Delta P}{P} = \frac{3v + v_0}{v + v_0} \cdot \frac{\Delta v}{v}$$

which gives us a definitely more complicated dependence to reflect on. If we experience no headwind ($v_0 = 0$) the relative uncertainty in P is three times as large as that in our observed variable v. This dependency can change dramatically, though, if we have a stronger following wind ($v_0 < 0$), e.g. in the order of half the speed over ground. In that case, a miss on our measured velocity will raise the propagation of relative uncertainty to a factor of five, far away from what the common mind would expect!

Although this effect of error propagation may come unexpected for a person not regularly involved in technical calculations, things tend to grow much worse in case of self-evolutive models. In these models, frequently applied to simulate the temporal development of a system, the calculative result of a value at a certain time interval Δt_i is needed to determine its value at then next interval Δt_{i+1} and later stages. In such cases simple error propagation techniques will soon lead to accumulated errors well beyond the actual parameter mean values, thus rendering their calculation worthless. In such cases, more elaborated techniques, based on in-depth probability and density function considerations, must be put to work.

Well known since years, but still advancing, is the probabilistic approach or the concept of randomness which captures uncertainty through random variables and/or random processes in an objective fashion, excluding subjective views on the problem. Thus, based on long-lasting observations or experiments, an effective probabilistic assessment of the quantities, patterns, actions and the structural response within the engineering systems and system processes can be made available. For stochastic variables, the key ingredients are expected values, mean values, variances, and also moments of higher order, quantile values, probability density functions (pdf) and cumulative density functions (cdf). For the wide variety of time-variant stochastic processes, functions for the average behavior, auto-correlations, auto-covariances, covariances and auto-correlations as well as cross-correlations are of interest. All of them can appropriately map the stochastic phenomena of uncertainty in engineering scenarios. Two problems, however, have to be realized if conventional randomness is applied in reliability analysis or optimal reliability-based optimum design: (i) a sufficiently large sample size is required if the pdf and cdf of a random variable have to be accurately defined, (ii) often, it is not an easy task to find the characteristic description of a real world process, with respect e.g. to stationarity, ergodicity, spectral moments, etc. . Consequently, if unfounded assumptions are made when a stochastic simulation model of a technical system or process is created, then additional uncertainties are induced on top of the attempted incorporatations of uncertainty into a model (i.e. uncertainty of the uncertainty).

In the past years, fuzziness has tremendously increased in prosperity as a new paradigm to differently computerize uncertainty in engineering (Fig. 14, right). Introduced already by ZADEH [ZADE65] around 1965 despite numerous hostilities, the fuzzy-based optimal decision finding and fuzzy control have been extremely successful in many engineering applications. In particular, this applies for highly complex problems of automatic control engineering where the solution of non-linear partial differential equations represented the regular modus operandi for a long time. As a matter of fact, many highly non-linear or near-chaotic problems, which could not appropriately be solved through numerical simulation (e.g. by means of non-linear finite element models), proved manageable by applying the fuzziness paradigm. The nucleus for fuzziness is the human attitude, even if no viable numerical representation of a problem is possible, to express solution behaviors in terms of fuzzy rules which connect uncertain input with uncertain output quantities. Both, input and output quantities are modeling informal and lexical uncertainties, i.e. non-statistical properties, where particularly subjective aspects are entered into the consideration. To this end, specified crisp input and output quantities of a basic set are transformed by fuzzification into fuzzy sets using so-called membership functions of differnt functional shapes (e.g. triangular, trapezoidal or curved). Based on the above tripod 'fuzzification', 'fuzzy rules ad defuzzification', 'fuzzy in fuzzy control', and based on the α-discretization and α-level optimization in structural analysis (see book published by Moeller [MOELL04]), data as well as model uncertainty of engineering systems and processes can be verifiably be handled. It should be mentioned that fuzziness can analogously be expanded from elementary quantities to functions leading to fuzzy functions, fuzzy processes and fuzzy fields.

Fuzzy randomness has to be introduced if neither randomness nor fuzziness is sufficient to describe the 'crude reality'. This happens when the rigorous preconditions and laws of randomness can not be matched, e.g. because a stochastic quantity is affected by informal and/or lexical uncertainties. Typical for this are (i) only small numbers of samples such that the type of the pdf or cdf has to be approximated with considerable uncertainty and (ii) the violation of fundamental principal of probability, i.e. the constant reproduction condition, by which it must be guaranteed that all samples of the taken universe obey identical boundary conditions. In practice, both premises are often violated. The universe may contain only a few samples, in many cases the boundary conditions are varying with respect to time and location. Also, the assumption of statistically independent stochastic variables and/or stochastic processes, used in many engineering application, does not correspond with real world scenarios. The effect of how a random quantity is fuzzificated can be seen in (Fig. 15) where a cdf is superponed with fuzziness.

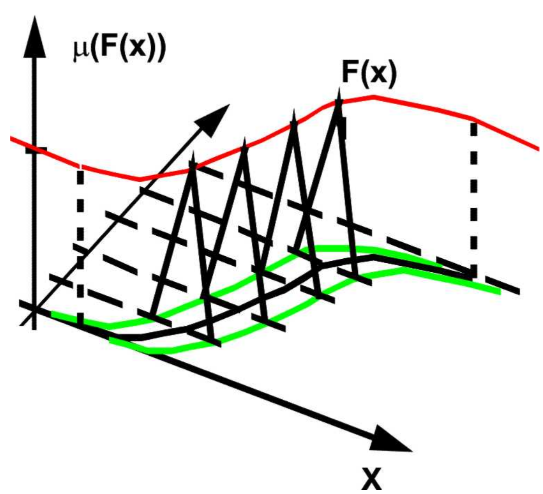

Abb. 15. Fuzzification of a Stochastic

In view of these unavoidable uncertainties all model predictions necessarily assume the character of a range of values, depending on the sensitivity of the parameter uncertainties to the macroscopic predictions. Any rational application of a model, including forecasts on the optimal target function values and settings of parameters, must therefore take such uncertainties and their consequences on the conclusions into account.

5 General Aspects of Model Analysis

We now proceed to the fundamentals of numercial analysis of a model. The motivation for that is basic understanding and, hopefully, melioration of a structure or a structure-generating process. The latter assumes that a goal can be defined. This almost trivial prerequisite, taken for granted in most cases, does in fact require further consideration if we leave the usual textbook examples. We will show several practically important cases where the definition of the ’correct’ target function is by no means stringent. There are even systems to which such a thing as a target function cannot be attributed in the common understanding.

5.1 A reflection on optimality

What is the optimal society, or an optimal public transportation system? Even if we leave out any subjectivity-related optimality definition, a similar question might be: What is an optimally adapted plant or animal?

If we ponder on those questions we will notice some properties of systems that make them hard to treat in any optimizing framework. The predominant problem with the mentioned systems is their multitude of principally sensible quality measures that all seem more or less arbitrary, both by definition and relative weighting, if some of them are to be considered at the same time. If we look at biology, the concept of an optimally adapted animal has been ever-changing over time, space and context, as the phylogenetic tree [82] and the diversity of the living world tell us quite impressively. Nevertheless there seems to be a kind of directional evolution in the development of a certain species along its historical course. This observation in turn tempts into assuming some kind of teleological positivism [83] by interpreting these developments as directed towards a final state — implying that a targeted melioration or even optimization is feasible. There are objective observations, however, that suggest this assumption not to be realistic.

If we compare different answers of biological evolution to very similar environmental conditions in spatially far away corners of the world we find quite disparate concepts that obviously have proven as well fitting: The rather large number of different marsupial species on the Australian archipelago inhabit habitats very similar to those presently dominated by ordinary (placental) mammals in the rest of the world. There is an interesting *a posteriori* interpretation of the differing developmental traits of those concepts, though [80]: ’The early birth of marsupials removes the developing young much sooner than in placental mammals, and marsupials did not need to develop a complex placenta to protect the young from its mother’s immune system. Early birth places the tiny new-born marsupial at greater risk, but significantly reduces the risks associated with pregnancy, as there is no need to carry a large fetus to full-term in bad seasons. Because a newborn marsupial must climb up to its mother’s nipples, the otherwise minimally developed newborn has front limbs that are much better developed than the rest of its body. This requirement is responsible for the more limited range of locomotory adaptations in marsupials than placentals; marsupials must retain a grasping forepaw and cannot develop it into a hoof, wing, or flipper as some groups of placental mammals have done.’

It may be estimated that quite a number of systems that a human optimizer would like to meliorate fall into a similar category: Interpreting a historical development in terms of an identified melioration effect succeeds quite well, but a prediction of further development, or even the derivation of an optimality point is not viable.

Even if a multitude of potentially important targets is considered in an optimization effort there is another spoilsport aspect to be considered: Targets themselves may appear or disappear during the process of ongoing melioration, when effects neglected at the start of the consideration become decisive in later stages, or limiting boundary conditions eventually dissolve. Let us inspect the century-long development process of a rather technical system — the melioration of power plants.

When the vast expansion of electrical power usage started in the earlier decades of the last century (cite???) the foremost quality of plants was the cycle efficiency and as secondary aspects the total power output of a given design and its adaptability to changing load situations. Due to still lacking large-scale transportation facilities usually near-by primary energy sources, as e.g. local coal mines played an important role. If exhausts were regarded at all it was mostly limited to the pollution by particulates in the nearest neighborhood of the power plant, usually resulting in quite high chimneys to enforce better mixing of exhausts with clean air (cite???). With growing size of the energy market and individual request amounts the cost efficiency became more and more important as the electrical energy entered a steadily growing range of end consumer products (like household aids, electrically operated industrial machines etc.). The growing demand created dependencies on far-away suppliers of primary energy sources that led eventually to the first world-wide "oil crisis´´ in the seventies of the last century — a new competing goal of power station (or at least: power provision) valuation sprang into existence: the provision security.

A relief from this situation was sought by the alternative construction of nuclear energy plants, lessening the dependency of the industrial countries on fossile energy imports. Only regarded from a technical point of view this emerging technology promised large amounts of cheap electrical energy. But at least in various countries the opposition to nuclear power production was under-estimated[1]. The additional expenditures for shielding and securing the required infrastructure against opposing groups, as well as fulfilling politically imposed requirements of strongly raised security assessments, made the price of electrical energy production explode relative to earlier calculations. Had the social and political side effects been foreseen they might have been counteracted on a definitely lower impact level. But as history tells this chance had been missed.

As a short-term effect the return of the fossile energy burning power plants, meanwhile very cleanly operating with respect to particulate emissions, collided with the rather recently emerging concerns of volatile exhausts leading to the emergence of a strong renewable energy movement. While on the long run this method of consumable energy production on the basis of renewables will be the only one proving practicable, even on a mid-range perspective it might lead to raising costs in the immediate future, potentially leading to undesired side effects on economic and social development.

Summarizing this historical development of striving for the best (i.e. cheapest, most productive, stable-operating) power plants we obtain some insights into the emerging optimization process peculiarities. The persecution of a central main objective for certain periods leads to the upcoming of additional goals, not having been regarded before, that grow to play limiting and counteracting roles in the whole process. Even if later-on appearing goals had been taken into account at the start of the long-term development they would not have contributed to a more pointed and less crisis-driven development, as the additional targets would have led to *an inferior outcome at an earlier point in time*. Alternatively proposed solutions, anticipating later objectives, would not have been competitive at earlier points in time. This, finally, leads to the conclusion that the system 'effective electrical energy preparation' is one without a objectively defineable long term target.

There are many more systems and optimization objectives that are comparable in this respect. Here, just one other shall be discussed that is surely familiar to (almost) any reader: the development of easy and efficient individual locomotion over larger distances.

Earth-bound individual locomotion is strongly coupled to the improvement of passenger cars. Due to the technical success of the respective automobile development and

[1] This description is mainly abstracted from the historical development in Germany. Other countries, like France, exhibited significantly different developments.

the present exorbitant use of them (at least compared to the times of early development) traffic jams, too few parking space, fuel costs, and air pollution became severe limiting factors, only counteracted by additional enforced technical and logistic developments. Those changing boundary conditions in turn modify the development paths modern automotive technology is taking. While in earlier times maximum velocity, spacyness and comfort were anticipated future development goals, public attention and customer focus has strongly shifted towards economical operation, crash stability, unobtrusiveness and endurance, with an upcoming perspective on navigation support and automatized, instantaneous avoidance of traffic jam conditions.

While this description primarily holds for passenger cars, an almost opposite direction of development is observed for the two-wheeled versions of automotive devices, the motor bikes. In earlier times they were mainly seen and used as a simple means of individual transportation, with mostly cost-effectiveness, solidity and ease of maintenance being the predominant goals. Contemporary aspects strongly deviate from this view, bringing individuality, enforced demonstration of power etc. to front. Extrapolating views of the past, a typical recent motor bike would have been quite probably attributed as too heavy, too difficult to service, and too expensive in former times.

We may interpret these views in terms of static vs. dynamical structure generation and sustenance. We may regard the goals of a technical system, eventually in combination with an (averaged) subjective view of a potential evaluator, as static as long as we deliberately blind out longer-range perspectives. A power station or a car *may be* optimized with respect of a certain goal, or even a set of goals, if its or their apprehension is regarded as constant. Changes of apprehension frequently manifest themselves in their distribution within a society, quite often with a gradient on individual age, leading to a certain type of ‘generation conflicts´.

So, in an unrestricted view, neither rational, sociological nor technical structures lend themselves to an analysis of optimality. However, they do so in a restricted view. in a limited time horizon as well as in a limited location technical structures can be optimized. We shall discuss the constraints(???) of the optimality algorithms in the next section.

5.2 Constraints in optimality definition

In a partial view, many structures and structure generating processes lend themselves to melioration and even optimization endeavors. Their actual executions are determined by a number of constraints, also frequently referred to as boundary conditions.

Parameter ranges as boundary conditions

Preparing a system for (numerical) optimization, we first model the causal dependencies of one or more target function values upon a set of configuration parameters. These dependencies need not necessarily be direct, sharp, or simple. In most practical cases complex multiple, non-linear and unsharp dependencies are quite common. Configuration parameters are typically *bounded,* i.e. they cannot assume arbitrary values: Modeling the uplift of an airship we have to lower-bound the weight of the lift gas to that of helium, as there is no one lighter than that. Modeling the performance of a gas turbine power station (depending on the properties of its operative parts, like turbines, pumps, heaters etc.) we have to limit the maximum temperatures of the underlying thermodynamical process to the maximum values that the material contact faces of the structural elements can tolerate.

These constraints usually limit the attainable values of the target qualities as well. In many cases it is not obvious, though, what extremal target function values can be obtained. Here the underlying model must be queried, either by direct mathematical evaluation of dependencies, or by numerous samplings of target values depending on respective sets of input variable settings.

The notion of limitations or bounds in the model definition, implying attainable target values, seems rather trivial at first sight. As soon as the potential ranges of configuration variables depend on each other, this grows into a sometimes very complex problem, though. Revisiting the power station example, we may observe that the maximum surface temperature on the turbine blades depends on the amount of additional cooling gas input through fine air ducts inside the blades. Additional cooling gas reduces the maximum output of the turbine again. Furtheron the available cross section for such air ducts is limited by the required material stability for the very fast spinning blades, thus limiting even more design variables. This description is by no means exhaustive, though, but it should suffice to demonstrate the point of mutual dependencies of target function value limiting configuration factors of a system, making it rather difficult to estimate the influence of them on available configuration parameter spans. It directly brings us to the problem of potentially excluding interesting regions of (unbounded) configuration parameter space without the ability to grasp it.

Bounds are not necessarily maximum or minimum limiting conditions. They may as well appear in the interior of an otherwise already limited configuration parameter range. DO WE HAVE A NICE PRACTICAL EXAMPLE FOR THAT? Maybe the non-existence of intermediate density state of water at standard conditions?

Superstructures

As soon as structural variants exist for the realization of a system we enter the domain of discrete optimization. For the matter of argument we only restrict our view to a problem class containing integral as well as continuous configuration variables. A typical representative of such a problem is the chemical engineering process optimization shown in section 4.3, page 28. Besides several continuous variables there are structural alternatives, each of them defining their individual subset of dependent parameters.

A frequently adopted method to manage the space of potential structural alternatives is the definition of a so-called superstructure[2]. It contains all allowed structural variants as binary switching options, expressed in allowed parameter values of 0 and 1 respectively. As with continuous variables, simple decision variants are readily intelligible. But as soon as intermingled and mutually dependent switch settings are required the system may become highly complicated. Again, we note the problem of ascertaining that every desired parameter setting, here with respect to structural alternatives, should be able to be reached. If this is not the case we may unknowingly exclude important ranges of the “true´´ (i.e. pragmatically feasible) configuration space from systematic optimization.

In general, such superstructures occur in combination with continuous variables, thus creating mixed-integer linear or non-linear optimization problems. Especially the latter are known for their algorithmic complexity with respect to rigorous and affirmed location of their respective global optima. They continue to be subject of current research and development.

One of the major problems in such systems is the fact that configuration parameters are related to each other in a hierarchic way: The continuous parameters defining the properties of a chemical engineering plant in some switchable sub-tree (like pressures, temperatures, sizing, material selections) only *exist* if the sub-tree is switched on. Otherwise they are completely meaningless.

[2] The concept of superstructures is much more encompassing than the pure interconnections of structural alternatives if we use it in a more transdisciplinary context, but for the reason of argument we only understand its meaning in the narrower sense usually attributed in the area of chemical engineering.

Open-ended structures

A boundary condition of a system's structure can also be defined implicitly and self-adaptingly. In that case the system's higher-level semantic definition must be interpreted and translated into a functional model amenable for computational evaluation. As an example we may re-inspect the chemical engineering flowsheet as shown in fig. 13 on page 29. The conventional method to model such a system is the super-structure approach discussed in the last paragraph. Here connections of pre-determined apparatuses are fixated, each being defined by its respective set of design variables. Mathematical algorithms may then be applied to identify a best possible setting of the variables, leading to the respective value(s) of target function(s). An alternative semantic definition of the flowsheet may be given by just describing the effect that is desired from a concatenation of various kinds of generalized unit operations. We might say: "The lower-boiling fraction separated in 'Flash 1´ is to be fed into a distillation column (representing y4 $\longrightarrow$ 1)´´ and so forth, of course favourably in a more formalized language structure, but with the same level of (im)preciseness. Layout changes can be expressed relatively easy in such an abstracting language, as on this level of generalization it is, per definitionem, not necessary to define precise configuration parameter settings in details. Instead we rely on computational intelligence for re-adapting them at least to a certain measure of quality. If we restrict ourselves to such system definitions there is ample room for modifying the layout of the interconnection structure, just by changing some elements of our description. We may introduce additional items or take items or whole branches out.

Compared to the more mathematically oriented super-structure approach it is somewhat simpler to assure the definability of all sensible *structural* variants. But as the optimizing effort needs the assistance of computerized hierarchically underlying continuous parameter definition intelligence, we in turn cannot be sure that for all structural variants the complete scope of continuous parameter settings is realized, so the general problem of potentially excluding interesting configuration space ranges remains.

It is very difficult to subject this kind of semantics-based structural redefinition to a rigorous, algorithmic system optimization. In the rather few documented cases this approach was put to work evolutionary concepts were applied. CITATIONS?!?

5.3 The chore of choice: Trading off conflicting goals

Many, if not almost all, melioration problems of the real world exhibit more than one target function. Cars are not only valued for their maximum speed but also for a low comsumption value, a high reliability, low maintenance costs etc. . A delivery service not only needs to deliver the goods in time but also is expected to handle them carefully, to be competitive in its costs, or to obey time windows for pick-up and delivery as close as possible. Usually such sets of target aspects cannot be satisfied optimally for each criterion at the same time. Let us, in this respect, reconsider the passenger car example: Satisfying the desire for high maximum speed will require large and complex engines, as well as elaborated braking equipment, which in turn leads to higher investment and maintenance costs. A part of the raised costs may be caught by applying less solid constructive layouts but will in turn lead to lower reliability. Almost any scenario of constructive change will stress different aspects of the arbitration effort of „optimizing everything at the same time".

Accordingly, the systematic, objective optimization of a given system will come to an end sooner or later, and the domain of the subjective choice sets in: If two or more conflicting goals cannot be resolved at the same time, the individual opinion or valuation is the last resort to choose *the one* solution suggestion that is to be realized in actual utilization. Nevertheless there is a *set of best solutions*, frequently called the „Pareto Set" (fig. 16) of the given problem. It represents those solution suggestions where you can

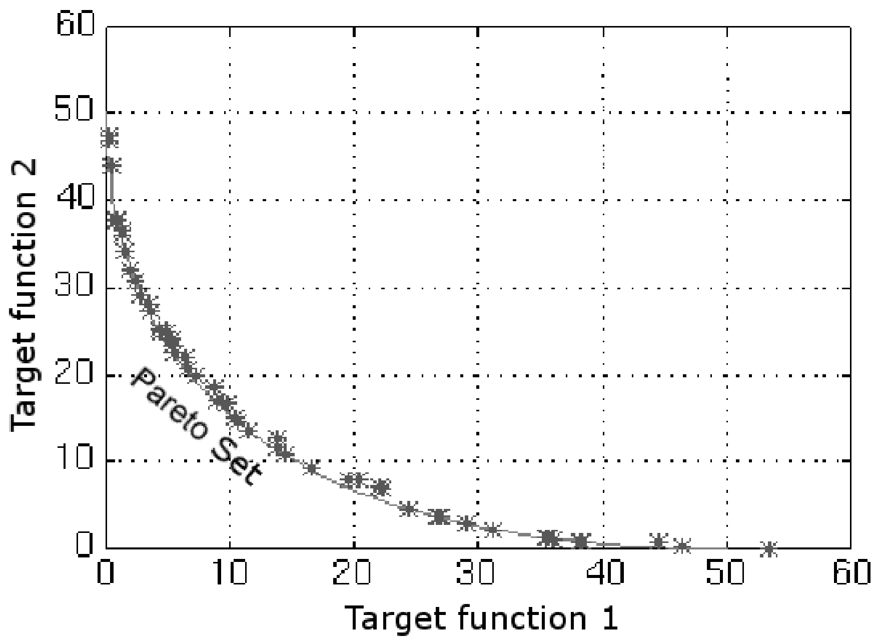

Abb. 16. Graphical representation of a two-dimensional pareto front. Both the actually determined target function combinations (starred positions) and an idealized, continuous front line are given for the non-dominated solution propositions.

improve one target aspect only by impairing at least one other. Mathematically, it creates a mapping

$$\mathbf{x} \,\epsilon\, D \quad \Longrightarrow \quad \mathbf{f}(\mathbf{x})$$

with $\mathbf{x}$ representing the vector of system design variables and $\mathbf{f}(\mathbf{x})$ the vector of dependent individual target function values. This mapping can be discrete (in the case that only distinct configuration parameter settings are feasible) or continuous (if there are arbitrarily adjustable configuration parameters like temperatures or pressures in a technical system). If there is a large number of potential distinct settings the individual frontier points may be represented by an idealizing continuous boundary representation that may be handier for consecuting evaluation purposes. These pareto sets exhibit some properties that shall be discussed further.

If we begin with scrutinizing a single pareto front for a technical system with continuous variables, we frequently observe a qualitative target arbitration behaviour schematically shown in fig.16. As an example we may take a residential heating system, with the inlet temperature into the hot water piping as the governing configuration value to trade heating costs, f_{costs}, vs. lagging of temperature adaptation in the heated rooms, f_{lag}. Both targets should be minimized. Raising the inlet temperature will yield a faster response of the system, but at the same time raised energy losses and hence higher costs and vice versa. The dependency of both target functions on the temperature is, due to the underlying thermodynamic laws, a smooth but counteracting one.

As a certain setting must be chosen, at least for a given point in time, it is the question how the arbitration is individually performed. A classic method is the linear superposition of target function values, defined by a combined quality function Q_{eff} (see fig. 17):

$$Q_{eff} = \frac{\alpha f_{costs} + \beta f_{lag}}{\alpha + \beta} \quad .$$

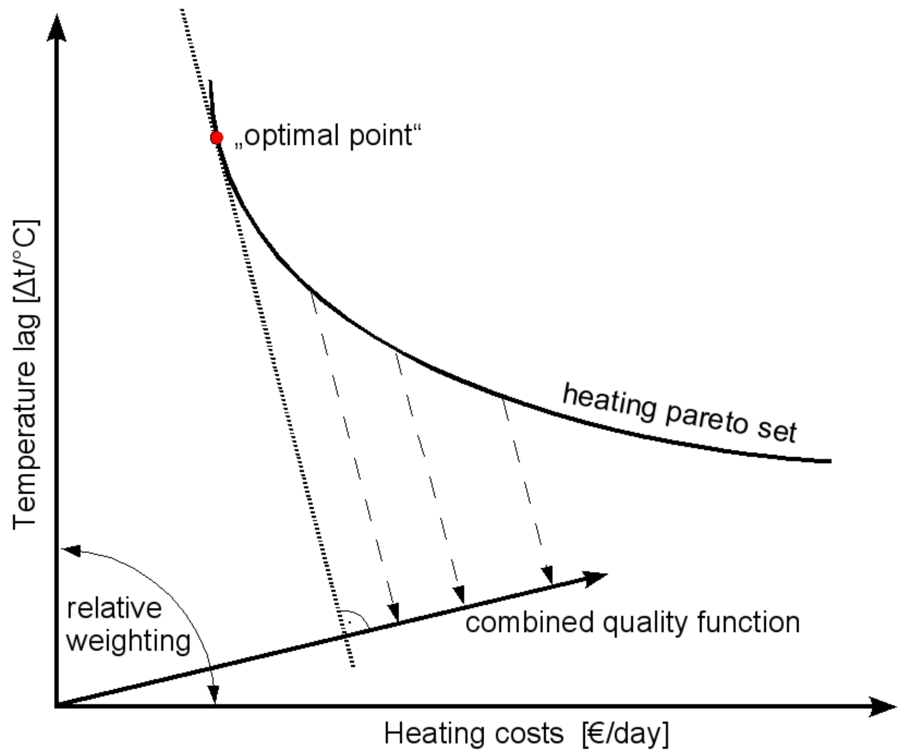

Abb. 17. Classical arbitration of two target functions by linear superposition. The general concept is transferrable to more than two targets. Explanation see text.

Any such aggregation will define a linear superposition of the originally separate functions that may be expressed as a tilted new “optimization axis”, equivalently expressed by arbitration iso-lines perpendicular to that axis. The angle of such iso-lines therefore may as well define the individual arbitration model. If, as an extreme example, almost only operation costs are considered, the arbitration axis would almost coincide with the abscissa of our plot, and iso-lines would be near to parallel to the ordinate. Every single pair of system responses in f_{costs} and f_{lag}, resulting from a respective setting of the piping inlet temperature, is projected onto the new optimization axis, leading eventually to the identification of an ’optimal point’. Changing the subjective focus and hence the relative weighting of the individual goals will result in a shifted optimum and, together with it, in a different temperature setting for the heating system.

If there are discrete configuration parameters in a pragmatically defined optimization problem, up to the point that strongly differing solution concepts compete with each other, the combined pareto front tends to become more complex and exhibits additional features. Let us investigate a rather simple but illustrative, idealized example: If we have to decide which means of transportation to use, we are interested in minimizing two target functions: the duration of the trip, represented by the inverse velocity, and the costs of the trip, represented by the relative price. It is evident that taking a higher velocity transportation method will usually induce raised costs, as the investment, fuel, and maintenance costs of an airplane, for example, exceed that of a bicycle, or the repetitive consultation of a shoemaker’s services.

Mainly depending on the total distance to be traveled, quite different vehicles can be chosen for a certain trip. Only for very small distances the plane is no alternative, and only for the really long distance trips walking and bicycle riding is ruled out. For each vehicle an individual pareto set may be expected: By applying a high technological input (like low friction bearings, aerodynamical spokes, streamlined frames etc.), a respective

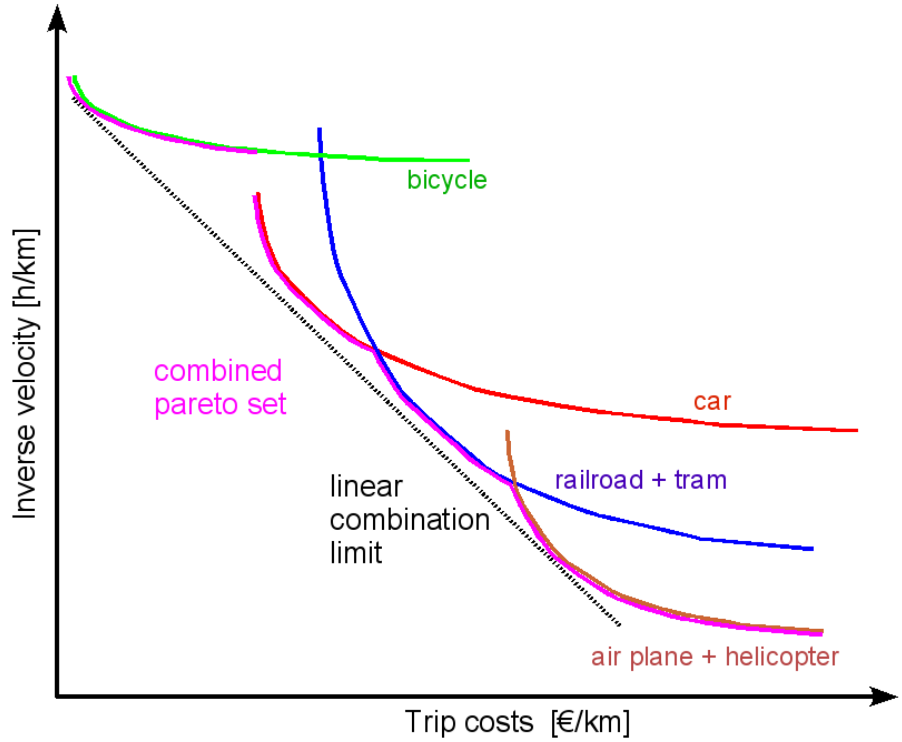

Abb. 18. Creation of a structured, qualitative pareto set of two target functions by superposition of individual sets of structurally different solution approaches. Explanation see text.

bicycle can be made substantially faster compared to cheap ones from the superstore. But even for such high-tech bikes the ultimate limiting condition is the power input by the human, defining the maximum velocity even for highest-priced items. Similar arguments hold for the other vehicles as well, thus creating individual pareto sets of their own, expected to penetrate each other at distinct, but case-dependent, front intersection points.

The resulting pareto set, in the figure shown as the purple-marked border line, exhibits some interesting properties. First, it is as a superposition of the individual pareto front lines not necessarily differentiable, as the frontiers usually intersect with non-zero angles. Second, it does not even need to be steady, as adjacent partial pareto fronts need not touch each other. In our qualitative example, the pareto curve for the car starts already at higer velocities than a bicycle rider will ever reach. So, if at the position of that unsteadiness a small increase of trip costs is accepted there is a leap in transportation velocity by the switch to the other method of transportation. Such leaps can happen in any direction.

If we apply our method of linear superimposing target functions to this plot we will, in this very restricted quality assessment view, arrive at the conclusion to use either plane or bicycle. Tilting the superposition assessment axis from one extreme to the other will leave out any other means of transportation. On the other hand this implies that rather extreme values of costs and velocity are are regarded as equally sufficing at the switching point from one solution alternative to the other. In a practical point of view that is obviously not what we intended to obtain. The solution key is the acceptance of *non-linear* assessment functions: In practice we honor the *simultaneous* fulfillment of target function qualities, as long as it does not lead to a major deterioration of the linearly superimposed solutions. Mathematically this may be represented by defining the combined quality function Q_{eff} with an additional, non-linear term

$$Q_{eff} = \frac{\alpha f_{costs} + \beta f_{lag}}{\alpha + \beta} + \gamma \cdot f_{costs} \cdot f_{lag} \qquad (\gamma < 0 \text{forminimization}),$$

With increasing non-linearity factor γ we honor the simultaneity of advantageous individual target function settings stronger. It is almost impossible, though, to express such subjective, nonlinear preference settings in an objectivized plot. Rather, it is sensible to report the complete (quantative) pareto set calcluated for an optimization problem to an individual or a group of human decision takers. On that basis every involved person can decide how far he or she would compromise on a back-laying, but simultaneously near-advantageous configuration of target values.

Apart from regions of intermingled subsystem pareto fronts, most neighboring pareto set elements originate from neighboring configuration parameter settings. Small changes in target function values are usually caused by equivalently small changes in the system parameter settings, as is obvious in our residential heating example.

If a number of n target functions is considered the dimension of the „pareto front“ is $(n - 1)$. Therefore, presenting pareto sets for two conflicting targets is not too elaborate: One can — in case of discrete system layouts — just either sketch the possible realizable target function combinations for the non-dominated solution suggestions, or present an abstracted 'front line' of those solutions for densely populated sets (fig. 16).

Addressing more than two targets is more difficult: Even if there still is a pseudo three-dimensional representation for three targets, its visualization and interpretation is not so simple. A fast display, freely rotatable in every direction by computer mouse interactions, can help to get a correct notion of the non-dominated front face items for the given problem. Even if an interaction of the decision maker with a computer is necessary for this, the display of the pareto front is basically static. Besides view angle changes no active work of the computer is required.

Proceeding to four and more targets changes this situation significantly: There is no static display possibility for a 'pareto front' with so many dimensions. How is a decision maker then able to adhere to the concept of pareto optimization; how can he get a notion on the pareto elements of a given problem, to be able to trade off the goals according to his or her individual preferences, taking into account the limits of feasibility?

Here the decision maker support requires active work on the computer's side. As the pareto set cannot be displayed as a whole, individual interactions must help investigating it. A first approach might be to set all but two of the targets to fixed values temporarily and see what (partial) pareto front shows up for the two unrestricted values. The drawback of this approach is evident: For n criteria there are $\binom{n}{m} = n!/2! \cdot (n - 2)!$ possibilities of combining them. So four criteria create six two-dimensional views, five criteria ten, and six criteria already 15. In addition this method only creates 'cuts' perpendicular to the variables' axes, so interesting combinations of criteria results are not considered.

In its last resort, multi-criteriality always is a matter of subjective choice: If two or more conflicting goals cannot be resolved at the same time, with every objective criterion being already considered, the individual taste is the only guide to choose *the one* solution suggestion that is to be realized.

Another possibility of identifying relative decision makers' preferences is the pre-emptive request on several (assumed, not necessarily existing) combinations of target values. In choosing between pairs of suggested properties relative weightings may be derived. This is no way to perform the choice itself, though, and the real-life selection may be influenced by other objectives than isolated pair-wise decisions.

5.4 Objective and subjective goals

Show typical cases of subjective goals, like "good soundör comfort". Demonstrate how subjective perception can partially be objectivized, like the Fanger comfort model.

Besides technical optimizations with objective aspects there are others with almost purely subjective goal functions. There are lots of optimization processes in which human perception and senses play a crucial role. Examples are the production of a coffee mixture with a special taste and smell, the development of perfumes, or digestibles with intended smells or tastes.

For illustration, a taste optimization is presented briefly (WEISS following HERDY 2000, partly private communication). How can a coffee mixture with optimal taste be created? As an appropriate methodology an evolution strategy has been identified, working on subjective assessments by comparative taste attributions of testers. There are no objective quantities which can be measured, but coffee testers nevertheless have the task to produce a coffee with an intended taste. This presupposes a sense of taste which is trained in long times of practice.

An evolutionary strategy for optimal taste determination may be set up by the following steps: Five randomly mixed coffee mixtures are given to the coffee testers. The mixtures ratios are not known by the testers. The task is to find out the coffee mixture coming closest to an offered reference mixture. The mixture of the coffee approaching the reference taste best is taken as a basis for five new mixture variations. This looped procedure converges after few follow-ups — the testers cannot differentiate anymore between the quite similarly tasting test mixtures, although there are objectively differing mixing ratios, also with some distance to the goal mixture.

From this example we may abstract some characteristic conditions for optimization procedures with subjective goal functions:

- Optimization goals usually possess a considerable fuzziness.
- An optimization goal is crucially influenced by the group of assessing persons and their subjective notion of quality fulfillment.
- Subjectively shaped goal functions are usually multi-criteria functions (see below).
- Objectivity of goal criterion which is often asked for and aimed at can only be realized in a limited way.

These aspects gain increasing importance for the design of such optimization processes. Beside the treatment of the individual problems themselves there is an important aspect of procedural development in treating subjectivity-biased melioration processes. It is highly desirable to identify more or less objectively determinable goal definitions that might replace the present subjective ones. Some expectedly remunerating examples of presently subjectivity-dominated issues from the field of technical acoustics are:

- Acoustic quality of larger rooms, like concert halls, auditoria, or classrooms,
- adjustment of hearing aids for hear-impaired individuals,
- walking noise emission of modern laminate floors,
- acoustic quality of loudspeakers, musical instruments, bells, etc.,
- sound design as well as sound quality, used as acoustic visiting-card, for special technical products, like automobile doors.

Some subjectively perceived sound qualities are already relatable to objective measurable quantities, although the overall perception of sound as an acoustic quality is very complex in nature. There are influential subjective factors as individual conceptions, hearing habits, etc. Consequently, sound perception cannot simply be described by objectively measurable sound pressure levels (or loudness), pitch (or timbre), and reverberation in a room. They correlate in some cases quite well, but in other cases there are striking discrepancies. As of today's knowledge objective assessments can replace subjective assessments only in exceptional cases. However, subjectively superior acoustic qualities often coincide with certain settings of objectively measurable properties.

Two facets of subjectivity play important roles in this respect. First, human beings do not perceive sound events in complex situations as a separable input. Instead, acoustic perceptions are embedded into non-acoustic environmental influences: The acoustic perceptions in an concert hall will also be dependent on optical perceptions, on architectural aesthetic effects, the expectations in acoustic performance, the comfort of the taken position (climatic conditions, disturbing influences, seat comfort), and on the psychosocial condition of the listener (missing acceptance of a musical piece, personal physical condition). Second, the assessment is influenced by the group of test listeners and by its type. The assessment changes when performed by a group of laymen or experts, trained or untrained listeners, active or passive participants (listeners to music or musicians).

Insights from the field of acoustics may be transferred to other subjectivity-influenced domains, like textual relevance attribution or climatic comfort issues. If we try, for example, to identify similar texts in a larger collection of writings, the *practical* attribution of similarity is strongly dependent on the relevance attribution of included passages as set by the assessor. Even if larger parts of two compared texts diverge, individual sections with subjectively decisive keywords found in both may lead to the conclusion that they are closely related. In the area of text comparison there is a chance of objectification by applying a frequency-of-incidence measure on individual words. But as ultima ratio only the appraising subject can judge the similarity.

The climatic comfort constitutes a more objective, although not sharply defined subjective target function, e.g. in the assessment of buildings. While the individual perception of coziness may differ substantially in a given room there are at least well-tested statistical comfort models, taking into account different dimensions like environment temperature, radiation temperature imbalances of surrounding walls, environmental humidity, airflow around the human body, and some other influencing factors. Even though the resulting comfort model does not represent the perception of a certain single individual in a room, it very well describes the probability density distribution of the comfort attribution of a larger number of people. So, if a building is not designed for a certain small number of nameable persons, but to serve an initially not known number of anonymous inhabitants like office workers, these quasi-objective quality measures may well serve the needs of building designers in their effort of trading off different target qualities like costs, comfort, flexibility and such.

6 How to find the needle in the haystack: Practical Methods for Multicriterial Optimization

Rigorous mathematics (Grötschel/Küfer) Explain typical methods of rigorous multicriterial(!) optimization, together with demands in model structure upon which the optimization algorithm works. Show typical limitations of algorithms.

6.1 Challenges for Contemporary Evolutionary Algorithms (Bartz-Beielstein, Preuss, and Schwefel)

Does one need more than one optimization method? Or, stated differently, is there an optimal optimization method? Following from the No Free Lunch theorem (NFL, [85]), in the general case —without clearly specified task— there is not. For every single task, creating a specialized method would be advantageous. Alas, this requires (i) a lot of effort, and (ii) extensive knowledge about the treated problem, and is thus rarely done. Alternatively, two strategies are usually followed when tackling a 'new' optimization problem:

- Adapt an existing algorithm to the problem in its current form, and/or
- model/formulate the problem appropriately for an existing algorithm.

Whereas 'traditional' mathematical optimization approaches mostly favor the second approach, it may provoke unwanted side-effects: One has to make sure that the most important features of the original problem are taken over into the model. E.g., matching the problem to an existing algorithm may obscure its original global optimizer(s) or good local ones so that they become unreachable for the optimization algorithm. Besides, many algorithms require the problem to fulfill properties it obviously or possibly does not, e.g. continuity and differentiability. Particularly, in cases where computing the quality value of one solution candidate requires running a complex simulation software, one seldomly knows in advance which properties the underlying (unknown) objective function possesses.

When nothing more than quality determining response values for any set of input variables are known for a problem, we speak of *black box* optimization. In the single-objective case, the common notion of an objective function and its global optimum/global optimizers —as given in equation 1 for unconstrained problems— is still useful. However, global optimizers, the set of input vectors **x** for which f(**x**) gets minimal, cannot be determined analytically; empirics is the only way to find them.

$$f^{*G} = \min\{f(\mathbf{x})|\mathbf{x} \in X\} \tag{1}$$

The black box concept immediately leads to *direct search* methods — such a method only utilizes objective function responses and "does not 'in its heart' develop an approximate gradient", as [86] puts it. As far back as in the 1960s, many direct search methods have been invented, e.g. the famous *Nelder-Mead simplex algorithm* [52]. At the same time, the first steps into the world of *evolutionary computation* (EC) happened, presenting very simple versions of what is now put under the unified denotation *evolutionary algorithms* (EA). These do not only use bio-inspired heuristics, they also employ randomness. Even nowadays, people may be alienated by the extensive use of random numbers in and the fragmentary theory supporting EAs. Nevertheless, in spite of these issues perceived as problematic, these optimization methods have demonstrated their ability in now numerous real-world applications and are thus far from outdated.

Interestingly, in recent years, the mathematical optimization community has again shown increased interest in direct search methods [39]. This may have to do with (i) the fact that these techniques simply did not go extinct on the practitioners side, and (ii) improved theoretical analysis methods that now help tackling heuristic algorithms. In computer science, the growing field of *randomized algorithms* is exclusively dealing with algorithms employing random numbers — not only in optimization. [50] give an overview.

This section targets at introducing the main EA concepts and highlight specialized techniques for three important application areas: Multiobjective optimization, optimization in presence of noise, and multimodal optimization. Additionally, we demonstrate these techniques by utilizing them on a simplified real-world test case. However, we start with a glimpse on the history of EAs, and a concrete description of how such an optimization method works.

Historical roots

Although there have been precursors in proposing utilization of evolutionary concepts for optimization tasks, as e.g. [16] (also see the fossil record, [29]), invention and development of the first evolutionary algorithms is nowadays attributed to a handful of pioneers who independently suggested three different approaches.

- Fogel, Owens, and Walsh introduced evolutionary programming (EP) [30], at first tasked at evolving finite automata, later on modified into a numerical optimization method.
- Genetic algorithms (GA), as laid out by [36], first focused on combinatorial problems and consequentially started with binary strings, inspired by the genetic code found in natural life.
- Evolution strategies (ES) as brought up by [56] and [69] began with solving engineering problems and thus mostly used a real-valued representation.

In the early 1990s, a fourth branch of evolutionary algorithms emerged, explicitly performing optimization of programs: Genetic programming (GP), suggested by [40]. Since about the same time, these four techniques are collectively referred to as evolutionary algorithms, building the core of the evolutionary computation (EC) field.

What is an evolutionary algorithm?

Today, there is little doubt about components and general structure of an EA. It is understood as population based stochastic direct search algorithm —not excluding population sizes of one as e.g. featured in simple evolution strategies— that in some sense mimics the natural evolution.

Besides initialization and termination as necessary constituents of every algorithm, EAs consist of three important factors: A number of search operators, an imposed control flow (Figure 19), and a representation that maps adequate variables to implementable solution candidates.

Whether different EAs may put different emphasis on the search operators mutation and recombination, their general effects are not in question. Mutation means neighborhood based movement in search space that includes the exploration of the 'outer space' currently not covered by a population, whereas recombination rearranges existing information and so focuses on the 'inner space'. Selection is meant to introduce a bias towards better fitness values; GAs do so by regulating the crossover via mating selection, ESs utilize the environmental selection.

A concrete EA may contain specific mutation, recombination, or selection operators, or call them only with a certain probability, but the control flow is usually left unchanged. Each of the consecutive cycles is termed a *generation*. Concerning the representation, it should be noted that most empiric studies are based on canonical forms as binary strings or real-valued vectors, whereas many real-world applications require specialized, problem dependent ones.

For an in-depth coverage on the defining components of an EA and their connection to natural evolution, see [28] and [27].

Basic ES variants

In the following, we introduce the most important canonical variants of evolution strategies for single objective optimization, to serve as basis for more specialized algorithms later on.

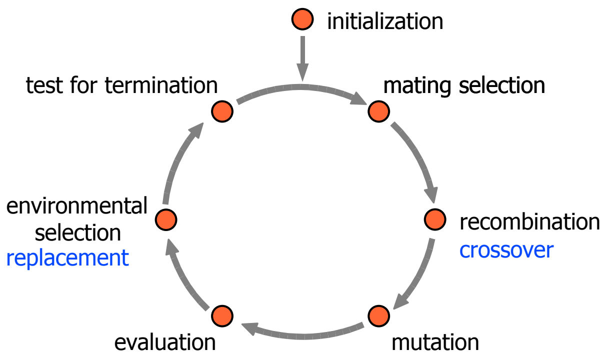

Abb. 19. The evolutionary cycle, basic working scheme of all EAs. Terms common for describing evolution strategies are used, alternative terms are added below in blue.

The first ES, the so-called $(1+1)$-ES or two membered evolution strategy, uses one parent and one offspring only. Two rules have been applied to these candidate solutions:

1. Apply small, random changes to all variables simultaneously.
2. If the offspring solution is better (has a better function value) than the parent, take it as the new parent, otherwise retain the parent.

[66] describes this algorithm as "the minimal concept for an imitation of organic evolution". The $(1+1)$-ES (Figure 20) is applied by many optimization practitioners to their optimization problem and included in our analysis for three reasons: (i) It is easy to implement, (ii) it requires only few exogenous parameters, and (iii) it defines a standard for comparisons.

Procedure $(1+1)$-ES.

Initialization: Initialize the iteration counter: $t = 1$. Determine: (i) a point $X_1^{(t)}$ with associated position vector $x_1^{(t)} \in \mathbb{R}^d$ and (ii) a standard deviation $\sigma^{(t)}$. Determine the function value $y_1 = f(x_1^{(t)})$.

`while` some stopping criterion is not fulfilled `do`

Mutation: Generate a new point $X_2^{(t)}$ with associated position vector $x_2^{(t)}$ as follows:

$$x_2^{(t)} = x_1^{(t)} + z, \tag{2}$$

where z is a d-dimensional vector. Each component of z is the realization of a normal random variable Z with mean zero and standard deviation $\sigma^{(t)}$.

Evaluation: Determine the function value $y_2 = f(x_2^{(t)})$.

Selection: Accept $X_2^{(t)}$ as $X_1^{(t+1)}$ if

$$y_2 < y_1, \tag{3}$$

otherwise retain $X_1^{(t)}$ as $X_1^{(t+1)}$. Increment t.

Adaptation:

$$\text{Update} \quad \sigma^{(t)}. \tag{4}$$

`done.`

Abb. 20. The two membered evolution strategy or $(1+1)$-ES for real-valued search spaces. The symbol f denotes an objective function $f : \mathbb{R}^d \to \mathbb{R}$ to be minimized. The standard deviation σ will be referred to as *step-width* or *mutation strength*.

The first $(1+1)$-ES used binomially distributed mutations [64]. These have been replaced by continuous variables and gaussian mutations, which enable the $(1+1)$-ES to generate larger mutations and thereby possibly escape from local optima [55]. The factors (parameters) of the $(1+1)$-ES are summarized in Table 1.

Tabelle 1. Factors of the two membered evolution strategy. Based on the default values, the step size σ is multiplied by 0.85, if the success rate is larger than $1/s_r = 1/5$ or equivalently, if more than 20 out of 100 mutations have been successful.

Symbol	Factor	Range	Default
s_n	adaptation interval	$\mathbb{N}$	100
s_r	$1/$success rate	$\mathbb{R}_+$	5
s_a	step size adjustment factor	$\mathbb{R}_+$	0.85
$\sigma^{(0)}$	starting value of the step size σ	$\mathbb{R}_+$	1
$s_{1/5}$	step size update rule	{intv, cont }	cont

Rechenberg introduced the first multimembered ES, the so-called $(\mu+1)$-ES. It uses μ parents and one offspring and is referred to as the steady-state ES. Schwefel introduced the $(\mu+\lambda)$-ES, in which $\lambda \geq 1$ candidate solutions are created each generation, and the best μ out of all $\mu+\lambda$ individuals survive, and the (μ, λ)-ES, in which the parents are forgotten and only the μ best out of λ candidate solutions survive. A birth surplus is necessary for the (μ, λ)-ES, that is $\lambda > \mu$. [68] and [14] provide a comprehensive introduction to evolution strategies.

Selection in ES

Selection should direct the evolutionary search to promising regions. In ES, only candidate solutions with good function values are allowed to reproduce. The mating selection process is deterministic in contrast to the random processes used in genetic algorithms (cf. Section **??**). This selection scheme is known as *truncation* or *breeding selection* in biology.

The κ-selection scheme takes the age of the candidate solutions into account: only candidate solutions that are younger than κ are allowed to reproduce.

For $\kappa = 1$ this selection method is referred to as *comma-selection*: only offspring individuals can reproduce. Consider the following situations:

- If $\mu = \lambda$, then all offspring are selected and the population performs a random walk, because the selection process provides no information to guide the search.
- A necessary condition of convergence towards an optimal solution is $\mu < \lambda$.

The κ-selection is referred to as *plus-selection* for $\kappa = \infty$: both the offspring and the parents belong to the mating pool. The plus-selection is an elitist selection scheme, because it guarantees the survival of the best individual found so far. *Steady-state*-ES that use μ parents and one offspring are used in asynchronous parallel systems.

Variation in ES

[14] propose some guidelines derived from the philosophy of Darwinian evolution to design variation operators.

1. A *state* comprises a set of object and strategy parameter values ($x^{(t)}$, $s^{(t)}$). *Reachability* demands that any state can be reached within a finite number of iterations. This feature is necessary to prove (theoretically) global convergence.
2. Variation operators (mutation and recombination) should not introduce any bias, e.g. by considering only good candidate solutions. Variation operators are designed to *explore* the search space in contrast to selection operators that exploit the gathered information. Table 2 summarizes the features of the ES operators.

3. *Scalability* is the third criterion that should be fulfilled by variation operators in ES: Small changes of the representation should cause small changes in the function values.

Tabelle 2. Guidelines from the philosophy of Darwinian evolution in ES.

Operator	Creation	Information
variation	stochastically	unbiased
selection	deterministically	biased (use function values)

The standard ES recombination operators produce from a family of ρ parent individuals one offspring. Note, that the standard GA crossover operator uses two parents to create two offspring.

Consider a set of parental vectors representing either object or strategy parameters:

$$\{(x_{11},\dots,x_{1d}),(x_{21},\dots,x_{2d}),\dots,(x_{\mu 1},\dots,x_{\mu d})\}. \tag{5}$$

Two recombination schemes are commonly used in ES. Both use a set $\mathcal{R} = \{r_1, r_2, \dots, r_\rho\}$, that represents the indices of the mating partners. It is constructed by randomly (uniformly) chosing ρ values (without replacement) from the set $\{1, 2, \dots, d\}$.

1. *Discrete recombination* can be described as follows. The recombined vector is determined as
$$x_{\text{discrete}} = (x_{u_1 1}, x_{u_2 2}, \dots, x_{u_d d}\}, \tag{6}$$
where u_i is randomly (uniformly) chosen from $\mathcal{R}$.
2. To implement *intermediate recombination,* Equation 6 has to be modified as follows:
$$x_{\text{intermediate}} = \frac{1}{\rho}\left(\sum_{i=1}^{\rho} x_{r_i 1}, \sum_{i=1}^{\rho} x_{r_i 2} \dots, \sum_{i=1}^{\rho} x_{r_i d}\right). \tag{7}$$

Mutation —which can be considered as the basic variation operator in evolution strategies—is applied to the recombined solution, cf. Figure **??**. Mutation in ES is a self-adaptive process, that relies on the individual coupling of endogenous strategy parameters with object parameters. After being recombined as described above, the strategy parameters (standard deviations, step width, or mutation strength) are adapted. These adapted values are used in a second step to mutate the object parameters.

We consider algorithms with one step width σ first. To prevent negative standard deviations, the adaptation of the step width must be multiplicatively. [14] discuss an additional argument for a multiplicative mutation of the mutation strength on the sphere model f_2, see Equation **??**. It can be shown, that in expectation σ should be changed by a constant factor. Therefore, the mutation operator can be implemented as

$$\sigma^{(t+1)} = \sigma^{(t)} \cdot \exp(\tau z), \tag{8}$$

where z is a realization of the $\mathcal{N}(0,1)$ distributed random variable Z. The parameter τ is the so-called *learning rate*. The object variables are mutated next:

$$x^{(t+1)} = x^{(t)} + w, \tag{9}$$

where w is a realization of the $\mathcal{N}(0,\sigma^{(t+1)})$ distributed random variable W.

Taking the logarithm on both sides of Equation 8 results in

$$\log(\sigma^{(t+1)}) = \log(\sigma^{(t)}) + \tau z. \tag{10}$$

Hence, mutations of the strategy (on a logarithmic scale) and object parameters are structural similar.

This multiplicative mutation scheme for one σ can be extended to several strategy parameters $\sigma = (\sigma_1, \ldots, \sigma_d)$. **(author?)** [65] proposes the following extended log-normal rule:

$$\sigma^{(t+1)} = \exp(\tau_0 z_0) \cdot \left(\sigma_1^{(t)} \exp(\tau z_1), \ldots, \sigma_d^{(t)} \exp(\tau z_d)\right), \tag{11}$$

where z_i are realizations of the $\mathcal{N}(0,1)$ distributed random variables Z_i, $i = 1, \ldots, d$. This mutation scheme uses a global and a local learning parameter τ_0 and τ respectively. The whole vector is scaled by the random factor $\exp(\tau_0 z_0)$ after each component has been mutated.

Algorithm Designs for Evolutionary Algorithms

We consider the parameters or control variables from Table 3. This table shows typical parameter settings. **(author?)** [5] presents a kind of default hierarchy that includes four parameterizations for simple and complex algorithms and suggests to perform experiments. Hence, our approach can be seen as an extension of Bäck's methods. The reader is referred to **(author?)** [8] for a detailed description of these parameters.

Tabelle 3. Default settings of exogenous parameters of a "standard" evolution strategy. Source: **(author?)** [5]. Bäck does not recommend to use this "standard" without reflection. Problems may occur, when these "standards" are blindly adopted and not adjusted to the specific optimization problem.

Symbol	Parameter	Range	Default
μ	Number of parent individuals	$\mathbb{N}$	15
$\nu = \lambda/\mu$	Offspring-parent ratio	$\mathbb{R}_+$	7
$\sigma_i^{(0)}$	Initial standard deviations	$\mathbb{R}_+$	3
n_σ	Number of standard deviations. d denotes the problem dimension	$\{1, d\}$	1
c_τ	Multiplier for individual and global mutation parameters	$\mathbb{R}_+$	1
ρ	Mixing number	$\{1, \mu\}$	2
r_x	Recombination operator for object variables	$\{i, d\}$	d (discrete)
r_σ	Recombination operator for strategy variables	$\{i, d\}$	i (intermediary)
κ	Maximum age	$\mathbb{R}_+$	1

Ways to Cope with Uncertainty

In many real-world optimization problems, function values can only be estimated but not determined exactly. Falsely calibrated measurement instruments, inexact scales, scale reading errors, etc. are typical sources for measurement errors. If the function of interest is the output from stochastic simulations, then the measurements may be exact, but some of the model output variables are random variables.

Computer simulations are a suitable means to optimize many actual real-world problems. We will concentrate on stochastic simulation models in the following. The reader is referred to [59] for deterministic models.

Consider e.g. a sequence of traffic signals along a certain route or elevators' movements in high-rise buildings. *Optimization via simulation* subsumes all problems in which the performance of the system is determined by running a computer simulation. As the result of a simulation run is a random variable, we cannot optimize the actual value of the simulation output, or a singular performance of the system Y. One goal of optimization via simulation is to optimize the expected performance

$E[Y(x_1, x_2, \ldots, x_n)]$, where the x_i's denote the controllable input variables [70, 4, 6]. The stochastic nature of the simulation output forces the optimization practitioner to apply different methods than in the deterministic counterparts.

The stochastic output in optimization via simulation complicates the selection process in direct search methods. The efficiency of the evaluation and selection method is a crucial point, since the search algorithm may not be able to make much progress if the selection procedure requires many function evaluations.

From our point of view the following case is fundamental for the selection procedure in noisy environments:

> Reject or accept a new candidate, while the available information is uncertain. Thus, two errors may occur: An α error as the probability of accepting a worse candidate due to noise and a β error, as the error probability of rejecting a better candidate.

In the context of selection and decision making the terms "candidate" and "point" will be used synonymously. A well established context where these error probabilities are analyzed is hypothesis testing.

We concentrate our investigations on the selection process when the function values are disturbed by additive noise, cf. Equation **??** on p. **??**. Noise that affects the object variables is not subject of our investigations.

Noise Model

Our analysis is based on the following statistical assumptions. Let $\{Y_{ij}\}$, $1 \leq i \leq r$, $1 \leq j \leq k$, denote r independent random samples of observations, taken from $k \geq 2$ candidates. The Y_{ij} can denote function values taken from candidate solutions $X_1, \ldots, X_k$ or individuals (particles) of some evolutionary algorithm. Candidate X_i has a (fitness) function value with unknown mean μ_i and common unknown variance $\sigma^2_{\epsilon,i} = \sigma^2_\epsilon$, $1 \leq i \leq k$. The *ordered means* are denoted by

$$\mu_{[1]} \leq \mu_{[2]} \leq \ldots \leq \mu_{[k]}, \tag{12}$$

where $\mu_{[1]}$ denotes the mean of the best candidate (minimization). Generally, normal response experiments are considered.

Noise makes it difficult to compare different solutions and select the better ones. Noise affects the selection process in evolutionary algorithms: In every iteration, the μ best out of λ candidate solutions have to be determined.

Wrong decisions can cause a *stagnation* of the search process: Over-valuated candidates—solutions that are only seemingly better—build a barrier around the optimum and prevent convergence. The function value at this barrier will be referred to as the *stagnation level*. Or, even worse, the search process can be *misguided:* The selection of seemingly good candidates moves the search away from the optimum. This phenomenon occurs if the noise level is high and the probability of a correct selection is very small. How strongly the noise affects the overall performance of evolutionary algorithms is not clear. For example for evolution strategies, it has been shown that increasing the population size may help the algorithm to cope with the noise [13]. A comprehensive overview of various evolutionary algorithm variants in the presence of noise is given in [37].

Common Techniques

One may attempt to reduce the effect of noise explicitly. The simplest way to do so is to sample a solution's function value n times, and use the average as estimate for the true expected function value. This reduces the standard deviation of the noise by a factor of $\sqrt{n}$, while increasing the running time by a factor of n.

Further means used by evolutionary algorithms to cope with noise are averaging techniques based on statistical tests, local regression methods for function value estimation, or methods to vary the population size [72, 12, 58, 3, 15, 10].

A Taxonomy of Selection Methods

As introduced above, noise affects the selection process of evolutionary algorithms. In the following, a comprehensive taxonomy of elementary selection methods is developed.

Depending on the prior knowledge, selection schemes can be classified according to the following criteria:

Threshold: subset selection – indifference zone.
Termination: single stage – multi stage (sequential).
Sample size: open procedures – closed procedures.
Variances: known – unknown, equal – unequal.

The goal of subset selection is the identification of a subset containing the best candidate. It is related to screening procedures. *Subset selection* is used when analyzing results, whereas the *indifference zone* (IZ) approach is used when designing experiments. The sample size r is known in subset selection approaches, it is determined prior to the experiments in the indifference zone approaches.

Single stage procedures can be distinguished from *multi stage* procedures. The terms "multi stage" and "sequential" will be used synonymously. The latter can use *elimination*: If inferior solutions are detected, they are eliminated immediately. Selection procedures are *closed*, if prior to experimentation an upper bound is placed on the number of observations to be taken from each candidate. Otherwise, they are *open*. Furthermore, it is important to know whether the variance is common or known.

Threshold Selection

Threshold rejection (TR) and *threshold acceptance* (TA) are complementary strategies. Threshold rejection is a selection method for evolutionary algorithms, that accepts new candidates if their noisy function values are significantly better than the value of the other candidates [42]. "Significant" is equivalent to "by at least a margin of τ". Threshold acceptance accepts a new candidate even if its noisy function value is worse. The term *threshold selection* (TS) subsumes both selection strategies.
The basic idea of threshold selection is relatively simple and already known in other contexts:

- [43] introduced a threshold operator (with some errors, see [23]) for a (1+1)-evolution strategy and objective functions without noise.
- [74] proposed a threshold strategy that accepts only random changes that result in a specified minimum improvement in the function value.
- [25] presented a threshold acceptance algorithm.
- [51] stated that a similar principle, the so-called truncation selection, is very important in plant and animal breeding: "Only individuals with phenotypic value at least as great as some number c are permitted to reproduce." Truncation selection is important for breeders, but it is unlikely to occur in natural populations.

Threshold selection is also related to Fredkin's paradox: "The more equally attractive two alternatives seem, the harder it can be to choose between them—no matter that, to the same degree, the choice can only matter less"[48]. Regarding the distinction between rules of inductive behavior and learning rules discussed in **(author?)** [9], TS as presented here is an automatic test rule and belongs to the former type of rules.

Procedure Threshold Selection

1. Given: A candidate X_1 with a related sample Y_{1j} of r observations and sample mean $\overline{y}_1 = \sum_{j=1}^{r} y_{1j}/r$.
2. Take a random sample of r observations Y_{2j}, $1 \leq j \leq r$, in a single stage from a new candidate X_2.
3. Calculate the sample mean $\overline{y}_2 = \sum_{j=1}^{r} y_{2j}/r$.
4. Select the new candidate X_2 if and only if

$$TR : \overline{y}_2 + \tau < \overline{y}_1, \text{ with } \tau \geq 0 \tag{13}$$

or

$$TA : \overline{y}_2 + \tau < \overline{y}_1, \text{ with } \tau \leq 0. \tag{14}$$

Abb. 21. Threshold selection. This basic procedure can be implemented in many optimization algorithms, for example evolution strategies or particle swarm optimization.

The Threshold Selection Procedure

The experimental goal is to select the candidate associated with the smallest mean $\mu_{[1]}$. Figure 21 shows the threshold selection algorithm. As can be seen from Equation 13, threshold rejection increases the chance of rejecting a worse candidate at the expense of accepting a good candidate. It might be adequate if there is a very small probability of generating a good candidate. Equation 14 reveals that threshold acceptance increases the chance of accepting a good candidate at the risk of failing to reject worse candidates.

Threshold Selection and Hypothesis Testing

The calculation of a threshold value for the TR scheme can be interpreted in the context of classical hypothesis testing as the determination of a critical point [11]. The critical point $c_{1-\alpha}$ for a hypothesis test is a threshold to which one compares the value of the test statistic in a sample. It specifies the critical region CR and can be used to determine whether or not the null hypothesis is rejected. We are seeking a value $c_{1-\alpha}$, so that

$$Pr\{S > c_{1-\alpha} \quad |H \text{ true } \} \leq \alpha, \tag{15}$$

where S denotes the test statistic and the null hypothesis H reads: “There is no difference in means.” The threshold acceptance selection method can be interpreted in a similar manner.

Generally, hypothesis testing interpreted as an automatic rule as introduced in **(author?)** [9] considers two-decision problems in which a null hypothesis is either accepted or rejected. A false null hypothesis H can be rejected 50 % of the time by simply tossing a coin. Every time that heads comes up, H is rejected. The rejection procedures considered so far can be applied to k-decision problems. Here, larger sample sizes are required than for the two-decision problem. The probability of a correct selection for $k > 2$ is smaller than 50 % if the decision is based on the roll of a fair k-sided die. To avoid too large sample sizes r for fixed k the indifference zone δ^* can be increased or the probability of a correct selection P^* can be reduced. To avoid too large sample sizes r for fixed k the indifference zone δ can be increased or the probability of a correct selection $Past$ can be reduced.

Multiple Objectives

beginnings (see Zitzler Tutorial paper)

why EA for this task

alternatives (e.g. aggregation)

new algorithms: SPEA2, NSGA2, IBEA, SMS-EMOA

difficulties and currently main field of work: estimating quality of an approximated pareto set, s-metric

Multimodal Problems

Although, during the last three decades, many empirical and most of the theoretical studies have been devoted to simple test problems with only one extremal point, the great majority of practical applications requires optimization in far more complex landscapes. Multimodality —the presence of more than one locally optimal point— requires a shift from a hill-climbing oriented towards a global perspective.

local optimality condition (Toern/Addis) difficulties in deciding local optimality (uncountably infinite neighborhood)

similarities to constraint handling: constrained transformed to multimodal problems by adding penalty function

blind / information acquisition techniques (remove the mist by collecting information about the problem topology)

different sorts of crowding, island models with location control, cea, restart techniques, and clustering

Conclusions

Methods of Bounded rationality (Gigerenzer) Describe methodology. Differentiate to Mathematics and EA. Explain areas of application for BR methods. Differentiate towards typical numerical methods

Fuzzy decisions (Hüllermeier)

3

Case studies (each 10-15 pages => ca. 175 pages)

Every contributed case study should address the formerly already listed questions (see enclosed list). Each author should try to integrate (most of) those aspects in his or her contribution in order to create link pointsfor the methodological comparison discussion in the entailing chapter.

7 Production Engineering: The conflicting facets of injection molding tools

J. Mehnen

The layout of mold temperature control designs is decisive for efficient die casting. Die casting and injection casting are typical technologies in mass production. Most plastic components from micro switches to bumpers of cars are manufactured via injection molding. Metallic structures such as aluminum gear boxes or stairs of escalators are produced via die casting. All these products are generated by injecting a hot material into a die. In zinc die casting the injected metal has a pouring temperature of about $435°C$ and the puring temperature of aluminum varies from $620°C$ (thick walled) to $730°C$ (thin walled) [35]. Aluminum and zinc are examples of the most prominent metallic casting materials. The material cools down, solidificates, and the workpiece is ejected from the die before a new process cycle can start. The efficient cooling of a die is one of the main factors to reduce the cycle time and, therefore, to increase the cost effectiveness of a tool. Improvements of up to 50 percent [78] are possible, because, today, the layout of the bores in casting tools is typically done manually. The layout design is based on expert knowledge and rules of thumb. Experts have a good estimate about where to position cooling bores properly. Heuristic rules are quite common in practice. These rules are fast and work very efficiently. Unfortunately, only little of this knowledge is available in the form of explicit mathematical rules [88]. Solutions for complex problems may not be satisfying. Problems with blowholes or insufficient surface qualities need remanufacturing of the die. This process can be very expensive and time consuming. The manufacturing of many dies is not reasonable because each die can be very expensive. Die casting tools cost up to five hundred thousand Euro. Therefore, well designed cooling circuits help to reduce manufacturing costs of the die, minimize cycle times, increase tool life and improve the workpiece quality.

Optimization and particularly algorithmic optimization demand for quantitative measures of the quality of a solution. Therefore, the experts knowledge together with physical properties have to be modeled. The modeling of a problem leads to the exact definition of quality criteria, the choice for efficient problem spaces, data structures, and efficient quality criteria evaluation methods. Due to the multiobjective character and the high complexity of the problem, powerful optimization algorithms have been used. Very useful techniques for difficult problems are evolutionary algorithms.

These algorithms are able to find surprisingly good solutions by clever analysis of the quality functions. Sometimes even experts are surprised about the new solutions found and get motivated to use new approches. There are several advantages of evolutionary algorithms. The algorithms do not have to be changed when the fitness functions are varied or changed. Many classic optimization algorithms – this is also true for evolutionary algorithms – are only able to solve single objective problems. Although often the single objective case is complex enough, many real world problems have a multiobjective character. Problems may have contradictory aims, i.e. one objective value cannot be improved without deteriorating an other. The reduction of all criteria to only one value is possible only in special cases. In the case of uncertainty about the preferences of a user toward one or an other criterion, a set of best compromise solutions should be presented.

Multiobjective evolutionary algorithms can be used either with a posteriori or with a priori techniques. A posteriori approaches follow the idea of uncertainty about a perfect solution. Their goal is to generate best compromise solutions, so called Pareto optimal solution sets, from which an expert can choose according to non-formalized ad hoc knowledge. In case of well defined multiobjective goals sometimes also scalar evolutionary algorithms are useful. Several multiobjective evolutionary algorithms are mentioned in literature. A very good evolutionary scalar optimizer is the Evolution

Strategy (ES) [67]. It can be used for multiobjective problems technique via aggregation. State of the art a posteriori multiobjective optimizers are the NSGA II [19, 20] and the SPEA 2 [87]. All three techniques have been tested on the mold temperature design problem.

Another important issue when evolutionary algorithms are applied to real-world applications is computational speed. Stochastic algorithms tend to need a lot of fitness function evaluations. Especially for difficult problems with high dimensions, restrictions, a difficult multimodal character etc. lead to high population sizes or long runs of the stochastic search algorithm. Therefore, an efficient design of the evaluation functions is necessary. The definition of efficient quality functions is particularly needed in the case of mold temperature control design. Surrogate functions or well motivated simplification of the problem can lead to a significant speed up of the evaluations without much loss in quality.

Finally a systematic search for good parameter adjustments of the multiobjective search methods is necessary to improve the search results. The best algorithmic parameter settings for various problems often differ from the standard. In order to compare multiobjective optimization algorithms that generate Pareto fronts, it is common to introduce scalarization methods, so called metrics. A systematic approach to meta-optimize the parameter settings experimentally is to apply statistical design methods. The statistical design helps finding robust, statistically sound and optimized parameter adjustments. Here, existing classical and new statistical approaches have been applied to multiobjective as well as to single criterion optimization of the mold temperature design problem.

Modeling Aspects

Geometric Aspects

Molding tools are composed of various parts such as sliders, bores, pillars, the die surface etc. The cooling or heating of a local area of a die is realized by introducing channels into the tool that lead water or oil near the die surface. These channels form circuits. The number of bores and circuits depend on the complexity of the molding tool. Simple structures need few bores and a single circuit only. Complex tools may contain more than thirty bores and more than five circuits. The circuits are independent from

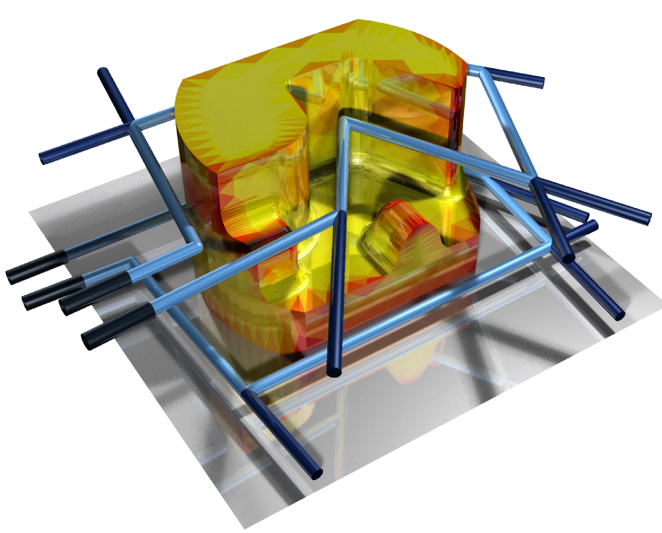

Abb. 22. Example of a design with two independent cooling circuits.

each other and can be used to heat up or cool down the tool. Heating up is important for reducing thermal stresses particularly for large tools in die casting. Dynamically changing strong thermal stresses can lead to increased tool wear or even to tool breakage. In the following, cooling and heating circuits are subsumed under the term cooling circuits because here only the geometric structure of the bores is important.

The number of the bores should not be too high because manufacturing of deep hole drilling bores is cost and time intensive. Deep hole drilling machines for large molding tools show restrictions with respect to the orientation of the bores. Often only horizontal orientations of the bores are possible while the die can be rotated on a table freely around the z-axis. Although some modern deep hole drilling machines do not have these restrictions, a design tool for realistic bores should be able to take these restrictions into account.

Another restriction in the bore directions is related to machining problems of intersecting bores. Bores that intersect each other in small angles cannot be manufactured safely by deep hole drilling. Therefore, it is preferable to design bores that intersect each other in large angles – ideally orthogonal.

The number of bores should also be kept small because any bore implies machining risks and costs and also weakens the molding tool. A certain wall thickness around each bore is necessary to reduce deformation or even breakage of the tool. In the design of large die casting tools rules of thumb are fast and well established but also very case depending and cannot be transferred linearly for small tools because the physics of the temperature follows nonlinear laws.

The simples structure of a cooling circuit is a consecutive sequence of bores. This is also the most prominent structure. More complex cooling strategies such as the branching of bores or so called finger bores for local cooling are more complex but lead to a higher efficiency of the cooling. Shortcuts with other cooling circuits are not allowed. A grouping of inlets and outlets of several circuits is also desired from a practical point of view.

Dead ends of bores used in circuits with more than two segments are sealed with plugs. Dead ends appear because each bore of a circuit has to be machined by deep hole drilling from outside. Plugging is necessary to seal a circuit from leakage. Plugs can be placed in any part of a bore. After setting a plug it seals a hole permanently. In some cases a plug can also be used to seal a die surface that has been penetrated by a bore. In general, drilling through the die surface is not allowed and should be avoided whenever possible because the subsequent machining, e.g. grinding, is very cost intensive.

Typically, dies and molds are designed with CAD systems. There exist various techniques for describing the geometric elements of a tool in the CAD system. The most general way is the triangulation of a surfaces. CAD-surfaces can be exported from CAD systems, e.g. in standard STL or IGES format and can be imported into a simulation software. The simulation software used in the following utilizes triangulation for describing the complete molding tool. The bores are modeled as virtual cylinders. Actually, it is possible to use arbitrary shaped triangulations of a die surfaces. For simplification reasons the tools are generally assumed to be rectangular blocks. The cooling cycles are modeled as sequences of consecutive cylinders. In the mathematical abstraction these sequences are simple polylines. Each cooling circuit can be defined uniquely by the position of the vertices of the polyline. The start and end points of the polyline always lie on the border surface of the molding tool. The model contains the complete bores for the testing of collisions and efficient machining. Plugs are defined by their position at the polyline. The drilling direction for the machining of a bore is not unique. Therefore, always both possible drilling directions can be modeled and analyzed.

Quality Criteria Definition and Evaluation Aspects

The cooling of an injection mold has to satisfy several aspects [44]. The most important idea is to machine cooling circuits that keep the cycle times as low as possible while the machining of the circuits should be cost efficient. Already these two basic aspects are contradictory. On the one hand the amount of bores increases the cooling effect and hence reduces the cycle time. On the other hand the costs increase with the number of bores. The efficiency increases with the number of bores but the costs do also. Hence, the basic problem is multiobjective with conflicting criteria. The efficiency of the cooling of the circuits can be increased by changing the layout of the cooling bores. The corresponding design problem also has a multiobjective character because the cooling efficiency of each bore should be uniform and strong. A uniform cooling distribution along the die surface is important e. g. to reduce internal stresses within the workpiece. An intense cooling is necessary for short cycle times. The relative distance and orientation of a bore to a point on the die surface is decisive for the cooling efficiency of the bore. The larger the distance more homogenous is the temperature distribution on the die surface but also the worse is the absolute cooling intensity and vice versa. The length of a bore also has a certain influence on the cooling effect. The longer the bore the better the cooling effect but the higher the costs for the manufacturing of the bore.

The complete expert knowledge for modeling the layout of cooling circuits can hardly be captured by an algorithm. Anyway, a lot of aspects of the current 'fuzzy knowledge' can be formalized and made available for mathematical use. Many values can be deduced from physical or economical properties of the machining process. Some knowledge can be formalized as quality criteria, e.g. average temperature of the die, others as restrictions, e.g. possible bore directions of the drilling machine. Basic properties such as the bore directions may also be encoded directly in the data structures. For example if the bores would have been restricted to lie in a single plane the problem would have been encoded different from the arbitrary 3D case as in the model used here.

The definition of the fitness functions can also benefit from efficient encodings. Here, the meaning of efficient encoding refers to fast fitness function evaluation as well as to a reduced difficulty for the optimization strategy. A low dimension of the search space and a continuous fitness landscape with only few restrictions is preferable.

For a technical realization of the efficient description of the casting expert's knowledge physical properties of the thermal flow in metal, economical assumptions about the manufacturing process as well as mathematical heuristics have been used. Adequate estimations about a sufficient level of detail is a helpful technique to reduce computation times. These approximations can be used to choose parameters of the geometric model of the tool such as the number and size of the triangles used to describe the die surface. Approximations have also been applied to realize a very fast estimation of the cooling effect of the bores.

The quality of the cooling depends on the local intensity of the cooling effect of the bore to the die surface. Additionally, a globally uniform intensity of the cooling distribution is important to get e.g. workpieces with a minimum of internal stress and, therefore, high workpiece qualities. Global and local cooling are conflicting goals. Additional conflicting goals such as the number and length of bores have to be taken into account. The layout of the bores has to fulfill technical and machining restrictions. The bores may not intersect with the die surface nor with any other bore. Of course, all bores have to be within the tool geometry. This restriction seems trivial, but it is a necessary condition that all solutions have to fulfill. Additional machining restrictions such as limited bore direction angles have been addressed in the optimization tool.

The evaluation of the cooling effect can be calculated very exactly by Finite Element Methods (FEM). Unfortunately, this approach needs long run times. FEM has the advantage that it calculates the complete temperature distributions within the tool for all times. It is necessary to use high numbers of nodes, because a small resolutions can lead to relevant discretization errors when the bores change their positions only slightly. In a static case and a sufficient number of nodes FEM is a tool for precise calculations that is very helpful for analyses of the final construction design. Actually, some companies use FEM to check the quality of the manually generated bores. This is especially necessary in die casting where blowholes – caused by insufficiently cooling – have to be avoided.

For manual layouts often only the mean temperature in the time period between injection and ejection is used to estimate the actually dynamic temperature distribution in the tool [88]. Using only the static case simplifies the analysis a lot and helps to increase the calculations.

Due to the long evaluation times of the FEM, a fast surrogate model was introduced. The new model approximates the temperature flow mechanism by modeling heat radiation. The exponential decrease of the cooling properties with increasing distance between bore and die surface is covered by both models. The quality of the FEM solutions can be reached in many cases. Only when the bore are shaded by an object or two bores are lying very near to each other, the approximation is not exact. In most cases the shading effect and the superposition problem can be disregarded because of the strongly decreasing local character of the cooling effect. Furthermore, the superposition of the bores introduces an overrating of the cooling effect. This leads in the optimization process to bores with a larger distances to the die surface. The cooling distribution criterion introduces a separation of the bores and, hence, the overestimation error reduces exponentially again.

The basic idea of the radiation approach is to calculate the temperature distribution and intensity on a die surface radiated by a beaming cylinder. Each bore can be interpreted as a ’neon lamp’, which illuminates the die surface. Mathematically, the radiation effecting each triangle $\tau_{i,j}$ – i.e. a small part of the die surface – is defined via the integral of the squared reciprocal distances along each polyline – i.e. bore – between the consecutive vertices P_i and P_{i+1} [77].

$$\tau_{i,j} = |P_i - P_{i+1}| \int_0^1 \frac{1}{(P_i + t \cdot (P_{i+1} - P_i) - M_j)^2} \, dt \tag{16}$$

The total local cooling effecting at triangle τ_j is the sum of all partial effects generated by each bore i:

$$\tau_j = \sum_i \tau_{i,j} \tag{17}$$

The global cooling effect f_l is calculated by the arithmetic average value of all τ_j. f_n is the normalized minimum effect value over all τ_j.

Function (2) can also be used to characterize the cooling distribution by calculating the statistical variances of the cooling values around each triangle τ_j. The global temperature distribution f_d is the mean of the variance values calculated over all triangles.

In the basic model a constant initial die surface temperature is assumed. In reality some areas are hotter than others, e.g. at the spray or where material accumulates. Demands for additional cooling is modeled by temperature values that are attached the triangles. These cooling demands are integrated into the model (see fig. 23).

The aggregated temperature effect f_t, which is used by the optimization process, is composed of three parts: the arithmetic average sum over all radiation influences of all bores on all triangles (describing the absolute cooling effect) f_l, the temperature standard deviation (modeling the uniformity of the illumination) f_d and the normalized minimum effect value f_n. All values are normalized and about similarly sensitive to changes in the bore design.

In multiobjective optimization it is possible to use a priori or a posteriori techniques. Both techniques have been used here. The a priori method is represented by an aggregation or sclarization technique. This method can be used adequately, if weights

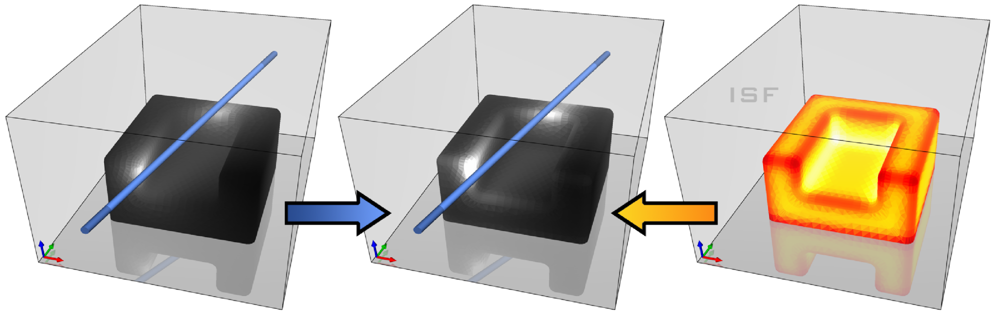

Abb. 23. Cooling effects modeled with the radiation approach (left). Highlighted areas are cooled better than dark. Additional cooling demands (right) can be included into the evaluated model (middle).

can be defined a priori by the user. It is also known that linear aggregation should only be used in the convex case of the Pareto front. Actually, there are only very few practical cases known where the Pareto front is concave. The a posteriori analysis of the Pareto front of the mold temperature control problem proves this fact again, as shown later.

The aggregated fitness f is a function of the standardized restrictions f_{pen}, the manufacturing costs f_c (i.e. the length of the bores), and the temperature effect f_t:

$$f = f_t \cdot (1.0 + f_{pen}) \cdot (1.0 + d \cdot f_c) \tag{3}$$

$$f_{pen} = a \cdot f_{bc} + b \cdot f_{sc} + c \cdot f_{ot}$$

The allover penalty f_{pen} is a linear combination of the number of collisions between two bores f_{bc}, the factor f_{sc} describes the collision between the bores and the tool surface and f_{ot} is a penalty for the bores that reach out of the tool. a,b,c,d are arbitrary real value weighting factors.

The radiation approach has the advantage of very short calculation times. Compared to the FE method, which takes about 10 minutes for a simple geometry, the calculation time for analysis of the cooling design using the radiation technique takes about $1ms$ on a standard PC. Short evaluation times and good approximation results allow to efficiently utilize stochastic algorithms and to solve the complex problem of mold temperature control design.

Multiobjective Evolutionary Algorithms

Since the last about five years the field of multiobjective evolutionary optimization is in the focus of increasing interest. The problem of multiobjective optimization is not new. Already in 1871 Vilfredo Pareto investigated in his work the problems of multi objectiveness and suggested the mathematical concept of best compromise solutions, the later called Pareto fronts and Pareto sets [53]. The classical single criterion optimization is a special subclass of the more general multiobjective optimization. In general, single criterion techniques cannot be applied to multiobjective cases without changes. Even the comparison of solutions is not as easy as in the scalar case, because sets of compromise solutions have to be compared with each other. Therefore, the optimization techniques have to be adapted to this larger class of problems. Optimization means a gradual increase in quality. A multiobjective optimization technique has to improve the quality of solution sets. Furthermore, not only the approximation of the true Pareto fronts and sets is a necessary precondition of a good multiobjective optimization algorithm. The structure of the solution has to be well spread along the Pareto front or at least well spread in a certain region of interest. Additionally, of course, the approximation has to be calculated fast. In multiobjective optimization a lot of problems are very complex in an algorithmic sense, i.e. they cannot be solved in polynomial time. In combinatorial problems even the management of the solution sets can be NP-hard. Nonlinearity, a multimodal character of single or all functions, discontinuity and function areas with no inclination, many nonlinear restrictions, high dimensions of the decision space and even dynamic features are characteristics of real world problems that make multiobjective optimization a very difficult field in mathematics [44].

There exist several deterministic methods for multiobjective optimization. Most of them are relatively old. Nearly all techniques find only one solution on the Pareto front per run. The theory in continuous multiobjective optimization is also not as well developed as in the single objective case, although there exist very good literature (see e.g. [26, 47]).

Multiobjective optimization techniques can be grouped into a priori, a posteriori and interactive techniques. Many approaches use the a priori approach, where all necessary knowledge about the fitness criteria and their relative importance is known in advance. In this case, typical approaches are scalarization of the multiobjective optimization problem via aggregation or lexicographical ordering. There are various aggregation methods that use e.g. weighting sums or sums of weighted factor combinations. Lexicographic max-ordering is an interesting alternative, because this technique allows to find points that are Pareto optimal [26]. This property is not true for linear weighting techniques. Today, a posteriori problems are generally solved by stochastic optimization as described later. Interactive methods are an interesting alternative to both extreme cases, where expert knowledge influences the search process directly [44].

Deterministic techniques have wonderful properties if the problem can be solved with these approaches. It is possible to give approximation estimations and deterministic methods are also very fast. Of course it is not easy to find an efficient method for a specific problem. Furthermore, the NFL-theorem [84] tells that there is no best optimization algorithm for any class of problems at all, although, reducing the general set of all problems to smaller classes, it is possible to find at least 'appetizers' [24].

Looking for algorithms that can find good solutions and also cover large areas of the Pareto front needing only one run leads to multiobjective evolutionary algorithms. In the single criterion case evolutionary algorithms have proved their effectiveness in many cases, either in theoretical or in practical applications. They may not always find the theoretically best solution, but on the one hand in the multiobjective case it is difficult to prove the optimality of a solution anyway – especially if only simulations are available – and on the other hand in practice gaining a relevant improvement is often that difficult that already a certain melioration is highly desirable.

Multiobjective evolutionary algorithms (MOEA) are developing since the last decade [32]. A relevant increase in research and application came with the introduction of the Pareto dominance based MOEA. Algorithms such as the NSGA-II [20] or the SPEA 2 [87] belong to the classics of the current state of the art. A description of current algorithms is available in [17] or [44]. A discussion of MOEA is also given in the chapter 'Challenges for Contemporary Evolutionary Algorithms' of this book.

Multiobjective evolutionary algorithms have the advantage that they do not have to be re-programmed when fitness functions are changed or new criteria or restrictions are introduced. They can be used either in scalar mode (classic single criterion EA) or in Pareto dominance mode (Pareto fronts and sets). Both approaches have been tested in the following.

In order to compare Pareto fronts with respect to their quality, various scalarization methods – so called metrics – have been introduced. For a well founded comparison see e.g. [75, 18]. Metrics help to calculate a quality measure in the objective space of a Pareto front with respect to a given reference point, a known best Pareto front or a relative measure between two Pareto fronts. Of course the definition of the quality is not unique. Of course good approximations of the true Pareto front and well spread solutions are desired. Often the true Pareto front is not known. Only best solutions found so far may be at hand. Furthermore, solution sets are discrete in nature. Metrics have to be able to cover all these multiobjective problems and to be as intuitive as possible. A technique that matches many of these issues is the attainment surfaces method [38]. A variant of this method is the MMBBH-measure [45]. This method has been used to assign a distribution measure and an approximation measure to one value which is needed for the comparison of algorithms. The solution sets for the comparison may come from different MOEAs or the same MOEA with different parameter settings. This value was necessary for the application in the univariate design of experiments as used here to meta-optimize the MOEA parameter settings.

Statistical Design and Analysis

Uncertainty, over adaptation and generalization abilities are conflicting properties in modeling and design of experiments. Experiments are necessary for the verification of thesis or mathematical models. The model is generally only an approximation of reality. It has to cover certain properties and should help interpreting aspects of reality. Some models can also be used for extrapolation in time, space or contexts. Most models are restricted to a certain area of interest and use data sets as a basis to describe input-output relations. Usually only a restricted number of data is available, some data is missing or sparse, or the data contains errors or noise. In general, the design of adequate models is a very difficult task. This is especially true for the optimization in real-world applications because often no models, i.e. no mathematical functions, exist at all. The selection of the data for the training of a model is also non trivial in general, because the number and distribution of the learning data depends on the model and on the availableness of the data.

A well established method for the analysis and optimization of problems, especially when the number of available data is low, is statistical design of experiments. Design of experiments (DoE) methods need a minimum of data. The results can be used to describe interactions between parameters and support optimization purposes as well [76]. Grid design plans have the disadvantage that the number of experiments grow strongly exponential with the number of dimensions of the parameter space. One-factor-at-a-time approaches are very popular but they do not allow to make statements about the interactions between parameters [76]. Full factorial design plans use two or three parameters per dimension, fractional factorial design plans need even less experiments and allow to characterize factor interactions and optimization. The standard DoE models are linear or quadratic. Planar linear models, i.e. models without factor interactions, are used for screening. This means that in problems with high numbers of factors (parameters) the most relevant are filtered with respect to the influence on the responses (scalar outcome). In the modeling phase a linear model with interactions, i.e. factors can influence each other multiplicatively, is used to describe interdependencies between factors. In the modeling phase of DoE only the relevant factors, that were previously filtered by the screening phase, are used because the design plans are a little more expensive. In the optimization phase of DoE a square model is used to fit the measured data via regression. Quadratic optimization can be solved efficiently. Algorithms such as the conjugate gradients method are advisable. During the screening phase Plackett-Burman design plans are used. These designs can be calculated by systematic schemes [49]. Factorial design plans are used in the modeling phase. They also follow schemes that are listed in the literature [49]. Central composite design plans are typical for optimization [76].

Design and Analysis of Computer Experiments (DACE) [60] is a technique for the generation of design plans and evaluation of experiments with deterministic behavior and quantitative character. The analysis of nominal factors, nondeterministic responses

[7] and sequential introduction of new design points (SPO) [46] are current state of the art extensions that improve this method.

DACE needs a space filling design. A typical plan is the latin hypercube design (LHD). Compared with DoE, LHD generates similarly few design points. The points are spread on a grid in a way that each column and line of the grid is occupied by one point only. Additionally, in the experiments the points are distributed in a maximum space filling fashion in the search space. The responses of the corresponding experiments following the LHD designs are interpolated by Kriging. Kriging introduces a correlation between the measured points and yields a smooth interpolating response surface. The problem of noise in the responses can be solved by boot strapping [46].

SPO (sequential parameter optimization) introduces an easy to use and efficient way to improve the design plans by iterative introduction of new design plans. New design points are set following a minimum trade off principle that minimizes the uncertainty of a prediction error (global search) and maximizes the local improvement to find better function values (local search) [63]. Starting from an optional DoE screening of the relevant factors, the first step in SPO is to perform usual DACE calculations. The Kriging model in the DACE approach plays the role of a statistical prediction tool. In the subsequent SPO iterations, only relatively few new points are introduced to improve the model fit step by step. Compared to the number of points used during the first SPO step, generally in the following steps only few (only 10 percent had been necessary in our experiments) new points are used.

The DACE/SPO technique is important when the number of fitness function evaluations has to be small. This is especially true when the evaluations are expensive. This technique can also be used for meta-optimization of EA parameters. Although the EA's behavior is relatively invariant against changes of the fitness functions, a tuning of the parameters can still help improving the quality of the results with a statistical significance.

Results

The evolutionary optimization of the model temperature control designs starts with an initial definition of a population of possible solutions. Each solution called individual is a vector of vertices of a polyline modelling the cooling circuit and the respective bore directions. The initial population contains values that are arbitrarily distributed in the volume of the tool. In the beginning it is possible that the bores may penetrate the surface of the die. These solutions have a relatively bad quality so theses individuals are selected from the population during the evolutionary process. The initial parameters such as the number of bores, the geometric properties of the bores and the die, parameters that influence the optimization such as the population size, the selection pressure and many more values can be tuned with a graphical user interface. Selected bore designs can be visualized online during the run and can be modified interactively. The system supports a reevaluation of modified solutions in real time. This allows the user to watch the optimization process and to tune the parameters of the simulation according to the individual impression interactively. In the analysis standard die geometries were used. Alternatively the system can import, visualize and analyze STL-CAD-geometries. Due to the modular structure of the tool, the evaluation process can be linked to an evolution strategy or to other MOEA-tools. The interfaces between the optimization strategy, the visualization and the evaluation module are realized as an internet socket connection. Therefore, several system components can be exchanged with respect to the respective aims of research.

The ES has been modified slightly. The experiments show that using correlated mutations yield a slight improvement in the convergence properties. A significant influence on the probability to converge toward good solutions was realized by introducing a limited minimum step size that is declining slowly. This variation was useful because the standard step size adaptation is to fast for this application. An increase in the freedom of search, i.e using relatively large step sizes, helps to avoid getting stuck in local minima that are typical for the strongly restricted search space. Due to the fact that bores must not intersect each other, structures can get into constellations that cannot be resolved with small geometric step sizes. Large step sizes are an effective solution for this problem. Additionally the penalty functions for the intersection of fixed geometric restrictions such as pillars etc. that must not be intersected by bores can increase gradually in the simulation. This also improves the ability of the algorithm to explore the complete search space.

Uncertainties about the preferences of the user toward different quality properties can be modeled via a priori and a posteriori techniques. The ES approach aggregates the fitness functions using a product of monomes using normalized fitness values. This method allows to model fitness function interactions and helps to tune the values of the influence of the functions with respect to the preference of the user. The single objective evolution strategy as described above has used. This has the advantage that good initial parameter settings of the ES can be chosen according to good experience. The behavior of the ES is well known and the analysis of the optimization process can therefore be focused more on the convergence behavior of the bore structures toward good solutions. The fast fitness evaluation of the radiation approach is a necessary precondition for the efficient optimization using population based techniques. FEM analyses in the ES loop have been tested but the long evaluation times allowed only small population sizes and few generations.

Uncertainties using a posteriori approaches are supported via a MOEA tool called KEA (Kit for Evolutionary Algorithms) that contains current state of the art MOEA algorithms such as the NSGA II and SPEA 2. Two multiobjective particle swarm algorithms are also included in KEA. The MOEA optimizing tool is written in Java and communicates via an internet socket interface with the evaluation tool. It provides a graphical user interface (GUI) for choosing between different MOEA algorithms and various fitness functions. The GUI also supports the tuning of the MOEA parameters. The Pareto-fronts of two criteria multiobjective functions can be displayed online during an optimization run.

A typical optimization of the mold temperature control designs using the ES working on the standard test problem of the spherical die takes about five minutes on a standard PC. The MOEA optimizers are similarly fast. Of course in the MOEA approach not

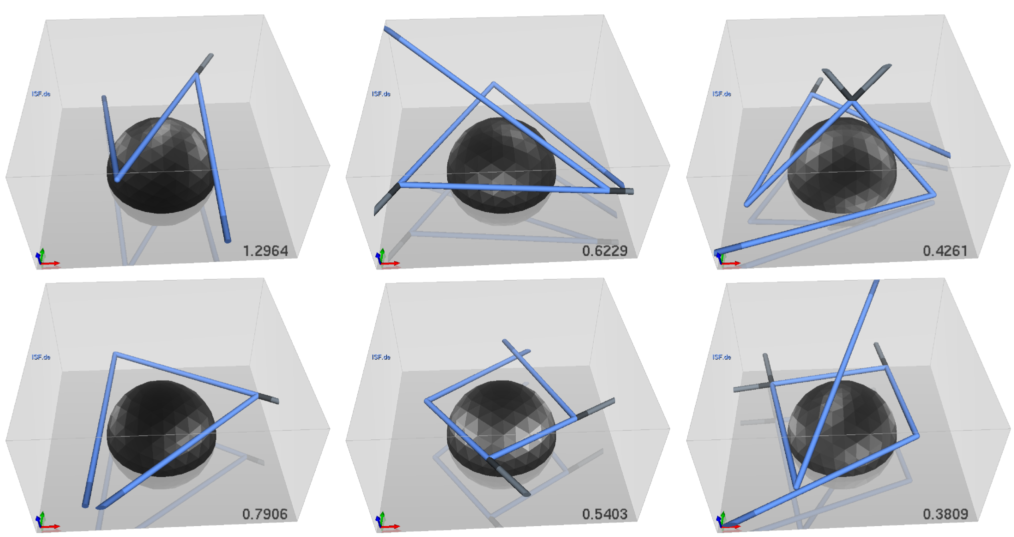

Abb. 24. Two typical design solutions each for 3, 4, and 5 bores.

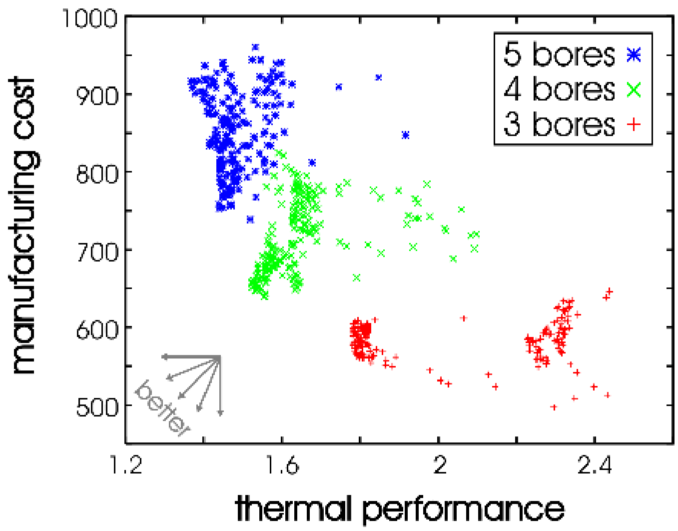

Abb. 25. 200 solutions for runs with 3, 4, and 5 bores using the aggregating approach with an ES approximating a Pareto front.

every Pareto-optimal solution (i.e. bore structure) is displayed online during the run. Single solutions have to be selected from the Pareto-front by the user after the run and have to be displayed by the visualization tool of the evaluator separately (and Fig. 25).

The MOEA optimizers are not limited to two criteria. The mold temperature control design optimization typically uses four quality criteria (cooling distribution, cooling intensity, costs) together with an aggregated value of the restrictions. Using a higher number of criteria than four or five – actually more than twelve criteria are implemented – is not advisable due to practical reasons. The Pareto-front becomes much too complex to be analyzed by the human user. In literature about realistic real-world applications a maximum of seven criteria seems to be advisable [17].

Looking at the geometric results of the bore designs when applied to a half sphere die surface, it is striking that a relatively high number of local optima appear. The solutions can be grouped to clusters of designs with a similar structure (see Fig. 24). The structures between the clusters differ significantly. This is an indication that the fitness landscape is multimodal. The average quality of the results of each cluster also differ. Even for a human expert it is not easy to tell what structure has a better quality than another. Therefore, the objective evaluation is a valuable help for the designer to compare designs and to learn more about efficient structures.

The application of the standard statistical design of experiments proved to be an adequate technique to improve the values of the ES as well as the MOEA results. The MOEA results have been compared using the MMBBH-measure [45]. The Meta-optimization using DoE assumes a quadratic behavior of the response values. In many practical applications and also near the optimum this assumption is very useful and an optimum can be found efficiently. Often an optimum lies on the border of the min-max-limits of the factors. Then the area of interest should be shifted to a more promising region and the experiments have to be repeated. Applying DoE to the ES and MOEA increased the quality of the results significantly, i.e. the values of the measure qualities improved more than ten percent compared to results with standard settings.

A more efficient approach for the design of experiments is provided by the DACE/SPO-method [46]. Although, also here extrapolation is also not allowed, a larger region of interest can be scanned and modeled with free form surfaces (see fig.26). The Kriging model shows very intuitively the interdependencies between the parameter settings and their influence on the quality. The strong stochastic character and the mul-

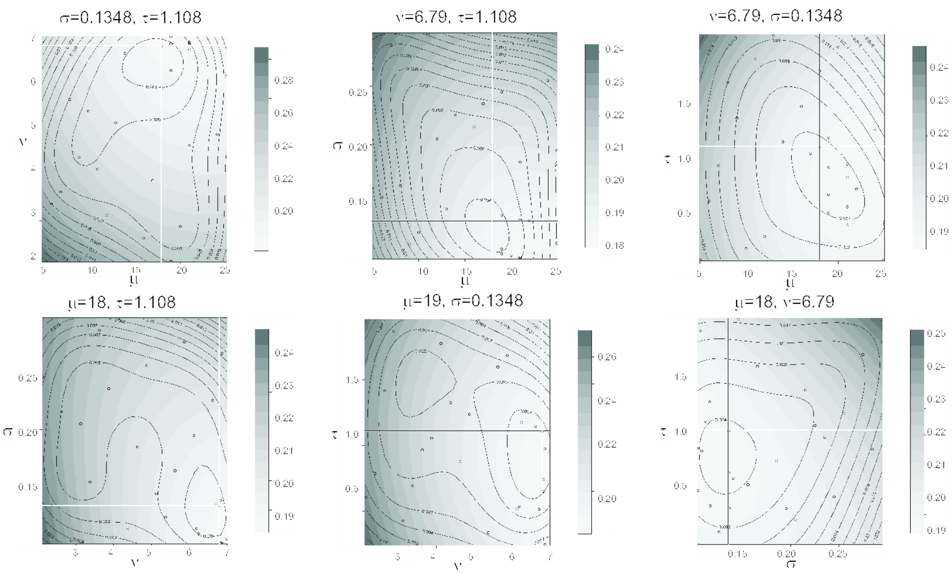

Abb. 26. DACE response fitness landscapes (grayscale) and errors (contour lines) for the parameter optimization of the ES. The best parameters are marked by a cross.

timodal fitness function is a problem. DACE usually needs a deterministic behavior of the responses. The evolutionary algorithms generate a spectrum of results for the same parameter setting. The distribution of the experimental results can be handled efficiently with a boot strapping method. The number of experiments per parameter setting can be kept surprisingly small, although the variance of the results is quite high. The boot strapping method generates single representative response values that were used in the DACE model as deterministic responses. The number of additional new design points that are used to improve the DACE model in the SPO step is also quite small. Therefore, the DACE/SPO technique is very efficient and supports an intuitive interpretation due to the flexible Kriging visualization. Near quadratic dependencies between the parameters are displayed for the spherical test design problem. In DoE only quadratic hypotheses can be confirmed or rejected. The application of the DACE/SPO technique to the mold temperature control design optimization algorithm improved the EA results significantly.

Summary

The mold temperature design optimization problem shows a high structural complexity, the fitness functions are multimodal, complexly restricted, high dimensional and multiobjective and the uncertainty about the users preferences make the search for good solutions extremely difficult. An adequate geometric modeling as well as the objective modeling of qualitative and quantitative quality criteria is an important precondition for realistic real-world optimization. This problem is even more challenging when the evaluation of the criteria is time consuming and many evaluations have to be performed.

The modeling of the fitness functions follows an approximation of the real temperature flow evaluation using a radiation technique and the modeling of practical heuristics. The multiobjective optimization is performed via evolutionary algorithms. Different techniques such as scalarization and Pareto dominance were applied alternatively to provide different solution methods to the user. Speed, visualization and interactive optimization are important issues for real-applications. The systematic improvement and verification of the results have been done using statistical approaches such as classical design of experiments and modern DACE/SPO methods.

8 Medicine: Arbitration of benefits and harms in radiation therapy treatment plans

Karl-Heinz Küfer, Philipp Süss

Show the necessity of trading of treatment effect vs. side effect. Discuss the possibility of mathematical aids in finding a parato set. Show necessity of custom approach, to exploit mathematical benefits.

4

Decision support for intensity modulated radiotherapy planning

Karl-Heinz Küfer

Department of Optimization
Fraunhofer Institute for Industrial Mathematics (ITWM)
Fraunhofer Platz 1, 67663 Kaiserslautern, Germany
E-mail: *kuefer@itwm.fhg.de*

No Institute Given

4.1 Intensity modulated radiation therapy planning

Radiotherapy is, besides surgery, the most important treatment option in clinical oncology. It is used with both curative and palliative intention, either solely or in combination with surgery and chemotherapy. The vast majority of all radiotherapy patients is treated with high energetic photon beams. Hereby, the radiation is produced by a linear accelerator and delivered to the patient by several beams coming from different directions (see figure 4.1).

Abb. 4.1. The gantry moves around the couch on which the patient lies. The couch position may also be changed to alter the beam directions.

In conventional conformal radiation therapy, only the outer shape of each beam can be smoothly adapted to the individual target volume. The intensity of the radiation throughout the beam's cross section is uniform or only modified by the use of pre-fabricated wedge filters. This, however, limits the possibilities to fit the shape of the resulting dose distribution in the tissue to the shape of the tumor, especially in the case of irregularly shaped non-convex targets like para-spinal tumors, prostate carcinoma located close to the rectum, or head-neck tumors in the proximity of the parotid glands and the spinal chord.
A significant advance in treating such difficult cases was the development of intensity
7 modulated radiation therapy (IMRT) where the intensities on the cross-section of a beam can be varied. Using multi-leaf collimators (MLCs) (see figure 4.2), the intensity is modulated by uncovering parts of the beam only for individually chosen opening times and covering the rest of the beam opening by the collimator leaves. This permits shaping highly irregular dose distributions in the target volume.

Abb. 4.2. A Multileaf Collimator (MLC). The square opening in the center of the machine is partially covered by leafs, each of which can be individually moved. (picture from [62])

An IMRT plan is physically characterized by the beam arrangement given by the angle of the couch relative to the gantry and the rotation angle of the gantry itself, and by the intensities on each beam. The treatment aim is - in lose terms - to deliver sufficient radiation to the tumor while sparing as much of the healthy tissue as possible. A major
1 challenge in IMRT planning is to cope with these fuzzy goals and demands.
Concerning the beam arrangement, in most cases a concentric irradiation geometry is chosen: the beams meet in one iso-center as depicted in figure 4.3. Additionally, the beams are generally chosen to lie in the same plane, as positioning the couch to

Abb. 4.3. Beam directions in a concentric irradiation geometry with an equiangular setup

realize a full 3D geometry increases treatment time substantially. Since the mathematical formulation of the planning problem including the irradiation geometry poses a much more difficult problem description[1], the beam directions are usually fixed manually. In

[1] This problem is not convex and hence can not be guaranteed to be solved exactly in reasonable time.

most cases, there are fewer than 10 beams and often they are arranged in equiangular distance around the patient.

The remaining planning parameters - the description of the intensity modulations of the individual beams - may now be determined by mathematical optimization procedures. However, the choice of an appropriate objective function to optimize presents one of the biggest hurdle in formulating a mathematical optimization problem for IMRT planning. Many quality measures for a treatment plan based on different approaches have been proposed by medical physicists and oncologists. Common to all measures is that they are based on the distribution of the realized dose in the patient's body. Some measures typically used are, for example, the average dosages received by organs and tumor volumes. Often a variation from ideal reference doses are measured. Yet another function is the variance around the mean dose in the target. Structurally, these commonly used quality measures are descriptive and comparable for the planner.

In most clinical cases, one or two objective functions are specified for the clinically relevant structures. This way, a multi-criteria formulation of the planning problem arises naturally. The quest for a solution that has a single optimal quality score is thus extended to a search for solutions that are *Pareto optimal* or *efficient*. An efficient treatment plan has the property that none of the individual criteria can be improved while at least maintaining the others. In our context, given a Pareto optimal treatment plan, lowering 7
the dose in one organ can only be achieved if the dose in the tumor is also lowered, or at least one more organ receives a higher dose, for example. In general, there is a multitude of such solutions to any given multi-criteria optimization problem. We refer to all efficient solutions as the *Pareto set*.

The subjective nature of IMRT planning is in part due to the specific choice of objective functions by a treating oncologist. A different planner may prefer similar but different quality measures for a given case. Since all quality indicators in IMRT planning are 9
designed to achieve the same fuzzy goal - allowing control over the spread of the tumor - they measure more or less similar effects. It may therefore safely be assumed that the quality measures are correlated with each other: good plans under one objective function are also good plans with respect to another objective function. In other words, a Pareto set obtained using one specific set of quality measures is most likely not altered too much if a different set of correlated quality indicators were used. This diminishes to some extent the severity of the fuzzy environment IMRT planning is situated in. As long as the planner uses quality indicators that correlate with commonly agreed on sensible measures, it is most likely that the treatment plans he obtains are of high clinical quality.

The software MIRA (Multi-criteria Interactive Radiotherapy planning Assistant) developed by the Fraunhofer institute for industrial mathematics (ITWM) in Kaiserslautern is designed to calculate Pareto sets and, more importantly, provides an interactive environment to select an efficient plan that is according to the notions of the oncologist. The latter is the implementation of a support system for a highly complex decision problem. Detecting planning possibilities and physical limitations in a database of efficient solution while considering multiple clinical preferences at the same time is a very challenging task. 9

The optimization algorithms implemented in MIRA calculate a database of Pareto optimal treatment plans that cover a clinically meaningful range. After the decision-maker selected some quality measures, MIRA begins with the calculation of solutions. With a high number of criteria comes a relatively large Pareto set. Since this set is continuous, the possibility of displaying the infinitely many solutions is not an option. To still select a treatment from the Pareto set, it suffices to approximate the set appropriately. To achieve this, a so-called planning horizon is determined to exclude clinically irrleveant solutions. The Pareto set will contain a lot of plans where the trade-off between overdosage in the organs and a specific dose in the tumor will be too high to consider. MIRA deploys *extreme compromises* as corner points for the planning horizon in the Pareto set. They are the compromise of all possible combinations of the

(a) The lines depict the dose distribution in the body

(b) The same distribution as in (a) with a more detailed resolution - the original picture shows different colors for each level of dose

Abb. 4.4. Visualization of the dose distribution in one of more than 100 CT-slices of a patient

objective functions. The idea and its mathematical consequences are described in more detail in [41].

After this planning region has been marked, individual solutions are placed between the extreme compromises to approximate the Pareto set in more detail. Calculating an extreme compromise or another solution from the Pareto set requires solving a high dimensional optimization problem. Specialized optimization routines developed at the ITWM enable

efficient calculations and creation of the databases in reasonable time. The strategy called the *adaptive clustering method* solves a sequence of approximate problems, which iteratively converge to the original problem. The solution of each previous problem is used to adaptively refine the approximation. The method is explained more fully in [61].
After the calculation of a database of treatment plans, the interaction with the decision-maker begins. MIRA has an interface called the *Navigator* (see figure 4.5) which enables the user to explore the Pareto set using visual controls. All user interaction with the Navigator are internally translated as optimization problems on the Pareto set and solved in real-time. This allows a smooth interaction with MIRA and explains the names “assistant” and “Navigator”.

Abb. 4.5. The Navigator. The star on the left hand side shows the planning horizon and the current solution. On the right the dose visualization and other statistics for the current solution are shown.

The star on the left hand side is composed of axes for the different quality indicators. The objective functions associated with the organs at risk are combined into a radar plot, whereas the functions associated with tumor volumes are shown as separate axes. The interval on the axes corresponds to the range of values contained in the database for the individual objective functions. The white polygon marks the indicator function values for the currently selected plan. The shaded area represents the planning horizon.
The interaction with MIRA is characterized by two fundamental mechanisms which are patented for the use in IMRT planning by the ITWM:

- *restriction*, to alter the planning horizon, and
- *selection*, to alter the current solution.

Restriction allows the user to set the bounds of the planning region in both directions: to narrow it and to expand it. Unwanted parts of the Pareto set can be excluded from planning this way. The planning horizon displays a valuable piece of information for the decision-maker: the inner boundary close to the center of the star represents the highest ambition in an individual objective the user may have, while the outer boundary depicts the absolute highest “cost” in a criterion that has to be paid to improve another objective. The line representing the currently selected plan has handles at each intersection with an axis and triangles for the tumor related axes. Both can be grabbed with the mouse and moved to carry out the selection mechanism. Internally, optimization routines ensure that the changes in the other objectives are at a minimum. This enables the greatest possible control over the navigation through the Pareto set: MIRA tries to change the current

Abb. 4.6. A conceptional view on the functionality of the Navigator: while the user changes the current solution, the “distance” of the current solution on the surface to the point represented by a star is minimized. The line connecting the stars shows the path the user has dragged in the navigation by moving a slider. This is where the users would like to the solution to be. The shaded region depicts the planning region on which the path casts a “shadow”. The coordinates of this shadow are displayed in the Navigator.

solution only in the direction the decision-maker wants it to.
The right hand side of the screen displays the corresponding plans concurrently. It is the visualization of the dose distribution parallel to the interaction with the Navigator that conveys the possibilities and limitations in a clinical case.

Several optimization problems are solved each second to allow a smooth interaction with the user. The sensitivity of the quality measures that are not actively changed to the objective that is currently altered is shown by the magnitude of the changes in the current solution while the handle is pulled. If small changes in one quality measure lead to rather large changes in some other criterion, the objective functions are particularly sensitive to each other in the region around the current solution. The decision-maker gathers a lot of insight and feeling for the clinical case using this valuable information and is thus able to accurately determine the trade-offs involved with each treatment plan.

We now demonstrate a small planning example for a head-neck case. These are typically difficult to plan, since the primary tumor can be located anywhere in the naso- and oropharyngeal area, and regularly the lymphatic nodal stations have to be irradiated because they are at risk of containing microscopic tumor spread. This results in big, irregular shaped target volumes with several organs at risk nearby.

The salivary glands are such organs at risk that are quite radio-sensitive. The tolerance dose of the biggest salivary gland, the parotid gland, is approximately a mean dose of 26 Gy. The goal should be to spare at least one of the parotid glands. Otherwise the patient might suffer from xerostomia (a completely dry mouth) which can significantly reduce the quality of life. Other normal structures that have to be considered are (depending on the specific case) e.g. the brain stem, the spinal cord, the esophagus and the lungs.

The objective functions to two tumor volumes are the homogeneity of the dose around a specified mean. The axes of the star depict objective function values of the spinal chord, unclassified tissue, parotid glands, the left eye, the right eye and the brain stem. Now the interactive part of the planning process begins. The required compromises between the selected objectives can be evaluated by pulling the handles as described above. A first
4 intuitive feel for the possibilities in the case is developed this way.

While the user alters the solution, the Pareto set and its dose distributions are explored in real time. MIRA also provides a locking mechanism for an organ. The user can fix a value on an axis, which will henceforth not be breached unless the lock is removed. All treatment plans that score worse than the fixed values are excluded from the planning region. This effect can be seen by the reduced planning horizon in 7(b).

Complex planning scenarios can be interactively explored in this manner and the best clinical treatment can be found within shortest time.

The design of the Navigator and the functionality of MIRA are targeted to spare the user from mathematical formulations of optimization problems, and rather provide clinically descriptive information that is easy to interpret by experts. These aspects play a decisive role in acceptance of MIRA among treatment planners.

MIRA and the navigator present a tremendous opportunity for improvement of the planning workflow in hospitals all over the world. The graphical user interface for navigating a complex solution set and the flexibility in modeling provided by MIRA also facilitate interaction between physicians and medical physicists during the planning stage.

All optimization processes in commercially available planning packages for IMRT are
8 based on mixing the objective functions if they even provide the opportunity for adding more than one. No currently available software provides multi-criteria decision support to the extent that MIRA can realize.

(a) Navigator at the beginning of the planning process

(b) The Navigator after some restrictions have been made - note the significant difference in the remaining planning domain.

Abb. 4.7 The Navigator for the head and neck

4.2 Multi-criteriality in the public perception: Experiences with a multicriterial decision support by an interactive program tool

P. Roosen and K.-H. Küfer

How do ’uneducated, ordinary people’ tackle multi-criterial decision making? What strategies do they pursue to ’obtain what they want’? Professionals regularly and explicitly dealing with these kind of problems know several methods of approach, each of those being thoroughly investigated and assessed with respect of their predications and limitations (cf. section 5.3). In spite of the frequency of such problems also in everyday life this knowledge is not widely spread, though, so it is interesting to investigate whether non-professionals can make use of a support tool oriented towards this decision domain. The general idea is that the last resort of objectivity is the Pareto set of a given problems, so its display will yield the best possible decision support with respect to a self-restrained potential influencing of the decider by the tool provider.

While most practical problems present more than two target qualities to be optimized there is no static presentation tool that would provide a intelligible, ideally printable, selection display for the decider to work on. Therefore a computer-based tool was created to support interactive, visual pareto set testing poking. As the principal display mode is based on a star plot, the system was called ’Pareto Star’ accordingly. Its applicability is not limited to pure pareto sets, though, as in practical applications there is not necessarily a sharp distinction of important decision criteria vs. irrelevant ones. Instead, a kind of continuous transition of criteria more or less crucial for the decision is frequently observed.

The designed decision support system is not a tool to *identify* a pareto set — it requires the existence of a database of realizable solutions with their individual target function values. There are cases, though, that exhibit continuous settings or operation modes and hence a dense pareto set of unlimited member count, so no distinct, objectively chosen points in target function space can be provided. Here the principle of uncertainty should be kept in mind, however: Small changes in the relative trade-off target settings usually do not make large differences in the eyes of the decider. So it suffices to identify sensible representants of the originally dense Pareto set and present them as individual target function (and accordingly configuration parameter) combinations in the selection tool.

As a practical test bed the selection of an handheld GPS receiver was chosen. There are several practical selection criteria, and no all-purpose devices serving the whole range of private navigation tasks exist. So the potential buyer is confronted with the selection problem, being an ubiquitous obstacle when scanning respective electronic discussion fora for answers and opinions. The tool devised as decision support is implemented as a web service, extendable to almost arbitrary multi-criterial decision problems. It operates on three basic tables, implemented by means of a database system: i) the list of potentially interesting criteria, ii) the list of available devices, and iii) the cross-combination table of their individual characteristics.

For the given decision domain there are about 20 parameters potentially relevant to different usages. Those may be as different as navigating a sensible automotive way in the streets, with upcoming traffic jam information being considered in a dynamic re-routing along the way, and trying to find small treasury caches somewhere in nature as recreative outdoor activity [34]. As the respective items tend to be relatively costly (ranges in autumn 2005 between about 200 and 1000 EUR) many users try to cover different application domains by choosing just one device, sensibly supporting potentially conflicting usage scenarios. This in turn individualizes the target function weighting to an always new set.

Decision support tool setup

The decision support system was set up as a (German) web service, to reach widely distributed people as best as possible. It was placed on a non-commercial, topic related website of one of the authors with sufficient traffic to be found by chance. Additionally it was announced in various fora concerned with recreational GPS use, to attract the largest number of people possible. As there is no chance of forwarding a detailed usage manual to the casual visitor, the entry page was set up as such (fig. 4.8).

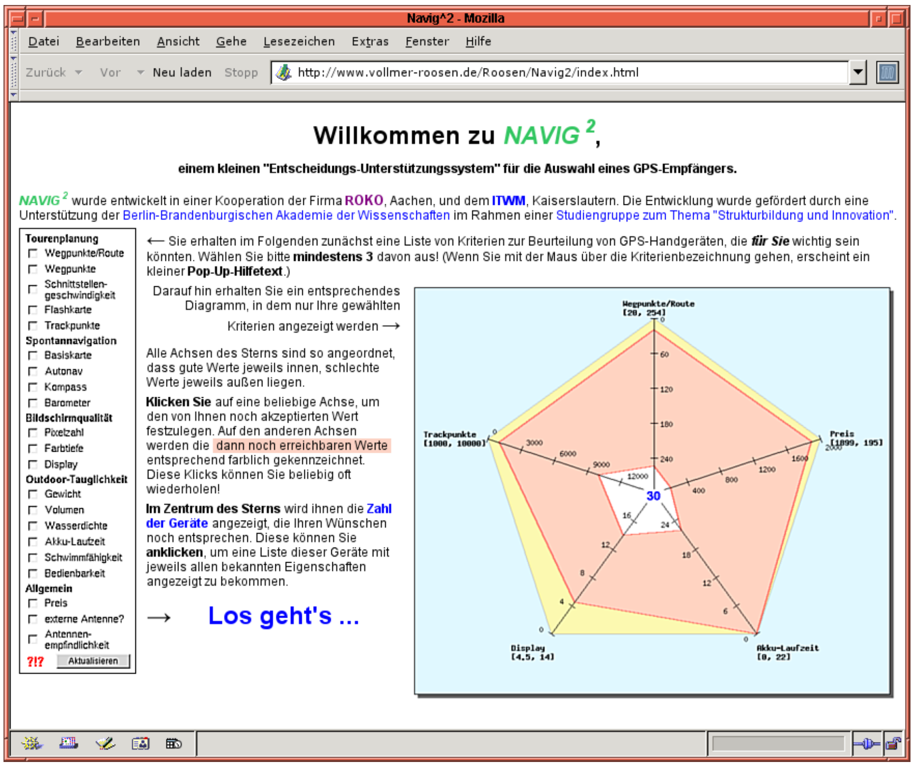

Abb. 4.8. Introductory page for the 'Pareto Star' GPS handheld receiver selection aid. A small explanation of the usage is given.

In a very brief explanation the user is informed how to choose his favorite target values and how to maneuver in the star plot. Finally a hint is given that the resulting list of complying receivers is available via the central number in the star, indicating how much of the basically available devices are still left.

Starting the application by clicking the appropriate link ("Los geht's") leads to the personal criteria definition page (fig. 4.9). Here the user is requested to select at least three criteria that he regards as important. For every criterion a short information box, shortly explaining what is evaluated and which possible values can appear for it, pops up if the user moves his mouse pointer on its name. After pressing the 'Aktualisieren' (update) button the first star plot is presented (fig. 4.10).

The axes of the plot are directed so that the better values always lie in the center of the star. (Non-orderable properties, like housing colors or geometries, cannot be handled within the scope of the tool presently.) The different colors of the plot areas indicate different things. On any axis there are transitions from light blue to yellow to red to white. The transition from light blue to yellow indicates an individually set limit of acceptance (see below) that the user tolerates for that criterion. Without any prior interaction that limit is automatically set to the lowest value the respective axis can

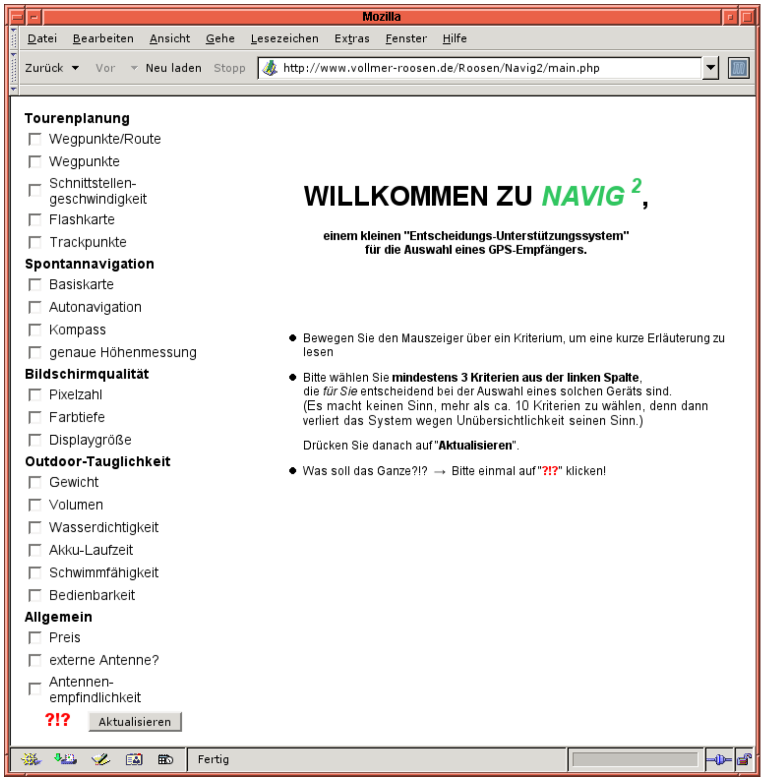

Abb. 4.9. When entering the GPS receiver selection aid the user is asked for his personally most important properties of the devices. A small step-by-step guide is presented to help him starting. By clicking on the blinking '?!?' he obtains information on the background of the service.

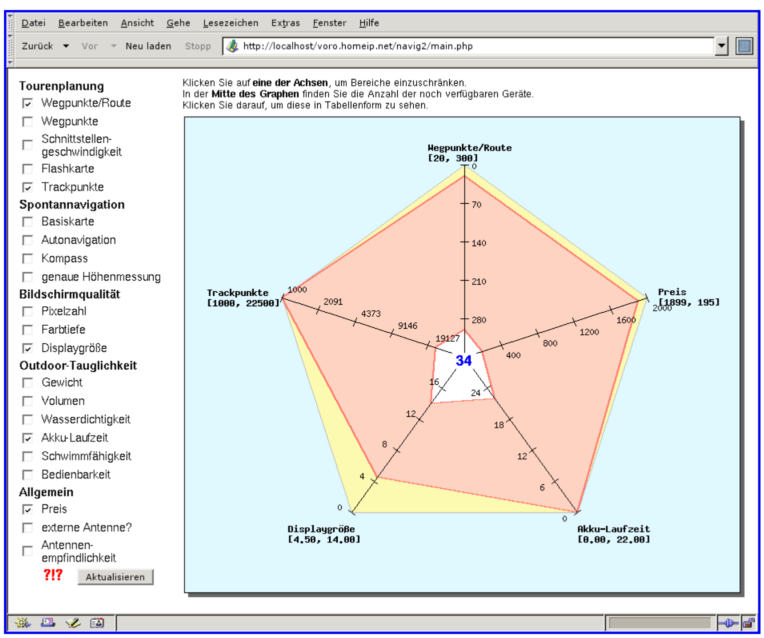

Abb. 4.10. After entering at least three personally relevant criteria. In this example the five criteria 'waypoints per route', 'track points', 'display size', 'accumulator runtime' and 'price' were chosen.

display for the given set of choices. The consecuting yellow part on any axis indicates the range of values that is not covered by the choices' list, without being limited by a user's action though.

The red part on the axis is the range that the real properties of the discussed GPS devices cover. As the display represents *all* devices not excluded by an individual acceptance limit, only the range for the whole set, not the individual properties of the devices, are displayed. The premise here is that the decider should concentrate on his basic individual demands, not on the concrete criteria of the devices.

The white range on any axis finally indicates a range that is not or no more accessible, either due to generally limited performance in the whole set of choices, or to a limitation that was defined for other criteria by the decider.

In the center of the star plot there is a thick blue number. It displays the number of choices left when the Pareto Star system honors the limitations set by the decider. For the GPS handheld devices there is an unlimited number of 34. The definition of a personal acceptance limit for any criterion selected beforehand will reduce the count of remaining devices respectively. The limits of the other target functions will change as well, both for the lower and the upper bound of the red range, as is evident when investigating the principally counteraction physical properties of (largest possible) display size and (smallest possible) device weight. Specifying a low value for the maximally accepted weight will reduce the maximum available screen sizes of the remaining items.

The point and click procedure on the accepted values, representing immediately the preferences of the the decision taker, is re-iterated for arbitrary axes until the number of remaining items is reduced to a tolerable quantity, quite often around 5 to 8. During this process the decision seeker may as well drop criteria from his selection as choose additional ones from the continuously displayed left criteria list column. Choosing an additional one will initialize another axis without a set boundary value, but obeying limitations on its criterion range imposed by set limitations on other criteria. After that the central number should be clicked to obtain a simple table of those devices, listing *all* their criteria properties (fig. 4.11).

Besides the potential display of the criteria description popups there are additional functions and bits of information in this list. First of all it may be sorted in ascending or descending order for any criterion by clicking on the respective small arrows on the left of the criterion's name. This supports the notion that usually there is a dominant criteron (mostly: the price) even if a multitude of relevant ones has been selected before.

Secondly the columns of the list are colored differently. Green columns indicate pareto-optimal items *with respect to the criteria set and the individual accepted ranges settings of the decision seeker*. It must be stressed that this pareto set is highly individual, as a change in the criteria list or alternative thresholds of acceptance may significantly change it. Colums colored white & blue indicate that the respective item is not a pareto set member but still complies with the imposed limits. In each non-pareto column the dominating devices are shown at the bottom of the table (row 'dominierende Geräte'), so the decider can immediately compare them with each other. The reason to include the non-pareto optimal devices in the list is the lack of a practical sharp distinction between relevant and irrelevant criteria in the selection process. Even if a criterion has not been included into the interactive Pareto Star selection that does not indicate that it's value is completely irrelevant. If, for a given device, such a factor becomes very favorable while the chosen 'important' ones strongly resemble each other it may as well serve as a secondary, hierarchically less stringent decision argument. This leads to a discussion on parity and hierarchy of optimization goals and will be revived in a later section.

Usage analysis

The decision support system was designed in order to monitor 'uneducated' users' reactions on the availability of an individually determined pareto set indicating filtering

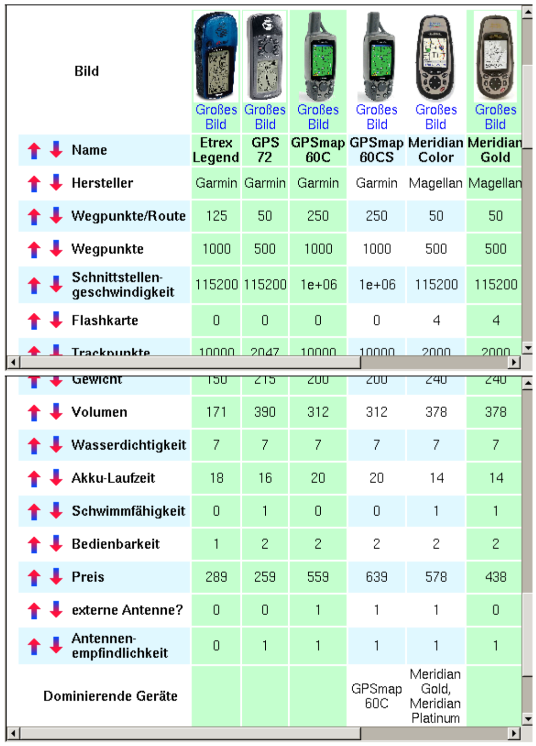

Bild	Großes Bild	Großes Bild	Großes Bild	Großes Bild	Großes Bild	Großes Bild
Name	Etrex Legend	GPS 72	GPSmap 60C	GPSmap 60CS	Meridian Color	Meridian Gold
Hersteller	Garmin	Garmin	Garmin	Garmin	Magellan	Magellan
Wegpunkte/Route	125	50	250	250	50	50
Wegpunkte	1000	500	1000	1000	500	500
Schnittstellen-geschwindigkeit	115200	115200	1e+06	1e+06	115200	115200
Flashkarte	0	0	0	0	4	4
Trackpunkte	10000	2047	10000	10000	2000	2000
Gewicht	150	215	200	200	240	240
Volumen	171	390	312	312	378	378
Wasserdichtigkeit	7	7	7	7	7	7
Akku-Laufzeit	18	16	20	20	14	14
Schwimmfähigkeit	0	1	0	0	1	1
Bedienbarkeit	1	2	2	2	2	2
Preis	289	259	559	639	578	438
externe Antenne?	0	0	1	1	1	0
Antennen-empfindlichkeit	0	1	1	1	1	1
Dominierende Geräte				GPSmap 60C	Meridian Gold, Meridian Platinum	

Abb. 4.11. Display of the properties' list of the remaining items, after reducing their number via the interaction with the star diagram. A combined image is shown because the computer screen is not large enough to contain the table as a whole.

system since this kind of decision domain is ubiquitous in normal life. To attract a sufficient number of users the tool was placed on a well-frequented web site dealing with the application of smaller GPS devices. By means of small introductory notes it was announced on several related web-based fora.

On the web site itself a visually quite dominant teaser (red, bold and blinking '?!?' click field, placed directly next to the main display refresh button) was placed to motivate users to express direct responses. If a user clicks it he becomes a short note on the background of the tool and a text entry field to submit his impressions on the tool. In addition to those (rather few) direct responses the usage of the system was monitored by analyzing the server log files over an interval of about 3.5 months.

About 30 direct responses were received. Half of them just helped correct some faulty decision data on the ever-so-changing specifications of the individual receivers (due to frequent firmware revisions that the manufacturers provide for their customers). Those responses indicated that the users observed the provided data with thorough scrutiny and were eager to clean up the data base in the sense of a community effort. The other half stated that they thought the tool really helped deciding on a device to be bought. Some of those responders remarked, though, that they already bought their devices prior to finding the decision support tool. The retrospection of their purchase decision was mixed: Some stated they might have decided otherwise if they had known about the Pareto Star beforehand, some reported an affirmation of their stomach based former decision. No response discussed the tool and the method by itself, though. Obviously the content aspect was too dominant to focus on the methodology.

In order to obtain more of those methodological data the server log file have been analyzed. Due to the setup of the service as LAMP system (containing the components Linux, Apache, PHP, MySQL) and the input data submission mostly via the http GET method most of the user settings with respect to choice and number of criteria could be reconstructed (table 4.1). The total number of 2600 visitors should provide an adequate sample size to consider the data as trustworthy.

Number of visitors (total)	2600
⟶ number per day	24
Star interactions per visit	2.7
Result table requests per visit	1.8
Mean number of criteria chosen for the star diagram	5.5

Tabelle 4.1. Statistical access values for the Pareto Star decision support system for GPS handheld receivers.

The anticipated essence of the tool — the toying with the star diagram — was used rather sparsely with the averaged 2.7 interactions per visit. Here a definitely larger number was anticipated. There is an explanation, however, in the supposition that the users tend to be somewhat result-oriented (in the sense that they want to decide on a concrete device) and that the playing around with the tool did not satisfy their immediate curiosity just *which* devices were still in the race and which had already been left behind.
Another point may be the non-familiarity of the users with the density distribution of the individual criteria. If a scale indicates that a large range of values is available it does not show that perhaps 95 % of the devices are clumped in a (possibly non-favorable) small part of the span. This soon leads to rather sparse result lists. The average number of less than two result table requests per visit once again shows that the users do not tend to play with the tool, but instead are very target-oriented. After receiving a result table they rather seldom reiterate the star plot selection process with changed criteria settings.
This in turn may be interpreted as result of economizing the selection process in the sense of a frugal decision method: The users seem to obtain a convincing final selection list by just focusing on the first result table. The fact that usually familiar brands and types are listed will enhance that notion, once again supported by the Bounded Rationality view of typical human selection methods, here 'take the best known'.
These interpretations, derived from the average system's usage, do not hold for some very eager users who indeed played around quite intensely with the tool, reaching well above ten or fifteen interactions with the star diagram and a respective number of result lists.

Real-world applications

The developers presented the tool to several hardware and service providers in very different application domains, like used automobiles reselling, real estate brokerage, recreational accomodation services, or last minute flight ticketing. In almost every discussion the tool was honored as a very interesting method of information retrieval and assessment. Nevertheless the tool was not used for a respective innovative information preparation because a conflict of interests was regularly perceived. This may well be explained by discussing a potential application in the automotive reselling business.
For an internet-based vehicle reseller the usual business model is to provide a presentation platform for an arbitrary number of private vendors that publish their offers on it. For larger sites well above a million vehicles are offered at a given time. So even if the list size of criteria usually deemed relevant is estimated to five or six there are lots of offers not nearly approaching pareto optimality in the sense of the potential purchasers. If the trade platform offers a method to easily identify the most interesting ones in the sense of constructing a pareto front, most offers would never (or at least only very rarely) be shown. Vendors intending to use the platform would turn away from using it. This in

turn conflicts the business model of the web site provider who would be the one to offer the additional pareto set filtering service — hence the aversion to put such a tool to work. It may be guessed that most potential vendors have a certain kind of qualitative notion about the competitiveness of their offers, so the availability of a selection tool would shy them away to other platforms *not* offering it.
This situation holds for most points of electronic sales, therefore it is not to be expected that this methodology of multi-criterial decision support will find its penetration into a broader usage. More promising fields of application are areas where all participating players do have a sincere interest in an objective target criteria assessment. One of them is presumably the consumer consulting service that does not earn its money by selling items but with offering the best possible objectivity in consumer information requests.

4.2.1 Refinement of the selection tool

The experiences with the usage behaviour led to the conclusion to redesign the tool. It should both familiarize people with a multi-criterial selection process support and better meet their expectations with respect to their desire to obtain almost immediate practical suggestions. The new 'Pareto Star' system is consequently not a star plot diagram any more, but rather a semi-hierarchical table of bar charts with similar, but enhanced colorization ranges relative to the existing implementation.
While the final result table display will not be changed very much, the multi-criterial selection process is significantly changed towards a faster 'real results' display, partially implying a partially hierarchical expectation concept. Starting with an empty list of selection-relevant criteria the user will be asked to choose a first criterion (fig. 4.12).
After selection of one or several personally relevant criteria the result bars (fig. 4.12, left side) display a differentiated colourization (see below). A set of adequate receivers is presented on the right side of the interaction window, being background-colored in light orange to display a Pareto set membership as in the older version of the Pareto star system final result table.
The order of parameter bars on the left side is treated hierarchically, with the importance of the criteria decreasing from top to bottom. Accordingly the order of actual receiver displays is sorted in a strictly hierarchical way. Receivers exhibiting identical settings for the most important goal are sorted by the second important one and so on. In case of an additionally chosen criterion the list on the right is resorted accordingly. It is as well possible to switch places and hence the order of importance for the selected target criteria, with a respective effect on the order of the actual receiver display.
The graphical bars describing the single criteria contain additional information relative to the Pareto Star. Favourable settings are always on the right side of the bars, so the user will tend to click into the bar to limit the respective value with increasing rigidity from left to right. His personal limit is indicated by a thick black vertical bar in the horizontal criterion distribution, with the chosen value being displayed on top of it. On the left of that marker there is a white background. On the right usually a light green range is situated, indicating an interval that does not contain items with respective values. A light blue range follows, hinting towards actually available items. Contrary to the Pareto Star an additional background colored range follows: The red section indicates the interval of target value settings that will lead to a complete emptiness of the receiver list. This range is influenced by the user's limit settings in other criteria, due to the fact that those restrict the available choice in others.
In the former Pareto star system the user did not receive information about the density and distribution of the criteria values independent of his personal choices. Therefore an additional piece of information was introduced in the new hierarchical system: Small black dots in the center of the horizontal bar indicate the actually available values in the underlying database.

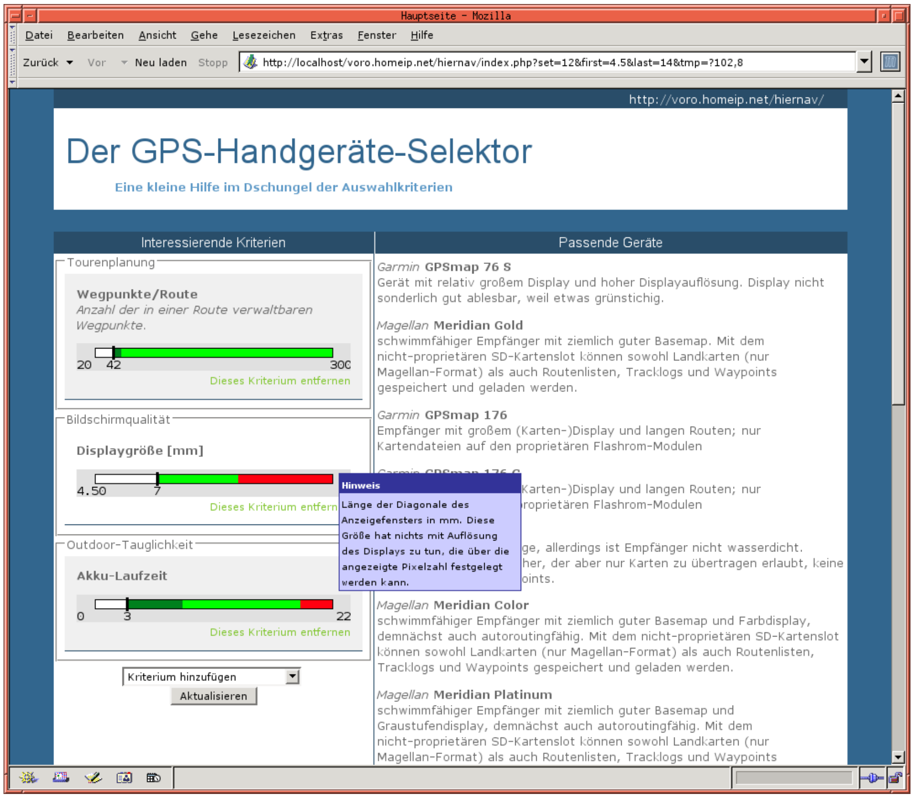

Abb. 4.12. New multi-criterial decision support system with instantaneous semi-hierarchical result display. (The visual appearance of the interface itself represents an intermediate development phase.)

In the end the user will choose to display the resulting items table that is presented in a similar fashion as in the presently available Pareto Star system.
Finally, the new system contains visual procedures to work on discrete and unsortable list criteria (fig. 4.13). Usually there are values that cannot be ordered by their values, like the color of a car body, or the selection of edge shapes for a chair or a desk. Here the user will be able to define his or her personal preferences, with the displayed order of potential values reordered upon each new change of choice, separating the desired ones from the undesired. As soon as this selection system is in a stable condition it will replace the present Pareto Star, and its availability will be announced once again in the respective fora to attract new users.

Abb. 4.13. Treatment of discrete or non-sortable target functions in the new semi-hierarchical selection support tool. The choice of background colouring in drop-down menus is chosen according to the definition in sortable parameters. The order of entries can change due to prior selections and implied restrictions in other criteria.

4.3 Molecular Biochemistry: RNA structural optimization on large neutral nets (Schuster)

Explain neutral nets and their impact on optimizing respective systems. Show differences for the different kinds of neutrality.

4.4 Energy Engineering: Multicriterial residential heating design (Roosen)

(What target aspects are there for heating of rooms? Dependence of usage characteristics on the choice of the heating system. Overlapping pareto sets of several base configurations. Distinction of easily changeable parameters versus constructive parameters in optimality striving.)

Problem definition and scope

Developing innovative space heating concepts necessitates consideration of various factors influencing the pragmatic target function ‘warm, comfortable housing space’. Additional goals of the individual user must be considered, e.g. the accepted energy consumption, the costs conjoined with that, the flexibility of utilization, and many more. Besides the purely technical development of a respective heating system it is thus necessary to identify a suitable application niche relative to already established systems: It is not to be expected that a new design will outperform existing ones in every respect. There are too many application profiles and boundary conditions for the requested space heating demands. Prior to an extensive development of a novel solution the primary advantages need to be identified.

For every heating system there are several configuration parameters (like area proportions, insulation thicknesses, temperature niveaus, operation control concepts) exhibiting specific effects on the usually contradicting pragmatic target functions, like overall costs, cosyness primary energy consumption etc. . The targets are affected differently, depending on the foreseen usage scenarios (like constant usage of a flat vs. modulated usage of a utility building vs. stochastic usage of an event hall) clothing habits (strict business dress code vs. casual wear) and unchangeable parameters for a given building like window area ratio. Therefore the optimal system configuration has to be determined individually with respect to the potential relative target function ratios.

Varying the relative target function weighting and monitoring resulting major system parameter changes, like a switching of the basic heating system type, yields a notion on *decision stability*. A stable decision in this sense is a constant selection of one certain system type even in the light of slightly changing subjective target function weightings.

The method of choice in this paper for identifying dominant and stable ranges of a heating system is the pareto optimiziation. The set of optimal solutions for two or more partial targets, the pareto set, is determined, covering all possible relative weightings of the considered partial targets. Due to the fact that the individual target function value is calculated independently from any other and not traded against them there is no need to require an a-priori weighting. The targets are instead pursued simultaneously and independently.

In the scope of this exemplary study the following **target values**:

- a comfort measure
- the primary energy consumption

for the **simulated space heating variants**:

- gas fired floor heating
- gas fired wall radiator heating
- dynamic planar electric heating

are considered. The analysis is carried out for three **usage scenarios** and a floor area of 78 m^2:

- Continuosly used flat (living room, sleeping room, kitchen, hall) with differenciated usage profiles and respective heat gains in the different rooms

- Same flat with same relative usage profiles, but pure weekend usage (Friday from 18:00 h to Sunday 18:00 h)
- open-plan office with same ground area, 15 computer work places and flexitime usage (workstart in the interval 7:30 h to 8:30 h, combined with respective distributed work end in the afternoon).

The diversity of usages and the respective inner gains with their varying amounts and temporal distributions will show the differences in heating systems performance and their selective advantages for those tasks.

Simulation and Optimization

The aim of the reported work was to identify effects of specific usage scenarios and boundary conditions on the optimality of the target functions 'energy consumption' and 'comfort'. The desire to obtain significance for practical purposes lead to the conclusion that the underlying system simulation be as realistic as possible. Several energetic simulation systems, each cumulating several tens years of development times, are suitable for this purpose. Taking such a complete simulation package usually leads to the fact, though, that there is no possibility to interact with inner structures of the simulative calculations from the outside, i.e. the user interface. It is rather necessary to treat the simulation package as a whole as a kind of 'black-box' performer that may respond to parameterized simulation task definitions with its specific way of result response. The response, typically made available in the form of files, has to be scanned and parsed for significant target function information by the optimization procedure. A re-implementation of a simulation core is usually prohibitive due to the complexity of such a task.
The simulation system ENERGYPLUS (e^+) [22], provided by the US-american Department of Energy, seems especially suitable for optimizing the energetic and comfort behaviour of residential space heating systems. It combines a very great detailedness of simulated effects with a relatively simple system specification language in input files that may be artifically created and/or modified by a custom optimization module. A simulation task may be invoked by a command line call, referencing a respective pure ASCII input file of some 200 to 300 kB in size. The input file containes the individual settings of (a part of) the building to be simulated, like geometries, geographic orientation, wall structures, usage profiles, efficiency values for heating devices and such. They are defined by comma-separated lists depending in their structures on precursory keywords, relating to each other. Simulations are performed for real weather data sets defined with respect of temperature, humidity, and irradition data for numerous cities of the world. Consequently solar irradiation gains through windows are considered in detail, with their geographic orientation being taken into account. We chose a weather data set of Düsseldorf/Germany for the optimization calculations documented below.
Some configuration parameters require a somewhat detailed intermediate treatment in transferring their settings from optimization to simulation module. As an example, the shift in an onset time of a heater needs to be reformulated into a complete rewritten 24 hourse schema in order to make e^+ recognize the change. For this task the macro preprocessor m4 [71] was used. It creates the actual simulation input file with valid 24 h schemata by changing a template containing respective macros.
Upon request e^+ creates very detailed //Lastgänge//, energy flux reports etc., and also time-resolved comfort calculations according to the Fanger model [21] that are written into respective output files. In order to provide the optimization module with the necessary target function values for each previously defined system setting the output files have to be pre-processed in an appropriate fashion.
The interpretation of the Fanger comfort value provided by e^+ needs some additional comments. In contrast to the discrete model originally devised by Fanger the comfort

value returned by the simulator is a continuous one. But the general idea is preserved by modeling coldness discomfort notions as negative, hotness discomforts as positive values. In order to attribute comfort values to heating system configurations the hour-wise and room-resolved absolute Fanger values are averaged after weighting them with the person's density. Hence the goal of the optimization process is to strive for a positive bound value of zero as comfort measure. Non-declinable periods of overheating during summertime limit the lower bound of it to values well above zero, though, as the investigated spaces are only equipped with heating, not climatization devices. Underheating and overheating periods in unpopulated rooms do not contribute to discomfort as there is no person to notice that discomfort.

The inhabitants' comfort assessment is closely coupled with their (simulated) clothing thickness. Changes in the range from light summer casual wear to thick winter pullovers yield significant differences in evaluation in otherwise identical thermal environment conditions. For practical purposes it is therefore not sensible to assume constant clothing conditions: It is definitely more realistic to assume the typical dynamical behaviour of the inhabitants to partially undress or add pieces of clothing within the respective socially controlled limits if slight notions of discomfort are experienced. Such an adaptive clothing schema is not feasible in e^+, though. Here a fixed hourly schema for a set clothing thickness is to be provided.

To achieve a practically realistic statement a calculational trick has been implemented. The number of inhabitants of every room is halved, with one half being clad as light as possible and the other half as thick as possible. As the ideal comfort value is zero, a change in the sign of the Fanger value for the two halves is taken as an indicator that there exists a clothing thickness in the foreseen range providing maximum comfort. Accordingly, for that interval the practical comfort target function value is set to zero, designating the optimal one in the positive definite range of comfort assessment absolute values. If the signs of both halves are same, either the thinnest clothing is still to warm or the thickest is still not warm enough to obtain the maximum comfort. In that case the lowest absolute value will be taken as target value, assuming that the inhabitants will adopt their clothing to the nearest socially acceptable level. The latter changes significantly in context with the simulated scenario: The variation span for private living conditions is definitely larger than the one for the bureau business condition.

Since the simulator's internal organisation and the functional structure of the target function in its dependence on the available configuration parameters remain hidden to the user's scrutiny, the optimization module that is to cooperate with the simulator should react as insensitive as possible against non-differentiabilities and unsteadiness. Therefore a custom developed evolutionary optimization system was put to work [57]. It provides a dynamic list population management, a Pareto set oriented multicriterial assessment schema and could quite simply be parallelized with respect to the invoked simulator runs by using the PVM system [54, 31]. This simple parallel execution of up to seven simulation runs on separate TCP/IP network coupled computers was sufficient in the context of the given task as is shown in fig. 4.14. Without changing the relatively inflexible generation paradigm the dead times imposed by idling processors at the end of a generation's simulation runs do not add up and reduce the overall calculational speed too much: The ratio $1/2 \cdot CPUcount/individualpergeneration$, representing the statistical idle time per generation, is simply too small. The calculation time of one simulation run is typically about 40 sec (simulator machines' performance category: Celeron 1 GHz). Accordingly the speed loss due to necessary information transfer times was neglegible for the less than 1 kB communication requirements per simulation invocation. Even when using more powerful processors the described concept will still be favourable and applicable without larger performance losses.

Abb. 4.14. CPU load curves of four of the cluster of six computers working in parallel. The dead times of individual units appear statistically during the closure runs of each generation.

Heating system comparison for different usage szenarios

All scenarios were optimized for the same basic boundary conditions of floor area and geographic orientation. Furtheron it is assumed that identical usage profiles apply for adjacent units above and below the calculated one, so no net heat transfer is calculated via floor and ceiling. The residential units are simulated as four zone models (see fig. 4.15: living room, sleeping room, kitchen, hallway without heating). The bureau is rendered as one zone (open-plan office).

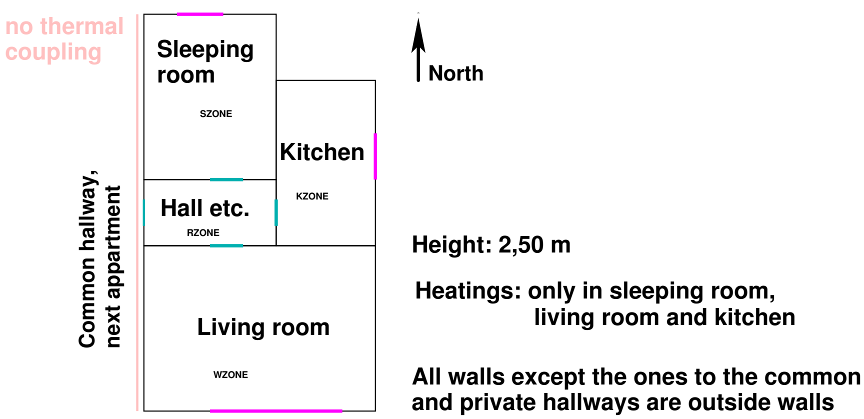

Abb. 4.15. Structure of the residential units, both continuous usage and weekend use scenario. There is no heat exchange via floor, ceiling, and adiabatic walls to next flat on same floor.

According to the different usage profiles some hourly schemas were adapted analogously: The forced venting by outside air ('infiltration') was significantly reduced in times of absence. The temperature schema was accordingly adapted for bureau (Mo – Fr) and weekend flat (Fr – Su) usage. The temperature settings for nightly drawdowns were applied during the other times.
Since the different heating systems exhibit very different heating rates — with the floor based heating being known as exceptionally slow — different pre-heating times have been introduced as optimizable parameters. So competitive comfort values can be reached, possibly conjoined with significantly increased energetic inputs, though.
The following configuration parameters were changed within sensible limits:

Technical section

Thickness of inner walls. This value describes the the thickness of the walls between adjacent rooms inside the calculation region. Since massive walls act as heat buffers this value has direct influence on the temperature characteristics resulting from short-term stochastic heat flux changes (solar irradiation through windows, ventilation events).

Thickness of wall insulation. The outer face of the facade is equipped with a respective heat insulation, interpreted as a polyurethane layer.

Maximum heating power. This value limits the energy output of the heating system, limiting the energy consumption as well influencing the comfort.

Heat transfer coefficient of the radiators. This value only exists for hydraulically operated heating systems and describes the heat transfer coefficients (UA) for the individually simulated rooms. In the case of floor heating it describes the total length of heating tubes.

Temperature settings (regarded as constant heating control system targets over the whole simulated period)

Daytime temperatures in individual rooms. Temperature targets of living room and kitchen (together) and sleeping room (separate) are specified. The hallway is only passively heated by its energetic coupling to the other rooms. Living room/kitchen presets are interpreted as setting for the bureau one zone modelling.

Nighttime temperatures in individual rooms. Same specifications as for daytime temperatures, but given for the drawdown times. Same treatment in the one zone model, like with daytime temperatures.

Begin of daytime operation. This is, like the other temperature settings, interpreted as time of day at which the heating is switched to daytime operation; living root and kitchen.

Stop of daytime operation. Time of day when nighttime operation starts.

Preheating period. This period denominates the advance switch-on interval for the weekend usage simulation, to achieve a sufficient comfort level at the time of personal presences.

Absence temperature. Prevention of freezing and the required preheating period upon arrival is influenced by this temperature being set during times of absence .

So all scenarios contain at least 10 configuration parameters that are subject to optimization.

Continuously inhabited flat.

The continuously inhabited flat accomodates a varying number of people. Following the assumption that the sleeping room is practically uninhabited during daytime, while the kitchen and the living room is non used during night time, there are periods of potential temperature lowering and hence energy consumption reduction without loss of comfort. Depending on the characteristics of the heating supply the rooms react differently if one tries to put this saving potential to work. The simulation periods contain two weeks per season.

First of all, fig. 4.16 shows clearly differing Pareto fronts for the individual systems. While floor heating and EDDH develop continuously bent curves the radiator heating seems to exhibit only piecewise steady ones that strongly jumps at a Fanger value of about 1. This jump coincides with respective error messages of the simulator that is no more comprehendable from the furtheron produced numerical results. The very small energy inputs observed in that range of the Pareto front seemingly lead to a non-converged mode of the simulation. Therefore the wall radiator results should only be considered at Fanger values less than one. This first qualitative result should be taken as a rather important issue in simulation-based numerical optimization: The optimization process itself may drive the simulator into parametric regions where it defects and deliveres unreliable results without the tutoring human to notice it. So, in complex and unclear situations additional inspection facilities beyond the pure target function values

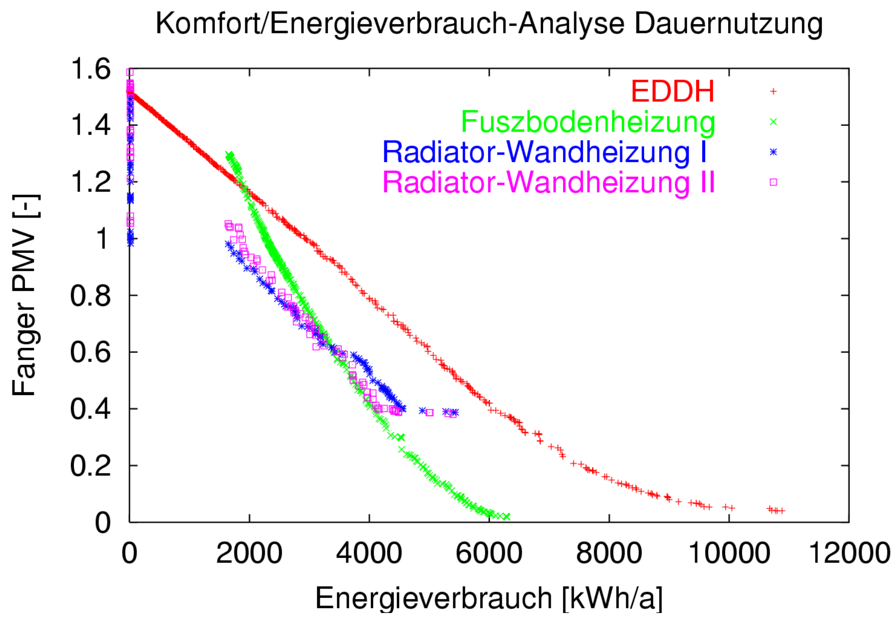

Abb. 4.16. Comparison of the Pareto fronts of three space heating systems with respect to comfort and primary energy consumption, under the condition of a continuous but differentiated usage of the available rooms. For discussion see text.

should be provided to help the supervisor classify the conciseness of the calculated results.

Furtheron the figure shows two separate wall radiator optimization run result sets to illuminate the statistical behaviour of results. Both runs yield the ’ideal’ Pareto front in varying parts and stay inferior in others. Here the phenomenon of optimizational stagnation is evident: with evolutionary optimization there is no proof that the optimum really has been approximated. Repeated optimizations are mandatory, and even then there is no security on how far the real optimum frontier is still away, even if in most cases some asymptotic behaviour is likely.

In the continuously inhabited flat the unpopulated intervals in the different rooms seem to be too short to bring in some observable primary energy consumption effect from the rather rapid and dynamically operating EDDH, or the wall radiation system in a reduced amount, relative to slow floor heating appliance. Even though the *absolute* EDDH energy input (not shown in the diagram) is less than those of the two competing systems, the conversion losses in the production of the electrical energy (assumed effectivity of 40 %) leads to a primary energy related domination of the EDDH over the whole investigated range. The flat seems to be kept on a kind of quasi-static temperature level. The general, mean heat losses through the outside walls and windows are not evened out by inner gains and dominate the summed-up consumption picture.

Flat with weekend usage.

The partitioning of the weekend flat is identical to the continuously inhabited one. During its times of usage the same inner gains are assumed, but only from Friday 18:00 h to Sunday 18:00 h. During the rest of the week all temperatures are set to the nighttime operation level, the forced ventilation is set to a very low rate, and no inner gains are assumed. No comfort rating is performed as long as no inhabitant is present. As the flat is severely chilled out just before the weekend inhabitance starts again a freely choosable preheat time is provided and subject to optimization.

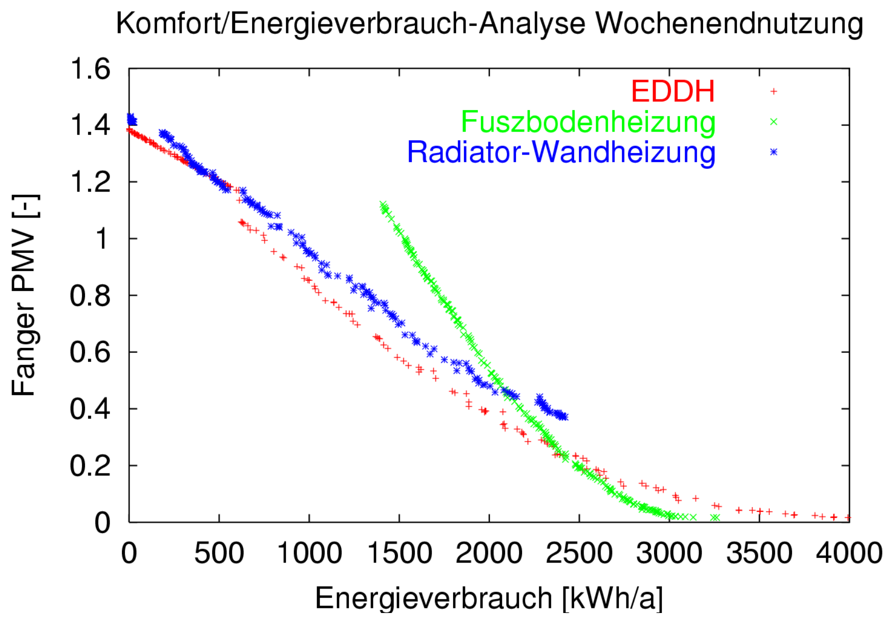

Abb. 4.17. Comparison of the Pareto fronts of three space heating systems with respect to comfort and primary energy consumption, under the condition of a weekend usage. See text for discussion.

Since the few usage periods in the simulated observation times of the continuosly used room systems would presumably lead to erroneous results the complete simulation time was raised to 200 days per year, again assuming the weather reference data of Düsseldorf, Germany, provided by the e^+ system. This adaptation yields a number of comfort-assessed days comparable with the continuous usage situations. Nevertheless it is important to consider the intermediate, unpopulated days as well: Their settings influence both primary energy consumption and achievable comfort levels during inhabited periods. In the weekend usage mode the electro-dynamical direct space heating concept can show its advantages: It shows a leading position in most parts of the Pareto tradeoff curve and surpasses the radiator solution in any case. Only in the high-comfort (but accordingly also high energy input) domain the floor heating concept is slightly superior. In the Pareto range around 1500 kW/a most EDDH parameter sets show a very short preheat time prior to the 18 h arrival time of the inhabitants (17 h and later). This suffices due to the small thermic masses that need to be warmed up to obtain reasonable comfort values.

Bureau situation.

Relatively large inner gains are observed in the bureau situation with the same floor area as the residential usage: An occupation density of 15 computer workplaces with the system switched on during normal bureau times is assumed. Accordingly a remarkable (dis-)comfort baseline is observed in the Pareto sets of all heating concepts (fig. 4.18). The heat produced by inner gains cannot be eliminated during summer days since no climatization, but only ventilation is assumed and simulated to preserve a certain level of comparativity. In addition, in the bureau environment the clothing order is somewhat more restricted with respect to very light clothing, leading also to a raised level of discomfort during hot summer days.
The relative positioning of the heating systems is somewhat similar to the results for the weekend used flat: While the floor heating concept dominates the range of high comfort

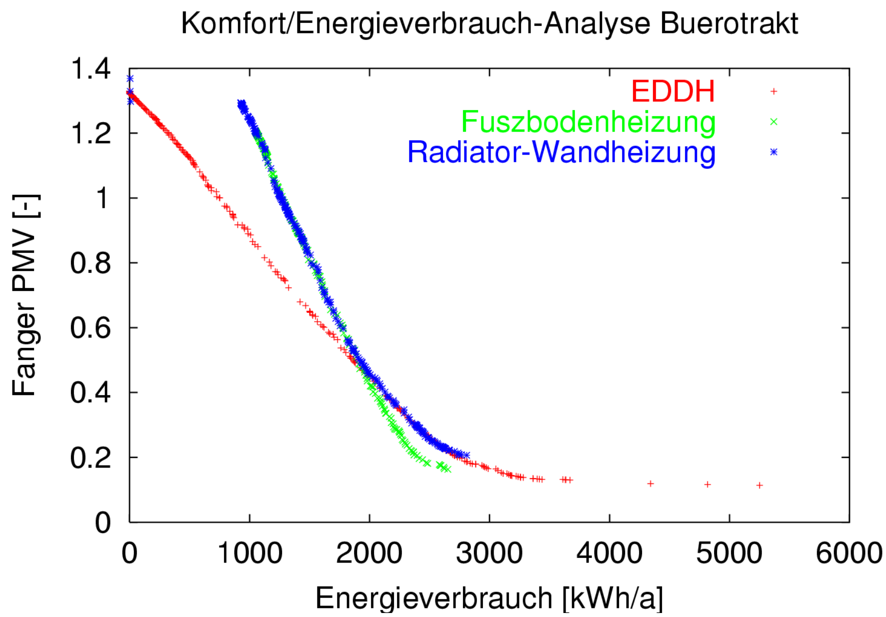

Abb. 4.18. Comparison of the Pareto fronts of three space heating systems with respect to comfort and primary energy consumption, under the condition of a bureau usage. See text for discussion.

(and high energy consumption) the low energy input (and combined with that, the low comfort) range is the domain of the EDDH. The radiator wall heating concepts exhibits an interesting distribution: Its Pareto set mostly conincides with the dominated parts of the other two concepts, without ever being a leader itself.

Interpretation of results

The results allow some insights into the potential, but also the limitations, of space heating system optimization. First of all, the anticipated outcome is confirmed. 'There is no such thing as a free lunch' — or in other words: For each system in question one has to balance comfort requirements with an appropriate (primary) energetical input, although the amount of the latter may be systematically different across heat provision system types. Each heating system produces its own Pareto front in the comfort vs. energy input plots.

There is no general leader system completely dominating other existing concepts. Depending on the structure of the heating request — mainly continuous, intermediate with relatively long on/off periods, or short periodic changes as three example situations — the advantages and disadvantages of the supply systems become visible.

In a continuously populated residence the determining factor for primary energy consumption is the heat loss through walls and windows that has to be balanced by almost steady heat input during phases of cold weather. Here the overall conversion efficiency of the heat provision system plays the most important role for the primary energy input calculation. Since the production of the very high exergy carrier medium 'electricity' only coincides with relatively high conversion losses, this in turn leads to a general energetical inferiority of ohmic heat generation when same levels of comfort are to be reached. In the continuously used flat there are no major dynamical effects the highly adaptable electric heat production can utilize to lessen this general disadvantage.

The weekend use of a flat is almost a complete antithesis. Long chillout periods during the absence of the residents lead to the necessity to heat up relatively large volumes of thermically capacious, heat absorbing material. Here the established systems of radiators

or floor heating need long preheat times in order to get the living space itself sufficiently cozy. After the inhabitants once again leave the flat, the heat energy mainly introduced into the building's walls slowly dissipates into the environment without any further benefit. Contrasting to this behaviour the EDDH system mainly heats this space itself, leaving the surrounding walls mainly in their cold states during the rather short usage periods. So the total input heat energy per weekend is significantly smaller than with the other systems which overbalances the conversion losses.
The slow start profile of the bureau scenario, due to the slow crowding of the workplace according to the simulated flexitime usage, once again seems to promote the relatively slow floor heating system, at least for sufficiently high comfort levels. The dynamic effect that would presumingly promote the EDDH seems not large enough to overcompensate the general electrical conversion loss on the primary energy target function.
Although these results are quite convincing and can well be interpreted in terms of general scientific concepts, they rely strongly on the modeling input flexibility of providing the simulation model with practically relevant detailedness — and therefore on the effort in setting up the model. Some of those issues shall be scrutinized in more detail in the next paragraphs.
One major aspect of an EDDH application is the *very* fast response of the comfort value on any control action performed to the actual heat input into the room. This response is in the order of minutes, while the simulation system will only yield reasonable comfort values on the basis of 10 minute steps. A faint approximation is the assumption that each change in population of the simulated space, and any change of inner gains, just coincides with full hours for which the time schedules must be defined for the e^+ simulator. A more detailed simulation would require both many more time steps and a simulation system explicitly taking into account dynamic heat distribution processes in the surrounding walls. This in turn would multiply the required simulation time which was already at its limit for the available hardware.
e^+ provides some automatisms for adapting the actually input heating energy into the room, as well as for the venting in case of overheating etc.. Although these procedures are already quite elaborate with respect to realistic scenario modeling, it does not suffice to describe pragmatic personal actions. Just to give some examples, what is *not* sufficiently described: i) When people populate a flat the doors between rooms tend to be left open in case of rather similar temperature distributions while they are usually closed more rapidly if temperature gradients occur. ii) The comfort feeling in a surrounding with a temperature gradient — both in time and space — is not covered. Even if people feel slightly uncomfortable due to a slightly too cold surrouding this feeling is somewhat compensated if the surrounding tends to get warmer. iii) If there are different temperatures in different rooms a person wanders frequently between the sensation will be different compared to his stay in just one of them. If a room is used only for short times (like some minutes in each case), but this rather often, a discomfort value attributed to this room will be in practice less perceived in comparison to a person staying in that environment for a longer time.
There is one additional aspect that cannot be covered even by the best possible situation-aware modeling: The potentially changing room assignment in a given complete space. Exchanging a bedroom against a living room, or splitting a large living room into smaller subsections that become somewhat autonomous in their usages, changing cooking behaviour — like switching from conventional cooking with large inner heat gains to microwave ovens with almost no external heat production, or exchange of slow bureau computers to faster but more energy demanding ones: Those usage changes cannot be foreseen, so another pragmatic optimization target should be kept in mind as very important for most cases, the usage flexibility.
Trying to integrate this aspect leads to the frontier of modeling complexity and the possibility to interpret obtained results, as there is no objective method in sight presently to economically (with respect to modeling setup) take this into consideration. The only

resort that may adress this problem might be the qualitative comparison of the magnitude of calculated scenario differences and their attribution to the probability of happening scenario switches.

4.5 Chemical Engineering: Optimal development of a chemical production site (Schembecker)

Describe the various aspects to be considered in the practical development of a production site. Show that different goals - not only technical ones - exist. Explain that important decisions have to be taken on rather poor informational data.

4.6 Civil Engineering: Structural Design and Process Shaping in Consideration of Uncertainty Phenomena

Dietrich Hartmann

(Demonstrate different kinds of uncertainties. Show their effects on the prediction of deconstruction processes. Show effect on their optimization. Is there a theoretical and/or practical limit of predictability?)

Introduction

Structural design of contemporary systems in civil engineering, as well as the shaping of the processes associated with such systems, is characterized by an increasing comlexity. This constitutes a big challenge to engineers which can only be mastered by means of rigorous computer-aided approaches. Customarily, engineers also attempt, to figure out what are the best solutions for their systems and processes. It comes, therefore, not as a surprise that, to obtain practicable results, the computer models established must to represent largely unaltered real world scenarios. Needless to say, that this general demand holds for many other scientific diciplines beside engineering. Thus, working solution concepts can be analogously transformed to other domains.
Due to the complexity, however, computationally intensive simulations are indispensable, with regard to the number of simulations and for the purpose of verification and validation. Not until that, a systematic evaluation of the simulations may allow for optimized systems or processes. To this end, the optimization or simply the amelioration of initial proposals is measured with respect to desired objectives, serving as a quality criteria subject to prescribed constraints. The high complexity additionally calls for multidisciplinary solutions and necessitates multi-level, multi-scale and even multi-paradigm models, leading to micro-, meso-, macro- and/or super-models interwoven to each other.
In the following, three different examples are dealt with which are to demonstrate how the handling of uncertainty in complex structural engineering problems can be accomplished, according to the general discussion in section 4.4, on page 32 ff. The first example is concerned with the lifespan-oriented design of large scale structures subject to time-variant stochastic loading, taking into account deteriorations. In the second example, the computer-aided destruction and collapse of complex structural systems using controlled explosives is discussed with respect to the realization of fuzzy randomness. Finally, the third example addresses the structural optimization taking into account stochastic imperfection in the geometry induced during the erection process.

Lifespan-oriented design

The lifespan-oriented design of structural systems in civil engineering is in the focus of a Collaborative Research Center (SFB 398), funded by the German Research Association (DFG). In this center, an own specific research project among others is dealing with the optimum design of steel structures accounting for the deteriorations and damages induced during the utilization of a structural system. This is to say that the design problem is transformed into an equivalent structural optimization problem.
For steel structures, in particular, the design against failures due to fatigue phenomena often becomes crucial. To give an example, industry halls composed of frames and used as a first reference system may fail under repetitive, stochastically multiply correlated as well as in stationary wind actions.
A further example are steel arched bridges (Langer's beams) which are considered as the second type of a reference system.

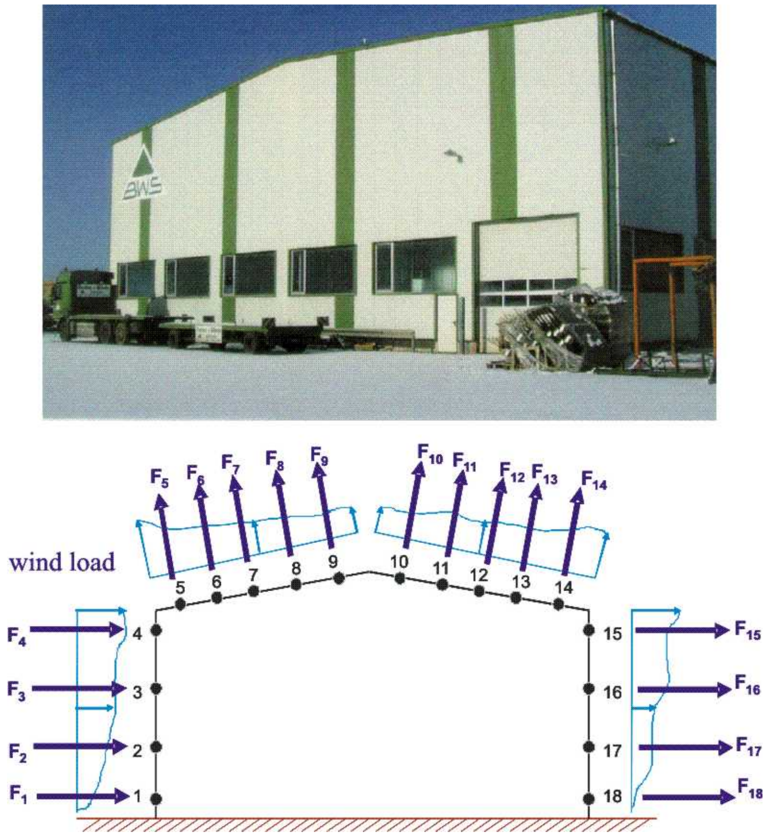

Abb. 4.19. Industry hall subjected to wind loading

Abb. 4.20. Steel arched bridges

Surprisingly, it has been assessed that even newly erected bridges show failures in the hanger connection plates linking the vertical arch hangers with the main girder of the bridge.

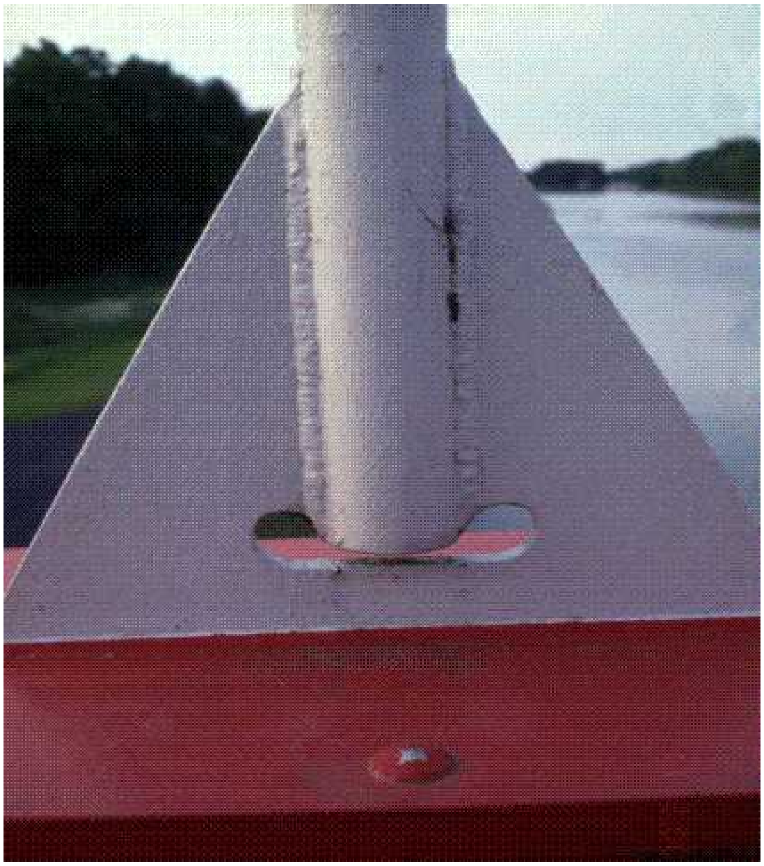

Abb. 4.21. Hanger connection plate

Extensive research and testing have figured out that mostly vortex induced across vibrations caute the damage (cracks in the plate), not rarely shortly after the erection. Again, the randomness of the wind actions, even if stationarity is prevailing, affects the damage-sensitive vibrations within a so-called lock-in domain.
From these remarks it can be concluded that the lifespan-oriented structural analysis and design is substantially governed through a plenitude of uncertainties in the stochastic load processes (actions) with regard to appearance, duration and intensities. Accordingly, the structural response in terms of displacements, vibrations, strains and stresses have to be regarded as time-variant stochastic processes themselves yielding stochastic deteriorations and a degradation in the total system. Of course, the relevant structural data have also to be treated stochastically in terms of basic variables (simple stochastic variables).
To elucidate, in principle, how randomness as one of the categories of uncertainty is incorporated into engineering design the afore mentioned industry halls are contemplated, again (Fig. 4.19). In this case a multi-level approach is applied to appropriately capture the real world behavior of the interactions between loading, structural response and deteriorations. By that, the time-variant computation of the deteriorations and the damage is accomplished by two consecutive analysis models (i) a first-model (Fig. 4.19) is used by which the stresses due to extreme turbulent wind velocities in a micro-time

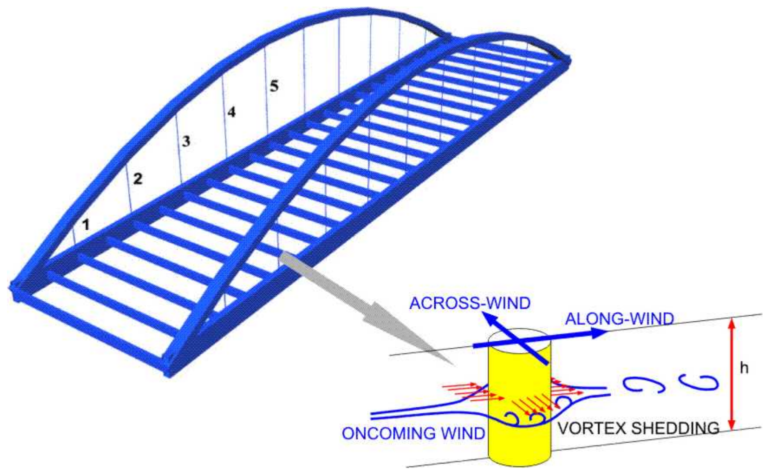

Abb. 4.22. Vortex-induced vibrations (lock-in effects)

scale (T ≈ 10 minutes) are computed (ii) a second model is taken to estimate the stresses due to wind lading in the macro-time scale (Fig. 4.23).

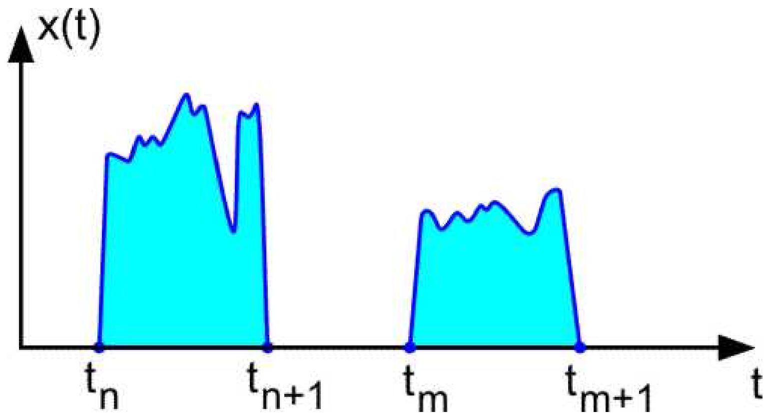

Abb. 4.23. Intermittend Continious Stochastic Process

Based upon this micro-macro separation, the fluctuations in the micro-model can be assumed as piecewise stationary and Gauss-distributed while the actions over the long term are represented as a stochastic intermittent continuous pulse process (Markov renewal process). The deteriorations induced in the micro-time scale are then accumulated, according to the pulse process and using stochastically a modified linear Palmgren/Miner (S/M)-rule. By means of the above described damage accumulation, representing a first passage problem, the failure probability P_f due to fatigue can be assessed. The value for P_f is not allowed to exceed a given admissible limit P_{adm}. This requirement, then, forms a stochastically non-linear and time-dependent constraint for the structural optimization problem besides side constraints for the selected optimization/design variables. For the steel frames considered here, according to (Fig. 4.23) only two sizing variables are defined: The height X_1 and the width X_2.
The objective function is the cross-sectional area $A(X_1, X_2)$. Thus we have the following optimization problem
The solution of the exemplary design problem can be learned from (Fig. 4.26) where the optimization domain is depicted in the X_1/X_2-space.

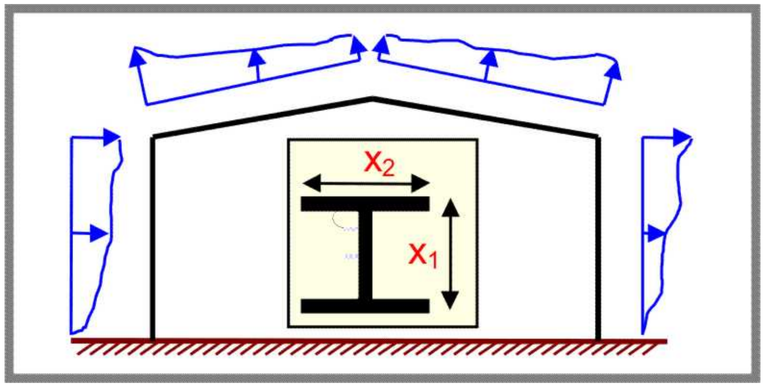

Abb. 4.24. Optimization model for framed system (industry hall)

$$min\left\{ Area(X_1, X_2) \;\middle|\; \begin{array}{c} 0{,}1 \le X_1 \le 1{,}0 \\ 0{,}1 \le X_2 \le 0{,}3 \\ g(\mathbf{X}, \mathbf{Y}, \mathbf{Z}(t)) = P_f(\mathbf{X}, \mathbf{Y}, \mathbf{Z}(t)) - P_{adm} \le 0 \end{array} \right\}$$

Abb. 4.25.

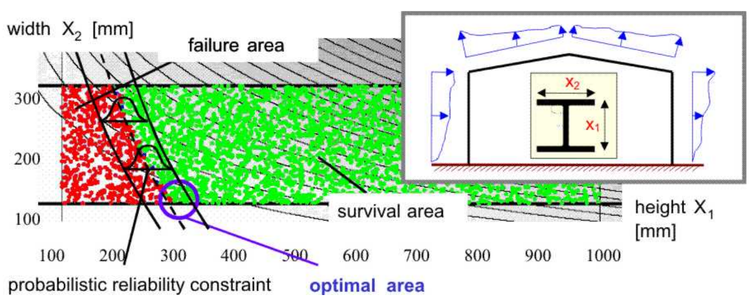

Abb. 4.26. Solution domain for stochastic optimization

As illustrated, the stochastic constraints for the failure probability scatters around expectation values (see bell-shaped curves in (Fig. 4.26) and creates two seggregated sub-domains: (i) a first domain containing green points (no failures) and a second domain of red points representing failures. The optimum is a set of points (encircled in blue) the size of which depends on the selected critical fractile.

Computer-based demolition of complex structural systems

n many suburban densely populated areas there are existing buildings and constructions that arrive the expiration of their service life because of insufficient structural quality, changed requirements regarding utilization or simply inacceptable layout. As a consequence, an increasing number of buildings (departments store, administration buildings, multi-storey buildings etc.) have soon to be demolished. With respect to emerging costs, time and obstructions of people, traffic or inner-city infrastructures the demolition by using controlled explosives is mostly the best technique to tear down structures. This, however, only applies if the collapse, triggered through the explosive charges, elapses according to schedule and does not cause the slightest collateral damages, neither for human beings nor in the adjacent infrastructure. The following figure

Abb. 4.27. Instances of explosions in the practise

exemplifies characteristic blast scenarios that took place in intercity areas such as the explosions for the demolition of (i) a library in Dortmund, (ii) a high-rise residential building in Hamburg and (iii) an administration building in Wuppertal. All of these demolitions have been attended and analysed for research purposes with the objective of creating a computer-based simulation system by which the real world collapse due the explosion can be mimiced. Based on this simulation, an optimal demolition strategy is

Library Dortmund   Millerntor Hamburg   Sparkasse Wuppertal

Abb. 4.28.

to be sought in such a way that the discrepancies between the provoked debris pile and a prescribed collision domain is minimal". The research is carried out in the Research unit Computer-based destruction of buildingsthat, again, is supported by funds of the DFG, linking together Structural Mechanics, Computational Mechanics, Structural Concrete and Engineering Informatics.

Everybody who has watched real world blast demolitions of complex buildings on the TV, or as spectator, can imagine that a plethora of imponderabilities may prevent the desired collapse result. Such imponderabilities may be caused by both data uncertainties (structural data, explosion parameters e.g. regarding volumes of charges, time sequences of ignition, efficiency of explosive charges) as well as model uncertainties (range of structural behaviors, contact and impact phenomena during the collapse, wind actions, etc.). If all relevant aspects are to be modeled realistically, then, the total aforementioned subcategories for uncertainty modeling have to be materialized, i. e.

- randomness
- fuzzyness and
- fuzzy Randomness.

The application of randomness is possible if stochastic variables (basic variables) can be identified, if a sufficiently large universe is available and the mathematical principles of randomness are valid. Material properties, like the modulus of elasticity, Young's module, etc. are good examples for that. By contrast, fuzziness is more suitable if qualitative issues have to be captured, based upon subjectively defined membership functions for uncertain quantities. Herewith, a new type of interval arithmetic can be introduced instead of using conventional numerical concepts. Finally, if items or parts of the stochastic quantities (basic variables or random processes) contain incomplete information or violate the principles of pure randomness, then, fuzzy randomness fills the gap between objective randomness and subjective fuzzyness within the same common body of representation.

Beside the representation of uncertainty a further fundamental solution concept is mandatory. Due to the complexity of the simulation-guided optimization problem, the original problem has to be decomposed and distributed into subproblems, representing different problem levels with respect to space and time. In the Research Unit project, four distinct interacting levels are recognized to map the real world behavior of the collapse cascade appropriately. The subsequent figure (Fig. 4.29) illustrates the spatio-temporal multi-level approach more detailed.

Some few elucidations: As can be seen, on the micro-level the local effects of the explosive charges are modeled using a smooth particle hydrodynamics code and a net-free Galerkin method (hereby time scale ranging from micro- to milliseconds while space dimensions [L] are between mm and cm). In a micro-/meso-level sub-model the shock waves and vibrations are computed based on a finite element method

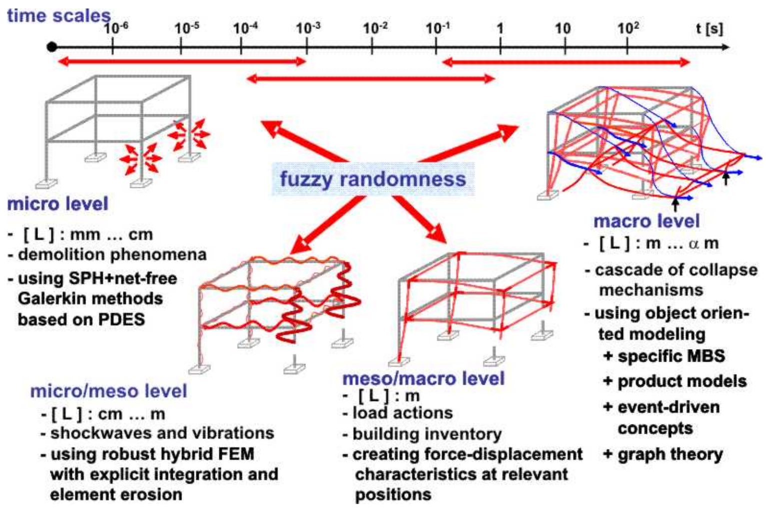

Abb. 4.29. Spatiotemporal in blast simulation

associated with explicit time integration and element erosion ($\Delta t \approx$ milliseconds, [L] between cm and m). On the meso-/macro-level of the structural system an inventory control determines the random character of the structural data (concrete parameters, reinforcement assembly, etc.). Also, three dimensional strain-stress characteristics at critical sections are established which are employed in the collapse cascade of the total system level. The domain of influence is in meters. On the total system level the collapse cascade is simulated, which spans from a few hundredth to about hundred seconds, and comprehends the total structure and its fragments after the explosion. The complexity of the collapse cascade requires a flexible approximation model where a multi-body approach fits best, taking into account the essence of the three preceding sub-models. According to the multi-body model, bodies are regarded as objects interacting with other objects: Thus, an object-oriented implementation is accomplished that also allows for a mapping of the event sequences during collapse.
From the graphical diagram in Fig. 4.29 it can be recognized, that on all four sub-levels fuzzy randomness is integrated to embed the relevant uncertainties as already described above. For a prototypical system a high-rise concrete skeletal building, which covers the most significant collapse mechanisms, the first simulation results of the chosen fuzzy-random multi-level concept are available. In Fig. 4.30 the collapse of the building after the explosion is visualized . Also, the debris hill is portrayed.

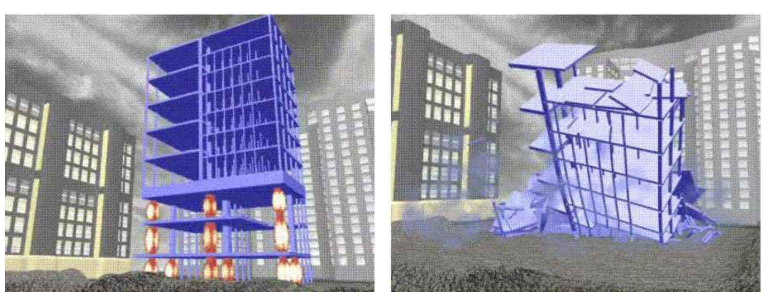

Abb. 4.30. Collapse simulation using fuzziness

Within the test simulation simplisticly two quantities, the rigidity of the structure and the friction coefficient, have been fuzzificated in terms of two triangular membership functions. This approach yields an uncertain range for the debris radius (Fig. 4.31) where the α-value represents the level of membership: The value α=1 means full membership while α=0 (left and right of α=1), marks vanishing membership.

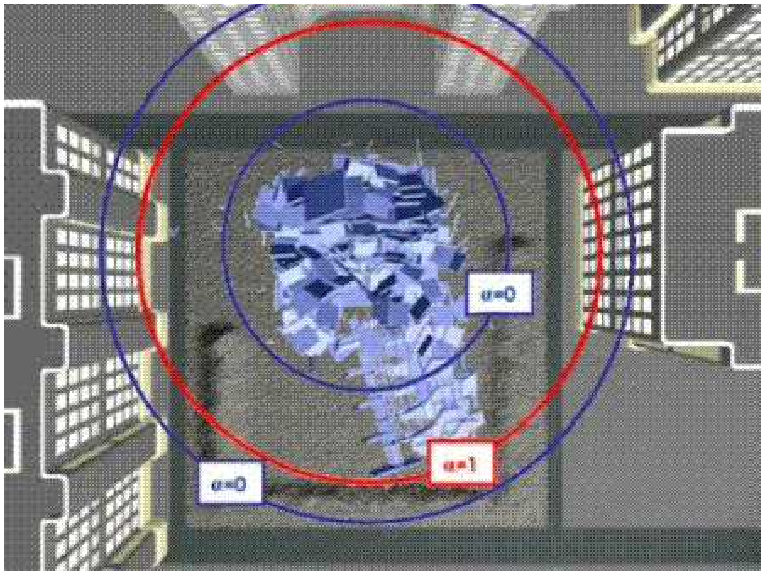

Abb. 4.31. Uncertain debris radius

Imperfection-sensitive structural optimization

Imperfections in a structural system, induced through construction and erection, are a natural phenomenum that never ever can be totally excluded. In general, as geometric imperfections may occur. Here, the focus is on geometric imperfections which lead to more or less aberrations from the planned geometric shape of a structure. Geometric imperfections can become dangerous in structures subjected mainly to compressive stresses because sudden collpase can happen. Typical compression-subjected structure are shown in Fig. 4.32.

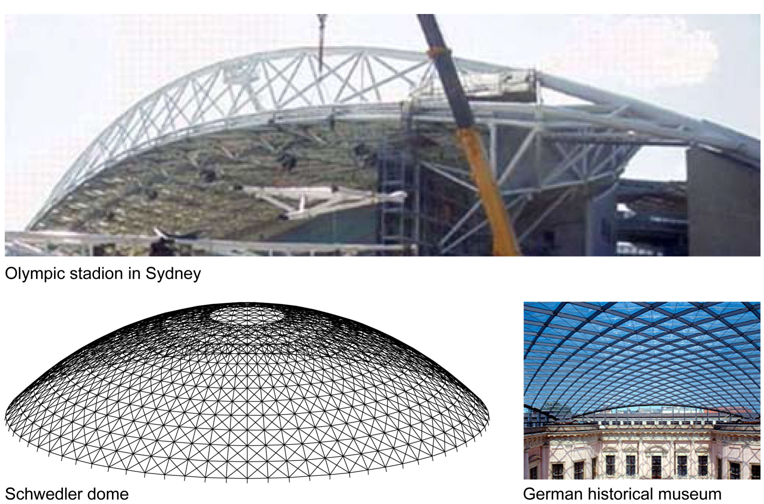

Abb. 4.32. Examples for compression-subjected structures

In particular, if it is attempted to provide an optimum design for the aforementioned structures, using a structural optimization model, the optimum solution computed may be counterproductive. The reason for that is that the öptimalßhape obtained by numerical optimization may show an extremely sensitive behavior and fail due to the already smallest geometric imperfections. To explicitly exclude such catastrophic failures with a reasonable degree of probability, the uncertainties due to geometric imperfections have to be scrutinized and incorporated into the shape optimization model. This model has been created in the dissertation project of M. BAITSCH [BAIT03] at the author's institute where again a multi-level approach is required. For the reference structure displayed in Fig. 4.32, (left) the relevant uncertainties due to geometric imperfections have been successfully represented by means of stochastic fields, in association with a property chosen pdf and correlation functions. Starting with the prescribed perfect geometric shape of the structure the introduction of random variables (imperfection variables) for the stochastic fields allow the assessment of the worst possible imperfect geometric and, accordingly, the corresponding most inappropriate structural response. To find out the unknown worst imperfect shape it is assumed that the amplitudes against the perfect geometry cannot be arbitrarily large. In harmony to that fact, the q-dimensional envelope (ball) in the space of the uncorrelated imperfection variables is determined. Hereby, the characteristic length (radius) of the envelope represents a prescribed fractile being is a measure for the gradual assessment of uncertainly. Based on this concept, on a first level of the total optimization model the worst possible imperfect shape is computed by a maximum optimization (see Fig. 4.33).

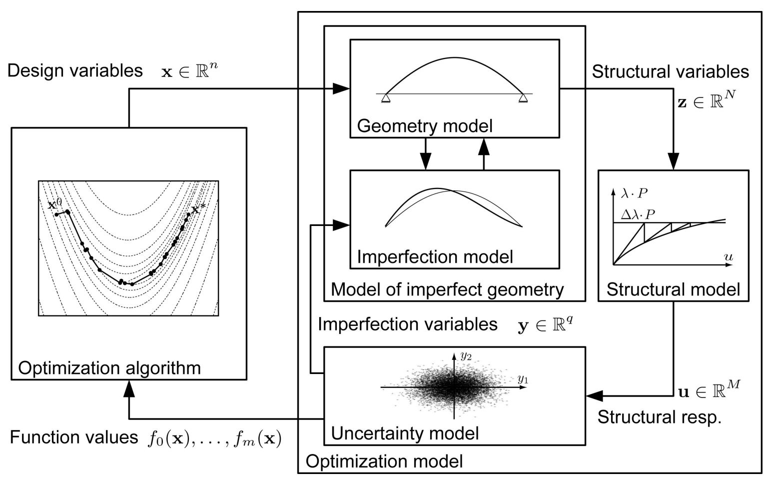

Abb. 4.33. Two level optimization concept

Having determined the worst imperfect geometric a conventional structural optimization is carried out, on the second level (see Fig. 4.33). As illustrated, the uncertainty model yields the imperfection model for which a geometrically non-linear finite element analysis is performed. This analysis forms the kernel for the structural optimization in particular for the constraints of the optimization iteration. The design variable are mode coordinates of the arched girder. As objective criterion the strain energy is taken by which the robustness of the structure against sudden failure can be ensured. In Fig. 4.34 the optimization history is mirrored.
The picture at the top of Fig. 4.34 demonstrates the cascade-shaped reduction of the initial solution where the toggling between the anti-optimization and the structural optimization takes place. The bottom charts of Fig. 4.34 showing the initial and optimized structure, the corresponding displacements as well as the govering imperfect shape.

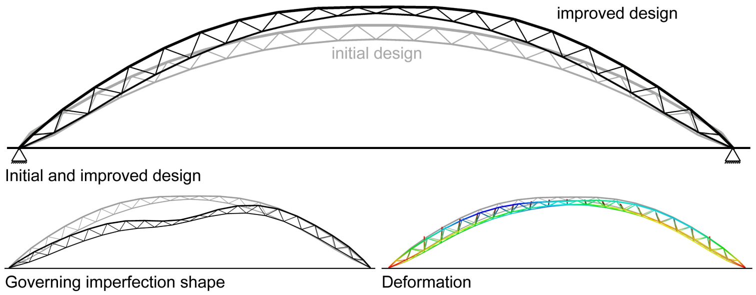

Abb. 4.34. Optimization results

Conclusions

The actual endeavors in scientific structural engineering to understand highly complex systems and processes, realistically, definitely necessitales the introduction of different categories of uncertainty as well the integration of subjective assessments described in a mathematized fashion. By that, conventional numerical methods deployed in Computational Mechanics already for a long time can be linked to methods stemining from the Probability Theory and Computational Intelligence. The consequence is that such an enhancement of Computational Engineering opens new directions to cope with what has been denoted in Cognitive Science as restricted rationality". In the above chapters it could be demonstrated that there are effective and sufficiently accurate ways and means to capture the sophisticated non-deterministic behavior of selected structural engineering, prudentially. Three applications examples have been scrutinized in which a wide band width of computer models for the acquisition of uncertainty phenomena could be pointed out: (i) the lifespan-oriented design of structures, (ii) computer-based demolition of complex buildings and (iii) the imperfection-sensitive structural optimization.
The essential benefit in all three cases is the sustainable improvement of the safety of the systems used and the processes activated. Hence, the vulnerability of the systems and the endangerment of human lives can be drastically scaled down. In this context, it is significant that the improvements or reductions can not only evaluated qualitatively but also quantitatively. That is to say uncertainty can be better gauged and better detected than before, in advance to hazardous events.

4.7 Logistics: The Complexity of Operational Transport Optimisation (Kopfer)

Describe the multitude of boundary conditions observed in real-world freight delivery problems (time windows, packing/stacking conditions, ...) and the methods that have been established to treat them. Elaborate on the nature of respective heuristical approaches and their limitations. Mention uncertainty aspects, like stop-ups and other casuals during delivery tours, and their effect on tour optimalities.

4.8 Innovation management: Assisting the development of optimal innovation processes (Möhrle)

Show that different goals exist when innovating processes. Discuss method of probabilistic outcome optmization. Discuss reception of probabilistic prediction with decision makers in innovation processes.

4.9 Online Problems in Logistics: Optimal activity management for the ADAC Yellow Angels (Grötschel)

Discussion of typical online-problem aspects relativ to static optimization. Which limits exist in treatment of such problems? What minimal bounds can be given as guaranteed maximum distance to a theoretical optimum?

4.10 Social Systems: Self-Optimizing Interaction Organisation in Communities (Strübing)

Which kinds of self-organization are to be observed in the selected area of computer technological communities? Is there some social "goal"discernible in the interaction process?

4.11 Evolutionary Medicine: Limitations on system optimizability, imposed by historic development (n.n.)

The evolution of a biological species leads to fixations in the variability space that are (near to) impossible to transcend by an ongoing natural development process. Instead, side branches are followed, optimizing the setup into sub-optima. Medical examples illustrate this phenomenon.

4.12 Optimization in Acoustics: Selected Examples

Peter Költzsch, Volker Bormann

Description of the problem of creating a loudspeaker with good acoustic properties. Definition of objective and subjective targets. Which targets are most relevant in practice? Choice of target functions by higher-level goal of maximizing sales.

Description of the problem of creating a good concert hall acoustics, a loudspeaker with high sound quality, an optimal adjustment of hearing aids, an optimal walking sound of special floors.
Definition of objective and subjective goals.
Examples of sound design.
Which targets are most relevant in practice?
Is it necessary to replace subjective by objective design goals?
Which influence do the test person's skills (experts, nonprofessionals) have on the optimal results?
Definition of objective and subjective goals.
Which influence do have test person's skills (experts, non-professionals) on optimal results?

4.12.1 "Good acoustics´´ of a concert hall or a opera house

The concept of "good acoustics´´ of a room is not precisely defined. Fuzzy expressions are used for the description of such a condition. These expressions are audibility, acoustic properties of a room, acoustical intimacy, liveliness, reverberance, fullness of tone, sound volume, timbre, acoustic "warmth´´ of the room, spatial impression, auditory spaciousness, and acoustic envelopment of a listener ("a listener should be bathed in sound from all directions´´).
Room acoustic criteria have a subjective and an objective aspect (AHNERT et al. 1993, ANDO 1985, 1988, ANDO et al. 1997, BERANEK 1992, 1994, 1996, 2003, FASOLD et al. 1984, 1987, 1998, KUTTRUFF 2000, REICHARDT 1979, SCHROEDER 1999). The subjective aspect concerns the perceived acoustic quality of a room. For example at music performances this means the perception of acoustic spaciousness or clarity of the sound of the orchestra. The objective aspect involves the physical description of sound field parameters for the subjective quality assessment. Room acoustic criteria are also dependent on type and use of a room and on the type of a acoustic performance. Criteria for different kinds of music and for speech are to be distinguished.
The acoustic design of a room or a hall for the purpose of musical performances is usually part of an overall design. Thereby objective methods are used including objective criteria as well as subjective elements which accompany them. "Good acoustics´´ of a concert hall results from the interaction of subjectively assessed acoustic impressions in general and of subjective sound perceptions in detail as well as objectively sound field parameters, and that at points where listeners and musicians are placed.
A function sufficient for the goal "good acoustics´´ results from individual and from collective experiences with acoustically good or bad rooms. Furthermore such function of the goal can be derived from hearing tests in a room and from tests in a synthetic sound field in an anechoic room by mainly investigating separated sound effects.
Methods of computer simulation (with subsequent auralization) and investigations with acoustic models of a room, and applications of synthetic sound fields in an anechoic room can be used to calculating in advance or even to measuring the objective criteria in a not yet existing room.
As before mentioned at present no standard is at hand for the overall acoustic quality of a room or for "good acoustics´´. Especially a clear connection between objective factors which can be measured to a subjective assessment is still partly unsolved. The acoustic literature offers a set of criteria for the assessment of the acoustic quality of

rooms regarding musical performances (for example by FASOLD et al. 1987, 1993, 1998): - Spaciousness, spatial impression (temporal and spatial structure of a sound field), - Clarity (transparency), sound entry, absence of echo (temporal structure of a sound field), - Loudness, listening level, balance (dynamical structure of a sound field, - Timbre (spectral structure of a sound field).
The valence of these criteria approximately corresponds to the mentioned order. However it is absolutely dependent on genre and style, e.g. classical chamber music, romantic symphonic music, orchestral music, opera.

Importance	BARRON (1988)	SCHMIDT (1990)	BERANEK (1992)
1.	reverberation clarity	reverberation time	transparency time
2.	spaciousness	spatial loudness	impression
3.	loudness	loudness	spaciousness
4.	clarity timbre	clarity transparency	transparency
5.	intimacy	-	intimacy
6.	-	-	timbre (warmth)

Tabelle 4.2. Comparison of evaluation of room acoustic quality characteristics of concert halls (FASOLD et al. 1993)

Descriptions of room acoustic quality characteristics (following FASOLD et al. 1998, AHNERT et al. 2002, MORFEY 2001, BERANEK 1996 etc.):

Reverberation time is the duration of audibility of reverberations. With it is connected the reverberation time which can be measured as time for a decay of sound energy of 60 dB after switching off the sound source. It is a rough quantity for the room acoustic quality of a concert hall, but it is not the only criterion.

Reverberance is the perception of reflected sound supplementary to direct sound which is not perceived as a repetition of direct sound.

Clarity, transparency is the possibility to perceive tones in a sequence of time (transparency in time) or in distribution of instruments (transparency in register).

Loudness, listening level is the decrease of sound pressure level in a room, in comparison with front places

Auditory spaciousness is the apparent enlargement of a source of sound (or orchestra) mainly in lateral direction to the audience in spite of its view (apparent source width)(BLAU 2001, 2002).

Auditory spatial impression is a being "wrapped in music´´, an envelopment in music, that is the listener's feeling, that the strength and directions of the reverant sound seems to arrive equally from all directions – forward, overhead, and behind (BERANEK 1996); a perception of a combination of all sources of sound with the environment of room and taking into account the audience therein.

Timbre is pitch feeling, sensation, warmth, brilliance, brightness.

Intimacy is the acoustic perception of the size of a hall, the feeling of proximity of a sound source, and interval in time between the arrival of direct sound und first reflections to the listener.

Definitions of the main acoustical criteria are given below, furthermore optimal ranges of large concert halls (FASOLD et al. 1998, BERANEK 1996, MORFEY 2001, AHNERT et al. 2002):

Reverberation time in seconds for decreasing the sound energy or the sound pressure level in a room from stop the source of its steady state level to one millionth or a level decay by 60 dB. The optimal range is 1.6 to 2.1 s. Early decay time is also applied as time in s for decreasing the sound pressure level in a room to the first 10 dB of the decay process. This time is most important for the perception of reverberance.

Reverberance measure is the ratio of reverberation energy in the time between 50 ms and oo (infinite) to that in the time between 0 and 50 ms. The optimal range is 0 to 4 dB for concert hall, 3 to 8 dB for symphonic music and –2 dB to +4 dB for musical theatres with optional use for concerts. A similar meaning has reverberation distance.

Clarity index, early-to-late sound index (fig. 6) is the ratio of early to late energy of sound energy arriving at a point in a room. For music is C80 valid. The limit between "early´´ and "late´´ is taken at 80 ms following the direct sound. The optimal range is –1 to +3 dB, depending strongly on the musical genre.

Early lateral energy fraction, lateral efficiency is the ratio of sound energy at a point in a room that arrives from the sides of a room to that which arrives from all directions, both within 80 ms. The optimal range is approximately 20

Spaciousness, spatial impression measure is defined as reinforcement of the spatial perception by reflections coming from the sides (LEHMANN 1974, SCHMIDT 1967, TRAUTMANN 1986, 1989, LEHNERT 1993). This index takes into account the spatial perception which is determined by reverberance and diffusivity (listeners envelopment) as also from the spaciousness which is caused by early lateral sound reflections (source broadening or apparent source width). For the acoustic quality of a concert hall are especially important the reflections from the side walls which are delayed to 25 to 80 ms compared to direct sound. The optimal range is +1 to +7 dB.

Interaural cross-correlation coefficient – IACC (MORFEY 2001) is a measure of the similarity between the signals received by the two ears and therefore a quantity of perception of a room. It is obtained by cross-correlating of the impulse response functions of a room which is measured at the positions of the two ears over a sample period. This is usually defined as the first 80 ms after the arrival of direct sound. With increasing influence of lateral reflections compared to direct signals of sound the IACC becomes less.

Loudness, sound level difference is the decay of a sound pressure level in a room. The level at any seat in a hall is compared with that in front of the podium (distance 5 to 10 m).

Sound entry (attack) is the perception of build-up time or the steepness of build-up of musical instruments. This is strongly influenced by technology of playing the instruments. It is characterized by the time of excitation. For shorter times the build-up of sound is steeper or the sound entry is harder.

Timbre (MORFEY 2001) is the tone colour or quality of musical sound. The term timbre can be applied to the sound of individual instruments or groups of instruments. It can also refer to the balance between low, middle and high frequencies in the musical sounds. Timbre is also a aggregate of attributes that allows a listener to distinguish a sound, in terms of subjective impression, from any other sound having the same loudness, pitch, and duration as well as the same direction of arrival.

Bass ratio (MORFEY 2001) Is the ratio between the reverberation times at low frequencies (at octave centre frequencies 125 and 250 Hz) to reverberation times at middle frequencies (at octave centre frequencies at 500 and 1000 Hz) in a concert hall. The bass ratio is considered to relate to perceived warmth, that is the liveness or fullness of bass tones. Musicians sometimes describe a hall as „dark" which has to prominent a bass or where high frequencies are largely attenuated (BERANEK 1996). The optimal range is 1.2 to 1.3.

Diffusivity is the local uniformity of a sound field distribution with respect to intensity and direction of incidence. It is also a temporal quantity for a statistical temporal distribution of a sound field. High temporal diffusivity means low appearance of harmonic natural frequencies.

Immediacy of response, attack (BERANEK 1996) is a feature from the musician's standpoint. A hall should give the performers the feeling that it responds immediately to a note. This term is related to the manner in which the first reflections from surfaces in the hall arrive back at the musician's ear. If the reflections occur too long after the note is sounded, they will be heard as an echo. If reflections are heard only from the nearby surrounding stage walls of the stage around him, the musician will fail to sense the acoustics of the hall altogether.

Results of assessments for two examples of well-known concert halls will be dealt with (FASOLD et al. 1987, 1993)

	Neues Gewandhaus Leipzig (1981)	Schauspielhaus Berlin (1985)
Hall:	fan-like plan	classical shoebox form
Volume:	21,000 m^3,	15,000 m^3,
	1900 seats, 11 m^3/seat	1430 — 1670 seats, 10.5 — 9 m^3/seat

In these examples subjective assessments of acoustic quality were investigated with help of 50 to 60 test listeners at 5 selected representative seat areas. The assessments were required by the test listeners about the criteria: loudness, duration of reverberation, clarity, and spatial impression. These criteria seemed to be most important. For the assessments a scale of estimation was used. It was divided into 5 classes (middle class: without complaints, then: complaints are perceptible, and: complaints are disturbing; in both directions from low to high impressions). From the results conclusions were drawn of the acoustic quality of the room at individual seats as well as of the whole hall.
In fig. 1 the results of subjective tests are shown. The results are averaged over the seat areas.
Fig. 1: Selected results of subjective assessments of acoustic quality of two concert halls. The diagram shows the percentage distribution of these assessments for different acoustic qualities (FASOLD/STEPHENSON 1993)
As another example some results of optimization shall be dealt with for the new Semper Oper in Dresden. The optimization was carried out with subjective goal functions. The optimization was necessary because of the reconstruction of the opera-house after its destruction in World War II. It was reopened at February 13th, 1985. The large hall of this opera-house is of the type of a theatre hall with semicircular plan and four balconies. Volume: 12,500 m3, 1,300 seats, 9.6 m3/seat. The hall can be used for performances of operas as well as for concerts. The investigations were carried out by Prof. Walter Reichardt and Prof. Wolfgang Kraak with their co-workers of the Institut für Technische Akustik der Technischen Universität Dresden in the time of 1970 and 1980 (KRAAK 1984).
On basis of existing documents a model of the historic hall of the Semper Oper was prepared on an scale 1:20, see fig. 2.
Fig. 2: View into the auditorium of the model (1:20) of the historic Semper Oper Dresden. Absorbing materials to imitate the audience are shown.(photo: R. DIETZEL, TU Dresden)
In this model the impulse responses of the room were measured as objective criteria. The measurements took place binaural with an artificial head. It was measured in the empty room and in the case of a opera-house with the fictitious audience. From the measurements the characteristics of sound quality were determined. These are reverberation quantity, clearness index C50, clarity index C80, spatial impression index, loudness, echo criteria, centre time, and reflection measure.
Hearing tests were carried out before opening the rebuilt Semper Oper. The tests included 86 persons in case of opera tests and 81 persons in case of concert tests. Overall 28 tests were carried out with different variants (spatial details, musical motives etc.). The test persons were 28

The criteria of assessments were dynamics (loudness, balance), temporal structure (reverberance, clarity (transparency), blend, echo), spatial structure (spatial perception, spaciousness), frequency composition (timbre, change in timbre), and acoustical overall perception. The quantity of hearing impression related to these characteristics had to be marked on a scale.

Following results were achieved: For concert cases the objective characteristics which were derived from impulse measurements are located on average in a optimal range. The differences between various seats in the audience and the differences between an occupied room and an empty room are relatively small. According to subjective assessments the majority of criteria on an average is found as optimal. In case of a concert it was found as acoustically extremely good.

For the opera case the objective tests showed the reverberation time as substantially dependent on the equipment of the stage. The other objective criteria are found partially in the optimal range and partially slightly off the optimal range. The subjective assessments showed in summary the following results: - Loudness, reverberation duration, mixing and spatial perception were estimated as optimal or as nearly optimal. - The transparency was assessed lower than for the concert case. - The speech intelligibility was assessed between good and satisfying. - The intelligibility of recitatives was assessed as better than good.

The comparison of acoustics of the old with the new Semper Oper was interesting (fig. 3). This comparison leads to the conclusions: The similarity in architecture of both opera houses had resulted in optimal values for spatial perception, mixing, and loudness. Therefore it can be stated the new opera house had attained the nearly former acoustic quality.

Fig. 3: View from the stage into the auditorium of the new Semper Oper (photo: R. Dietzel, TU Dresden)

The auditorium of the new opera house is slightly increased and therefore the compactness was slightly reduced. In consequence the reverberation time is increased. Now the slope of the parquet in the background and the view to the stage had improved the acoustic conditions. That's the reason for the good subjective assessments.

From now on the following investigation methods are used to designing acoustically a concert hall. This is necessary to follow the optimal acoustic criteria before a final design can be established.

Computer simulations as a method for acoustic design of a room The models of simulations are based on physical laws of geometrical room acoustics (ray acoustics). That means wavelength are considered as short in comparison to characteristic dimensions of a room. There are in detail: mirror image source method, ray tracing models respective sound particle simulation methods (fig.4) including Monte-Carlo methods (VORLÄNDER/MECHEL 2002, STEPHENSON 2004)

Fig. 4: Computer simulation with a sound particle method (STEPHENSON in FASOLD et al. 1998)

The essential goal of computer simulations is to find out room impulse responses, fig. 5. From this the interesting acoustic criteria of a room will be calculated, e.g. clarity index C80 in fig. 6.

Fig. 5: Computer modelling in room acoustics At the left side is the acoustic system and on the right side the impulse response h(t): from source to receivers (VORLÄNDER 2004)

Fig. 6: Results of a computer simulation: Distribution of the clarity index C80 in an concert hall with three balconies (VORLÄNDER 2004)

The computer simulation can be finished with an auralization. That means the audibility of a sound field can be realized at a seat in the auditorium of a virtual room by computational realization (MORFEY: "Auralization is the conversion of a digital waveform into audible form.´´). This auralization is carried out with help of calculated room impulse response, further with a "dry´´ (free of echo) recorded sound (e.g. music signals in anechoic rooms) as well as with outer ear transfer functions. Computer simulation with auralization thus enables us to hear into a yet not existing room!

Experimental model measuring technique for room acoustics The procedure is based on the similarity criterion of Helmholtz number He = idem (KÖLTZSCH 2003). According to the reduction of geometry of a room (usual scale 1:20) the sound wave length are reduced in the same relationship. The frequency range is reciprocally transformed to the geometrical scale reduction (see example Semper Oper in fig. 2).

In a such model room the reflection and absorption properties of room surfaces must also be modelled. Sound sources are used e.g. shock excitation type spark transmitter with sound radiation characteristics which can be selected. The receiver is a model artificial head. It has two sound pressure receivers. Room impulse responses are determined in model measurements. This is the same as with computer simulations. Therefore the desired room acoustic characteristics are calculated (e.g. reverberation time, clarity index etc.). Furthermore the model offers the possibility of auralization (hearing into the model).

Model measurements are very expensive and time consuming. The advantage however consists in a clearness of the model. This is a crucial advantage in co-operation of acousticians and visually oriented architects. Diffraction and scattering processes in a room can be reproduced physically accurate which is a further advantage of model measuring techniques. These are the problem cases of a computer simulation. Details of configurations within a room can be varied fast and simply with model measurements. Examples are inclinations of walls or cover sections.

Synthetic sound field Natural sound fields can be reproduced in anechoic rooms (free-field rooms) with electro-acoustic equipments. For this purpose synthetic sound fields are generated with help of loudspeaker devices. An example shows fig. 7.

Fig. 7: Experimental set-up in an anechoic room to generating a synthetic sound field. The problem which is to be investigated in this case is the subjective investigation of the apparent source width of an orchestra (following M. BLAU in KÖLTZSCH (Ed.) 2003). This research method optimizing room acoustics has the advantage to consist of the fact of systematically varying components of a single sound field. Furthermore the influence of variations of these components on room acoustics quality can be subjectively judged.

Optimal adjustment of hearing aids to rest-hearing ability of a deaf person Modern hearing aids have a large number of adjustable parameters. With this the loss of hearing can be individually compensated for with most different circumstances. Therefore the problem is to find out an optimal parameter constellation “by hand´´ in order to maximally use the rest-hearing ability. Such a hearing aid adaptation is successively carried out in a “dialogue´´ between acoustician of hearing aids and hearing impaired persons. The optimization includes for developers and manufacturers of hearing aids absolute objective goal functions, among others: miniaturization of a hearing aid, minimization of energy consumption, expansion of the range of adjustable parameters, and adaptation to requirements of users.

Against a multi-criterial optimization by a hearing impaired person is primarily connected with subjective characteristics of his hearing impression. These are loudness, pleasantness (pleasant impression, being agreeable), timbre, distortions, speech intelligibility, noise, background and ambient noise. Furthermore non-acoustic, subjective criteria play frequently a crucial role, e.g. comfort of wearing, low maintenance, being user-friendly, service comfort, low purchase price, visibility in public.

For processes of adjustment is often used a so-called characteristics spider as a suitable tool. With it the multi-criteria optimization with subjective goal functions can be demonstrated, fig. 8.

Fig. 8: Characteristics spider for judgements of natural sound samples with an evaluation example (following G. FUDER in KÖLTZSCH (Ed.) 2003)

4.12.2 Optimization of the quality of speech communications

Optimization goal in speech communications is the highest possible degree of speech intelligibility. This has a considerable influence on successes into learning in schools, on aspects of occupational safety, and on the entire psychosocial conditions of an individual. Quantifiable and thus measurable objective criteria are the dimensions of a room, reverberation times, speaker-listeners distances, signal-to-noise ratio, type of background noise and its frequency spectrum, temporal structures, and information contents. Additionally measurable criteria are related to useful signals. These are speech levels, speech spectra, and speech velocities. Moreover for this optimization there are numerous subjective criteria. These cannot be measured but with them are connected important factors of influence: speech quality, speech melody (intonation, prosody), accents as well as dialects, possible hearing impairments, speakers and listeners knowledge of vocabulary of used languages in dependence on age, education, upbringing, interests, training and practice.

The optimization strongly is influenced by conditions as are types and places of speech presentations, are spoken texts as sentences, as single words, as numbers, as rhymes, and as sung texts. Places of speech presentations as churches, auditoriums, concert halls, or class or living rooms also are of influence.

Furthermore in context with speech presentations following visual subjective factors are of important influence for high speech intelligibility. These are visual contact to a speaker, gesture, facial expressions, mouth movements, and supporting graphical aids.

4.12.3 Acoustic optimization of modern floors in buildings (walking noise, impact sound)

Optimization of modern floors is highly multi-criterial. Primarily following goal functions are of use to manufacturers: profitability, competitiveness or marketability, acceptance of products by users, prices which can be achieved, quality control (guarantee, warranty), investment costs (production plants, buildings), material costs, production costs, and manpower costs.

Secondly for users the criteria of optimization with regard to acoustic characteristics are feeling at walking (elasticity and substructure of floors), walking noise (loudness, timbre, sound feeling), and impact noise (excitation and propagation of structure-borne sound and its radiation into the surrounding air) (SARRADJ et al. 2000, 2001, 2004, 2005).

In a principle sketch are shown walking and impact noises (Fig. 9).

The process of optimization of these acoustic criteria including subjective judgements is very time consuming because of the dependence of walking noises from materials of a floor (elasticity, structure of sandwich-layers, mass), substructures, quality of installations, room acoustic conditions (especially reverberations), footwear, and walking habits (step rhythm, easy or sturdy steps).

Fig. 9: Walking noise and impact sound (principle sketch)(following E. SARRADJ 2004)

Subjective acoustic goal functions (quality criteria) of walking noise were fixed, concerning the potential buyers and people exposed to such unpleasant noise. This includes perceived loudness as objectively measured sound pressure level in average in a succession of pairs of steps, perceived timbre (pitch) as objectively measured amplitude spectrum of walking sound, and overall perception of walking noise with regard to pleasantness at sound exposure.

At hearing tests the walking sound of a person was measured for subjective assessments. 8 laminate floors from different manufacturers were investigated. They included different materials and layer structures. The before digitally recorded walking noises were offered in hearing tests to approximately 70 individuals. As method pair comparisons were chosen. Two probes of walking noises were offered over headphones. The test listeners

had to mark on three scales which noise is quieter, is in timbre higher, and at least is more pleasant.
Results for all test persons were analysed with reference to sex and to age of a test group. These groups were divided into group 1 (under 30 years), group 2 (30 to 50 years), and group 3 (more than 50 years). An example of results shows fig. 10.
Fig. 10: Correlation of walking noise which is quieter and offers a more pleasant feeling (L-A) as well as a noise which is higher and offers a more pleasant feeling (H-A) (following R. EMMLER, see also SARRADJ 2000, 2004).
The results lead to the conclusions: there is no walking noise on laminate floors which is ideal, the connection from perceived loudness and timbre to assessed pleasantness clearly is of subjective influence, and another influence is of sex and age group. Regarding the laminates judged as most unpleasant (loudest and dullest walking noise) the assessments of all test persons were quite similar. On average of all test persons laminates were assessed as most pleasant which were quietest and made the highest timbre. Test persons of group 2 (30 to 50 years) gave the most stable assessments. With them the correlation between quiet and high to pleasant was the highest one. Male testers were more stable in their assessments. Female testers at least partly prefer louder and duller walking noises. Group 3 (more than 50 years) exhibits more weakly correlations (probable cause: their hearing impairments may have just started).
Conclusions: Loudness and pitch (timbre) as criteria which can be objectively measured are only a limited standard for pleasantness and subjective sound feeling. The loudness most influences pleasantness. Subjective assessments of timbre (deeper or higher pitch) vary individually. That means there is no distinct trend in the relationship of the measured criterion of timbre and the goal function of pleasantness.
Furthermore it is very interesting that the acoustic perceptions also depend on optical perceptions, in this case on décor (surface pattern) of laminate floor. Those laminates were assessed as acoustically most pleasant which also were assessed as more pleasant with reference to décor of laminate. This is also right regarding the order of pleasantness (SARRADJ et al. 2005).
Obviously an optimization of modern floors (laminate) cannot be carried out only with objective goal criteria. Of crucial importance also is the very complex subjective optimization criterion of pleasantness of walking noises.

4.12.4 Acoustic quality of loudspeakers

At least assessments of acoustic quality of loudspeakers are performed by subjective criteria. In hearing tests the overall subjective impression is assessed as subjective goal function of an acoustic optimization (KLIPPEL 1987, 1988, BECH 1992, OLIVE 2003). Objective criteria for processes of optimization are sharpness, bass reinforcement, clarity, change in timbre, and distortions. Defect quantities are derived from these criteria as deviations from an optimum of each of them. The sum of all partial objective criteria is not identical to a subjectively perceived quality by the way of this optimization.
The acoustic qualification of test listeners is of great importance to this problem. Since this problem is generally relevant for optimizations with subjective goal functions here it should be dealt with the example of optimizations of loudspeakers. Trained experts as listeners are used in hearing tests of manufacturers. Such listeners represent subjective “measuring instruments´´ (KLIPPEL). The requirements to the listeners are hearing experience (experienced HiFi hearing, regular visit to concerts, musical activities by themselves). These experiences determine critically and discriminately the listeners conceptions of an ideal sound expression and the skill to being able to estimate a sound impression. Additionally test experience of trained listeners is an essential advantage. This includes familiarity with scaling methods and other conditions of hearing tests. In tests for loudspeakers as consumer products are preferred untrained persons (laymen). They can be seen as representatives for buyers. Demographic and socio-economic factors

are of concern in these tests. Already the training process can influence the listener. Trained listeners are in general more critical than untrained listeners. For example they estimate acoustically bad loudspeakers worse than laymen. That means judgements of trained listeners can not be representative for unbiased and untrained listeners. This is valid for the large group of buyers and users. Expensive acoustic improvements of loudspeakers which were assessed by trained testers would often not be noticed by untrained listeners. Trained listeners are best in subjective tests with regard to reliability of assessments. At such tests high stress of testers has to be considered with regard to temporal and physical exhaustion. This may cause mental and physical fatigue and thus reduced attention and motivation.

Some information about a selected example follows now. In an extensive investigation four loudspeakers were tested (OLIVE 2003). For the purpose to estimate objective criteria for an optimization in an anechoic room sound signals were measured at conditions of a right angle to the diaphragm areas of the loudspeakers and in a hearing windows of horizontal ≈ 30 degrees and vertical ≈ 10 degrees. Additionally measured quantities were sound power, directivity index of sound radiation, and transfer functions. In hearing tests an assessment scale was used for a perceived quality as a general impression (scale of assessment: 0 to 10 which means absolute rejection to total preference). The results distinctly correlate subjective judgements of the order of loudspeakers concerning their sound quality and the results of the measurements.

A significant result was the preference of some loudspeakers by assessments of untrained listeners. This is very similar to the assessments by trained listeners. Experts of these tests conclude hearing tests are the last crucial arbitrator of acoustic quality of loudspeakers.

Sound characteristics of loudspeakers should not only be assessed but also be described and scaled (KLIPPEL). Verbal terms are used for sound characteristics as volumes, sharpness, clarity, height or bass reinforcement, spatial feeling, change in timbre, bass clarity, and brightness. Objective assessments refer to descriptions of loudspeakers at listening conditions which were determined by measurements. The goal is to express the impression of listening by acoustic conditions which can be measured at a place of listening.

Subjective tests of loudspeakers can be represented with these results in a property space. Objective results of attributes of loudspeakers are combined in such a property space. The comparison of these two spaces shows whether the subjective test results can be explained with objective characteristics. As further result it can be shown whether subjective quality factors can be forecasted. Subjective loudspeaker tests should principally serve less to quality assessments according to opinions of manufacturers. These tests should serve more to subjective sound evaluations for users and buyers/customers.

A long path of research shall be gone to explain sound impressions of loudspeakers from physically measured quantities at conditions of listening places. This example seems to show the objectivation of subjective goal functions to be in a good way!

Acoustic quality of products and acoustic design (sound quality, sound design) The emitted sound is a quality feature of a set of products (JEKOSCH et al. 2004, 2005, BLAUERT et al. 1996, 1998, BODDEN et al. 2002, 2003). It is relevant to customers to purchasing a product. Manufacturer and buyer aim at and expect a sound which is a typical trade mark. Acoustic visiting card, “product sound´´ or corporate sounds are terms which are spoken of. Usually the following definition is used (BLAUERT/JEKOSCH 1996): Product-sound quality is a descriptor of the suitability of the sound attached to a product. It results from judgements upon the totality of auditory characteristics of the said sound – the judgements being performed with reference to the set of those desired features of the product which are apparent to the users in their actual cognitive, actional and emotional situation.

Nevertheless this problem is ambivalently judged. Some people prefer values in use in preference to structures of emitted sound. Others prefer quiet products. There are still others preferring louder noise and typical sounds of a product.

Further factors are important because the emitted sound has to have some content of information. This content is of concern to safety factors (e.g. in case of deviations from a normal machine noise or noise of an approaching vehicle as a warning signal to pedestrians). Noise can indicate operating conditions and necessities of maintenance or repair.

Emotional and very different meanings are linked to sounds in everyday life as is unquestionable. This depends on age, living standards, attitude towards life, hearing habits, and experiences. A consequence is the preference of some types or marks of products by different parts of a population. As an example some prefer extravagant and striking sound properties and others prefer particularly discreet sound qualities.

In actual cases it will be asked: What power of noise and which sound a mixer in a kitchen is allowed to have? What strength of noise, how loud, and which timbre a vacuum cleaner is allowed to have according to suction power or normal operating conditions? Which acoustic signal suggests overloading, necessity to empty the dust bag or even to repair? What sound of motor vehicle shall be established which is specific for a type of car (e.g. the so-called “Porsche sound´´)? What kind of sound shall be heart in a car coming from the windscreen wipers or the direction indicator? What kind of motor sound should be designed?

In an actual case these considerations lead to the question what really is the goal of optimization with so many subjective and objective criteria. Additionally in some cases must be taken into account limits of noise (e.g. noise with motor vehicles).

These problems of acoustic quality and acoustic design are rather similar to before mentioned ones. This leads to questions more in general: Can the chosen functions of subjective goals be definitely formulated? Is it possible to arrange a goal function in a form which can be measured? How large is the fuzziness of goal functions? How are to be selected testers for listening tests? How are to be trained testers?

Problems of acoustic qualities and acoustic design are characterized by the criteria (JEKOSCH et al. 2004, 2005) which limits of sound emissions are to be considered with respect to hearing damage and annoyance, which threshold values are to be considered as a safety factor to avoid risks or dangers and to judge operating conditions, which subjective preferences of noise are to be considered in very different situations of everyday life, and which quality of noise at a product and at production is to be considered according to requests of manufacturers and users.

To arrive at progresses of these problems empirical investigations are necessary in scientific research and in developments. Methods of evaluation must be achieved for acoustic product quality and acoustic design by means of field studies and laboratory tests. Goals of such investigations should be to derive rules for ingenious design of product sound. The aim is a code of practice for acoustical optimizations using subjective goal functions which are adapted to the products.

4.12.5 Subjective evaluations of acoustic qualities of musical instruments

For musical instruments the important acoustic criteria are pitch, response, timbre, projection etc. These can be derived from quantities of different measurements. Acoustic evaluations of musical instruments are traditionally conducted in a subjective manner up to now. But more and more a mix of subjective and objective criteria is used which complement each other. However in line with the trend objective criteria are applied which can be physically measured.

The statement of an expert is a generally accepted opinion: “An optimization is principally possible only if subjective criteria are described objectively. . . . (But) the condition arrived at is still completely insufficient today.´´ Therefore in this case the correlation between objective criteria and subjective evaluations is of crucial importance (BALTRUSCH 2003, GÄTJEN 1995, MEYER 1977, ZIEGENHALS 1995, 1996, 2000, 2002).

For subjective test procedures it is important to include judgements as to value by musicians and by listeners. However the test results of listeners and of musicians can deviate considerably from each other. As to the order of judgements musicians have a plain opinion: Goal of acoustic optimizations have to be fixed on ears of professional musicians. Objective criteria can only be used as a support to subjective impressions.

The discrepancies are well-known which arise from assessments of pitch and sound of guitars by listeners and musicians. Both use different criteria for their assessments. From the example with guitars of an acoustic optimization can be concluded to try more and more to correlate measured quantities and subjective judgements. A wide variety of instruments, large production output, and coercion to arrive at high acoustic quality gives the motivation to try this procedure.

Therefore in these multi-criteria optimization processes in most cases the poorly defined subjective criteria must definitely be described by physical quantities. Beforehand the subjective criteria should be assessed by trained listeners and professional musicians. Additionally, musical experts, manufacturers, and if necessary musical historians have to contribute. An influence of non-acoustic criteria takes part as well.

A further example of optimization is the quality judgement of bells (FLEISCHER 2000, HOUTSMA et al. 1987). According to expert opinions only subjective judgement is important. As an expert puts it: “The subjective judgement of a bell expert is above all results of physical measurements.´´ (FLEISCHER 2003) In practice acoustic judgements on basis of subjective criteria are exclusively carried out mainly by special bell experts, church musicians, historians, but not by physicists or laymen. Both of the last would only be able to assess the sensory pleasantness of sound of a bell but poorly defined. The acoustic goal function of a bell is very subjective and of a multi-criteria nature. At present there is no objective procedure with a same result as subjective evaluations. This is also true if in an objective procedure would be included a set of objective optimization criteria which can be measured.

5

Transdisciplinary perspective (Lucas/Roosen, 10 pages)

Summarizing chapter for deriving the added value of the BBAW study group. Try to point out transdisciplinary fecundation areas.

Commonalities Common aspects and effects observed in (nearly) all participating fields of structure research and development.

Specialties Point out methods and problems peculiar to the various disciplines (e.g. use of mathematics)

Transdisciplinary perspective

Examples of border-crossing methodological approaches (??)

6

Literature

Collected citations of all preceding sections, possibly with very short comments on their contents. Citations should be chosen on the base of easy availability for the potential reader. So favour articles in commonly available journals against those in highly specialized disciplinary journals.

Request for discussion: Should we refer online articles, e.g. articles of Wikipedia? This would strongly help the readers in finding additional and up-to-date material. If yes, a dedicated web page should be set up for actualized online references, as web contents tend to change continually.

Literaturverzeichnis

[1] A. Winfree. *Science*, 175:634–636, 1972. Experimental realisation of a Belousov-Zhabotinskii reaction.

[2] Al Hensel. Game of Life Homepage. *http://www.ibiblio.org/lifepatterns/*, (March 2006).
Original site for Conway's Game of Life. Contains a Java applet and some links to elaborate pattern providers.

[3] Dirk V. Arnold. Evolution strategies in noisy environments — A survey of existing work. In L. Kallel, B. Naudts, and A. Rogers, editors, *Theoretical Aspects of Evolutionary Computing*, Natural Computing, pages 239–249. Springer, Berlin, 2001.

[4] F. Azadivar. Simulation optimization methodologies. In P.A. Farrington, D. T. Nembhard, D. T. Sturrock, and G. W. Evans, editors, *Proceedings of the 1999 Winter Simulation Conference*, pages 93–100, Piscataway, New Jersey, 1999. IEEE.

[5] T. Bäck. *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, New York, 1996.

[6] J. Banks, J. S. Carson, B. L. Nelson, and D. M. Nicol. *Discrete Event System Simulation*. Prentice Hall, 2001.

[7] T. Bartz-Beielstein. *Experimental Research in Evolutionary Computation - The New Experimentalism*. Natural Computing Series. Springer, Berlin, 2005.

[8] Thomas Bartz-Beielstein. Experimental analysis of evolution strategies – overview and comprehensive introduction. Interner Bericht des Sonderforschungsbereichs 531 *Computational Intelligence* CI–157/03, Universität Dortmund, November 2003.

[9] Thomas Bartz-Beielstein. *New Experimentalism Applied to Evolutionary Computation*. PhD thesis, University of Dortmund, April 2005.

[10] Thomas Bartz-Beielstein and Sandor Markon. Tuning search algorithms for real-world applications: A regression tree based approach. In G. W. Greenwood, editor, *Proc. 2004 Congress on Evolutionary Computation (CEC'04), Portland OR*, volume 1, pages 1111–1118, Piscataway NJ, 2004. IEEE Press.

[11] Thomas Beielstein and Sandor Markon. Threshold selection, hypothesis tests, and DoE methods. Interner Bericht des Sonderforschungsbereichs 531 *Computational Intelligence* CI–121/01, Universität Dortmund, Dezember 2001.

[12] H.-G. Beyer. Evolutionary algorithms in noisy environments: Theoretical issues and guidelines for practice. *CMAME (Computer methods in applied mechanics and engineering)*, 186:239–267, 2000.

[13] H.-G. Beyer. *The Theory of Evolution Strategies*. Natural Computing Series. Springer, Heidelberg, 2001.

[14] H.-G. Beyer and H.-P. Schwefel. Evolution strategies – A comprehensive introduction. *Natural Computing*, 1:3–52, 2002.

[15] J. Branke, C. Schmidt, and H. Schmeck. Efficient fitness estimation in noisy environments. In L. Spector et al., editor, *Genetic and Evolutionary Computation Conference (GECCO'01)*. Morgan Kaufmann, 2001.

[16] Hans J. Bremermann. Optimization through evolution and recombination. In M.C. Yovits, G.T. Jacobi, and G.D. Goldstein, editors, *Self-Organizing Systems*. Spartan Books, Washington DC, 1962.

[17] C.A. Coello Coello, D.A. van Veldhuizen, and G.B. Lamont. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Academic Publishers, New York, 2002.

[18] Y. Collette and P. Siarry. *Multiobjective Optimization. Principles and Case Studies*. Cost Engineering in Practice. Springer, Berlin, 2003.

[19] K. Deb, S. Agarwal, A. Pratah, and T. Meyarivan. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. KanGAL report 200001, Indian Institute of Technology, Kanpur, India, 2000.

[20] K. Deb, S. Agarwal, A. Pratah, and T. Meyarivan. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In M. Schoenauer , editor, *Proceedings of the Parallel Problem Solving from Nature VI Conference*, pages 849–858, Paris, France, 2000. Springer Lecture Notes in Computer Science No. 1917.

[21] Department of Energy of the USA. Discussion of thermal comfort models. In *EnergyPlus Engineering Manual*, pages 171–190. 2003.

[22] Department of Energy of the USA. EnergyPlus (e^+) Building Energy Simulation Software Home Page. *http://www.eere.energy.gov/buildings/energyplus*, (Stand Oktober 2003).

EnergyPlus is a elaborated simulation system for modeling the energetic flows and requirements in buildings, taking into account environmental conditions (solar irradiation, geographical orientation etc.) and usage dependent energetic inputs and losses. It is, in it's basic version, distributed free of charge.

[23] M. Driml and O. Hanš. On a randomized optimization procedure. In J. Kožešnik, editor, *Transactions of the 4th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes (held at Prague 1965)*, pages 273–276, Prague, 1967. Czechoslovak Academy of Sciences.

[24] S. Droste, Th. Jansen, and I. Wegener. Perhaps Not a Free Lunch But At Least a Free Appetizer. Technical Report CI-45/98, Universität Dortmund, Dortmund, 1998.

[25] G. Dueck and T. Scheuer. Threshold accepting: a general purpose optimization algorithm appearing superior to simulated annealing. *Journal of Computational Physics*, 90:161–175, 1990.

[26] M. Ehrgott. *Multicriteria Optimization*, volume 491 of *Lecture Notes in Economics and Mathematical Systems*. Springer, Berlin, 2000.

[27] A. E. Eiben and J. E. Smith. *Introduction to Evolutionary Computing*. Springer, Berlin, Heidelberg, 2003.

[28] Agoston E. Eiben and Marc Schoenauer. Evolutionary computing. *Information Processing Letters*, 82(1):1–6, 2002.

[29] David B. Fogel. *Evolutionary Computation: The Fossil Record*. Wiley-IEEE Press, New York, NY, 1998.

[30] Lawrence J. Fogel, A. J. Owens, and M. J. Walsh. Artificial intelligence through a simulation of evolution. In A. Callahan, M. Maxfield, and L. J. Fogel, editors, *Biophysics and Cybernetic Systems*. Spartan Books, Washington DC, 1965.

[31] G. A. Geist and V. S. Sunderam. Network based concurrent computing on the PVM system. *Journal of Concurrency: Practice and Experience*, 4 (4):293 – 311, June 1992.

[32] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley, Reading, MA, USA, 1989.

[33] Monika Grötzner and Peter Roosen. Thermodynamic evaluation of chemical processes as a base of structure-optimizing process synthesis. In *Europ. Symp. on Computer-Aided Process Engineering*, Hungary, 1999.

[34] Groundspeak Inc. Geocaching Portal. http://www.geocaching.com, September 2005.
Introduction web site to the recreational use of the GPS technique for finding small hidden treasury chests. Being a relatively new occupation this idea seems to develop towards a kind of mass sports.

[35] S. Hasse. *Taschenbuch der Geißerei-Praxis*. Schiele und Schön, Berlin, 2006.

[36] John H. Holland. Genetic algorithms and the optimal allocation of trials. *SIAM Journal of Computing*, 2(2):88–105, 1973.

[37] R. Jin, X. Du, and W. Chen. The use of metamodeling techniques for optimization under uncertainty. Journal of Structural & Multidisciplinary Optimization (in press), 2005.

[38] J. Knowles. *Local-Search and Hybrid Evolutionary Algorithms for Pareto Optimization*. PhD thesis, The University of Reading, Reading, UK, 2002.

[39] Tamara G. Kolda, Robert Michael Lewis, and Virginia J. Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review*, 45(3):385–482, 2003.

[40] John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, 1992.

[41] KH Küfer, M Monz, A Scherrer, P Süss, , FV Alonso, AS Sultan, TR Bortfeld, and C Thieke. Multicriteria optimization in intensity modulated radiotherapy planning. *Berichte des ITWM*, 77, 2005. http://www.itwm.fhg.de/zentral/download/berichte/bericht77.pdf.

[42] Sandor Markon, Dirk V. Arnold, Thomas Bäck, Thomas Beielstein, and Hans-Georg Beyer. Thresholding – A selection operator for noisy ES. In J.-H. Kim, B.-T. Zhang, G. Fogel, and I. Kuscu, editors, *Proc. 2001 Congress on Evolutionary Computation (CEC'01), Seoul*, pages 465–472, Piscataway NJ, 2001. IEEE Press.

[43] J. Matyáš. Random Optimization. *Automation and Remote Control*, 26(2):244–251, 1965.

[44] J. Mehnen. *Mehrkriterielle Optimierverfahren für produktionstechnische prozesse*. Schriftenreihe des ISF. Hrgs.: K. Weinert. Vulkan Verlag, Essen, 2005.

[45] J. Mehnen, Th. Michelitsch, Th. Bartz-Beielstein, and N. Henkenjohann. Systematic Analyses of Multi-objective Evolutionary Algorithms Applied to Real-World Problems Using Statistical Design of Experiments. In R. Teti , editor, *Intelligent Computation in Manufacturing Engineering, 4th CIRP International Seminar on Intelligent Computation in Manufacturing Engineering, CIRP ICME '04*, pages 171–178, University of Naples, Italy, 2004.

[46] J. Mehnen, Th. Michelitsch, Chr. Lasarczyk, and Th. Bartz-Beielstein. Multi-Objective Evolutionary Design of Mold Temperature Control using DACE for Parameter Optimization. In H. Pfützner and E. Leiss, editors, *Short Paper Proceedings of the ISEM 2005, 12th Interdisciplinary Electromagnetic, Mechanic and*

Biomedical Problems, volume L11-1 of *Vienna Magnetics Group Reports*, pages 464–465, Bad Gastein, 2005.

[47] K. Miettinen. Interactive Nonlinear Multiobjective Procedures. In M. Ehrgott and X. Gandibleux , editors, *Multiple Criteria Optimization*, pages 227–276. Kluwer Academic Publishers, 2002.

[48] M. Minsky. *The Society of Mind*. Simon and Schuster, New York, 1985.

[49] D.C. Montgomery. *Design and Analysis of Experiments*. John Wiley, New York, 5 edition, 2001.

[50] Rajeev Motwani and Prabhakar Raghavan. *Randomized algorithms*. Cambridge University Press, New York, NY, 1995.

[51] Thomas Nagylaki. *Introduction to Theoretical Population Genetics*. Springer, Berlin, 1992.

[52] J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7(4):308–313, 1965.

[53] V. Pareto. *Manual of Political Economy*. Augustus M. Kelley (Original in French 1927), New York, 1971.

[54] Home Page of PVM — Parallel Virtual Machine.

[55] I. Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. PhD thesis, Department of Process Engineering, Technical University of Berlin, Germany, 1971.

[56] Ingo Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann-Holzboog, Stuttgart, 1973.

[57] P. Roosen. Concept and application of the evolutionary parameter optimization toolbox EPO. *ASME-Conference "Modeling and Simulation"*, 1999.

[58] Y. Sano and H. Kita. Optimization of Noisy Fitness Functions by Means of Genetic Algorithms using History of Search. In M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J.J. Merelo, and H.-P. Schwefel, editors, *Parallel Problem Solving from Nature (PPSN VI)*, volume 1917 of *LNCS*, pages 571–580, Berlin, 2000. Springer.

[59] T.J. Santner, B.J. Williams, and W.I. Notz. *The Design and Analysis of Computer Experiments*. Springer, Berlin, 2003.

[60] T.J. Santner, B.J. Williams, and W.I. Notz. *The Design and Analysis of Computer Experiments*. Springer Series in Statistics. Springer, New York, 2003.

[61] A Scherrer, KH Küfer, M Monz, FV Alonso, and TR Bortfeld. IMRT planning on adaptive volume structures - a significant advance in computational complexity. *Berichte des ITWM*, 60, 2004. http://www.itwm.fhg.de/zentral/download/berichte/bericht60.pdf.

[62] W Schlegel and A Mahr. 3D Conformal Radiation Therapy - Multimedia Introduction to Methods and Techniques. Multimedia CD-ROM, Springer, 2001.

[63] M. Schonlau, W. Welch, and R. Jones. Global Versus Local Search in Constrained Optimization of Computer Models. *New development and applications in experimental design*, 34:11–25, 1998.

[64] H.-P. Schwefel. Kybernetische evolution als strategie der exprimentellen forschung in der strömungstechnik. Master's thesis, Technical University of Berlin, Germany, 1965.

[65] H.-P. Schwefel. *Numerische Optimierung von Computer–Modellen mittels der Evolutionsstrategie*, volume 26 of *Interdisciplinary Systems Research*. Birkhäuser, Basel, 1977.

[66] H.-P. Schwefel. *Evolution and Optimum Seeking*. Sixth-Generation Computer Technology. Wiley Interscience, New York, 1995.

[67] H.-P. Schwefel. *Evolution and Optimum Seeking*. Sixth-Generation Computer Technology Series. Wiley-Interscience, 1995.

[68] H.-P. Schwefel, G. Rudolph, and T. Bäck. Contemporary evolution strategies. In F. Morán, A. Moreno, J. J. Merelo, and P. Chacón, editors, *Advances in Artificial*

Life – Proc. Third European Conf. Artificial Life (ECAL'95), pages 893–907, Berlin, 1995. Springer.

[69] Hans-Paul Schwefel. *Evolutionsstrategie und numerische Optimierung*. Dr.-Ing. Dissertation, Technische Universität Berlin, Fachbereich Verfahrenstechnik, 1975.

[70] Hans-Paul Schwefel. Direct search for optimal parameters within simulation models. In R. D. Conine, E. D. Katz, and J. E. Melde, editors, *Proc. Twelfth Annual Simulation Symp., Tampa FL*, pages 91–102, Long Beach CA, 1979. IEEE Computer Society.

[71] R. Seindal. Gnu m4 development site. *http://www.seindal.dk/rene/gnu/*, (Stand Frühjahr 2000).

[72] Peter Stagge. Averaging efficiently in the presence of noise. In A.Eiben, editor, *Parallel Problem Solving from Nature, PPSN V*, pages 188–197, Berlin, 1998. Springer.

[73] Steven Pinker. The Seven Wonders of the World. http://pinker.wjh.harvard.edu/articles/media/1999_06_07_mcgill.html, 7 June 1999. Steven Pinker's convocation address at McGill University.

[74] E. C. Stewart, W. P. Kavanaugh, and D. H. Brocker. Study of a global search algorithm for optimal control. In *Proceedings of the 5th International Analogue Computation Meeting, Lausanne*, pages 207–230, Aug.-Sept. 1967.

[75] D.A. van Veldhuizen. *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*. PhD thesis, Air Force Institute of Technology, USA, 1999.

[76] C. Weihs and J. Jessenberger. *Statistische Methoden zur Qualitätssicherung und -optimierung in der Industrie*. Wiley-VCH, Weinheim, 1999.

[77] K. Weinert, J. Mehnen, Th. Michelitsch, K. Schmitt, and Th. Bartz-Beielstein. A Multiobjective Approach to Optimize Temperature Control Systems of Molding Tools. *Production Engineering ? Research and Development, Annals of the German Academic Society for Production Engineering (WGP)*, XI(1), 2004.

[78] Wieder GmbH. Werkzeugkühlung erhält Impulse. *MaschinenMarkt, MM Plastics*, 5, 2003.

[79] Wikipedia article. Civilization (computer game). http://en.wikipedia.org/wiki/Civilization_computer_game. Short introduction in a series of computer games called 'Civilization', with a discussion of their features and game systematics.

[80] Wikipedia article. Marsupial. http://en.wikipedia.org/wiki/Marsupial.

[81] Wikipedia article. Master equation. http://en.wikipedia.org/wiki/Master_equation. The article presents a short introduction into the concept of the time evolution of a system with discrete states into its most probable state.

[82] Wikipedia-Article. Phylogenetics. http://en.wikipedia.org/wiki/Phylogenetics. Explanation of the concept of phylogenesis, i.e. the systematic order of ancestral relationsships of biological species.

[83] Wikipedia article. Teleology. http://en.wikipedia.org/wiki/Teleology. Explanation of various philosophical concepts of an assumed targeted development, mostly of nature.

[84] D. H. Wolpert and W. G. Macready. No Free Lunch Theorems for Optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, 1997.

[85] David H. Wolpert and William G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, April 1997.

[86] Margaret H. Wright. Direct search methods: Once scorned, now respectable. In *Numerical Analysis 1995 (Proceedings of the 1995 Dundee Biennial Conference in Numerical Analysis)*, volume 344 of *Pitman Res. Notes Math. Ser.*, pages 191–208, Boca Raton, FL, 1995. CRC Press.

[87] E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: Improving the Strength Pareto Evolutionary Algorithm. Technical Report 103, Computer Engineering and Communication Networks Lab (TIK), Swiss Federal Institute of Technology (ETH), Gloriastrasse 35, CH-8092 Zürich, 2001.

[88] O. Zöllner. Optimierte Werkzeugtemperierung. Technical Report ATI 1104, Bayer, Plastics Business Group, 1997.

7

Contributing Scientists

Half a page per person. (Optional, should be discussed)