DESIGN SHOULD BE MANAGED AS A PHYSICAL PROCESS, TOO

Glenn Ballard¹ and Lauri Koskela²
(1) University of California, Berkeley, USA, (2) Salford University, Manchester, U.K.

ABSTRACT
The mainstream stand in textbooks on design and design management is that design is a problem solving process, starting from the perceived problem and ending with a detailed solution. We contend that these accounts overlook one necessary and important conceptualization of design, namely design as a physical flow process. This view holds design as a spatio-temporal process, where information (on whatever media) is traversing through a network of designers and other stakeholders. Among the characteristics of a design process, its duration, cost and often output quality can only be explained through this view. Accordingly, it is important to manage design as a physical process, based on the unique features of this view. However, due to a relative neglect of this view of design we have the situation that practical prescriptions and approaches to design management contain, at most, partial or fragmentary methods and tools towards management of the physical process side. Fortunately, the theoretical and practical development of this concept carried out in the framework of production management can be advantageously used also for the design context. However, because of the fundamental differences between material production and design, the concepts emanating from production have to be adapted for being valid in design. Thus, for example, while in material production time reduction is the primary goal for management, in design, besides design time reduction, the elimination of making-do in design tasks must be taken as an equally significant goal. A framework for conceptualizing management of design as a physical process has been presented, and practical development and trialing of a number of related methods in the context of building design accounted. Future work is needed for developing a seamless set of methods for design management, incorporating the concept of design as a physical process, besides other requisite concepts.

Keywords: design, physical process, flow process, design structure matrix, Last Planner, pull-push

INTRODUCTION
The mainstream stand in textbooks on design and design management is that design is a problem solving process, starting from the perceived problem and ending to a detailed solution. We contend that these accounts arguably overlook one necessary and important conceptualization of design, namely design as a physical process. This view holds design as a spatio-temporal process, where information (on whatever media) is traversing through a network of designers and other stakeholders. Among the characteristics of a design process, its duration, cost and often output quality can only be explained through this view. Accordingly, it is important to manage design as a physical process, based on the unique features of this view. However, due to a relative neglect of the view of design as a physical process we have the situation that practical prescriptions and approaches to design management contain, at most, partial or fragmentary methods and tools towards management of the physical process side.

We endeavor to present an overview on the present theoretical and practical status of managing design as a physical process, and in so doing, to pinpoint the related research challenges. The paper is based on theoretical and applied research by the authors in the last fifteen years. This work has mostly been done in the context of building design.

The paper is structured as follows. The alternative concepts of design are first reviewed and a comparison is made to operations management. Then, the insights accumulated in production management regarding operations as physical flow processes are interpreted from a design viewpoint.
Next, a framework for managing design, also as a physical flow process, is presented. To conclude, the findings are briefly discussed and future research challenges outlined.

ALTERNATIVE CONCEPTS OF DESIGN

Concepts of design in the design literature
Many textbooks on design, e.g., Pahl et al. [1] and Cross [2], present a prescriptive process model of design, understood as problem solving. This model, usually progressing from abstract and general to concrete and specific, contains different stages and activities seen as necessary in the complete cycle of design, as well as their interactions. The differing characteristics of the various stages, for example named as “analysis” and “synthesis”, are emphasized. The underlying conceptualization sees design is a problem solving process, starting from the perceived problem and ending to a detailed solution. This is reflected in the definition of design by Dym [3]: Engineering design is the systematic, intelligent generation and evaluation of specifications for artifacts whose form and function achieve stated objectives and satisfy specified constraints.

However, these general models are not very helpful in practical situations of design management. For this, Pahl et al. [1], for example, suggest moving on to planning of activities, as well as timing and scheduling of these activities. However, this means that the traditional approach of project management is followed. The underlying conceptualization of project management is that of a transformation (Koskela & Howell [4]): the design process, as a total transformation, is decomposed into more manageable pieces, and these are organized into a plan, assigned, executed and controlled.

However, there is a third conceptualization of design, but this view has been acknowledged only rarely, and in other than the mainstream literature on design. This approach sees design as a spatio-temporal physical process, related to the movement of information. Augustin and Ruffer [5] suggest using logistic thinking in the analysis of product development. Adachi et al. [6] suggest conceiving concurrent engineering as the application of JIT ideas to design. In their book on design improvement, Sekine and Arai [7] focus on what happens to information in design: “things are made through the flow of information”. The unit of analysis is the total flow of information. Reinertsen [8] develops an approach on design management based on queueing theory.

Parallel to these proposals, there are accounts of the difficulty to acknowledge iteration in design, a fundamental feature in the view of design as a physical process. For example, Friedrich et al. [9] strongly criticize the customary notion that large projects can be measured using yardsticks viewed as simple summations of individual yardsticks taken discipline by discipline, system by system, or component by component. They argue that the overall effects of revisions, repairs, and rework on large projects can be very significant, even when the individual effects of specific functions and disciplines appear small and within “normal” acceptable practices. Somewhat later, Cooper [10] gives an explanation to this phenomenon, coining it “rework cycle”, and claiming that it has been customarily mismanaged.

On the other hand, there have been a number of initiatives to partially include the view of design as a physical process into the design arena. The Design Structure Matrix (DSM) method (Eppinger et al. [11]) allows the representation of information flows between design tasks, and makes it possible to order the design tasks in such a way that the number of cases where a task has to send feedback to an earlier task is minimized. Thus it is possible to minimize the waste due to unnecessary iterations. In terms of concurrent engineering, such tools as collocation and integrated teams have been advanced, which influence the information flow through physical layout or social engineering. Also, there are Systems dynamics approaches for accounting for the iteration and rework in design (for example, Lee et al. [12]). However, generally it can be said that the design specific approaches towards managing design as a physical flow are partial and poorly grounded in theory.

Comparison to production management
It is contended that the existence of three different foundational concepts of design is not surprising. In prior work (Koskela [13]), it has been shown that production management has been based on three
complementary concepts of production, termed transformation, flow and value generation. Although slightly differing terminology is used here, these are the same concepts as found in design (flow corresponds to the physical process, and value generation to problem solving).

Thus there are three major concepts of production, and each of them has produced practical methods, tools and production templates. It has been argued that these three concepts of production are not alternative, competing theories of production, but rather partial and complementary (Koskela [13]). What is needed is a production theory and related tools that fully integrate the transformation, flow, and value concepts. As a first step towards this, we can conceptualize production simultaneously from these three points of view: transformation, flow and value- the associated concepts and principles have been called the TFV theory of production.

APPLYING THE GENERIC PRINCIPLES OF PRODUCTION AS PHYSICAL PROCESSES TO DESIGN MANAGEMENT

The evolution of the concept of production as a physical spatio-temporal process, flow, is fascinating. The idea goes back to Ford [14], who understood time as one production factor among others: “The time element in manufacturing stretches from the moment the raw material is separated from the earth to the moment when the finished product is delivered to the ultimate consumer.” And “Time waste differs from material waste in that there can be no salvage. The easiest of all wastes, and the hardest to correct, is this waste of time, because wasted time does not litter the floor like wasted material.” Ford's production concept was subject to wide interest and publicity. However, it seems that the flow aspects of Ford's mass production were generally misunderstood or simply not conceived. The template of mass production that started to diffuse contained primarily those features of Ford's mass production model that belonged to the domain of the transformation concept of production. Nevertheless, as accounted by Wada [15], somehow the idea of "flow production" kept alive in Japan and became a focus of interest in the late 1930s. The breakthrough in the implementation of flow production ideas happened in the Toyota Company, and the resultant production template became known as the Toyota production system. The next breakthrough happened through the establishment of queuing theory as a basis for the flow view.

Next, the major descriptive and prescriptive insights deriving from this conceptualization of production are commented regarding their interpretation from a design management point of view.

Waste

The first observation through which this production concept can be practically implemented was that time is consumed by two types of activities when viewed from the point of view of the product: transformation activities and others, apparently non-transformation activities, called waste by Ford and categorized by the Gilbreths [16] as transfer, delay and inspection activities. Obviously, these non-transformation activities are unnecessary from the point of view of the transformation. So, the less of them the better; best if there are none of them. As Shingo indicates [17], the approaches to improving these two types of activities are totally different: making the one more efficient; trying to eliminate the other.

The predominant wastes in design are somewhat different in comparison to physical production. As noted by Cooper [10], unnecessary iteration is one typical design waste. As it is easy to carry out a design task based on assumptions, the waste of making-do is prevalent. In making-do, an activity is started before all its inputs are at hand (Koskela [18]). The waste caused by batching is equally common in making and in design – in the latter, it is often caused by stage-gating.

Queueing theory

A second important insight is that it is possible to model the behavior of production as a physical process using appropriate models. As Hopp and Spearman [19] have convincingly shown in their book "Factory Physics", queueing theory provides the basic, physical model of production. These
authors show that by means of queueing theory, various insights, which have been used as heuristics in the framework of the Toyota Production System, can be mathematically proven. In totality, 15 laws on the behavior of production flow lines are presented. One important result of a queueing theory analysis is that variability early in the line is more disruptive than variability late in the line. However, maybe the most fundamental result regarding production control is that in view of a certain level of variability in production, there is always a penalty in one form or another, even if the control is the best possible. One has to select among three alternatives [19]:

1. Buffering of flows (for increasing the probability that all parts are available at a workstation when needed), which leads to long cycle times and high work-in-progress levels.
2. Accepting lower utilization levels of resources, which equates to acquisition of extra capacity.
3. Accepting lost throughput (due to starvation of workstations).

In practice, the most common way of accommodating variability is through buffering. Thus, the original goal of the Toyota Production System to decrease inventories enforces the reduction of variability. Because of the relation of inventories to the cycle time, as defined by Little’s Law [19], this can be also interpreted as follows: reduction of the cycle time leads to reduction of variability.

However, Factory Physics has been developed based on repetitive production, and it turns out that extensions are needed for other production situations. Thus, for example, in construction type of production, there is another way of accommodating variability: making-do (as mentioned above), not discussed by Hopp and Spearman. In the framework of lean construction, a production control method geared towards eliminating making-do has been developed, namely Last Planner (Ballard [20]).

The applicability of the Factory Physics framework in design has not been much explored. Probably the majority of its conclusions hold regarding repetitive, mature design processes. For more emergent design processes, where the contents or order of tasks is not predictable (Levardy & Browning [21]), obviously other conceptualizations are needed. However, intuitively some of the results, like the greater disruption caused by variability early in the line, seem applicable to all design processes.

**Improvement**

Factory Physics explains a third, earlier insight made at Toyota, namely the importance of eliminating, through learning, such variation from production, which gets transformed into variability in the queueing theory sense. For this, the suggestion of Shewhart [22] was put into practice, through standardization of tasks, visual management and continuous improvement:

“It may be helpful to think of the three steps in the mass production process as steps in scientific method. In this sense, specification, production, and inspection correspond respectively to making a hypothesis, carrying out an experiment, and testing the hypothesis. The three steps constitute a dynamic scientific process of acquiring knowledge. […] Mass production viewed in this way constitutes a continuing and self-corrective method for making the most efficient use of raw and fabricated materials.”

From the viewpoint of design, prototyping can be interpreted to be based on the same idea. Also otherwise, there are no obvious obstacles of introducing standardization, visual management and continuous improvement into design.

**Conclusions from design viewpoint**

As justified by queueing theory, the reduction of cycle time and making-do can be used as the drivers for improving design. To achieve these goals, such means as elimination of unnecessary steps, rework and iteration, and ultimately, variability reduction, can be utilized. Reduced lead time is intrinsically beneficial; reduced making-do is beneficial through improved design outcome. Due to the multitude of interdependencies in design, the elimination of unnecessary iteration is among the weighty means towards the main goals for design, to be realized through an optimal ordering of the tasks. However, design process may be emergent, rather than pre-determined; this may prevent predicting the contents
of design, but not necessarily in the short or medium term. Also, it may be worthwhile to increase the number of interdependencies, for the sake of improved problem-solving.

**FRAMEWORK OF MANAGING DESIGN AS A PHYSICAL PROCESS**

Current thinking and practice regarding design system management is now described, using the general functions of production systems management, each with its characteristic managerial principles: (1) design system design (2) design system operation, and (3) design system improvement. Design system operation can further be decomposed into three functions, each with its characteristic managerial principles: planning, control and correction. The physical flow process nature of design, as explored above, is especially emphasized.

**Design system design**

The design of any system is in large part a matter of deciding how the work to be done by the system is to be structured; i.e., how the work will be divided into pieces, allocated to various specialists, and ultimately assembled into valued goods and services (Ballard [23]). Corresponding to this structure of work, the organization is structured and equipped to do the work. There are a multitude of issues to decide when we design the design system. We decide, for example, on the physical layout of designers, IT and communication support, data standards, design representation (drawings or Building Information Modeling), contractual relations, incentives, decision making structure, validation and verification structure, targeted social system structures and whether non-conventional solutions, such as set based design or target costing, are adopted. All these will have an impact on the physical flow of design information – however, it is not possible to treat these issues in detail here.

Target costing is of special interest in the design of design systems because it involves self-imposing constraints (dependencies) in order to create value: thus a more complex physical process is deliberately chosen for the sake of value generation. As such, it may be a step toward integration of the transformation, flow and value conceptualizations of production understood as designing and making things. Target costing was developed as a method of managing product profitability by Japanese manufacturers of consumer and industrial products (Cooper & Slagmulder [24], [25]). It has subsequently been widely applied in manufacturing throughout the world, and has been adapted for use in the construction industry (Ballard & Reiser [26], Ballard [27], [28]). The basic concept in target costing is to specify the customer’s allowable cost by subtracting target profit (more generally, expected benefits from use of the facilities to be produced) from forecast revenues. It is one application of the general principle to self-impose necessity as a means to spur improvement (for Toyota’s application of this principle, see Liker [29]).

One advocate of target costing in construction is Sutter Health [28]. The largest healthcare company in Northern California, Sutter Health first applied target costing on design and construction of an Acute Rehabilitation Center project in 2005-6, and succeeded in reversing an upward trend in costs and cost overruns on their capital projects. The second application was on their Fairfield Medical Office Building, which was completed 18% below market. In early 2009, Sutter Health had three major hospital projects in design, all using target costing. Sutter Health is also in the vanguard in the area of contractual relations, having introduced its Integrated Form of Agreement, intended as an application of the theory of relational contracting (Lichtig [30]).

**Design system operation**

---

1 A number of other healthcare systems have implemented target costing with similar results, including SSM in St. Louis, and Thedacare, a healthcare company headquartered in Appleton, Wisconsin. Thedacare completed its Shawano Clinic project 18% below a market-based benchmark and 3.5 months ahead of schedule [28].
As previously noted, design system operation is divided into planning, controlling and correcting; each of which is treated in turn in the following.

**Planning**

Planning is here understood as structuring and sequencing design tasks. This is done at different levels of detail. For example, a design task might be expressed as “Design the building structure”. That task can be broken down into phases such as conceptual design, design development, and detailed engineering. In turn, each phase can be broken down into a sequence of generic tasks:
- determine the design criteria for the building structure
- generate alternatives
- evaluate alternatives against criteria
- select from alternatives

The generic tasks within each phase can be expressed in terms specific to the product being designed, but prior to doing so it is critical to determine the interdependence of work in each design discipline, say structural design with the work of other design disciplines such as architecture, mechanical engineering and electrical engineering.

Different types of work have different characteristics. One of the differences between designing and making involves the type of dependencies between tasks. Tasks dedicated to fabricating and assembling products are not reciprocally interdependent. Tasks dedicated to designing products may be reciprocally interdependent; for example, structural, heat and energy loads. Determining the load placed on a structure is a function of the weight, location and footprint of mechanical and electrical equipment. The specification of electrical equipment is a function of the energy required. The specification of mechanical equipment is a function of heat load, to which the equipment itself contributes. Mechanical and electrical equipment add to energy requirements and to heat load. Finally, the structure itself impacts heat loads and power requirements. The determination of loads cannot be done in a linear sequence, but rather requires a type of conversation and design iteration between the relevant specialists in which they search for the combination of values that satisfies known requirements with the best approximation of desired product characteristics. The design structure matrix shown in Figure 1 has been used to reveal such interdependencies and to restructure and resquence tasks to reduce interdependencies.

![Design Structure Matrix](image)

*Figure 1: Design Structure Matrix (Austin, et al. [31])*

If the design task is recurrent and relatively stable, a plan can be made based on the optimal order of tasks, as suggested by the DSM method. However, often the design process is emergent, in the sense of earlier design decisions influencing the interdependencies of later tasks. In such cases, it is only
possible to create an optimal order for the near term chunk of design. In addition, as only the designers themselves well understand the interdependencies, it is often a good strategy to create a plan through negotiation between the designers, and improve it through DSM.

How best develop the sequence of tasks that is then to be analyzed and improved using DSM? According to some theorists, a principle of planning is that those who are to do the work should plan how to do it (Ballard and Howell [32]). This principle is based on the fact that the various design disciplines build on each others’ work, and hence that one specialist must understand the conditions of satisfaction of another when providing a drawing, model, calculation, range of values, or other work product.

Collaborative planning may occur in a variety of forms. One approach used in design of buildings, bridges, and such has been called “pull scheduling” [32]. A team of design specialists creates a logic network, working backwards from a target milestone such as ‘Select structural system’. Working backwards is done to develop a network in which the completion of each task releases successor tasks; a network with minimum float. The conversation between specialists is facilitated by asking each person what they need from others in order to perform their task. This results in a set of sticky notes on a wall, arranged in sequence, with connecting lines indicating dependencies, as shown in Figure 2.

![Figure 2: Collaborative Planning](image)

Knots of reciprocally interdependent tasks are assigned to the relevant specialists. Participants are asked to leave their task durations unpadded; to use average durations. Once a network is completed, duration of the entire network can be calculated. The objective is to develop a network with a shorter duration than the time allowed for the phase of work being planned. That surplus time is treated as schedule contingency to be allocated to buffering individual tasks that are both critical and variable. If there is additional time remaining after this allocation, it can be reserved in a phase buffer to be used as needed, or the target duration for the phase can be reduced.

In conclusion, collaborative planning produces a work plan that is tested and improved through DSM analysis. This can be done at any level of detail, but must be done in order to provide the goals that production control seeks to achieve; namely, the handoffs between specialists expressed in the logic network.

---

2 Photograph courtesy of Alan Mossman, Director, The Change Business Ltd.
Buffer management is another aspect of design system planning. In planning, decisions are made regarding what type of buffers to use, where best to locate them, and how best to size them. Consider, for example, the approach advocated by Goldratt [33]; namely, to include in a schedule of tasks a time buffer that can be drawn on like a bank account when tasks are not completed when scheduled. The rule of thumb proposed is that such time buffers be half as large as the schedule they are intended to buffer from variation in task duration. The objective is to assure completion of the network of tasks on or before the scheduled end date. One consequence is that the start dates of all tasks on the critical path slip each time ‘money’ is withdrawn from the buffer, introducing variation in the timing of work load for the various specialists with tasks to perform. In fact, this is a disguised use of capacity rather than time as the buffer. To the extent that specialists cannot find alternative work on which to apply their capacity, that capacity is lost. A labor hour not productively employed cannot be recovered.

After generating schedule contingency (float), an alternative approach to Goldratt’s is to have the team of specialists collaboratively agree how to allocate that contingency to individual tasks, based on task criticality and the probability and extent of variation in duration [32]. This approach yields more reliable start dates and thus does not require sacrificing capacity in order to achieve schedule objectives.

**Control**

Control is often understood, especially in project production systems such as construction and product development, as after-the-fact detection and correction of variances from target outcomes. That is the explicit conceptualization in critical path method approaches to planning and control. Variances between ‘should’ and ‘did’ are signals for management action; an application of the theory of management by exception. In our terminology, that rather belongs to Correction, to be discussed below. Control is here understood as steering toward targets, which include scope, quality, schedule and cost. Work flow control is vital for all these targets because design work occurs through a network of interdependent specialists. Having a well structured plan, whether produced using DSM or other means, does not assure plan execution.

Preparation for performing scheduled tasks includes identification and removal of constraints such as securing needed information from others, acquiring resources, and designing and testing work methods.

Release of work from one specialist (individual or group) to another is best done through pull mechanisms; in other words, in response to a signal from the immediate customer. ‘What work should I do next?’ This question is answered by referring to the logic network to identify immediate customers, then asking those customers if they are ready for my input. The commitments that emerge are documented and shared among the team members, then reviewed periodically to identify plan failures, commitments not kept, to understand the underlying causes, and to mobilize learning to prevent reoccurrence in the future3.

Where control is practiced as described here, the percentage of planned tasks completed increases substantially. An early study found on engineering and construction projects found that routinely 45% of planned tasks are not completed as planned, and hence cannot support the project logic network. Implementation of elementary control functions such as collaborative planning, constraints analysis, and securing commitments has improved that metric to 65-70%, with corresponding improvements in design quality, cost and schedule [20].

**Correction**

---

3 For a more complete description of work flow control using the Last Planner method, see Ballard, et al., 2002 Ballard, 1999 and Ballard, 2000.
Correction is the immediate response to detection of a breakdown; an unintended deviation from target outcomes. For example, if an error is detected in a drawing, the process that produced the drawing is stopped until the error has been corrected. Correcting the flawed drawing is necessary but not sufficient. That drawing may have been used as a basis for other design work, so it is necessary to understand the ripple effects of the detected error and identify and correct errors in connected work products.

Other breakdowns are handled in the same way. For example, the design program may have incorrectly specified an air conditioning system. Detection of this error triggers an investigation to identify and correct calculations and inferences drawn from that specification.

Correction is cleaning up the mess. Adjusting to the consequences of breakdown may involve changes in near term goals or plans for goal achievement in order to achieve project objectives. Prevention of reoccurrences of breakdowns belongs to design system improvement.

Design system improvement

Improvement of the design output may be assumed to be more important than improvement of the design system. However, Sverlinger [34] found that half of the disturbances in design are due to design organizations themselves, rather than the design process in question. A similar order of magnitude for this factor was found by Koskela, Ballard and Tanhuanpää [35]. Further, even small system improvements can have large impact on performance over time.

How are production systems (and specifically design systems) methodically improved? Opportunities for improvement may be revealed by breakdowns; unintentional deviations from target outcomes. In design, typical breakdowns include failure to complete scheduled tasks on time, errors in calculation or representation, failure to integrate component designs, and nonconformance of a selected option with ‘must have’ criteria. Methodical improvement occurs from analysis of breakdowns to actionable causes, followed by practical experiments to test the effectiveness of proposed changes. An ‘experiment’ in this sense is an intended deviation from process. These practical experiments can be understood as applications of Shewhart’s plan-do-check-act cycle [22], which requires specification of a basis for comparison; i.e., process standardization. If we don’t understand how the breakdown occurred, it is not possible to prevent reoccurrence.

A ‘hypothesis’ to test may emerge from intuition and imagination, without an actual breakdown. Consequently, we can say that improvement in a design system is learning from breakdowns and from experiments. Methodical improvement requires a culture that encourages experimentation and that sees breakdowns as indications that the existing understanding of causality is flawed.

CONCLUSIONS

Arguments and evidence have been forwarded towards the view that the concept of design as a physical, spatio-temporal process has been relatively neglected in both design theory and design practice. Fortunately, the theoretical and practical development of this concept carried out in the framework of production management can be advantageously used also for the design context. However, because of the fundamental differences between material production and design, the concepts emanating from production have to be adapted for being valid in design. Thus, while in material production time reduction is the primary goal for management, in design, besides design time reduction, the elimination of making-do in design tasks must be taken as an equally significant goal.

A framework for conceptualizing management of design as a physical process has been presented, and practical development and trialing of a number of related methods in the context of building design accounted. Future work is needed for developing a seamless set of methods for design management, incorporating the concept of design as a physical process, besides other requisite concepts. Issues for future research comprise, for example:
• implications of queuing theory applied to design management;
• extension of managing flows to generate value beyond target costing in product development and construction;
• application of visual management principles to design management, to provide total project visibility to all stakeholders
• the extent to which computer modeling now or soon will enable agile prototyping to be applied in the design of all types of products, including those of the built environment;
• learning to differentiate between value-adding and non-value-adding complexity and variety in design;
• reducing non-value-adding variation and matching buffers thereto;
• and the use of contracts and commercial terms to align the interests of specialists with project objectives.

REFERENCES
