Review articles

Struct Multidisc Optim 21, 90–108 © Springer-Verlag 2001

Aims, scope, methods, history and unified terminology of computer-aided topology optimization in structural mechanics*

G.I.N. Rozvany

Abstract Topology optimization of structures and composite continua has two main subfields: **L**ayout **O**ptimization (LO) deals with grid-like structures having very low volume fractions and **G**eneralized **S**hape **O**ptimization (GSO) is concerned with higher volume fractions, optimizing simultaneously the topology and shape of internal boundaries of porous or composite continua. The solutions for both problem classes can be exact/analytical or discretized/FE-based.
This review article discusses FE-based generalized shape optimization, which can be classified with respect to the types of topologies involved, namely **I**sotropic-**S**olid/**E**mpty (ISE), **A**nisotropic-**S**olid/**E**mpty (ASE), and **I**sotropic-**S**olid/**E**mpty/**P**orous (ISEP) topologies.

Considering in detail the most important class of (i.e. ISE) topologies, the computational efficiency of various solution strategies, such as SIMP (**S**olid **I**sotropic **M**icrostructure with **P**enalization), OMP (**O**ptimal **M**icrostructure with **P**enalization) and NOM (**N**on**O**ptimal **M**icrostructures) are compared.

The SIMP method was proposed under the terms "direct approach" or "artificial density approach" by Bendsøe over a decade ago; it was derived independently, used extensively and promoted by the author's research group since 1990. The term "SIMP" was introducted by the author in 1992. After being out of favour with most other research schools until recently, SIMP is becoming generally accepted in topology optimization as a technique of considerable advantages. It seems, therefore, useful to review in greater detail the origins, theoretical background, history, range of validity and major advantages of this method.

Received October 5, 1999
Revised manuscript received May 2, 2000

G.I.N. Rozvany

Dept. of Structural Mechanics, Faculty of Civil Engineering, Budapest University of Technology and Economics, Müegyetem rkpt. 3, Kmf 35, H-1521 Budapest, Hungary
e-mail: rozvany@eik.bme.hu

* A shorter version of this review article was presented at the NATO ARW "Topology Optimization of Structures and Composite Continua" (Budapest, Rozvany 2000)

Key words topology optimization, generalized shape optimization, finite elements, SIMP method, perforated plates, composites, homogenization

1 Introduction

Topology optimization is a relatively new and rapidly expanding field of structural mechanics, which can result in much greater savings than mere cross-section or shape optimization. Owing to its complexity, it is an intellectually challenging field; its progress, however, has often been hampered by conceptual inconsistencies and terminological confusion. For this reason, a critical and systematic re-examination of the relevant issues seems warranted.

This review deals mainly with mechanical, structural and computational aspects, whilst investigations of purely mathematical interest are outside its scope.

For very *low volume fractions*, important principles of topology optimization were established already around the turn of the century, in the context of trusses, by the versatile Australian inventor Michell (1904). These were extended to grillages (beam systems) some seventy years later by Rozvany (e.g. 1972a,b). Drawing on these applications, the basic principles of *optimal layout theory* were formulated by Prager and Rozvany (e.g. 1977a) and generalized considerably by the latter in the eighties and nineties (e.g. Rozvany 1992; Rozvany and Birker 1994).

Topology optimization for *higher volume fractions* is now termed **G**eneralized **S**hape **O**ptimization (GSO) (Rozvany and Zhou 1991) or Variable Topology Shape Optimization (Haber *et al.* 1996). It involves the simultaneous optimization of the topology and shape of internal boundaries in porous and composite continua.

In the context of discretized mechanics, this development was prompted by the observation of Cheng and Olhoff (e.g. 1981) that optimized solid plates contain systems of ribs which are similar to optimized grillages. For compliance design of perforated plates (disks) in plane stress, optimal microstructures were studied by various mathematicians (e.g. Lurie *et al.* 1982; Kohn and Strang 1986; Vigdergauz 1986).

The first exact *analytical* solutions for optimal perforated plates and the correct expressions for the rigidity tensor of homogenized optimal microstructures were obtained by Rozvany, Olhoff, Bendsøe *et al.* (1985/87), and Ong, Rozvany and Szeto (1988).

The birth of practical, *FE-based* topology optimization for higher volume fractions was brought about by extensive pioneering research of (e.g. 1989, for a review see Bendsøe 1995), and his "homogenization" school (e.g. Bendsøe and Kikuchi 1988, Bendsøe, Diaz and Kikuchi 1993). This was followed by a parallel exploration of the SIMP approach, suggested orignally by Bendsøe (1989) and used extensively by Zhou and the author (e.g. Rozvany and Zhou 1991, presented in 1990), who also suggested the term "SIMP" (Rozvany, Zhou and Birker 1992).

One aim of this review article is to show that for the most important classes of topology problems (ISE and IS topologies) the so-called SIMP method has decisive advantages. The history and theoretical foundations of this method are also explained.

2 Classes of problems in FE-based generalized shape optimization (GSO) and their fields of application

In order to forestall possible conceptual and terminological misunderstandings in this field, it is necessary to define clearly classes of problems in FE-based GSO.

2.1 Types of elements used in GSO

In the terminology of this paper, an element is called

- **S**olid (S), if it is filled entirely with one material;
- **E**mpty (E), if it contains no material;
- **P**orous (P), if it contains one material and void (i.e. cavities or empty space);
- **C**omposite (C), if it contains more than one material but no void; and
- **C**omposite-**P**orous (CP), if it contains more than one material and void.

2.2 Topologies containing only isotropic-solid or empty finite elements (ISE topologies)

For simplicity, our discussion of ISE topologies is restricted to a *given finite set of elements of fixed shape* which may be *empty or filled entirely with one of several isotropic materials* of given properties.

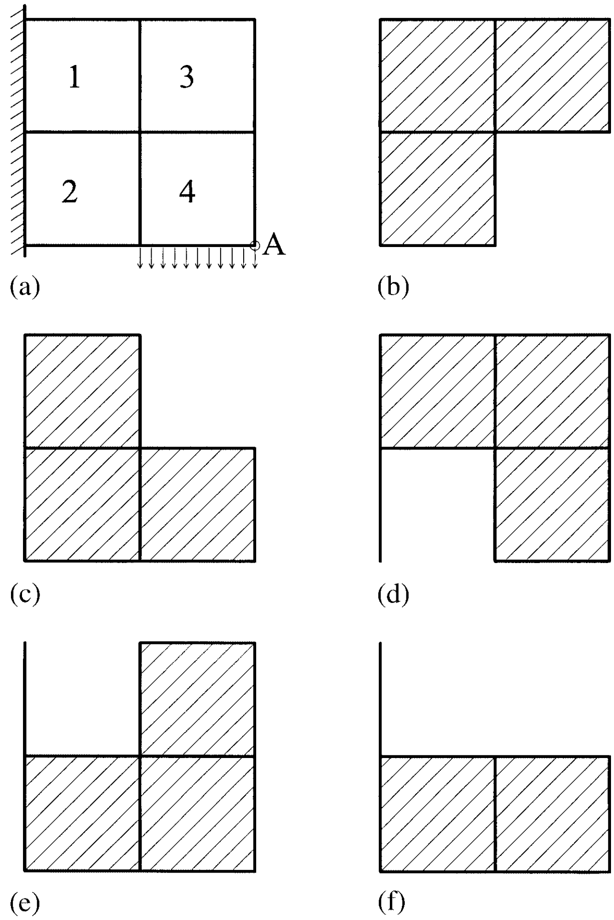

Fig. 1 Simple example involving an 1ISE topology: (a) problem statement, (b) infeasible solution, (c) optimal solution, (d–f) feasible but nonoptimal solutions

Using the terminology in Sect. 2.1, ISE topologies contain **I**sotropic-**S**olid or **E**mpty elements.

The number of base materials (or "phases") involved may also be indicated: a topology consisting of four possible materials and void will be denoted 4ISE topology. If all elements must be filled with material (i.e. there are no empty elements) then the letter E is omitted from the identifier for that topology. For example, a topology having solid elements out of four materials (phases) but no empty elements will be called a 4IS topology (choice of **4** materials for **I**sotropic-**S**olid elements).

The simplest subclass of problems in generalized shape optimization is a 1ISE (or "black-and-white" or 0-1) topology involving the optimal distribution of a single material within the design domain. An example of this type of problems is a *perforated plate* in which the plate thickness for any element is restricted to either zero or a given nonzero value.

An elementary example of this class of problems is given in Fig. 1a, in which we have $2 \times 2 = 4$ square finite elements in plane stress; two elements (1,2) are supported along their left edge and one element (4) is loaded uniformly along its bottom edge. The load is to be transmitted to the supports, so that

- the vertical displacement at the bottom right corner of element 4 (point A) is minimized,
- any element is either solid or empty, and
- the volume fraction does not exceed 0.75.

At the limiting volume fraction of 0.75, we have four possible solutions: one is infeasible[1] (Fig. 1b), one is optimal (Fig. 1c), and two are feasible but nonoptimal (Figs. 1d and e). At the volume fraction of 0.5, we have only one feasible solution (Fig. 1f) which is nonoptimal. At the volume fraction of 0.25 we have no feasible solution.

It is important to note the following.

- In practical topology optimization problems, the number of finite elements involved is very large, but the governing basic principles are exactly the same as in the above elementary example.
- The considered problem does not change essentially if we prescribe the maximum displacement at Point A and minimize the total volume or weight (with variable but uniform thickness for all nonvanishing plate elements), subject to a limit of 0.75 on the volume fraction.
- We can easily add other design constraints, restricting e.g. the stresses, natural frequencies, buckling loads, etc., but all these refer to the finite element model. This means that, for example, instead of the (theoretically) infinite stress at a reentrant corner, we consider only discretized stresses at the nodes of the finite elements or average stresses for each element. However, stress concentrations in the actual design can be avoided by a second stage shape optimization.

Optimization of a 1ISE (or "black-and-white") topology is a *discrete variable* (0-1) problem, involving 2^N possible solutions where N is the given number of the finite elements. As can be seen from the example in Fig. 1, many of the possible solutions may be infeasible. The number of possible solutions for a nISE (n material) topology is $(n+1)^N$.

Finally, it is to be noted that ISE topologies can be regarded as a special case of IS topologies, in which the density and mechanical properties of one material tend to zero and hence that material degenerates into void. For example, a 1ISE (black and white) topology is a special case of 2IS topologies.

2.2.1
Applications of ISE topologies

Discrete ISE topologies are used for practical design problems in which we want to finish up with chunks of given isotropic materials which are at least as big as the size of the elements used. Important applications are abundant in all branches of manufacturing and construction industries.

[1] "infeasible" here implies that the loaded edge is not connected by nonvanishing elemets to the the support

2.3
ASE topologies (anisotropic material, solid or empty elements)

Anisotropic elements may also be employed in discretized generalized shape optimization if we wish to finish up in our design with relatively large chunks of materials, but our choice of the latter includes anisotropic materials.

This class of topologies will be termed *ASE topologies* (involving **A**anisotropic-**S**olid or **E**mpty elements) in which the orientation and magnitude of mechanical properties (e.g. elements of the rigidity tensor $\mathbf{E}_{ijk\ell}$) are constant for each element. A general formulation of this class of problem was recently developed by Rodrigues *et al.* (1999), see also Guedes and Taylor (1997) and Taylor (1998). In ASE topologies, the relation between the usually continuously variable mechanical properties ($\mathbf{E}_{ijk\ell}$) and the element cost (or weight or resource ρ) is given and, therefore, we can use *optimization with continuous variables*. However, an element is allowed to degenerate into an *empty element* (with zero cost and zero rigidities) during the optimization process.

Figure 2 shows an elementary conceptually optimal ASE topology for the problem in Fig. 1a, assuming that only the material properties E_{1111} and E_{2222} are relevant.

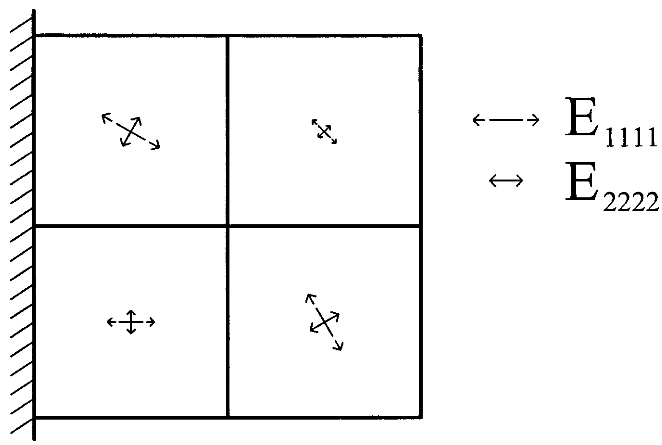

Fig. 2 Simple example of an ASE topology

2.4
ISEP, ISEC and ISECP topologies (isotropic base material, solid, empty or porous elements)

Using an FE formulation, we may try to approximate the "exact", continuum-type optimal topology for a given problem (in which the number of internal boundaries

Table 1 Fundamental properties of various types of topologies in generalized or variable-topology shape optimization

	ISE	IS	ASE	ISEP, ISEC, ISECP
void allowed	yes	no	yes	yes
base material(s)	isotropic	isotropic	anisotropic	isotropic
original elements	homogeneous	homogeneous	homogeneous	(optimally) nonhomogeneous
homogenized elements	–	–	–	anisotropic homogeneous
types of optimization problem	discrete value (0-1 for 1ISE)	discrete value	continuous†	continuous†
number of possible solutions*	$(n+1)^N$	n^N	infinite	infinite

†can be formulated as discrete value problem
* some may be infeasible, N = number of elements, n = number of base materials (phases)

is allowed to tend to infinity). In this case, for each finite element we use the *homogenized anisotropic properties* of originally nonhomogeneous elements. The latter contain an *optimal microstructure* consisting of void and one or several *isotropic* materials. The above topology is called (before homogenization) an ISEP topology (**I**sotropic base material; **S**olid, **E**mpty or **P**orous elements). After homogenization, an ISEP topology reduces to an ASE topology for computational purposes.

An elementary example of a conceptually optimal ISEP topology is given in Fig. 3, having elements with layered rank-2 microstructures.

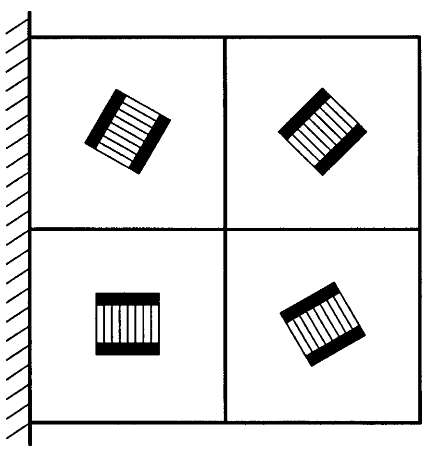

Fig. 3 Simple example of an ISEP topology

ISEC topologies are similar to ISEP topologies, but their porous elements are replaced by composite ones. Their elements are therefore solid, empty or composite. In ISECP topologies, composite-porous elements are also admitted.

2.5
Applications of ISEP topologies

ISEP (ISEC and ISECP) topologies have two very important applications.

- They may indicate the exact optimal topology for a given problem, which can be compared with exact analytical solutions for verification purposes (e.g. Jog, Haber and Bendsøe 1994). Exact solutions can be used as the absolute limit of material economy for a given design problem, providing a basis for assessing the relative degree of material economy of practical designs.
- We may wish to manufacture fibre-reinforced or other densely structured composites on the basis of "exact" optimal solutions.

The basic features of the above topologies are summarized in Table 1.

ISE and ISEP topologies were termed, respectively, SE and SEP topologies in earlier studies (e.g. Rozvany, Zhou and Birker 1992; Rozvany, Kirsch and Bendsøe 1995).

Naturally, we could restrict the material properties and microstructure parameters in ASE and ISEP topologies, respectively, to given discrete values and then we would again finish up with discrete value optimization.

3
Solution strategies of generalized shape optimization with ISE and IS topologies

Since most practical problems are associated with ISE and IS topologies at present, we will only discuss in detail methods for these problems. As before, our investigation

Table 2 Methods used for large ISE or IS topologies in generalized shape optimization

	SIMP	OMP	NOM	DDP
Microstructure of elements	solid, isotropic	optimal nonhomogeneous	nonoptimal nonhomogeneous	solid, isotropic
additional penalization	yes	yes	no	not necessary
homogenization necessary	no	yes	yes	no
no. of free parameters per element	1	2D*: 3 or 4 3D: 5 or 6	> 1	1
available for:	all combinations of design constraints	compliance	all combinations of design constraints	compliance
penalization adequate	yes	yes	no	–

*orthogonal or nonorthogonal rank-2 laminates

is restricted to topologies with a given finite number of elements.

3.1 The SIMP method

In this method, we are using **S**olid **I**sotropic **M**icrostructures with **P**enalization for intermediate densities.

Note. Some authors interpret "S" and "M" in SIMP as "Simple" and "Material". In SIMP's original definition (Rozvany, Zhou and Birker 1992; Rozvany, Bendsøe and Kirsch 1995), "S" stood for "**S**olid" (as opposite to "Porous", as e.g. in "solid gold", meaning "entirely filled with that material" and *not* as opposite to "Fluid"). The meaning of "Simple Material" or "Simple Microstructure" would be too vague. "Microstructure" in this article means the material configuration in a nonhomogeneous element or "cell". "Solid Microstructure" in SIMP was meant to refer to the limiting (degenerate) case of (nonhomogeneous) microstructures in which the entire element is occupied by one material (without cavities). Bendsøe and Sigmund (1999) use "M" for "Materials", which is a very useful alternative, but the term "materials" in this article is already used for the *base material(s)*, i.e. phase(s) in porous or composite elements.

The justification of the SIMP method can be easily understood if we consider the example of a perforated plate in plane stress, in which the plate thickness is either zero or a given value (t_0). In order to explore all possible solutions for a large number of elements (e.g. 40 000), we would have to carry out a prohibitively large number ($2^{40\,000} \cong 10^{12\,041}$) of analyses, and therefore we must resort to iterative methods with initially continuous variables.

We can, for example, assume that the plate thickness t may vary continuously between zero and t_0 (after Rossow and Taylor 1973). The mechanical properties of the plate (e.g. stiffnesses[2] for in-plane forces, permissible values of in-plane forces etc.) are linearly proportional to its thickness. We can easily minimize the weight of the above plate by using either an optimality criteria (OC) or a mathematical programming (MP) method. In fact, for a compliance constraint, this problem is convex and therefore the (only) optimum is easily and quickly calculated.

[2] In other studies these are expressed in terms of the "rigidity tensor $\mathbf{E}_{ijk\ell}$". In structural mechanics, the term "stiffness" is used instead of "rigidity"

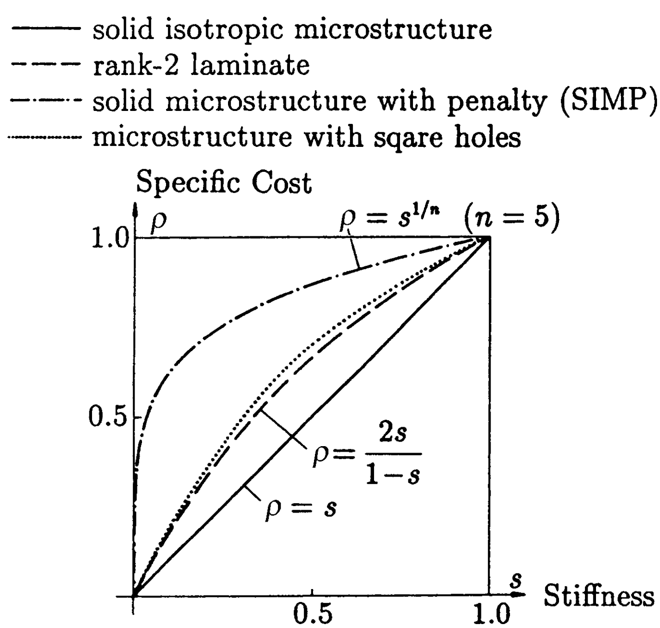

Fig. 4 Stiffness (s)/specific cost or density (ρ) relation for various types of microstructures (after Rozvany, Zhou and Birker 1992)

The catch is that the resulting solution will contain all sorts of thicknesses and in ISE topologies we want only thicknesses of zero or t_0. Our unwanted result can be largely improved if we penalize intermediate thicknesses (with $0 < t < t_0$).

This procedure is shown graphically in Fig. 4, in which the relation between the normalized plate stiffness s (in plane stress) and the specific cost or density ρ (here: plate thickness) is indicated. For a plate of varying thickness without penalty, this relation is a straight line. We can penalize the intermediate thicknesses by using the relation (after Bendsøe 1989)

$$s = \rho^p \, , \tag{1}$$

where $p > 1$. In our unpenalized procedure (after Rossow and Taylor 1973), we had $p = 1$ (continuous line in Fig. 4). The inverse of the penalized relation in (1) is shown graphically for $p = 5$ (dash-dot line in Fig. 4).

The above penalization will effectively suppress intermediate thickness values but the problem becomes nonconvex even for compliance design and therefore globality of the optimum cannot be guaranteed. The results are usually improved if we start at the beginning of the iteration with $p = 1$ and then increase p *gradually* to a higher value (say $p = 5$).

3.2 The OMP method

In this method **O**ptimal **M**icrostructures with **P**enalization for intermediate densities are used (e.g. Allaire 1997). This means that first the solution is optimized using for each finite element an optimal microstructure, derived rigorously for the particular type of design constraints and objective function(al). Such microstructures have been studied completely at present only for "compliance" design, in which the total amount of external work is either minimized or constrained.

For a 2D problem, the optimal (rank-2, layered) microstructure of an element has three free parameters (two layer densities and one orientation) and for 3D problems it has five free parameters (three layer densities and two orientations). For so-called nonselfadjoint problems (e.g. Lurie 1995), the rank-2 layered microstructures are in general nonorthogonal and hence the number of free parameters must be increased to four (2D) and six (3D).

However, optimal microstructures do not provide sufficient penalization for ISE (black-and-white) topologies (see Fig. 4, broken line). For this reason, some additional penalization is usually introduced.

3.3 The NOM method

In this method **N**on**O**ptimal **M**icrostructures or **N**ear **O**ptimal **M**icrostuctures (e.g. Bendsøe and Kikuchi 1988)

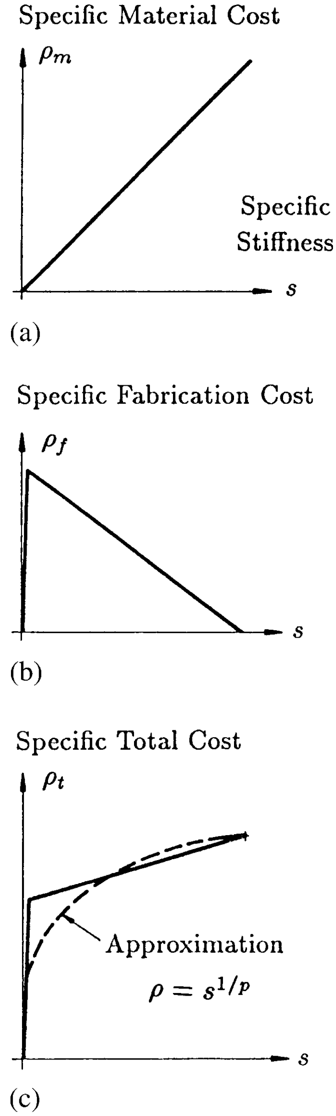

Fig. 5 Justification of the SIMP procedure on the basis of fabrication costs (Rozvany and Zhou 1991)

are used without penalty. The fact that the microstructure is nonoptimal assures a certain degree of "fixed" penalization, but this is often not adequate for an ISE or IS topology. For example, if we use square holes in our microstructure, we obtain the curve indicated in Fig. 4 (dotted line), which is far from the strongly penalized relation (with $p = 5$).

In the NOM method, the number of free parameters per element may be somewhat lower than in the OMP method (for 2D problems with square hole, for example, two instead of three).

3.4 The DDP method

The SIMP, OMP and NOM methods may be used in combination with either **O**ptimality **C**riteria (OC) or **M**athematical **P**rogramming (MP) methods. The DDP

(**D**ual **D**iscrete **P**rogramming) method is discussed separately from SIMP, because it *does not require penalization*, although it uses solid isotropic microstructures.

In this method, convex and separable approximation schemes are used to generate a sequence of explicit approximate subproblems. Each of them is solved in the dual space with a subgradient based algorithm. The very high efficiency of this method has been demonstrated by Beckers and Fleury (e.g. 1997; see also Beckers 1999), but so far this technique has only been used for compliance problems.

The four main methods used for ISE or IS topologies are summarized in Table 2.

3.5 Other ("emerging") methods for GSO of ISE topologies

Other methods in the literature for the considered class of problems include **G**enetic **A**lgorithms (GA) and "hard-kill" methods such as "**E**volutionary **S**tructural **O**ptimization" (ESO).

GA has the *advantages* of *robustness* and the potential capability of locating the *global optimum*. It is not very practical for large numbers of elements because the population for such problems is enormous (see Sect. 3.1) and the number of renalyses also very high.

A heuristic method termed alternately "hard-kill", ESO (**E**volutionary **S**tructural **O**ptimization) and ABG (**A**daptive **B**iological **G**rowth) is discussed in another review article of this issue (Rozvany 2001) under the more appropriate term SERA (**S**equential **E**lement **R**ejections and **A**dmissions method).

4 Justifications of the SIMP-type procedure by various authors

4.1 Material interpolation schemes

One possible approach to justify the $s = \rho^p$ relation in (1) is to find ranges of microstructures which generate the correct value of s and ρ for various p values. This was demonstrated recently in a very elegant study by Bendsøe and Sigmund (1999), who

- determined the limits of p values within which a micromechanical model of relation (1) can be realized, using variational bounds on the homogenized material properties of mixtures of materials, e.g. the Hashin-Shtrikman (1963) bounds; and
- constructed actual microstructures realizing relation (1) for various values of p and Poisson's ratios ν, using the "inverse homogenization" method by e.g. Sigmund (1994a).

4.2 Allowance for fabrication costs

In an independent derivation of the SIMP approach, justified the relation (1) by including fabrication costs (in addition to material costs) in the total cost.

As long as we include the final thicknesses or densities ($t = 0$, $t = t_0$) in our initially feasible set of solutions for a structure, we can choose any physical model for intermediate values. By selecting a plate varying continuously between the above values (Rossow and Taylor 1973), we obtain the specific material cost ρ_m shown in Fig. 5a. If we then assume that elements with $0 < t < t_0$ are "machined down" from an original thickness of t_0 (Fig. 6), the fabrication cost of such machining will be, say, $\rho_f = \beta(t_0 - t)$, see Fig. 5b. However, for elements of zero thickness, the fabrication cost reduces suddenly to zero (we neglect the cost of "sawing out" empty elements). If we then superimpose the specific material and specific fabrication costs, we arrive at a specific total cost ρ_t function in terms of the stiffness s which can be approximated by the relation in (1), see Fig. 5c.

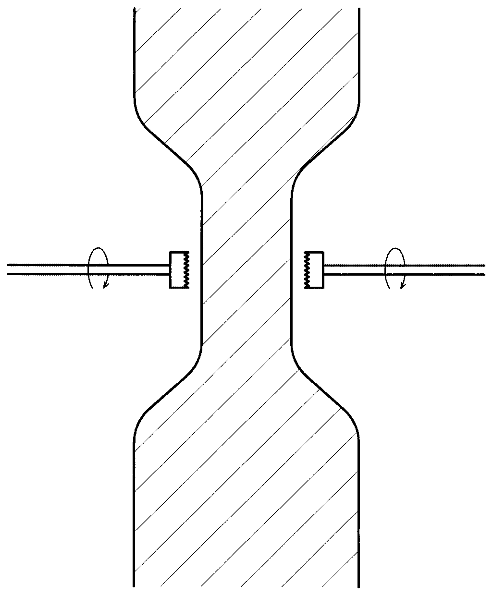

Fig. 6 Fabrication cost based on "machining" plates of varying thickness

Another possibility is to use, for example, square holes in a plate and including the length of the perimeter of the square holes multiplied by a given constant as fabrication costs ("sawing out" the square holes). In this case, the specific material cost ρ_m will be

$$\rho_m = \alpha(1 - a^2) \to a = \sqrt{1 - \rho_m/\alpha}\,, \tag{2}$$

where α is the material cost per unit plate area and a is the specific side length of the square holes. The fabrication cost becomes

$$\rho_f = \beta(4a)\,. \tag{3}$$

The ratio of the total cost (ρ_t) to material cost then becomes

$$\frac{\rho_t}{\rho_m} = \frac{\alpha(1-a^2)+\beta(4a)}{\alpha(1-a^2)}\,. \tag{4}$$

Figure 7 is a modified version of a diagram which appeared in the author's 1989 book (Rozvany 1989). The original diagram contained the lower three curves which compared the specific cost (weight) of microstructures with square holes and rank-2 laminates (with percentage difference) for given equal principal stiffnesses (rigidities). The new top curves show the total cost ρ_t for a fabrication cost factor of $\beta/\alpha = 0.2$ and the SIMP-curve for $s = \rho^5$, which are clearly similar.

One could also use circular holes in a perforated plate and cubic or spherical cavities in a 3D continuum and would arrive at relations close to (1) by allowing for the manufacturing cost of internal boundaries.

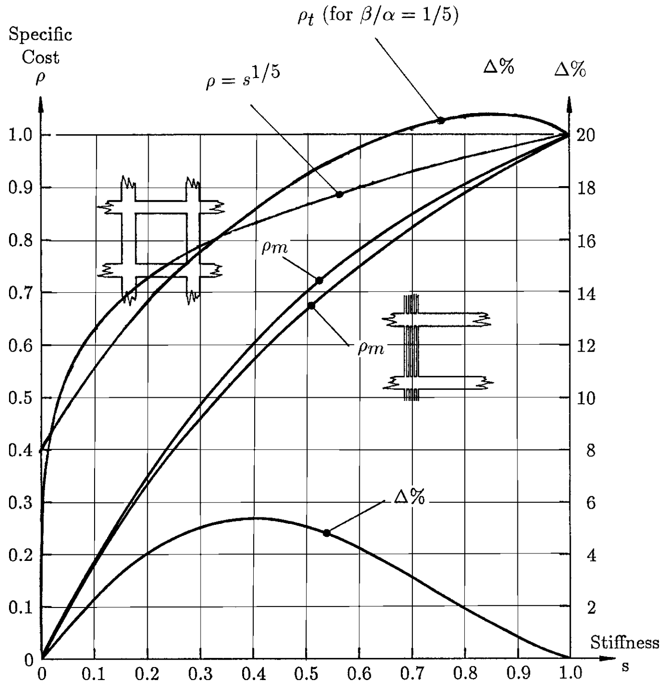

Fig. 7 Specific cost of perforated plates taking cost of "sawing out" into consideration (see ρ_t)

4.3 Penalization as computational tool in discrete value optimization

Penalization in between given discrete design values is a standard method in nonlinear optimization (e.g. Shin *et al.* (1990; Bauer 1994) which does not necessarily rely on a physical model to justify it.

5 Advantages and disadvantages of various methods for ISE (and IS) topologies

As noted previously, ISE/IS topologies are the most useful in practice because they provide a sharp, black-and-white material layout and do not contain porous regions with infinitesimal fibres, which are difficult to realise in industrial applications. Moreover, SIMP has a very large, and constantly increasing, element number capability and hence relatively dense systems of "members" can be obtained by this method in the optimal topology if such a solution is preferred. If the number of internal boundaries is to be kept low, then "perimeter control" (e.g. Haber *et al.* 1996) or mesh independent filtering (Sigmund 1994b) can be combined with SIMP.

5.1 The SIMP method

Obvious advantages of SIMP for optimizing ISE/IS topologies are as follows.

(a) *Computational efficiency* in terms of storage capacity and CPU time, since only one free variable is used per element (OMP requires up to 6, see Sect. 3.2).
(b) *Robustness* in the sense that SIMP can be readily used for *any combination of design constraints*. OMP is presently restricted to compliance or equivalent designs, because the optimal microstructure is not known for more complicated design conditions.
(c) *Penalization can be adjusted freely*, and hence the computationally optimal penalization can be used, which is not the case with NOM, for example.
(d) *Conceptual simplicity*, since the algorithm *does not require derivations involving higher mathematics*.
(e) Since the p-value in (1) is increased progressively, we can start SIMP with a solution for $p = 1$ for which some problems (e.g. compliance) are *convex* and the solution a *global optimum*. The subsequent gradual incrementation of the p-value is not likely to move the solution too far from the global optimum, but this is only an "experimental" finding at this stage.
(f) SIMP *does not require homogenization* of the microstructure.

A disadvantage of SIMP is that the solution depends on the degree of penalization (p-value) and it does not necessarily converge to the optimal solution (Stolpe and Svanberg 2001). However, other methods for ISE topologies have the same disadvantage to at least the same extent.

5.2
The OMP method

The OMP method has the following obvious disadvantages.

(a) It is *computationally* complicated and relatively *inefficient*, requiring much more free variables per element than SIMP.
(b) The optimal microstructures are developed fully at present for compliance design only, which makes OMP a *highly nonrobust* method.
(c) Derivation of the exact optimal microstructure for any new design conditions requires *advanced mathematical treatment*.
(d) The OMP problem is *intrinsically nonconvex*, even for compliance design.
(e) OMP *requires homogenization* of the microstructure which is an extra operation in comparison to SIMP.
(f) As for SIMP, the solution is dependent on the degree of penalization.

OMP has a potential advantage, if we wish to calculate both optimal ISE and ISEP topologies for a given design problem, because both can be obtained in an extended single operation (by starting with an unpenalized optimal microstructure). This, however, happens rarely in practice.

5.3
The NOM method

The NOM method may potentially have a smaller number of variables than the OMP method (e.g. square holes), which is an advantage. However, it has the following disadvantages in comparison with SIMP.

(a) NOM involves *more variables per element* than SIMP (but possibly less than OMP).
(b) *Penalization* is *fixed* and often inadequate, resulting in "grey" regions and/or a nonoptimal ISE/IS topology.
(c) As OMP, NOM is also inherently *nonconvex*.
(d) NOM also *requires homogenization*.

The relative advantages and disadvantages of various methods for ISE and IS topologies are summarized in Tables 3 and 4.

6
Theoretical objections to SIMP and their irrelevance to ISE/IS topologies

Broader adoption of the computationally superior SIMP method was delayed by almost a decade owing to certain theoretical objections to this technique.

6.1
Lack of physical interpretation of SIMP

In first introducing the SIMP method, Bendsøe (1989) commented on the efficiency of this method but noted that "it is impossible to give a physical meaning for intermediate values" (of density). For this reason, Bendsøe (1989) refers to an "artificial material" with the appropriate density-stiffness relation. At the time, Bendsøe's comments were absolutely justified.

As noted already by Rozvany and Zhou (1991) at a meeting in 1990, a certain physical interpretation of the SIMP relation in (1) can be achieved by the inclusion of manufacturing costs (e.g. the length of the perimeter of the holes) in the cost function.

As mentioned in Sect. 4.1, the above objection was completely removed by Bendsøe and Sigmund (1999), who constructed composites realizing the relation in (1) within limits on the p-value. Moreover, these authors correctly point out:

> "if a numerical method leads to black-and-white designs, one can, in essence, ignore the physical relevance of 'grey' and in many situations a better computational scheme can be obtained if one allows the violation of bounds on properties of composites".

6.2
Mesh-dependence of SIMP results

As another possible disadvantage of the SIMP method (then termed "direct approach"), Bendsøe (1989) correctly noted that the "scheme is very dependent on the mesh".

Mesh-dependence can be mostly avoided by constraining the length of the internal boundaries or "perimeter", as demonstrated in an outstanding paper by Haber *et al.* (1996). For each constrained perimeter value, the ISE topology remains stable below a critical mesh size. Alternatively, a mesh-indpendent filtering method can be used (Sigmund 1994b; Sigmund and Petersson 1998).

Whilst perimeter control is absolutely necessary if we want to restrict a practical design to a simple topology, *mesh-dependence actually becomes beneficial if we want to demonstrate that ISE topologies tend to a known exact analytical solution*.

To demonstrate this, we refer to an optimal grillage solution by Prager and Rozvany (1977b), which was also confirmed using a FE formulation by Sigmund *et al.* (1992). In this beam layout problem (Fig. 8), a rhombic grillage has simple line supports along two edges (double lines) and the other two edges (single lines) are unsupported, with a point load P at the unsupported corner. It was shown rigorously by Prager and Rozvany (1977b) that for the above problem the optimal layout consists of an infinite number of beams, having the type of beam layout shown in Fig. 8, in which continuous and broken lines

Table 3 Relative advantages of the SIMP, OMP and NOM methods in optimizing ISE and IS topologies

SIMP	OMP	NOM
(a) computational efficiency (one variable per element) (b) robustness, suitable for (almost) any design condition (c) penalization can be adjusted freely (d) conceptual simplicity, no higher mathematics required (e) convexity can be preserved for the early iterations with $p = 1$ (f) no homogenization necessary	additional information about the optimal ISEP topology	potentially smaller no. of variables per element than OMP

Table 4 Relative disadvantages of the SIMP, OMP and NOM methods in optimizing ISE and IS topologies

SIMP	OMP	NOM
solution dependent on the degree of penalization	(a) greater computational effort than SIMP (b) highly nonrobust (restricted presently to compliance design) (c) requires advanced mathematics for deriving optimal microstructures (d) intrinsically nonconvex (e) requires homogenization (f) solution dependent on the degree of penalization	(a) more variables per element than SIMP (b) penalization fixed and often insufficient for reaching the correct ISE/IS topology (c) intrinsically nonconvex (d) requires homogenization

indicate beams under positive and negative moments, respectively. For a finite number (n) of long beams, the structural weight has been shown to be (Prager and Rozvany 1977b)

$$W_n = \frac{kPa^2}{2}\left(\frac{7}{4} + \frac{3}{4n}\right), \tag{5}$$

where k is a constant. The optimal layout and optimal weight for a given n-value is *strongly mesh-dependent*, with the absolute optimal weight (for $n \to \infty$) of

$$W_{\text{opt}} = \frac{7}{8}kPa^2 . \tag{6}$$

In verifying this optimal layout by *FE-based methods* (e.g. Sigmund *et al.* 1992), we find that the strongly mesh-dependent structural weight clearly converges to the value in (6) as we use finer and finer meshes. Similarly, it is necessary to allow mesh-dependence if we wish to confirm by ISE topologies (SIMP method) any analytical solutions consisting of an infinite number of internal boundaries (see, for example, Lewinski, Zhou and Rozvany 1994, p. 417). This procedure would not be possible, if we put an effective constraint on the perimeter (i.e. total length of internal boundaries).

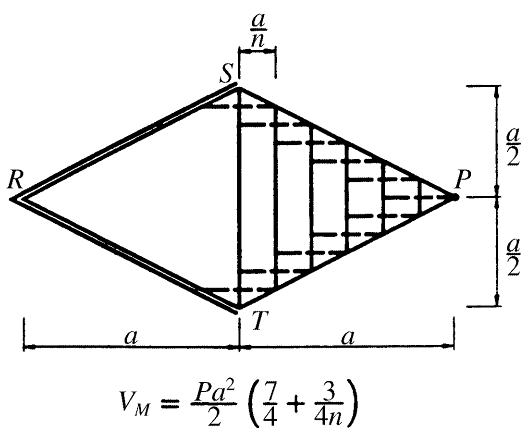

Fig. 8 A strongly mesh-dependent optimal topology

6.3 "Nonexistence" of the solution or "ill-posedness" of the problem

The SIMP method was often criticized in the past as being an "intuitive" algorithm, owing to

- "nonexistence of the solution", or
- "ill-posedness" of the problem.

Here we only mention a balanced statement by Haber *et al.* (1996), who quite correctly point out that SIMP (then called "engineering approach") "does not directly address the ill-posedness of the underlying continuum problem".

(a) *Unquestionable well-posedness of ISE/IS topology problems*. In order to forestall any misunderstanding, we must state from the outset that *discrete* optimization of ISE/IS topologies is clearly well-posed. For example, for one material/void (1ISE) problems we have 2^N (0-1)-type solutions where N is the given number of finite elements. If we generate all 2^N solutions, one (or several) of these must have the lowest objective function value and hence it (they) represents the global optimum.

Nonexistence may occur only in the sense that within a given limiting volume fraction some loaded nodes cannot be connected to the supports with nonempty elements. For this reason, we may restrict ISE/IS type problems to those cases for which at least one solution is feasible.

Within this restriction, the solution *for ISE/IS type problems* (by definition: with a finite N value) the solution clearly exists and the problem is intrinsically well-posed, even when N is extremely large (but finite). The foregoing conclusion should, in itself, disperse any doubt about the validity of the ISE/IS topology formulation for *any* practical problem: with the rapid improvement of computational capabilities, we can employ the SIMP procedure for ground structures with millions of DF's, and the problem is still be well-posed.

(b) *Irrelevance of well-posedness of the underlying continuum-type problem*. Turning to the problem of exact, continuum type problems, it will be explained that

- for a rigorous mathematical solution existence and well-posedness must be established, but
- this fact has nothing to do with the validity and computational efficiency of methods for *discrete* ISE/IS topologies (e.g. the SIMP method).

The objections associated with exact, continuum-type solutions can be best elucidated by considering again the optimal solution in Fig. 8 (with $n \to \infty$). It should be clarified that this example is a zero volume fraction problem with rank-1 microstructures but the principles to be discussed are independent of the volume fraction or the rank of the microstructure.

The force field corresponding to the optimal solution in Fig. 6a with $n \to \infty$

- has an infinite number of discontinuities, and
- takes on locally an infinite value (in mechanics: "concentrated forces"), as in the delta- or impulse-"function".

If a mathematician were to restrict his solutions to some "better-behaved" functions (e.g. functions with only a finite number of discontinuities), then he would declare that the solution to the problem in Fig. 8 "does not exist". To an engineer or "mechanician" (e.g. Prager), this statement is less important, because his optimal solutions containing an infinite number of internal boundaries and impulse-like material/force concentrations can be derived and described by using principles of mechanics. Moreover, more rigorously derived solutions by mathematicians have not established a single precedence where the mechanician's optimal solution was found incorrect.

However, rigorous mathematical studies of the underlying continuum-problem are

- of considerable theoretical interest, and
- very satisfying, since they usually confirm independently solutions derived from principles of mechanics.

In addition to identical end results, mathematical and mechanical solutions also use similar intermediate steps. The mathematician's "G-closure" (e.g. Allaire 1997, p. 110) is equivalent to including in the feasible set for the problem in Fig. 8 the solution with $n \to \infty$. Admittedly, Prager and Rozvany (1977b) did not consider some theoretical implications of this step, but they regarded the latter as less important for their frame of reference.

Moreover, mathematicians obtain a "well-behaved" class of functions by replacing the original nonhomogeneous microstructure of an element with the "homogenized" properties of an anisotropic but homogeneous element. Such "effective properties" of structures with infinitely dense microstructures were used considerably earlier in mechanics (e.g. Prager and Rozvany 1977a; to some extent also Michell 1904).

On the other hand, difficulties seem to arise in rigorous homogenization when force fields or rib densities contain impulse-like concentrations, as along the free edges in Fig. 8 with $n \to \infty$, or often in Michell trusses. For this reason, Strang and Kohn (1983), for example, used an upper constraint on rib densities in their rigorous treatment of Michell trusses.

It is to be remarked that the treatment of plane trusses by mathematicians (e.g. Strang and Kohn 1983, p. 121; Allaire 1997) is somewhat imcomplete from a mechanical viewpoint because they consider a *plane stress field* for which the integral of the sum of the absolute values of the *principal stresses* ($|\sigma_1| + |\sigma_2|$) must be minimized. This problem is equivalent to the least-weight truss problem *only if the members are restricted to the principal stress directions*. Optimal layout theory (e.g. Michell 1904; Prager and Rozvany 1977a) starts off with the much more general problem of potential truss members in all possible directions and then *proves* that in (at least one) optimal solution the members are oriented in the principal directions. However, in certain regions all directions are equally optimal and hence members in the optimal layout may be positioned in more than two directions.

7 SIMP vs. OMP – intuitive arguments and experimental evidence

In order to present a balanced view on computer-aided topology optimization, we must also mention arguments in favour of OMP.

There have been numerous claims in the literature and at conferences (e.g. Sesimbra 1992, see Bendsøe and Mota Soares 1993) that the considerable extra effort required by OMP is justified, because OMP gives more correct results. These arguments were lucidly summarized by a leading homogenization expert, (Allaire 1997, p. 129), who compared what we now call SIMP and OMP methods:

> "Of course, such an approach" (in our terminology: SIMP) "has the real advantage of being straightforward to implement. However, as we shall see, its results are not as good as the ones of the homogenization method".

After this, (Allaire 1997, p. 129) develops a type of SIMP algorithm for the particular case of compliance design, using convexification of the stress-based formulation. Then he concludes

> "The algorithm converges quickly and smoothly ... In general, the fictitious penalized design" (i.e. SIMP) "fails to have the same degree of complexity and detailed pattern as the homogenized penalized design" (i.e. OMP).

This means that in the above quotation the main argument in favour of the more complicated OMP method is that in some test examples it gave a better resolution than the SIMP method.

Clearly, the performance of the SIMP and OMP methods strongly depends on the correct level of penalization and it is not clear how convexification affected Allaire's SIMP-type solutions.

Some other numerical experiments could be quoted in favour of SIMP. For example, in extensive studies since 1990, Zhou and the author have found that *SIMP results* are actually better structured, and for low volume fractures *much closer to the corresponding Michell trusses*, than those obtained by OMP in the literature (e.g. some examples of the latter in the conference proceedings for the NATO ARW in Sesimbra 1992). Considering for example the so-called MBB-beam problem, the solution in Fig. 9 was obtained very early (Rozvany and Zhou 1991, presented in 1990) and yet it was much closer to the corresponding analytical solution (Lewinski, Zhou and Rozvany 1994) than many solutions by OMP much later.

One should also mention rational arguments in the literature, why OMP *should be* better than SIMP. Quoting again Allaire (1997):

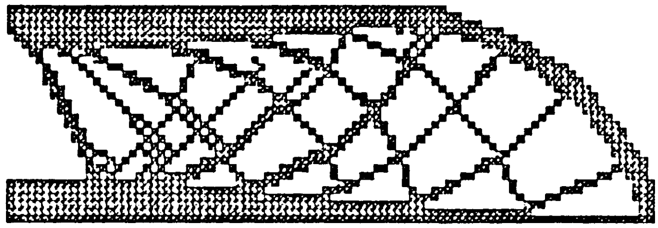

Fig. 9 Early SIMP solution for the "MBB beam" by Rozvany and Zhou (1991, presented in 1990)

> "This sensibly worse behaviour of the fictitious material approach takes its roots in the fact that there are no implicit microstructure hidden at the submesh level like for the homogenization method."

This argument would certainly be valid for the optimization of an ISEP topology *without* penalization. It could be debated, however, for ISE topologies which are *inherently* "black-and-white" ones. Since OMP is also aiming at an ISE topology at the end, it indeed "throws away" (Allaire 1997) all the information it collected about optimal anisotropy through *much extra computational effort*. Many examples show that the *optimal ISE and optimal ISEP topologies are significantly different* and hence it is not necessarily justified to get to the former via the latter.

Further evidence of the reliability of SIMP-type results come from the optimal topologies obtained recently by Beckers and Fleury (1997) and Beckers (1999) because

(a) their method of discrete dual programming is aimed directly at an optimal ISE topology without considering the optimal ISEP topology (i.e. optimal microstructures) first and in this respect it is very similar to SIMP; and

(b) their results demonstrate an *excellent resolution capability* and have been *fully confirmed* by

- 2D examples derived by Zhou and Rozvany using SIMP (for a review, see Rozvany, Kirsch and Bendsøe 1995)
- 3D examples of Olhoff *et al.* (1998) who actually used OMP, and
- exact analytical solutions (e.g. Rozvany 1998).

8 The changeable history of SIMP and related methods

8.1 Introductory comments

As mentioned in the Introduction, *optimal layout theory* deals with structures of very low (theoretically zero) volume fractions, whilst *generalized shape optimization* considers problems with higher volume fractions. It has been shown by various methods (Rozvany, Olhoff, Bendsøe *et al.* 1985/87; Allaire and Kohn 1993a, Bendsøe and Haber 1993) that for particular classes of problems the optimal topology of generalized shape optimization converges to those of optimal layout theory as the volume fraction tends to zero.

8.2 Early roots of generalized shape optimization – optimal layout theory

As noted in Sect. 1, practical FE-based topology optimization for *higher volume fractions* was pioneered by Bendsøe in Denmark (also fostered and promoted by Kikuchi) at the end of the eighties. Topology optimization for *low volume fractions* started almost nine decades earlier, and about twenty thousand km to the South-East, by a versatile Australian inventor in Melbourne. Michell's (1904) theory of least-weight trusses included in essence many features of what we now call *optimal layout theory*. Also in Melbourne, many extensions of this theory (optimal grillages and shell grids, general principles of layout theory) were explored by the research team of the author in the seventies. The earliest papers on these new applications were by Rozvany (1972a,b), but later Prager developed a keen interest in these extensions and wrote several joint papers with the author about them (e.g. Prager and Rozvany 1977a,b; Rozvany and Prager 1976, 1979).

Layout theory was generalized considerably in the eighties and nineties, most recently to multiload, multiconstraint structures (e.g. Rozvany 1992; Rozvany and Birker 1994).

The exact solutions of layout theory usually contain an intersecting system of an infinite number of members with an infinitesimal spacing, which Prager (1974) termed e.g. "truss-like continua" or "grillage-like continua". These solutions can easily be approximated numerically by optimizing ISEP topologies for low volume fractions. Moreover, optimal ISE/IS topologies with a finer mesh also clearly converge to these solutions (see excellent examples by Lewinski, Zhou and Rozvany 1994, p. 417). An early version of "perimeter control" was introduced in grillage layout optimization (Rozvany and Prager 1976), enforcing a finite number of beams.

The optimal microstructures of optimal layout theory being rank-1, the optimization procedure consists of determining at any point of the design domain the optimal directions and specific strength or stiffness of truss elements (in $\mathbb{R}^2$ or $\mathbb{R}^3$) or shell elements (in $\mathbb{R}^3$). Although optimal layout theory deals with grid- or honeycomb-type structures, these are special subclasses of problems of generalized shape optimization and hence they can be used as reliable benchmark examples for verifying the result of the latter discipline.

The field "generalized" or "variable topology shape optimization" was originally termed "advanced layout optimization", realizing that for higher volume fractions it is necessary to optimize the microstructure (for a review, see Rozvany and Ong 1987).

8.3 Beginnings of generalized shape optimization

The first step in the direction of generalized or variable topology shape optimization was a paper by Rossow and Taylor (1973). Although the use of $p = 1$ in (1) did not result in an ISE (black-and-white) topology, in private communications Taylor explained to the author already in the early seventies that his intention was to obtain cavities or perforations in the plate in some areas. Taylor's milestone contribution was therefore probably the first conceptualization of generalized shape optimization in the history of mechanics.

Progress in generalized shape optimization was brought about by two important developments.

First, a new class of optimal structures emerged in the early eighties. It was found by Cheng and Olhoff (1981) in FE-based optimization of solid plates that the optimal plate design develops a system of ribs and Prager pointed out in a letter shortly before his death in 1980 that the layout of these ribs is almost identical to the one of optimal grillages (e.g. Rozvany and Adidam 1972). Similarly, Kohn and Strang (1983) found solid, empty and porous regions in plastically designed cross-sections for torsion. The densely ribbed nature of plastically designed optimal solid plates was investigated in greater detail by Rozvany, Olhoff, Cheng and Taylor (1982).

Second, various mathematical studies established optimal microstructures for perforated plates, the most popular of which were rank-2 laminates (e.g. Lurie *et al.* 1982, Kohn and Strang 1986, Avellaneda 1987).

Drawing on the above results, Rozvany, Olhoff, Bendsøe *et al.* (1985/87)

- derived the homogenized rigidity tensor of rank-2 layered laminates for zero Poisson's ratio; and
- obtained exact optimal topologies for axisymmetric plates in flexure.

The methods used in deriving these optimal solutions were strongly influenced by the optimal layout theory discussed in Sect. 8.2.

It is important to note that in the above papers the effective stiffness calculations were based on quite simple mechanical principles which are as follows. Since we are dealing here with intersecting ribs of first and second-order infinitesimal width, the following conclusions can be reached from St. Venant's principle of elasticity.

- The second-order ribs only cause a nonuniform stress distribution in the (second-order) boundary layer of

first-order ribs. It can be shown then that the effect of such boundary layer on the deformations of a first-order rib tends to zero. In other words, it can be assumed in effective stiffness calculations that *the first-order rib is stressed biaxially but uniformly over its entire volume*.
- Similarly, nonuniformity of stresses at the end of a second-order rib only inlfluence the stress in that rib over a second-order distance from the end and the rest of the rib is uniformly stressed in its longitudinal direction only. Since the influence of the end disturbance can be shown to tend to zero, in effective stiffness calculations it can be assumed that the *second-order ribs are uniformly stressed uniaxially*.

Based on the above conclusion, the rest of the rigidity/compliance calculations is an *elementary exercise* but its results (Rozvany, Olhoff, Bendsøe *et al.* 1985/ 87) have been fully confirmed by more rigorous homogenization studies (for a review, see e.g. Bendsøe 1995).

The same results were extended to nonzero Poisson's ratio and composite two-material plates (e.g. Ong, Rozvany and Szeto 1988), and all rigidity tensors derived by "mechanical" or "engineering" methods agree with those obtained later by homogenization.

8.4
The birth of SIMP

8.4.1
The original introduction of SIMP by Bendsøe in 1989

There is not a shadow of doubt that the SIMP method, under the name "the direct approach" or "artificial density" method, was first described by Bendsøe (1989). He also noted that for $p = 1$, this method reduces to that for a variable thickness sheet, as described by Rossow and Taylor (1973).

For very good reasons mentioned earlier, SIMP was clearly *not* Bendsøe's choice of method at the time and he objected to

- net dependence and
- ficititiousness of material properties

in SIMP (see Sections 6.1 and 6.2).

In his Conclusions Bendsøe (1989) reiterated:

> "Weighting cost and complexity against generality, it seems that the most satisfactory method is to employ the porous material approach, using simple square voids at the microscale"

(in our current terminology, a NOM method). We must emphasize that the above (temporary) preference of Bendsøe for NOM does not diminish the momentous importance of his discovery of SIMP. In the same paper Bendsøe (1989) expresses a preference for NOM (square holes) over second-order laminates owing to the simplicity of the former.

8.4.2
An independent derivation of the SIMP method in 1990

As explained in Sect. 4.2, SIMP was also derived on the basis of allowance for *manufacturing costs* and announced at a meeting in Karlsruhe in 1990 (proceedings published later, see Rozvany and Zhou 1991). At the time, the above authors did not notice the computational identity of their technique with Bendsøe's (1989) "direct method", because they expressed the SIMP relation in a somewhat different form. However, there is no excuse for this oversight because the above authors even referred to Bendsøe's 1989 paper in another context.

In the paper by Rozvany and Zhou (1991), advanced examples were presented and the advantages of SIMP over the OMP and NOM methods explained. SIMP was also extended to *stress constraints* (Rozvany, Zhou and Birker 1992).

8.5
Coupling of SIMP with DCOC

The COC (**C**ontinuum-type **O**ptimality **C**riteria) method was initially used with SIMP (Rozvany and Zhou 1991; Zhou and Rozvany 1991). However, the capability of SIMP increased considerably when Zhou reformulated COC in terms of matrix methods of FE-computations and derived powerful independent proofs (Zhou 1992; Zhou and Rozvany 1992/93). The now fully FE-based algorithm is termed DCOC, adding **D**iscretized to COC.

The DCOC method was soon extended to a variety of design constraints, including multiple load conditions, combined stress, displacement, system stability, natural frequency constraints, elastic supports, temperature strains, variable prestrain, support settlements, variable loads, passive control, etc. (e.g. Rozvany and Zhou 1992; Rozvany, Zhou and Sigmund 1994). This means that the *SIMP-DCOC combined method has had the capability of optimizing ISE/IS topologies for some very complex design conditions since about 1991, whereas OMP has been restricted mostly to compliance and related designs*. Moreover, Zhou and Haftka (1994) developed an advanced, derivative based DCOC algorithm and Zhou (1995; see also Zhou and Rozvany 1996) introduced improved approximations for DCOC with eigenvalue and other problems.

It should be noted, however, that the above papers showed the power and efficiency of SIMP-DCOC only on test examples. The development of complete algorithms/software based on this method requires much further research effort.

A sample derivation of a SIMP-DCOC redesign formula is given in the Appendix.

8.6 Period of SIMP's limited popularity

The merits of the SIMP method were discussed in greater detail in a paper entitled "Generalized shape optimization without homogenization" by Rozvany, Zhou and Birker (1992), in which SIMP results were compared with designs by NOM in the literature, a SIMP output for two alternate load conditions was verified by comparision with the corresponding analytical solution and a SIMP algorithm for combined stress and displacement constraints was outlined. SIMP was also explained in great detail at a NATO ASI in Berechtesgaden in 1991 (Rozvany, Zhou, Birker and Sigmund 1993). In spite of all this effort, SIMP remained *largely out of favour* with most research schools due to objections outlined in Sect. 6.

At a NATO ARW on topology optimization in Sesimbra, 1992, much time was devoted to homogenization (i.e. OMP and NOM) methods for ISE topologies (e.g. Bendsøe, Diaz and Kikuchi 1993; Allaire and Kohn 1993b). However, the organizers (Bendsøe and Mota Soares) very fairly allocated a session to "other methods", including layout theory and SIMP.

At this session Mlejnek (1993) also presented an interesting paper in which in effect SIMP was used, coupling it with the method of moving asymptotes (Svanberg 1987).

Following a paper on the advantages of SIMP (Rozvany, Zhou, Birker and Sigmund 1993), a longer debate on SIMP vs. OMP/NOM type methods took place in Sesimbra. Most authors using homogenization methods (including eminent mathematicians) were still in favour of OMP or NOM, whilst some working in other areas (e.g. Kirsch, Haftka) supported the argument that it is a disadvantage of OMP that it is restricted to compliance design. Haftka actually coauthored a paper (Sankaranaryanam *et al.* 1993) showing that topology optimization for compliance can be highly nonoptimal for other constraints.

8.7 The "come-back" of SIMP

During the last three years, the advantages of SIMP have become quite obvious and this technique has gained a fairly general acceptance. The most important factors in SIMP's complete acceptance were probably the paper by Bendsøe and Sigmund (1999) on material interpolation schemes as well as continual and imaginative utilization of this method by Sigmund. SIMP is being employed and called often "SIMP" in a wide range of applications by most leading researchers in this field (e.g. *stress constraints:* Duysinx and Bendsøe (1997), Duysinx (1999), *optimality of bone microstructures:* Sigmund (1999), *pressure loading:* Hammer and Olhoff (2000), *material interpolation:* Bendsøe and Sigmund (1999), *multiphysics systems:* Sigmund (2000a), *a new class of composites*: Sigmund (2000b), Gibiansky and Sigmund (2000), *MEMS design:* Sigmund (2001a,b), *geometric nonlinearities:* Buhl *et al.* (2000), Pedersen *et al.* (2000), *fluid-solid interaction:* Chen and Kikuchi (2000), *software development:* Thomas and Schramm (2000), *general review:* Sigmund (2000c), *education:* Sigmund (2001a), Tcherniak and Sigmund (2001) *convergence studies:* Stolpe and Svanberg (2001), *combined optimization of material and structure:* Rodrigues, Guedes and Bendsøe (2001).

8.8 Alternative terms for SIMP

Although the term "SIMP" is used by most investigators, four alternative terms, namely "material interpolation", "artificial material", "power law" or "density" method still appear occasionally in publications. To keep the terminology uniform, it would be preferable to use "SIMP" consistently. We may note that if we use the materials derived through inverse homogenization by Bendsøe and Sigmund (1999), then neither S (for "solid") nor P (for "penalization") in SIMP is valid any more. However, the original interpretation of SIMP does not restrict this method to the variational bounds for composites.

9 On the historic achievements of Martin Bendsøe

The field of computer-aided topology optimization has been conceived, developed and brought to fruition mostly by this ingenious researcher. Out of many of his vital contributions, at least the following three represent a revolutionary break-through:

(a) development of FE methods for generating optimal ISE (black-and-white) topologies (e.g. Bendsøe and Kikuchi 1988);
(b) the concept of SIMP type (power-law) methods (Bendsøe 1989);
(c) close approximation of "exact" (micro-structured) optimal topologies (e.g. Jog, Haber and Bendsøe 1994). This latter can be used as a limit of material economy and a basis for assessing the relative economy of practial designs.

Any one of the above achievements would be sufficient to ensure immortality in the history of structural mechanics.

10 Concluding remarks

(a) Types of problem classes and solution strategies of FE-based generalized shape optimization (GSO) have been outlined.
(b) It was shown that the SIMP methodology has several advantages over other techniques (OMP, NOM).
(c) Theoretical objections to the SIMP approach have been fully removed (e.g. Bendsøe and Sigmund 1999).

Acknowledgements The author is indebted to Dr. M. Zhou for useful suggestions and for proposing the type of derivation used in the Appendix. Financial support from the Hungarian Ministry of Education, Grant FKFP 0308/2000 is gratefully acknowledged.

References

Allaire, G. 1997: The homogenization method for topology and shape optimization. In: Rozvany, G.I.N. (ed.) *Topology optimization in structural mechanics* (Proc. CISM Course, held in Udine, 1996), pp. 101–133

Allaire, G.; Kohn, R. 1993a: Explicit bounds on the elastic energy of a two-phase composite in two space dimensions. *Q. Appl. Math.* **51**, 675–699

Allaire, G.; Kohn, R. 1993b: Topology optimization and optimal shape design using homogenization. In: Bendsøe, M.P.; Mota Soares, C.A. (eds.) *Topology design of structures* (Proc. NATO ARW, Sesimbra, Portugal 1992), pp. 207–218. Dordrecht: Kluwer

Avellaneda, M. 1987: Optimal bounds and microgeometries for elastic two-phase composites. *SIAM J. Appl. Math.* **47**, 1216–1228

Bauer, J. 1994: A survey of methods for discrete optimum structural design. *Comp. Assist. Mech. Eng. Sci.* **1**, 27–38

Beckers, M. 1999: Topology optimization using a dual method with discrete variables. *Struct. Optim.* **17**, 14–24

Beckers, M.; Fleury, C. 1997: Topology optimization involving discrete variables. In: Gutkowski, W.; Mroz, Z. (eds.) *Proc. WCSMO-2, 2nd World Congress on Structural and Multidisciplinary Optimization* (held in Zakopane, 1996), pp. 533–538. Warsaw: Inst. Fund. Tech. Research

Bendsøe, M.P. 1989: Optimal shape design as a material distribution problem. *Struct. Optim.* **1**, 193–202

Bendsøe, M.P. 1995: *Methods for the optimization of structural topology, shape and material*. Berlin, Heidelberg, New York: Springer

Bendsøe, M.P.; Diaz, A.; Kikuchi, N. 1993: Topology and generalized layout optimization of elastic structures. In: Bendsøe, M.P.; Mota Soares, C.A. (eds.) *Topology design of structures* (Proc. NATO ARW, Sesimbra, Portugal 1992), pp. 159–206. Dordrecht: Kluwer

Bendsøe, M.P.; Haber, R.B. 1993: The Michell layout problem as a low volume fraction limit of the perforated plate topology optimization problem: an asymptotic study. *Struct. Optim.* **6**, 263–267

Bendsøe, M.P.; Kikuchi, N. 1988: Generating optimal topologies in structural design using a homogenization method. *Comp. Meth. Appl. Mech. Eng.* **71**, 197–224

Bendsøe, M.P.; Mota Soares, C.A. 1993: *Topology design of structures*. (Proc. NATO ARW, held in Sesimbra, Portugal, 1992). Dordrecht: Kluwer

Bendsøe, M.P.; Sigmund, O. 1999: Material interpolation schemes in topology optimization. *Arch. Appl. Mech.* **69**, 635–654

Buhl, T.; Pedersen, C.B.W.; Sigmund, O. 2000: Stiffness design of geometrically nonlinear structures using topology optimization. *Struct. Multidisc. Optim.* **19**, 93–104

Cheng, K.-T.; Olhoff, N. 1981: An investigation concerning optimal design of solid elastic plates. *Int. J. Solids Struct.* **17**, 305–323

Chen, B.-C.; Kikuchi, N. 2000: Topology optimization of structures and compliant mechanisms subject to pressure loads. *Proc. OptiCON 2000*, pp. 1–10. Irvine, CA: Altair Engineering, Inc.

Duysinx, P. 1999: Topology optimization with different stress limit in tension and compression. *Proc. WCSMO-3, 3rd World Cong. on Structural and Multidisciplinary Optimization* (held in Buffalo, NY), pp. 626–628

Duysinx, P.; Bendsøe, M.P. 1997: Topology optimization of continuum structures with stress constraints. In: Gutkowski, W.; Mroz, Z. (eds.) *Proc. WCSMO-2, 2nd World Cong. on Structural and Multidisciplinary Optimization* (held in Zakopane), pp. 527–532. Warsaw: Inst. Fund. Tech. Res.

Gibiansky, L.V.; Sigmund, O. 2000: Multiphase elastic composites with extremal bulk modulus. *J. Mech. Phys. Solids* **48**, 461–498

Guedes, J.M.; Taylor, J.E. 1997: On the prediction of material properties and topology for optimal continuum structures. *Struct. Optim.* **14**, 193–199

Haber, R.B.; Bendsøe, M.P.; Jog, C.S. 1996: A new approach to variable-topology shape design using a constraint on the perimeter. *Struct. Optim.* **11**, 1–12

Hammer, V.B.; Olhoff, N. 2000: Topology optimization of continuum structures subjected to pressure loading. *Struct. Optim.* **19**, 85-92

Hashin, Z.; Shtrikman, S. 1963: A variational approach to the theory of the elastic behaviour of multipurpose materials. *J. Mech. Phys. Solids* **11**, 127–140

Jog, C.S.; Haber, R.B.; Bendsøe, M.P. 1994: Topology design with optimized self-adaptive materials. *Int. J. Numer. Meth. Eng.* **37**, 1323–1350

Kohn, R.V.; Strang, G. 1983: Optimal design for torsional rigidity. In: Atluri, S.N.; Galagher, R.H.; Zienkiewicz, O.C. (eds.) *Hybrid and mixed finite element methods*, pp. 281–288. Chichester: Wiley & Sons

Kohn, R.V.; Strang, G. 1986: Optimal design and relaxation of variational problems. I, II, III. *Comm. Pure Appl. Math.* **39**, 113–137, 139–182, 353–377

Lewinski, T.; Zhou, M.; Rozvany, G.I.N. 1994: Extended exact solutions for least-weight truss layouts. Part I: Cantilever with a horizontal axis of symmetry. Part II: Unsymmetric cantilevers. *Int. J. Mech. Sci.* **36**, 375–398, 399–419

Lurie, K. 1995: An optimal plate. In: Olhoff, N.; Rozvany, G.I.N. (eds.) *Proc. First World Congress of Structural and Multidisciplinary Optimization, WCSMO-1* (held in Goslar, Germany), pp. 169–176. Oxford: Pergamon

Lurie, K.A.; Cherkaev, A.V.; Fedorov, A.V. 1982: Regularization of optimal design problems for bars and plates. I, II, III. *JOTA*, **37**, 499–522, **37**, 523–543, **42**, 247–282

Michell, A.G.M. 1904: The limits of economy of material in frame structures. *Phil. Mag.* **8**, 589–597

Mlejnek, H.P. 1993: Some explorations in the genesis of structures. In: Bendsøe, M.P.; Mota Soares, C.A. (eds.) *Topology design of structures* (Proc. NATO ARW held in Sesimbra, 1992), pp. 287–300. Dordrecht: Kluwer

Olhoff, N.; Rønholt, E.; Scheel, J. 1998: Topology optimization of three-dimensional structures using optimum microstructures. *Struct. Optim.* **16**, 1–18

Ong, T.G.; Rozvany, G.I.N.; Szeto, W.T. 1988: Least-weight design of perforated elastic plates for given compliance: nonzero Poisson's ratio. *Comp. Meth. Appl. Mech.* **66**, 301–322

Pedersen, C.B.W.; Buhl, T.; Sigmund, O. 2000: Topology synthesis of large-displacement compliant mechanisms. *Int. J. Numer. Meth. Eng.* (to appear)

Prager, W. 1974: *Introduction to structural optimization* (CISM course 212, held in Udine). Vienna: Springer

Prager, W.; Rozvany, G.I.N. 1977a: Optimization of structural geometry. In: Bednarek, A.R.; Cesari, L. (eds.) *Dynamical systems*, pp. 265–293. New York: Academic Press

Prager, W.; Rozvany, G.I.N. 1977b: Optimal layout of grillages. *J. Struct. Mech.* **5**, 1–18

Rodrigues, H.; Soto, C.; Taylor, J.E. 1999: A design model to predict optimal two-material composite structures. *Struct. Optim.* **17**, 186–198

Rodrigues, H.; Guedes, J.M.; Bendsøe, M.P. 2001: Hierarchical optimization of material and structure. *Struct. Multidisc. Optim.* (accpeted)

Rossow, M.P.; Taylor, J.E. 1973: A finite element method for the optimal design of variable thickness sheets. *AIAA J.* **11**, 1566–1569

Rozvany, G.I.N. 1972a: Grillages of maximum strength and maximum stiffness. *Int. J. Mech. Sci.* **14**, 651–666

Rozvany, G.I.N. 1972b: Optimal load transmission by flexure. *Comp. Meth. Appl. Mech. Eng.* **1**, 253–263

Rozvany, G.I.N. 1989: *Structural design via optimality criteria*. Dordrecht: Kluwer

Rozvany, G.I.N. 1992: Optimal layout theory: analytical solutions for elastic structures with several deflection constraints and load conditions. *Struct. Optim.* **4**, 247–249

Rozvany, G.I.N. 1998: Exact analytical solutions for some popular benchmark problems in topology optimization. *Struct. Optim.* **15**, 42–48

Rozvany, G.I.N. 2000: Problem classes and solution strategies in FE-based topology optimization. In: Rozvany, G.I.N.; Olhoff, N. (eds.) *Proc. NATO ARW on Topology optimization of structures and composite continua* (Budapest, 2000), pp. 19–36. Dordrecht:

Rozvany, G.I.N. 2001: Stress ratio and compliance based methods in topology optimization – a critical review. *Struct. Multidisc. Optim.* **21**, 109–119

Rozvany, G.I.N.; Adidam, S.R. 1972: Rectangular grillages of least weight. *J. Eng. Mch. ASCE* **98**, 1337–1352

Rozvany, G.I.N.; Birker, T. 1994: On singular topologies in exact layout optimization. *Struct. Optim.* **8**, 228–235

Rozvany, G.I.N.; Birker, T. 1995: Generalized Michell structures – exact least-weight truss layouts for combined stress and displacement constraints. Part I: General theory for plane trusses. *Struct. Optim.* **9**, 178–188

Rozvany, G.I.N.; Kirsch, U.; Bendsøe, M.P. 1995: Layout optimization of structures. *Appl. Mech. Rev.* **48**, 41–119

Rozvany, G.I.N.; Olhoff, N.; Bendsøe, M.P.; Ong, T.G.; Sandler, R.; Szeto, W.T. 1985/87: Least-weight design of perforated elastic plates I, II. DCAMM Report 306, *Int. J. Solids Struct.* **23**, 521–536, 537–550

Rozvany, G.I.N.; Olhoff, N.; Cheng, K.-T.; Taylor, J.E. 1982: On the solid plate paradox in structural optimization. *J. Struct. Mech.* **10**, 1–32

Rozvany, G.I.N.; Ong, T.G. 1987: Minimum-weight plate design via Prager's layout theory. In: Mota Soares, C.A. (ed.) *Computer aided optimal design: structural and mechanical systems* (Proc. NATO ASI, Troia, 1986), pp. 165–179. Berlin, Heidelberg, New York: Springer

Rozvany, G.I.N.; Prager, W. 1976: Optimal design of partially discretized grillages. *J. Mech. Phys. Solids* **24**, 125–136

Rozvany, G.I.N.; Prager, W. 1979: A new class of structural optimization problems: optimal archgrids. *Comp. Mech. Appl. Mech. Eng.* **19**, 127–150

Rozvany, G.I.N.; Zhou, M. 1991: Applications of COC method in layout optimization. In: Eschenauer, H.; Mattheck, C.; Olhoff, N. (eds.) *Proc. Conf. "Eng. Opt. in Design Processes"* (held in Karlsruhe 1990), pp. 59–70. Berlin, Heidelberg, New York: Springer-Verlag

Rozvany, G.I.N.; Zhou, M. 1992: Extensions of new discretized optimality criteria methods to structures with passive control. *Proc. 4th AIAA/USAF/NASA/OAI Symp. Multidisc. Anal. Optim.* (held in Cleveland, OH), pp. 288–297. Washington D.C.: AIAA

Rozvany, G.I.N.; Zhou, M.; Birker, T. 1992: Generalized shape optimization without homogenization. *Struct. Optim.* **4**, 250–254

Rozvany, G.I.N.; Zhou, M.; Birker, T.; Sigmund, O. 1993: Topology optimization using iterative continuum-type optimality criteria (COC) methods for discretized systems. In: Bendsøe, M.P.; Mota Soares, C.A. (eds.) *Topology design of structures*. Dordrecht: Kluwer

Rozvany, G.I.N.; Zhou, M.; Sigmund, O. 1994: Optimization of topology. In: Adeli, H. (ed.) *Advances in design optimization*, pp. 340–399. London: Chapman and Hall

Sankaranaryanam, S.; Haftka, R.T.; Kapania, R.K. 1993: Truss topology optimization with simultaneous analysis and design. *Proc. 33rd AIAA/ASME/ASCE/AHS/ASC Struct. Dyn. Mat. Conf.* (held in Dallas), pp. 2576–2585. Washington D.C.: AIAA

Shin, D.K.; Gürdal, Z.; Griffin, O.H., jr. 1990: A penalty approach for nonlinear optimization with discrete design variables. *Eng. Opt.* **16**, 29–42

Sigmund, O. 1994a: Materials with prescribed constitutive parameters: an inverse homogenization problem. *Int. J. Solids Struct.* **31**, 2313–2329

Sigmund, O. 1994b: *Design of material structures using topology optimization*. Ph.D. Thesis, Dept. Solid Mechanics, Technical University of Denmark

Sigmund, O. 1999: On the optimality of bone microstructures. In: Pedersen, P.; Bendsøe, M.P. (eds.) *IUTAM Symposium "Synthesis in Boid Solid Mechanics*, pp. 221–234. Dordrecht: Kluwer

Sigmund, O. 2000a: Systematic design of mechanical systems using topology optimization. In: Sienz, J. (ed.) *Engineering Design Optimization* (Proc. 2nd ASMO UK Conf., held in Swansea), pp. 5–12

Sigmund, O. 2000b: A new class of extremal composites. *J. Mech. Phys. Solids* **48**, 397–428

Sigmund, O. 2000c: Topology optimization: a tool for the tailoring of structures and materials. *Phil. Trans.: Math. Phys. Eng. Sci.* **358**, 211–228

Sigmund, O. 2001a: Design of multiphysics actuators using topology optimization. Part I: one-material structures. *Comp. Meth. Appl. Mech. Eng.* (to appear)

Sigmund, O. 2001b: Design of multiphysics actuators using topology optimization. Part II: two-material structures. *Comp. Meth. Appl. Mech. Eng.* (to appear)

Sigmund, O. 2001c: A 99 line topology optimization code written in Matlab. *Struct. Multidisc. Optim.* **21**, 120–127

Sigmund, O.; Petersson, J. 1998: Numerical instabilities in topology optimization: a survey on procedures dealing with checkerboards, mesh-dependencies and local minima. *Struct. Optim.* **16**, 68–75

Sigmund, O.; Zhou, M.; Rozvany, G.I.N. 1992: Layout optimizatioh of large FE systems by optimality criteria methods: applications to beam systems. In: Haug, E.J. (ed.) *Proc. NATO ASI: "Concurrent Engineering Tools and Technologies for Mechanical System Design"* (held in Iowa). Berlin, Heidelberg, New York: Springer

Stolpe, M.; Svanberg, K. 2001: On the trajectories of penalization methods in topology optimization. *Struct. Multidisc. Optim.* **21**, 128–139

Strang, G.; Kohn, R.V. 1983: Hencky-Prandtl nets and constrained Michell trusses. *Comp. Meth. Appl. Mech. Eng.* **36**, 207–222

Svanberg, K. 1987: The method of moving asymptotes – a new method for structural optimization. *Int. J. Numer. Meth. Eng.* **24**, 359–373

Thomas, H.L.; Schramm, U. 2000: Issues of commercial optimization software development. In: Sienz, J. (ed.) *Engineering Design Optimization* (Proc. 2nd ASMO UK Conf., held in Swansea), pp. 13–18

Taylor, J.E. 1998: An energy model for optimal design of linear continuum structures. *Struct. Optim.* **16**, 116–127

Tcheniak, D.; Sigmund, O. 2001: A web-based topology optimization program. *Struct. Multidisc. Optim.* (to appear)

Vigdergauz, S. 1986: Effective elastic parameters of a plate with a regular system of equal strength holes. *Mech. Solids* **21**, 162–166

Zhou, M. 1992: *A new discretized optimality criteria method in structural optimization*. Doctoral Dissertation, Essen University, FB 10. Also: Düsseldorf: VDI Verlag

Zhou, M. 1995: An efficient DCOC algorithm based on high-quality approximations for problems including eigenvalue constraints. *Comp. Meth. Appl. Mech. Eng.* **128**, 383–394

Zhou, M.; Haftka, R.T. 1994: A comparison of optimality criteria methods for stress and displacement constraints. *Proc. 35th AIAA/ASME/ASCE/AHS/ASC Struct. Dyn. Mat. Conf.* (held in Hilton Head, SC). Washington D.C.: AIAA

Zhou, M.; Rozvany, G.I.N. 1991: The COC algorithm, Part II: topological, geometrical and generalized shape optimization. *Comp. Meth. Appl. Mech. Eng.* **89**, 309–336

Zhou, M.; Rozvany, G.I.N. 1992/93: DCOC: an optimality criteria method for large systems. Part I: theory. Part II: algorith. *Struct. Optim.* **5**, 12–25

Zhou, M.; Rozvany, G.I.N. 1996: An improved approximation method for the DCOC method of sizing optimization. *Comp. Struct.* **60**, 763–769

Appendix. Sample derivation of SIMP-DCOC redesign formulae: displacement constraints

For illustration purposes, we consider the simple case of

- perforated plates with element thicknesses of either $t = 0$ or $t = t_0$,
- a single load condition,
- displacement constraints $(d = 1, \ldots, D)$.

The equivalent SIMP problem can be stated as follows:

$$\min \; \gamma_e A_e \sum_e (\widehat{t}_e)^{1/p} \,, \tag{7}$$

$$\text{with} \quad \widehat{t}_e = t_e / t_0 \,, \tag{8}$$

subject to

$$\sum_e \frac{A_e}{E_e} \tilde{\mathbf{f}}_{de}^T \mathbf{K}_e^{-1} \mathbf{f}_e - \mathit{\Delta}_d \leq 0 \quad (d = 1, \ldots, D) \,, \tag{9}$$

$$\widehat{t} - 1 \leq 0 \,, \tag{10}$$

where *for the element* e: γ_e is the specific weight, A_e the element area, t_e the element thickness, E_e Young's modulus, $\tilde{\mathbf{f}}_{de}$ the virtual nodal forces for the displacement constraint,[3] $\mathbf{K}_e$ the element stiffness matrix and $\mathbf{f}_e$ the element nodal forces. Moreover, $\mathit{\Delta}_d$ are the limiting displacement values and p is the penalization factor. Since in (9)

$$\mathbf{K}_e^{-1} = \frac{1}{\widehat{t}_e} \widehat{\mathbf{K}}_e^{-1} \,, \tag{11}$$

where $\widehat{\mathbf{K}}_e$ is the "normalized" element stiffness matrix for $t_e = t_0$, the Kuhn-Tucker condition becomes

$$\frac{\gamma_e A_e}{p} (\widehat{t}_e)^{(1-p)/p} - \frac{A_e}{E_e} \frac{1}{\widehat{t}_e^2} \left(\sum_e \nu_d \, \mathbf{f}_{de}^T \right) \widehat{\mathbf{K}}_e^{-1} \mathbf{f}_e + \omega = 0 \,, \tag{12}$$

where ν_d and ω are Lagrange multipliers for (9) and (10). Then by (10) we have the redesign formula

$$\widehat{t}_e = \left[\frac{p}{\gamma_e E_e} \left(\sum_d \nu_d \, \widehat{\mathbf{f}}_{de}^T \right) \widehat{\mathbf{K}}^{-1} \mathbf{f}_e \right]^{p/(p+1)} \,. \tag{13}$$

For computational convenience, the above redesign formula can easily be expressed in terms of virtual nodal displacements $\hat{\mathbf{u}}_e$, nodal displacements $\mathbf{u}_e$ and the normalized stiffness matrix $\hat{\mathbf{K}}_e$ of the elements $(e = 1, \ldots, E)$.

A *compliance constraint* is a special case of the above displacement constraints with $\widehat{\mathbf{f}}_{de} = \mathbf{f}_e$ and $\mathit{\Delta}_d = C$ where C is the limiting compliance value.

In the above derivation, the dependence of $\widehat{F}_{de}$ and F_e on $t_e (e = 1, \ldots, E)$ was ignored. This is because it was shown by Zhou and Rozvany (1992/93) that the "statically determinate" approximation is in fact exact if at least one displacement constraint is active in the above problem.

The values of the Lagrange multipliers ν_d $(d = 1, \ldots, D)$ can be determined iteratively by using a suitable (reciprocal) approximation (Zhou and Rozvany 1992/93).

[3] virtual forces are caused by "unit dummy loads" at the constrained displacements