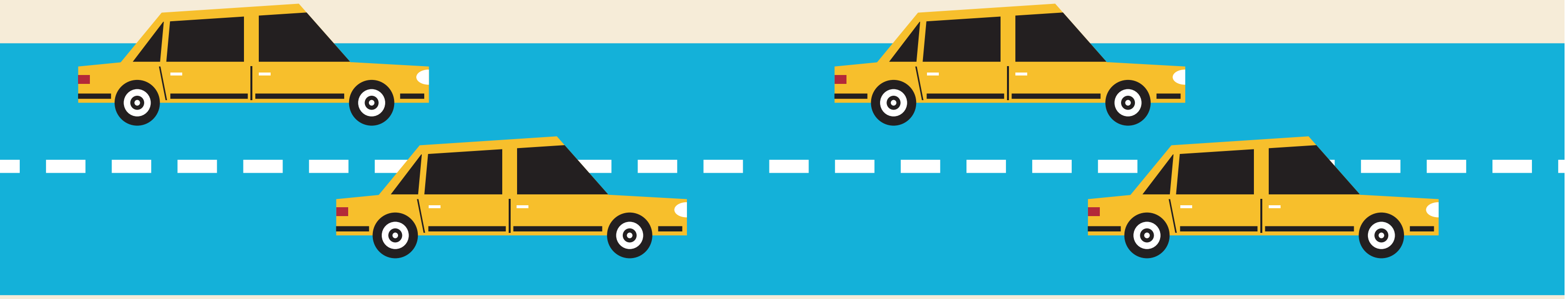
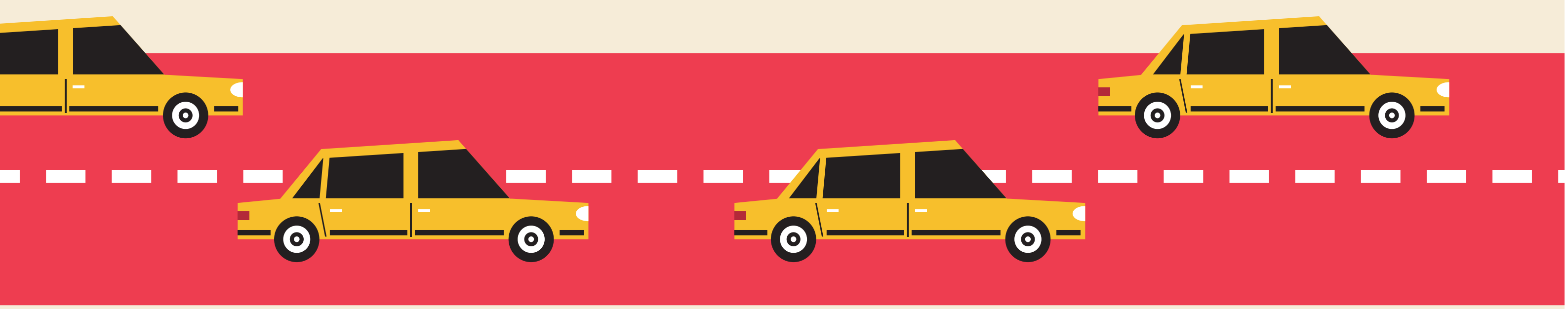
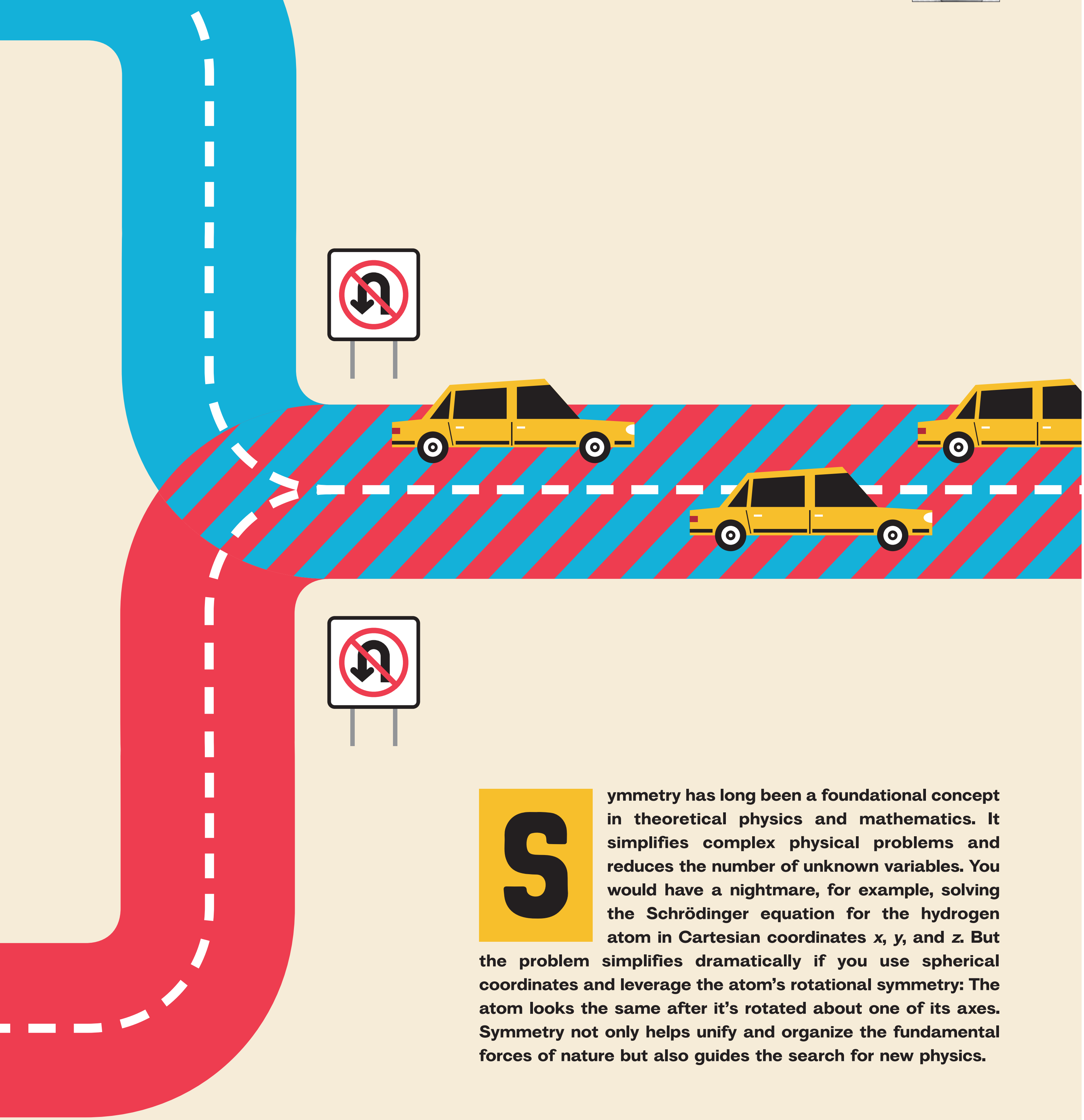


NONINVERTIBLE SYMMETRIES: WHAT'S DONE CANNOT BE UNDONE

Recent research has shown that the traditional notion of symmetry is too limited. A new class of symmetries is bringing surprising insights to quantum systems.

Shu-Heng Shao

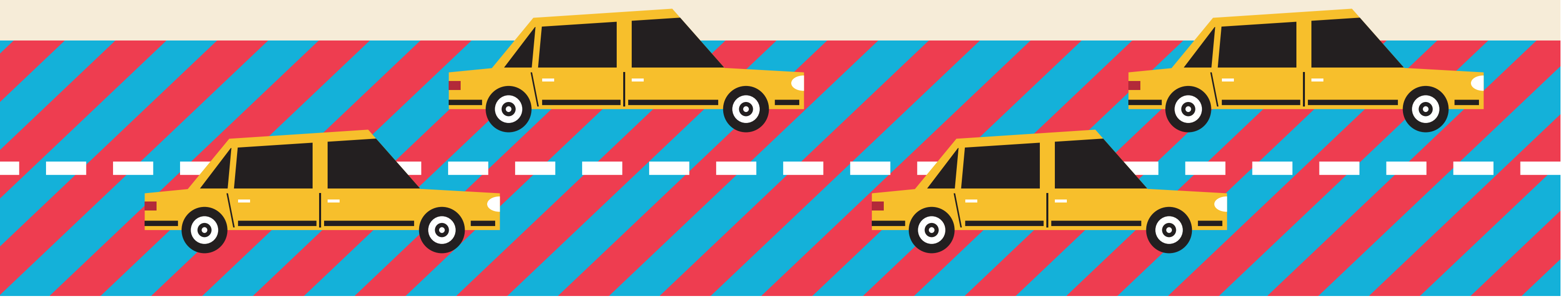




Shu-Heng Shao is an assistant professor at the MIT Center for Theoretical Physics—a Leinweber Institute in Cambridge, Massachusetts. He works on quantum field theory and quantum many-body systems with applications in high-energy physics, condensed-matter physics, and quantum gravity.

Symmetry has long been a foundational concept in theoretical physics and mathematics. It simplifies complex physical problems and reduces the number of unknown variables. You would have a nightmare, for example, solving the Schrödinger equation for the hydrogen atom in Cartesian coordinates *x*, *y*, and *z*. But the problem simplifies dramatically if you use spherical coordinates and leverage the atom's rotational symmetry: The atom looks the same after it's rotated about one of its axes. Symmetry not only helps unify and organize the fundamental forces of nature but also guides the search for new physics.

"The fundamental reason that symmetries can be noninvertible is quantum superposition."



Symmetry transformations are those that leave a system looking and behaving the same. Conventional symmetry transformations are invertible. If we rotate a square by 90°, for example, the transformation can be undone by a −90° rotation. Such intuition is formulated rigorously by Wigner's theorem, which implies that every symmetry transformation in quantum mechanics has an inverse. The mathematical language used to describe conventional symmetry transformations is called group theory, a foundational concept that has shaped modern physics for more than a century (see, for example, the article by Martin Rodriguez-Vega, Maia Vergniory, and Greg Fiete, Physics Today, May 2022, page 42).

One way only

In recent years, however, researchers have shown that the traditional notion of symmetry is too limited in quantum field theory and quantum many-body systems. A new class of symmetries—noninvertible—has been identified in various physical systems, including lattice models describing magnetism and quantum field theories of strong interactions between quarks. As the name suggests, noninvertible symmetries are implemented by transformations that do not have inverses—that is, what's done cannot be undone.

The fundamental reason that symmetries can be noninvertible is quantum superposition. In deterministic classical physics, a cat is either alive or dead. In quantum physics, Schrödinger's cat can be both alive and dead simultaneously. The wavefunction describing Schrödinger's cat is a superposition of two individual wavefunctions—one for an alive cat and one for a dead cat. Superposition introduces more possibilities for symmetries in quantum physics: A symmetry transformation can cause the wavefunction of a single cat to become a superposition of two. If the transformation is repeated, the result is a superposition of increasingly many cat wavefunctions, and no inverse transformation reverts to a single cat.

As paradoxical as it may sound, the new symmetries lead to new conservation laws, which serve as novel tools to study strongly coupled physical systems. They also point to alternative physical models and beg for a new mathematical framework to describe symmetries in quantum physics.

Noninvertible symmetry of a magnet

Noninvertible symmetries already exist in physicists' favorite toy model for ferromagnetism: the Ising model in one spatial dimension. The 1D model consists of an array of qubits placed on a circle, as illustrated in figure 1. Each qubit can be spin up $|\uparrow\rangle$, spin down $|\downarrow\rangle$, or any quantum superposition of the up and down states, such as $|\rightarrow\rangle \propto |\uparrow\rangle + |\downarrow\rangle$.

The state $|\uparrow\uparrow\cdots\uparrow\rangle$, in which every spin is pointing up, corresponds to a magnet whose north pole is pointing up. Similarly, the state $|\downarrow\downarrow\cdots\downarrow\rangle$ corresponds to a magnet whose south pole is pointing up. On the other hand, the state

$$|\rightarrow\rightarrow\cdots\rightarrow\rangle \propto |\uparrow\uparrow\cdots\uparrow\rangle + |\downarrow\uparrow\cdots\uparrow\rangle + |\uparrow\downarrow\cdots\uparrow\rangle + \cdots + |\downarrow\downarrow\cdots\downarrow\rangle \quad (1)$$

represents a superposition of all possible spin configurations. Since that is a state with no preference for spin up or spin down, magnetization is lost. The transition from $|\uparrow\uparrow\cdots\uparrow\rangle$ to $|\rightarrow\rightarrow\cdots\rightarrow\rangle$ models the process of heating up a magnet: As the temperature reaches a critical value, the magnet loses its magnetization.

What are the symmetries in the toy model for a magnet? Because the north and south poles are on the same footing, an ordinary symmetry can transform one to the other. It flips all the spins from up to down and vice versa: $|\uparrow\uparrow\cdots\uparrow\rangle \to |\downarrow\downarrow\cdots\downarrow\rangle \to |\uparrow\uparrow\cdots\uparrow\rangle$. That is an invertible symmetry—apply it twice, and we return to the starting point. The demagnetized state $|\rightarrow\rightarrow\cdots\rightarrow\rangle$ is symmetric under the spin-flip symmetry because there is no notion of north versus south.

At the critical temperature, an additional symmetry emerges whose effect is $|\uparrow\uparrow\cdots\uparrow\rangle \to |\rightarrow\rightarrow\cdots\rightarrow\rangle$ and $|\downarrow\downarrow\cdots\downarrow\rangle \to |\rightarrow\rightarrow\cdots\rightarrow\rangle$. The additional symmetry transformation acts identically on the up and down states, as illustrated in figure 2. Whether the output state was initially in the up or down state before the transformation isn't knowable. Relatedly, if we apply the transformation a second time, we find a superposition of up and down states: $|\rightarrow\rightarrow\cdots\rightarrow\rangle \to 1/\sqrt{2}\ (|\uparrow\uparrow\cdots\uparrow\rangle + |\downarrow\downarrow\cdots\downarrow\rangle)$. The symmetry transformation cannot be inverted and thus it is a noninvertible symmetry.

The technical details

We now examine more carefully the noninvertible symmetry in the Ising model, which has L qubits, labeled by $j = 1, 2, \ldots, L$, arranged on a 1D closed, periodic ring. (It has a counterpart in two dimensions.[1]) On each qubit, a quantum operator, denoted as $\mathbf{Z}$, can be applied to measure the spin: $\mathbf{Z}|\uparrow\rangle = +|\uparrow\rangle$, $\mathbf{Z}|\downarrow\rangle = -|\downarrow\rangle$. Alternatively, another quantum operator, denoted as $\mathbf{X}$, can be applied to flip the spin, where $\mathbf{X}|\uparrow\rangle = |\downarrow\rangle$, $\mathbf{X}|\downarrow\rangle = |\uparrow\rangle$.

If we represent a qubit's spin-up and spin-down states as two column vectors, $\binom{1}{0}$ and $\binom{0}{1}$, then the operators become the Pauli matrices $\mathbf{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and $\mathbf{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. When we have multiple qubits, we can similarly define $\mathbf{Z}_j$ and $\mathbf{X}_j$ as operators that measure or flip the jth qubit while leaving the others unchanged.

In quantum mechanics, the time evolution of a system is governed by an operator called the Hamiltonian, which features in the Schrödinger equation. At the critical temperature, the Hamiltonian for the Ising model takes the form

$$\mathbf{H} = -\sum_{j=1}^{L} \mathbf{X}_j - \sum_{j=1}^{L} \mathbf{Z}_j \mathbf{Z}_{j+1}\,. \tag{2}$$

The first term models a transverse magnetic field, and the second term models the coupling between the spins of neighboring qubits.

What are the symmetries of the critical Ising model? A necessary condition for a symmetry is that it must lead to a transformation that leaves the Hamiltonian invariant. The Hamiltonian is invariant under the transformation $\mathbf{X}_j \to \mathbf{X}_j$ and $\mathbf{Z}_j \to -\mathbf{Z}_j$.

That transformation is the spin-flip symmetry implemented by the operator $\mathbf{V} = \mathbf{X}_1\mathbf{X}_2 \ldots \mathbf{X}_L$. It commutes with the Hamiltonian, which means that it does not change over time. In other words, it's a conserved quantity.

Is there an additional symmetry in the Ising model? Another transformation that leaves the Hamiltonian invariant is

$$\mathbf{X}_j \to \mathbf{Z}_j\mathbf{Z}_{j+1}\,,\ \mathbf{Z}_j\mathbf{Z}_{j+1} \to \mathbf{X}_{j+1}\,, \tag{3}$$

which is known as the Kramers–Wannier transformation.[2] What exactly do the arrows in equation 3 mean? Even though the literature has commonly suggested that the Kramers–Wannier transformation is invertible, it's not. To see why, let us assume that the transformation is implemented by conjugating an operator by an invertible operator $\mathbf{U}$. The Kramers–Wannier transformation would thus be written as

$$\mathbf{U}\mathbf{X}_j\mathbf{U}^{-1} \stackrel{?}{=} \mathbf{Z}_j\mathbf{Z}_{j+1}\,,\ \mathbf{U}\mathbf{Z}_j\mathbf{Z}_{j+1}\mathbf{U}^{-1} \stackrel{?}{=} \mathbf{X}_{j+1}\,. \tag{4}$$

Let us apply the invertible transformation on the spin-flip operator $\mathbf{V}$ of the Ising model: $\mathbf{U}\mathbf{V}\mathbf{U}^{-1} = \mathbf{U}(\mathbf{X}_1\mathbf{X}_2 \ldots \mathbf{X}_L)\mathbf{U}^{-1} = (\mathbf{Z}_1\mathbf{Z}_2)(\mathbf{Z}_2\mathbf{Z}_3)\ldots(\mathbf{Z}_L\mathbf{Z}_1) = 1$, where in the last step, we have used $\mathbf{Z}_j^2 = 1$. When we multiply by $\mathbf{U}^{-1}$ from the left and $\mathbf{U}$ from the right, the spin-flip operator $\mathbf{V}$ becomes a trivial operator, which is a contradiction.

The Kramers–Wannier transformation, therefore, cannot be implemented by an invertible operator, as Wigner's theorem suggests. Rather, the meaning behind the arrows in equation 3 is answered by the following operator $\mathbf{D}$:[3]

$$\mathbf{D} = \underbrace{e^{-\frac{2\pi i L}{8}} \left(\prod_{j=1}^{L-1} e^{\frac{i\pi \mathbf{X}_j}{4}} e^{\frac{i\pi \mathbf{Z}_j\mathbf{Z}_{j+1}}{4}}\right) e^{\frac{i\pi \mathbf{X}_L}{4}}}_{\text{invertible but not conserved}} \times \underbrace{\frac{1 + \prod_{j=1}^{L} \mathbf{X}_j}{2}}_{\text{conserved but noninvertible}}\,. \tag{5}$$

It is rather complicated, but the only thing we need to know is that $\mathbf{D}$ is a product of an invertible but not conserved operator and a conserved but noninvertible operator. Because of the second factor, $\mathbf{D}$ is a noninvertible matrix that has some zero eigenvalues. The noninvertible operator implements the Kramers–Wannier transformation in the following precise sense:

$$\mathbf{D}\mathbf{X}_j = \mathbf{Z}_j\mathbf{Z}_{j+1}\mathbf{D}\,,\ \mathbf{D}\mathbf{Z}_j\mathbf{Z}_{j+1} = \mathbf{X}_{j+1}\mathbf{D}\,. \tag{6}$$

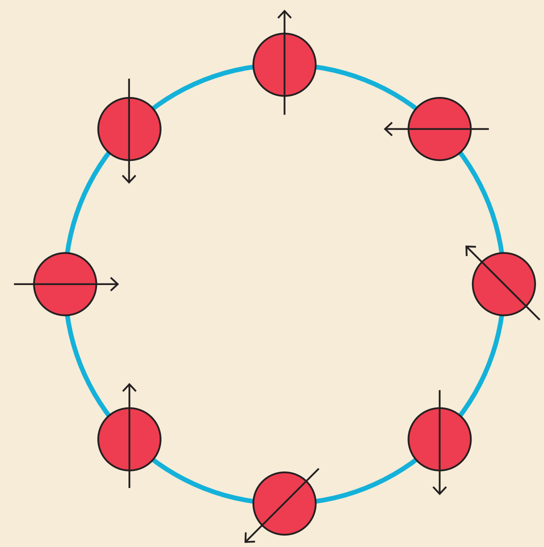


FIGURE 1. THE ISING MODEL consists of an array of qubits in one spatial dimension. Each one can be in a spin-up state $|\uparrow\rangle$, a spin-down state $|\downarrow\rangle$, or a superposition of the two. The model is an archetypal system that explores the differences between ordinary, invertible symmetries and noninvertible symmetries, which, once applied, cannot be undone. (Illustration by Three Ring Studio.)

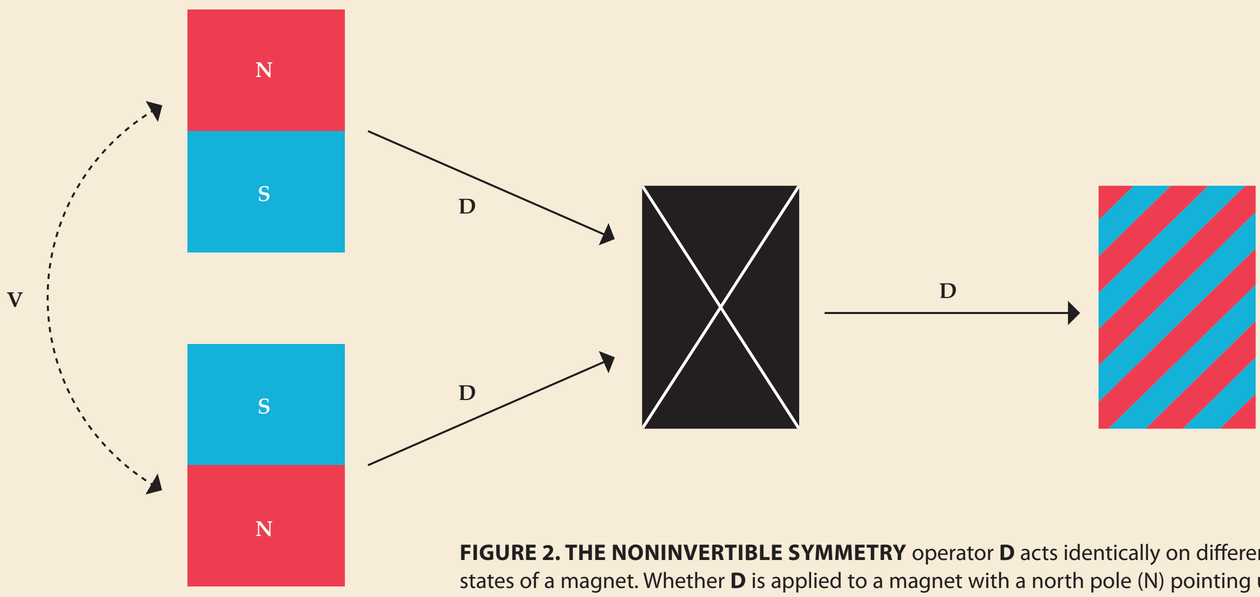


FIGURE 2. THE NONINVERTIBLE SYMMETRY operator **D** acts identically on different states of a magnet. Whether **D** is applied to a magnet with a north pole (N) pointing up or a south pole (S) pointing up, the result is the same: a demagnetized output state, shown in black. When **D** is applied a second time, the magnet enters a quantum superposition of the two initial states. (Illustration by Three Ring Studio.)

Since $\mathbf{D}^{-1}$ does not exist, however, the equation cannot be written in the form of equation 4, and thus no contradiction exists. The operator commutes with the Hamiltonian, and it is therefore a conserved quantity that does not change over time. It is a noninvertible symmetry.

The square of an ordinary symmetry is another symmetry: If a 90° rotation is applied twice, the result is a 180° rotation. But what about for the noninvertible symmetry? From equation 4, we see that applying **D** twice moves $\mathbf{X}_j$ forward to site $j + 1$. It appears that the noninvertible symmetry behaves like a lattice translation by half a site. That's inaccurate, however: Nothing exists between two lattice sites!

In fact, the operator **D**, the spin-flip operator **V**, and the one-site lattice translation operator **T** obey the algebraic relation $\mathbf{D} \times \mathbf{D} = \mathbf{T} \times (1 + \mathbf{V})/2$.

The action of **D** on a state, therefore, is conditioned on the response of that state to the spin-flip symmetry. The square of **D** corresponds to a one-site lattice translation **T** on states that are symmetric under **V**, such as the state in equation 1. But it is zero on states that flip signs under **V**. That algebra, which was recently derived,[3] does not fit in the mathematical framework of group theory and goes beyond the paradigm of Wigner's theorem. The new, noninvertible symmetry brings fresh insights into physical systems more generally.

So what is it good for?

One important task in condensed-matter physics is to characterize the phase diagram of a physical system. A familiar example is water at atmospheric pressure, which has gas, liquid, and solid phases. The task is often challenging because of strong interactions among the microscopic particles and atoms. Symmetry provides one of the few powerful analytic tools available to study such strongly coupled systems.

In particular, the noninvertible symmetries of systems that are invariant under the Kramers–Wannier transformation bring fresh insights into quantum systems. In the critical Ising model and a large class of related models, for example, magnetization and demagnetization coexist, which implies nontrivial entanglement properties. A formalization of that intuition[4-7] shows that the presence of a noninvertible symmetry forbids a featureless phase in which there is no entanglement. Moreover, the symmetry constrains the number of ground states with the lowest energy. Such a constraint would not have been possible if the Kramers–Wannier transformation were mistaken for an ordinary, invertible symmetry.

For many years, the discussion of noninvertible symmetries was confined to toy models in one spatial dimension, such as the Ising model for magnetization. A pair of papers[8,9] from a few years ago, however, led to many developments. Inspired by earlier work,[10] they introduced a construction of noninvertible symmetries applicable in three or more spatial dimensions. The key to the construction was the connection to another type of novel symmetry, known as the higher-form symmetry, that acts on extended objects such as strings.[11]

The ideas led to a rapid discovery of new symmetries across various physical systems, including quantum electrodynamics.[12,13] The symmetries provide tantalizing insights into other topics too, including particle physics; point to mistakes in the literature on scattering amplitudes;[14] and consolidate conjectures in quantum gravity and string theory. Beyond high-energy physics, noninvertible symmetries have also been applied to lattice models in condensed-matter theory and quantum information, which has led to the discovery of novel topological phases of quantum matter and constraints on phase diagrams. The new symmetries have emerged as a unifying language that brings together researchers of high-energy physics, condensed matter, and quantum information.

The idea of the noninvertible symmetry has emerged from interdisciplinary developments in physics and math. Because it goes beyond the framework of group theory, it calls for a new mathematical language. In some cases, the correct language has been identified as category theory, a generalization of group theory. Such advancements are fostering vibrant collaborations between mathematicians and physicists and mark a new chapter in the alignment of the two fields.

Throughout history, symmetries have contributed to major breakthroughs in physics. The symmetry discovered in 1941 by Kramers and Wannier[2] is now understood as a special example of noninvertible symmetries, and it predicted the critical temperature of the Ising model. The result encouraged Lars Onsager in 1944 to find the exact solution of the Ising model.[15] In recent years, the discovery and application of new noninvertible symmetries has led to a wave of progress across various areas of physics, and many more promising breakthroughs are on the horizon.

REFERENCES

1. D. Aasen, R. S. K. Mong, P. Fendley, *J. Phys. A: Math. Theor*. **49**, 354001 (2016).
2. H. A. Kramers, G. H. Wannier, *Phys. Rev*. **60**, 252 (1941).
3. N. Seiberg, S.-H. Shao, *SciPost Phys*. **16**, 064 (2024).
4. E. O'Brien, P. Fendley, *Phys. Rev. Lett*. **120**, 206403 (2018).
5. C.-M. Chang et al., *J. High Energy Phys*. **2019**, 26 (2019).
6. M. Levin, *Commun. Math. Phys*. **378**, 1081 (2020).
7. N. Seiberg, S. Seifnashri, S.-H. Shao, *SciPost Phys*. **16**, 154 (2024).
8. Y. Choi et al., *Phys. Rev. D* **105**, 125016 (2022).
9. J. Kaidi, K. Ohmori, Y. Zheng, *Phys. Rev. Lett*. **128**, 111601 (2022).
10. M. Koide, Y. Nagoya, S. Yamaguchi, *Prog. Theor. Exp. Phys*. **2022**, 013B03 (2022).
11. D. Gaiotto et al., *J. High Energy Phys*. **2015**, 172 (2015).
12. Y. Choi, H. T. Lam, S.-H. Shao, *Phys. Rev. Lett*. **129**, 161601 (2022).
13. C. Córdova, K. Ohmori, *Phys. Rev. X* **13**, 011034 (2023).
14. C. Copetti, L. Córdova, S. Komatsu, *Phys. Rev. Lett*. **133**, 181601 (2024).
15. L. Onsager, *Phys. Rev*. **65**, 117 (1944).

PT