A Survey of Categorical Symmetry

Da-Chuan Lu

1Department of Physics, University of California at San Diego, La Jolla, CA 92093

This paper reviews some recent developments in the categorical symmetry. The categorical symmetry is a generalization of the ordinary global symmetry in the sense that the operators which impose the symmetry form a fusion category instead of the ordinary group algebra. Categorical symmetries in conformal field theories and gapped phases are analyzed and the consequences of imposing the categorical symmetry is discussed.

INTRODUCTION

Global symmetry places many constraints on the kinematics of a physical system and is endowed with prediction power by Noether theorem (for conservation law), Ward identity (for correlation functions), Mermin–Wagner theorem (for spontaneous symmetry breaking) and so on.

Global symmetry with anomaly further constrains the renormalization group (RG) flow. For example, the ‘t Hooft anomaly is an obstruction for gauging the global symmetry, if a global symmetry has the ‘t Hooft anomaly, then there cannot be a symmetric RG flow to unique gapped phase, the possible phases are topological order, gapless or spontaneous symmetry breaking phase.

Another important concept is anomaly inflow, which allows us to put the theory with ‘t Hooft anomaly on the boundary of symmetry protected topological phases (SPT). SPT is a gapped phase, if no symmetry is imposed, the SPT can be connected to the trivial gapped phase with unique ground state. When imposing the symmetry, there will be a quantum phase transition between the SPT and the trivial gapped phase. In the language of anomaly inflow, an SPT phase in spatial dimension $d$ has a global symmetry $G$ such that, when placed on a manifold with boundary, the $(d-1)$-dimensional theory on the boundary has a ‘t Hooft anomaly for $G$.

All the above discussions suggest various applications of the global symmetry $G$. The question is can we generalize the notion of global symmetry and may have more or other constraints on the system?

Before the generalizations, let’s see the two manifestations of global symmetry in a system,

1. The states and local operators are classified by representations of the symmetry group $G$, they are usually called particles and observables charged under $G$, denoted by $V_{\lambda}$, $\lambda$ for some representation of $G$. In the Lagrangian formalism, the field contents are representations of $G$ and each term in the Lagrangian transforms as a singlet of $G$.

2. The operators $U_g$ supported on the $(d-1)$-dimensional manifold $M^{(d-1)}$ impose the symmetry, meaning that $U_g$ commute with the Hamiltonian and $U_g U_h = U_{gh}$ and $g, h, gh \in G$. Such operators are usually called charge operator.

It turns out the charged operator $V_\lambda$ may be not clear but the charge operator $U_g$ still exist (“d” makes differences).

Nonetheless, the first manifestation can be used to make meaningful analytic continuation of the symmetry group, for example, continuous group $O(N), Sp(N), U(N)$ and discrete group $S_N$ can be analytic continued to $N \in \mathbb{Z} \to n \in \mathbb{R}$, the price is to use category language. This is explored in [1], by generalizing the charged objects $V_\lambda$ which form representations in $G$ to objects in category $\text{Rep}(G)$ (this category contains rep as object, irep as simple object and intertwiner as morphism, fusion rule is the usual tensor product of reps) and the invariant tensors in the Lagrangian to “birdtrack” string diagrams in $\text{Rep}(G)$ [2].

More fruitful generalizations of global symmetry are based on the second manifestation, namely generalizing the charge operator $U_g$. Note that the charge operator $U_g$ associated to symmetry means that the dependence on its supported $(d-1)$-manifold $M^{(d-1)}$ is topological, but the charged objects $V_\lambda$ is not always topological. There are two main generalizations in the literature,

1. High-form symmetry[3]: for $q$-form symmetry, the charge operator is now supported on the $(d-q-1)$-manifold, denoted as $U_g(M^{(d-q-1)})$ and they still satisfy the group multiplication,

$$U_g(M^{(d-q-1)}) U_h(M^{(d-q-1)}) = U_{gh}(M^{(d-q-1)}) \quad (0.1)$$

for $q > 0$, the group must be abelian and the charged objects are supported on $q$-dimensional manifold. The ordinary symmetry is thus $0$-form symmetry.

2. Categorical symmetry in 2d [4–10]: as noted previously, the charge operator $U_g$ supported on $M^{(d-1)}$ is topological, they form a subset of the most general topological defect line operators (TDLs) $X_a, a \in \mathcal{C}$ where $\mathcal{C}$ is some label set. These operators commute with the Hamiltonian and in general only satisfy the fusion algebra, namely,

$$X_a \otimes X_b = \bigoplus_{c \in \mathcal{C}} N_{ab}^c X_c \quad (0.2)$$

$\mathcal{N}_{ab}^c$ is some fusion matrix.
where \( N_{ab}^c \in \mathbb{Z}_{\geq 0} \). For ordinary symmetry, we require \( N_{ab}^c = \begin{cases} 1 & \text{if } c = ab \in G, \\ 0 & \text{otherwise} \end{cases} \), thus \( X_a \otimes X_b = X_{ab} \) gives back the group multiplication law. Note that \( X_a \) is not invertible for general \( N_{ab}^c \). It is possible to generalize the notion of global symmetry by having certain subset of the TDLs.

In a short, the categorical symmetry is defined by the presence of the topological defect line operators which form a fusion category (kind of a generalization of finite group). The conditions will be explicit later.

3. Combining above two: In dimension \( d > 2 \), it is possible to have topological operator supported on different dimensions corresponding to various high form symmetries. Specifying what conditions should these topological operators satisfy will lead to more exotic global symmetry.

We will focus on the second generalization in this paper, and most of the systems we studied are 2d. The topological defect line operators in the 2d conformal field theory (CFT) are a good starting point to understand categorical symmetry \([5, 7]\), the categorical symmetry can also be generalized to other gapped phases.

### TOPOLOGICAL DEFECT LINE OPERATORS IN CFTS AND CATEGORICAL SYMMETRY

See Appendix for a short review on the concepts and notations in CFT. We denote the modes of the holomorphic and anti-holomorphic component of the stress tensor by \( L_m \) and \( \bar{L}_m \), we call the defect line operator \( X_a \) topological if and only if \([L_m, X_a] = 0 = [\bar{L}_m, X_a] \) for all \( m \in \mathbb{Z} \) \([6][20]\), meaning that the topological defect line operators are transparent to all components of the stress tensor, and they can be deformed without change the correlation functions. Since the Hamiltonian \( H \propto (L_0 + \bar{L}_0) \), the TDLs commute with the Hamiltonian, but the inverse needs not to be true.

One can think of the TDL as a circle surrounding the origin, the composition (or fusion) of two TDLs can be realised by placing the defect \( X_a \) on the unit circle, the defect \( X_a \) on a circle with radius \( r > 1 \) and taking the limit \( r \to 1 \), and in the limit (well-defined for TDLs) one obtains the fused TDL \( X_a \otimes X_b \) at \( r = 1 \). In this way, one can get the fusion algebra of the TDLs \([6, 11, 12]\). Interestingly, for all A-type Virasoro minimal models, this fusion algebra coincides with the fusion rules of Virasoro highest weight representations \([11]\). The TDLs and their fusion rules for free boson are also known \([13]\).

With the correspondence between the fusion rule of TDLs and highest weight representations in the A-type minimal models, one can have some applications, for example, the Ising CFT contains 3 primary field operators \( 1, \sigma, \epsilon \), and the fusion rule is

\[
\sigma \times \sigma = 1 + \epsilon, \quad \sigma \times \epsilon = \sigma, \quad \epsilon \times \epsilon = 1, \tag{0.3}
\]

the correspond TDLs have the fusion rule,

\[
X_\sigma \otimes X_\sigma = X_1 \otimes X_\epsilon, \quad X_\sigma \otimes X_\epsilon = X_\sigma, \quad X_\epsilon \otimes X_\epsilon = X_1. \tag{0.4}
\]

We would like to further specify the types of the TDLs. Following \([13]\), the group-like TDLs are those invertible TDLs, \( X_a \otimes X_a = X_1 \) and \( X_a \otimes X_b = X_{ab} \) with \( a, b, ab \in G \), and bar denotes the orientation reversal. They actually impose the symmetry transformation of local operators in the ordinary way, e.g. \( X_\epsilon \) in the Ising CFT is the group-like TDL which generates the \( \mathbb{Z}_2 \) symmetry. We denote this set of TDLs by \( \mathcal{G} \).

![FIG. 1: The blue line is the group-like TDL, the green dot is the bulk field, when taking the TDL past the bulk fields, it will yield a transformed bulk field. Note figure reprints from [5], I can’t draw a better one.](image)

Note that till now both sides of the TDLs we discussed are the same CFT, in general they can be different. The duality TDLs are those separate the CFT(A) and CFT(B), also satisfy: for every bulk fields in CFT(A), first taking \( X_a \) past the bulk fields and then taking another \( X_b \) past the resulting in general disorder fields in CFT(B), gives back a sum over bulk fields in CFT(A). The iff condition is the duality TDLs fusing with themselves yields a sum of group-like TDLs, \( X_a \otimes X_a = \bigoplus_{b \in \mathcal{G}} X_b \). The set of the duality TDLs is denoted by \( \mathcal{D} \), note that \( \mathcal{G} \) is the subset of \( \mathcal{D} \).

![FIG. 2: The blue lines are the dual TDLs, the green dot is the bulk field, when taking the TDL past the bulk fields, it will yield disorder fields as the middle step, then taking another TDL past, results in a summation of bulk fields and the duality TDLs no longer separable. Note, figure reprints from [5].](image)
can be expressed in terms of torus amplitudes with defect lines of CFT(A) as follows,  

\[
\begin{array}{c}
\text{B} \\
\text{A}
\end{array}
= \frac{d_{\text{in}(A)} d_{\text{in}(B)}}{d_{\text{in}(\gamma)}^{\lambda}} \sum_{\sigma} \chi_{\sigma}
\]

(0.5)

The CFT(B) obtained in this way is the orbifold of CFT(A), one can also think this as a gauging process. If an RCFT possesses a duality defect, it automatically also has the "auto-orbifold" (or self-duality) property which means the model is equivalent to its orbifold or G-gauged version.

In the Ising CFT, Kramer-Wannier duality is imposed by \( X_\sigma \) which transforms the order operators to the disorder operators, and \( X_\sigma \otimes X_\sigma = X_1 + X_\sigma \) suggesting \( X_\sigma \) is the duality TDL by definition. The presence of \( X_\sigma \) means that the Ising CFT is equivalent to its \( \mathbb{Z}_2 \) gauged version which is summing over \( \epsilon \) the \( \mathbb{Z}_2 \) lines. Indeed the Ising model is equivalent to the \( \mathbb{Z}_2 \) gauge theory. In the Ising model, \( G = \{ 1, \epsilon \} \), \( D = \{ 1, \epsilon, \sigma \} = C \).

The Kramer-Wannier duality of Ising CFT can be generalized to system with abelian symmetry by the Tambara-Yamagami category. The Tambara-Yamagami category consists of TDLs \( X_g \) associated to an abelian group \( G \) as well as a duality line \( X_\sigma \), satisfies the following fusion rule,  

\[
X_g \otimes X_N = X_N, \quad X_N \otimes X_N = \bigoplus_{h \in G} X_h
\]

(0.6)

These categories are \( \mathbb{Z}_2 \) extensions of \( \text{Vec}_G \) where \( G \) is an abelian group and we denote the Tambara-Yamagami category induced by abelian \( G \) as \( \text{TY}(G) \) (more detail need specifying, we omit for the purpose of demonstration).

To wrap up what we learned about the topological defect lines in CFTs,

1. The ordinary symmetry transformation is imposed by group-like topological defect lines which are invertible TDLs.

2. The most general TDLs form a fusion category and the fusion rule is the same as the fusion rule of primary field operators in the type-A minimal model, fusion rules in other minimal models and free boson theory are also known. (One can think of a fusion category as a non-commutative, non-cocommutative generalization of a finite group, see the link)

3. The duality TDL satisfies \( X_a \otimes X_a = \bigoplus_{h \in G} X_h \), and when an RCFT has such duality TDL, it is automatically equivalent to its orbifold or \( G \)-gauged version.

4. The \( G \)-orbifold or \( G \)-gauging is obtained by summing over a sufficiently fine network of group-like TDLs associated to \( G \).

CATEGORICAL SYMMETRY OF GAPPED PHASES

We have discussed that the TDLs are common in the CFTs, they are in general non-invertible and form a fusion category, we denote the CFT with a set \( C \) of TDLs as having the categorical symmetry \( C \). It is also possible to have categorical symmetry in gapped phases, there are two known systems that have categorical symmetry,

1. Nonabelian-\( G \) gauge theory.

2. Gapped boundary of 3d topological field theory (TFT) with categorical "gauge" symmetry.

Nonabelian \( G \)-gauge theory

Let's first review two relevant fusion categories,

1. \( \text{Vec}_G^\omega \): for finite group \( G \) and 3-cocycle \( \omega \in H^3(G, U(1)) \), \( \text{Vec}_G^\omega \) contains a simple objects for every element \( g \in G \) and fusion rules defined by group multiplication, with one-dimensional fusion spaces and \( F_{g_1, g_2, g_3} = \exp(i\omega(g_1, g_2, g_3)) \). If \( \omega \) is trivial, then \( \text{Vec}_G \) describes the ordinary group.

2. \( \text{Rep}(G) \): for finite group \( G \), the objects of \( \text{Rep}(G) \) are representations of \( G \) and simple objects are the irreducible representations, the fusion is the usual tensor product of representations.

In a gapped phase of \( G \)-gauge theory where \( G \) is a finite (possibly nonabelian) group, the Wilson lines are topological operators generating an integral fusion category \( \text{Rep}(G) \) (the quantum dimension of a Wilson line is the dimension of its representation). When \( G \) is abelian, then \( \text{Rep}(G) = \text{Vec}_G \) and the symmetries generated by the Wilson line are grouplike symmetries \( \hat{G} \) known as the magnetic symmetries of the gauge theory, where  

\[
\hat{G} = \{ \chi : G \to U(1) \mid \chi \text{ is an irrep} \}
\]

is the Pontrjagin dual of \( G \), for example, \( \hat{\mathbb{Z}_n} \cong \mathbb{Z}_n \).

The nonabelian \( G \)-gauge theory is more interesting, since now \( \text{Rep}(G) \) contains simple objects which are not invertible, corresponding to irreps of dimension greater than one, so one needs the fusion category perspective to analyze the symmetry properly. One can further sum over a fine enough mesh of lines in \( \text{Rep}(G) \) to get back to the theory with global symmetry \( G \). The gauging and regauging the nonabelian symmetry is discussed in [10].
Gapped boundary of 3d TFTs

One incarnation of the holographic principle in quantum field theory is the correspondence between 3d $G$-Chern-Simons theory as the bulk field theory and the 2d Wess-Zumino-Witten CFT on a suitable Lie group $G$ as the boundary field theory [14]. This fits in the more general anomaly inflow argument, namely when putting the non-trivial bulk theory on the manifold with boundary, there is some inconsistency that needs to be canceled by the boundary theory, the inconsistency is captured by the ‘t Hooft anomaly of the symmetry $G$.

Actually the CFT is one consequence of $G$ with ‘t Hooft anomaly, all possibilities are,

1. gapless,

2. $G$ symmetry spontaneously broken (for discrete symmetry $G$ in $d > 1$ and continuous $G$ in $d > 2$, $d$ is spacetime dimension.)

3. non-trivial gapped phase preserve $G$ (e.g. topological order)

For finite group, the above discussion is replaced by Dijkgraaf-Witten theory which can be thought of as finite group version of Chern-Simons theory, and they admit SPT or spontaneous symmetry breaking phase as the gapped boundary. By generalizing the ordinary symmetry to categorical symmetry, it is possible to have gapped phase preserve the categorical symmetry $C$ via the anomaly inflow. In [8], the authors argue that a 1+1D theory with a fusion category symmetry forms a boundary condition of a 2+1D topological quantum field theory known as Turaev-Viro/Levin-Wen theory, which is a construction induced from a fusion category $C$. They also prove several theorems, one is

If $C$ has an object with non-integer quantum dimension, then the Turaev-Viro theory defined by $C$ does not admit a gapped, non-degenerate, $C$-symmetric boundary condition.

For example $TY(\mathbb{Z}_2)$ having $\dim(X_N) = \sqrt{2}$ does not admit gapped boundary but $TY(\mathbb{Z}_2 \times \mathbb{Z}_2)$ having $\dim(X_N) = 2$ admits symmetric gapped boundary and this is associated to the self-dual property of the SPT order on the 1+1d boundary [8]. However, there are some other examples of anomalous fusion categories which have all integer quantum dimensions, such as $TY(\mathbb{Z}_n)$ with $n$ being a perfect square. (The Tambara-Yamagami category is introduced around Eq. 0.6, it also needs more data to specify the category, they are omitted for demonstration. With specified data, $TY(\mathbb{Z}_2 \times \mathbb{Z}_2)$ be equivalent to $\text{Rep}(H_3), \text{Rep}(D_5), \text{Rep}(Q_8)$ as analyzed in [8, 10, 15]).

SUMMARY

The invertible topological defect lines which forms a $\text{Vec}_G$ impose the ordinary symmetry, this is generalized to the categorical symmetry by allowing the TDLs to form a general fusion category and the TDLs are in general non-invertible. The interesting fusion categories that possess the non-invertible TDLs are

1. $\text{Rep}(G)$ with $G$ being nonabelian, this can be obtained by gauging the nonabelian $G$ or equivalent to certain Tambara-Yamagami category.

2. Tambara-Yamagami category $TY(G)$, which is relevant to the self-duality/auto-orbifold of a system with any abelian ordinary symmetry $G$, is constructed by having the group-like TDLs associated to the abelian group and one additional duality TDL. This is a generalization of the Ising Kramer-Wannier duality.

Conversely, if a system admits duality line, then it is automatically self-dual, if the dimension of the duality line is not integer, the system cannot have symmetric gapped phase. The duality line places a rather strict and interesting constraint on the behavior of the system. Apart from the group-like and duality TDLs, other subsets of TDLs may also place interesting constraints on the system, these need to be explored.

Other physical relevant fusion categories need to be found, one is to generalize the duality to $n$-ality, for example, the $n$-ality line may have the fusion algebra $X_N^{\otimes n} = \bigoplus_{h \in G} X_h$ and $\dim X_N = \sqrt{|G|}$. Another one is to further understand the TDLs corresponding to the $T$-duality of free boson theory and level/rank duality of the WZW theory, the TDLs in free boson theory are investigated in [13].

Another interesting generalization is towards higher dimension. It is known that when gauging a $p$-form symmetry given by an abelian group, the dual abelian group generates the $(d-2-p)$-form symmetry. However for non-abelian 0-form symmetry, the gauged version will have $\text{Rep}(G)$ as the dual $(d-2)$-form “categorical” symmetry, and for $d = 3$ the 1-form symmetry seems to require including the modular tensor categories[10]. It is also interesting to study the duality lines/surfaces in higher dimension, for example, the electromagnetic duality exchanging electric field and magnetic field is an analogue of the 2d Kramer-Wannier duality. Recent developments in duality web of the 2+1d quantum field theories offers more examples [16].

Acknowledgements My warmest thanks to Prof. John McGreevy for teaching us the group theory and exposing us to various deep and interesting topics, the discussion with John instigates and helps through this
APPENDIX

A short review on CFT and minimal models

We review some notations in the CFTs [17]. The \( 2d \) plane is parametrized by complex number \( z, \bar{z} \), the fields are separated to holomorphic and anti-holomorphic parts, in many cases they decouple, one only need to focus on the holomorphic part and the antiholomorphic part is easy to restore. The energy-momentum tensor \( T(z), \bar{T}(\bar{z}) \) generates the local conformal transformation, and the mode expansions is \( T(z) = \sum_{n \in \mathbb{Z}} z^{-n-2} L_n \). The \( L_n, \bar{L}_n \) satisfy the well-known Virasoro algebra,

\[
[L_n, L_m] = (n - m)L_{n+m} + \frac{c}{12}n(n^2 - 1)\delta_{n+m,0} \\
[L_n, \bar{L}_m] = 0 \\
[\bar{L}_n, L_m] = (n - m)\bar{L}_{n+m} + \frac{c}{12}n(n^2 - 1)\delta_{n+m,0}.
\]

where \( c \) is the central charge. Similar to the representations of Lie algebra. The highest weight state is defined by,

\[
L_0 |h, \bar{h}\rangle = h |h, \bar{h}\rangle, \quad \bar{L}_0 |h, \bar{h}\rangle = \bar{h} |h, \bar{h}\rangle \\
L_n |h, \bar{h}\rangle = 0, \quad \bar{L}_n |h, \bar{h}\rangle = 0 \quad \text{if } n > 0.
\]

As mentioned previously, one can focus on the holomorphic part, the descendant states are obtained by applying the raising operators in all possible ways,

\[
L_{-k_1}L_{-k_2}...L_{-k_n} |h\rangle, \quad (1 \leq k_1 \leq ... \leq k_n).
\]

The highest weight state and the descendant states form a representation of the Virasoro algebra, called Verma module, labeled by the central charge and highest weight \( V(c, h) \).

It is possible that \( V(c, h) \) is reducible, there exists a singular vector, the norm of it and its descendant are zero. By identifying states that differ only by a state of zero norm, one can get the irreducible representations \( M(c, h) \). The minimal models are constructed by \( M(c, h) \), the Hilbert space is

\[
\mathcal{H} = \bigoplus_{1 \leq r < p', 1 \leq s < p} M(c, h_{r,s}) \otimes M(c, h_{r,s})
\]

and this diagonal minimal model is labeled by \( M(p, p') \) with \( p > p' \). And,

\[
c = 1 - \frac{(p - p')^2}{pp'} \\
h_{r,s} = \frac{pps - (p - p')^2}{4pp'}
\]

As discussed in the main text, one can think of the topological defect line as a circle surrounding the origin, the composition (or fusion) of two TDLs can be realised by placing the defect \( X_b \) on the unit circle, the defect \( X_a \) on a circle with radius \( r > 1 \) and taking the limit \( r \to 1 \), and in the limit (well-defined for TDLs) one obtains the fused TDL \( X_a \otimes X_b \) at \( r = 1 \). In this way, one can get the fusion algebra of the TDLs [6, 11, 12]. Diagrammatically, fusing blue and red lines to a green line is represented as,

The fusion category is determined by the fusion rule

\[
X_a \otimes X_b = \bigoplus_{c \in C} N_{ab}^c X_c
\]

and the \( F \)-symbols associated to the \( F \)-move,

\[
F^{-13}_{12} = (F^{-13}_{12})_{ab}
\]

The \( F \)-symbol is \( N \)-by-\( N \) matrix if \( a, b \) are \( N \)-dimensional, for \( \text{Vec}_G^\omega \), \( F \) is 1-dimensional and \( F^{g_1, g_2, g_3, g_4}_{g_1, g_2, g_3, g_4} = \exp(i\omega(g_1, g_2, g_3, g_4)) \). The \( F \)-symbols satisfy the consis-

paper. We also thank Zhengdi Sun for various discussion and Prof. Yi-Zhuang You for explaining gauging the symmetry of anyons and self-duality in various systems.

The minimal models are relevant to many physical applications, e.g. \( M(4,3), c = \frac{1}{2} \) describes Ising transition, \( M(5, 4), c = \frac{7}{10} \) describes tricritical Ising criticality and so on.

Thanks to the state-operator correspondence in the CFT, the highest weight states correspond to primary field operators. The operator product expansion of these primary field operators in the minimal models satisfies the fusion algebra. For example, in the Ising CFT, the primary fields are \( 1, \sigma, \epsilon \) with \( (h, \bar{h}) = (0, 0), (0, \frac{1}{16}), (\frac{1}{16}, \frac{1}{16}) \), the fusion rules for these primary operators are,

\[
\sigma \times \sigma = 1 + \epsilon, \quad \sigma \times \epsilon = \sigma, \quad \epsilon \times \epsilon = 1
\]
tency condition, which is called the “pentagon” equation, stating that the upper way and lower way to get the right side from the left side should be equivalent. Knowing the fusion rule, one can solve the pentagon equation to find the $F$-symbols.