

## MIT Open Access Articles

*Category theory based solution for the building block replacement problem in materials design*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Giesa, Tristan, David I. Spivak, and Markus J. Buehler. "Category Theory Based Solution for the Building Block Replacement Problem in Materials Design." *Advanced Engineering Materials* 14.9 (2012): 810–817.

**As Published:** <http://dx.doi.org/10.1002/adem.201200109>

**Publisher:** Wiley Blackwell

**Persistent URL:** <http://hdl.handle.net/1721.1/77560>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike 3.0



# Category theory based solution for the building block replacement problem in materials design

Tristan Giesa<sup>1</sup>, David I. Spivak<sup>2</sup>, and Markus J. Buehler<sup>1,3\*</sup>

<sup>1</sup> Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Mass. Ave. Room 1-235A&B, Cambridge, MA, 02139, USA

<sup>2</sup> Department of Mathematics, Massachusetts Institute of Technology, 77 Mass. Ave, Cambridge, MA, 02139, USA

<sup>3</sup> Center for Computational Engineering, Massachusetts Institute of Technology, 77 Mass. Ave. Room 1-235A&B, Cambridge, MA, 02139, USA

\* Corresponding author, E-mail: [mbuehler@MIT.EDU](mailto:mbuehler@MIT.EDU)

**ABSTRACT:** An important objective in materials design is to develop a systematic methodology for replacing unavailable or expensive material building blocks by simpler and abundant ones, while maintaining or improving the functionality of the material. The mathematical field of category theory provides a formal specification language, which lies at the heart of such a methodology. In this paper, we apply material ologs, category-theoretic descriptions of hierarchical materials, to rigorously define a process by which material building blocks can be replaced by others while maintaining large-scale properties, to the extent possible. We demonstrate the implementation of this approach by using category-theoretic techniques to predict concrete conditions needed for building block replacement. As a perspicuous example, we specify structure-function relationships in two systems: a laminated composite and a structure-function analogue, a fruit salad. In both systems we demonstrate how ologs provides us with a mathematical tool that allows us to replace one building block with others to achieve approximately the same functionality.

**Abbreviated Title:** Category theory and materials design

**Keywords:** Materials design, category theory, building blocks, theory, modeling, mathematics, hierarchical materials

**Submitted to:** *Advanced Engineering Materials*

## 1. Introduction

Thousands of labs across the world have been studying various biological and engineered synthetic materials from a molecular perspective. This research provides a bottom-up approach to materials which can be used to predict properties from a fundamental molecular-level point of view,<sup>[1, 2]</sup> establishing the field of materiomics.<sup>[3, 4]</sup> Typically, for each material, one aims to identify how material function (*e.g.*, strength and toughness) emerges based on the interplay of basic material building blocks such as molecules, mineral particles, or chemical bonds. New research in the mathematical field of *category theory* provides a precise framework to capture such relationships in an abstract space, as demonstrated in a recent paper.<sup>[5]</sup> When building a formal model of a material system using category theory, the result is a so-called material olog, which looks like a graph or network equipped with additional information. The olog precisely expresses how structure and functionality can be composed, combined, and interconnected throughout the system.

One of the most urgent questions for material scientists is how one can design high performance materials by use of abundant and cheap material constituents,<sup>[6]</sup> similar to natural designs found in biological materials.<sup>[7]</sup> In more abstract terms, the key problem is to determine the conditions necessary for replacing one or more of a system's basic building blocks by other building blocks, while maintaining its overall functionality. This is referred to as the "*building block replacement problem*". For example, if we consider a polymer with covalent crosslinks, it is of interest whether the covalent cross-links can be replaced by other molecular structures, such as weak van der Waals or hydrogen bonds, without compromising the material's properties. What types of structures, based on the new building blocks, need to be introduced to achieve this? A simple example from linguistics illustrates this problem. Consider the sentence:

I drive my car from home to work. [A]

The sentence contains eight basic building blocks, the words used to construct it: "I", "drive", "my", "car", "from", "home", "to", and "work". Suppose that one of the building blocks, say "home", becomes unavailable. There is a way to solve the building block replacement problem in this case; *i.e.* to express the same meaning without using the unavailable word. We could, for example say:

I drive my car from [*the place where I sleep and meet my family*] to work. [B]

It is clear that “the place where I sleep and meet my family” describes “home”, and hence, we have expressed the same meaning as before. To do this we used different building blocks, resembling the utilization of a hierarchical structure. Obviously there are many solutions to this problem, and one important question is if the same meaning of sentence [A] can be expressed with a particular set of available building blocks. In the case of [B] we have silently assumed that we have the availability of a set of new building blocks: “place”, “sleep”, “meet”, and “family”.

An approach based on category theory can help us in the design process by creating a mathematically rigorous formulation of the problem. Using this approach grants access to a range of techniques for finding solutions to the building block replacement problem under varied constraints, such as specific types and amounts of available building blocks.

In this paper, we introduce the use of material ologs for solving the building block replacement problem by way of two elementary materials: a laminated composite and a fruit salad. These are chosen because they exhibit a functorially isomorphic structure and function, while being simple enough to be easily understood. In each case we relate the structural makeup (collection of layers / fruits) to its function (five-dimensional mechanical property / taste vector). We then show how to “solve the olog” in each case. For example, we explain how to maintain the mechanical properties of the laminate (respectively the taste of the fruit salad) even in the absence of certain materials (resp. fruits). The analogy and impact for materials science is clear: there is much to be gained by replacing energetically-expensive material building blocks by other ones, which are possibly simpler and more abundant, while maintaining the functionality of the whole system.

## **2. Methods**

Ologs are based in the mathematical field of category theory, and use that formalism to capture strict relationships between various objects of study.<sup>[5, 8, 9]</sup> For example, material



structure and function can both sit inside a single olog where the complex interactions between them are laid bare, in accordance with the principles of materiomics. Ologs can be easily translated into database schemas which can store research findings.<sup>[10]</sup>

Database theory, in particular the ability to query datasets, can be applied to any olog that has been loaded with data. In typical applications, database queries are commands like “return the set of persons  $x$  where the first name of  $x$  is Mary and the salary of  $x$  is greater than 100”. In the context of material ologs, queries cannot be so simply and colloquially interpreted. For example, one might query “return the set of ductile systems of bricks and glue for which the brick element is carbon nanotubes”. To understand precisely what is meant by ‘ductile systems’ in this context, one would have to look carefully at the olog in which they are defined (for example a ‘ductile system’ has been formally defined in an olog<sup>[9]</sup> as a one-dimensional system whose failure extension is much greater than the failure extension of its glue element, a term that is also defined within that olog). This is a significant advantage of the olog formalism: notions such as ‘ductility’ can be defined in a context-specific way, so that researchers can precisely interpret the jargon used by other labs and correctly interpret their findings. If conventions are established for such notions in each context (axial loading for beta-sheets, etc.), large data sets can be compiled from multiple labs, and the possibility for solving the building block replacement problem is enhanced. Additional methods needed to solve these types of problems are introduced in Section 3.

The mathematical formalism of ologs also allows for the specification of so-called aggregation functions, which extract meaningful invariants from large amounts of data. These functions include mathematical operations that can range from counting or averaging, to complex numerical procedures such as solving partial differential equations.

Category theory enables the specification of formal analogies between different systems. Two systems, described by ologs, are related by functors of varying “qualities”. For example, a functorial isomorphism is the strongest relation that may be obtained. A functor  $F: C \rightarrow D$  is the mapping of boxes and arrows in olog  $C$  to the boxes and arrows in olog  $D$ . Combining  $F$  with a functor  $G: D \rightarrow C$ , such that the “round-trip” compositions map each box and arrow to

itself, is called an isomorphism. In this case, no category-theoretic information is lost under the transformation.

### **3. Results**

In the present section we discuss the process of “solving an olog” for the case of a fruit salad and a laminated composite (**Figure 1**), which both include features of a hierarchical system. Structurally, each laminated composite (respectively, each fruit salad) is composed of a collection of layers (respectively, fruits). We can represent each in a hierarchical olog.<sup>[8]</sup>

To tackle the building block replacement problem, one needs to be able to abstract from the details of the system to a higher level of representation and replace the specific building blocks by more general objects. In our example, the fruit salad could be abstracted to a mixture consisting of various substances, such as a gravel mixture in concrete.

The most ubiquitous sort of numerical implementation of “solving the olog” is given by database query. Here (as elsewhere<sup>[5]</sup>), an olog is considered as a database schema, which in turn is simply a finitely presented category  $C$ . Hierarchical categories<sup>[8]</sup> allow for aggregation functions, as described in the Methods section. Call the olog ‘MaterialOlog’, and imagine that ‘MaterialOlog’ is filled with data already, call it ‘Dat’. (The notion of recording data in an olog is presented in <sup>[5]</sup>; it is a functor  $C \rightarrow Set$ .) This olog has boxes (**Figure 1**), labeled  $A, A1, B, C, \dots, H$ , each of which refers to a table in the database. For example,  $Dat(A)$  refers to the table of layered composites (resp. fruit salads) we have observed and documented. This table has a column for the material name, a column for its constituent building blocks, and a column for its resulting property-vector.

Suppose we are given the following building block replacement problem: ‘given a material system  $X$ , which happens to contain an unavailable building block  $B$ , display all *known* systems that have the same functionality as  $X$  (or are within some tolerance interval of  $X$ ), but do not include  $B$ .’ In other words, we hope to replace building block  $B$  by a collection of other building blocks while maintaining functionality.

This problem can be tackled numerically by applying a category-theoretic database query against the MaterialOlog/Dat database. For experts in the language of category theory and

its application to databases<sup>[10]</sup>, we pull the data back *via* a functor  $\mathbf{F}$  to a subschema consisting of only two objects: material and property vector, see as visualized in **Figure 2a**. We then perform a regressive update, using the data migration functors  $\mathbf{F}^*$  and  $\mathbf{F}_*$ , to remove building block  $B$  from the material set while keeping the property vector constant (or within a specified tolerance). Finally we perform a right-pushforward of the result and base change with the unit of the right-pushforward adjunction applied to the original dataset. In other words, our query is the fiber product ‘Query-Result’ in **Figure 2b**.

In general, a database query can solve building block replacement problems of the following two sorts:

- I. Fix functionality and then query the data store on the olog to determine the possible building blocks and the structural setups that achieve the desired functionality; or
- II. Fix one or more building blocks and explore the design space for the others in terms of structure and functionality.

In both cases, we have access only to what is known or stored in the olog dataset; in general this information must be obtained by physical experiments or simulation.

Apart from a data query, a key interest in materials science involves using mathematical methods to predict conditions under which hypothetical building blocks can be exchanged for existing ones while maintaining or even improving system functionality. In the next sections, we will discuss this by means of representative calculations.

### 3.1 Fruit salad

A fruit salad is a material that exhibits the structure and function of a general composite material, and hence exemplifies material ologs. Each fruit is a particular instance of a fruit species (arrow 1). Functionally, each fruit species has a specific (average) taste vector (arrow 3) and each fruit salad has a total taste (sum of the individual fruit tastes, arrow 2). A taste is a five-dimensional vector of real numbers, where the entries reflect a measure for the five taste sensations: umami (savory, meat-like taste), sweetness, sourness, bitterness, and saltiness.

In the case of the fruit salad, we imagine a scenario where we begin with a fruit salad  $FS1$ , whose taste we want to emulate with a new fruit salad  $FS2$ . The constraint is that  $FS2$  cannot contain certain building blocks (fruits) and must contain others. More precisely, suppose we are given a fruit salad  $FS1$  composed of, say, the fruit set (fruit1, fruit2, fruit3, fruit4) and having a total taste vector (arrow 2) called  $T(FS1)$ . “Solving the olog” means that we must find a fruit set  $FS3$  to add to some given fruit set  $FS2$ , say  $FS2 = (\text{fruit4}, \text{fruit5})$ , such that the taste will be the same as that of the original fruit salad  $FS1$ , i.e.  $T(FS1) = T(FS2) + T(FS3)$ . However, neither  $FS2$  nor  $FS3$  is allowed to use building block fruit1.

The solution in this specific case is obtained by solving a system of equations. Above,  $T$  denotes a linear transformation from the real vector space with basis  $S$ , the set of fruits, to the taste vector space  $\mathbf{R}^5$ . Considering a concrete example, we assign to each fruit species a taste vector with entries from 0 to 10 indicating the intensity of this taste with respect to other fruit species. Specifically,  $FS1$ , as part of the data base, could consist of a bananas with taste vector  $Ban = (3; 9; 1; 3; 1)$ , apples with taste vector  $App = (2; 7; 7; 2; 1)$ , papaya with  $Pap = (6; 6; 2; 3; 0)$  and lemons with  $Lem = (0; 3; 9; 5; 1)$ . The total taste is a linear combination of the ingredients with  $T(FS1) = a \cdot Ban + b \cdot App + c \cdot Pap + d \cdot Lem$ , where e.g.,  $a = 0.2, b = 0.6, c = 0.1, d = 0.1$ , such that  $T(FS1) = (2.4; 6.9; 5.5; 2.6; 0.9)$ . Similarly,  $FS2$  consists of papaya with  $Pap = (6, 6, 2, 3, 0)$  and tomatoes with  $Tom = (8; 2; 5; 3; 1)$  where  $T(FS2) = e \cdot Pap + f \cdot Tom$  and  $e = 0.5, f = 0.5$ , such that  $T(FS2) = (7.0; 4.0; 3.5; 3.0; 0.5)$ . We need to solve for  $FS3$  in the equation  $T(FS3) = T(FS1) - T(FS2) = (-4.6; 2.9; 2.0; -0.4; 0.4)$ . If, for instance, tomatoes become available and bananas become unavailable, we solve for the linear combination of fruit ingredients in  $FS3$ . The best fit (least-square) solution is over-determined for less than five taste-independent fruits and under-determined for more than five. In many cases, matrix inversion *via* QR decomposition (e.g. with a Householder reflection) predicts a set of solutions, the taste conditions, for this problem. In our setup, the equation system is given by

$$(1) \quad \begin{bmatrix} 2 & 0 \\ 7 & 3 \\ 7 & 9 \\ 2 & 5 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} -4.6 \\ 2.9 \\ 2.0 \\ -0.4 \\ 0.4 \end{pmatrix}.$$

7

The numerical solution yields  $\alpha = 0.1419$  and  $\beta = 0.1001$  (representing *e.g.* a volume percentage in the mixture with respect to the original volume) and a residuum of 5.28. Depending on the context, such a calculation may require additional constraints, for example that all coefficients must be positive.

### 3.2 Laminated composites

Similar to the olog of the fruit salad, we can solve the building block replacement problem in the case of layered composites. A functorial isomorphism (uniquely determined up to permutations of the entries in the property vector) converts all arrows and objects in the fruit salad olog to those in the laminated composite olog (**Figure 1**). This isomorphism corresponds to a high degree of similarity in the calculation for solving the building block replacement problem in the two systems.

The only difference in the calculations lies in the aggregation function represented by arrow 2. The mathematical model – based on laminate theory – that connects the properties of the whole laminate to the properties of the single layer is more complex. Indeed, while the fruit salad mathematical model represented by arrow 2 is a simple weighted summation of the ingredients, the junction of material laminates requires a procedure *e.g.* based on elasticity theory (this is not a limitation but merely a consequence of the particular physics that governs the problem at hand, and which must be incorporated into the description in the olog).<sup>[11]</sup> Here, we consider only the in-plane part of the so-called ABD-matrix to simplify the calculation and enable the straightforward connection to the fruit salad system.

As in Section 3.1, we now analyze a symmetric stack of four material layers  $MS1$  composed of (layer1, layer2, layer3, layer4)<sub>s</sub> with distinct mechanical properties  $T(MS1)$  and a stack of two material layers  $MS2$  composed of (layer4, layer5) with distinct mechanical properties  $T(MS2)$ . Again we wish to determine what material layers  $MS3$  we can add to the material layers  $MS2$ , say  $MS2 = (\text{layer4}, \text{layer5})_s$ , so that the mechanical properties will be the same as that of the material layer stack  $MS1$ , *i.e.*  $T(MS1) = H(T(MS2), T(MS3))$ . We require that neither  $MS2$  nor  $MS3$  is allowed to use building block layer1 or layer3.

Again,  $T$  denotes a linear transformation from the real vector space with basis  $\mathcal{S}$ , the set of laminates, to the in-plane mechanical properties space for a symmetrical laminate  $\mathbf{R}^5$ , uniquely given by  $E_{xx}, E_{yy}, G_{xy}, \nu_{xy}, \nu_{yx}$ . We assign to each laminate a mechanical properties vector (already homogenized). Specifically,  $MS1$ , as part of the data base, could consist of eight equally thick laminates with total thickness  $h = 1.6$  mm that are stacked symmetrically. All data is stored in appropriate units, *i.e.* MPa: laminate1 with  $lam1 = (130,000; 9,000; 500; 0.3; 0.02)$ , laminate2 with  $lam2 = (120,000; 10,000; 700; 0.4; 0.03)$ , laminate3 with  $lam3 = (60,000; 60,000; 200; 0.3; 0.3)$  and laminate4 with  $lam4 = (100,000; 3,000; 900; 0.15; 0.05)$ . The elastic behavior of the composite is determined by its laminate stiffness matrix (ABD-matrix) dependent on its ply thicknesses  $h_k = |z_{k+1} - z_k| = h/8 = 0.2$  mm with  $z_k$  being the respective coordinate. Then  $A_{ij} = \sum_{k=1}^8 \bar{Q}_{ij,k} (z_{k+1} - z_k)$ ,  $B_{ij} = 1/2 \sum_{k=1}^8 \bar{Q}_{ij,k} (z_{k+1}^2 - z_k^2)$  and  $D_{ij} = 1/3 \sum_{k=1}^8 \bar{Q}_{ij,k} (z_{k+1}^3 - z_k^3)$ . For the in-plane composite properties we consider only the A-matrix such that  $\bar{Q}_{ij,k} = A_{ij,k}/h_k$  as the effective single layer properties and  $E_{xx} = (A_{11} - A_{22}^2/A_{12})/h$ ,  $E_{yy} = (A_{22} - A_{11}^2/A_{12})/h$ ,  $G_{xy} = A_{33}/h$ ,  $\nu_{xy} = A_{12}/A_{22}$ . Consequently  $T(MS1) = H([A]_{ij}, h) = H(H^{-1}(lam_k, h_k), h)$  where  $H$  is the linear transformations described above. Similarly, the symmetric composite  $MS2$  consists of four equally thick laminates: laminate4 with  $lam4 = (100,000; 3,000; 900; 0.15; 0.05)$  and laminate5 with  $lam5 = (90,000; 90,000; 900; 0.3; 0.3)$ . We need to solve for the distribution of thicknesses in  $MS3$  consisting of building blocks from  $MS1$  and  $MS2$  that yield an equal behavior to  $MS1$  under the constraint that only laminate2, laminate4, and laminate5 are available. Similar to the fruit salad, we can decouple the transformation of  $MS1$  and  $MS2$  due to the linearity of  $A_{ij}$  with respect to  $z_k$ . The over-determined system of equations can again be solved by Householder reflection

$$(2) \quad 2 \begin{bmatrix} \bar{Q}_{11,2} & \bar{Q}_{11,4} & \bar{Q}_{11,5} \\ \bar{Q}_{22,2} & \bar{Q}_{22,4} & \bar{Q}_{22,5} \\ \bar{Q}_{12,2} & \bar{Q}_{12,4} & \bar{Q}_{12,5} \\ \bar{Q}_{33,2} & \bar{Q}_{33,4} & \bar{Q}_{33,5} \end{bmatrix} \begin{pmatrix} h_2 \\ h_4 \\ h_5 \end{pmatrix} = \begin{pmatrix} A_{11}^{comp1} \\ A_{22}^{comp1} \\ A_{12}^{comp1} \\ A_{13}^{comp1} \end{pmatrix} \Leftrightarrow \begin{bmatrix} 24.3 & 2.01 & 19.8 \\ 2.03 & 0.60 & 19.8 \\ 0.81 & 0.09 & 5.93 \\ 0.24 & 0.14 & 0.04 \end{bmatrix} \begin{pmatrix} h_2 \\ h_4 \\ h_5 \end{pmatrix} = \begin{pmatrix} 13.1 \\ 3.53 \\ 1.08 \\ 0.11 \end{pmatrix}.$$

The numerical solution yields  $h_2/h = 0.517$ ,  $h_4/h = 0.325$  and  $h_5/h = 0.158$  (representing the thickness fraction of each ply) and a residuum of 467. Depending on the context, such a calculation may require additional constraints, for example that a certain total thickness is

required. In case the determination of out-of plane properties is desired, a non-linear equation system has to be solved (*e.g.* fixed-point iteration).

This demonstrates the surprising abstraction power of ologs: ostensibly unlike systems trace back to a very similar category-theoretic representation, which guides the design with context-specific mathematical tools, in these simple examples QR decomposition and Householder reflection (**Figure 2c**). All systems having the same olog can be solved by similar techniques, despite large differences in their physical makeup. This provides a powerful avenue for example for bioinspired materials design where the key concepts from a biological system must be transferred to a synthetic one with likely different building blocks or other constraints.

### **3.3 Predictions arising from ologs**

In case the building block properties are not fixed, the conditions to maintain or improve the system's functionality can be determined under various constraints. The prediction might result in material properties that are currently unavailable for engineer, but are achievable through genetic engineering or other synthetic manufacturing processes.

Considering the laminated composite example presented in section 3.2, the mechanical properties of the engineered layer can be determined exactly by the equation  $\bar{Q}_{ij,*} =$

$(A_{ij}^{comp1}/2 - \bar{Q}_{ij,4}h_4 - \bar{Q}_{ij,5}h_5)/h_*$  under the constraint that all layers have equal thickness.

The solution of this equation yields the mechanical property vector of the new material phase  $lam_* = (64,146; 31,452; 125; 0.287; 0.141)$ . Such a composite laminate could be engineered through ply rotation of a standard laminate.

## **4. Discussion**

After studying a diverse set of materials one realizes that many of the analyzed systems follow similar patterns by which material function is realized, such as hierarchical structures (*e.g.*, self-similar or dissimilar architecture), size or confinement effects (*e.g.*, smaller is stronger), or the formation of composite structures with certain patterns (*e.g.* high-aspect ratio platelets or inclusions in a matrix material).<sup>[12]</sup> While such similarities between silk, diatoms, and bone can be observed and discussed informally amongst researchers, there

have been few, if any, tools to express these similarities mathematically, and hence limited capacity to perform a systematic study. The situation is reminiscent of linguistic theory in the early twentieth century, when patterns in disparate languages were observed. Noam Chomsky fundamentally transformed the field by finding a mathematical description of the syntax (structure) and semantics (function) of human languages.<sup>[13]</sup> Like linguistics, materiomics may greatly benefit from a rigorous mathematical description of the relationship between structure and function in materials.

Such a formal description and repository for discovered similarities may have profound impacts on the field of materiomics. Indeed, the ability to replace functional groups within a molecule or a hierarchical material and then understand its resulting functionality depends on an intimate knowledge of the design space and access to large data resources (as in the Protein Data Bank). As research develops, the design space and data resources grow with it, often in unforeseen ways, making the need for a mathematical theory all the more pressing. The category-theoretic formalism provides the necessary structure to encode the structural makeup of complex materials and the corresponding piece-by-piece functionality of their subunits. It also scales with progress in the field, allowing the research community to process new developments almost seamlessly, applying functorial connections from old categories to new ones.

The techniques discussed above constitute a new approach to modularity for materials engineering using a systems approach. In a category-theoretic framework, the structure-function relationship in a system can be recorded, and the necessary conditions for system assemblability and functionality are elucidated. In case of building block exchange we may then come up with solution steps to deduce the conditions to fit in a new block, at least up to the frontier of our scientific understanding. Ologs are a tool to compile information about how materials function in a mathematically rigorous way, which allows for comparative analyses between different materials. Ologs can guide us in the design of technologies that resemble natural processes. Capturing how function emerges in different systems is of extreme value for engineers, who have not had this type of tool available. Ologs provide such a cross-disciplinary methodology for capturing function and its breakdown at the building block level.



The use of ologs also allows us to identify and extract powerful mechanisms from nature for adoption in engineering. One particularly intriguing aspect of natural materials is that their functions do not rely on the superiority of building blocks. Rather, function emerges because of hierarchical structures, in spite of the weakness, and perhaps because of the simplicity of the constituent building blocks.<sup>[14]</sup> The beta-sheet protein structure is the building block used in spider silk, but may not be the building block of choice in engineering. For example, an engineer may instead want to use a carbon nanotube, a clay mineral, or some other naturally occurring structure – one that is inexpensive or abundantly available. What “solving the olog” means in practice is that we can learn how to make a material that behaves like silk (*e.g.*, extreme strength, toughness, deformability, etc.) from synthetic building blocks like carbon nanotubes.

In this paper we solved relatively simple ologs for fruit salads and for composites of material layers. In a categorically similar manner, it is possible to solve the more complex protein material ologs, such as those in recent literature<sup>[9]</sup>. Suppose we have a hierarchical protein material that includes alpha-helices at the nano-scale, and we would like to replace them with carbon nano-tubes (CNTs) while maintaining the ductility, at least within some specified tolerance. With the olog in place, a database query or complex aggregation functions can be applied. For example, in order to replace alpha-helices with CNTs, the olog expresses the need for a much weaker glue element that can connect pairs of CNTs, together with a strong lifeline with hidden length (see **Figure 3**). All this is precisely described by the olog, which is simultaneously computer-readable, as a precise mathematical construct, and human-readable, as a natural-language presentation of concepts. Challenges may arise if the given olog is too simplistic or if there is insufficient data in the store. If this is the case, the olog can be extended or additional data collected, and functorial data migration applied to move all the old data to the new olog.<sup>[10]</sup>

Here, we introduced a set of strategies for solving certain material ologs. To transition from simple systems such as mixes and layers, to more complex materials, *e.g.* involving interaction terms or hierarchical structures, would require more resources. Of particular necessity is more data. A single lab, even a well-subsidized one, simply cannot produce the

amount of structured data necessary to make the kind of discoveries we believe are possible. Sharing data resources is a key step toward increasing efficiency and productivity.

In 2009 Fields medalist Timothy Gowers proposed<sup>[15]</sup> that a massive research collaboration effort could be employed to solve long-standing mathematical questions, for example to find a new combinatorial proof of the Hales-Jewett theorem. This initiative resulted in several successful articles in a very short amount of time.<sup>[16-18]</sup> Similarly, recent projects such as SETI@home, distributed.net, and Fold-it, have employed “crowd sourcing” to harness the labor of thousands of individual people and computers, eventually resulting in solutions that may not have been possible with conventional methods.<sup>[19, 20]</sup> A similar program could be undertaken by the materials science community to synergistically tackle complex problems such as the above introduced “building block replacement problem”. Ologs, or some similarly expressive modeling formalism, provide a necessary tool to adequately present and share findings across multiple research groups.

It is clear that several of the impacts listed above (*e.g.*, the development of high functioning, sustainable, and environmentally benign materials) are at the heart of civil and environmental engineering. Collaborative research efforts based on a foundation of formal mathematics may pave the way for such large scale engineering projects.

## **5. Conclusion**

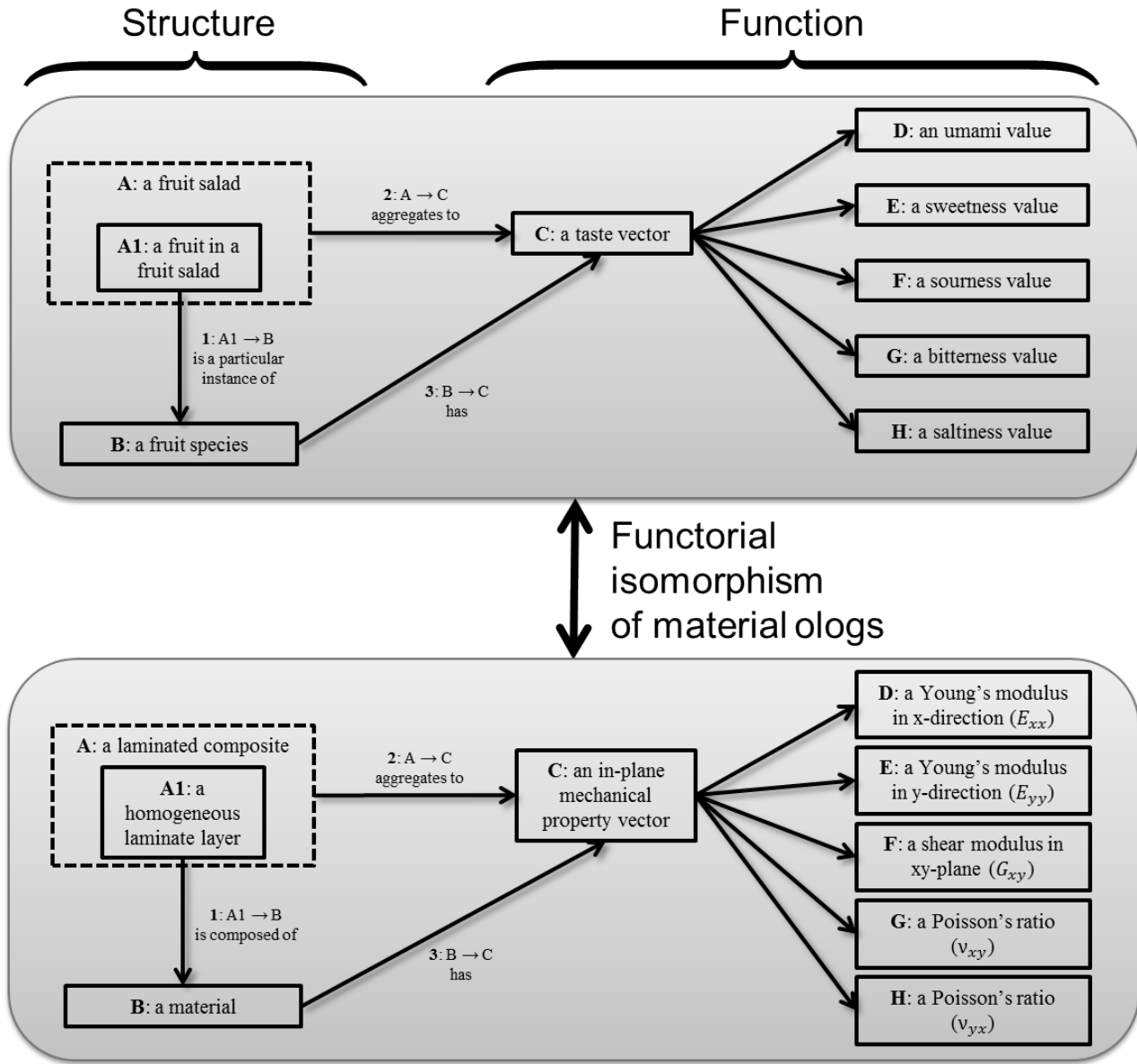
The present study provides a framework for the solution to the building block replacement problem in materials science through a novel methodology called “olog solving”, which is concretely exemplified in the case of a fruit salad and a laminated composite. Category theory based ologs combined with database theory and context-based numerical tools are shown to be beneficial toward the design of *de novo* materials. Furthermore, the use of such tools may become enhanced through the use of data from massively collaborative research projects. The result could be a database to find functionally optimized material solutions including nanotechnology, self-assembled or genetically engineered protein, and other biomaterials, to meet specific engineering requirements, similar to classical macroscale engineering catalogues.

**Acknowledgements:** Support from AFOSR and ONR is acknowledged.

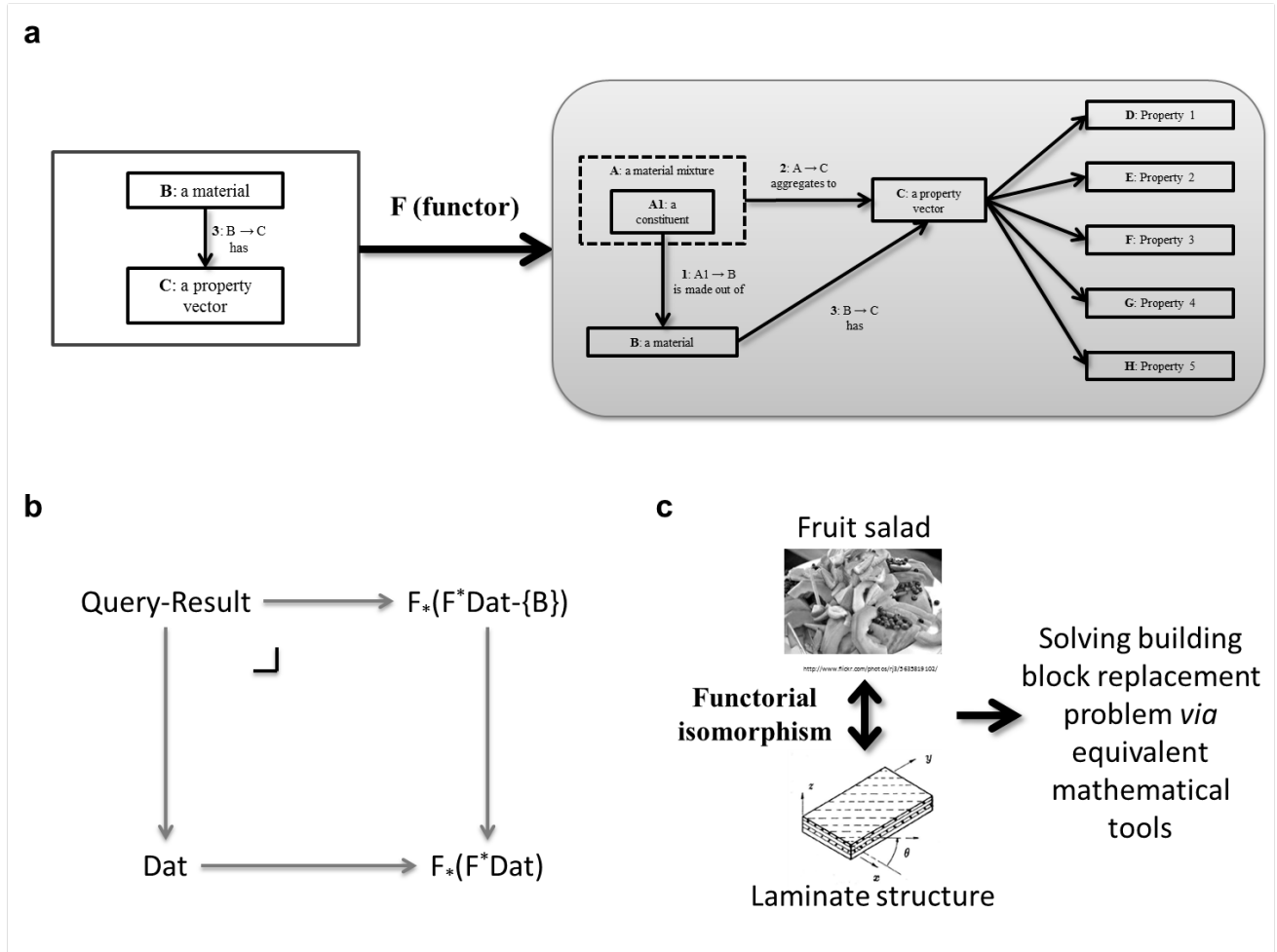
## **References**

- [1] A. Pandey and M. Mann, *Nature*, **2000**,405.
- [2] S. G. Zhang, *Nat Biotechnol*, **2003**,21.
- [3] S. W. Cranford and M. J. Buehler, *Nanotechnology, Science and Applications*, **2010**,3.
- [4] S. Cranford and M. J. Buehler, in: *Biomateriomics*, Materials Science Springer, New York, **2012**.
- [5] D. I. Spivak and R. E. Kent, *PLoS ONE*, **2012**,7.
- [6] P. T. Anastas and M. M. Kirchhoff, *Accounts Chem Res*, **2002**,35.
- [7] N. Li, J. Lv and D. K. Niu, *J Mol Evol*, **2009**,68.
- [8] T. Giesa, D. I. Spivak and M. J. Buehler, *BioNanoScience*, **2011**,1.
- [9] D. I. Spivak, T. Giesa, E. Wood and M. J. Buehler, *PLoS ONE*, **2011**,6.
- [10] D. I. Spivak, *Accepted for publication in Information and Computation*, **2012**,Also available at: <http://arxiv.org/abs/1009.1166>.
- [11] O. O. Ochoa and J. N. Reddy, in: *Finite element analysis of composite laminates*, Solid mechanics and its applications, Kluwer Academic Publishers, Dordrecht ; Boston, **1992**.
- [12] M. J. Buehler, *Nat Nanotechnol*, **2010**,5.
- [13] N. Chomsky, in: *Syntactic structures*, Mouton de Gruyter, Berlin ; New York, **2002**.
- [14] M. J. Buehler, *Nano Today*, **2010**,5.
- [15] T. Gowers, Is massively collaborative mathematics possible?, in: *Gowers's Weblog*, Cambridge, 2009.
- [16] D. H. J. Polymath, *ePrint arXiv:0910.3926*, **2009**.
- [17] D. H. J. Polymath, *ePrint arXiv:1002.0374*, **2010**.
- [18] D. H. J. Polymath, *ePrint arXiv:1009.3956*, **2010**.
- [19] M. Shirts and V. S. Pande, *Science*, **2000**,290.
- [20] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popovic and F. Players, *Nature*, **2010**,466.

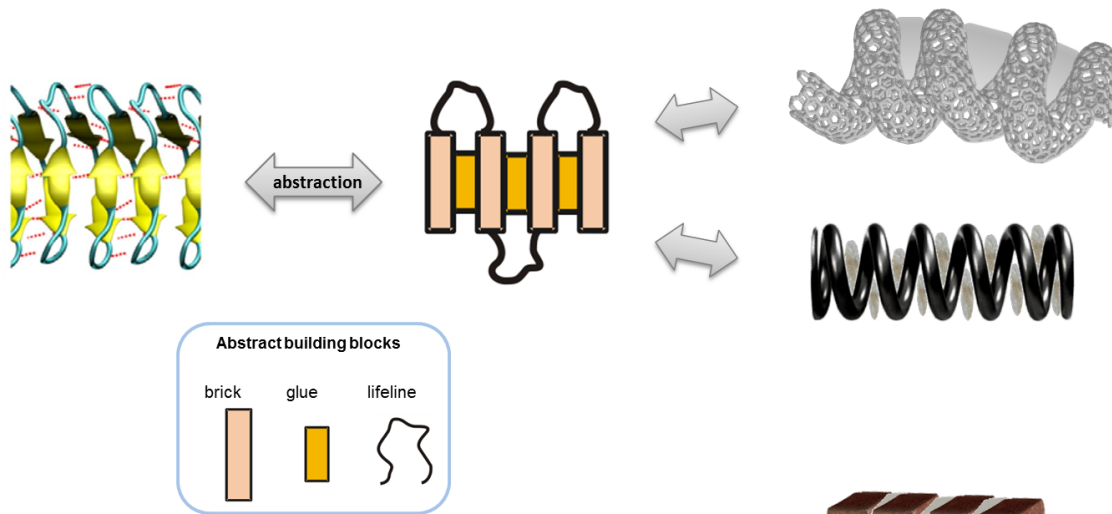
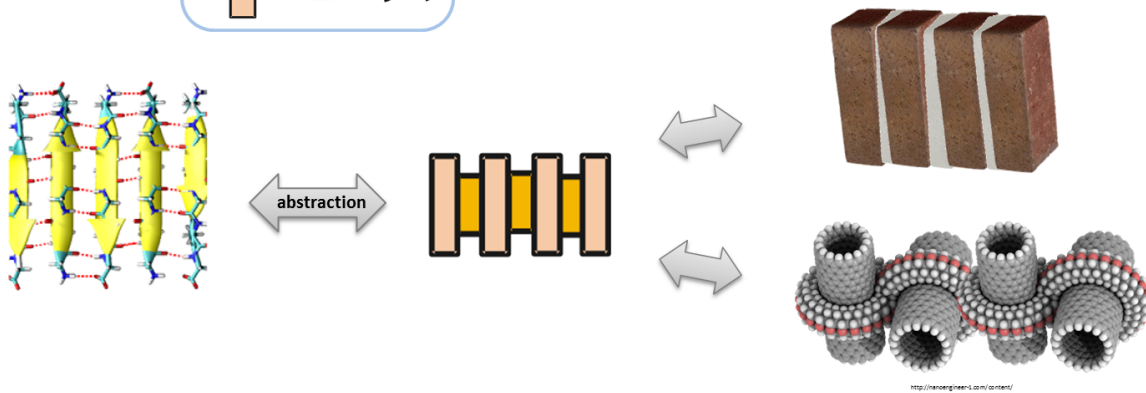
## Figures and captions



**Figure 1 | Material ologs used to describe the structure and function relationship for two functorially isomorphic materials, fruit salads and laminated composites.** Structurally, each material mixture is composed of a collection of constituent building blocks (fruits and laminate layers). Each constituent is made out of a particular material (arrow 1). Functionally, each material has a property vector (arrow 3) and each material mixture aggregates to a property vector (arrow 2). The property vector is a five dimensional vector of real numbers. The two paths A1 $\rightarrow$ C, one given by arrow 1 followed by arrow 3 and the other given by the implicit arrow A1 $\rightarrow$ A followed by arrow 2, are equivalent.



**Figure 2 | Querying a database to solve a building block replacement problem. (a)** In this building block replacement problem the property of a known material mixture must be mimicked by substituting an unavailable constituent  $B$  with available ones. The olog containing all the relevant structure and functionality is compared by a functor  $F$  to an olog consisting only of the structure and function to remain constant. **(b)** Given a state called  $Dat$  on the material olog, the building block replacement problem can be tackled by applying a category-theoretic database query against the database. This is achieved by removing building block  $B$  from the set of materials, keeping the property vector constant, and applying the fiber product displayed here, where  $F^*$  and  $F_*$  are the pullback and right-pushforward data migration functors for  $F$ . **(c)** We demonstrate the surprising abstraction power of ologs: ostensibly unlike systems trace back to functorially isomorphic category-theoretic representations. All systems having the same olog can be solved by similar techniques, despite large differences in their physical makeup. Fruit photo by Ben Alman, laminate image from Ochoa and Reddy.<sup>[11]</sup>

**a****b**

**Figure 3 | The building replacement problem as an abstraction process. (a)** A  $\alpha$ -helix is abstracted to a brick-glue-lifeline system as in [9]. Possible replacements include a helical carbon nanotube and an ordinary spring. An adequate olog can predict the necessary conditions for the glue, under which this replacement yields a functional behavior similar to the original system. **(b)** An amyloid fibril is abstracted in a similar way as panel (a). In this case, possible replacements include either clay bricks with mortar or carbon nanotubes with crimp junctions (image source: Nanorex Inc.).