AI Research Associate for Early-Stage Scientific Discovery

Morad Behandish[1]*, John T. Maxwell III[1], Johan de Kleer[1]
[1]Palo Alto Research Center (PARC)
3333 Coyote Hill Road, Palo Alto, California 94304 (www.parc.com)

Abstract

Artificial intelligence (AI) has been increasingly applied in scientific activities for decades; however, it is still far from an insightful and trustworthy collaborator in the scientific process. Most existing AI methods are either too simplistic to be useful in real problems faced by scientists or too domain-specialized (even dogmatized), stifling transformative discoveries or paradigm shifts. We present an AI research associate for early-stage scientific discovery based on (a) a novel minimally-biased ontology for physics-based modeling that is context-aware, interpretable, and generalizable across classical and relativistic physics; (b) automatic search for viable and parsimonious hypotheses, represented at a high-level (via domain-agnostic constructs) with built-in invariants, e.g., postulated forms of conservation principles implied by a presupposed spacetime topology; and (c) automatic compilation of the enumerated hypotheses to domain-specific, interpretable, and trainable/testable tensor-based computation graphs to learn phenomenological relations, e.g., constitutive or material laws, from sparse (and possibly noisy) data sets.

Introduction

Data-driven AI methods have been applied extensively in the past few decades to distill nontrivial physics-based insights (scientific discovery) and to predict complex dynamical behavior (scientific simulation) (Stevens et al. 2020). Notwithstanding their effectiveness and efficiency in classification, regression, and forecasting tasks, statistical learning methods can hardly ever evaluate the soundness of a function fit, explain the reasons behind observed correlations, or provide sufficiently strong guarantees to replace parsimonious and explainable scientific expressions such as differential equations (DE). Hybrid methods such as constructing "physics-informed/inspired/guided" architectures for neural nets and loss functions that penalize both predication and DE residual errors (Raissi, Perdikaris, and Karniadakis 2019; Wei and Chen 2019; Daw et al. 2020) and graph-nets based on control theory and combinatorial structures (Cranmer et al. 2019; Seo and Liu 2019; Sanchez-Gonzalez et al. 2020) are all important steps towards explainability; however, the built-in ontological biases in most machine learning (ML) frameworks prevent them from *thinking outside the box* to discover not only the known-unknowns, but also unknown-unknowns, during early stages of the scientific process.

*Corresponding Author, e-mail: moradbeh@parc.com.

Contributions

We present 'cyber-physicist' (**CyPhy**), our novel AI research associate for early-stage scientific process of hypothesis generation and initial (in)validation, grounded in the most invariable mathematical foundations of classical and relativistic physics. Our framework distinguishes itself from existing rule-based reasoning, statistical learning, and hybrid AI methods by:

(1) an ability to rapidly enumerate and test a diverse set of mathematically sound and parsimonious physical hypotheses, starting from a few basic assumptions on the embedding spacetime topology;

(2) a distinction between non-negotiable mathematical truism (e.g., conservation laws or symmetries), that are directly implied by properties of spacetime, and phenomenological relations (e.g., constitutive laws), whose characterization relies indisputably on empirical observation, justifying targeted use of data-driven methods (e.g., ML or polynomial regression); and

(3) a "simple-first" strategy (following Occam's razor) to search for new hypotheses by incrementally introducing latent variables that are expected to exist based on topological foundations of physics.

Background

AI-assisted discovery of scientific knowledge has been an active area of research (Langley 1998) long before the rise of GPU-accelerated deep learning (DL). As computational power and data sources are becoming more ubiquitous, model-based, data-driven, and hybrid AI methods are playing an increasingly more important role in various scientific activities (Kitano 2016; Raghu and Schmidt 2020).

Related efforts to our approach to scientific hypothesis generation and evaluation are mostly engineered after how humans approach scientific discovery, including sequential rule-based symbolic regression (Schmidt and Lipson 2009; Udrescu and Tegmark 2020), latent space representation learning via deep neural net auto-encoders (Iten et al. 2020;

Nautrup et al. 2020) and strategic combinations of divide-and-conquer, unsupervised learning, simplification by penalizing description lengths in the loss function, and a posteriori unification by clustering (Wu and Tegmark 2019). While these and other efforts have shown great promise for elevating AI to the role of an autonomous, creative, and insightful collaborator that can offer human scientists a set of viable options to consider, their applications have remained limited to rather basic examples.

On the more domain-specialized end, DL has been widely successful in classification, regression, and forecasting tasks in scientific areas as diverse as turbulence (Miyanawala and Jaiman 2017; Wang et al. 2020), chaotic particle dynamics (Breen et al. 2020), molecular chemistry and materials science (Butler et al. 2018), and protein engineering (Yang, Wu, and Arnold 2019), among others. Most specialized DL architectures are ad hoc, designed (by humans) using narrow, domain-specific, and (by construction) biased knowledge and expertise, stifling innovation and surprise. Moreover, DL models that successfully capture nontrivial patterns in data are often difficult to explain, lack guarantees even within their training space, and poorly extrapolate to out-of-training scenarios (Mehta et al. 2019). Training such models for high-dimensional physics problems requires enormous data, which is either unavailable or too costly to obtain in many experimental sciences.

The Cyber-Physicist

We introduce an AI tool that can bridge multiple levels of abstraction, using a *domain-agnostic* representation scheme to express a wide range of mathematically viable physical hypotheses by exploiting common structural *invariants* across physics. Our approach entails:

(a) defining a relatively unbiased ontology rooted in fundamental abstractions that are common to all known theories of classical and relativistic physics;

(b) constructing a constrained search space to enumerate viable hypotheses with postulated invariants, e.g., built-in conservation laws that are consistent with the presupposed spacetime topology; and

(c) automatically assembling interpretable ML architectures for each hypothesis, to estimate parameters for phenomenological relations from empirical data.

At the core of (a) is a powerful mathematical abstraction of physical governing equations rooted in algebraic topology and differential geometry (Frankel 2011). This abstraction leads to an ontological commitment to the relationship between physical measurement and basic properties of the embedding spacetime—but nothing more, to leave room for innovation and surprise. This relationship has been shown to be responsible for the *analogies* and *common structure* across physics (Tonti 2013), exploited in (b), along with search heuristics based on analogical reasoning. Each viable hypothesis is *automatically* compiled to an interpretable "computation graph"—tensor-based architecture, akin to a neural net with convolution layers to compute differentiation/integration and (non)linear local operators for constitutive equations—for a given cellular decomposition of embedding spacetime using well-established concepts from cellular homology (Hatcher 2001) and exterior calculus of differential and discrete forms (Bott and Tu 1982; Hirani 2003) that are under-utilized in AI.

Topological Foundations of Physics

The key enabler of our AI framework is a simple type system for (a) physical *variables*, based on how they are measured in spacetime; and (b) physical *relations*, based on their (topological vs. metric) nature, and the variables they connect. Following the ground-breaking discoveries by a number of mathematicians, physicists, and electrical engineers (Kron 1963; Roth 1955; Branin 1966) towards a general network theory, Tonti explained the fascinating analogies across classical and relativistic physics in his pioneering lifelong work (Tonti 2013) by reframing them in the language of cellular homology, leading to informal classification diagrams. Tonti diagrams can be formalized as directed graphs with strongly typed nodes for variables and edges for relations. The variable are typed as $(\mathrm{d}_1, \mathrm{d}_2)-$forms based on their measurement on d_1- and d_2-dimensional submanifolds (d_1- and d_2-cells) of space and time, respectively. For instance, to model heat transfer in (3+1)D spacetime, temperature is typed as a $(0,1)-$form because it is measured at spatial points ($0-$cells) and during temporal intervals ($1-$cells), whereas heat flux is typed as a $(2,1)-$form because it is measured over spatial surfaces ($2-$cells) and during temporal intervals ($1-$cells). In classical calculus, both of these variables reduce to scalar and vector fields, probed at spatial points and at temporal instants, to write down compact pointwise DEs; however, keeping track of the *topological and geometric character* of DEs is key to a deeper understanding of how known physical theories work, and building on top of it for AI-assisted discovery of new physics grounded in mathematical foundations.

The spatiotemporal cells (or embedding manifolds) are further classified as primary or secondary, endowed with inner or outer orientations, respectively, depending on how the variables change sign in a hypothetical reversal of spacetime orientation (Mattiussi 2000). The cells are related by topological *duality* (Fig. 1 (a)). For example, an inner-oriented curve ($1-$cell, $\overline{\sigma}^1$) sitting in primary space, along which temperature variations are measured, is dual to an outer-oriented surface ($2-$cell, $\widetilde{\sigma}^2$) sitting in secondary space, over which heat flux is measured, and the two cells are spatially registered and consistently oriented, if we embed them in two co-located "copies" of 3D space.

The relations among variables on the Tonti diagrams are typed based on the pairs of variables they relate, as well as the nature of the relation itself:

- *Topological* relations map spatiotemporal forms to forms of one higher dimension in space or time via incidence relations, and are responsible for propagation of information in spacetime through incident cells.
- *Metric* relations locally map forms defined over dual cells to one another based on phenomenological properties, spatial lengths, and temporal durations, and are responsible for local distortion of information.

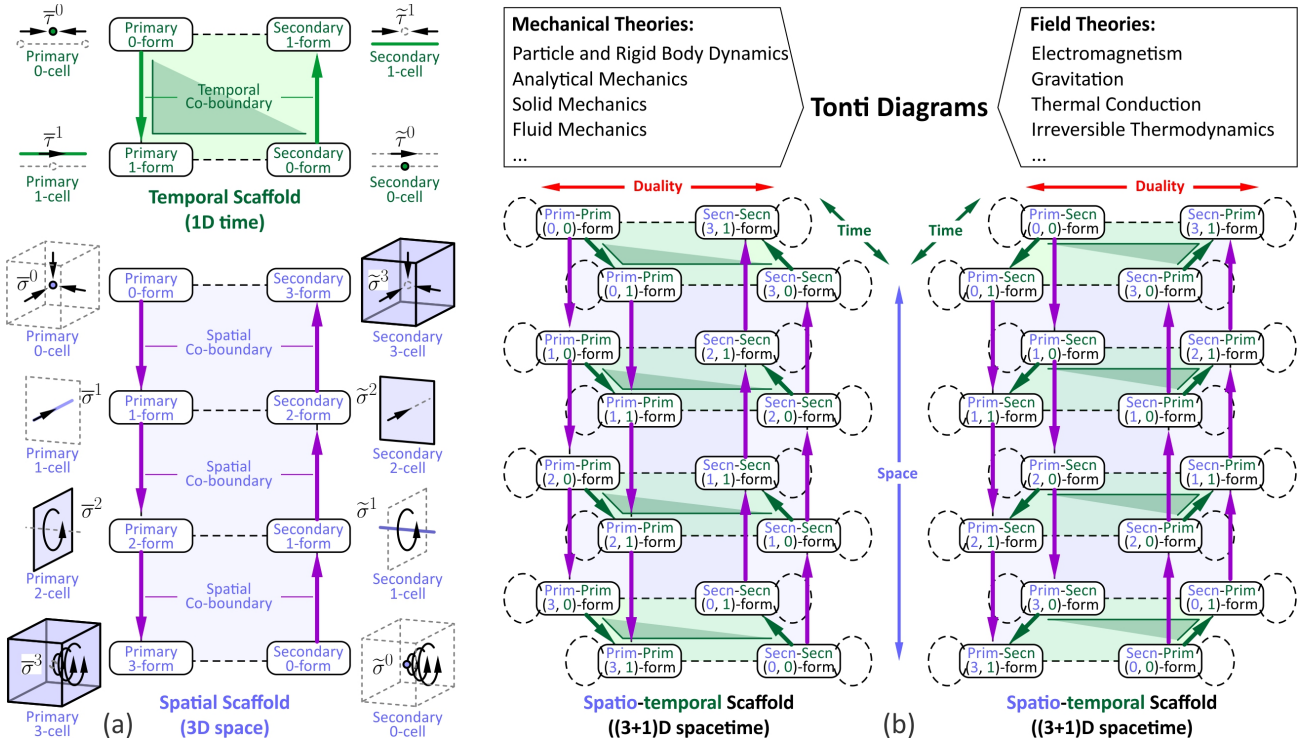

Figure 1: A topology-aware representation for physics (Tonti 2013): (a) variables associated with spatial and temporal cells of various dimensions give rise to primary forms and secondary forms (also called pseudo-forms); (b) resulting in 32 possible types for spatio-temporal forms, and an underlying structure for fundamental theories of physics.

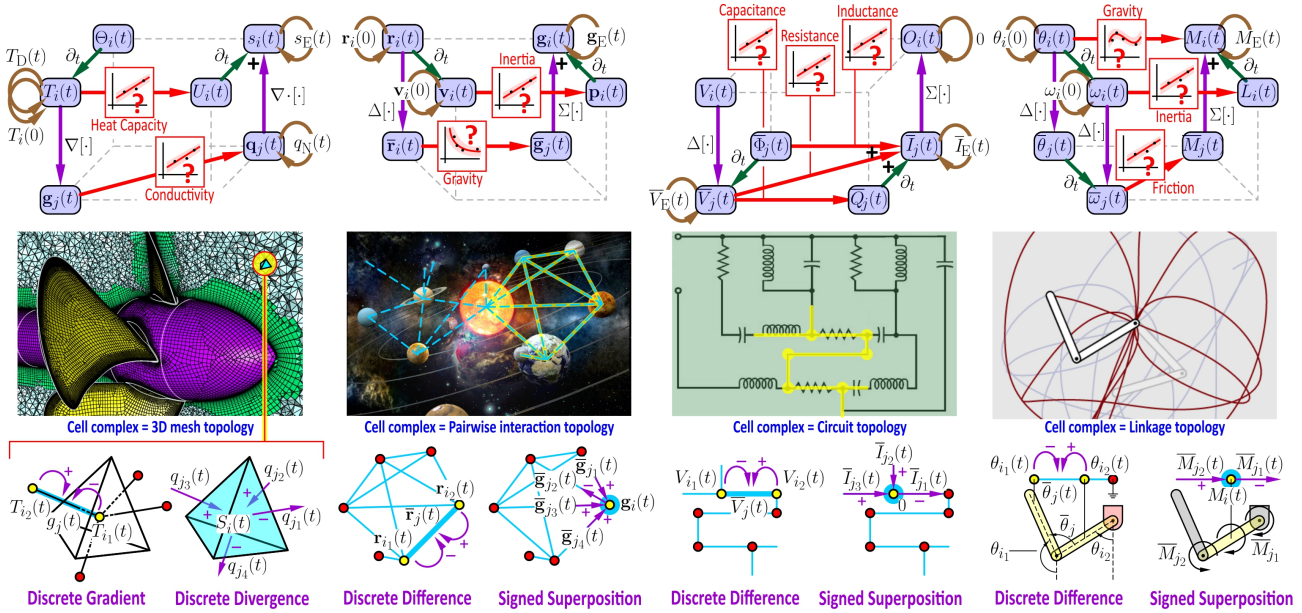

Figure 2: Tonti diagrams are recipes to generate governing equations in different contexts, defined by a continuum, discrete, or semi-discrete setting and a topological embedding of the variables based on how they are measured. The conservation laws in terms of co-boundary operators result directly from assumed properties of space (or spacetime), while constitutive relations must be learned from data (e.g., via regression/ML).

- *Algebraic* relations are in-place, i.e., map a given form to another from of the same type, and can be used to capture initial/boundary conditions, external source/sink terms, or cross-physics couplings between variables of the same type on different diagrams.

The relations are drawn in Fig. 1 as vertical arrows, horizontal (or horizontal-diagonal) arrows, and loops ($1-$cycles), respectively. The *interpretation* of these relations to symbolic or numerical operations depends on the choice of a cellular decomposition of spacetime on which they operate. For example, using a continuum spacetime with infinitesimal cells, the variables are viewed as *differential forms* and the topological operators on them are interpreted as *exterior derivatives* (Bott and Tu 1982). In elementary calculus, these operators give rise to gradient, curl, and divergence in space and partial derivative in time in terms of *scalar and vector fields* that are proxy to these forms, leading to partial DEs (PDEs). In a discrete (or semi-discrete) setting, on the other hand, the same diagram can be used to produce integral (or integro-differential) equations that capture the same fundamental conservation and constitutive realities, where the variables are viewed as *co-chains*, also called *discrete forms* (or mixed forms, e.g., discrete in space, differential in time, or vice versa) and the topological operators become *co-boundary operators* that are fundamental in cellular homology (Hatcher 2001). For example, using a semi-discretization in space with integral quantities associated with $0-$, $1-$, $2-$ and $3-$cells on a pair of staggered unstructured meshes in 3D, while keeping time as a continuum, the semi-discrete form of the heat equation as a system of ordinary DEs (ODEs) (Fig. 2 (a)). Upon discretization of time, one obtains algebraic equations that can be solved or parameter estimated via tensor-based ML.

It is important to note that 3D meshes in space and 1D time-stepping are not the only ways to provide a combinatorial topology to interpret Tonti diagrams in a discrete setting. Another example is a directed graph representation of lumped-parameter networks such as system models in Modelica or electrical circuits in Spice. The variables in this case are associated with nodes, edges, and meshes (i.e., primitive cycles) and incidence relations are obtained from graph connectivity and edge directions. The same topological operator that leads to a spatial divergence, discretized by a sum of fluxes on the incident faces of a volume in a 3D mesh, also leads to a superposition of forces on interacting planets, sum of currents in/out of junctions in electrical circuits, and superposition of torques on kinematic chains (Fig. 2 (b, c, d)). Both ODEs and PDEs and their integral or integro-differential forms upon full or semi-discretization can be captured with the same (abstract) operators, and Tonti diagrams serve as *recipes* to compose them to generate governing equations.

Figure 3 shows a few other examples of Tonti diagrams for fundamental theories in classical and relativistic physics. The differences amount to (a) topological and metric context; (b) relevant variables and their dimensions/units; and (c) phenomenological relations.

An Ontology for Scientific Process

We present a novel representation, called 'interaction networks' (**I-net**s), based on a generalization of Tonti diagrams that is expressive and versatile enough to accommodate novel scientific hypotheses, while retaining a basic commitment to philosophical principles such as parsimony (Occam's razor), measurement-driven classification of variables, and separation of non-negotiable mathematical properties of spacetime (homology) from domain-specific empirical knowledge (phenomenology). Data science is employed to help only with the latter.

- We conceptualize three levels of abstraction related by inheritance: abstract (symbolic) **I-net**s $\rightarrow$ discrete (cellular) **I-net**s $\rightarrow$ numerical (tensor-based) **I-net**s.
- At each level, an **I-net** instance is contextualized by user-defined assumptions on spacetime topology, semantics of physical quantities, and structural restrictions on allowable diagrams based on analogical reasoning and domain-specific insight (if available).
- Every **I-net** instance distinguishes between topological and metric operators; however, it has additional degrees of freedom (beyond Tonti diagrams) for the latter to allow for phenomenological relations among variables that may not be dual to each other.

The latter is motivated by the observation that some existing middle-ground theories use phenomenological relations to capture a combination of topological and metric aspects.

We define an abstract (symbolic) **I-net** on a single $\mathscr{D}-$space as a finite collection of primary and/or secondary co-chain complexes that are inter-connected by phenomenological links, as shown in Fig. 4 (a). Each co-chain complex is a sequence of (symbolic) $\mathrm{d}-$forms related by (symbolic) co-boundary operators from $\mathrm{d}-$forms to $(\mathrm{d}+1)-$forms $(0 \leq \mathrm{d} \leq \mathscr{D})$. The interpretation of $\mathrm{d} \rightarrow (\mathrm{d}+1)$ maps depends on the embedding dimension $\mathscr{D}$; for instance, if $\mathscr{D} = 1$ the only option for the input is $\mathrm{d} = 0$ leading to a simple partial derivative $(0 \rightarrow 1)$, whereas for $\mathscr{D} = 3$, we can have $\mathrm{d} = 0, 1, 2$ leading to gradient $(0 \rightarrow 1)$, curl $(1 \rightarrow 2)$, and divergence $(2 \rightarrow 3)$ operations, respectively.

These sequences may represent different (mechanical, electrical, thermal, etc.) domains of physics. Although, for most known physics, each domain's theory appears as one pair of (primary and secondary) sequences in tandem, connected by horizontal (or horizontal-diagonal) constitutive relations leading to Tonti diagrams, we do not make any such restriction when looking for new theories. The cross-sequence links can thus represent both single-physics constitutive relations and mutli-physics coupling interactions. Conservation laws, on the other hand, are represented by a balance between the output of a topological operator and an external source/sink, the latter being represented by a loop.

It is often more convenient to define product spaces (e.g., separate 3D space and 1D time, as opposed to 4D spacetime) in which conservation laws are stated as sums of incoming topological relations being balanced against an external source/sink. To accommodate such representations, we define abstract (symbolic) **I-net**s on a product of a $\mathscr{D}_1-$space

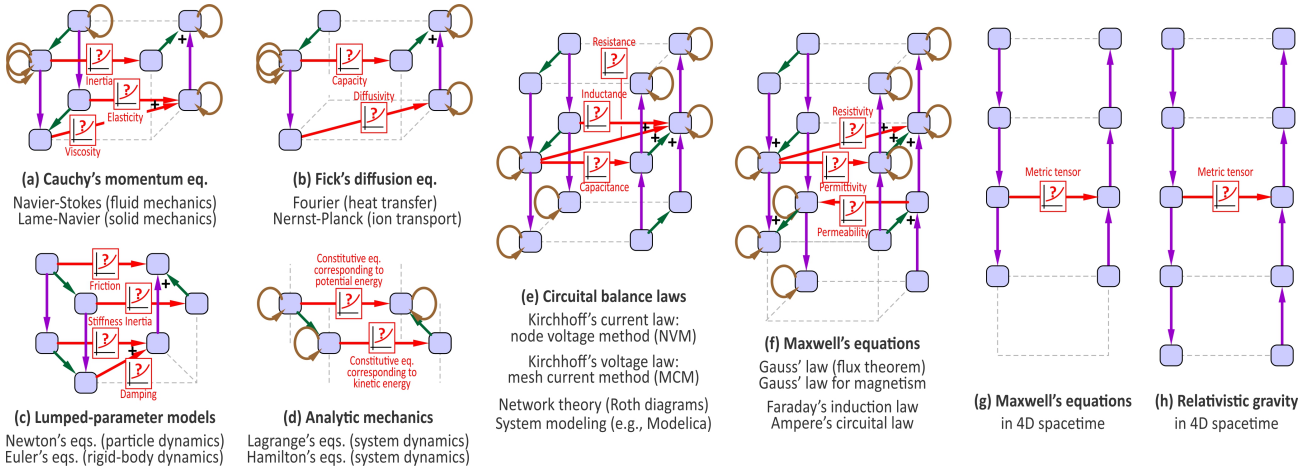

Figure 3: Tonti diagrams capture the common structure responsible for analogies across classical and relativistic physics with a clear distinction between topological and phenomenological relations that follow certain rules.

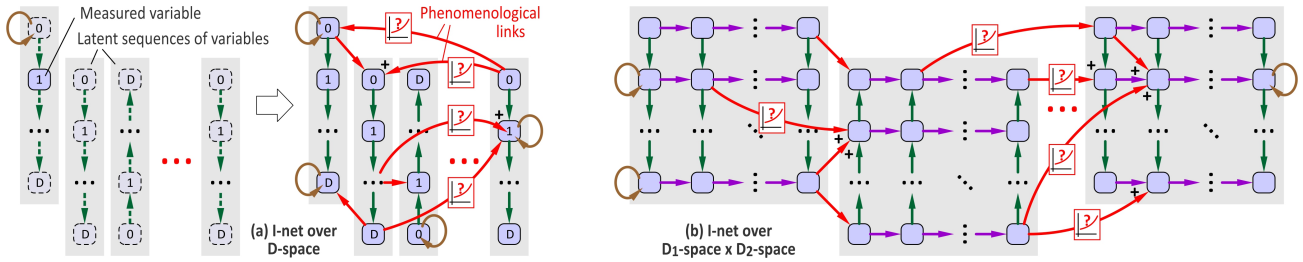

Figure 4: **I-net**s are generalizations of Tonti diagrams for finite topological products of finite-dimensional spaces with relaxed rules for feasible phenomenological links to accomodate middle-ground theories.

and a $\mathscr{D}_2-$space as multi-sequences of co-chains, connected by phenomenological links, as before. It is possible to form $2^2 = 4$ possible such multi-sequences with various orientation combinations, two of which lead to so-called mechanical and field theories (Tonti 2013), shown in Fig. 1 for (3+1)D spacetime and repeated in Fig. 4 (b) for higher-dimensional pairs of abstract topological spaces. This construction is generalized to products of more than two spaces in a straightforward combinatorial fashion.

Based on the topological context, the semantics for co-boundary operators is unambiguously determined by the dimensions of the two variables (i.e., co-chains) they relate. However, phenomenological links require specifying a parameterization of possibly nonlinear, in-place, and purely metric relations they represent, using unknown parameters that must be learned from data.

Once one or more hypotheses are specified in the language of abstract (symbolic) **I-net**s with unknown phenomenological parameters (e.g., thermal conductivity in the earlier heat transfer example), the parameters can be optimized to fit the data and the regression error can be used to evaluate the fitness of hypotheses.

A Search for Viable Hypotheses

Having defined a combinatorial representation of viable hypotheses that are partially ordered in terms of complexity, the next step is to generate and test the hypotheses in a "simple-first" fashion. The search space is defined by a directed acyclic graph (DAG) whose nodes (i.e., 'states') represent symbolic **I-net** instances. The edges (i.e., state transitions) represent generating a new **I-net** structure by incrementally adding complexity to the parent state. Each action can be one or composition of (a) defining a new symbolic variable, in an existing co-chain complex, by applying a topological operator to an existing variable; (b) defining a new variable in a latent co-chain complex; and (c) adding phenomenological links of prescribed form and unknown parameters, connecting existing variables. The search is guided by a loss function determined by how well the hypotheses represented by these **I-net** structures explain a given dataset. The algorithm may also be equipped with user-specified heuristic rules to prune the search space or prioritize paths that are perceived as "more likely" due to structural analogies with existing theories.

The input to the search algorithm includes the bare minimum contextual information such as the assumed underlying topology, a preset number of physical domains, and the types of measured variables, e.g., spatiotemporal associations, tensor ranks and shapes, and dimensions/units. The search starts from an "initial" **I-net** instance (i.e., the 'root') that embodies only measured variable(s) with no initial edges except the ones that are asserted a priori, e.g.,

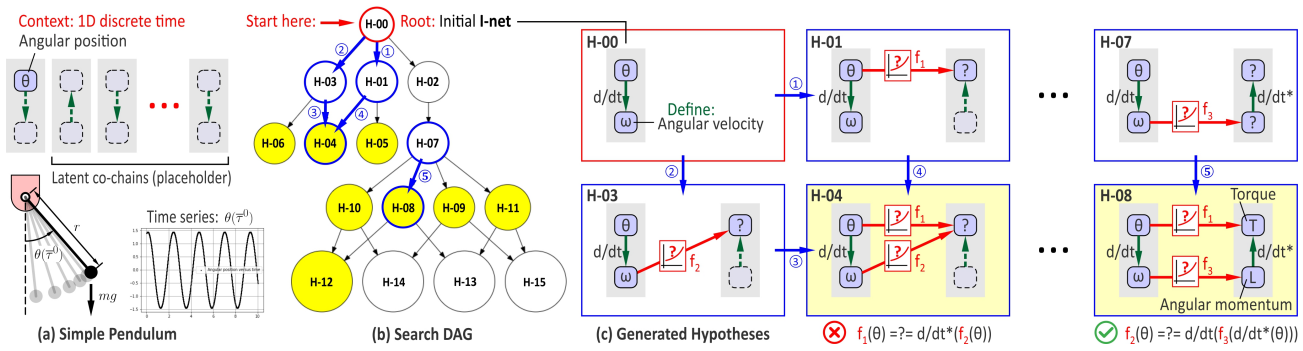

Figure 5: The search space for the dynamics of a pendulum in 1D time. The complete hypotheses (yellow nodes) correspond to **I-net** structures that pose new nontrivial equations to be tested against data, whereas incomplete hypotheses (white nodes) have "dangling" branches that are completed in their child states.

ID	Equation	Penalty	Test Error
H-04	1 = -0.90659802·cos(θ(t)) + 0.04620761·(d/dt θ(t))^2	2.1	0.00079403
H-08	sin(θ(t)) = -0.10048319·(d/dt^2 θ(t))	1.6	0.51283429
H-14	d/dt sin(d/dt θ(t)) = -0.03675202·(d/dt^2 θ(t)) + -1418.61437500 + -2.18736535·cos(d/dt θ(t)) + -49.82946585·θ(t) + -568.20850566·(d/dt θ(t))^2 + 11203.60425779·cos(θ(t)) + 24.59512553·θ(t)^2 + 65.76742005·sin(θ(t))	3.0	0.97938609
H-09	θ(t) = -0.01661234·(d/dt sin(d/dt θ(t))) + -0.13494041·(d/dt^2 θ(t))	1.6	18.54692244
H-06	1 = -0.04373119·sin(d/dt θ(t)) + 0.03373455·θ(t) + 0.10676103·(d/dt θ(t))^2 + 0.5080396·cos(d/dt θ(t))	2.1	34.22083192
H-10	cos(d/dt θ(t)) = -0.00528888·(d/dt* sin(d/dt θ(t))) + -0.08040678·(d/dt θ(t))^2 + 0.32158535	2.7	34.23619927
H-05	d/dt θ(t) = -0.75787578·θ(t)^2 + -0.90966452·sin(θ(t)) + -2.28240998·cos(θ(t)) + 0.61502711·θ(t) + 1.8389267	2.1	70.41833973
H-11	d/dt* sin(d/dt θ(t)) = -0.04651204·(d/dt θ(t)) + -0.09307301·(d/dt d/dt* θ(t)) + 0.0073915·(d/dt* (d/dt θ(t))^2) + 0.02052373·(d/dt* cos(d/dt θ(t)))	2.7	771.1561706

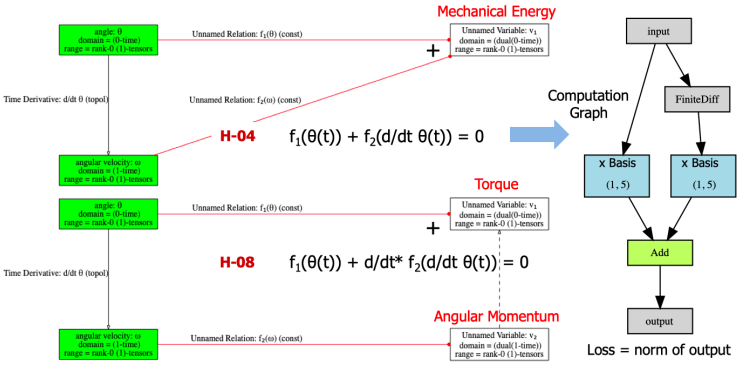

Figure 6: The hypotheses H-04 and H-08 of Fig. 5 are enumerated and visualized by the software and evaluated against data (split 0.7-0.3 for training/testing). Both energy (first-order) and and torque (second-order) forms of the governing equation are discovered without human intervention. The former was quite unexpected, since its **I-net** structure does not correspond to a Tonti diagram. The latter has a larger error due to finite difference discretization.

loops for initial/boundary conditions or source terms, if applicable. The spatio-temporal types and physical semantics for these variables are provided by the experimentalist.

For example, consider a simple pendulum (Fig. 5 (a)). We have only 1D time, leading to a topological space of interconnected time instants $\overline{\tau}^0, \widetilde{\tau}^0 = \overline{\tau}^0 + \epsilon/2$ and time intervals $\overline{\tau}^1 = (\overline{\tau}^0, \overline{\tau}^0 + \epsilon), \widetilde{\tau}^1 = (\widetilde{\tau}^1, \widetilde{\tau}^1 + \epsilon)$ to which data may be associated. Suppose we are given time series data for angular position $\theta(\overline{\tau}^0)$. The initial **I-net** instance is a single symbolic variable for this $0-$form, which can be differentiated only once in primary 1D time to obtain angular velocity as a $1-$form: $\theta(\overline{\tau}^0) \rightarrow \omega(\overline{\tau}^1) = \delta[\theta](\overline{\tau}^1)$ at the root of the search DAG (Fig. 5 (b)). The DAG is expanded by adding new phenomenological links, either between two existing variables, or between an existing variable and one in a newly added latent co-chain sequence (Fig. 5 (c)). In this example, the hypotheses are numbered H-00 (the root) through H-15, enumerating all possible **I-net** structures formed by at most one latent co-chain complex in 1D time. The user can specify the maximum number of latent variables that the algorithm may consider, to keep the search tractable.

Not every introduction of new variables or relations makes nontrivial statements about physics. For example, the hypothesis H-01 produces a new variable typed as a $1-$pseudo-form $T(\widetilde{\tau}^1) = \not{f}_1(\theta(*\widetilde{\tau}^1))$, where the $*-$operator takes $\widetilde{\tau}^1$ to its dual: $*(\widetilde{\tau}^0, \widetilde{\tau}^0 + \epsilon) = \widetilde{\tau}^0 + \epsilon/2$. However, until this new variable is reached through another path to close a cycle and pose a nontrivial equation, we do not have a complete hypothesis to (in)validate against data. Further down the search DAG, H-08 defines a new variable typed as a $0-$pseudo-form $L(\widetilde{\tau}^0) = \not{f}_2(\omega(*\widetilde{\tau}^0))$ where $*\widetilde{\tau}^0 = (\widetilde{\tau}^0 - \epsilon/2, \widetilde{\tau}^0 + \epsilon/2)$. The co-boundary operation $L(\widetilde{\tau}^0) \rightarrow T(\widetilde{\tau}^1) = \delta[L](\widetilde{\tau}^1)$, closes the cycle and produces a commutative diagram (Fig. 5 (c)) leading to:

$$\mathscr{E}_{\text{H-08}}(\theta; \not{f}_1, \not{f}_2) = \not{f}_1(\theta) - \delta^*[\not{f}_2(\delta[\theta])] = 0, \qquad (1)$$

where $\not{f}_1, \not{f}_2$ are selected from restricted function spaces $\mathscr{F}_1, \mathscr{F}_2$ to avoid overfitting (e.g., parameterized by a linear combination of domain-aware basis functions) and their parameters must be determined from data to minimize the residual error $\mathscr{E}_{\text{H-08}}$ over the entire period of data collection. A loss function can, for example, be defined as a mean-squared-error (MSE) to penalize violations uniformly over the time series period:

$$\text{Loss}_{\text{H-08}} = \min_{\not{f}_1 \in \mathscr{F}_1} \min_{\not{f}_2 \in \mathscr{F}_2} \left\| \mathscr{E}_{\text{H-08}}(\theta; \not{f}_1, \not{f}_2) \right\|_{\widetilde{\tau}^1}, \qquad (2)$$

where $\|\cdot\|_{\widetilde{\tau}^1}$ is an L_2-norm computed as a temporal integral, i.e., sum of squared errors $\mathscr{E}^2_{\text{H-08}}(\theta;\ell_1,\ell_2)$ over time intervals $\widetilde{\tau}^1$ where (1) is evaluated. In this example, it turns out that the best fit is achieved with $\ell_1(\theta) = c_1 \sin\theta$ and $\ell_2(\omega) = c_2\omega$ where $c_2/c_1 = -g/r$. The latent variables $L(\widetilde{\tau}^0)$ and $L(\widetilde{\tau}^0)$ turn out to be the familiar notions of angular momentum and torque, respectively, although the software need not know anything about them to generate and test what-if scenarios about their existence and correlations with angular position and velocity. Hence, *interpretability* of the discovered relationships by a human scientist does not require predisposing the AI associate to such interpretations, enabling unexpected discoveries.

In general, every state in the search DAG can be classified as complete or incomplete hypotheses. The former are **I-net** structures with "dangling" branches that carry no new nontrivial information in addition to their parent states. Every time such a branch is turned into one or more closed cycles by adding enough new variables and/or relations, a new constraint is hypothesized that can be evaluated against data. When adding new dangling branches to the **I-net** structure, the search algorithm prioritizes actions that produce **I-net** structures similar to existing Tonti diagrams by assigning a penalty factor to every violation of the common structure (e.g., diagonal phenomenological links connecting non-dual cells). The loss for complete hypotheses can be computed as the sum of penalties for the **I-net** structure and the sum of residual errors for each of the independent constraints, implied by converging paths, multiplied by use-specified relative weight of the penalties and errors. We use an A* algorithm to search the space of hypotheses. Since we cannot compute the error for incomplete hypotheses, we can only prune them when the increase in their penalty is large enough that it would fail even if it had no error at all.

Generating Symbolic Expressions

One of the practical features of our implementation in Python is its ability to automatically convert **I-net** instances to symbolic DE expressions in SymPy, when the co-boundary operators are interpreted in a differential setting for infinitesimal cells ($\epsilon \to 0^+$); for example, equation (1) can be rewritten as a nonlinear ODE:

$$\mathscr{E}_{\text{H-08}}(\theta;\ell_1,\ell_2) = \ell_1(\theta) - \frac{\partial}{\partial t}\left[\ell_2\left(\dot{\theta}(t)\right)\right]. \qquad (3)$$

As a result, the generated hypotheses can be evaluated using any number of existing ML or symbolic regression frameworks that standardize on ODE/PDE inputs. For example, using non-orthogonal basis functions $\{1, x, x^2, \sin x, \cos x\}$ to span both function spaces $\mathscr{F}_1, \mathscr{F}_2$, we can substitute for both symbolic functions:

$$\ell_1(\theta) := c_0^1 + c_1^1\theta + c_2^1\theta^2 + c_3^1\sin\theta + c_3^1\cos\theta, \qquad (4)$$

$$\ell_2(\dot{\theta}) := c_0^2 + c_1^2\dot{\theta} + c_2^2\dot{\theta}^2 + c_3^2\sin\dot{\theta} + c_3^2\cos\dot{\theta}, \qquad (5)$$

into (3) to obtain a symbolic second-order (non)linear ODE in SymPy. Next, the software performs algebraic simplification to identify *equivalence classes* of hypotheses that, despite coming from different **I-net** structures, lead to the same ODE upon differential interpretation of the **I-net**s. For ODEs which, after simplification, are linear combinations of nonlinear (differential/algebraic) terms that are computable from data, we can apply symbolic regression to estimate the coefficients from data; for example, we tried LASSO-regularized least-squares regression in PDE-FIND (Rudy et al. 2017) where each term involving a derivative is evaluated using finite difference or polynomial approximation, whose results are shown in Fig. 6.

There are at least two issues with this approach:

First, more sophisticated regression or nonlinear programming methods are needed if the DE has terms that have nested nonlinear functions, i.e., cannot be represented as a linear combination of nonlinear terms because of unknown coefficients embedded within each term. We solve this problem by directly mapping **I-net** structures to computation graphs in PyTorch, skipping differential interpretation to symbolic DEs altogether.

Second, numerical approximation of symbolic PDEs is a tricky business, as the discrete forms (in 3D space) may not obey the conservation principles postulated by the **I-net** structure after such approximations. It is difficult to separate discretization errors from modeling errors and noise in data. One of the key advantages of **I-net**s is the rich geometric information in their type system that is fundamental to physics-compatible and mimetic discretization schemes (Koren et al. 2014; Palha et al. 2014; Lipnikov, Manzini, and Shashkov 2014) that ensure conservation laws are satisfied *exactly* as a discrete level, regardless of spatial mesh or time-step resolutions. Such information is lost upon conversion to symbolic DEs. Retaining this information is even more important when dealing with noisy data, because discrete differentiation of noisy data (e.g., via finite difference or polynomial fitting) can substantially amplify the noise.

The good news is that we can *directly* interpret the same **I-net** instance in integral form to generate equations over larger regions in space and/or time, to make the computations more resilient to noise. For example, in the heat equation, the discrete divergence of heat flux over a single $3-$cell is replaced by a flux integral over a collection of $3-$cells, and is equated against the volumetric intgeral of internal energy within the collection. The cancellation of internal surface fluxes (discrete form of Gauss' divergence theorem) is built into the interpretation based on cellular homology. The integrals can be computed using higher-order integration schemes, e.g., using polynomial interpolation with underfitting to filter the noise.

Further details on directly and automatically mapping the abstract (symbolic) **I-net** structures to discrete (cellular) and numerical (tensor-based) **I-net** instance (e.g., computation graphs in PyTorch), learning scale-aware phenomenological relations, and physics-compatible discretization and denoising will be presented in a full paper.

Real-World Scientific Discovery

Figures 7 and 8 illustrate the application of our AI approach to an elastodynamics challenge problem provided by AFRL in the course of the DARPA AI Research Associate (AIRA)

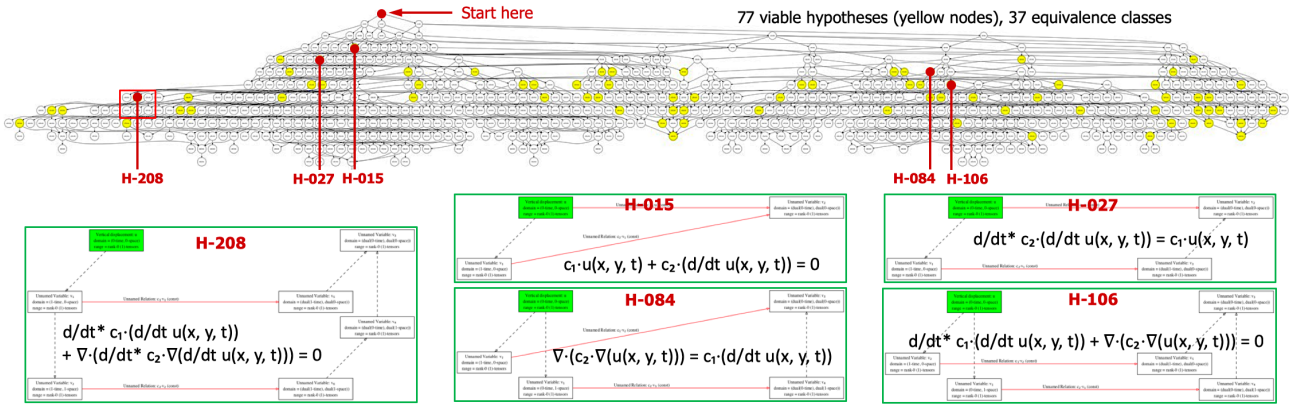

Figure 7: The search DAG and a number of viable hypotheses to explain ultrasound wavefield in metal parts.

ID	Penalty	Equation (symbolic PDE form)	RMSE
365	5.3	d/dt* c_1·(d/dt u(x, y, t)) + ∇·(d/dt* c_4·∇(d/dt u(x, y, t))) = c_1·u(x, y, t) + c_2·(d/dt u(x, y, t))	9.40E-07
237	3.9	d/dt* c_2·(d/dt u(x, y, t)) + ∇·(d/dt* c_3·∇(d/dt u(x, y, t))) = c_1·u(x, y, t)	9.40E-07
331	5.0	d/dt* c_2·(d/dt u(x, y, t)) + ∇·(d/dt* c_3·∇(d/dt u(x, y, t))) = c_1·(d/dt u(x, y, t))	1.05E-06
208	3.6	d/dt* c_1·(d/dt u(x, y, t)) + ∇·(d/dt* c_2·∇(d/dt u(x, y, t))) = 0	1.05E-06
301	4.8	∇·(d/dt* c_2·(d/dt ∇(u(x, y, t)))) = c_1·u(x, y, t) + c_2·(d/dt u(x, y, t))	1.52E-06
121	3.0	∇·(d/dt* c_2·∇(d/dt u(x, y, t))) = c_1·u(x, y, t)	1.52E-06
171	4.1	∇·(d/dt* c_2·∇(d/dt u(x, y, t))) = c_1·(d/dt u(x, y, t))	1.60E-06
215	3.7	d/dt* ∇·(c_1·∇(d/dt u(x, y, t))) + ∇·(d/dt* c_1·∇(d/dt u(x, y, t))) = 0	1.60E-06
230	4.8	d/dt* c_2·∇(d/dt u(x, y, t)) = c_1·∇(u(x, y, t)) + c_3·∇(d/dt u(x, y, t))	2.08E-06
116	3.9	d/dt* c_1·∇(d/dt u(x, y, t)) = c_2·∇(d/dt u(x, y, t))	2.22E-06
72	2.8	d/dt* c_2·(d/dt ∇(u(x, y, t))) = c_1·∇(u(x, y, t))	2.29E-06
468	6.5	d/dt* c_2·(d/dt u(x, y, t)) + ∇·(c_4·∇(u(x, y, t)) + c_5·∇(d/dt u(x, y, t))) = c_1·u(x, y, t) + c_3·(d/dt u(x, y, t))	6.87E-06
443	6.2	d/dt* c_2·(d/dt u(x, y, t)) + ∇·(c_3·∇(u(x, y, t)) + c_4·∇(d/dt u(x, y, t))) = c_1·(d/dt u(x, y, t))	6.88E-06
284	5.6	d/dt* c_2·(d/dt u(x, y, t)) + ∇·(c_4·∇(d/dt u(x, y, t))) = c_1·u(x, y, t) + c_2·(d/dt u(x, y, t))	6.90E-06
265	5.3	d/dt* c_2·(d/dt u(x, y, t)) + ∇·(c_3·∇(d/dt u(x, y, t))) = c_1·(d/dt u(x, y, t))	6.91E-06
332	5.1	d/dt* c_2·(d/dt u(x, y, t)) + ∇·(c_3·∇(u(x, y, t)) + c_4·∇(d/dt u(x, y, t))) = c_1·u(x, y, t)	9.79E-06
296	4.8	d/dt* c_1·(d/dt u(x, y, t)) + ∇·(c_2·∇(u(x, y, t)) + c_3·∇(d/dt u(x, y, t))) = 0	9.80E-06
180	4.2	d/dt* c_2·(d/dt u(x, y, t)) + ∇·(c_3·∇(d/dt u(x, y, t))) = c_1·u(x, y, t)	9.83E-06
159	3.9	d/dt* c_1·(d/dt u(x, y, t)) + ∇·(c_2·∇(d/dt u(x, y, t))) = 0	9.83E-06
191	4.3	d/dt* c_3·(d/dt u(x, y, t)) + ∇·(c_4·∇(u(x, y, t))) = c_1·u(x, y, t) + c_2·(d/dt u(x, y, t))	2.77E-05
120	2.9	d/dt* c_2·(d/dt u(x, y, t)) + ∇·(c_3·∇(u(x, y, t))) = c_1·u(x, y, t)	2.77E-05
166	4.0	d/dt* c_2·(d/dt u(x, y, t)) + ∇·(c_3·∇(u(x, y, t))) = c_1·(d/dt u(x, y, t))	2.88E-05
106	2.6	d/dt* c_1·(d/dt u(x, y, t)) + ∇·(c_2·∇(u(x, y, t))) = 0	2.88E-05
278	5.6	∇·(c_3·∇(u(x, y, t)) + c_4·∇(d/dt u(x, y, t))) = c_1·u(x, y, t) + c_2·(d/dt u(x, y, t))	3.00E-05
247	5.1	∇·(c_3·(d/dt ∇(u(x, y, t)))) = c_1·u(x, y, t) + c_2·(d/dt u(x, y, t))	3.00E-05
71	2.8	∇·(c_3·∇(d/dt u(x, y, t))) = c_1·(d/dt u(x, y, t))	3.00E-05
259	5.3	∇·(c_3·∇(u(x, y, t)) + c_4·∇(d/dt u(x, y, t))) = c_1·(d/dt u(x, y, t))	3.00E-05
115	3.8	∇·(c_2·∇(u(x, y, t)) + c_3·(d/dt ∇(u(x, y, t)))) = c_1·u(x, y, t)	4.71E-05
86	3.3	∇·(c_2·∇(d/dt u(x, y, t))) = c_1·u(x, y, t)	4.72E-05
58	3.1	c_1·∇(u(x, y, t)) + c_2·(d/dt ∇(u(x, y, t))) = 0	7.79E-05
55	3.0	d/dt* c_3·(d/dt u(x, y, t)) = c_1·u(x, y, t) + c_2·(d/dt u(x, y, t))	0.00053997
27	1.6	d/dt* c_2·(d/dt u(x, y, t)) = c_1·u(x, y, t)	0.00053997
46	2.7	d/dt* c_1·(d/dt u(x, y, t)) = c_2·(d/dt u(x, y, t))	0.00059487
28	1.6	∇·(c_2·∇(u(x, y, t))) = c_1·u(x, y, t)	0.000671
95	3.4	∇·(c_3·∇(u(x, y, t))) = c_1·u(x, y, t) + c_2·(d/dt u(x, y, t))	0.00068679
84	3.1	∇·(c_2·∇(u(x, y, t))) = c_1·(d/dt u(x, y, t))	0.0007597
15	2.1	c_1·u(x, y, t) + c_2·(d/dt u(x, y, t)) = 0	0.019795

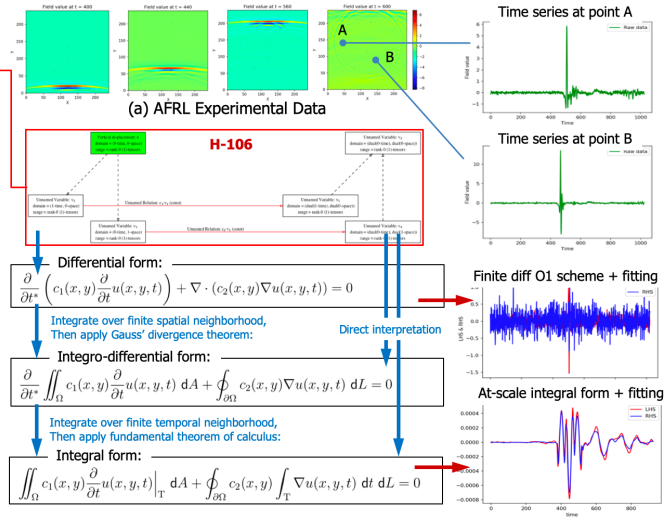

Figure 8: The AI associate discovers the (integral form) of the wave equation as well as the proper length/time scale at which the heterogeneous material properties (in this case, speed of sound) must be defined.

program that supported the development of **CyPhy**. The input is noisy data obtained by ultrasound imaging, measured in (2+1)D spacetime over the surface of several material samples with heterogeneous properties.

Figure 7 illustrates the search DAG along with a number of **I-net** structures for viable hypotheses, each postulating the relevance of a conservation law and existence of a few phenomenological relations. Figure 8 shows the ranking of these hypotheses based on their residual errors when tested against data. Each hypothesis can be interpreted in differential, integral, or integro-differential forms. The results demonstrate that integral forms applied to wide spatial and temporal neighborhoods (of $\sim$ 25 grid elements along each axis) with high-order polynomial underfitting (up to cubic in each coordinate), resulting in a length/time scale-aware definition of (nonlocal) phenomenological relations as well as physics-compatible (i.e., mimetic) discretization and de-noising, are preferable to strictly local numerical schemes such as finite difference discretization.

Conclusion

Statistical learning methods, despite their accuracy and efficiency in narrow regimes for which they are carefully engineered, are not sufficient to independently acquire *deep understandings* of the scientific problems they are applied to. Human scientists continue to handle most of knowledge-centric aspects of the scientific process based on domain-specific insight, experience, and expertise.

Our novel approach to early-stage scientific hypothesis generation and testing demonstrates a path forward towards context-aware, generalizable, and interpretable AI for scientific discovery. Our AI associate (**CyPhy**) distinguishes between non-negotiable mathematical truism, implied by the relationship between measurement and presupposed spacetime topology, and phenomenological realities that are at the mercy of empirical learning. Data-driven regression is targeted at the latter to enable distilling governing equations from sparse and noisy data, while providing deep insights into the mathematical foundations.

Acknowledgment

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Agreement No. HR00111990029.

References

Bott, R.; and Tu, L. W. 1982. *Differential Forms in Algebraic Topology*. Springer Science & Business Media.

Branin, F. H. 1966. The Algebraic-Topological Basis for Network Analogies and the Vector Calculus. In *Symposium on Generalized Networks*, 453–491. Polytechnic Institute of Brooklyn, NY.

Breen, P. G.; Foley, C. N.; Boekholt, T.; and Zwart, S. P. 2020. Newton versus the Machine: Solving the Chaotic Three-body Problem Using Deep Neural Networks. *Monthly Notices of the Royal Astronomical Society* 494(2): 2465–2470.

Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; and Walsh, A. 2018. Machine Learning for Molecular and Materials Science. *Nature* 559(7715): 547–555. doi:10.1038/s41586-018-0337-2.

Cranmer, M. D.; Xu, R.; Battaglia, P.; and Ho, S. 2019. Learning Symbolic Physics with Graph Networks. *arXiv preprint arXiv:1909.05862* .

Daw, A.; Thomas, R. Q.; Carey, C. C.; Read, J. S.; Appling, A. P.; and Karpatne, A. 2020. Physics-Guided Architecture (PGA) of Neural Networks for Quantifying Uncertainty in Lake Temperature Modeling. In *Proceedings of the 2020 SIAM International Conference on Data Mining*, 532–540. Society for Industrial and Applied Mathematics (SIAM).

Frankel, T. 2011. *The Geometry of Physics: An Introduction*. Cambridge University Press.

Hatcher, A. 2001. *Algebraic Topology*. Cornell University.

Hirani, A. N. 2003. *Discrete Exterior Calculus*. Ph.D. dissertation, California Institute of Technology.

Iten, R.; Metger, T.; Wilming, H.; Del Rio, L.; and Renner, R. 2020. Discovering Physical Concepts with Neural Networks. *Physical Review Letters* 124(1): 010508.

Kitano, H. 2016. Artificial Intelligence to Win the Nobel Prize and Beyond: Creating the Engine for Scientific Discovery. *AI Magazine* 37(1): 39–49.

Koren, B.; Abgrall, R.; Bochev, P.; Frank, J. E.; and Perot, B. 2014. Physics-Compatible Numerical Methods. *Journal of Computational Physics* 257(Part B): 1039–1039.

Kron, G. 1963. *Diakoptics: The Piecewise Solution of Large-Scale Systems*, volume 2. MacDonald.

Langley, P. 1998. The Computer-Aided Discovery of Scientific Knowledge. In *International Conference on Discovery Science*, 25–39. Springer.

Lipnikov, K.; Manzini, G.; and Shashkov, M. 2014. Mimetic Finite Difference Method. *Journal of Computational Physics* 257: 1163–1227.

Mattiussi, C. 2000. The Finite Volume, Finite Element, and Finite Difference Methods as Numerical Methods for Physical Field Problems. *Advances in Imaging and Electron Physics* 113: 1–147.

Mehta, P.; Bukov, M.; Wang, C. H.; Day, A. G. R.; Richardson, C.; Fisher, C. K.; and Schwab, D. J. 2019. A High-Bias, Low-Variance Introduction to Machine Learning for Physicists. *Physics Reports* 810: 1–124.

Miyanawala, T. P.; and Jaiman, R. K. 2017. An Efficient Deep Learning Technique for the Navier-Stokes Equations: Application to Unsteady Wake Flow Dynamics. *arXiv Preprints* ISSN 0264-6021. doi:10.1016/j.eswa.2008.08.077.

Nautrup, H. P.; Metger, T.; Iten, R.; Jerbi, S.; Trenkwalder, L. M.; Wilming, H.; Briegel, H. J.; and Renner, R. 2020. Operationally Meaningful Representations of Physical Systems in Neural Networks. *arXiv preprint arXiv:2001.00593* .

Palha, A.; Rebelo, P. P.; Hiemstra, R.; Kreeft, J.; and Gerritsma, M. 2014. Physics-Compatible Discretization Techniques on Single and Dual Grids, with Application to the Poisson Equation of Volume Forms. *Journal of Computational Physics* 257: 1394–1422.

Raghu, M.; and Schmidt, E. 2020. A Survey of Deep Learning for Scientific Discovery. *arXiv preprint arXiv:2003.11755* .

Raissi, M.; Perdikaris, P.; and Karniadakis, G. E. 2019. Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations. *Journal of Computational Physics* 378: 686–707.

Roth, J. P. 1955. An Application of Algebraic Topology to Numerical Analysis: On the Existence of a Solution to the Network Problem. *Proceedings of the National Academy of Sciences* 41(7): 518–521.

Rudy, S. H.; Brunton, S. L.; Proctor, J. L.; and Kutz, J. N. 2017. Data-Driven Discovery of Partial Differential Equations. *Science Advances* 3(4): e1602614.

Sanchez-Gonzalez, A.; Godwin, J.; Pfaff, T.; Ying, R.; Leskovec, J.; and Battaglia, P. W. 2020. Learning to Simulate Complex Physics with Graph Networks. *arXiv preprint arXiv:2002.09405* .

Schmidt, M.; and Lipson, H. 2009. Distilling Free-Form Natural Laws from Experimental Data. *Science* 324(5923): 81–85.

Seo, S.; and Liu, Y. 2019. Differentiable Physics-Informed Graph Networks. *arXiv preprint arXiv:1902.02950* .

Stevens, R.; Taylor, V.; Nichols, J.; Maccabe, A. B.; Yelick, K.; and Brown, D. 2020. AI for Science. Technical report, Argonne National Laboratory(ANL).

Tonti, E. 2013. *The Mathematical Structure of Classical and Relativistic Physics: A General Classification Diagram*. Modeling and Simulation in Science, Engineering and Technology. Birkhäuser. ISBN 9781461474210.

Udrescu, S. M.; and Tegmark, M. 2020. AI Feynman: A Physics-Inspired Method for Symbolic Regression. *Science Advances* 6(16): eaay2631.

Wang, R.; Kashinath, K.; Mustafa, M.; Albert, A.; and Yu, R. 2020. Towards Physics-Informed Deep Learning for Turbulent Flow Prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1457–1466.

Wei, Z.; and Chen, X. 2019. Physics-Inspired Convolutional Neural Network for Solving Full-Wave Inverse Scattering Problems. *IEEE Transactions on Antennas and Propagation* 67(9): 6138–6148.

Wu, T.; and Tegmark, M. 2019. Toward an AI Physicist for Unsupervised Learning. *arXiv preprint arXiv:1810.10525* .

Yang, K. K.; Wu, Z.; and Arnold, F. H. 2019. Machine-Learning-Guided Directed Evolution for Protein Engineering. *Nature Methods* 16(8): 687–694. doi:10.1038/s41592-019-0496-6.