Simultaneous Form Generation and Performance Evaluation: A “Two-Way” Inference Approach

Ulrich Flemming
Ardeshir Mahdavi

Department of Architecture
Carnegie Mellon University
Pittsburgh, PA 15213 USA

The conventional approach toward performance evaluation transforms given design attributes into performance indicators, and designers can improve these indicators only indirectly through the manipulation of design attributes. This paper outlines a contrasting “two-way” inference approach that allows designers to also manipulate the performance indicators directly and observe the resulting changes in design attributes. The advantages of this approach and its limitations are outlined. Methodological and implementation difficulties that arise from it are introduced, and possible solution strategies are described. A first prototype for a system that implements and demonstrates this approach is outlined. The larger debate about “functionalism” touched by this approach, and its response to it, are briefly reviewed.

Keywords: form generation, performance evaluation, form-function mapping, two-way approach.

1 Introduction

In this paper, we use the term design variables to denote the geometric (size, shape, and location) and material properties of a building and its components. These are the variables directly manipulated by designers and shown directly in construction documents such as drawings and specifications, for example, the length or thickness of a wall or the material composition of a column. We call a particular value of a design variable a design attribute, for example, the particular width of a window. Taken together, the design attributes determine the form of a building.

We use the term performance variable to refer to any aspect of function, purpose, or performance that a building or component may or may not accommodate or satisfy, for example, the daylight factor at a point in a room, or the thermal properties of a wall. Performance variables are usually expressed as performance criteria or indices. These definitions are to be understood in a very broad sense, and there is no assumption that all indices are numerical.

Given this usage, the distance between two rooms is a design variable, while the ease of moving equipment between the two rooms is a performance variable that is influ-
enced by this distance along with other design variables, such as width of openings, level changes, etc. That is to say, a performance variable is often influenced by more than one design variable. The converse is also true: one design variable is likely to influence more than one performance variable; for example, the distance between two rooms may influence not only the ease of moving equipment but also the degree of sound transmission between the two rooms. Design and performance variables thus interact in complex ways.

Design or performance variables should therefore not be considered in isolation during the design process. On the other hand, neither human designers nor computers can handle all of these variables and their interactions simultaneously. Designers are therefore faced with the challenge to devise a design process that divides the overall task into more manageable subtasks by focusing on selected variables at any time, while allowing for due consideration of interactions and dependencies.

In building design, this decomposition into subproblems occurs in two main forms:

1. The overall process is divided into phases (such as preliminary design, design development, construction planning), where each phase concentrates on a limited set of design variables which are indicated by the specific scale at which a building is considered in that phase; only those design variables are considered that can be drawn or displayed with ease at the given scale.

2. Within a phase, function usually follows form; that is, design attributes are determined first, and performance indices are then established for the given set of design attributes, possibly with assistance from analysis and simulation software; that is, the process follows a general “generate-and-test” approach.

These statements are not meant to deny the occurrence of feedback within a phase or between phases: design attributes can be changed when the evaluations are unsatisfactory, and flaws that are discovered during design development may lead to revisions of earlier decisions.

But in practice, these types of feedback occur only infrequently for various reasons. Many analysis and simulation programs have problematic user interfaces and are not integrated with a general CAAD environment. In addition, many programs require comprehensive input data (which is usually not available at the early design stage) and commonly do not assist the designer in terms of expert knowledge (input preparation, parametric analysis, result interpretation). As a result, simulation tools are often used only in the final phase of the design process and then mainly for dimensioning purposes. But there is general agreement that the early phases can benefit specifically from careful analysis and simulation, including the comparison of conceptually different design alternatives.

Propositions and efforts to improve the usability of existing simulation tools include the graphical enhancement of the user-interface; integration efforts within CAAD environments; and the development of knowledge-based systems. We introduce in this paper a “two-way” inference approach that aims at coupling form generation and performance evaluation as closely as possible, because it allows for the simultaneous or parallel execution of these tasks. We outline implementation possibilities for a tool based on this approach to allow designers to explore the interactions between design and performance variables, or between the form and function of a building, with greater ease and depth.

The next section reviews some attempts made to address this problem and the general debate about functionalism or functional determinism in light of which these efforts should be seen. Section 3 outlines the two-way inference approach and discusses methodological difficulties that arise from it. Section 4 introduces a pilot system that demonstrates
a framework for the implementation of these ideas. Section 5 briefly returns to some of the issues raised in Section 2.

2 Background

At first glance, Louis Sullivan's famous dictum that “form follows function” (Sullivan, 1988) appears to reinforce, in general, the motivation introduced above. But a closer reading of his article reveals that he was interested in a very specific formal issue intimately connected to his practice, the general form of tall office buildings that were emerging as a new building type and for which no precedents existed. He generally observed that these buildings naturally divide into three vertical sections. They start with a lower, very public floor or two. This part is followed by a middle section, in which “tiers of typical offices, having the same unchanging function, shall continue in the same unchanging form.” The top floor, finally, is reserved mainly for mechanical equipment. From this, he concluded that the resulting building should be formally divided into three vertical sections and not appear like several buildings, each of a more familiar height, “piled one upon the other until the top of the pile is reached” (Sullivan, 1988:110).

A more openly “functionalist” argument was made by Hannes Meyer, who led the Bauhaus from 1928 to 1930. He greatly admired works produced by engineers, which he characterized as “function times economy.” He declared that “building is a technical not an aesthetic process” and that “the function diagram and economic program are the main guiding principles in a building scheme” (Schnaidt, 1965). This “functional determinism” has come under attack by architects and theoreticians who view it as reductionist (Colquhoun, 1967) or fallacious (Steadman, 1979:186-208). Their arguments cannot be reviewed here in detail, but it is important in the present context to point out that a narrowly-conceived functional determinism does not hold up under its own premises. A closer inspection of the function of a building reveals that the criteria by which it is assessed are extremely varied; they may conflict with each other because they compete for the same limited resources (area, costs), or contradict each other for technical or other reasons, such as conflicting interests of client, users or the public at large. This situation has been demonstrated by Radford and Gero (1980), who show that for the design of a window in a specific situation, daylight performance and summer thermal performance conflict in the sense that improving one criterion might lead to a point where further improvements can only be achieved by diminishing the other criterion. The designs that have reached such a point form a Pareto-optimal set or “frontier” of alternatives each of which exhibits a particular trade-off between the two criteria. None of these alternatives is prima facie optimal; that is, a decision is by no means as automatic as the radical functionalists assumed and must rely on value judgments.

Even if one excludes aspects of judgment, intricate methodological problems remain when a designer is asked to find a form that satisfies a given combination of performance indices. It is, for example, often not even clear at the outset what types of components may be involved in the design of a building. In Kahn's Kimball Art Museum in Ft. Worth, Texas, for example, the roof structure consists of a series of beams with a curved cross-section, two of which can be combined to form a vault that admits light from the top and diffuses it in combination with a suspended screen which also diffuses sound. Thus we have two components that perform structural, lighting, and acoustic functions aside from playing a major role in the overall architectural composition. But we can imagine alternative solutions in which the same functions are performed by different types, even different numbers of components; e.g., the function of light diffusion may be carried
out by skylights that are not constitutive parts of the roof structure. Thus, transformations from function to form are generally not unique, and a straightforward functional determinism appears not only undesirable, but actually impractical.

The general problem that arises in this connection has been described in (March, 1976) as follows. In the classical syllogism, we have a law linking a case to an outcome and a case, and can then infer the outcome. For example,

\begin{align*}
\text{Law:} & \quad \text{All humans are mortal.} \\
\text{Case:} & \quad \text{Socrates is human.} \\
\text{Outcome:} & \quad \text{Socrates is mortal.}
\end{align*}

The inference that leads to an outcome, given a law and a case, is normally called \textit{deduction}. But within the framework of a syllogism, we can use two other forms of inference that lead to the third part of a syllogism if the two other parts are given: \textit{induction}, the process that infers a law given a case and an outcome, and \textit{abduction}, the process that infers a case given a (desired) outcome and a law. The term \textit{abduction} was coined by the philosopher Charles Pierce, and it is from him that March took the term and argued that the process of finding a design (a case) with desired properties (the outcome) that are specified up-front can be viewed as a form of abduction. Furthermore, the evaluation of a design for certain performance indices that relies on laws, such as the laws of physics, can be viewed as a form of deduction, and the process that establishes the laws to be used in this deduction from empirical observations can be viewed as a form of induction. The view of design as abduction is given a more extensive treatment in (Coyne et al., 1990).

Abduction is not straightforward if different cases can have the same outcome; that is, an outcome can be achieved in multiple ways. This introduces the problem of ambiguity to which we will return in the following discussion. The situation is aggravated in design, where the many performance indices that, taken together, indicate the desired outcome interact in intricate ways with the many design attributes that describe a design (cf. Mahdavi and Berberidou, 1992).

It is not surprising, then, that attempts to systematize and formalize architectural design in connection with computer-aided design underplay the abductive aspects of design and tend to favor some form of generate-and-test; that is, they assume that a design must be specified to a certain level of completeness before it can be evaluated. This approach solves some of the problems created by the complex interactions between design and performance variables that were described above: it is generally easier to evaluate a design deductively according to multiple criteria and to discover conflicts than to develop a design abductively directly from multiple and conflicting criteria.

In Stiny's elegant (but rather abstract) formalization of a generative-and-test approach (1981), two descriptions of an evolving design are constructed and maintained in parallel, one describing the form of the design (and constructed by a shape grammar) and one that extracts properties of the design and describes it “in terms pertaining to, for example, purpose, function, and use, meaning, type or form;” this second description can then form the basis for analysis and evaluation. The LOOS system implements this approach for a practically relevant problem domain (Flemming et al., 1988). The idea of maintaining two parallel descriptions, one in terms of design attributes and one in terms of performance indicators, will assume a specific form in the approach outlined below.

A more general treatment of form/function issues in architecture based on a generate-and-test approach can be found in Mitchell (1990). This and similar treatments envisage an iterative process that cycles through synthesis and analysis (including simulation)
Simultaneous Form Generation and Performance Evaluation

steps and explores the interactions between design and performance variables in the concrete context of the given design problem. That is, design is essentially viewed as search through a space of possibilities that are first generated and then evaluated.

But this is an ideal situation. In practice, the decoupling of generation from evaluation leads to the drawbacks mentioned in the introduction: evaluations may occur too late, or the more promising combinations of design attributes may not be generated in the first place.

A recent survey suggests that the methodological problems outlined above are also confronting the engineering disciplines (Flemming et al., 1991). Of particular interest in this connection is the experimental system described by (Tidd et al., 1992), which displays not only the geometry of a mechanical device, but also performance indices, either through scales, or through curves (graphic representation of functional relationships), where the design attributes (also called design parameters) and performance indices are interrelated in complex ways. The interesting feature is that designers can directly manipulate any of these variables and observe the implications for the other variables in real time. For example, they can change design attributes and observe changes in the values of performance indices; this is the traditional way. But designers can also change the value of a performance index and observe the resulting changes in the other indices and design attributes; this is what we call a two-way or bidirectional inference mechanism, which essentially allows for simultaneous form generation and performance evaluation.

This experiment is interesting because it suggests an inference mechanism that is complementary to the conventional one and able to counteract some of the drawbacks connected with it. Note, however, that it can only succeed when the relations between form and performance variables can be formalized in computationally efficient terms (for a truly interactive system, they must also be solvable in real time). In this case then, a bidirectional approach may become possible. An outline of this complementary approach in the context of building design is the main topic of this paper.

Mahdavi et al. (1991) present a related approach for the partially automated generation of artificial lighting configurations based on daylight availability. The possibility of reversing the conventional application of simulation capabilities is also explored in (Brown, 1991). A two-way approach in architectural daylighting simulation and the related concept of “fast-response” computational modules are introduced in (Mahdavi and Berberidou, 1992). Convergence strategies for two-way simulation environments are also outlined by Mahdavi (1993a) with focus on applications in the domain of architectural acoustics. Techniques of partial design optimization, which can become important for the two-way approach, have been discussed among others (Radford and Gero, 1980a, 1988). A preference-based “on-the-fly” optimization technique has been introduced (Mahdavi, 1993b) as a general computational strategy for the design of multidirectional simulation environments.

3 The Conventional and Two-Way Inference Approach

3.1 The Conventional Approach

Conventional analysis and simulation procedures are generally deductive mechanisms that transform pertinent design attributes (geometry, materials, environment) into performance indices. This transformation, shown in Figure 1, produces theoretically unique results. An example is the transformation of building, room and aperture geometry, surface reflectance values, and window transmittance into daylight distribution data under given sky conditions, as described in Table 1.
3.2 The Two-Way Inference Approach

The two-way inference approach intends to enable the designer to generate configurations of building components (geometric configurations, material constellations) and observe their interactions with desired performance indices in a “quasi real-time” fashion; it offers, in particular, the possibility to directly manipulate performance indices and to observe the corresponding changes in design attributes. It thus partially implements an abductive design process in March’s sense. It has a strong interactive component and does not intend to automate whole design sequences.

We suggest that this approach can significantly increase the effectiveness of computer-aided analysis and simulation tools by (1) reducing the number of parametric iterations for design variables a designer may need to explore; and (2) enhancing the designer’s understanding of and experience with the interdependencies and complex interactions between the numerous variables involved in a particular design context, especially in the early phases.

But the implementation of a comprehensive two-way inference approach is a highly complex task. This is particularly evident when many design variables with large degrees of freedom are involved. Moreover, the general issue of ambiguity associated with abduction, which will be discussed in the following section in more detail, introduces significant methodological and implementation problems. However, we suggest that despite

---

**Table 1. Daylight Factor Computation as a Transformation of Variables**

| Formal variables | e.g., room geometry (width, depth, height), aperture geometry (number and dimensions of openings) |
| “Semantic” variables | e.g., photometric properties (reflectances of walls, ceiling and floor, visible transmittance and reflectance of transparent building components) |
| Contextual variables | e.g., sky conditions, sun position (standard overcast, uniform skies, other isotropic skies, non-isotropic skies) |
| Performance Indicators | e.g., illuminance values and/or daylight factors on different reference surfaces |

---

Figure 1. Conventional simulation approach.
these limitations and problems, two-way inference techniques hold enough promise for a considerable number of well-defined activities in building design to warrant a closer study.

### 3.3 The Problem of Ambiguity

In the abductive design mode, ambiguity arises when the same performance criterion may be met by different configurations of design attributes. For example, higher daylight factors in a room can be achieved by increasing the window area, by increasing the visible transmittance of the window glazing, by increasing the surface reflectances of room interior, or by various combinations of these measures. Thus, the transformation of lighting performance criteria into concrete design configurations cannot be formalized deterministically. The ceiling configuration in the Kimball Art Museum described in Section 2 is another example.

Within the conventional approach, certain strategies have been developed for the rule-based generation of a permutative set of design configurations. This set can then be evaluated either directly by the designer or through complex performance-based weighting systems. However, similar strategies would not be appropriate for the two-way inference approach as envisioned above. The time-consuming generation of numerous potential solutions and a posteriori evaluation do not facilitate the quasi real-time (dynamic) observation of the interdependencies between performance indices and design attributes.

Due to the complex interaction patterns involved, we cannot envision a single successful transformation strategy. Rather, we suggest using a set of flexible and locally effective strategies, which are modular and multifaceted, and thus able to reduce and/or efficiently re-group the set of design attributes that have to change in a particular situation in order to realize desired performance indices (Mahdavi and Berberidou, 1992). In this context, two concepts are particularly important:

- **The lock concept:** the interactive selection of those design attributes that are allowed to respond to successive changes in performance indices, while other design attributes remain unchanged.
- **The priority concept:** the interactive definition of a set of priorities that imply a different degree of freedom for each design attribute. From an operational point of view, a number of strategies can be implemented to facilitate the selective response to modifications of performance indices, whereby heuristic rules, mathematical multi-criteria optimization techniques, and regression-analytical methods play an important role.

### 3.4 An Example

These ideas are demonstrated in the following example. Figure 2a shows the schematic cross-section of a room with a single window. The curve in this figure illustrates the typical distribution of the daylight factors along the section line on a horizontal reference plane. This curve demonstrates the basic problem of side-lit rooms, namely the rather large illuminance gradient between areas close to the window and the deeper portion of the room. One method to achieve a more uniform illuminance distribution is the proper use of light reflectors as shown in Figure 2b.

The design of effective light reflectors is a rather complex task. Aside from the traditional approach based on physical modeling, complex one-way simulation techniques (e.g., ray-tracing) have to be used based on complete design information. As a result, extensive iterations are needed to approach adequate dimensions, shapes, and locations for light redirection devices. The two-way inference approach, in contrast, would enable the designer to manipulate the daylight distribution curve directly and to observe immediately changes in the design attributes of the light reflector.
In order to do this, however, the relationship between daylight distribution and certain attributes of a light redirection device (such as its length and its vertical location relative to the window) must first be established. This relationship is rather complex due to the twofold effect of light redirection devices: (1) obstruction, which reduces the illuminance levels in the area close to the window, and (2) light redirection towards the ceiling, which contributes to higher illuminance values in the deeper portion of the room. In addition, light reflectors may have different shapes and reflectances and may be translucent.

This complex set of relationships must be formalized either through regression-analytical procedures or complex mathematical modelling. The functional relationships can then be used to facilitate a real-time transformation of performance values into changes in design attributes and vice versa. An example of regression-analytical techniques used for the formalization of such functional relationships is discussed in (Mahdavi and Berberidou, 1992). This formalization, along with the strategies discussed above (selective locking, multi-criteria optimization), create the computational “intelligence” behind a two-way inference system.

Figure 2. Schematic cross-section of a side-lit room without light reflector (a) and with light reflector (b) along with the corresponding curves of light distribution along the section line on a horizontal reference plane (after Mahdavi and Berberidou, 1992).
4  A First System Prototype

We outline in this section a system prototype to demonstrate the capabilities and concepts introduced in the last two sections. The prototype is comprised of the following major components (see Figure 3):

• an effective interface that allows the designer to manipulate directly both the design attributes (geometry, spatial relations, material attributes, etc.) and performance values (lighting, energy, thermal comfort analysis, etc.) of a building component

• a two-way inference module that is able to (a) predict changes in the performance values in response to changes in the design attributes; and (2) suggest changes in design attributes as a result of changes in the performance values

• a database of building components and associated inference procedures able to support the two-way feedback envisioned.

Figure 3. The components of a first prototype for a two-way inference system.

The “cases” the system knows about and that can be inferred abductively are restricted to those contained in the database; that is, the system works under what is known in the literature as the closed-world assumption (Coyne et al., 1990:282). Designers must be aware of this if they are to interpret correctly the results obtained by working with the system.

Figures 4a-d show the user-interface for the proposed prototype. As Figure 4a illustrates, the general user interface of the system is part of an overall CAD environment. The user can select specific parts of the design under consideration for interactive modification and evaluation. Figure 4b shows the general organization of the screen, which establishes a general framework for simultaneous form generation and performance evaluation. Separate windows display geometrical attributes (in this case, plan and section), component attributes, scalar and graphic representations of the current state of various performance indicators (in this case, daylighting and energy), as well as pointers to component databases. As Figures 4b and 4c illustrate, the user can install/modify components (in this case, a light-shelf) and observe their performance in real-time (e.g., in terms of the vertical light distribution profile). At the same time, changes in performance values can dynamically transform design attributes. Figure 4d demonstrates this for a change in room proportions that results from a modification of the two-point daylight factor.
Figure 4a. General user-interface of the system as integral part of a CAD environment. The user can select specific parts of a design.

Figure 4b. Interface for simultaneous form generation and performance evaluation. Separate windows display geometrical attributes, component attributes, and various performance indicators (daylighting, energy).
Figure 4c. Representation of component effects on daylighting performance. User can install/modify components and observe performance. Alternatively, changes in performance indicators are translated into design modifications.

Figure 4d. Example of a formal transformation (change in room proportions) as the result of modification of the two-point daylight factor.
5 Discussion

We started this paper with a brief review of functionalism and the arguments made against it because we want to dispel any notion that we advocate a revival of a strict functional determinism. The system that we describe should make the methodological and technical difficulties that arise from such a position abundantly clear. It is not meant to make the inference from function to form automatic. Rather, it intends to make available to designers as much knowledge as possible about the complex interactions between form and function when the components under consideration are known, and to do this in a form that is as direct and useful in the design process as possible.

In no case is the system meant to deny designers their role in finding an overall form and expression for a building that is uniquely appropriate for the given context. In the ideal case, the system would assist designers in finding out concretely the circumstances they want to express and give form to in the first place. This brings us back to Sullivan. In the quoted article, he did not stop with establishing the three basic parts of a skyscraper, but went on to show that it is precisely when we accept this division, especially the undifferentiated middle section, that unique expressive opportunities arise:

“It [the tall building] must be tall, every inch of it tall. The force and power of altitude must be in it, the glory and pride of exaltation must be in it. It must be every inch a proud and soaring thing, rising in sheer exultation that from bottom to top it is a unit without a single dissenting line–that it is the new, the unexpected, the eloquent peroration of most bald, most sinister, most forbidding conditions.

And thus the design of the tall office building takes its place with all other architectural types made when architecture, as has happened once in many years, was a living art. Witness the Greek temple, the Gothic cathedral, the medieval fortress.”

(Sullivan, 1988:108-109, o.e.)

Acknowledgments

The authors gratefully acknowledge the contributions of Paul Mathew, Suresh Bhavnani, and Stephen Lee to the design and implementation of the first interface prototype for the architectural two-way feedback system illustrated in Figures 4a-d.

References


Flemming, U., Adams, J., Carlson, C., Coyne, R., Fenves, S., Finger, S., Ganeshan, R., Garrett, J., Gupta, A., Reich, Y., Siewiorek, D., Sturges, R., Thomas, D., and


