Dealing with the combinatorial explosion of the morphological matrix in a "manual engineering design" context

Motte, Damien; Bjärnemo, Robert

Published in:
Proceedings of the 25th International Conference on Design Theory and Methodology - DETC/DTM'13

DOI:
10.1115/DETC2013-12040

Published: 2013-01-01

Document Version:
Publisher’s PDF, also known as Version of record

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
DEALING WITH THE COMBINATORIAL EXPLOSION OF THE MORPHOLOGICAL MATRIX IN A “MANUAL ENGINEERING DESIGN” CONTEXT

Damien Motte*, Robert Bjärnemo
Division of Machine Design
Department of Design Sciences
Lund University
P.O. Box 118, 221 00 Lund
Sweden
Email: damien.motte@mkon.lth.se,
robert.bjarnemo@mkon.lth.se

ABSTRACT

The morphological matrix is an important element of the engineering design methodology and is present in many textbooks. This method originally aimed at generating an exhaustive set of solutions for a given problem, by decomposing it into subproblems, finding solutions to each subproblem, and combining them. One issue associated with the morphological matrix has been the necessity to deal with the combinatorial explosions of solutions, especially at the conceptual design phase, when the still fuzzy nature of the design problem precludes the use of automated search for an optimal solution by means of specific algorithms (the “manual engineering design” context), apart from a few exceptions. Several heuristics based on the reduction of the number of combinations are investigated, and their efficiency is assessed. It is showed that the often-recommended compatibility matrix heuristic is the least efficient and can result in overlooking potentially interesting combinations. In fact all heuristics, even combined, generally fail to decrease the number of combinations to a level that can be handled by the designers, unless the original number of combinations is low. However, if one abandons the principle of an exhaustive investigation of the combinations in order to find the “best” solution, it can be showed statistically that the probability of ending up with a “good” concept among a very large number of combinations can be attained. Moreover, it is showed that the number of combinations one is willing to investigate also can contribute to increase this probability. Moreover the experience gained from the first round of investigation can serve as a guide to choose and assess other combinations. Based on those results, some recommendations for using the morphological matrix with all the different heuristics are given. Moreover, this paper discusses and relativizes the importance of the combinatorial explosion issue of morphological matrix compared with some other advantages and shortcomings of the method.

INTRODUCTION

Introduced in the 40s by Zwicky [1-4], the morphological approach, also called the morphological matrix, has quickly become an important element in engineering design methodology. Kesselring introduces it in Germany in 1955 [5], and makes it an integrated part of the systematic design approach [5]. In the Anglo-Saxon world, the paper by Norris [6] affirms the active use of the morphological matrix in the early 60s. The morphological matrix is now present in most textbooks on engineering design, such as Pahl and Beitz, [7], Ulrich and Eppinger [8], Ullman [9], Roozenburg and Eekels [10], Dym and Little [11] and Ehrlenspiel [12].

The morphological matrix aims principally at generating an exhaustive set of solutions for a given problem, by decomposing it into subproblems, finding solutions for each subproblem, and combining them. The method’s strength, its ability to propose a very large number of solutions, is also its Achilles’ heel: it presents the well-known drawback of a combinatorial explosion. The number of possible combinations

* Address all correspondence to this author.
The strategy adopted is Figure 1, that is the “manual engineering design context” (as expressed in [13;14]), that is, when the concepts are manually generated by engineering designers, alone or in team. This is still the way the morphological matrix is most widely used. In manual engineering design context, several heuristics have been proposed in the literature in order to reduce the number of combinations. This publication presents several of those heuristics and their relative efficiency is discussed.

In a first part (the four next sections), the use of the morphological matrix in the manual engineering design context is presented; it is also showed that the method presents several advantages other than that of a creative method, and other shortcomings than that of the combinatorial explosion. The main heuristics from the literature are then presented in larger detail (the automated or semi-automated tools and methods are also briefly discussed) and their efficiency discussed together with an illustration. Finally, in the light of some statistical considerations, some recommendations are proposed.

**MAIN USAGE OF THE MORPHOLOGICAL MATRIX: A CREATIVE METHOD**

The morphological analysis, also called morphological matrix or morphological box, is one of the central methods of systematic design methodology, that aims among other things at finding an optimal product-to-be, while not overlooking some potentially interesting solutions. The strategy adopted is breadth-first top-down, which means first finding the largest possible number of abstract solutions (breadth-first) and then more concrete ones (top-down). This concretization follows the model of the technical system, or TS [7;15;16]. The TS is described in terms of an overall function (purpose of the system) that can be decomposed into several subfunctions. Technical solutions that can realize these functions are described at an abstract level, called working principles. A product’s working principle refers here to the technical realization of the basic laws of nature (emanating from biology, chemistry or physics) which, alone or in combination, generates the function of the product – its way of working or functioning. The combination of the working principles constitutes a solution principle [7] or a concept [8]. These solutions are embodied in a component’s structure. The approach emphasized in systematic design methodology for finding an optimal solution principle is decomposition-combination, also called the factorization method [7, pp. 53, 61], that is the division of a problem into sub-problems, the finding of solutions for sub-problems and the combination of those solutions into an overall solution principle, or concept. Using the TS model, that amounts to dividing the overall function (the problem) into a structure of subfunctions, finding suitable working principles for each subfunction, combining them, evaluating them and selecting the best solution principle. The embodiment of the solution principle is then designed during the embodiment and detail phases (Figure 1).

The morphological matrix is naturally used at the step of working principle combination. The systematic design approach is the dominant design process model, and therefore the morphological matrix is found in many textbooks, e.g. [7-10;12;17;18].

Basically, the morphological matrix is no more than a classification scheme [19] with the functions of the TS present in the first row followed by the working principles that achieve the functions (Figure 2). This scheme, however, forces the designer to search for an exhaustive set of solutions for each function. Then, by going through every possible combination and selecting the best one, the designer is ensured to have found the best concept for the problem.

![Figure 1](image1.png)

**Figure 1.** The systematic design process model, with emphasis on the decomposition-combination steps (from [7] and [20])

| Function 1 | WP_{11} | WP_{12} | WP_{13} | WP_{14} |
| Function 2 | WP_{21} | WP_{22} | WP_{23} |
| ... | ... | ... | ... |
| Function f | WP_{f1} | WP_{f2} |

![Figure 2](image2.png)

**Figure 2.** Example of a morphological matrix

The number of combinations of working principles increases exponentially with the number of functions and solutions. A method often recommended in order to decrease
the number of combinations is the search for incompatibilities between working principles, by comparing pairwise the compatibility of the working principles. This technique is sometimes called the compatibility matrix [7, Section 8.2.3;19, Section 5.3], or cross-consistency assessment [21, pp. 795-796]: the working principles are listed as entries in the top row and top column, and each cell of the matrix contains the result of one comparison.

Dartnall and Johnston’s [22] case study provides an example of a thorough use of the morphological analysis. The method was applied to the search for novel down-hole water lifting piston pumps. They could retrieve designs of down-hole pumps that were patented over the last 100 years, as well as some configurations that were not found in the patent literature. Using this method for jet propulsion systems, Zwicky generated technical solutions that resulted in 16 patents [3, p. 199;4]; he also describes other technical areas were the exhaustive set of possible combinations was determined, such as radio-waves measurement techniques (radar), and physico-chemical processing techniques of single threads (monofilaments) like silk or nylon [3, p. 195;4]. According to Jones, the method has been used successfully in such engineering problems as transporting oil without tankers and providing a rapidly moveable rain cover for a cricket pitch [23, p. 295]. See Norris [6] for further examples.

Compared to other solution generation methods, the morphological matrix is a relatively efficient creative method. In an empirical study, Ekvall [24] compared four creative problem-solving methods: analogy (A), brainstorming (B), the discussion method (D), and the morphological analysis. (The discussion method, also known as "creative management", or "problem-solving group discussion", is based on the central role of the discussion leader for a fruitful session: make sure that all participants speak their mind, re-launch the discussion, propose new angles, guide the idea discussions, etc. [25;26]). 24 engineers of various experience levels (not students) participated (4 groups of 6 persons) and had to solve 18 technical problems during one week. The problems were composed of 9 inventive problems and 9 improvement problems; 4 problems per method and 2 problems with the methods of their choice. The morphological matrix produced the smallest number of ideas that were judged "useful" by the participants among all methods, but the highest proportion among all generated ideas (69%; A: 17%, B: 41%, D: 42%). The ideas selected for all problems were also judged by a panel of 3 experts within engineering design for their usefulness. For inventive problems, the morphological analysis got the highest score (1.67 on a scale from 0 to 4); D and B got 1.42, and all three methods were significantly better than A (.73). For improvement problems, the morphological analysis method came second with 1.50 (D: 2.00, A: 1.27, B: 1.42).

OTHER USAGES OF THE METHOD

The morphological matrix possesses other indirect but essential advantages beyond that of finding the best concept for a given overall function. It obliges the designer to structure her work and systematically search for variants for each subfunction. It prevents the designer from focusing on one idea and overlooking potentially better ones. This is one of the early issues that triggered the need for a systematic design process [27, p. 22].

The morphological matrix also makes it possible to spot solution principles that were not represented in the developed function structure. Pahl and Beitz present an example of conceptual design, the impulse-loading test rig, where the morphological matrix used is illustrated [7, Section 6.6.2] and presented below in the section "The impulse-loading test machine application". Several of the variants developed from the morphological matrix do not correspond to the different developed function structures; for example Variant 7 has only 3 working principles while the function structure consists of 4 functions (compare Figure 6.47, p. 218 with Figure 6.45, p. 216).

By using the morphological matrix, the designer automatically documents the different working principles she has considered, both the combinations that are relevant and the ones that are not, and the incompatibilities between working principles. She can motivate many decisions on firm grounds, and she can re-use a large part of what has been done in future project.

The morphological matrix is also a powerful collaboration and communication tool. It allows presenting one's work in a synthetic and understandable way. In larger projects, when different teams are working on different functions, the morphological matrix is a good foundation for discussion [28]. In a series of workshops, Zeiler and colleagues investigated the collaborative aspect of the morphological matrix [30]. They showed that the morphological matrix was extensively used for communication support by architecture students — 64% of the time, vs. 29% of the time for design students. These are designers that have difficulties to get a shared understanding with other professionals due to their lack of experience. For experienced designers, the figures are inversed: 71% of the time is dedicated to design activity and 31% of the time for communication [29].

SHORTCOMINGS

Combinatorial explosion

The combinatorial explosion is an important drawback of the method. In most, the number of solutions to investigate is quickly overwhelming. If \( n_i \) is the number of working principles for each function \( i, i = 1, \ldots, f \), the total number of possible combinations \( c \) is generally estimated by the following equation [7, p. 104;22;30, p. 60]:

\[
c = n_1 \cdot n_2 \cdot \cdots \cdot n_i \cdot \cdots \cdot n_f
\]  

Consequently, the number of possible combinations augments exponentially with the number of functions and the number of working principles. There are different ways to deal with them, and this will be discussed in the next section.

The combinatorial explosion is not the only shortcoming. Other issues arise with the use of the method.
Completeness of the solution set

There is a whole set of arguments showing that the morphological matrix does not ensure completeness of the solution in most cases.

The very fact that the morphological matrix allows finding other function structures (as showed with Pahl and Beitz’ example above) is a strong case against the supposition that the method delivers an exhaustive set of possible solutions. The solutions found with the help of the morphological matrix depend on the original function structures. Many different function structures can be developed, and it is difficult to ensure that some are not overlooked, see e.g. [31, pp. 203-204]. Thus it does not ensure that there is not a much better concept to find elsewhere.

The decomposition-combination approach also forces the designer to think in term of modules and can prevent her from finding very different concepts.

Another issue pointed out by Ullman [9, p. 135] is that the method "erroneously assumes that each function of the design is independent and that each [working principle] satisfies only one function. Generally, this is not the case." Often, the same working principle can solve more than one function.

Likewise, there are cases where no working principle for a subfunction can be found (see e.g. [32, pp. A.15ff]) and the subfunction has to be further divided.

There is also no way of knowing whether the set of working principles found for each function is exhaustive. Design catalogues have been developed for that matter, e.g. [33], but they do not cover the whole range of possible specific functions.

It is also often proposed to eliminate the non-compatible combinations or sub-combinations (e.g. with the compatibility matrix), but another alternative is to propose an intermediary function that would play the role of interface between the non-compatible elements. This of course must be translated into a new function structure.

In the same vein, it is possible that an unwanted physical effect occurs when two working principles are put together. It is not always possible, with the compatibility matrix, to predict all possible undesired effects: some of them are detected at a detailed level of development, or are determined by the layout or embodiment of the TS. For example, a developed subsystem can generate more heat than planned and make some other subsystems malfunction. Or, the heat was planned, but the geometric constraints of the final TS make it impossible to evacuate, and a new cooling function must be added to the system.

Franke [34] exposes an even more general shortcoming. Generally, there is nearly no one-to-one correspondence between functions, working principles and components. Completely modular TSs, e.g. hydraulic and pneumatic systems, or electric systems, are more of an exception. The choice of an organ often leads to the addition of a function, which changes the function structure. In the same way, the choice of the components will change the organ structure, which in turn will change the function structure. Franke shows the necessary iterations with the development of a boiler feed pump [34, p. 920]. To the function “increase the pressure”, a specific pump system is proposed (multi-step centrifugal pump in a synchronized arrangement with radial separated housing and a common shaft). This solution requires the subfunction “provide shaft sealing”. One sub-solution is a “gliding ring sealing”; for this, however, a function “Protect elastomer and gliding ring from too high temperature” is needed. This can be solved by adding a "cooling system", which will require the new subfunction “control the closed cooling cycle”, etc. Claiming that the morphological matrix is a "generally valid" method [7, p. 105] is thus far from unproblematic.

This lack of completeness makes doubtful the utility of going thoroughly through the very time-consuming activities of searching for working principles, decreasing the number of combinations, investigating alternatives and evaluating them.

Learning and usability

The learning and ease of use of the method have also been experienced differently. Jones reports that “experienced designers in mechanical and structural engineering have quickly learned to use it with enthusiasm and success in areas in which they have some knowledge of problem structure and feasibility” [23, p. 295]. On the other hand, in Ekvall's [24] study the morphological method got a very low evaluation score (the lowest) on the degree of difficulty of learning (2.46 out of 7; A: 4.17, B: 6.04, D: 5.88). This was interpreted as a difficulty to have to structure a whole problem before developing solutions. Only 37.5% of the participants declared that they would probably use the morphological matrix at work for creative problems, far below A (66.7%) and B (62.5%) methods — but above D (20.8%). Likewise, Savanovic and Zeiler report that only 36% of 33 of the practitioners that participated in their workshops (see Section 3) were 'highly like' to re-use the morphological analysis [29]. 50% of their 25 students were 'highly likely' to re-use the method. Also importantly, although a panel of experts in Ekvall's study had ranked the solutions from the morphological matrix very high, the participants themselves did not. For both the inventive and improvement problems combined, the morphological method ranked last together with B. Similarly, the majority of the participants of Savanovic and Zeiler's study did not find their solution proposals beneficial (43% for the practitioners, 37% for the students). Finally, Jones reports another difficulty specific to the students: that of having to manipulate abstract elements as functions [23, p. 295]. This is not linked directly to the method itself, but hampers its use. All in all, it seems that the learning curve seems low for the morphological matrix, and this affects the appreciation of the methods and its results, even if the morphological matrix is an efficient method (see preceding section).

Finally, a minor remark concerning the combined use of the morphological matrix and compatibility is appropriate in this subsection. One of the mentioned advantages of the morphological matrix is to give the designer a good overview
of the solutions. Nevertheless, the non-compatible combinations are represented in a separate matrix (the compatibility matrix); the designer must constantly navigate between the two matrices, which can be tedious in term of usability.

**CONCLUSION OF THE FIRST PART**

To summarize the first three sections, the morphological matrix is not always used as intended and does not generally ensure completeness. That has the important implication that most of the time the “best” combination is not what the designer or the design team is looking for and therefore the combinatorial explosion is just one minor problem. There are cases, however, where the full application of the morphological approach is still interesting. In those cases, dealing with the combinatorial explosion issue is still relevant. This is developed in the next sections.

**DEALING WITH THE COMBINATORIAL EXPLOSION ISSUE**

Typically, when a technical system is well-known and well-defined, the morphological matrix can help in finding new possible solution principles. The examples of Dartnall and Johnston [22] and Zwicky [3, p. 199;4] presented above are cases in point. For those instances, the solution set can be considered exhaustive, and the combinatorial explosion of alternatives remains an issue.

The first section discusses briefly some tools and methods for automating or semi-automating the search for, and evaluation of solutions based on the morphological matrix. The subsequent section presents the impulse-loading test machine application that will be used to illustrate the heuristics proposed in the manual engineering design context.

**Automation and semi-automation of the exploration of the morphological matrix**

One possibility to automate or semi-automate the exploration of the morphological matrix is to re-use the information contained in past designs. Bryant et al. [35] proposes a computational concept generation algorithm that is based on a design repository system (similar to the NIST-repository design system [36]) where data of existing products are stored. The knowledge embedded in the design repository system can help reducing the number of alternatives, and “various measures of design needs (e.g. manufacturability, recyclability, failure etc.)[…] can be used to rank the resulting conceptual design solutions generated by this method” [35, p. 5]. The system is further enhanced [13] by a morphological matrix generator [14] searching for solutions to subproblems in the design repository system. The system of Kurtoglu and Campbell [37] is developed in the same line: design rules extracted from existing products help building new configurations.

Another possibility is to use mathematical models of the solutions, which would allow an analytic or numeric simulation of the different concepts. This was already used by Zwicky for jet engines activated by chemical energy [1, p. 125]. A more recent example is that of Gavel and others [38-41], who propose a computer-based system for aircraft concept design where each solution to subproblems is characterized by physical or statistical equations. It is possible through aggregation to evaluate each combination against a given set of criteria. A large number of solutions can rapidly be evaluated and an optimal concept can be obtained.

When these tools and methods can be used they present in obvious advantage regarding the other heuristics in drastically reducing the number of combinations or directly finding the optimal one. Often however they address some specific product types (such as aircrafts [39]). They are also time-consuming to develop and to learn. They are no yet widespread either, and therefore it is still important to consider the combinatorial explosion of the morphological matrix in the manual engineering design context.

**The impulse-loading test machine application**

The following discussion will be illustrated with Pahl and Beitz’ impulse-loading test rig example [7, Section 6.6.2]. It presents the advantage of being well documented and well spread, and although the test rig has been developed a long time ago it is still suitable to the problem at hand.

In this example, 2 function structures out of 5 have been chosen for further development (function structures 4 and 5). Each function structure has a different number of functions, and some functions appear repeatedly in the structure, see Figure 3.

In total, 4 different types of function are present: change energy (1-4), store energy (5), control energy in respect of magnitude and time (6), vary energy component (7). The numbers in parentheses correspond to Pahl and Beitz’s

---

**Figure 3.** Function structure variants 4 and 5 for impulse-loading test machine, after [7, p. 216].

Copyright © 2013 by ASME
numbering. The working principles proposed are further divided according to the type of energy they deal with (hydraulic, electrical or mechanical), see [7, p. 217]. Some working principles present in the morphological matrix have also been directly discarded. This data is summarized Table 1. The morphological matrix presented in Pahl and Beitz is incomplete, as it was given for illustrative purposes only: the function Store signal (function structure 4, Figure 3) is not included in the morphological matrix, and not all working principles are present. However, this partial morphological matrix suits the purpose of this illustration. For the sake of simplicity, the working principles for the subfunctions 5 to 7 that were divided according to the types of energy have been collapsed into one type for each subfunction.

Table 1. Number of working principles of the morphological matrix for the impulse-loading test rig

<table>
<thead>
<tr>
<th>Subfunctions</th>
<th>Total</th>
<th>Suppr.</th>
<th>Total</th>
<th>Suppr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change E (1-4)</td>
<td>27</td>
<td>20</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>El ↔ Me (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>El ↔ Hy (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Me ↔ Me (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Me ↔ Hy (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Store E (5)</td>
<td>8</td>
<td>6</td>
<td>-/-</td>
<td>8</td>
</tr>
<tr>
<td>Control E (6)</td>
<td>9</td>
<td>9</td>
<td>-/-</td>
<td>9</td>
</tr>
<tr>
<td>Vary E comp. (7)</td>
<td>5</td>
<td>5</td>
<td>-/-</td>
<td>5</td>
</tr>
</tbody>
</table>


Heuristics aiming at decreasing the number of combinations

It is first necessary to modify Eq. (1). It does not take into account the fact that the same morphological matrix can be used for several function structures, and that some of these functions can be used repeatedly in the same function structure; see application above. Working principles must be selected each time a function is used. Let \( c_j \) be the number of combinations of the function structure \( j, j = 1, \ldots, s \). Let \( a_{ij} \) be the number of times a function \( i \) is repeated within the function structure \( j; a_{ij} \) takes the value 0 when the function is not present in the function structure \( j \). The total number of possible combinations for one function structure is

\[
c_j = n_1^{a_{1j}} \cdots n_s^{a_{sj}} \cdots n_f^{a_{fj}}
\]

and the total number of combinations is:

\[
c = \sum_{j=1}^{s} c_j
\]

The following heuristics aim at decreasing the total number of combinations. They all require the designer to make a series of assessments of different kinds. This amount is estimated for two cases: the maximum number of possible assessments (use of the heuristic up to exhaustion) and the minimum number of possible assessments (minimum number of assessments required in order to observe at least one decrease of the number of combinations; this is not necessarily equal to 1).

1. It is possible to directly eliminate single working principles that for one reason or another are not interesting for the designer (lack of competence in a technical domain, not compatible with the company strategy…); see e.g. the application above. The downside of this approach is that it may exclude potentially interesting solutions. This reduces significantly the number of combinations by diminishing the values of the \( n_s \).

Number of assessments. Let \( n_{wp} \) be the total number of working principles: \( n_{wp} = \sum_{i=1}^{s} n_i \). Heuristic 1 requires examining at least \( n_{wp, min} = 1 \) working principle and at most all the \( n_{wp, max} = n_{wp} \) working principles.

2. Pahl and Beitz also recommend, for each subfunction, to arrange the subfunctions according to some extraneous parameter (e.g. type of energy) [7, p. 104]. A function repeatedly used generally accepts different inputs and outputs. For example, the first occurrence of the function "Change energy" of function structure 4 (Figure 3) of the application above accepts electrical or mechanical energy as inputs (given some TS requirements [7, p. 217]) and mechanical or hydraulic energies as outputs. The second occurrence of this function ("Change into torque") accepts mechanical or hydraulic energies as input and mechanical energy as an output. The first function needs only the solutions to subfunctions 1, 2, 4, and the second the solutions to the subfunctions 3 and 4 (see Table 1). For a function \( i \), let \( f_i \) be the number of different possible subfunctions. We have \( n_{ik} \leq n_i \), \( k = 1, \ldots, f_i \) and \( \sum_{k=1}^{f_i} a_{ij} = a_{ij} \), thus

\[
n_{ik} \leq n_i \Rightarrow n_{ik}^{a_{ik}} \leq n_i^{a_{ij}} \Rightarrow n_i^{a_{ij}} \cdots n_{ij}^{a_{ij}} \leq n_i^{a_{ij} + \ldots + a_{ij}},
\]

that is:

\[
\prod_{k=1}^{f_i} n_{ik}^{a_{ik}} \leq n_i^{a_{ij}}
\]

Equation (4) shows that applying this heuristic leads to a number of combinations always inferior or equal to the original one. Equation (2) may be rewritten as:

\[
c_j = n_1^{b_1} \cdots n_i^{b_i} \cdots n_g^{b_g}
\]

with \( g \) the total number of all instantiated functions in the function structures.

Number of assessments. Let \( b_i \) be the number of categories of each function. For each function structure \( j \), the designer will consider \( b_j \) categories \( a_{ij} \) times. The maximum number of assessments is:

\[
n_{H2,max} = \sum_{i=1}^{f} \sum_{j=1}^{s} b_i \cdot a_{ij}
\]

The minimum number of assessments is the assessment of the function which requires the least amount of comparisons, that is \( n_{H2,min} = \min(b_i \cdot a_{ij}) \).
3. The most often mentioned heuristic used to diminish the number of combinations is to identify incompatible combinations of working principles [7]-[9]. Pahl and Beitz propose using the compatibility matrix method, presented by Dreibholz [19, Section 5.3] as well as Hansen [42, Section 8.2.3]. Each working principle is compared with each other and the compatibility matrix documents the result of this comparison (abandon, defer...). There is a drawback with that heuristic that is usually not evoked. The fact two working principles are incompatible does not mean that they still won't be incompatible. For example, two working principles may not be compatible because of the heat generated by one of them, but together with a working principle from a "cooling system" function, they are.

**Number of assessments.** The working principles of one function type that is not used repeatedly do not need to be compared against each other. If the same function type is used repeatedly, they have also to be compared against each other. Finally, a working principle should not need to be assessed against itself. Let \( n_{i,rep} \) be the number of working principles of a function used repeatedly and \( n_{j,rep} \) be the number of working principles of the other functions, \( j \neq i \). The maximal number of pairwise compatibility assessments is equal to:

\[
\text{\( n_{H3,\text{max}} = \frac{n_{wp,rep}(n_{wp,rep} - 1)}{2} - \sum_{i} n_{i,rep} - \sum_{j} n_{j,rep}(n_{j,rep} - 1) \)}
\]

(7)

At the beginning of the investigation, for each pair \( \langle c,d \rangle \) of working principles that are not compatible, \( \prod_{i \in [1, g]} n_{i} \) combinations disappears: from Eq. (5), one deduces that one should begin by investigating the compatibility of the two functions that have the minimal \( n_{wp} \) to maximize the decrease of combinations. The minimum number of possible assessments is \( n_{H3,\text{min}} = 1 \).

4. It has been suggested to group functions into subsystems and to investigate these independently [22;43]. That presupposes that the designer knows that the working principles of each subsystem does not affect the other subsystems (heat, corrosion, etc.) and are compatible with each of them (if not, there is a chance that one subsystem is incompatible with the other). In particular domains, such as electronics, where each component/working principle is precisely defined, this method is applicable and Eq. (5) becomes

\[
\text{\( c_{j} = n_{\sigma(1)} \cdot \ldots \cdot n_{i} + \ldots + n_{j} \cdot \ldots \cdot n_{\sigma(g)} \)}
\]

(8)

which obviously is a number of combinations inferior to that of Eq. (5), as multiplicative expression is transformed to a partially additive one. It makes it very interesting even if one subsystem is constituted by only one function.

**Number of assessments.** There is no specific assessment, as the designer is supposed to already know about the compatibilities of the relevant working principles. The use of heuristic 4 is to be done at the beginning of the investigation as the other heuristics can be applied on the subsystems.

5. It has also been proposed to evaluate each working principle and to combine all the best ones of each function to obtain the overall solution [30, p. 60]. The number of assessments is then given as \( n_{HS} = n_{1} + \ldots + n_{l} + \ldots + n_{k} \) [30, p. 60]. However, it does imply that the dependences among working principles and the undetected effects are negligible, which makes it difficult to apply in a general case. Moreover it can be considered a special case of heuristic 4. It will therefore not be investigated further on.

6. Hansen [42, Section 8.2.3] proposes to reduce the use of the morphological matrix to the critical subsystems, which presupposes that the subsystems investigated are independent from the rest of the TS. This is therefore also a special case of heuristic 4.

**Application**

The heuristics 1 to 4 have been applied to the impulse-loading test machine example using the data presented Table 1.

Using the general formula, Eq. (2), the total number of combinations for the function structures 4 and 5 are 6,561,000 and 32,805 respectively. Counting away the suppressed working principles of the example, the number of combination becomes 2,700,000 and 18,000 resp. using the heuristic 2, the number of combinations is 2,739,000 and 13,680 with all the working principles, and 1,215,000 and 8,100 without the suppressed working principles.

If the compatibility matrix was to be applied, the number of compatibility assessments would amount to 1,044 for both function structures and 685 when combined with heuristic 1. The number of compatibility assessments is not affected by heuristic 2: organizing the morphological matrix into several groups of working principles does not impact the pairwise comparison of each one of them. At most, during the first applications of the heuristic, each assessment would delete 91,125 combinations (1.39% of the total), 50,000 (1.85%) combined with heuristic 1, 109,440 (4.00%) combined with heuristic 2, and 48,600 (4.00%) combined with heuristics 1 and 2, for the function structure 4. The results are summarized Table 2.

Heuristic 4 has been applied by dividing arbitrary the subfunctions of each function structure in two groups. For the function structure 4, the first group G1 consists in the subfunctions Change (1,2,4), Increase Energy (7), Store (5) and Release Energy (7). The second group G2 consists in Increase Energy (7), Control (6), Change into torque (3,4). This division follows the energy flow, see Figure 3. The first group G1 of the function structure 5 consists in the subfunction Control (6), the second group G2 in Change (1,2,4), Increase Energy (7), and Change into torque (3,4). Heuristic 4 has then been applied with and without the three first heuristics and the results are summarized Table 3.
consequently heuristic and \(a_{ij}\), she becomes, in order to choose \(a_{ij}\), it is difficult to compare the performance. By suppressing one working principle \(g - 1\) functions, are suppressed. If two working principles are proved incompatible (heuristic 3), all the potential combinations with the working principles of the other \(g - 2\) functions are suppressed. Consequently, heuristic 1 is always more powerful than heuristic 3. If heuristic 2 can be applied, the term \(n_i^{a_{ij}}\) becomes, at least \((n_i - n_i') \cdot n_i^{a_{ij}}\), with \(1 + a_{ij} = a_{ij}\) and \(n_i'\) the number of working principles that are not shared by the other instantiations of the function. That means that all the potential combinations of the \(n_i\) working principles with the working principles of the other \(g - 1\) functions are suppressed. Heuristic 2 is of the same order of magnitude as heuristic 1. Finally, heuristic 4 changes multiplicative terms to additive ones. Heuristic 4 is therefore most of the time much more efficient than the others. This is well illustrated by the example above.

In term of application times, it is difficult to compare the heuristics. It depends mainly on the knowledge of the designer. Heuristic 4 is the one that requires most knowledge, but also the one with the highest reward. The aim of these heuristics is to get a final number of combinations that is tractable by the designer and can be further investigated. However two problems appear. The first is that each heuristic takes time to make the necessary assessment, either because of the large number of assessments, or because of the amount of work necessary for performing each assessment. The function structure 4 in the example would require tremendous work using all heuristics to get down to say, a dozen of solution principles. The second problem is that applying all heuristics does not ensure that the remaining number of combination will be low. Pahl and Beitz recommend to "pursue only such solutions as meet the demands of the requirements list and fall within the available resources" or to "concentrate on promising combinations and establish why these should be preferred above the rest" [7, p. 105] which amounts to know already the best solutions in order to choose them. Moreover, if the designer is asked to rely on her intuition, then there is no need to apply any heuristic as she obviously will avoid the unfeasible ones. Should she pick two incompatible working principles without being aware of it, she can always abandon the development of the concept and select a new combination.

**Table 2. Number of possible combinations without heuristic 4**

<table>
<thead>
<tr>
<th>Function structure</th>
<th>Without H2</th>
<th>With H2</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>6,561,000</td>
<td>2,736,000</td>
</tr>
<tr>
<td>G2</td>
<td>1,215,000</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7,776,000</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: H: Heuristic

**Table 3. Number of possible combinations using heuristic 4**

<table>
<thead>
<tr>
<th>Function structure</th>
<th>Without H2</th>
<th>With H2</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>5,400</td>
<td>1,215</td>
</tr>
<tr>
<td>G2</td>
<td>3,000</td>
<td>900</td>
</tr>
<tr>
<td>Total</td>
<td>8,400</td>
<td>2,115</td>
</tr>
</tbody>
</table>

Abbreviations: H: Heuristic, N.A.: Not applicable

### Discussion and recommendations

**Efficiency of the heuristics**

The heuristics can have highly different levels of performance. By suppressing one working principle (heuristic 1), all the potential combinations of all the working principles of all other functions of the function structure, that is of \(g - 1\) functions, are suppressed. If two working principles are proved incompatible (heuristic 3), all the potential combinations with the working principles of the other \(g - 2\) functions are suppressed. Consequently, heuristic 1 is always more powerful than heuristic 3. If heuristic 2 can be applied, the term \(n_i^{a_{ij}}\) becomes, at least \((n_i - n_i') \cdot n_i^{a_{ij}}\), with \(1 + a_{ij} = a_{ij}\) and \(n_i'\) the number of working principles that are not shared by the other instantiations of the function. That means that all the potential combinations of the \(n_i\) working principles with the working principles of the other \(g - 1\) functions are suppressed. Heuristic 2 is of the same order of magnitude as heuristic 1. Finally, heuristic 4 changes multiplicative terms to additive ones. Heuristic 4 is therefore most of the time much more efficient than the others. This is well illustrated by the example above.

In term of application times, it is difficult to compare the heuristics. It depends mainly on the knowledge of the designer. Heuristic 4 is the one that requires most knowledge, but also the one with the highest reward. The aim of these heuristics is to get a final number of combinations that is tractable by the designer and can be further investigated. However two problems appear. The first is that each heuristic takes time to make the necessary assessment, either because of the large number of assessments, or because of the amount of work necessary for performing each assessment. The function structure 4 in the example would require tremendous work using all heuristics to get down to say, a dozen of solution principles. The second problem is that applying all heuristics does not ensure that the remaining number of combination will be low. Pahl and Beitz recommend to "pursue only such solutions as meet the demands of the requirements list and fall within the available resources" or to "concentrate on promising combinations and establish why these should be preferred above the rest" [7, p. 105] which amounts to know already the best solutions in order to choose them. Moreover, if the designer is asked to rely on her intuition, then there is no need to apply any heuristic as she obviously will avoid the unfeasible ones. Should she pick two incompatible working principles without being aware of it, she can always abandon the development of the concept and select a new combination.

**Statistical considerations and a new heuristic**

In order to assess whether applying any heuristic is useful at all, we can try to determine by how much these increase the probability that the designer will end up with a good solution principle [44].

The primary objective for the designer is to get the best solution principle but it is not always possible to ensure this due to the number of combinations. The designer may want at least to find one solution among the top combinations or to have a good probability that one of the investigated solution principle belongs there. Let \(w\) be the number top combinations that the designer targets. Let \(c\) be the total number of combinations. Let
Passing from $k = 10$ to $k = 20$ give $P(10, 10, 2,475) = 7.81\%$ and $P(10, 10, 909) = 20.04\%$, respectively.

This could be stated as heuristic 7:

7. Increase the number of solution principles to investigate.

Table 4. Table of probabilities that at least one sequence is in the set $w$ (rounded up)

<table>
<thead>
<tr>
<th>$w$</th>
<th>$c$</th>
<th>$k$</th>
<th>10</th>
<th>50</th>
<th>100</th>
<th>10$^3$</th>
<th>10$^4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>0.20</td>
<td>0.10</td>
<td>0.01</td>
<td>1.10^{-5}</td>
<td>1.10^{-7}</td>
</tr>
<tr>
<td>2</td>
<td>N.A.</td>
<td>0.36</td>
<td>0.19</td>
<td>0.02</td>
<td>2.10^{-4}</td>
<td>2.10^{-6}</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>N.A.</td>
<td>0.50</td>
<td>0.27</td>
<td>0.03</td>
<td>3.10^{-4}</td>
<td>3.10^{-6}</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>N.A.</td>
<td>0.69</td>
<td>0.42</td>
<td>0.05</td>
<td>5.10^{-4}</td>
<td>5.10^{-6}</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>N.A.</td>
<td>0.86</td>
<td>0.58</td>
<td>0.08</td>
<td>8.10^{-4}</td>
<td>8.10^{-6}</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>N.A.</td>
<td>0.97</td>
<td>0.67</td>
<td>0.10</td>
<td>1.10^{-3}</td>
<td>1.10^{-5}</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>N.A.</td>
<td>0.98</td>
<td>0.82</td>
<td>0.14</td>
<td>2.10^{-3}</td>
<td>2.10^{-5}</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>N.A.</td>
<td>1.00</td>
<td>0.90</td>
<td>0.18</td>
<td>2.10^{-3}</td>
<td>2.10^{-5}</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>N.A.</td>
<td>1.00</td>
<td>1.00</td>
<td>0.40</td>
<td>5.10^{-3}</td>
<td>5.10^{-5}</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>N.A.</td>
<td>1.00</td>
<td>0.65</td>
<td>0.01</td>
<td>1.10^{-4}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>0.89</td>
<td>0.02</td>
<td>2.10^{-4}</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4 illustrates the role of $c$ and $k$ in increasing the probability that at least one investigated solution principle belongs to the $w$ top solution principles.

$k$ be the maximal number of solution principles one is willing to investigate. The probability that at least one combination out of $k$ is among the top $w$ combinations is:

$$P(k, w, c) = 1 - \frac{c-w}{c} \cdot \frac{c-w-1}{c-1} \cdot \ldots \cdot \frac{c-w-k+1}{c-k+1} \quad (9)$$

That is:

$$P(k, w, c) = 1 - \prod_{m=0}^{k-1} \frac{c-w-m}{c-m} \quad (10)$$

When $k$ is negligible in front of $c$ and $w$, Eq. (10) becomes

$$P(k, w, c) = 1 - (1 - \frac{w}{c})^k \quad (11)$$

These equations show that this probability is driven by the three parameters $w$, $k$, $c$. In order to get an idea of the influences of the respective parameters on the probability, Table 4 presents the different probabilities of getting a solution principle for different $w$ in function of the number of randomly chosen solution principles $k$.

$w$ represents the ambition level of the designer. If $w$ is small, the total number of combinations $c$ to deals with need to be small and $k$ large for the designer to have a fairly high probability to get one of the targeted solution principles (see Table 4 for $w = 1$ and $w = 10$).

The total number of combinations $c$ has been the focus of the heuristics presented above, that is, trying to decrease its number. As Table 4 shows however, there must be a very large decrease of $c$ in order to have a significant increase in probability.

The impulse-loading test rig example above shows how difficult this is: the original numbers of possible combinations for the function structures 4 and 5 are 6,561,000 and 32,805 respectively, with 10 solution principles chosen out of these combinations, the probability of choosing at least a solution in the top 10 is $P(10, 10, 6,561,000) = 2.10^{-3}\%$ and $P(10, 10, 32,805) = 0.30\%$ respectively. Using heuristics 1, 2 and 4, one could only go down to 2,475 and 909 respectively. With this number of combinations, we have $P(10, 10, 2,475) = 6.37\%$ and $P(10, 10, 909) = 10.52\%$, respectively. Although an improvement, this is still quite low probability, and the designer needs to pursue the search for non-feasible combinations or lower his or her expectations, that is, increase $w$.

There is however another option. The designer can play with the third parameter $k$. Augmenting $k$ can significantly increase the probability of getting a good solution principle.
Figure 4. Probability that at least one investigated solution principle belongs to the $w = 10$ top combinations for different values of $k$ and $c$ (logarithmic scale). Note that the probability function is discrete but has been smoothed for readability.

Figure 5. Probability that at least one investigated solution principle belongs to the $w$ top combinations for different values of $w$, $k$, and $c$ (logarithmic scale). The maximal number of combinations for getting a .8 probability for each configuration is also indicated. Note that the probability function is discrete but has been smoothed for readability.
Importantly, Eq. (10) shows that the original number of combinations (the number of combinations at the beginning of the morphological matrix study) does not play any role. Whatever the original number of combination, the probability that at least one investigated solution principle belongs to the $w$ top combinations for different values is bounded to a certain number of combinations, given $w$ and $k$, see Figure 5. Although one cannot draw a definitive conclusion, one can speculate from Figure 5 that in a “manual engineering design” context, the number of remaining combinations must be quite low. For example, for an 80% probability that one of 20 investigated solution principles will end in the $w = 10$ top combinations, the total number of combinations cannot be superior to 139. In the case of design problems that have thousands of feasible combinations, no heuristic can directly help managing their investigations. Also, there is no need to search to decrease the total number of combinations at all costs if the designer rapidly understands that the final amount will not be manageable anyway (cf. Figure 4).

Increasing the number of investigated solution principles (heuristic 7) can however be helpful: Hansen proposes to develop and evaluate a few solutions; as there commonalities among many solutions, the experience accumulated can serve as basis to evaluate the remaining solutions. With this Bayesian approach, some strategies may be developed that would guide the design work and help choosing the first concept to develop, then the second, etc.

Selection of the solution principles to investigate

The selection of the remaining solution principles, once the heuristics have been applied, is not investigated in depth in this paper. The designer can choose solution principles based on his or her own experience, but with the risk of not finding a novel solution principle. From heuristic 7, one knows that random choice is actually a relevant strategy (note that this heuristic does not apply if the designer chooses specific combinations). Moreover, as discussed in the section on shortcomings, there is no insurance to find the best of all concepts with the morphological matrix; therefore a satisfying solution would be already a good result.

If the number of remaining combinations is relatively small, the designer can use the morphological matrix as originally intended, by studying all possible solutions (like in [22]).

Gilboa et al. [45] have drawn a parallel between the morphological matrix and designs of experiments. If one considers the morphological matrix rows (the functions) as factors with $n_i$ levels (i.e. the solution principles), the score of each combination can be considered as the “response Y that is modeled as the sum of main effects (factors) and first-order interactions” (p. 254). The minimal number of combination to estimate corresponds to the number of unknown parameters of the response model. Once the parameters are known, the scores (responses) of all combinations can be estimated and the most promising ones identified. In an illustration where 432 combinations were possible, the corresponding number of combinations to estimate was 70. This is much less than the total number of combinations; notice however that according to Eq. (9), the probability to be in the top 5% is $P(70, 5\%=21, 432) = 98.16\%$ ($P(70, 2.5\%=10, 432) = 86.05\%$); good solutions can be readily be obtained with less estimations. It is also not obvious why second-order interactions would not matter in the case of the morphological matrix.

Using the morphological matrix with the heuristics: some recommendations

It has been seen that the morphological matrix can be used as intended or more freely as a creative or collaborative tool. In the first case, it has been showed that the “best” solution can be outside the morphological matrix. In the second case, the heuristics can be used liberally to take full advantage of the morphological matrix. For example, heuristics 1 and 6 (reduce the use of the morphological matrix to the critical subsystems) are good starting points.

As a guideline, it can be recommended to begin by limiting the matrix to elements that necessitate a creative solution (heuristic 6) provided that the non-studied functions are independent of the working principles of the morphological matrix. Heuristic 4 then shall always be considered as it decreases the number of combinations by several orders of magnitude. Pahl & Beitz’ [7] matrix arrangement makes it easy to use. Then, before applying the heuristics 1, 2 and 3, it is always important to determine the total number of remaining combinations. If the number is too large, there is a little probability that the other heuristics will bring the number down to an acceptable level (see e.g. Table 3). Then heuristic 1 can be applied quickly as the criteria of elimination of working principles may be exogenous to the problem. Heuristic 2 can be also applied relatively quickly, especially for the functions that serve as input and output functions of the TS. Heuristic 3 can be used if some significant decrease is expected. Finally, the best way to increase the odds of ending up with a good concept is to increase the number of combinations to investigate (Heuristic 7). The selection of the remaining solutions has been discussed above.

CONCLUSION

This paper has presented the range of usages of the morphological matrix, its advantages and shortcomings. The latter go beyond the combinatorial problem: the morphological matrix is sometimes presented as a method which allows finding an exhaustive set of technical solutions for a given problem and is therefore presented as the method of choice if many textbooks. We have shown that this was not the case and that the morphological matrix should not be thought as a way to find the “best” concepts among all. Empirical studies have nevertheless found the method efficient in comparison to others and it is a good tool for documentation, communication and collaboration.

Several heuristics dealing with the combinatorial explosion of solutions have been investigated. The three most important
results are the following. First, it turns out that the compatibility matrix, often recommended, is the least efficient of them. Moreover, one has to remember that two incompatible working principles alone may be compatible together with others. Second, although this is counter-intuitive, trying to reduce drastically the number of total combinations when it is very large may well have no effect at all. Third, increasing the number of combinations to select can significantly increase the probability of getting a good final solution.

The discussion about the heuristics used in the manual engineering design context could also benefit the development of tools and methods semi-automating the search and evaluation of solutions based on the morphological matrix in order to reduce even more the number of proposed alternatives.

The morphological matrix had been developed in the spirit of being as exhaustive as possible and the method has been developed in that direction. The synthetic representation of so many potential concepts could perhaps be used to other purposes. Many engineering design problems are not concerned with finding the best solution among other, but have difficulties finding a solution that works. In a very scarce design space, the morphological matrix may serve as a visual support for the search. Similarly, instead of looking for feasible solutions, the morphological matrix could be used to search for bold, new designs by trying to combine very different working principles. Introduced in engineering design for more than 55 years, the morphological matrix has still a strong potential for further improvement.

REFERENCES

[22] Dartnall, J. and Johnston, S., 2005, "Morphological analysis (MA) leading to innovative mechanical design",


