


# homotopy.io: A Proof Assistant for Finitely-Presented Globular $n$ -Categories

**Nathan Corbyn** ✉   
University of Oxford, Oxford, UK

**Lukas Heidemann** ✉   
University of Oxford, Oxford, UK

**Nick Hu** ✉   
University of Oxford, Oxford, UK

**Chiara Sarti** ✉   
University of Cambridge, Cambridge, UK

**Calin Tataru** ✉   
University of Cambridge, Cambridge, UK

**Jamie Vicary** ✉   
University of Cambridge, Cambridge, UK

---

## Abstract

We present the proof assistant `homotopy.io` for working with finitely-presented semistrict higher categories. The tool runs in the browser with a point-and-click interface, allowing direct manipulation of proof objects via a graphical representation. We describe the user interface and explain how the tool can be used in practice. We also describe the essential subsystems of the tool, including collapse, contraction, expansion, typechecking, and layout, as well as key implementation details including data structure encoding, memoisation, and rendering. These technical innovations have been essential for achieving good performance in a resource-constrained setting.

**2012 ACM Subject Classification** Software and its engineering → Software notations and tools

**Keywords and phrases** Higher category theory, proof assistant, string diagrams

**Digital Object Identifier** 10.4230/LIPIcs.FSCD.2024.30

**Related Version** *Full Version*: <https://arxiv.org/abs/2402.13179>

**Supplementary Material** *Software (Source Code)*: <https://github.com/homotopy-io/homotopy-rs>  
archived at `swh:1:dir:836bb913a5b2d2556f124f1a260a631a1a7f7387`

*Audiovisual (Video Tutorial)*: <https://homotopy.io/braiding-example> [1]

**Funding** *Nathan Corbyn*: EPSRC Industrial CASE Studentship

*Nick Hu*: EPSRC Doctoral Training Partnership Scholarship [grant number 2218955]

**Acknowledgements** The authors would like to thank Anastasia Courtney, Yulong Huang, and Jasper Parish for their contributions during their summer internships, Akvilė Valentukonytė and Klaudia Urbanska for their contributions during their undergraduate projects, and Manuel Araújo, Wilf Offord, and Hao Xu for extensive testing of the tool and valuable feedback. We are also grateful to the students of the “Categorical Quantum Mechanics” course at Oxford, and the “Advanced Topics in Category Theory” course at Cambridge for testing and feedback.

## 1 Introduction

Higher category theory [26, 34] is a branch of mathematics that now has a wide range of applications, in areas as diverse as logic [18, 41], quantum field theory [3, 36], and geometry [27]. For working practically with higher categories, string diagrams are an



© Nathan Corbyn, Lukas Heidemann, Nick Hu, Chiara Sarti, Calin Tataru, and Jamie Vicary;  
licensed under Creative Commons License CC-BY 4.0

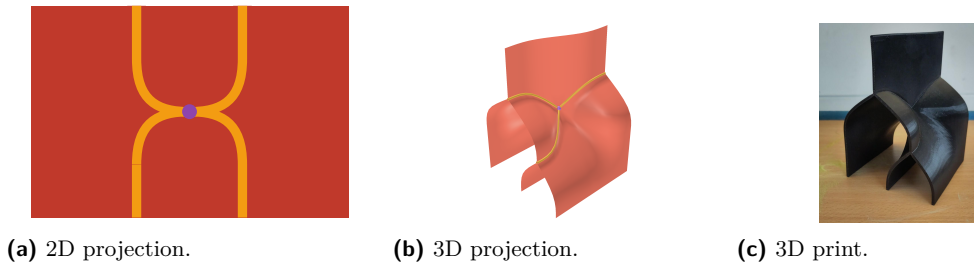
9th International Conference on Formal Structures for Computation and Deduction (FSCD 2024).

Editor: Jakob Rehof; Article No. 30; pp. 30:1–30:26



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



■ **Figure 1** The associator 3-diagram shown in both 2D and 3D (link to online workspace).

increasingly popular technique, introduced for monoidal categories by Joyal and Street [23], and since extended into higher dimensions by a range of authors [5, 7, 11, 13]. In a higher-categorical setting, string diagrams take the form of higher-dimensional manifold-like structures, which can efficiently encode complex compositional information. However, these structures can be hard to visualise, manipulate, or represent in research articles, limiting their effectiveness.

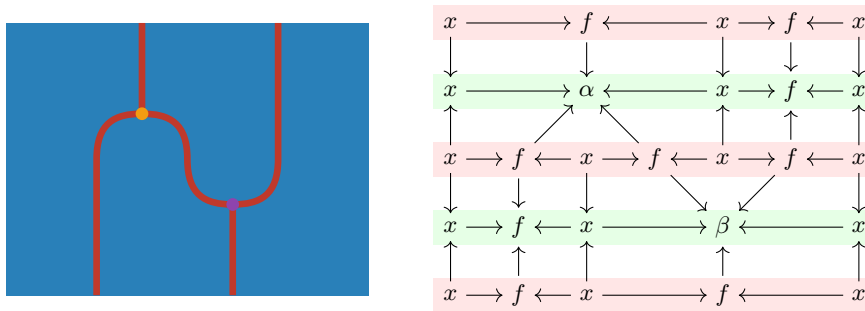
Our tool aims to bridge this gap, allowing string diagrams to become a practical technique for working higher category theorists. It runs in the browser, giving a low barrier-to-entry, and allows direct construction and manipulation of graphical representations of  $n$ -dimensional categorical structures, which we call  $n$ -*diagrams*, by direct point-and-click manipulation with the mouse or touch interface.

In this article, we give an overview of the user interface, and describe the following major subsystems, which give the tool its range of functionality.

- *Collapse* acts on a combinatorial  $n$ -diagram, constructing a quotient geometry by identifying points which can be considered topologically equivalent.
- *Contraction* allows a region of an  $n$ -diagram to be geometrically contracted to a point, yielding a homotopy that is itself encoded by an  $(n + 1)$ -diagram; the collapse algorithm gives its base case. This is the major mechanism for constructing all nontrivial diagrams in the theory.
- *Expansion* provides a limited converse to contraction, defined on a diagram with at least two vertices at the same height, with the effect that one vertex is moved to an adjacent height. The resulting diagram will always contract to yield the original diagram.
- *Typechecking* analyses an instance of our  $n$ -diagram data structure, and decides whether it represents a valid composite  $n$ -morphism in a free higher category.
- *Layout* generates a set of linear constraints representing the necessary coordinate relationships between all the parts of a diagram (such as the vertices, wires, and regions), which can be passed to a linear solver, and used by the rendering pipeline.

We also examine two significant aspects of the implementation.

- *Memoisation* is necessary since the  $n$ -diagrams stored by the tool have an intricate recursive structure, which in principle encodes all sub- $k$ -diagrams for  $k < n$ . If this data was stored separately in memory, the resource requirements of the proof assistant would grow exponentially with diagram dimension. Memoisation ensures each logically-distinct  $k$ -diagram is stored only once in memory.
- *Rendering* is a complex pipeline that produces suitable output on the screen (see Figure 1); we use SVG for output in dimension 0, 1, and 2, and WebGL for output in dimension 3 and 4. A subdivision algorithm is necessary to produce visually appealing output. We also render to STL for 3D printing, and to TikZ for convenient diagram export.



■ **Figure 2** A string diagram and the corresponding zigzag encoding.

The version of the tool presented in this article is a pre-release version, available here:

<https://beta.homotopy.io>

It is written in the Rust programming language, and compiled to WebAssembly to run in the web browser. The implementation is available under a free open-source licence on GitHub.

### 1.1 Mathematical Context

Our proof assistant implements the theory of higher categories known as *associative n-categories*, due to Dorn, Douglas and Vicary [12, 33]. This model is *globular*, in the sense that for  $n \geq 2$  the boundary data of any  $n$ -cell satisfies the globularity condition: the source of the source equals the source of the target, and the target of the source equals the target of the target. It is also strictly associative and unital, while retaining weak interchangers; in this sense it is a *semistrict* theory. It is conjectured that every weak higher category is equivalent to an associative  $n$ -category, although the proof of this remains out of reach.

The proof assistant allows users to build composite cells in the graphical language for freely generated semistrict globular  $n$ -categories which are free on a *signature*, a list of variables of specified types. For example, to define a monad-like structure, starting from the empty signature, we might first add an object  $x$ , followed by a 1-cell  $f: x \rightarrow x$ , followed in turn by a multiplication 2-cell  $\mu: f \circ f \rightarrow f$ , which we interpret as the monad multiplication.

The tool allows generators of non-zero dimension to be optionally tagged as invertible; this allows the user to work with directed higher categories (if no generators are tagged), higher groupoids (if all generators are tagged), or more general structures. The resulting invertible structure is coherent, meaning that it automatically satisfies the necessary higher-dimensional constraints, such as the adjunction equations.

In the implementation,  $n$ -cells are represented combinatorially as *n-diagrams*, simple inductive data structures which allow us to represent the mathematical zigzag construction [33]. We depict this in Figure 2, which illustrates the encoding of a 2-diagram (drawn on the left) in terms of an iterated sequence of cospans (drawn on the right). Here, a zigzag is a sequence of cospans taking value in some category of labels. We distinguish the *singular heights* drawn in green, where vertices might appear, from the *regular heights* drawn in red which are adjacent. Since zigzags and their morphisms themselves form a category, this construction can be iterated, and in this way higher-dimensional diagrams can be represented. This inductive nature of the construction is an essential requirement for our proof assistant: it allows us to describe  $n$ -cells as algebraic data types, as we will detail in Section 4, and to structure our key algorithms of Section 5 as recursive procedures.

■ **Table 1** Comparison of existing tools for string diagrams.

tool	generality	interactive	invertibility	visualisation
homotopy.io	$n$ -categories	✓	✓	up to 4D
Cartographer [37]	symmetric monoidal categories	✓	✗	2D
DisCoPy [10]	monoidal categories	✗	✗	2D
Globular [6]	4-categories	✓	partial	2D
rewalt [16]	$n$ -categories	✗	partial	2D
sd-visualiser [20]	traced cartesian closed categories	✓	✗	2D
Quantomatic [24]	compact closed categories	✓	✗	2D
wiggle.py [9]	monoidal 2-categories	✗	✗	up to 3D

## 1.2 Related Work

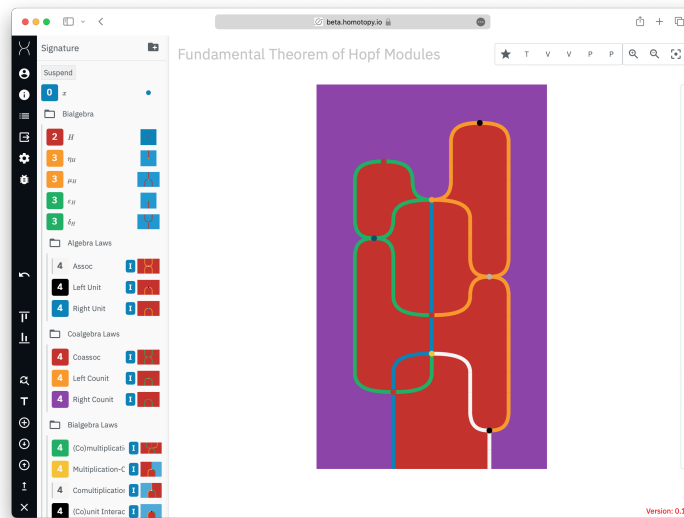
There are many tools for higher categories that use string diagrams as a visualisation method, and we summarise several of them in Table 1. Each takes a different categorical perspective, and is focused on a particular formalism. The most closely related tool is Globular, the direct precursor to homotopy.io, which was limited to 4-categories and lacked full support for coherent inverses. One other tool, *rewalt*, also allows manipulations at the level of  $n$ -categories for arbitrary  $n$ . Both homotopy.io and *rewalt* implement semistrict  $n$ -categories, but they have different notions of semistrictness: homotopy.io has strict associators and unitors, but weak interchanges; whereas *rewalt* has strict interchangers and associators, with weak identities. This difference of approach means that the corresponding notions of string diagram are quite different in each tool. The homotopy.io tool is the first string diagram proof assistant that can handle coherent invertible generators in all dimensions, an aspect it shares with traditional type-theoretic approaches to higher category theory, such as homotopy type theory [41], interpreted via proof assistants such as Agda or Coq.

This paper is the first detailed description of homotopy.io, with previous works focused on aspects of the theoretical foundations [33, 17, 39, 35, 40, 21]. The theory of associative  $n$ -categories, which is the basis of homotopy.io, was first developed by Dorn, Douglas, and Vicary and described in Dorn’s PhD thesis [12].

## 2 Using the Tool

The tool consists of two main components: the *signature* and the *workspace*. The signature stores a list of generators for an  $n$ -category, and the workspace stores an  $n$ -diagram in the free  $n$ -category generated by this signature. The tool implements a number of *actions* to modify the signature and/or workspace, and every state of the tool is determined by the list of actions that led to that state, starting from the empty signature and workspace. This makes it easy to implement an undo/redo system, and to reproduce a state by replaying the list of actions which is useful for debugging and testing.

All actions can be triggered by clicking on UI elements such as the buttons on the sidebar, interacting with the workspace diagram, or with keyboard shortcuts. We denote the keyboard shortcut associated to an action as A, which represents pressing the “A” key. Unlike a traditional proof assistant, there is no text-based aspect to the user-interface, except for metadata and for providing generator names. A screenshot of the user interface is shown in Figure 3. In this section, we describe its major components.



■ **Figure 3** The interface of the proof assistant.

**Adding generators.** There are two primary ways to add new generators to the signature:

- We can add new 0-dimensional generators by clicking “Add 0-cell” ( $\boxed{A}$ ) on the sidebar.
- We can add new  $(n + 1)$ -dimensional generators by constructing a source  $n$ -diagram in the workspace and clicking “Source” ( $\boxed{S}$ ), and then constructing a compatible target  $n$ -diagram in the workspace and clicking “Target” ( $\boxed{T}$ ). The order of these two actions is not important, and the first diagram will be stashed and displayed in the bottom-left corner of the workspace until the second one is constructed.

**Manipulating the workspace.** Clicking on any generator in the signature will bring that diagram into the workspace. We can raise the dimension of the diagram with the “Identity” ( $\boxed{I}$ ) action. We can construct composite diagrams by attaching another diagram to a boundary by clicking on the edge of the diagram. We can also use generators in the signature to perform rewrites by clicking *inside* the workspace diagram. Finally, we can perform homotopies, such as contractions and expansions which will be described in Section 5, by clicking and dragging.

**Theorems.** If the workspace is displaying an  $n$ -diagram  $D$  for  $n > 1$ , the “Theorem” ( $\boxed{H}$ ) action becomes available. This action creates a new  $n$ -dimensional generator  $T$  with the same type as  $D$ , and a new invertible  $(n + 1)$ -dimensional generator  $P: T \rightarrow D$ . This could be done by hand, and in this sense this action does not strictly add functionality, rather it adds a useful shortcut.

The idea is that  $T$  is an algebraic generator which axiomatises the existence of  $D$ , therefore allowing users to use it as a rewrite, and the generator  $P$  witness the fact that  $T$  is true (i.e. is inhabited), by rewriting it to  $D$ . This feature is useful for formalising complex proofs that depend on other lemmas, by saving each lemma as a theorem and then combining them to prove the main theorem. It can also be used to give definitions – i.e. we can think of  $T$  as a new generator which is *defined* to equal  $D$ , and  $P$  can then be used to expand the definition. This is similar to using the `abstract` keyword in Agda, to hide the “implementation detail” of a proof.

**View control.** Visualising  $n$ -diagrams for  $n > 2$  presents a fundamental challenge, since the geometry can be difficult for us to visualise. A number of features are therefore provided for manipulating the visualisation: projecting out certain dimensions, navigating to a subdiagram, and changing the rendering dimension (with a choice between 0 and 4 dimensions).


These manipulations are effected via the *view control* component in the top-right corner of the workspace. This consists of a list of  $n + 1$  buttons, where  $n$  is the dimension of the workspace diagram, corresponding to the following regular expression:

$$\star (S | T | Ri | Si)^k V^d P^{n-k-d} \quad (d \leq 4)$$

Here  $S$  means source,  $T$  means target,  $Ri$  means the  $i$ -th regular height, and  $Si$  means the  $i$ -th singular height. This means that we are viewing a  $k$ -dimensional subdiagram of the workspace diagram, which is an  $(n - k)$ -diagram obtained by recursively going into heights as specified by the first  $k$  symbols. This subdiagram is then projected to  $d$  dimensions; that is, we are *viewing*  $d$  dimensions, and *projecting* the remaining  $n - k - d$  dimensions. Clicking the star will reset the view to the original  $n$ -diagram (i.e. reset  $k$  to 0), and clicking on any of the  $k$  view buttons will reset the view to some prefix. Clicking on any  $V$  button will decrement  $d$  by 1, to a minimum of 0 (displayed as a single point) and clicking on any  $P$  button will increment  $d$  by 1, to a maximum of 4 (rendered as a movie of 3D geometries).

Users can descend into a height by clicking the chevrons appearing on the right-hand side of the diagram, or descend into the source height by pressing  $\boxed{\rightarrow}$ . Similarly, while inside a height, users can ascend to the parent height by pressing  $\boxed{\leftarrow}$ , and navigate to adjacent  $k$ -th height components by pressing  $\boxed{\uparrow}/\boxed{\downarrow}$ .

To illustrate how the projection works, consider Figure 1 which shows the 2D and 3D projections of a 3-dimensional diagram side-by-side. Note that the 2D projection can be understood as looking at the 3D projection from below, and projecting onto a 2D plane.

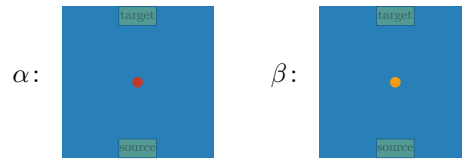
**Generator options.** When hovering over a generator, a  icon appears on the left which reveals a menu of options for that generator. This allows the user to rename the generator (with support for L<sup>A</sup>T<sub>E</sub>X) or change its colour or shape. Most importantly, it allows the user to mark a generator as invertible.

**Image export.** The sidebar has an “Image Export” button which reveals a panel for exporting the workspace diagram to different formats, such as SVG, TikZ, Manim, and STL.

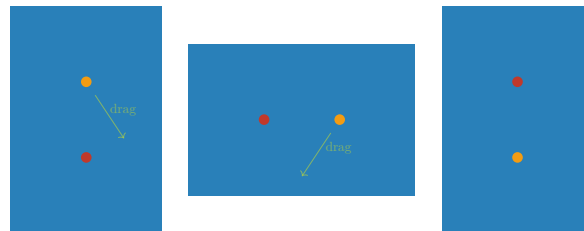
### 3 Example: Eckmann-Hilton

Here we will illustrate how the proof assistant may be used in practice to formalise results in higher category theory. Our running example will be the Eckmann-Hilton argument, the key result in the correspondence between braided monoidal categories and 3-categories which are doubly-degenerate, meaning they have a unique 0-cell and no non-identity 1-cells. This will be essential in Appendix A for our formalisation of Hopf algebras in braided monoidal categories. This section is accompanied by a video tutorial [1]. The resulting workspace can be loaded into the tool at <https://beta.homotopy.io/p/2402.00002>.

To formalise the Eckmann-Hilton argument, we load the tool and begin constructing our signature, which is given by a unique 0-cell  $x$  and two 2-cells  $\alpha, \beta$  which have source and target the identity on  $x$ . Since  $x$  is a 0-cell, we may immediately add this to the signature by pressing the “Add 0-Cell” button, and renaming the cell to “ $\$x\$$ ” – the use of “ $\$$ ”s is optional, but indicates to the tool that the generator’s name should be interpreted as L<sup>A</sup>T<sub>E</sub>X. To add



■ **Figure 4** The scalars  $\alpha$  and  $\beta$  in our signature for the Eckmann-Hilton argument.



■ **Figure 5** The Eckmann-Hilton argument.

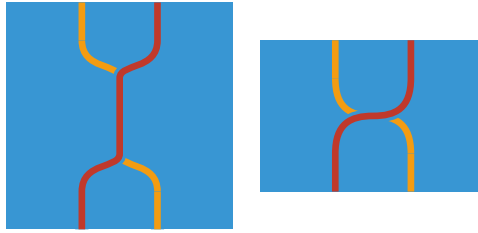
the two scalars  $\alpha$  and  $\beta$  to the signature, we need to provide the tool with their sources and targets, which in this case is  $1_x$ . Selecting the generator  $x$  we have added to the signature, we make it the current working diagram. We then click on the “Identity” button to get  $1_x$  from  $x$ , and take it as the source of a new 2-cell  $\alpha$  by clicking on “Source”. Repeating the construction of  $1_x$ , we can provide the tool with the target of  $\alpha$ . Since the diagrams have compatible boundaries, the globularity condition is satisfied and the tool allows to click on “Target” to create a 2-cell, which we rename to “ $\alpha$ ”. By repeating this procedure for  $\beta$ , we get the signature depicted in Figure 4.

With our signature in place, we can begin proving facts about the compositional behaviour of our scalars. The essence of the Eckmann-Hilton argument amounts to the commutativity and coincidence of vertical and horizontal composites of scalars, as depicted in Figure 5.

Let us detail how this proof can be built. We begin by constructing the right-most diagram in Figure 5 by selecting  $\alpha$  from the signature, clicking on the highlighted region marked as *target* in Figure 4 and selecting  $\beta$  to attach. This builds the vertical composite of  $\alpha$  and  $\beta$  and gives us the starting point or source for our proof. Since we wish to collect the steps of the proof into a 3-cell, we take the identity on the current diagram. This does not appear to change the current diagram displayed but adds an extra dimension in the view control.

We then drag  $\beta$  towards  $\alpha$  along the right to get the middle diagram, triggering a *contraction* procedure, which we will detail in Section 5.2. Next, we drag  $\beta$  further downwards, triggering the dual procedure of *expansion*, detailed in Section 5.3. Note that these procedures only succeed if they are sound, i.e. the resulting diagram has a valid type as checked by the procedure of Section 5.4.

We now inspect the proof from the top-dimension by clicking on the  $\star$  button on the right. From this view, our proof amounts to the construction of a braid as depicted in the first image of Figure 6. We can see this more clearly if we contract the middle part of the diagram by vertically dragging the bottom half-braid towards the top, obtaining the second image of Figure 6. We may also perform the proof entirely from this view, by horizontally dragging the legs of the braid past each other. A 3D visualisation of this proof can now be observed by pressing the P button in the view control.



■ **Figure 6** The Eckmann-Hilton proof (left) and its contraction (right).

## 4 Core Data Structures

In this section we outline the fundamental algebraic data types which are used to encode  $n$ -diagrams in the implementation. These are diagrams, rewrites, cones and cospans; their definitions are sketched in Figure 7. Note that the Rust type system, along with our choice of encoding, has the following limitation: there exist valid terms of the `Diagram` type which do not correspond to admissible  $n$ -cells of the free higher category generated by the current signature. For example, any `Diagram0` which consists of a generator with its `dimension` field set to a non-zero integer does not correspond to any 0-cell of any higher category. The typechecking procedure of Section 5.2 is used to determine validity of these terms.

```

1 type frame = int
2 type generator = { dimension: int; id: int }
3
4 type rewrite =
5   | Rewrite0Identity
6   | Rewrite0 of { source: generator; target: generator; label: frame }
7   | RewriteN of { cones: cone list }
8
9 and cone = {
10  index: int;
11  source: cospan list;
12  target: cospan;
13  slices: rewrite list;
14 }
15
16 and cospan = { forward: rewrite; backward: rewrite }
17
18 type diagram =
19   | Diagram0 of generator
20   | DiagramN of { source: diagram; cospans: cospan list }

```

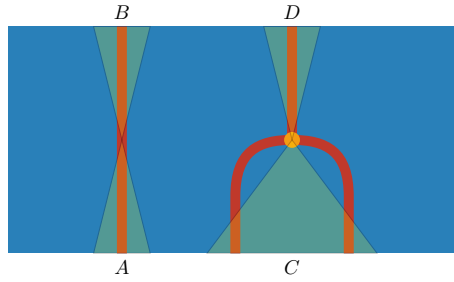
■ **Figure 7** The core data structures of `homotopy.io`.

### 4.1 Diagrams

The core data structure of `homotopy.io` is a recursive encoding for diagrams that is derived from the zigzag construction. A 0-diagram has trivial shape (a single point), and type is essentially the labelling which is an assignment of name to colour, which we call a *generator*. An  $(n + 1)$ -diagram is determined by an alternating sequence of singular and regular heights, which are  $n$ -diagrams, together with information on how the heights fit together:

$$r_0 \xrightarrow{f_0} s_1 \xleftarrow{b_0} r_1 \xrightarrow{f_1} s_1 \xleftarrow{b_1} r_1 \xrightarrow{f_2} s_2 \xleftarrow{b_2} r_2.$$



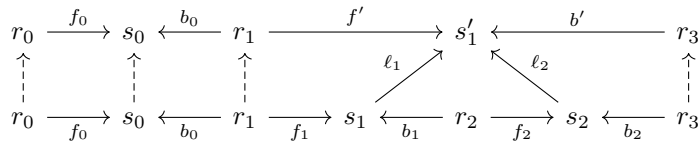


■ **Figure 8** A zigzag with four cones.  $A$  and  $B$  are *identity* cones that do not change the diagram, omitted in the sparse representation.

Each singular height  $s_i$  is equipped with rewrites  $f_i: r_i \rightarrow s_i$  and  $b_i: r_{i+1} \rightarrow s_i$  that encode the difference between  $s_i$  and its neighbouring regular heights  $r_i$  and  $r_{i+1}$ . We call the pair  $(f_i, b_i)$  a *cospan*, and describe the height of a diagram as the number of cospans (equally, singular heights) it contains. There is always one more regular height than singular heights. Instead of storing all the heights of a diagram, we only keep around the first regular height  $r_0$  and reconstruct the other heights when needed by applying the rewrites backwards and forwards.

### 4.2 Rewrites

Rewrites between diagrams are also encoded as a recursive data structure. A rewrite of 0-diagrams (0-rewrite) is either the identity rewrite or a rewrite between the underlying generators; in the latter case, it is also equipped with a *frame* which represents the directionality of the rewrite in some space associated to the ambient diagram. A  $(n + 1)$ -rewrite  $x \rightarrow y$  modifies the sequence of cospans in  $x$  by removing subsequences and replacing them with individual cospans. Each such modification is called a *cone* (see Figure 8). The following diagram, which encodes the bottom half of Figure 8<sup>1</sup>, is an example of a rewrite with a single cone that replaces the cospans  $(f_1, b_1)$  and  $(f_2, b_2)$  with the cospan  $(f', b')$ :



A cone in a rewrite of  $(n + 1)$ -diagrams also contains rewrites between the  $n$ -dimensional singular heights. In the example above these are the rewrites  $\ell_1: s_1 \rightarrow s'_1$  and  $\ell_2: s_2 \rightarrow s'_1$ . A rewrite does not store information for the parts of the diagrams that do not change. Since in practice most adjacent heights only differ in small parts, this sparse encoding leads to significant space efficiency.

The core data structures admit a series of auxiliary algorithms:

1. Given an  $n$ -diagram  $x$  we can apply a rewrite  $x \xrightarrow{R} y$  forwards to obtain  $y$ . Similarly, given an  $n$ -diagram  $y$  and a rewrite  $x \xrightarrow{R} y$  we can apply  $R$  backwards to  $y$ .

<sup>1</sup> This can be seen by overlaying the diagram along the bottom of Figure 8. The identity map  $s_0 \dashrightarrow s_0$  corresponds to cone A, and the right component consisting of  $s_1, r_2, s_2, s'_1$  and maps between them corresponds to cone C.

2. We can compute the heights of an  $(n + 1)$ -diagram at any singular or regular height. In particular, we can obtain the source and target of a diagram.
3. When  $x \xrightarrow{R} y$  and  $y \xrightarrow{R'} z$  are rewrites, we can compute the composite rewrite  $x \xrightarrow{R' \circ R} z$  while preserving sparsity.
4. For two  $n$ -diagrams  $s$  and  $t$  whose sources and targets agree, we can create an  $(n + 1)$ -dimensional diagram which represents an  $(n + 1)$ -morphism between  $s$  and  $t$ .
5. Given an  $n$ -diagram  $x$  and  $k \leq n$  we can compute the generators that would be visible in the projection of  $x$  to  $k$  dimensions.
6. For an  $n$ -diagram  $x$  and  $k \leq n$  we can produce a graph that is a dense encoding of the projection of  $x$  to  $k$  dimensions. The simplicial complex obtained as the flag complex from this graph has the geometry of the rendered diagram. The layout algorithm then assigns coordinates to the vertices (see Section 5.5).
7. Given a pair of diagrams  $x, y$  we can search for copies of  $y$  that are embedded into  $x$  and intersect a line through the projection. This allows us to find opportunities to rewrite a diagram, which itself can be realised as a higher cell.

## 5 Key Algorithms

Many of our key algorithms described in this section first compute the shape of the result, usually by recursion over diagram dimension, and then complete the type information. These data are encoded by directed graphs, for which we can utilise graph algorithms to implement our operations.

### 5.1 Collapse

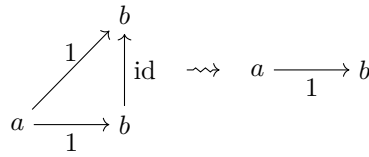
A string diagram has geometric properties (e.g. the length of a wire) which are not intended to be meaningful, as its meaning is captured entirely by topological properties (e.g. connectivity of a wire). Topology yields a natural compatibility with composition in the string diagram calculus: if two wires are placed in sequence, then the resulting string diagram is merely a longer wire, and therefore it looks topologically the same and represents the same mathematical content. However, both the tool and the theory are based on *combinatorial* encodings of string diagrams, and moreover there is a distinct encoding for the diagram of two wires in sequence versus one wire.

We describe the *collapse* of a diagram as a combinatorial representation of these topological invariants, essentially as a directed graph, which gives a normal form for the fully-exploded graph obtained from an  $n$ -diagram in a frame-preserving way. In more detail, given such a graph, every node has a neighbourhood which determines a set of outgoing frames and incoming frames (framing data); now, consider the largest equivalence relation on nodes determined by  $x \sim y$  when  $x$  is adjacent to  $y$  by an identity 0-rewrite, and  $x$  and  $y$  admit equal framing data; collapse is the quotient graph induced by  $\sim$ . Explicitly, we compute this by treating the graph as a simplicial complex and checking each 1-simplex (edge) which is an identity 0-rewrite to see if it is collapsible, whereby identifying both 0-simplex faces (endpoints) respects  $\sim$ , by checking all 2-simplices (triangles) for which that 1-simplex is a face. A simple example of this is given in Figure 9.

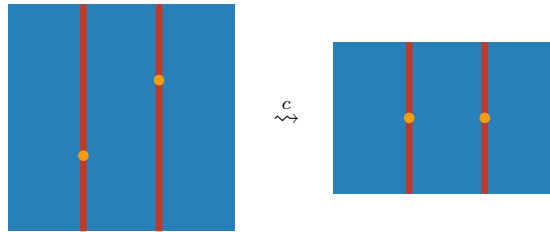
Collapse is used to compare when two diagrams differ as encodings but have the same topological data: many different diagrams may have the same collapse, for instance degenerating a diagram along any part<sup>2</sup> does not change its collapse, but two diagrams with distinct

---

<sup>2</sup> This corresponds to composition with a weak unit in the  $n$ -categorical model; combinatorially, this represents the insertion of redundant data.



■ **Figure 9** The edge  $b \xrightarrow{\text{id}} b$  is collapsible, so the graph can be simplified in a frame-preserving fashion.



■ **Figure 10** A contraction of two beads on two singular heights.

collapses are necessarily distinct. These degeneracies arise naturally in the course of the other operations described in this section, but can be often eliminated without rendering the diagram malformed. Collapse is a crucial component of typechecking, analogously to how the standard technique of deciding equality in a term calculus often amounts to computing normal forms for each term. A detailed account is given in [21].

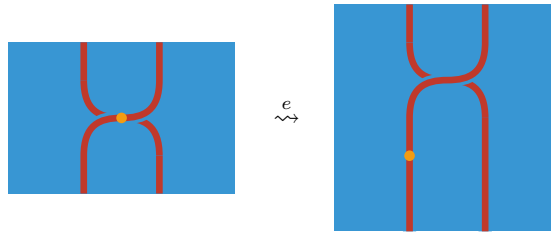
## 5.2 Contraction

*Contraction* is the process of shortening a diagram along a homotopy equivalence which reduces the height of a diagram locally, producing a contraction rewrite  $D \xrightarrow{c} C$ , which may or may not exist depending on  $D$ . An example of this is given by Figure 10, which combines both singular heights on the left into one on the right. The result of the contraction  $C$  always has a singular height of one. Informally, as a homotopy, the contraction rewrite  $c$  is *the* unique canonical way to shorten the diagram without making any arbitrary choices, and this is mathematically captured by its description as the computation of a categorical colimit.

Contraction works recursively on diagram dimension, with the 0-dimensional base case obtained via collapse, as in Section 5.1, and then ensuring that the maximal elements of the poset induced by the reachability relation on the resulting directed graph are compatible [21]. The higher-dimensional recursive case has been described theoretically by Reutter and Vicary [33]; it works by first determining a “ $\Delta$ -colimit”, which determines the shape that the result necessarily possesses, and then if that exists it then attempts to find a compatible labelling of each stratum to complete the typing information by a divide-and-conquer strategy.

In the base case, compatibility is the uniqueness of the labelling of the maximal element of the poset, combined with the condition that the framing data on each maximal element is identical. In the recursive case, for an  $(n + 1)$ -rewrite  $D \xrightarrow{c} C$ , the algorithm is as follows:

1.  $D$  is associated to a directed graph  $G$  whose nodes represent regular and singular heights of  $D$ , weighted by  $n$ -diagrams, and whose edges form cospans between regular heights and are weighted by  $n$ -rewrites;
2. this directed graph is then exploded, obtaining a larger directed graph  $E$  whose nodes are weighted by  $(n - 1)$ -diagrams, and whose edges are weighted by  $(n - 1)$ -rewrites, by replacing each node of  $G$  with a directed graph, as in the previous step, and each edge by a collection of connecting  $(n - 1)$ -rewrites;



■ **Figure 11** An anticontraction move.

3.  $E$  induces a subgraph  $\Delta_E$  which represents only the shape data of  $D$  consisting of only the singular-singular nodes; on  $\Delta_E$ , we obtain its condensation graph with respect to strongly-connected components<sup>3</sup>; if the edge relation of the condensation graph does not correspond to a linear order on nodes, then fail, otherwise, this determines:
  - a. subproblems determined by the subgraph of  $E$  induced by reachability for the nodes of a particular strongly-connected component of  $\Delta_E$ ;
  - b. a linear ordering of subproblems, which will determine how their solutions should be combined.
4. each subproblem represents an independent fragment of  $D$  which spans its entire height, and is solved recursively (step 2); its solution is a singular-singular height of  $C$ , which necessitates the linear ordering of subproblems; the rest of  $C$  and  $c$  are constructed from the remaining data;
5. the result is then typechecked, and if it fails, then the algorithm fails.

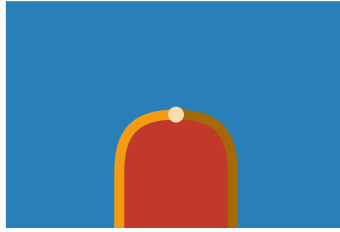
There are also auxiliary algorithms which propagate a contraction within some height of a larger diagram. We refer the interested reader to [33].

### 5.3 Expansion

*Expansion* is the dual of contraction, first described by Reutter and Vicary [33], and later refined by Tataru and Vicary [40]. It takes a diagram  $D$  and produces an expanded diagram  $E$ , such that  $E$  in fact contracts to give  $D$  via some rewrite  $E \xrightarrow{c} D$ . In this sense, expansion is a partial converse to contraction.

It is defined inductively, similarly to contraction. In the base case, it performs an interchanger move that separates two singular levels (the reverse of Figure 10). In the recursive case, expansion attempts to propagate an expansion of a sub-diagram to the diagram itself, yielding a diagram which will contract to the original. Since contraction is computed by a colimit process, this requires an algorithm that can reverse the ordinary colimit process, a procedure that we call *anticontraction* [40]. We illustrate this with the recursive expansion example in Figure 11, where a vertex is moved out of a singular height. Note that the expanded diagram on the right indeed contracts to give the original diagram on the left.

<sup>3</sup> This graph has a node for each strongly-connected component, and each edge represents reachability for strongly-connected components.



■ **Figure 12** The cancellation 2-cell  $f^{-1} \circ f \rightarrow 1_x$  generated by an invertible 1-cell  $f: x \rightarrow y$ .

## 5.4 Typechecking

Typechecking is the process of checking the validity of a diagram with respect to a given signature. Note that despite the name, in our context, “typechecking” does not refer to a type-theoretic procedure determined by inference rules. The underlying theory gives a clear perspective on this process [12, 17]: break the diagram into atomic “pieces”, and ensure that each piece collapses to the canonical neighbourhood of the respective generator. This tells us that the neighbourhood of each point in the diagram is fully described by the signature.

Collapse plays a crucial role here, because it ensures that higher-dimensional coherences of invertible generators also typecheck. For example, if  $f: x \rightarrow y$  is an invertible 1-cell, we can generate a 2-cell  $f^{-1} \circ f \rightarrow 1_x$  known as the *counit* (see Figure 12); this is well-typed because it collapses to the canonical neighbourhood of  $f$ .

## 5.5 Layout

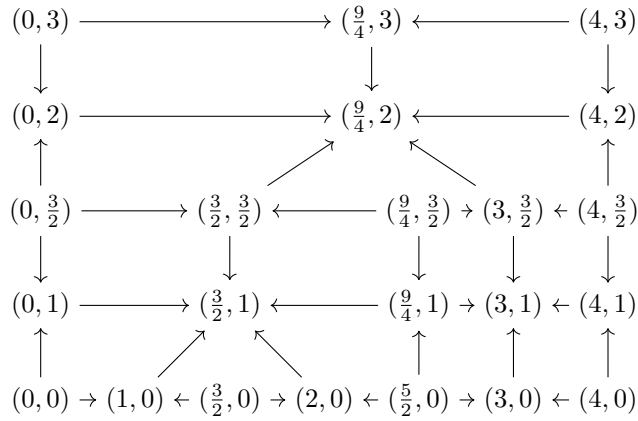
The layout algorithm is used to assign coordinates to every point in an  $n$ -diagram, and is a crucial component of the rendering pipeline described in Section 6.2, enabling rendering in 2D, 3D and 4D. We use a categorical construction based on factorisation systems and colimits to extract a set of linear constraints from the total order data of the diagram [39]. These constraints encode the necessary conditions for a layout to be well-defined. For example, in 2D, this would include the information that one wire is to the left of another wire; or, in 3D, that one surface is in front of another. We further impose aesthetic constraints, e.g. that wires and surfaces should be centred. Finally, these constraints are passed to the linear solver HiGHS [22] to find a layout that satisfies all constraints. An example layout for a 2-diagram is given in Figure 13: each node receives a cartesian coordinate  $(x, y)$  from the solver, which is used to determine where it should be rendered on-screen.

## 6 Implementation

Here, we describe further implementation details: the memoisation techniques used in our core data structures and aspects of the rendering pipeline. Whilst independent of the tool’s mathematical foundations, we have found many of the details here essential, particularly in enabling the tool to operate in a resource-constrained environment – i.e. the browser.

### 6.1 Memoisation

Our data structures for `Diagram` and `Rewrite` are immutable. Operations that would modify an object instead create a new one with the modifications applied. This allows us to apply a technique known as *hash consing* [14]: whenever a new `Diagram` or `Rewrite` is created, we



■ **Figure 13** The layout of a 2-diagram.

check in a hash table if a logically equal instance is already present and reuse it if possible. Hash consing enforces the invariant that any two instances that are logically equal must become physically equal – i.e. have the same representation and location in memory. To test for equality we may therefore perform a simple pointer comparison instead of traversing the entire deeply nested structure; similarly, we cache the hash value of every `Diagram` and `Rewrite` to avoid deep traversals. We incur a performance cost due to the hash table lookup involved whenever a new instance is created. However, we observe significant improvements in performance and memory usage overall. In practice, we have found that even under the sparse encoding of Section 4, directly representing diagrams remains highly redundant: applying hash consing makes memory usage almost negligible, even for large diagrams.

The algorithms from Section 5 are largely recursive over the structure of a diagram and perform many repeated recursive calls on logically equal substructures. We therefore memoise the results of recursive calls. Since our deduplicated representation allows for very fast equality checks and caches hash values, lookups in the memoisation table are comparatively cheap.

Our data structures are reference counted, so we know when the last reference to an instance goes out of scope and it is safe to remove it from the hash table. The recursive algorithms operating on diagrams temporarily materialise the structure that is implicit in the sparse encoding, leading to the same `Diagram` or `Rewrite` being created and destroyed many times during the execution of the algorithm. An eager approach to maintaining the deduplication hash table therefore leads to unnecessary churn. We therefore delay the removal of dead objects from the hash table to a batched garbage collection step that walks the table and removes instances that have no remaining references.

## 6.2 Rendering Pipeline

One of the most important parts of the implementation is the rendering pipeline, which allows for visualising  $n$ -diagrams in up to four dimensions. It consists of three components. First, we have the *layout* algorithm described in Section 5.5 that assigns real coordinates to every point of an  $n$ -diagram. Second, there is the *mesh generation*, which takes an  $n$ -diagram and computes a cubical mesh – i.e. a subdivision of the  $n$ -diagram into *abstract*  $k$ -dimensional cubes for  $k \leq n$ . Together, these data are called a *geometry* which is a collection of  $k$ -dimensional cuboids in the Euclidean space  $\mathbb{R}^n$ . Finally, we have a *subdivision*

procedure which takes a geometry and makes it smoother by recursively subdividing each cube into smaller cubes with the positions of the new vertices calculated by interpolating the positions of the old vertices.

The subdivided geometry can then be sent to a number of rendering engines. Amongst these, the most important are the SVG renderer, rendering 2D diagrams for all the user’s primary interaction, and the WebGL renderer, rendering 3D and 4D diagrams (with the 4D diagrams appearing as *smooth* animations of 3D diagrams). We also support a TikZ renderer, which can be used to generate high quality string diagrams for use in papers such as in this article.

In the 4-dimensional case, we obtain the animation by intersecting the geometry with an axis-aligned hyper-plane. This computation is easily implemented via a compute shader. However, to maximise *homotopy.io*’s compatibility with contemporary browsers, our shading pipeline is implemented in WebGL2, precluding the use of compute shaders.

This situation presents a challenge, as computing intersection geometries offline for large diagrams is prohibitively expensive in terms of memory consumption. Thus, we are forced perform the slicing in real-time, using a subtle rendering trick:

1. each cube is broken down into a disjoint collection of covering simplices (Figure 14, left);
2. each simplex is further decomposed into sub-simplices, each with one face whose vertices have equal  $w$ -components<sup>4</sup>,
3. of the six edges of each simplex, the three with non-zero  $w$ -components are converted to three GL vertices, and the simplex a corresponding polygon (Figure 14, right);
4. each GL vertex is supplied with the start and end coordinates of the corresponding edge;
5. the vertex shader interpolates between these points given a global  $w$ -coordinate, passed in as a uniform, dropping the vertex whenever  $w$  is outside the bounds of the edge.

The effect is that the animation is broken down into a collection of “birth” and “death” events for small polygons, which can be rendered straightforwardly, as if it were a 3D mesh. Animated wires in 4-dimensions are produced through a similar process, but the “pipe”-like appearance is produced via a deferred post-processing effect, to minimise the cost of storing geometry for the cylinders.

We additionally compute normals for intersection geometries in real-time via a barycentric interpolation across a volume-weighted average taken at each vertex. This is directly analogous to the calculation of normals in traditional 3D pipelines. However, much care is required when it comes to calculating and preserving the orientations of the elements of the mesh.

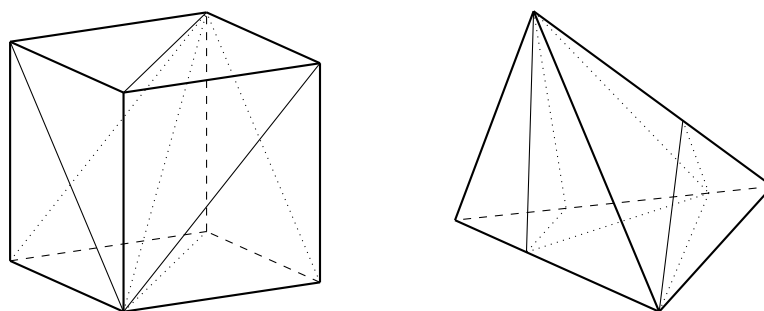
## 7 Future work

The tool is under active development, both mathematically as an end itself and as a piece of research software as a means to facilitate the development of higher-categorical mathematics. We sketch some ideas for future improvements to the tool:

**Functor boxes** It would be both interesting and practical to consider what a functor would mean in this context; graphically, it should allow for “boxing” pieces of diagrams, allowing those components to be treated as atomic; this would present the user with more flexibility in representing proofs.

---

<sup>4</sup>  $w$ -component refers to the fourth coordinate of the homogeneous coordinate system commonly used in computer graphics, and in our tool via WebGL.



■ **Figure 14** The decomposition of a 3-cube into six 3-simplices (left) and a 3-simplex into  $w$ -aligned sub-simplices (right).

**Dualisability support** The next logical step after the introduction of coherent inverses would be the introduction of coherent duals, which would allow for a mathematical exploration of important problems in higher algebra, such as the tangle hypothesis [4], and knot theory.

**Interoperability with type theory** Some arguments of logical flavour are more easily expressed type-theoretically, e.g. in homotopy type theory or CaTT [15], while more graphical arguments are better suited to homotopy.io; ideally, we could make these tools talk to each other to support bimodal reasoning.

**Linearity** Support for higher linear categories would make the tool amenable to directly performing calculations in topological quantum field theory.

**Gallery** An arXiv-like gallery can be presented to showcase published proofs, better enabling proof discovery.

**Cubical version** More general notions of string diagrams for cubical  $n$ -categories have been considered [32] and it would be an interesting question to extend homotopy.io to cubical shapes.

---

## References

- 1 Constructing a braiding in homotopy.io. URL: <https://homotopy.io/braiding-example>.
- 2 Nicolás Andruskiewitsch and Walter Ferrer Santos. The Beginnings of the Theory of Hopf Algebras. *Acta Applicandae Mathematicae*, 108(1):3–17, October 2009. doi:10.1007/s10440-008-9393-1.
- 3 Michael F Atiyah. Topological quantum field theory. *Publications Mathématiques de l’IHÉS*, 68:175–186, 1988.
- 4 John C Baez and James Dolan. Higher-dimensional algebra and topological quantum field theory. *Journal of mathematical physics*, 36(11):6073–6105, 1995.
- 5 John C Baez and Martin Neuchl. Higher dimensional algebra: I. braided monoidal 2-categories. *Advances in Mathematics*, 121(2):196–244, 1996.
- 6 Krzysztof Bar, Aleks Kissinger, and Jamie Vicary. Globular: an online proof assistant for higher-dimensional rewriting. *Logical Methods in Computer Science*, 14, 2018. doi:10.23638/LMCS-14(1:8)2018.
- 7 John W Barrett, Catherine Meusburger, and Gregor Schaumann. Gray categories with duals and their diagrams. *arXiv preprint*, 2012. arXiv:1211.0529.
- 8 Yuri Bespalov and Bernhard Drabant. Hopf (bi-)modules and crossed modules in braided monoidal categories. *Journal of Pure and Applied Algebra*, 123(1-3):105–129, 1998.
- 9 Simon Burton. String diagrams for higher mathematics with wiggle.py, 2023. Talk at SYCO 11. URL: [https://arrowtheory.com/wiggle\\_demo.pdf](https://arrowtheory.com/wiggle_demo.pdf).



- 10 Giovanni de Felice, Alexis Toumi, and Bob Coecke. Discopy: Monoidal categories in python. *arXiv preprint*, 2020. [arXiv:2005.02975](https://arxiv.org/abs/2005.02975).
- 11 Antonin Delpeuch. A complete language for faceted dataflow programs. *arXiv preprint*, 2019. [arXiv:1906.05937](https://arxiv.org/abs/1906.05937).
- 12 Christoph Dorn. Associative  $n$ -categories. *arXiv preprint*, 2018. [arXiv:1812.10586](https://arxiv.org/abs/1812.10586).
- 13 Christoph Dorn and Christopher L Douglas. Manifold diagrams and tame tangles. *arXiv preprint arXiv:2208.13758*, 2022.
- 14 Jean-Christophe Filliâtre and Sylvain Conchon. Type-safe modular hash-consing. In *Proceedings of the 2006 Workshop on ML*, pages 12–19, 2006. doi:10.1145/1159876.1159880.
- 15 Eric Finster and Samuel Mimram. A type-theoretical definition of weak  $\omega$ -categories. In *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–12. IEEE, 2017. doi:10.1109/LICS.2017.8005124.
- 16 Amar Hadzahasanovic and Diana Kessler. Data structures for topologically sound higher-dimensional diagram rewriting. *arXiv preprint*, 2022. [arXiv:2209.09509](https://arxiv.org/abs/2209.09509).
- 17 Lukas Heidemann, David Reutter, and Jamie Vicary. Zigzag normalisation for associative  $n$ -categories. In *Proceedings of the 37th Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 1–13, 2022. doi:10.1145/3531130.3533352.
- 18 Martin Hofmann and Thomas Streicher. The groupoid interpretation of type theory. *Twenty-five years of constructive type theory (Venice, 1995)*, 36:83–111, 1998.
- 19 Heinz Hopf. Über Die Topologie der Gruppen-Mannigfaltigkeiten und Ihre Verallgemeinerungen. *Annals of Mathematics*, 42(1):22–52, 1941. Publisher: Annals of Mathematics. doi:10.2307/1968985.
- 20 Nick Hu, Alex Rice, and Calin Tataru. sd-visualiser, 2024. URL: <https://github.com/sd-visualiser/sd-visualiser>.
- 21 Nick Hu, Calin Tataru, and Jamie Vicary. Coherent invertibility in associative  $n$ -categories. In preparation, 2024.
- 22 Qi Huangfu and JA Julian Hall. Parallelizing the dual revised simplex method. *Mathematical Programming Computation*, 10(1):119–142, 2018. doi:10.1007/s12532-017-0130-5.
- 23 André Joyal and Ross Street. The geometry of tensor calculus, i. *Advances in mathematics*, 88(1):55–112, 1991.
- 24 Aleks Kissinger and Vladimir Zamdzhiev. Quantomatic: A proof assistant for diagrammatic reasoning. In *Automated Deduction-CADE-25: 25th International Conference on Automated Deduction, Berlin, Germany, August 1-7, 2015, Proceedings 25*, pages 326–336. Springer, 2015. doi:10.1007/978-3-319-21401-6\_22.
- 25 Richard Gustavus Larson and Moss Eisenberg Sweedler. An Associative Orthogonal Bilinear Form for Hopf Algebras. *American Journal of Mathematics*, 91(1):75, January 1969. doi:10.2307/2373270.
- 26 Tom Leinster. *Higher operads, higher categories*. Number 298 in London Mathematical Society Lecture Note Series. Cambridge University Press, 2004.
- 27 Jacob Lurie. *Higher topos theory*. Princeton University Press, 2009.
- 28 V. Lyubashenko. Modular transformations for tensor categories. *Journal of Pure and Applied Algebra*, 98(3):279–327, February 1995. doi:10.1016/0022-4049(94)00045-K.
- 29 S. Majid. Algebras and Hopf Algebras in Braided Categories, September 1995. [arXiv:q-alg/9509023](https://arxiv.org/abs/q-alg/9509023). [arXiv:q-alg/9509023](https://arxiv.org/abs/q-alg/9509023).
- 30 Shahn Majid. Braided groups and algebraic quantum field theories. *letters in mathematical physics*, 22:167–175, 1991.
- 31 John W. Milnor and John C. Moore. On the Structure of Hopf Algebras. *The Annals of Mathematics*, 81(2):211, March 1965. doi:10.2307/1970615.
- 32 David Jaz Myers. String diagrams for double categories and equipments. *arXiv preprint*, 2016. [arXiv:1612.02762](https://arxiv.org/abs/1612.02762).

- 33 David Reutter and Jamie Vicary. High-level methods for homotopy construction in associative  $n$ -categories. In *2019 34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–13. IEEE, 2019. doi:10.1109/LICS.2019.8785895.
- 34 Emily Riehl and Dominic Verity. *Elements of  $\infty$ -Category Theory*, volume 194. Cambridge University Press, 2022.
- 35 Chiara Sarti and Jamie Vicary. Posetal diagrams for logically-structured semistrict higher categories. *arXiv preprint*, 2023. arXiv:2305.11637.
- 36 Christopher John Schommer-Pries. *The classification of two-dimensional extended topological field theories*. University of California, Berkeley, 2009.
- 37 Pawel Sobocinski, Paul W Wilson, and Fabio Zanasi. Cartographer: A tool for string diagrammatic reasoning (tool paper). In *8th Conference on Algebra and Coalgebra in Computer Science (CALCO 2019)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019.
- 38 Moss E. Sweedler. *Hopf algebras*. Mathematics lecture note series. Benjamin, New York, NY, 1969.
- 39 Calin Tataru and Jamie Vicary. A layout algorithm for higher-dimensional string diagrams. *arXiv preprint*, 2023. arXiv:2305.06938.
- 40 Calin Tataru and Jamie Vicary. The theory and applications of anticolimits. *arXiv preprint*, 2024. arXiv:2401.17076.
- 41 The Univalent Foundations Program. *Homotopy Type Theory: Univalent Foundations of Mathematics*. <https://homotopytypetheory.org/book>, Institute for Advanced Study, 2013.
- 42 Shouchuan Zhang. Braided hopf algebras. *arXiv preprint*, 2005. arXiv:math/0511251.

## **A** Case Study: Hopf Algebras and Hopf Modules

In this appendix, we will detail an extended case study to highlight how homotopy.io may be practically employed to formalise mathematical results of substantial character. The basis we will build on for this case study will be the discussion in Section 3, showing that the Eckmann-Hilton argument could be used to formalise braids in the tool. This will allow us to define Hopf algebras and Hopf modules in braided monoidal categories, and build towards a proof of the *fundamental theorem of Hopf modules*, asserting the freeness of Hopf modules when certain idempotents split. Since this formalisation will involve rich algebraic gadgets, long proofs and universal properties, we will build it up gradually. The final formalisation, consisting of the full proofs, is available at <https://beta.homotopy.io/p/2402.00006> so that the reader can inspect the proofs in detail and visualise them in 3D and 4D.

Hopf algebras originated in work by Hopf [19] in topology, developed on and popularised by Milnor and Moore [31] (see [2] for more detail). They have served to generalise results from group theory to objects of study in topology, knot theory, algebraic geometry, combinatorics and quantum theory [38]. Our discussion here is informed by the original proof of the fundamental theorem of Hopf modules for Hopf algebras over principal ideal domains is given in [25], its generalisation to braided monoidal categories in [28] and the alternative proof using Karoubi completions given in [8]. We also use the diagrammatic calculus for Hopf modules introduced by Majid in [30], which has been used to give a string diagrammatic translation of this result in [42], though the reader should be warned that the proof contains flaws.

We start by first defining a bialgebra  $H$  in a braided monoidal category. Recall from Section 3 that in order to obtain the braiding, we will start by adding a single 0-cell  $x$  to our signature, and then add our 2-cell  $H$  by specifying  $1_x$  as source and target. Hence,  $H$  is an object in our braided monoidal category. Moreover, we need to add four 3-cells: a unit  $\eta: 1 \rightarrow H$ , a multiplication  $\mu: H \otimes H \rightarrow H$ , a counit  $\varepsilon: H \rightarrow 1$ , and a comultiplication  $\delta: H \rightarrow H \otimes H$ , which are depicted in Figure 15.

For  $H$  to be an algebra,  $\eta$  and  $\mu$  must satisfy associativity and unitality laws, which are witnessed by the 4-cells depicted in Figure 16. We add these to the signature and mark them as invertible. For  $H$  to be a coalgebra, it will have to satisfy dual laws of coassociativity and counitality. Finally, for  $H$  to be a bialgebra, the algebra and coalgebra structure have to interact: units must propagate through comultiplications, counits through multiplication, counit and units must be one-sided inverses, and multiplication and comultiplication have to interact via the braiding. These interaction laws are specified as in Figure 17.

Thus,  $H$  is now a bialgebra. In order for it to be a Hopf algebra, we must additionally equip it with an antipode map  $\sigma: H \rightarrow H$ . This map needs to satisfy the two cancellation laws given in Figure 18, which make it an inverse to the identity map on  $H$  under the convolution product. These two laws are enough to establish the following lemmas on the interactions between the antipode and the bialgebra structure.

► **Lemma 1.** *For a Hopf algebra  $H$  in a braided monoidal category, we have  $\sigma \circ \eta = \eta$ .*

► **Lemma 2** ([29, Lemma 2.3]). *For a Hopf algebra  $H$  in a braided monoidal category, we have  $(\sigma \otimes \sigma) \circ \gamma_{H,H} \circ \delta = \delta \circ \sigma$ , where  $\gamma_{H,H}: H \otimes H \rightarrow H \otimes H$  is the braiding.*

**Proof.** The essential steps of the graphical proof are given in Figure 25. We start by constructing the left-hand side of the equation, as it involves a braid and thus is more complex than the right-hand side. This gradient of complexity helps us to use the contraction procedure effectively.

The first step of the proof is to introduce units at the top of the diagram, and a counit at the bottom. We then braid the right-most unit past the antipode on the left, multiply the units and use them with the counit to introduce an antipode on the left. Above this antipode, we then use the multiplication/comultiplication axiom. Here we note that for proof to go through, the orientation of this braid, the braid in the starting diagram and the braid in the multiplication/comultiplication interaction axioms must match.

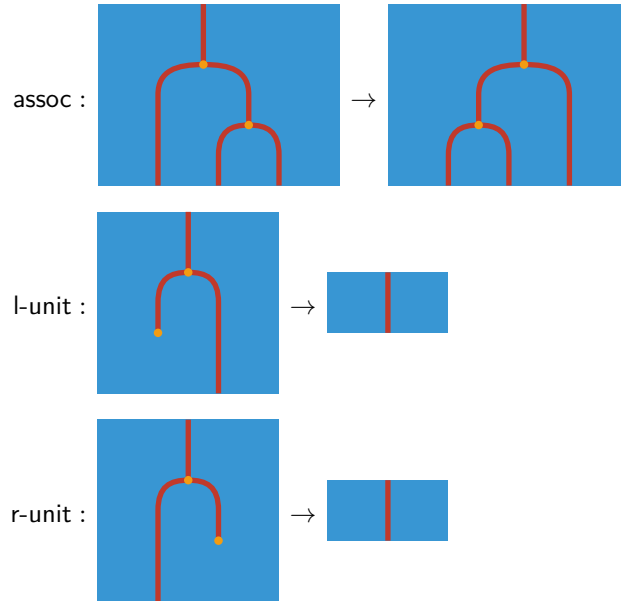
A sequence of (co)associativity and braid naturality moves follows, until we can isolate the right-most antipode into a bubble to eliminate it into a unit/counit pair, which we use to remove a pair of (co)multiplications. We then move to use the (co)associativity laws to isolate the right-most antipode into a bubble, and proceed to simplify the remaining diagram down to our target diagram by cancelling the (co)units. ◀

At this point, our signature contains our Hopf algebra  $H$  and two results about its properties, namely Lemma 1 and Lemma 2. Here we must note that the axioms for a Hopf algebra are self-dual, hence whenever we establish a result its dual statement, obtained by swapping unit and counit, multiplication and comultiplication and vice versa, will also hold. For instance, the dual of Lemma 1 would show that  $\varepsilon \circ \sigma = \varepsilon$ .

We will now add another object  $M$  with the same boundary as  $H$ , which is to be a (left)  $H$ -Hopf module. This means  $M$  is equipped with two maps, a (left) action  $\alpha: H \otimes M \rightarrow M$  and a (left) coaction  $\phi: M \rightarrow H \otimes M$ , which are compatible with the structure of  $H$ . Thus, we add invertible cells as in Figure 20 to impose interaction between the multiplication  $\mu$  and the action  $\alpha$ , between the unit  $\eta$  and  $\alpha$ , their duals for  $\delta, \varepsilon$  and  $\phi$ , and finally the interaction between the action/coaction pair.



■ **Figure 15** The 3-cells  $\eta$ ,  $\mu$ ,  $\varepsilon$ , and  $\delta$  (left to right) in the bialgebra signature.



■ **Figure 16** The associativity and unitality laws for the algebra  $(H, \eta, \mu)$ .

We now have an  $H$ -Hopf module  $M$  which we have manually added to the signature. The structure of the Hopf algebra  $H$  gives us another way to obtain  $H$ -Hopf modules: given any object  $B$  in our braided monoidal category, we may equip  $H \otimes B$  with the structure of an  $H$ -Hopf module by taking the action to be  $\mu \otimes 1_B: H \otimes H \otimes B \rightarrow H \otimes B$  and the coaction to be  $\delta \otimes 1_B: H \otimes B \rightarrow H \otimes H \otimes B$ . We call Hopf modules of this form *free*.

With the action and coaction of  $M$  and the antipode of  $H$ , we may define the endomorphism  $\nu := \alpha \circ (\sigma \otimes 1_M) \circ \phi: M \rightarrow M$  depicted in Figure 21. It turns out that this map is idempotent.

► **Lemma 3** ([8, Prop. 3.2.1]). *The map  $\nu: M \rightarrow M$  is idempotent – i.e.  $\nu \circ \nu = \nu$ .*

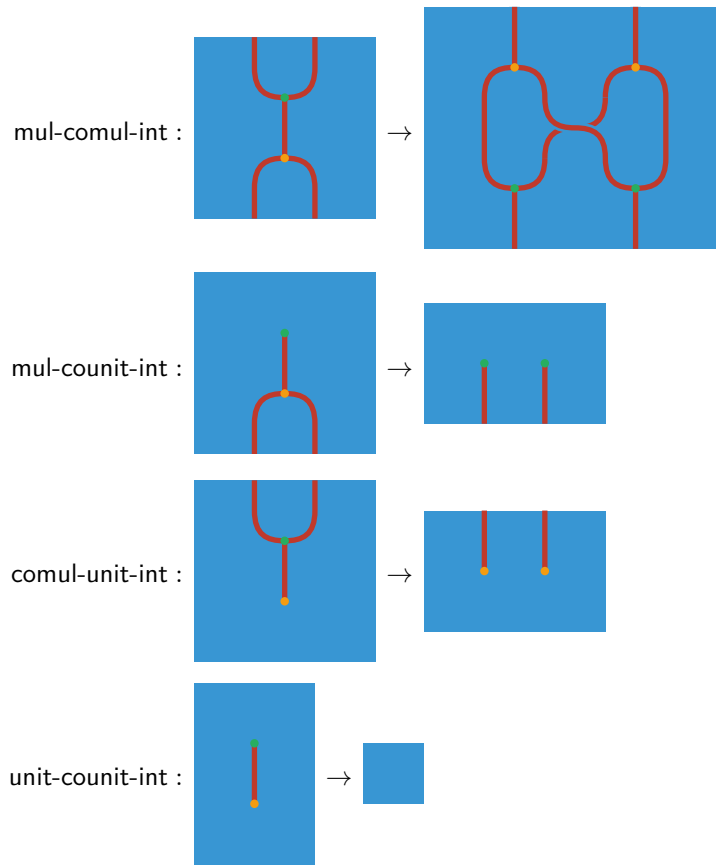
**Proof.** We first use Lemma 2 to establish that  $\phi \circ \nu = (\eta \otimes 1_M) \circ \nu$ . This is given in Figure 22, where the key step is to use the action/coaction interaction axiom in order to apply Lemma 3, and then simplify down the diagram. With this at hand, idempotency follows from Lemma 1 and routine simplifications. ◀

With Lemma 3 at hand, we assume now that we have a splitting for  $\nu$  – i.e. an object  $M^H$  and maps  $\iota: M^H \rightarrow M$ ,  $\nu: M \rightarrow M^H$  such that  $\nu \circ \iota = 1_{M^H}$  and  $\iota \circ \nu = \nu$ . It turns out that this splitting enjoys an additional universal property.

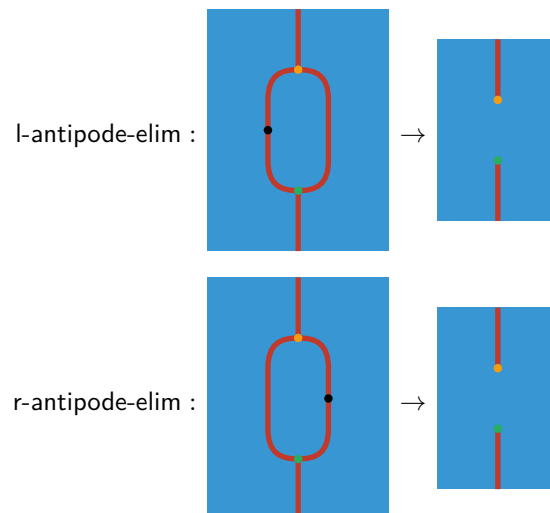
► **Lemma 4** ([8, Prop. 3.2.1]). *The map  $\iota: M^H \rightarrow M$  is the equaliser of  $\phi$  and  $\eta \otimes 1_M$ .*

**Proof.** We add an object  $P$  and a map  $\chi: P \rightarrow M$  such that  $\phi \circ \chi = \eta \otimes \chi$  to the signature. We then show that  $\chi = \iota \circ \nu \circ \chi$  by applying Lemma 1. But this means  $\chi = \iota \circ (\nu \circ \chi)$  and since  $\iota$  is monic this factorisation must be unique. ◀

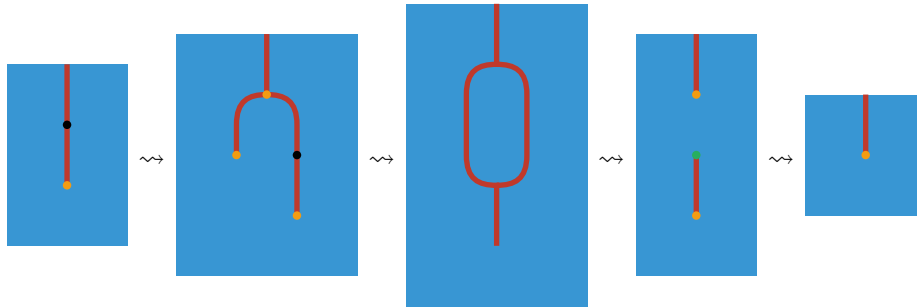
Thus,  $M^H$  enjoys a universal property which trivialises the coaction, we call it *object of coinvariants* of  $M$ . Note that by dualising Lemma 4 and its dependencies, we may also prove



■ **Figure 17** The bialgebra interaction laws for  $(H, \eta, \mu, \varepsilon, \delta)$ .



■ **Figure 18** The antipode cancellation laws for  $\sigma: H \rightarrow H$ .



■ **Figure 19** The proof of Lemma 1.

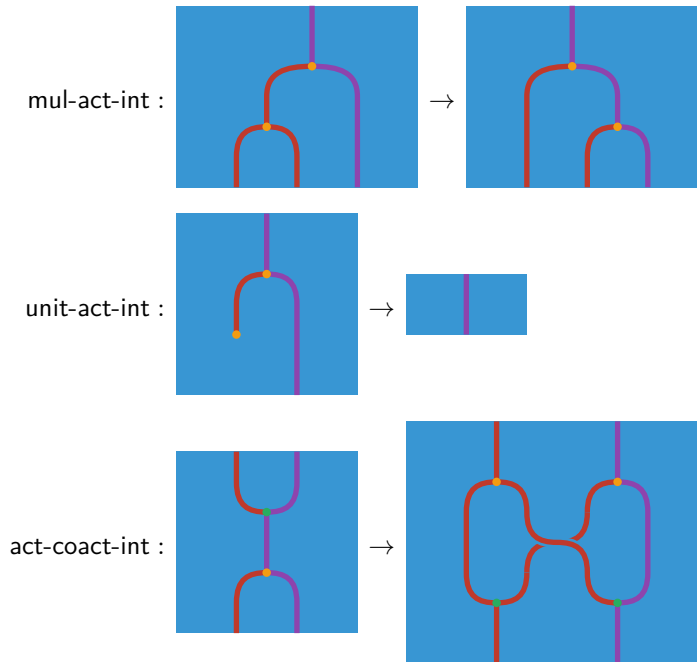
that  $v$  is the coequaliser of  $\alpha$  and  $\varepsilon \otimes 1_M$ , and thus  $M^H$  is also the *object of invariants* of  $M$ . Now the fundamental theorem of Hopf modules then asserts that if idempotents split in a braided monoidal category, up to isomorphism, every Hopf module is free.

► **Theorem 5** (Fundamental Theorem of Hopf Modules [25, Prop. 1],[8, Lemma 3.3.3]). *Let  $M$  be a left  $H$ -Hopf module in a braided monoidal category, and assume the map  $\nu: M \rightarrow M$  splits through  $M^H$ . There is an isomorphism of left  $H$ -Hopf modules  $H \otimes M^H \cong M$ .*

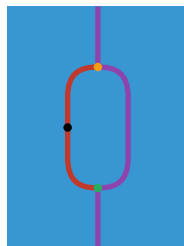
**Proof.** We use the data of the splitting  $\nu = \iota \circ v$  to construct two comparison maps  $\xi := \alpha \circ (1_H \otimes \iota)$  and  $\psi := (1_H \otimes v) \circ \phi$  between  $H \otimes M^H$  and  $M$ , which will show induce the isomorphism of Hopf modules. The string diagrams corresponding to  $\xi$  and  $\psi$  are given in Figure 23.

We claim  $\xi$  respects the Hopf module structure. Compatibility between  $\xi$  and the actions  $\alpha$  and  $\mu \otimes 1_{M^H}$  follows immediately from the multiplication/action interaction law. As for the coactions, we follow Figure 24 in using the action/coaction interaction law, followed by the equaliser property given by Lemma 4. This establishes the claim.

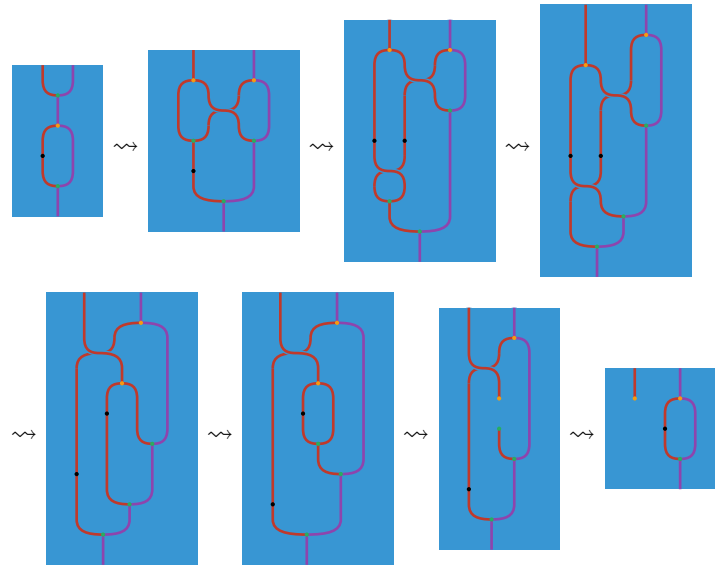
Finally, we prove that  $\xi$  and  $\psi$  are mutual inverses. The proof of  $\xi \circ \psi = 1_M$  is simpler, consisting mainly in the cancellation of the antipode, and is given in Figure 26. The proof of  $\psi \circ \xi = 1_{H \otimes M^H}$  is more involved, making use of  $\nu$  being idempotent (Lemma 3) as well as  $\iota$  being an equaliser (Lemma 4). This is given in Figure 27. ◀



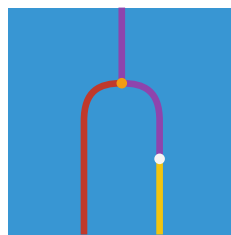
■ **Figure 20** The  $H$ -Hopf module laws for  $M$ , with the duals to the first two omitted.



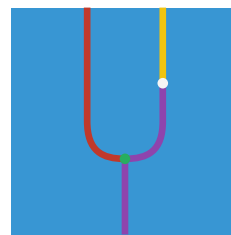
■ **Figure 21** The idempotent endomorphism  $\nu: M \rightarrow M$ .



■ **Figure 22** The proof of  $\phi \circ \nu = (\eta \otimes 1_M) \circ \nu$  in Lemma 3.

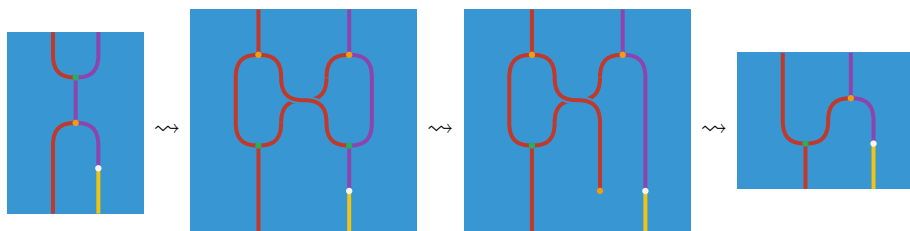


(a) The map  $\xi: H \otimes M^H \rightarrow M$ .



(b) The map  $\psi: M \rightarrow H \otimes M^H$ .

■ **Figure 23** The string diagrams for the mutually inverse  $\xi$  and  $\psi$ .



■ **Figure 24** The map  $\xi$  respects the coactions  $\phi$  and  $\delta \otimes 1_{M^H}$ .



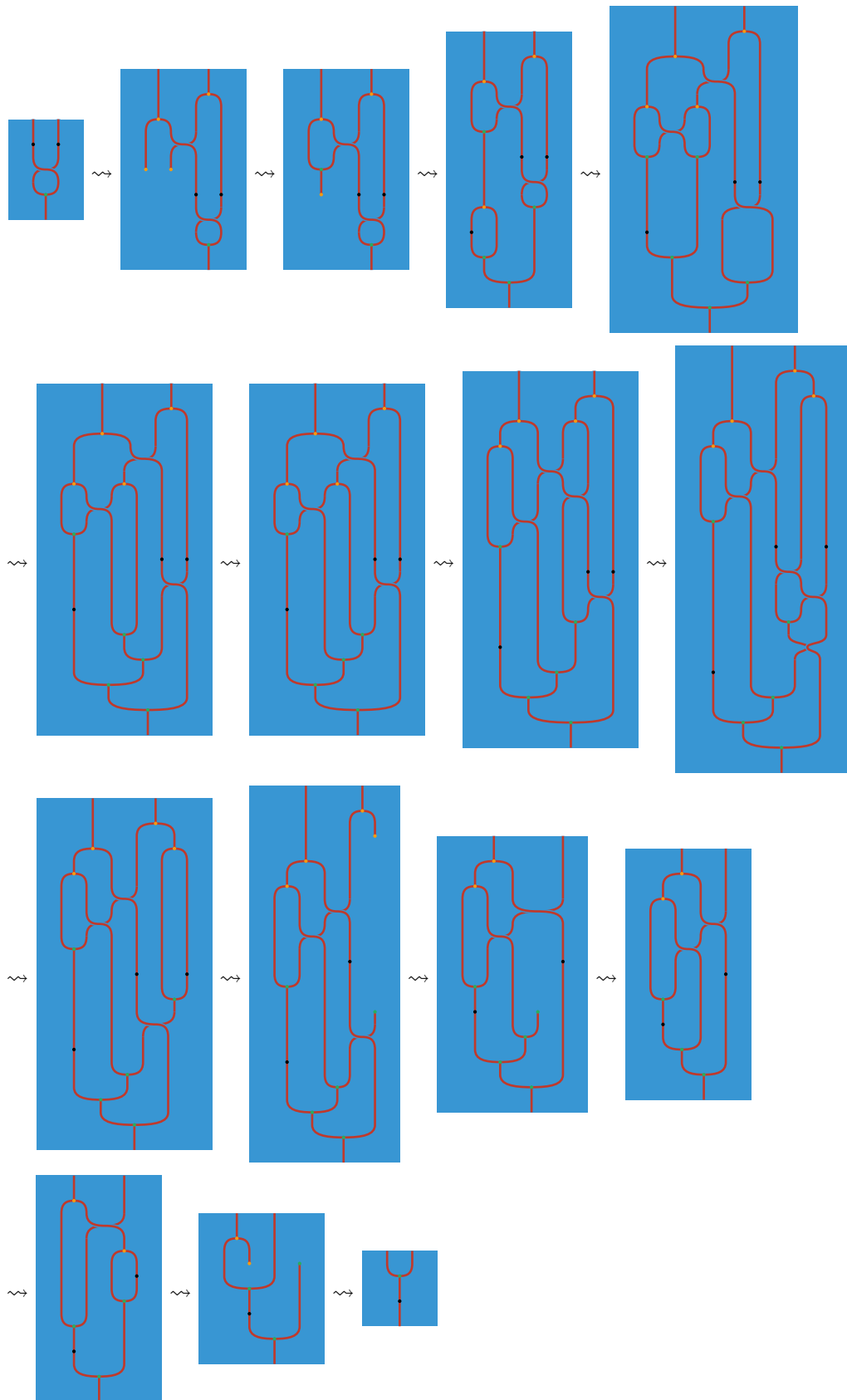
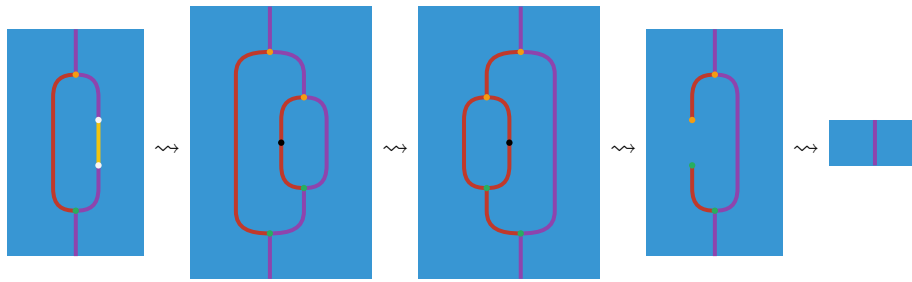
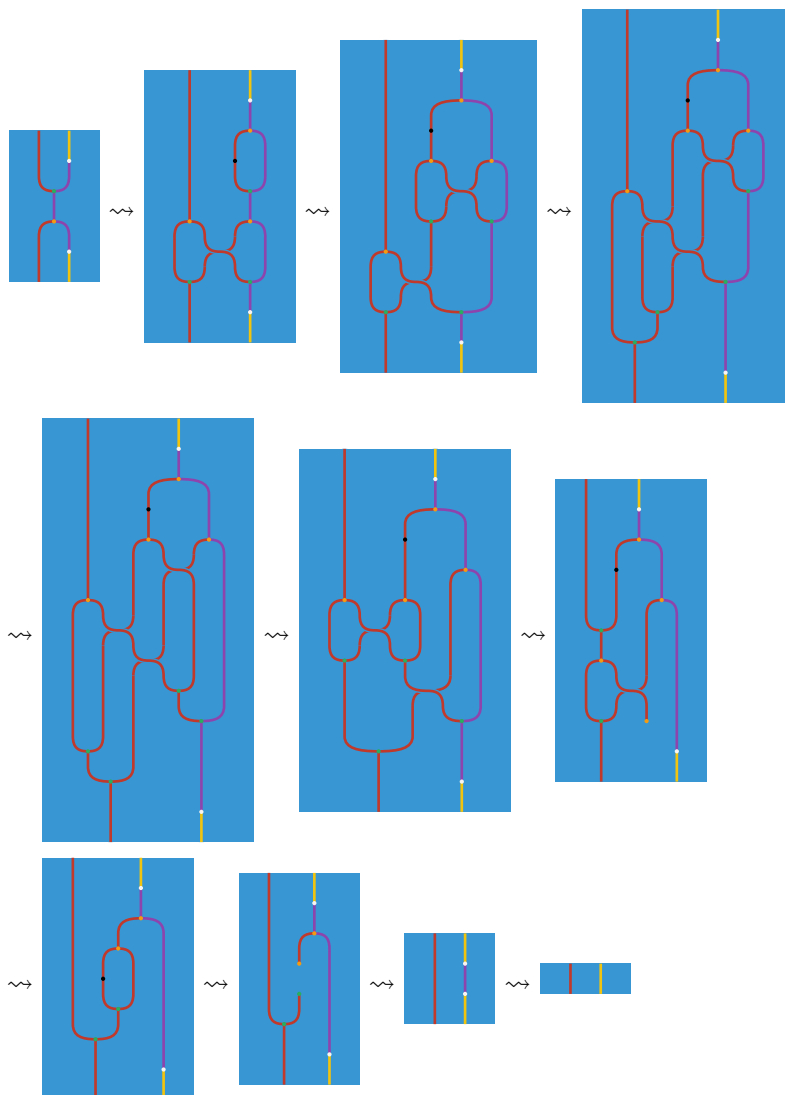


Figure 25 The proof of Lemma 2.



■ **Figure 26** The proof of  $\xi \circ \psi = 1_M$  in Theorem 5.



■ **Figure 27** The proof of  $\psi \circ \xi = 1_{H \otimes_M H}$  in Theorem 5.