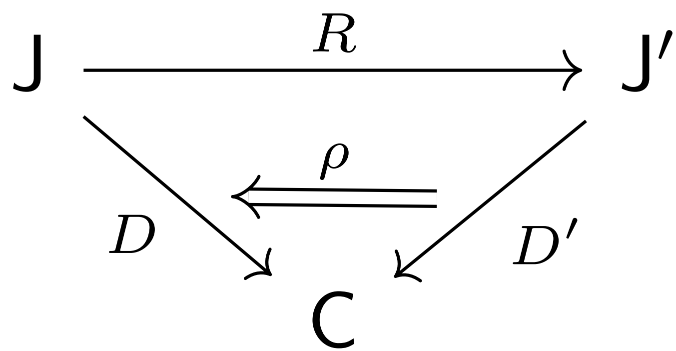
Categories of diagrams
in data migration and computational physics

Evan Patterson

Topos Institute Colloquium
December 2, 2021

Part 1: Diagrams and their morphisms



Diagrams

Diagrams are among the most fundamental notions of category theory.

Recall. A *diagram* in a category C is a functor $D\colon J \to C$, where J is a small category.

Two common classes of diagrams:

- *Free diagram*: diagram shape J is freely generated by a graph
- *Commutative diagram*: free diagram where images of any two paths with same source/target are equal in C

Example. A *span* in C

$$x \xleftarrow{f} w \xrightarrow{g} y$$

is a free diagram $D\colon J \to C$ of shape $J := \{1 \leftarrow 0 \to 2\}$ where $D(0) = w$ and

$$D(1) = x, \quad D(2) = y$$
$$D(0 \to 1) = f, \quad D(0 \to 2) = g.$$

Categories of diagrams: an incomplete history

Everyone knows and loves diagrams, but it is less appreciated that diagrams in C have a natural notion of **morphism** and so form a category (even a 2-category).

Two recent papers:

- Peschke & Tholen: "Diagrams, fibrations, and the decomposition of colimits" [PT20]
- Perrone & Tholen: "Kan extensions are partial colimits" [PT21]

Considerable work in the 70s by René Guitart, mainly in French [Gui73, Gui74, GVdB77].

But goes back to the earliest work on category theory:

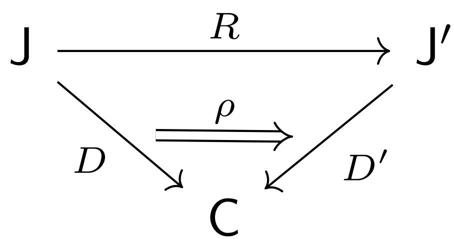
- Kock, PhD thesis: *Limit monads in categories* [Koc67]
- Eilenberg & Mac Lane: "General theory of natural equivalences" [EM45]

Categories of diagrams: definitions

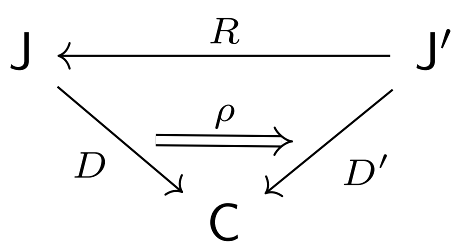
There are several notions of morphism of diagrams, hence several diagram categories.

Definition. The category $\mathsf{Diag}(C)$ has

- as objects, diagrams in C
- as morphisms from $D\colon J \to C$ to $D'\colon J' \to C$, a functor $R\colon J \to J'$ together with a natural transformation $\rho\colon J \Rightarrow J' \circ R$.

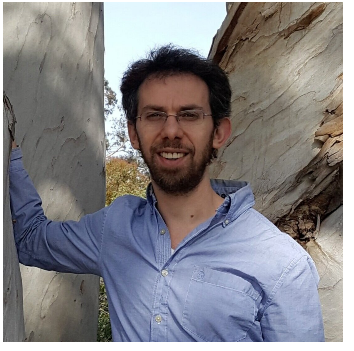


Similarly, the category $\mathsf{Diag}^{\mathrm{op}}(C)$ has the same objects and but the morphisms:



Part 2: Diagrams in functorial data migration

In collaboration with David Spivak



Categorical databases

The point of departure for the categorical databases story is that relational databases can be elegantly modeled by the basic concepts of category theory:

- database schema is a small category C, usually finitely presented
- database instance is a *copresheaf* on C, or C-*set*, namely a functor $X\colon \mathsf{C} \to \mathsf{Set}$
- homomorphism of databases X, Y is a natural transformation $X \Rightarrow Y$

Example. (Schema for graphs)

```
@present SchGraph(FreeSchema) begin
  V::Ob
  E::Ob
  src::Hom(E, V)
  tgt::Hom(E, V)
end
```

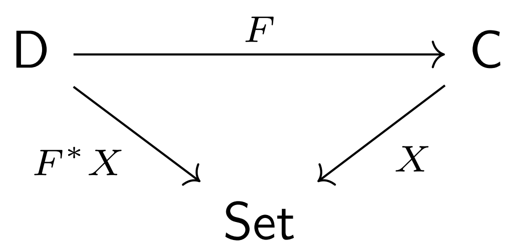
$$E \underset{\mathrm{tgt}}{\overset{\mathrm{src}}{\rightrightarrows}} V$$

Functorial data migration

The categorical viewpoint suggests the idea of functorial data migration [Spi12]:

> Functors between schemas induce functors between databases.

The simplest form of data migration is *pullback data migration*: given functor $F\colon D \to C$, precomposition with F defines a functor $\Delta_F := F^*\colon C\text{-Set} \to D\text{-Set}$.

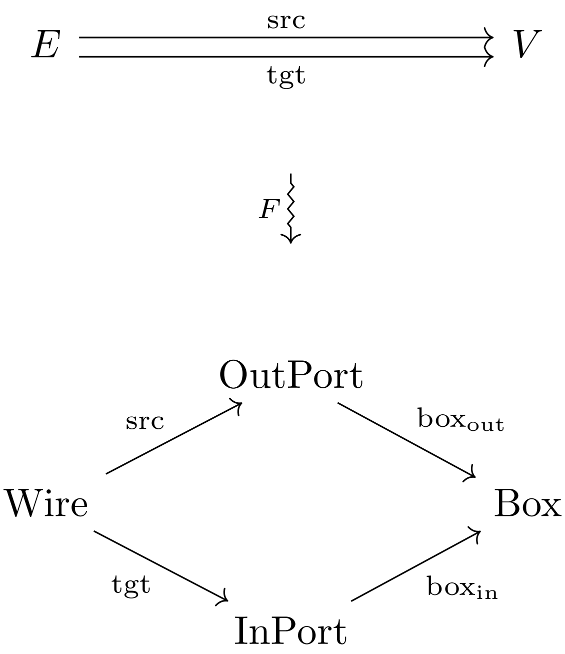


Useful for defining forgetful functors and other "projections."

Functorial data migration

Example. (Underlying graph of port graph)

```
F = @migration SchGraph SchPortGraph begin
  V => Box
  E => Wire
  src => src · box_out
  tgt => tgt · box_in
end
```



Covariant data migration

Pullback data migrations always have **left** and **right** adjoints: given $F\colon C \to D$, have

$$\textsf{C-Set} \;\underset{\Pi_F}{\overset{\Sigma_F}{\underset{\perp}{\overset{\perp}{\longleftarrow \Delta_F \longrightarrow}}}}\; \textsf{D-Set}$$

E.g., left pushforward migrations useful for free constructions.

Challenges:

- Can be very difficult to compute [SW15]
- Resulting databases can be infinite
- Difficult even for experienced users to predict the results

Instead, we consider generalizing pullback data migration:

- Less automated, but more flexible, more explicit, and easier to implement

Contravariant data migration with queries

Motivation. Since Cat is cartesian closed,

$$C\text{-Set} \to D\text{-Set} \qquad \leftrightsquigarrow \qquad D \to \text{Set}^{C\text{-Set}}$$

(ignoring size issues) where $\text{Set}^{C\text{-Set}}$ is the category of "all possible queries on C-sets."

Problem. It's too big in every sense.

Solution. Restrict to a tractable class of queries.

Let us start with the *representable* queries, of the form $X \mapsto C\text{-Set}(Q, X)$ for some Q.

- In database jargon, these are the *conjunctive queries*
- C-set Q is the *frozen instance* corresponding to the query

Problem. Finitely presentable queries can have infinite representing objects Q.

Solution. Replace $(C\text{-Set})^{\mathrm{op}}$ with more syntactical category: a category of diagrams!

Diagrams and limits

Original motivation for diagram categories is exposing the functorality of limits [EM45].

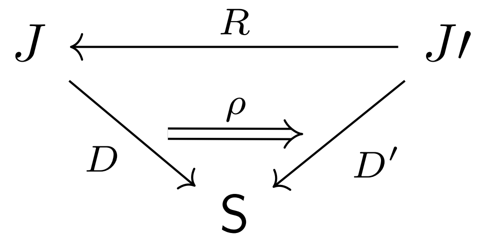
Theorem. When S is a complete category, taking limits gives a functor

$$\lim\colon \mathsf{Diag}^{\mathrm{op}}(\mathsf{S}) \to \mathsf{S}, \qquad \left(\mathsf{J} \xrightarrow{D} \mathsf{S}\right) \mapsto \lim D.$$

Dually, when S is a cocomplete category, taking colimits gives a functor

$$\operatorname{colim}\colon \mathsf{Diag}(\mathsf{S}) \to \mathsf{S}, \qquad \left(\mathsf{J} \xrightarrow{D} \mathsf{S}\right) \mapsto \operatorname{colim} D.$$

Operationally: define j'th leg of cone over D' as composite $\pi_j \cdot \rho_{j'}$, where $j := Rj'$.



Conjunctive data migration

Definition. A *conjunctive data migration* $C\text{-Set} \to D\text{-Set}$ is a data migration defined by a functor $F\colon D \to \mathsf{Diag}^{\mathsf{op}}(C)$.

Explicitly:

- Every object in D assigned a diagram in C, interpreted as a limit/conjunctive query
- Every morphism in D assigned a morphism of diagrams in C

Migration is evaluated by computing limits in Set: given $X\colon C \to \mathsf{Set}$, return

$$D \xrightarrow{F} \mathsf{Diag}^{\mathrm{op}}(C) \xrightarrow{\mathsf{Diag}^{\mathrm{op}}(X)} \mathsf{Diag}^{\mathsf{op}}(\mathsf{Set}) \xrightarrow{\mathsf{lim}} \mathsf{Set}.$$

Conjunctive data migration

Example. (Graph with edges the paths of length 2)

```
F = @migration SchGraph SchGraph begin
  V => V
  E => @join begin
    v::V; e1::E; e2::E
    tgt(e1) == v
    src(e2) == v
  end
  src => e1 · src
  tgt => e2 · tgt
end
```

In this query, object E is assigned to the diagram $E \xrightarrow{\text{tgt}} V \xleftarrow{\text{src}} E$.

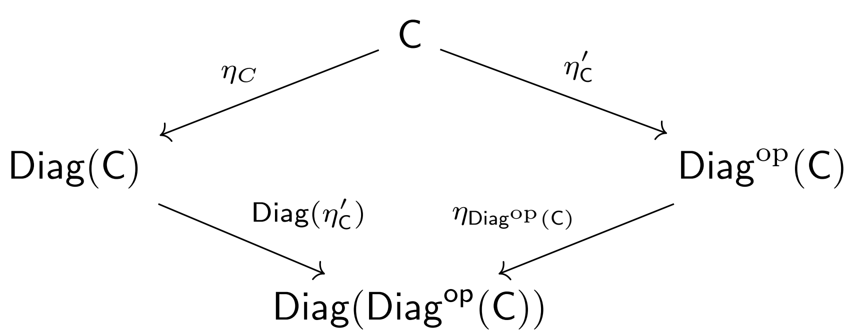
The hierarchy of queries

Dualizing to include colimits ("gluing queries"), we get the hierarchy of queries:

Query class	Category
Trivial queries	C
Conjunctive queries	$\mathrm{Diag}^{\mathrm{op}}(C)$
Disjoint unions	$\mathrm{Bun}(C)$
Duc ("disjoint union of conjunctive") queries	$\mathrm{Bun}(\mathrm{Diag}^{\mathrm{op}}(C))$
Gluing queries	$\mathrm{Diag}(C)$
Gluc ("gluings of conjunctive") queries	$\mathrm{Diag}(\mathrm{Diag}^{\mathrm{op}}(C))$

$\mathrm{Bun}(C)$ ("bundles in C") is full subcategory of $\mathrm{Diag}(C)$ spanned by discrete diagrams.

Remark. (Coercion) In practice, we implicitly convert between query classes:



Duc data migration

Example. (Graph with edges the paths of length $\leqslant 2$)

```
F = @migration SchGraph SchGraph begin
  V => V
  E => @cases begin
    v => V
    e => E
    path => @join begin
      v::V; e1::E; e2::E
      tgt(e1) == v
      src(e2) == v
    end
  end
  src => begin
    e => src
    path => e1·src
  end
  tgt => (e => tgt; path => e2·tgt) # Abbreviated for space.
end
```

Gluing data migration

Example. (Free symmetric reflexive graph on a reflexive graph)

```
F = @migration SchSymmetricReflexiveGraph SchReflexiveGraph begin
  V => V
  E => @glue begin
    fwd::E; rev::E
    v::V
    (fwd_refl: v → fwd)::refl
    (rev_refl: v → rev)::refl
  end
  src => (fwd => src; rev => tgt)
  tgt => (fwd => tgt; rev => src)
  refl => v
  inv => begin
    fwd => rev; rev => fwd; v => v;
    fwd_refl => rev_refl; rev_refl => fwd_refl
  end
end
```

Conclusion: diagrams in data migration

Advantages. This form of data migration offers two major advantages over SQL queries:

1. Results are general databases, not just tables
2. Queries can include general colimits, not just disjoint unions (and unions actually work properly)

Summary

- Diagrams are a combinatorial syntax for queries
- Morphisms of diagrams define foreign key relations between queries
- Working prototype available now in Catlab.jl, with blog post forthcoming

Future work

- Composing queries using the monad of diagrams [Koc67, PT21]
- Flexible transformation of data attributes, using arbitrary Julia functions
- Integration with recent work on grouping and aggregation [Spi21]

Part 3: Tonti diagrams and computational physics

In collaboration with:

James Fairbanks

U. Florida

Andrew Baas

GTRI

Jesus Arias

GTRI

Maia Gatlin

GTRI

Sophie Libkind

Stanford

Owen Lynch

Utrecht

Project overview

Objective. Develop compositional methods and software to reduce the very substantial engineering effort needed to build physics simulators (PDE solvers), including

- multiple interacting physics ("multiphysics")
- complex geometric domains

Elements.

1. *Category-theoretic*: diagrammatic formalism for specifying and composing physical theories, loosely inspired by Tonti diagrams
2. *Differential-geometric*: differential operators and their discretizations
   a. specifically, the discrete exterior caluclus (DEC) [Hir03, DHLM05]
   b. implemented in two dimensions in CombinatorialSpaces.jl
3. *Numerical*: PDE solvers based on these components (software forthcoming)

This talk will focus on the category-theoretic aspects.

What is a Tonti diagram?

- "Tonti diagrams" are a loose family of informal diagrams used to depict the quantities and differential equations in physical theories
- Promoted and popularized by Enzo Tonti but variations abound among other authors (Bossavit, Deschamps, Oden and Reddy, ...)

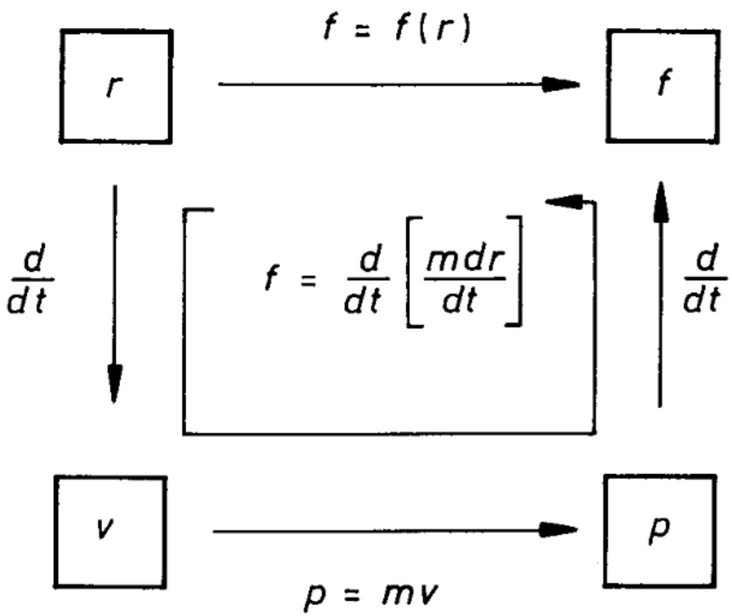


Figure. Tonti diagram for Newton's second law [LO91]

Maxwell's house: the origin of Tonti diagrams

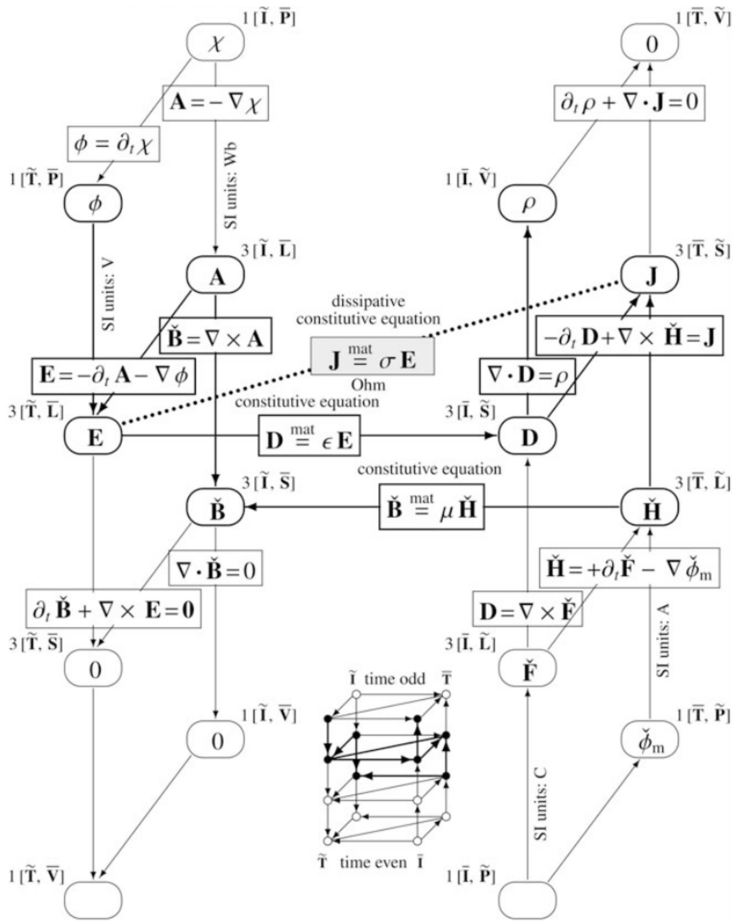


Figure. Tonti diagram for electromagnetism [Ton13]

Maxwell's house: the origin of Tonti diagrams

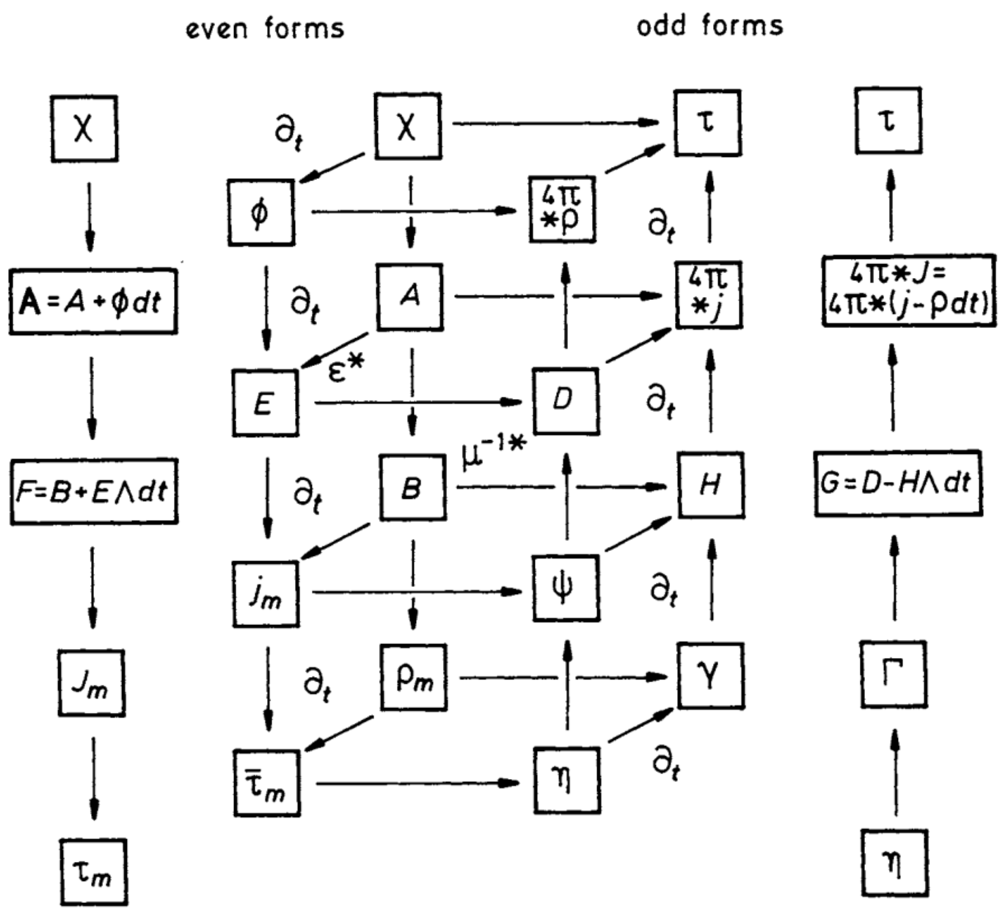


Figure. Tonti diagram for electromagnetism [LO91]

Maxwell's house: the origin of Tonti diagrams

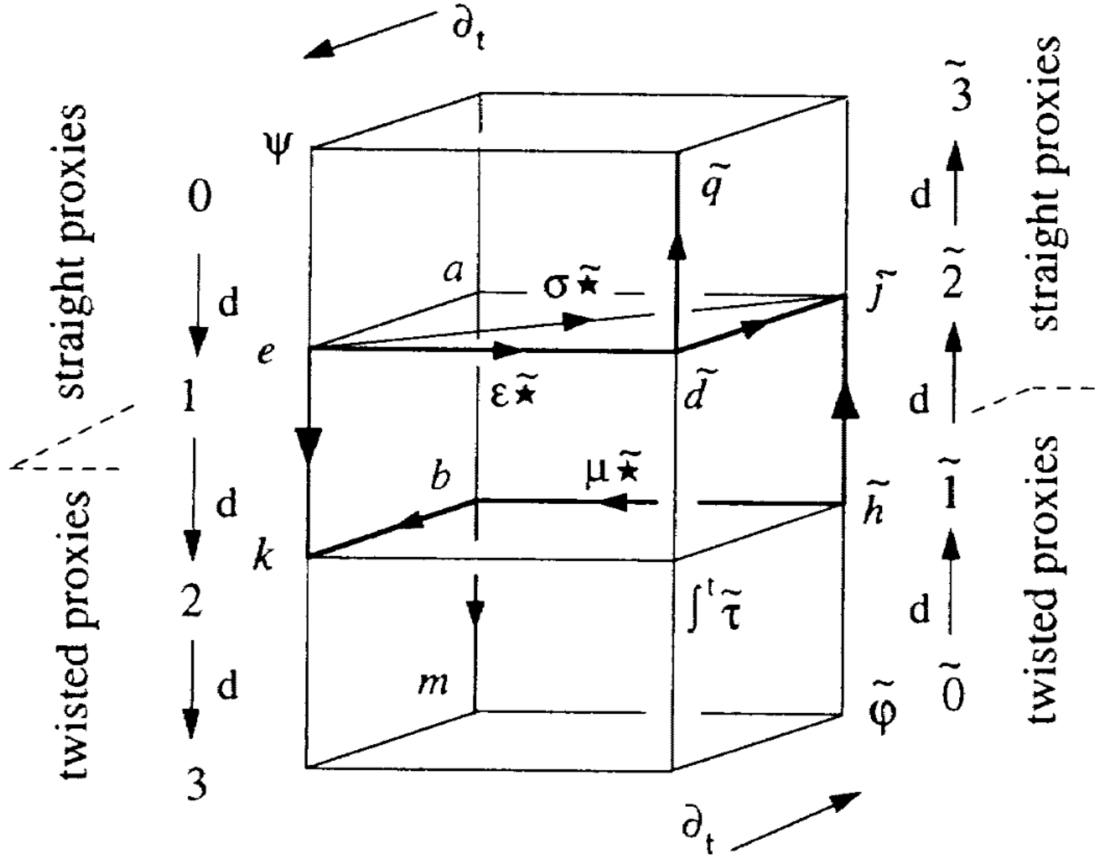


Figure. Maxwell's house according to Bossavit [Bos98]

Maxwell's house: the origin of Tonti diagrams

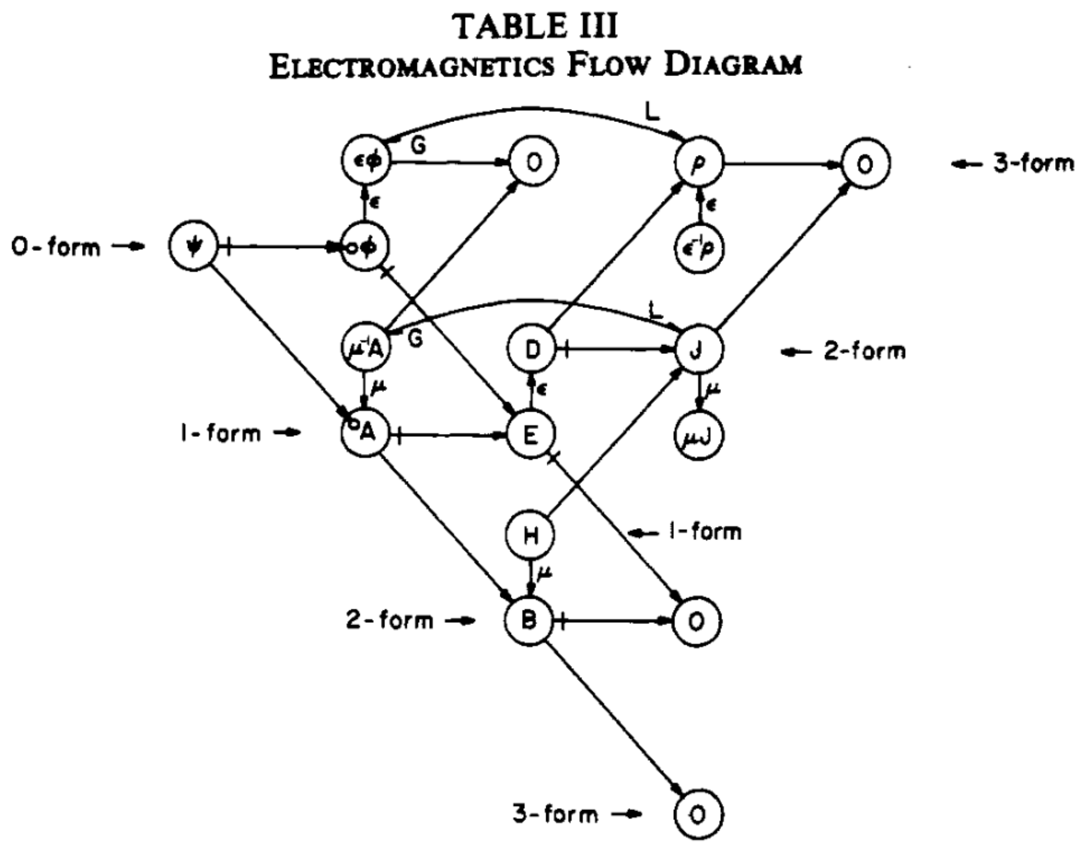


Figure. Electromagnetics diagram according to Deschamps [Des81]

Tonti diagrams in the wild

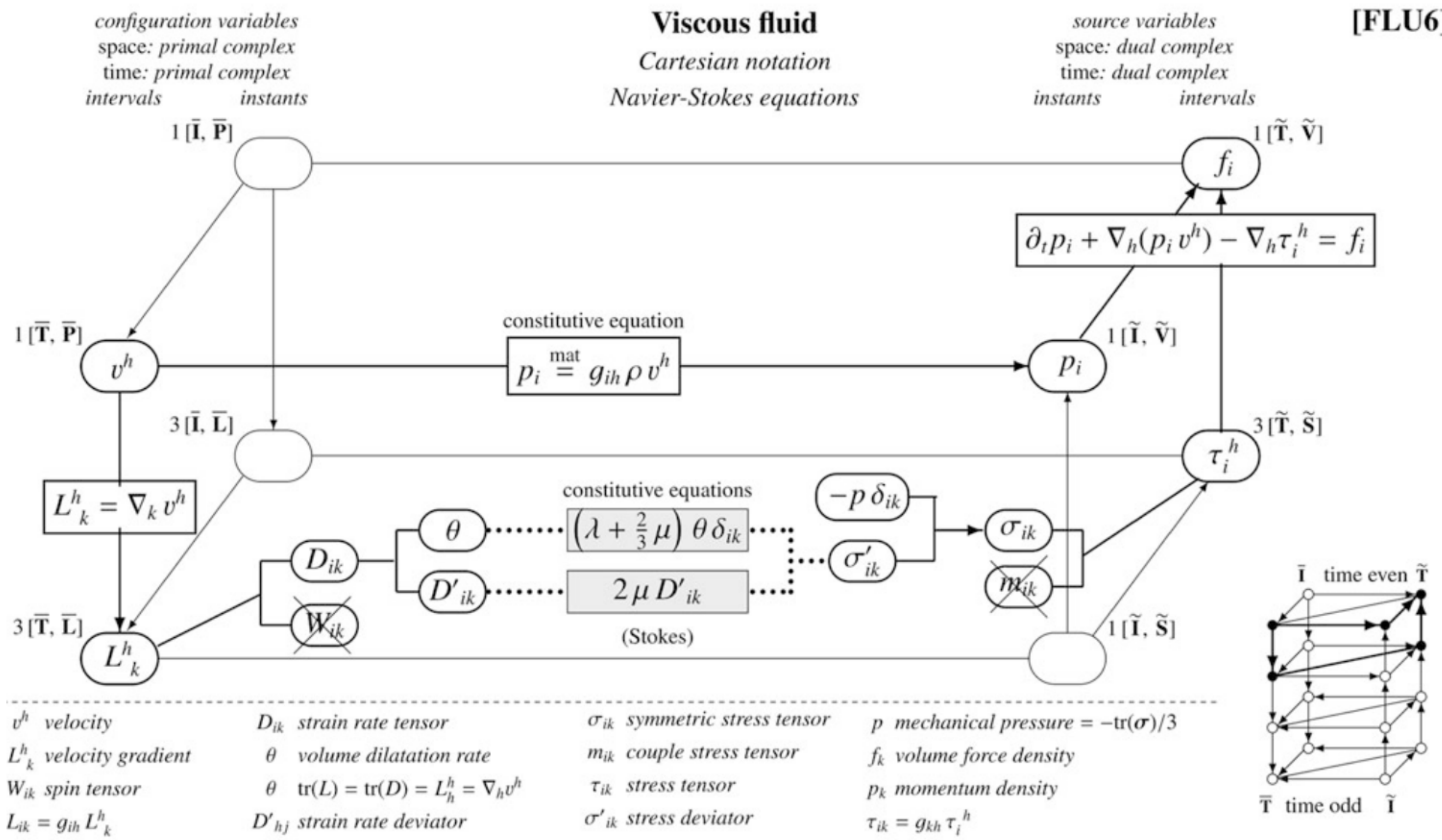


Figure. Tonti diagram for Navier-Stokes equation [Ton13]

But what is a Tonti diagram?

Question. Are Tonti diagrams "just" category-theoretic diagrams?

Answer. Almost: they are diagram lifting problems.

Background. Take as our setting a category C, having the interpretation:

- objects of C = spaces of physical quantities (scalar fields, vector fields, forms)
- morphisms of C = differential operators

For present purposes, we leave open the specifics.

- Minimalist choice is $C = \mathrm{Vect}_{\mathbb{R}}$
- More structure is present, e.g., in smooth case, objects are sheaves of vector spaces on Riemannian manifolds

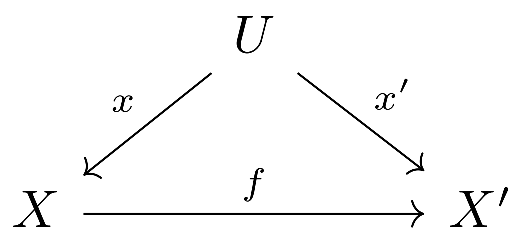
Diagrams of generalized elements

We are going to consider diagrams in a category of generalized elements of C.

Fix an object $U \in C$. (When $C = \mathrm{Vect}_{\mathbb{R}}$, take $U = \mathbb{R}$.)

Notation. Write $\mathrm{El}(C) := U \,/\, C$ for the coslice category having

- as objects, morphisms $U \xrightarrow{x} X$ (written "$x\colon X$") in C
- as morphisms $(x\colon X) \to (x'\colon X')$, $f\colon X \to X'$ in C forming a commuting triangle

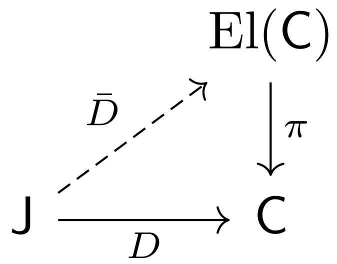


Idea.

diagram in C	$\leftrightsquigarrow$	system of equations
diagram in $\mathrm{El}\, C$	$\leftrightsquigarrow$	solution to a system of equations

Diagram lifting problems

Definition. The *lifting problem* associated with a diagram $D\colon J \to C$ is to find a diagram $\bar{D}\colon J \to \mathrm{El}\, C$ such that $\pi \circ \bar{D} = D$, where $\pi = \mathrm{cod}\colon \mathrm{El}\, C \to C$ is the canonical projection.



Equivalently, the lifting problem is to find a cone over D with apex U.

Remark. So, the limit of D, if it exists, is a "universal solution" of a class of lifting problems, where U ranges over C. In general:

- Limits are about *finding* solutions to equations
- Colimits are about *imposing* solutions

Example: diffusion equation

Phrased in exterior calculus, the diffusion equation is the lifting problem given by

$$\begin{array}{ccccc} C : \Omega_t^0 & \xrightarrow{\partial_t} & \dot{C} : \Omega_t^0 & \xleftarrow{\star^{-1}} & d\phi : \tilde{\Omega}_t^3 \\ {\scriptstyle d}\downarrow & & & & \uparrow{\scriptstyle d} \\ dC : \Omega_t^1 & & \xrightarrow[k\star]{\qquad\qquad} & & \phi : \tilde{\Omega}_t^2 \end{array}$$

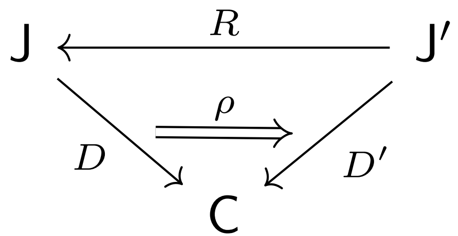
where

- we have fixed a three-dimensional Riemannian manifold M
- Ω_t^k (resp. $\tilde{\Omega}_t^k$) are the time-dependent straight (resp. twisted) smooth k-forms on M
- $C : \Omega_t^0$ is the concentration of the diffusing substance
- $\phi : \tilde{\Omega}_t^2$ is the negative diffusion flux
- $k > 0$ is the diffusivity, a constant

Warning. The diagrams in C or $\mathrm{El}\, C$ are free diagrams, and they do not commute!

Morphisms of diagrams

In view of the connection with limits, the correct category of diagrams is $\mathsf{Diag}^{\mathrm{op}}(C)$, where the morphisms look like:



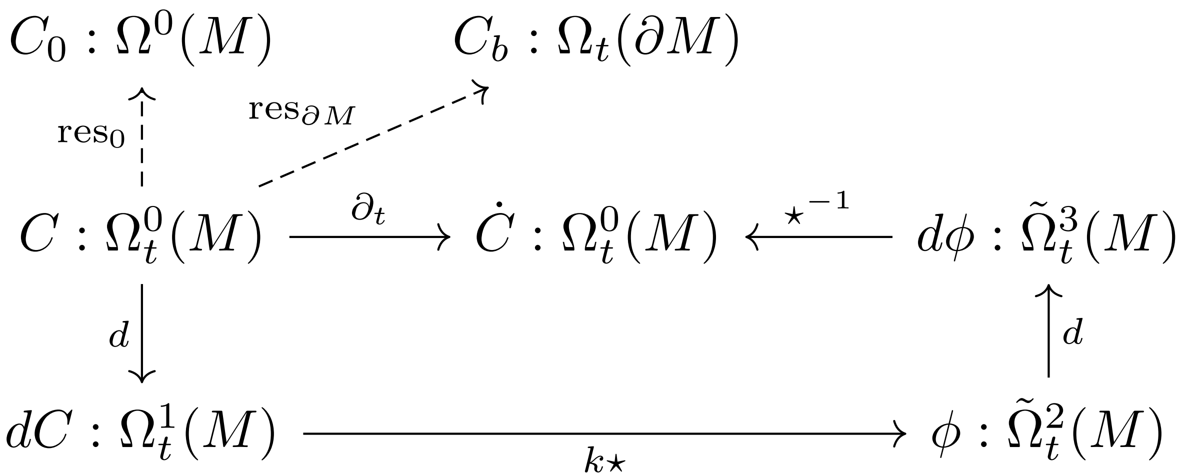
Two important uses for the morphisms:

1. Express boundary conditions and formalize boundary value problems as extension-lifting problems
2. Formalize relationships between different (presentations of) physical theories

Boundary conditions as diagram morphisms

Boundary conditions associated with a system $D \in \mathsf{Diag}^{\mathrm{op}}(\mathsf{C})$ can be represented by a morphism $D \to D_0$.

Example. (Diffusion equation with Dirichlet conditions)



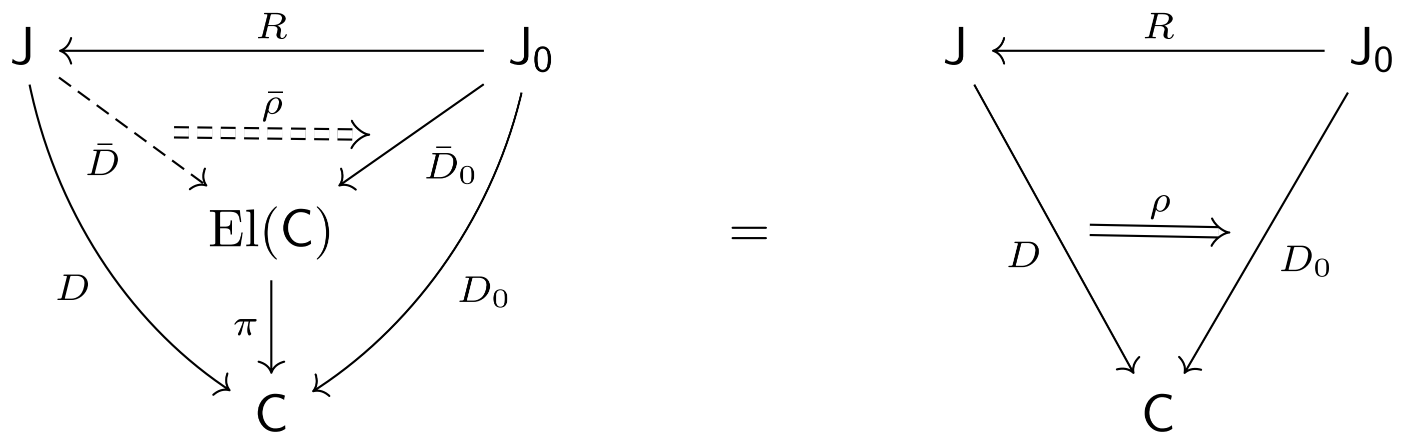
where the shape of D_0 is $\mathsf{J}_0 := \{\bullet, \bullet\}$ and

- $C_0 \colon \Omega^0(M)$ are initial conditions (C at time $t = 0$)
- $C_b \colon \Omega_t(\partial M)$ are boundary conditions (C on boundary of M)

BVPs as extension-lifting problems

Formally, a boundary value problem is a diagram extension-lifting problem.

Definition. Let $(R, \rho) \colon D \to D_0$ be a morphism of diagrams in C and $\bar{D}_0$ be a lift of D_0 to $\mathrm{El}\, C$. The *extension-lifting problem with data* $\bar{D}_0$ is to find a morphism of diagrams $(R, \bar{\rho}) \colon \bar{D} \to \bar{D}_0$ in $\mathrm{El}\, C$ such that $\mathrm{Diag}^{\mathrm{op}}(\pi)(R, \bar{\rho}) = (R, \rho)$.



- Lifting D to $\bar{D}$ through π = solving the equations
- Extending $\bar{D}_0$ to $\bar{D}$ through $R \colon J_0 \to J$ (up to ρ) = satisfying the boundary conditions

Transporting lifts along diagram morphisms

Proposition. Let $\pi\colon E \to C$ be a functor. Whenever π is a discrete opfibration, so is the functor $\mathrm{Diag}^{\mathrm{op}}(\pi)\colon \mathrm{Diag}^{\mathrm{op}}(E) \to \mathrm{Diag}^{\mathrm{op}}(C)$ given by post-composition with π.

Since $\pi = \mathrm{cod}\colon \mathrm{El}\, C \to C$ is a discrete obfibration, this means that:

- a diagram morphism $D \to D'$ transport lifts of D to lifts of D'
- in particular, given a BVP $D \to D_0$, the boundary values of a possible solution $\bar{D}$ can be computed, as one would expect!

Example: a variation on the diffusion equation

A strict morphism of diagrams connects two different presentations of the equation:

$$
\begin{array}{ccccc}
C : \Omega^0_t & \xrightarrow{\partial_t} & \dot{C} : \Omega^0_t & \xleftarrow{\star^{-1}} & d\phi : \tilde{\Omega}^3_t \\
{\scriptstyle d}\big\downarrow & & & & \big\uparrow{\scriptstyle d} \\
dC : \Omega^1_t & & \xrightarrow[k\star]{\qquad\qquad\qquad} & & \phi : \tilde{\Omega}^2_t
\end{array}
$$

$$\rightsquigarrow$$

$$
C : \Omega^0_t \;\underset{k\Delta}{\overset{\partial_t}{\rightrightarrows}}\; \dot{C} : \Omega^0_t
$$

Here $\Delta := \star^{-1} d \star d$ is the *Laplace-Beltrami* operator.

Extensions and applications

Extensions. Many extensions to formalism that I have not discussed:

- Weak equivalences of diagrams based on *initial functors* (cf. Street & Walters [SW73])
- Composition of free diagrams using *structured cospans* [Fon15, BC20]
- Diagrams involving cartesian products
- Diagrams involving monoidal products, e.g., tensor product in $\mathsf{Vect}_{\mathbb{R}}$

With the latter upgrades, we can express the major equations of mathematical physics, such as Maxwell's equations and the Navier-Stokes equations.

Applications. I have also not discussed our computational pipeline:

equations (diagrams) $\longrightarrow$ computation graphs (wiring diagrams) $\longrightarrow$ simulations (Julia)

For fun, I'll show a simulation anyway: evolution of electromagnetic fields (Maxwell's equations) with fully reflecting boundary.

Bibliography

[BC20] John C. Baez and Kenny Courser. Structured cospans. *Theory and Applications of Categories*, 35(48):1771–1822, 2020.

[Bos98] Alain Bossavit. On the geometry of electromagnetism (4): Maxwell's house. *J. Japan Soc. Appl. Electromagn. & Mech*, 6(4):12–20, 1998.

[Des81] Georges A. Deschamps. Electromagnetics and differential forms. *Proceedings of the IEEE*, 69(6):676–696, 1981.

[DHLM05] Mathieu Desbrun, Anil N. Hirani, Melvin Leok, and Jerrold E. Marsden. Discrete exterior calculus. *ArXiv:math/0508341*, 2005.

[EM45] Samuel Eilenberg and Saunders Mac Lane. General theory of natural equivalences. *Transactions of the American Mathematical Society*, 58(2):231–294, 1945.

[Fon15] Brendan Fong. Decorated cospans. *Theory and Applications of Categories*, 30(33):1096–1120, 2015.

[Gui73] René Guitart. Sur le foncteur diagrammes. *Cahiers de Topologie et Géométrie Différentielle Catégoriques*, 14(2):181–182, 1973.

[Gui74] René Guitart. Remarques sur les machines et les structures. *Cahiers de Topologie et Géométrie Différentielle Catégoriques*, 15(2):113–144, 1974.

[GVdB77] René Guitart and Luc Van den Bril. Décompositions et lax-complétions. *Cahiers de Topologie et Géométrie Différentielle Catégoriques*, 18(4):333–407, 1977.

[Hir03] Anil N. Hirani. *Discrete exterior calculus*. PhD thesis, California Institute of Technology, 2003.

[Koc67] Anders Kock. *Limit monads in categories*. PhD thesis, The University of Chicago, 1967.

[LO91] Ernesto A. Lacomba and Fausto Ongay. On a structural schema of physical theories proposed by E. Tonti. In *The Mathematical Heritage of C. F. Gauss*, pages 432–453. World Scientific, 1991.

[PT20] George Peschke and Walter Tholen. Diagrams, fibrations, and the decomposition of colimits. *ArXiv:2006.10890*, 2020.

[PT21] Paolo Perrone and Walter Tholen. Kan extensions are partial colimits. *ArXiv:2101.04531*, 2021.

[Spi12] David I. Spivak. Functorial data migration. *Information and Computation*, 217:31–51, 2012.

[Spi21] David I. Spivak. Functorial aggregation. *ArXiv:2111.10968*, 2021.

[SW73] Ross Street and R. F. C. Walters. The comprehensive factorization of a functor. *Bulletin of the American Mathematical Society*, 79(5):936–941, 1973.

[SW15] David I. Spivak and Ryan Wisnesky. Relational foundations for functorial data migration. In *Proceedings of the 15th Symposium on Database Programming Languages*, pages 21–28. 2015.

[Ton13] Enzo Tonti. *The mathematical structure of classical and relativistic physics*. Springer, 2013.