ORCID: 0009-0008-6449-3023
M. V. HERKALIUK
Ph.D. student, Institute of Physics and Technology, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine
m.herkaliuk@gmail.com

ORCID: 0000-0001-6473-53
A. V. GILCHUK
Ph.D., associate professor, Institute of Physics and Technology, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine
a.v.gilchuk@gmail.com

DOI: 10.31651/2076-5851-2025-11-57 PACS: 05.20.Dd;66.70.-f; 81.07.-b; 02.60.Cb;07.05.Mh

PHYSICS-INFORMED MACHINE LEARNING FOR NANOSCALE TRANSPORT PHENOMENA: CHALLENGES, APPROACHES, AND FUTURE PERSPECTIVES

Modeling nanoscale transport phenomena presents a critical challenge in which classical continuum equations fail, and high-fidelity solvers are computationally prohibitive. Physics-Informed Machine Learning (PIML) has emerged as a transformative approach to resolve this dilemma by synergistically fusing sparse experimental data with the governing laws of first-principle transport models. This review provides a comprehensive overview of how PIML — especially physics-informed neural networks (PINNs), operator-learning methods, and multi-fidelity frameworks — accelerates nanoscale transport analyses from BTE-based phonon transport to ballistic–diffusive heat transfer and near-field radiative effects. We address persistent data bottlenecks in nanomaterial research, including noisy measurements and high-dimensional partial differential equation (PDE) formulations, and present advanced strategies such as domain decomposition and hybrid mechanistic – Machine Learning (ML) methods to enhance the flexibility and scalability of these emerging approaches. Finally, we outline the current gaps in the field, from uncertainty quantification to the development of real-time digital twins, and chart future research directions poised to unify quantum-scale simulations, experimental metrology, and deep learning. By embedding physical constraints directly into the learning workflow, these physics-informed methods offer a transformative pathway for optimizing nanoscale transport, unlocking unprecedented opportunities in material design and device engineering.

Keywords: Nanoscale transport phenomena, Nanomaterials, Machine Learning (ML), Physics-Informed Neural Networks (PINNs), Deep learning (DL)

1. Introduction

1.1 Motivation and Scope

At nanometer-length scales, the mechanisms governing heat transfer, mass diffusion, and charge transport deviate significantly from the established macroscale paradigms. In crystalline semiconductors, for instance, heat is largely carried by phonon-quantized lattice vibrations, whose mean free paths can rival or exceed device dimensions; for instance, in silicon, the average phonon mean free path at room temperature can be over 40 nm, while the gate length of a modern transistor is already less than 10 nm. This significant mismatch means that heat transport is dominated by ballistic effects and boundary scattering rather than classical diffusion

[1]. Under such conditions, the standard Fourier law breaks down, and a host of additional effects comes into play: boundary scattering at interfaces, ballistic conduction through defect-limited regions, and quantum confinement in ultrathin layers. These nuances are not limited to phonons; electrons, especially in downscaled transistors, can also exhibit ballistic transport over distances comparable to their inelastic scattering lengths [2]. Even in layered van der Waals crystals, breaking the symmetries at the atomic level introduces unconventional conduction pathways that lead to pronounced non-linear or non-reciprocal transport responses [3].

The development of modern nanoelectronics, thermoelectric devices, and advanced materials hinges on the understanding of these exotic transport regimes. Accurately predicting the heat flow in integrated circuits is crucial for preventing thermal bottlenecks and improving reliability, whereas optimizing phonon transport can enhance the performance of thermoelectric generators or thermal barrier coatings [1, 4]. In phase-change memory cells, ultrafast nanoscale heating orchestrates the switching process, necessitating precise models of the thermal conduction and melting dynamics. Near-field radiative transport, relevant in sub-micrometer gaps, opens avenues for energy harvesting beyond the blackbody limit. Such phenomena are deeply tied to quantum mechanics, ballistic-diffusive transitions, and interfacial phenomena, placing them well outside the comfort zone of classical equations.

Traditional numerical methods have made tremendous strides in resolving these issues. Molecular Dynamics (MD) simulations offer atomistic insights into phonon scattering, but require formidable computational resources and may overestimate high-frequency mode populations at low temperatures [1, 2]. Monte Carlo schemes treat phonons or electrons as particles with probabilistic scattering rules, thereby elucidating boundary effects and phonon-electron interactions for both diffusive and ballistic regimes [2]. Similarly, solving the Boltzmann Transport Equation (BTE) has illuminated how boundary scattering and interface conductance shape the nanoscale thermal conductivities [4]. However, as device features continue to shrink into the quantum regime, balancing the accuracy and computational cost becomes more challenging. Moreover, many properties, such as interface scattering rates or anisotropic lattice constants, require parameter fitting, which introduces uncertainties that classical solvers alone cannot easily reconcile [5].

Simultaneously, experimentation has evolved to probe these finer scales. Techniques, such as time-domain thermoreflectance (TDTR), help isolate quasiballistic phonons by measuring localized temperature changes under laser heating [1, 4]. Scanning thermal microscopy (SThM) captures spatially resolved heat flux albeit with calibration complexities. Fabricating specialized nanostructures, such as grating heaters, enables the systematic characterization of phonon mean free path distributions [4]. However, the synergy between experimental insight and theoretical modeling remains hampered by data sparsity and the high dimensionality of the parameter spaces. This juncture has sparked growing interest in machine learning (ML) methods that can both incorporate physics-based constraints and leverage whatever data are available, whether from simulations, reduced-order models, or actual measurements [5–8].

Recognizing the multifaceted challenges at hand, ranging from quantum effects to short mean free paths and steep interface gradients, this review explores how emerging physics-informed ML techniques can bridge the gap between computationally intensive solvers and incomplete experimental data. This review argues that the rise of Physics-Informed Machine Learning (PIML) is more than an incremental advance; it represents a paradigm shift in the modeling workflow itself. The classical approach is a one-way street that defines physics, discretizes it, and solves for a single instance. In contrast, PIML enables a flexible, inferential workflow where partial physical laws, sparse data, and governing equations are fused to not only predict system behavior, but also to discover unknown parameters and unmodeled physics. These methods create a synergistic bridge between the rigor of first-principles models and

flexibility of data-driven approaches, paving the way for solving previously intractable problems in nanoscale transport. These approaches aim to preserve the interpretability of the governing equations while enhancing flexibility in parameter inference and uncertainty quantification. Our discussion spans nanoscale heat conduction, electron–phonon interactions, and near-field radiative exchange, outlining why conventional continuum-based models often fall short, and how ML-driven frameworks may offer more adaptive solutions.

To that end, the following sections delve deeper into fundamental transport models at the nanoscale and highlight recent breakthroughs. Ultimately, these complexities do more than call for a better solver, and they demand a new scientific workflow. This review demonstrates how Physics-Informed Machine Learning provides this new paradigm, guiding the design, interpretation, and optimization of next-generation nanoscale systems by transforming how we fuse theory with data. We will first establish the fundamental limitations of classical methods that created this need, and then explore how specific PIML approaches directly overcome these bottlenecks, paving the way for solving previously intractable problems in nanoscale science.

1.2 Overview of Nanoscale Transport Phenomena

Nanoscale transport phenomena encompass a rich array of physical processes that deviate significantly from traditional continuum-based descriptions of heat and mass transfer. At dimensions of the order of tens of nanometers or below, familiar concepts such as Fourier's law of heat conduction can become insufficient, as carriers (phonons, electrons, and even photons) experience scattering events and boundary effects in ways not typically observed at larger scales. The transition from diffusive to ballistic transport occurs because the dominant carrier mean free paths are comparable to or even exceed the material dimensions. As a result, carrier scattering can be strongly suppressed, yielding subcontinuum regimes in which classical laws break down. This section provides an overview of key nanoscale transport phenomena, highlighting ballistic versus diffusive behaviors, the role of phonon scattering and quantum size effects, and several industrial and research applications that motivate a deeper understanding of such processes.

A core distinction in nanoscale conduction is whether the heat carriers move diffusively or ballistically. Under diffusive conditions, energy transport follows a near-equilibrium picture, which is often described by Fourier or Fick equations with bulk material properties. Specifically, Fourier's law relates the heat flux vector, mathbfq, to the temperature gradient ∇T via the thermal conductivity k:

$$q = -k\nabla T \tag{1}$$

Similarly, Fick's first law describes the mass diffusion, linking the diffusion flux $\boldsymbol{J}$ to the concentration gradient $\nabla\varphi$ through the diffusion coefficient D:

$$J = -D\nabla\varphi \tag{2}$$

These linear, local relationships form the basis of classical continuum models, but fail to capture the non-local and non-equilibrium phenomena prevalent at the nanoscale.

In contrast, ballistic carriers traverse the medium with minimal scattering, rendering classical diffusion equations inaccurate, unless carefully modified. As described by Cahill et al. [1], ballistic effects become increasingly prominent when device features approach the mean free path of dominant phonons, which, for many crystalline semiconductors, can range from a few nanometers to several micrometers. Moreover, even within ostensibly diffusive materials, boundary scattering at surfaces or interfaces can create so-called "quasiballistic" regimes. Incorporating boundary reflections and partial phonon transmission is essential for predicting temperature fields and thermal resistance at the nanoscale.

Phonon scattering events, which include boundary, defect, and phonon-phonon scattering, critically affect the thermal conductivity of the nanostructures. Carbon nanotubes (CNTs), for instance, exhibit exceptionally high intrinsic thermal conductivities, yet real-world

devices often exhibit lower effective values because boundary conditions and substrate interactions shorten the phonon mean free paths [2]. In single-wall carbon nanotubes (SWCNTs), ballistic phonon transport can dominate for tube lengths of tens or hundreds of nanometers. However, as Lukes and Zhong note, simulation length and boundary condition choices strongly affect the perceived thermal conductivity, underscoring that “size matters” in a literal sense for these one-dimensional systems [9]. Similarly, in two-dimensional materials, wave-like phonon interference can either enhance or suppress heat conduction, depending on the arrangement of interfaces, defects, and atomic layers.

Beyond phonons, quantum size effects in electron transport also come to the fore in nanoscale systems, such as ultrathin films, nanowires, or quantum wells. When the electron de Broglie wavelengths become comparable to the device thickness, energy sub-bands form, and the scattering rates can differ markedly from those of the bulk. This condition is routinely met in contemporary devices. For example, the confinement layer in a silicon MOSFET can be as thin as 5 nm, which is significantly smaller than the electron's thermal de Broglie wavelength of approximately 17 nm at room temperature.

Such phenomena underpin the functionality of quantum devices including single-electron transistors and nanosensors. Non-linear or rectification effects can arise from broken symmetries in low-dimensional materials. Ideue and Iwasa elaborated on how inversion symmetry breaking in van der Waals heterostructures drives non-linear electric transport, resulting in phenomena such as non-reciprocal conduction or second-harmonic generation [3]. While these effects primarily involve charge carriers, the underlying principle—namely, that reduced dimensions expose quantum degrees of freedom— also applies just as well to phonon-based or photonic devices.

From an industrial and research standpoint, understanding nanoscale transport is indispensable for various applications. One prominent example is the transistor heat management in modern integrated circuits. As transistors scale down, localized hot spots can form at gate lengths of only a few nanometers, which limits the device performance and reliability. Here, ballistic and subcontinuum effects may exacerbate the thermal resistance in gate regions, prompting the need for advanced thermal metrology and modeling that go beyond traditional diffusive assumptions [1]. Thermoelectric materials, which convert heat into electrical energy, also rely on manipulating heat carriers at the subcontinuum level. Lowering the thermal conductivity while retaining good electrical conductivity often involves engineering nanostructures that scatter phonons more strongly than electrons. The success of this "phonon engineering" strategy is quantified by the dimensionless figure of merit, ZT, where nanostructuring has led to dramatic improvements, while bulk silicon has a negligible ZT of approximately 0.01, and silicon nanowires have demonstrated values approaching 1.0, an increase of nearly two orders of magnitude. Simultaneously, sensors and quantum devices leverage ballistic transport to achieve higher sensitivity or lower noise, capitalizing on a reduced scattering environment.

Practical examples of subcontinuum transport extend to measurements of phonon mean free path (MFP) distributions, which are crucial for accurately modeling heat conduction in nanostructured systems. Zeng et al. employed quasiballistic phonon transport in patterned grating structures to invert the measured thermal signals and reconstruct the MFP distributions of materials such as crystalline silicon [4]. By confining heat pulses to narrow metallic lines and tracking the resulting transient temperature signatures, one can observe how long MFP phonons escape the heated zone without scattering, thus reducing the apparent thermal conductivity. Repeating these measurements with varying line widths or geometries yields multiple “effective” thermal conductivity values, from which a Boltzmann Transport Equation-based model infers the underlying phonon spectrum. Such approaches are already being extended to complex systems, such as alloys, superlattices, and strongly disordered materials,

where direct theoretical predictions (e.g., from first-principles density functional theory) become computationally daunting.

Quantum-size effects are vital in emerging electronics and optoelectronics. Low-dimensional van der Waals systems—graphene, transition-metal dichalcogenides, and associated nanotubes—support a host of phase transitions, spin–orbit couplings, and topologically driven properties, each manifesting differently when conduction is confined to one or two dimensions [3]. These quantum features can be harnessed to create high-performance sensors, rectifiers, and spintronic elements that operate at energies far lower than those of conventional semiconductor junctions. Thus, understanding subcontinuum conduction bridges fundamental research into quantum phenomena with the practical engineering of next-generation devices.

Overall, at the nanoscale, the conduction phenomena reflect a tapestry of ballistic wave-like behavior, subcontinuum scattering, and quantum mechanical constraints. Investigations in this realm benefit from advanced experimental approaches such as time-domain thermoreflectance (TDTR) and scanning probe thermometry, together with robust theoretical frameworks grounded in the Boltzmann Transport Equation or atomistic simulations [1, 4, 9]. Designing materials and devices to exploit or mitigate these effects requires multidisciplinary expertise spanning material science, applied physics, and device engineering. Deeper insights into boundary scattering, phonon coherence, and non-linear electronic responses will likely shed light on how to tailor thermal and electrical conduction in ways that are deemed impractical. This new understanding in turn fuels the development of high-performance transistors, energy-efficient thermoelectrics, nanoscale sensors, and quantum-based components—each demonstrating that “less” can indeed mean “more” when it comes to leveraging unique size-driven transport phenomena.

By illuminating the interplay between ballistic and diffusive regimes, phonon scattering mechanisms, and quantum size effects, researchers are increasingly able to engineer conduction pathways. This sets the stage for the following sections, which delve more deeply into the models, measurement techniques, and design strategies that enable manipulation and application of nanoscale transport in real-world devices.

2. Background and Classical Approaches

2.1 Mathematical Models for Nanoscale Transport

At nanoscale dimensions, the foundational equations of heat transfer require modifications that account for non-classical effects, such as ballistic phonon transport, quantum confinement, and strong boundary scattering. Researchers have developed a suite of mathematical models and computational schemes to address nanoscale transport, including the heat equation with quantum corrections, Boltzmann Transport Equation (BTE), and various diffusion equation variants tailored to small-scale phenomena. The critical length scale governing the transition from classical to nanoscale transport is the phonon mean free path (MFP). As shown in Table 1, the MFPs for common materials can range from tens to hundreds of nanometers, which is a scale comparable to that of modern nanodevices.

Classical heat transfer formulations, rooted in the standard heat equation, treat energy carriers as diffusing particles in a local thermodynamic equilibrium. While this assumption suffices for macroscopic systems, it can break down when the characteristic length scales approach or fall below the mean free path of the phonons or electrons. Consequently, researchers have developed a suite of mathematical models and computational schemes to address nanoscale transport, including the heat equation with quantum corrections, Boltzmann Transport Equation (BTE), and various diffusion equation variants tailored to small-scale phenomena.

Table 1
Таблиця 1

Characteristic mean free paths (MFP) of phonons for some materials at room temperature (300 K).
Характерний середній вільний пробіг (MFP) фононів для деяких матеріалів за кімнатної температури (300 K).

Material	**Average MFP (nm)**	**Notes**
Silicon (Si)	~40 - 300	Depends on purity and structure
Diamond (C)	~300	High thermal conductivity
Gallium Arsenide (GaAs)	~20	Stronger phonon scattering
Graphene	> 750	Very high thermal conductivity

2.1.1 Heat Equation with Quantum Corrections

The classical heat diffusion equation provides the foundation for macroscopic thermal analysis:

$$\rho c_p \frac{\partial T}{\partial t} = \nabla \cdot (k_{\text{bulk}} \nabla T) + q \quad (3)$$

where ρ is the density, c_p is the specific heat capacity, T is the temperature, k_{bulk} is the bulk thermal conductivity, and q is the volumetric heat-source term. This equation assumes that the heat carriers (e.g., phonons) travel diffusively.

However, this assumption breaks down when the characteristic dimension of the system L becomes comparable to or smaller than the phonon mean free path Λ. In this subcontinuum regime, boundary scattering suppresses the contribution of long-MFP phonons, reducing the overall thermal conductivity. To account for this, the bulk conductivity k_{bulk} is often replaced by the size-dependent effective thermal conductivity k_{eff}. A common way to express this dependency is through the Knudsen number $Kn = \Lambda/L$.

For example, for a thin film, a simplified model derived from the Boltzmann Transport Equation yields:

$$k_{\text{eff}}(Kn) \approx k_{\text{bulk}} \left(1 + \frac{4}{3} Kn\right)^{-1} \quad (4)$$

Such modifications, while phenomenological, allow the heat equation to approximate subcontinuum effects without resorting to more computationally expensive kinetic solvers. The practical impact of this size effect is illustrated in Fig. 1, which shows a dramatic reduction in the thermal conductivity of silicon nanowires as their diameter decreases.

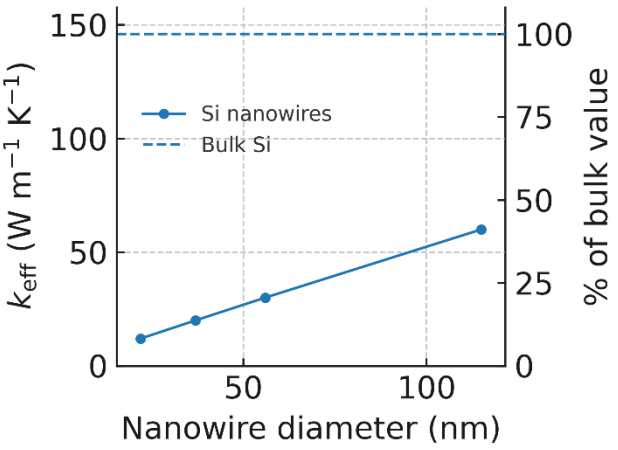

Fig. 1. Effective thermal conductivity k_{eff} of silicon nanowires at 300 K as a function of nanowire diameter. Solid line with circle markers shows nanowire values for diameters 22, 37, 56 and 115 nm (approximately 12, 20, 30 and 60 $\frac{W}{m\cdot K}$, respectively). The dashed horizontal line indicates bulk silicon, $k_{\mathrm{bulk}} \approx 146\frac{W}{m\cdot K}$. The right-hand axis reports the same data as a percentage of the bulk value, computed as $100 \cdot k_{\mathrm{eff}}/k_{\mathrm{bulk}}$ (the four points correspond to ~8%, ~13%, ~20% and ~40%).

Рис. 1. Ефективна теплопровідність k_{eff} кремнієвих нанодротів при 300 K як функція діаметра нанодроту. Суцільна лінія з круглими маркерами показує значення нанодротів для діаметрів 22, 37, 56 та 115 нм (приблизно 12, 20, 30 та 60 $\frac{W}{m\cdot K}$відповідно). Штрихова горизонтальна лінія вказує на об'ємний кремній, $k_{\mathrm{bulk}} \approx 146\frac{W}{m\cdot K}$. Права вісь показує ті ж дані у відсотках від об'ємного значення, обчисленого як $100 \cdot k_{\mathrm{eff}}/k_{\mathrm{bulk}}$ (чотири точки відповідають ~8%, ~13%, ~20% та ~40%).

2.1.2 Boltzmann Transport Equation (BTE)

Arguably, the most comprehensive framework for analyzing nanoscale heat conduction is BTE, which tracks the distribution function of energy carriers (e.g., phonons and electrons) in the phase space. In particular, phonon BTE provides a way to incorporate scattering mechanisms–phonon — phonon interactions, boundary scattering, and impurity scattering—to model heat transfer in crystals and nanostructures. The general form of the phonon BTE can be written as

$$\frac{\partial f}{\partial t} + v_g \cdot \nabla_r f = \left(\frac{\partial f}{\partial t}\right)_{\mathrm{coll}} \tag{5}$$

Here, $f(\boldsymbol{r}, \boldsymbol{k}, t)$ is the non-equilibrium phonon distribution function, which depends on position $\boldsymbol{r}$, wavevector $\boldsymbol{k}$, and time t. vg is the phonon group velocity, and the term on the right, $\left(\frac{\partial f}{\partial t}\right)_{\mathrm{coll}}$, represents the rate of change in f owing to scattering events (the collision operator).

Although BTE is conceptually straightforward, the collision term is a complex integral that accounts for all possible scattering mechanisms. To make the BTE tractable, researchers typically adopt simplifications such as Relaxation Time Approximation (RTA). Within the RTA, the collision operator is simplified to

$$\left(\frac{\partial f}{\partial t}\right)_{\mathrm{coll}} \approx -\frac{f - f_0}{\tau} \tag{6}$$

where f_0 is the equilibrium (Bose-Einstein) distribution and τ is the relaxation time, which represents the characteristic time for the distribution to return to equilibrium via scattering. Researchers typically adopt simplifications such as the relaxation time approximation (RTA) to make BTE tractable. In RTA, each phonon mode relaxes toward

equilibrium on a characteristic timescale. This simplification can yield relatively accurate results for many bulk materials, although it can underestimate the effects of normal (momentum-conserving) scattering processes and ballistic transport on small scales. Moreover, at nanoscale boundaries, phonon reflection, transmission, and partial specular scattering complicate the formulation of the boundary conditions. These effects may induce pronounced temperature jumps, flux slip, or interfacial resistances that deviate substantially from macroscopic intuition.

Progress in first-principles calculations has enabled parameter-free BTE solvers, such as those described in [10], which combine density functional theory with iterative solution schemes to capture three-phonon and isotope scattering across entire Brillouin zones. Such ab initio approaches, which are computationally expensive, have demonstrated predictive power for the lattice thermal conductivity in both bulk crystals and nanowires. On the more engineering-oriented side, approximate BTE solutions, hybrid Monte Carlo-diffusion techniques [12], and data-driven methods [13] provide flexible routes for modeling intricate nanoscale geometries without incurring the full expense of a purely ab initio framework.

2.1.3 Diffusion Equation Variants

Even with the availability of robust BTE solvers, diffusion-like equations remain pervasive in practical device simulations. In many situations, Fourier's law or slightly modified versions that allow for partial ballistic effects are employed. For example, in "two-step" or "multiscale" modeling strategies, a high-fidelity kinetic or Monte Carlo solver is applied near boundaries or in hot-spot regions where ballistic phenomena dominate, whereas a simpler diffusion model suffices in the bulk. As demonstrated in [11], such approaches are crucial for thin-film devices where the film thickness rivals the phonon mean free path. In these regimes, the classical equation often fails to predict the heat flux accurately, necessitating corrections for the boundary scattering and ballistic transport.

A further refinement is the inclusion of hyperbolic terms in the heat equation, which is sometimes introduced to account for finite thermal propagation speeds. While these hyperbolic "wave-like" models capture certain transient effects, they have limited approximations. Rigorous solutions require capturing the full phonon distribution, as indicated by the radiative transfer-based approaches discussed in [11]. However, for rapidly prototyped microscale designs, diffusion equation variants offer a balance between computational simplicity and partial accuracy, thereby guiding more advanced simulations or experiments.

2.1.4 Concluding Remarks

In summary, modeling nanoscale heat transport requires a spectrum of mathematical approaches, ranging from modest corrections to classical diffusion equations to full-phonon Boltzmann transport simulations. The appropriate choice depends on the geometry, temperature range, scattering mechanisms, and the desired accuracy. While the heat equation with quantum corrections can offer expedient solutions in near-diffusive conditions, the BTE is a more fundamental framework for capturing ballistic effects. Hybrid or multiscale strategies that integrate Monte Carlo or BTE solvers near boundaries with bulk diffusion are increasingly attractive for realistic device-level studies. Looking ahead, the expansion of first-principles BTE solutions [10], development of novel Monte Carlo-diffusion hybrids [12], and emergence of physics-informed neural networks [13] all point to a future where nanoscale heat conduction can be modeled with both speed and fidelity. Such advances are essential for guiding the thermal management and design of next-generation electronic and photonic devices.

The confluence of these challenges, from computational complexity to reliance on unknown parameters, creates an urgent need for a new modeling paradigm that can bridge the

gap between accuracy and computational cost. This is precisely the niche that physics-informed machine learning methods are poised to fill.

2.2 The Computational Bottlenecks Necessitating a New Paradigm

Traditional numerical methods, such as molecular dynamics (MD), Monte Carlo (MC), ballistic transport models, and hybrid continuum-atomistic approaches, have provided invaluable insights into nanoscale phenomena. However, each approach has inherent limitations when it comes to modeling the full complexity of nanoscale partial differential equations (PDEs) and multiphysics problems. MD methods, for instance, capture atomistic details with high fidelity but are constrained by immense computational costs and short time scales, typically on the order of nanoseconds or microseconds [14]. This method is based on numerically integrating Newton's second law of motion for a system of N atoms:

$$m_i \frac{d^2 \boldsymbol{r_i}}{dt^2} = \boldsymbol{F_i} = -\nabla_{\boldsymbol{r_i}} V(\boldsymbol{r_1}, \boldsymbol{r_2}, \dots, \boldsymbol{r_N}) \quad (7)$$

where m_i, $\boldsymbol{r_i}$, and $\boldsymbol{F_i}$ are the mass, position, and force acting on atom i, respectively. The force was calculated as the negative gradient of the interatomic potential energy function V, which is the most critical input to the simulation.

As simulations grow in size and complexity, whether investigating protein folding or thermal transport in semiconductor devices, the time-step requirements and sheer number of atoms become prohibitive. In addition, classical MD relies on force fields and potential functions that may not account for all relevant quantum effects. Although quantum corrections can improve accuracy, they add layers of computational overhead that are not always feasible to implement.

Monte Carlo (MC) methods approach the transport problem differently, particularly for scattering-dominated regimes, where phonons, electrons, or other carriers undergo numerous random collisions. By tracking particle trajectories through probabilistic scattering events, MC can capture a wide range of transport processes [1]. The simulation loop for a single particle (e.g., a phonon) follows a "free-flight and scatter" algorithm governed by the scattering rate $\Gamma = 1/\tau$. The probability of a particle traveling for time t without scattering is given by an exponential distribution:

$$P(t) = e^{-\Gamma t} \quad (8)$$

In each step, a free-flight time is sampled from this distribution, the particle's position is updated, and then a scattering event is simulated by randomly selecting a new state (e.g., new wavevector) based on predefined probabilities for different scattering mechanisms.

However, randomness at the core of the method can lead to significant variance in simulation outcomes, necessitating large sample sizes and longer runtimes to achieve statistically reliable results. This stochastic character becomes especially problematic when exploring delicate features, such as electron–phonon non-equilibrium in transistors or heat generation in carbon nanotubes. Although variance reduction techniques exist, they do not fully eliminate the trade-off between simulation accuracy and runtime.

Ballistic transport models are at another extreme, where scattering is largely neglected or drastically simplified [2]. These models have very small structures, such as nanowires or channels shorter than the mean free path, where carriers can traverse the domain without frequent collisions. However, once real devices extend beyond purely ballistic regimes or include multiple interfaces, the assumptions underlying purely ballistic equations fail to capture important scattering, boundary resistance, or interfacial phonon transmission phenomena. Consequently, ballistic treatments often yield overly optimistic estimates of the conductivity or underestimate the severity of localized heating in semiconductors.

Hybrid continuum-atomistic approaches aim to bridge scales by coupling classical PDE-based solvers (e.g., heat diffusion or fluid flow at the continuum level) with atomistic

descriptions (e.g., MD or ab initio methods) at crucial hotspots [9]. In principle, these techniques capture both large-scale device behaviour and local microscopic effects, offering a comprehensive view of phenomena such as localized heat generation in transistors and phonon scattering in nanocarbon materials. Nevertheless, building and maintaining such multi-resolution frameworks can be exceedingly complicated. Researchers must ensure seamless coupling between the continuum domain and atomistic region, exchanging boundary conditions, fluxes, and state variables without introducing spurious reflections or numerical instabilities. The computational overhead also increases because of the parallel management of the two distinct solvers, each requiring specialized algorithms and tight integration.

In many of the most demanding nanoscale applications, from biomolecular simulations to cutting-edge transistor design, these limitations can make standard methods insufficient for the accurate and efficient prediction of real-world behaviour. For example, short timescales in MD constrain researchers who want to model rare events such as protein conformational changes or device aging. In MC simulations, the cost of achieving statistically convergent results can balloon in large or complex geometries, particularly if near-field radiative effects or boundary scattering significantly influence heat transfer. Ballistic modeling oversimplifies these processes, ignoring essential scattering mechanisms and leading to incomplete energy dissipation. Finally, hybrid methods, which are conceptually powerful, often demand a level of expertise and computational resources that pose a barrier to their widespread adoption.

Taken together, these challenges underscore the need for novel frameworks that combine physical rigor with computational efficiency, potentially leveraging physics-informed machine learning or advanced reduced-order modeling. The ultimate goal is to capture the multiphysics nature of nanoscale systems without sacrificing accuracy or incurring prohibitive computational costs. As device miniaturization continues, and as multifunctional materials with intricate internal structures gain prominence, the drive toward more holistic and efficient numerical methods will only intensify. Despite their well-earned place in the researcher's toolkit, today's traditional methods alone often cannot meet the demands of modern nanoscale science and engineering, necessitating a fundamentally new strategy that can invert the traditional modeling process: one that can learn from sparse data while respecting physical laws, handle immense parameter spaces, and operate at the speed required for design and control. This is precisely the niche that physics-informed machine learning is poised to fill in.

See Table 2 for a comparitive overview of classical and emerging methods.

Table 2
Таблиця 2

Comparative overview: of major modeling approaches for nanoscale transport—essence, advantages, drawbacks (accuracy/speed/resources), and key limitations.
Порівняльний огляд основних підходів до моделювання наномасштабного транспорту — суть, переваги, недоліки (точність/швидкість/ресурси) та ключові обмеження.

Approach	**Essence**	**Advantages (accuracy/speed)**	**Drawbacks / Resource Needs**	**Key Limitations**
Molecular Dynamics (MD)	Atomistic simulation (no	Highest physics fidelity at atomic scale	Extremely high cost; limited simulated timescales	Requires interatomic potentials;

	continuum assumptions)			quantum effects costly to include
Monte Carlo (MC)	Particle-based transport incl. scattering	Captures ballistic↔diffusive regimes; flexible geometries	Moderate-to-high cost; variance/noise issues	Large samples for convergence; re-run for each scenario
Deterministic BTE	Phase-space PDE for carriers	Predictive with ab-initio scattering rates	Very high memory/CPU requirements	Complex boundaries; stability; detailed inputs required
Continuum (Fourier heat eq.)	Diffusive limit of transport	Very fast; highly scalable	Breaks down at sub-MFP scales	No ballistic/interface jumps; only near diffusive limit
Hybrid multiscale (MC/BTE + diffusion)	High fidelity where needed with reduced cost elsewhere	Lower cost than full high-fidelity everywhere	Coupling complexity and tooling overhead	Matching flux/temperature across interfaces is non-trivial
Physics-Informed NNs (PINNs)	NN + PDE residuals + (optional) data	Mesh-free; good for inverse problems; data-efficient	Training instability; retrain per instance	Stiff/high-freq solutions; hyperparameter sensitivity

3. Machine Learning Approaches for Nanoscale Transport Phenomena

3.1 Why Machine Learning in Nanoscale Transport?

Machine Learning (ML) methods have become pivotal in studying nanoscale transport phenomena, ranging from phonon and electron transport to energy conversion processes at the molecular or atomic level, owing to their ability to leverage partial experimental or simulation data for complex, high-dimensional problems. Traditional numerical techniques, such as direct simulation of the Boltzmann Transport Equation (BTE), require enormous computational resources when handling complicated geometries, boundary conditions, or numerous physical interactions across multiple length and time scales. Solving a non-linear BTE for a realistic 2D geometry can take several hours or even days on a high-performance computing cluster,

whereas a trained neural network surrogate can generate a solution of comparable accuracy in a matter of seconds on a single GPU. ML-based surrogate models alleviate some of these burdens by efficiently approximating the behavior of the system after training on data sampled from high-fidelity simulations or carefully designed experiments. In the “small data” regime often faced at the nanoscale, physics constraints and domain knowledge can further guide machine learning models to remain physically consistent even when direct measurements are sparse [5].

Another key advantage of ML for nanoscale transport is its real-time predictive capability. Many micro- or nanotechnological applications, such as nanoelectronic thermal management or drug delivery systems via nanofluidic channels, require rapid evaluation of transport properties under dynamically changing conditions. By embedding ML surrogates into multiscale modeling workflows, researchers can continuously update the boundary conditions and handle local variability without solving the underlying physics from scratch. This not only saves computational time but also enables data-driven decision-making in processes, such as adaptive design or online process control [15]. For instance, if a neural network is trained to predict the heat flux in a thermoelectric device under different contact resistances, then new configurations can be explored instantly rather than running full-scale simulations each time.

The development of multifidelity and physics-informed machine learning approaches further enhances the utility of ML for nanoscale transport. Multi-fidelity techniques combine low-accuracy but cost-effective simulations with high-accuracy and smaller-scale data to strike an optimal balance between computational efficiency and model reliability. Physics-informed strategies (e.g., physics-informed neural networks or PINNs) go a step further by incorporating governing partial differential equations directly into the learning process, ensuring that the output respects conservation laws and other physical constraints [5, 16]. Although such methods have been more extensively applied to continuum mechanics and fluid flow, their principles can be easily extended to modeling energy, mass, and charge transport at sub-continuum scales.

Ultimately, machine learning holds promise for accelerating nanoscale transport research by uniting partial observations, either from local measurements or partial micro- to macro-scale simulation outputs, with robust surrogate models that reduce computational costs and enable real-time analysis. As advanced manufacturing and nanotechnology continue to push design boundaries, these data-driven approaches will become indispensable for characterizing, predicting, and optimizing transport phenomena at the smallest scale. Future challenges include improving interpretability, ensuring generalization beyond training conditions, and achieving reliable uncertainty quantification. Nevertheless, ongoing progress in blending physical priors with flexible learning architectures underscores the transformative role that ML plays in nanoscale transport modeling and simulation.

3.2 ML Adoption in Physics community

The adoption of machine learning (ML) in physics has grown substantially, but the emergence of Physics-Informed Neural Networks (PINNs) since their popularization in 2018 marks a particularly transformative trend. Publication data illustrates this explosive growth, with the annual number of articles on PINNs increasing from approximately 120 in 2018 to over 2,000 by 2023. This momentum has continued, with the number of publications on track to match or exceed previous years in 2025, based on year-to-date data. This rapid expansion underscores a significant shift in computational physics, where PINNs evolve from a niche methodology to a foundational tool for solving complex, data-sparse problems by embedding physical laws directly into the learning process [40, 52]. The results are summarized in Fig. 2.

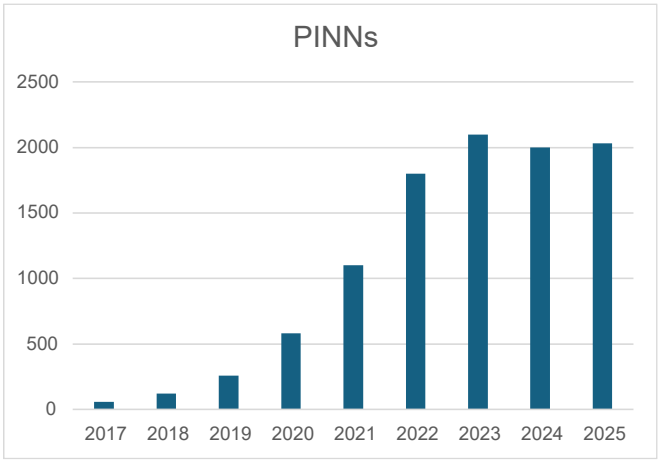

Fig. 2. Annual publications on Physics-Informed Neural Networks (PINNs) and variants, 2017–2025 (YTD). Counts aggregated from Dimensions.ai

Рис. 2. Річна кількість публікацій, присвячених фізично-інформованим нейронним мережам (PINNs) та їхнім варіантам, за 2017–2025 роки (на поточний момент).

We quantified the annual publication counts for PINNs and related variants (2017–2025 YTD) across physics, transport phenomena, and materials/nanomaterials using the Dimensions.ai API. Queries targeted titles, abstracts, and keywords (e.g., "physics-informed," PINN/XPINN/fPINN) with subject-area/domain filters when supported. Records from multiple endpoints were duplicated using the DOI or normalized titles. Counts were then aggregated by calendar year and subfiltered for transport-focused topics. The 2025 bar reflects data through August only; we mark it 'YTD' and avoid year-over-year inferences.

Within this trend, fields such as materials science, nanomaterials, and transport phenomena have become key areas for PINN application. While the overall volume of ML publications in these domains is large, exceeding 1,000 articles annually since 2021, PINNs are carving out a critical and rapidly growing niche. As shown in the accompanying chart, PINN-related publications on transport phenomena and materials science have increased steadily. Furthermore, within the broader paradigm of physics-informed machine learning (PIML), PINNs and their variants are overwhelmingly dominant, accounting for over 90% of publications in this subfield since 2021 and projected to reach 97% by 2025 based on partial data. The proliferation of specialized modifications such as XPINN, PIKAN, and ST-PINN further signals the maturation of the field as researchers actively work to enhance the stability, accuracy, and efficiency of these powerful models. [8]

3.3 Overview of Existing ML Methods in Physical Applications

Having established the computational bottlenecks of traditional methods, we now survey a portfolio of emerging machine learning techniques. Each method offers a unique approach to overcoming a specific classical limitation, from the high cost of re-simulation to the difficulty of discovering governing equations from the data.

Below is an overview of notable machine learning (ML) methods that have been applied to modeling physical systems, with particular attention to how these approaches incorporate physics-based principles, operator learning, and hybrid strategies. The discussion highlights key techniques—Hamiltonian/Lagrangian neural networks, operator-based methods (Fourier and Deep Operator Networks), sparse discovery frameworks (SINDy), neural ordinary differential equations (ODE), and partial MLphysics hybrids—followed by examples in quantum settings and a short note on limitations. While each method exhibits unique strengths, they collectively illustrate the emerging trend of physics-informed machine learning to tackle

the complexity of real-world dynamics and transport phenomena. See Table 3 for landmark applications and representative results.

Table 3
Таблиця 3
Landmark applications in nanoscale transport: method family, problem/feature, task, key result, and citation.
Знакові застосування в нанорозмірному транспорті: сімейство методів, проблема/особливість, завдання, ключовий результат та цитування.

Method	**Problem / Feature**	**Key Result**	**Authors (Year)**
PINN (Parametric)	Stationary, mode-resolved phonon BTE (1D–3D)	Accurate phonon transport without labeled data; solutions over parameterized spaces	Li R. et al. (2022)
PINN (Transient)	Time-dependent, mode-resolved phonon BTE (TDTR)	Excellent agreement with analytical/transient heat conduction benchmarks	Zhou J. et al. (2023)
Hamiltonian NN (HNN)	Learning conserved quantities	Enforces conservation laws; improved long-term stability/generalization	Greydanus S. et al. (2019)
Lagrangian NN (LNN)	Dynamics without canonical coordinates	Energy-preserving; more flexible than HNNs for some systems	Cranmer M. et al. (2020)
Fourier Neural Operator (FNO)	Parametric PDEs (Burgers, Darcy, Navier–Stokes)	Zero-shot super-resolution; up to ~10^3× faster than solvers at inference	Li Z. et al. (2021)
DeepONet	Learning non-linear operators	Operator-learning architecture with strong generalization to new inputs	Lu L. et al. (2019)
SINDy + Autoencoder	Discovering reduced coordinates & governing equations	Joint discovery of latent coordinates and sparse governing laws	Champion K. et al. (2019)

Δ-PINN	PDEs on complex geometries (e.g., Stanford bunny)	Laplace–Beltrami encoding enables solutions on complex topologies	Costabal F. S. et al. (2023)
fPINN	Fractional advection–diffusion (forward/inverse)	Extends PINNs to fractional operators	Pang G. et al. (2018)

3.3.1 Hamiltonian and Lagrangian Neural Networks

Classical Molecular Dynamics simulations, while powerful, often suffer from long-term energy drift owing to numerical integration errors. To address this fundamental stability issue, Hamiltonian (HNNs) and Lagrangian Neural Networks (LNNs) are designed to prioritize the conservation laws intrinsic to classical mechanics. HNNs learn a Hamiltonian function from data and enforce Hamilton's equations, thereby preserving an energy-like quantity across extended time horizons [17].

Specifically, a neural network is trained to approximate a scalar Hamiltonian function, $H_\theta(\boldsymbol{q}, \boldsymbol{p})$, where $\boldsymbol{q}$ and $\boldsymbol{p}$ are the generalized coordinates and momenta, respectively. The dynamics are then predicted by solving Hamilton's equations using the network output:

$$\dot{q} = \frac{\partial H_\theta}{\partial p}, \quad \dot{p} = -\frac{\partial H_\theta}{\partial q} \tag{9}$$

Derivatives were efficiently computed via automatic differentiation. This structure hard-codes the conservation of energy in the model's architecture. By taking position-momentum pairs as inputs and computing the partial derivatives of a learned scalar output, HNNs directly encode symplectic geometry into their structure. Consequently, they tend to maintain physically meaningful orbits in the phase space longer than generic neural networks, which often suffer from numerical energy drift.

LNNs adopt a complementary viewpoint, working in generalized coordinates and learning the underlying Lagrangian [18]. Under this paradigm, Euler–Lagrange equations are used to predict accelerations, ensuring adherence to conservation principles such as energy or momentum. One chief advantage of LNNs over HNNs is their coordinate-agnostic formulation, rendering them well suited to problems where canonical momenta may be difficult to define (e.g., in relativistic or constrained systems). Both Hamiltonian and Lagrangian frameworks demonstrate reduced long-term error accumulation compared with traditional architectures, thus providing stable predictions even for highly non-linear or chaotic dynamical systems.

3.3.2 Fourier Neural Operators

A critical bottleneck for traditional BTE or MD solvers is the 'one-shot' problem: a costly simulation must be re-run from scratch for every new set of boundary conditions or material parameters. To overcome this barrier to rapid design and optimization, operator-learning networks such as the Fourier Neural Operator (FNO) and DeepONet have been developed, which seek a resolution-invariant representation of partial differential equation (PDE) solutions [19]. This approach replaces the pointwise or convolution layers with integral kernels represented in the frequency domain. By performing a truncated Fourier transform of the hidden layer activations, frequency-dependent weights that act non-locally in the physical space are learned. The core of an FNO layer is the replacement of standard convolutions with a Fourier-domain spectral convolution. For an input function $v_t(x)$, the operation is defined as

$$v_{t+1}(x) = \sigma\left(W v_t(x) + \mathcal{F}^{-1}\left(R_\phi \cdot (\mathcal{F} v_t)\right)(x)\right) \tag{10}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$are the Forward and inverse Fourier transforms, respectively, R_ϕ is the learned linear transform in the frequency domain, W is a local linear transform, and σ is an activation function. This allows the model to learn global-resolution-independent patterns efficiently.

The primary appeal of FNOs is their speed and flexibility. Once trained on representative examples of parametric PDEs (e.g., different initial conditions or source parameters), the FNO can rapidly generate high-fidelity solutions for new parameter instances without retraining. Additionally, it exhibits mesh independence, meaning that a single model can be evaluated seamlessly at multiple spatial resolutions, an invaluable property when tackling multiscale or high-dimensional phenomena where standard solvers may become prohibitively expensive.

Despite their advantages, FNOs have significant practical limitations that must be considered. First, their reliance on the Fast Fourier Transform (FFT) inherently restricts them to uniform grids and periodic or rectangular domains, posing a major challenge for the complex irregular geometries common in nanodevices. Proposed solutions such as Geo-FNO aim to address this by using a learned coordinate transformation to map an irregular physical domain to a uniform latent grid, where the FFT can be applied. Second, because Fourier transform emphasizes global interactions, FNOs can struggle to capture important local spatial features. Training an FNO is a data-hungry process that often requires thousands of high-fidelity simulation examples to learn the operator, which can be a significant bottleneck in computationally expensive nanoscale problems.

3.3.3 Deep Operator Networks

Deep Operator Networks (DeepONets) also focus on learning operators rather than merely mapping a fixed-dimensional input to an output vector [20]. The key concept involves splitting the network into two parts: a “branch” network that encodes the input function (sampled at sensor points) and a “trunk” network that encodes the coordinates at which the output function is evaluated. By constructing an inner product of the branch and trunk embeddings, the final output captured the entire function-to-function mapping.

The architecture explicitly represents the operator G mapping an input function $u(x)$) to an output function $G(u)(y)$. The output was approximated as follows:

$$G(u)(y) \approx \sum_{k=1}^{p} b_k\big(u(x_1), \dots, u(x_m)\big) \cdot t_k(y) \qquad (11)$$

where the "branch" network produces coefficients b_k from the input function u sampled at m points, and the "trunk" network produces a basis of functions tk evaluated at the output coordinate y.

This structure is particularly advantageous in scientific contexts where one must predict not a single value, but rather a continuous field governed by PDEs, such as heat or mass distributions in transport problems.

DeepONets have garnered interest because they readily incorporate known boundary conditions or observational data while preserving their ability to extrapolate to new scenarios. They are further extensible to inverse problems and parametric studies, where one seeks to infer the material coefficients, force terms, or boundary conditions from sparse measurements. Their operator-centric perspective aligns well with the mathematics of PDEs and provides a compact representation that remains robust under domain or parameter changes.

The implementation hurdles for DeepONets also warrant further discussion. They can struggle to extract representative features from inputs with intricate structures, such as porous media, which are analogous to nanostructured materials. The data requirements for training can lead to significant memory (RAM) bottlenecks, particularly when both the input and output functions are defined over high-resolution spatial domains. Furthermore, for time-dependent problems, vanilla DeepONets can suffer from stability degradation and error accumulation during long-term prediction. This has necessitated architectural extensions such as Physics-

Informed Time-Integrated (PITI)-DeepONet, which is specifically designed to improve long-term accuracy by integrating the time dimension more robustly.

3.3.4 SINDy, Neural Implicit Flows, and Related Sparse Discovery Methods

While classical models assume that the governing equations are perfectly known, real-world nanoscale systems may exhibit unmodeled physics or emergent behaviors that are difficult to derive from first-principles. To tackle this challenge of model discovery, methods such as Sparse Identification of Non-linear Dynamics (SINDy) aim to discover governing equations or latent variables directly from the data. They merged sparse regression with a library of candidate functions to identify minimal, interpretable expressions describing the observed dynamics [21].

Given time-series data for state vector $x(t)$, SINDy constructs a library of candidate non-linear functions $\Theta(x) = [1, x, x^2, sin(x), \dots]$. It then solves a sparse regression problem to find a sparse matrix of coefficients $\boldsymbol{\Xi}$ that best fits the dynamics.

$$\frac{dx}{dt} = \boldsymbol{\Theta}(\boldsymbol{x})\boldsymbol{\Xi} \tag{12}$$

The nonzero elements of $\boldsymbol{\Xi}$ reveal the terms that constitute the governing differential equation.

When coupled with autoencoders or other neural components, SINDy extends to high-dimensional signals by learning a low-dimensional embedding that allows a concise set of governing equations. Such frameworks show promise in uncovering hidden variables or manifold structures, improving long-term predictive stability, and preserving interpretability, traits that are often compromised in purely black-box models.

Parallel to sparse regression initiatives are neural implicit flow models, which infer continuous transformations of probability densities without requiring the explicit inversion of large Jacobians. By adopting continuous formulations, these “normalizing flow” methods can embed physical constraints and invariants, offering yet another avenue to incorporate domain knowledge into generative modeling. Although these techniques have primarily appeared in computational fluid dynamics, they are increasingly being tested in broader contexts such as reactive transport in porous media or micro/nanoscale flows.

However, a critical discussion of SINDy highlights its primary weakness of extreme sensitivity to measurement noise, which is ubiquitous in nanoscale experiments. The core of the SINDy algorithm requires the estimation of time derivatives from data, a process that notoriously amplifies noise and can lead to the identification of spurious or incorrect physical terms in the discovered model. Several techniques have been developed to improve their robustness in real-world applications. These include Ensemble-SINDy (E-SINDy), which leverages bootstrap aggregating to build a more stable model from multiple fits on subsets of noisy data [21], and Bayesian-SINDy, which recasts the problem in a probabilistic framework to quantify uncertainty and improve model selection in noisy, data-scarce regimes.

3.3.5 Neural Ordinary Differential Equations

Neural ODEs [22] present an alternative method for building continuous-depth neural networks. Instead of stacking a discrete number of layers, NODE defines the transformation of its hidden state z as a continuous process governed by an ordinary differential equation (ODE), which is parameterized by a neural network f_θ:

$$\frac{dz(t)}{dt} = f_\theta(z(t), t) \tag{13}$$

The output of the "network" is then the solution to this ODE at a specific time T, found by integrating from the initial time t_0:

$$z(T) = z(t_0) + \int_{t_0}^{T} f_\theta(z(t), t)dt \tag{14}$$

This can be seen as a continuous analog of a Residual Network (ResNet), where a ResNet performs a discrete update:

$$z_{t+1} = z_t + f(z_t) \tag{15}$$

, whereas a NODE makes this update process continuous.

This continuous formulation provides two notable benefits: any modern adaptive ODE solver can be used for the forward pass, and the memory cost of backpropagation via the adjoint method is constant with respect to the depth. Neural ODEs and their stochastic or partial variants have attracted attention for tasks such as time-series forecasting, invertible transformations (continuous normalizing flows), and integration with physics-based constraints, where the learned function f_θ can represent unknown or complex parts of the dynamics of a physical system.

3.3.6 Hybrid ML–Physics Methods and physics-informed dynamic mode decomposition

As noted in Section 2.2, coupling different classical solvers in a hybrid multiscale framework is notoriously complex, often leading to instabilities at interfaces. PIML offers a more seamless path to the hybridization of classical physics solvers (e.g., finite element or spectral methods) with partial ML components to exploit the strengths of each domain. One prominent example is physics-informed dynamic mode decomposition (piDMD), where a low-order linear operator is sought but constrained to respect known physical principlessuch as energy conservation or shift invarianceenforced through manifold constraints [23]. By restricting the solution space to operators that satisfy domain knowledge, piDMD reduces overfitting, improves interpretability, and can match or surpass standard data-driven decompositions for fluid flows, waves, and other large-scale dynamic processes.

The core idea of Dynamic Mode Decomposition (DMD) is to approximate the system's evolution with an optimal linear operator A that maps the state x_k at one time step to the next: $x_{k+1} \approx Ax_k$. A physics-informed DMD (piDMD) then imposes physical constraints directly onto A—for instance, by requiring its eigenvalues to lie on the unit circle to enforce energy conservation in a non-dissipative system.

More broadly, hybrid strategies might embed a neural model for complicated subphenomena (e.g., boundary conditions or local constitutive laws) into a classical solver that handles global constraints. This modular design leverages the stability and mathematical rigor of well-established methods while harnessing the flexibility of ML to approximate difficult or uncertain sub-problems. Such partial MLphysics paradigms often align well with multiscale or multi-physics applications, a frequent scenario in nanoscale transport models. For example, in the simulation of a modern transistor, a classical finite-element solver can handle heat diffusion in the bulk silicon substrate, while a dedicated neural network learns the highly complex, non-linear thermal boundary resistance at the nanoscale metal-semiconductor interface, a value that is notoriously difficult to model from first principles.

3.3.7 Use Cases in Quantum Systems and Other Complex Domains

Quantum systems exemplify scenarios where ML-based approaches can reduce the computational overhead of traditional solvers, particularly for eigenvalue problems such as the time-independent Schrödinger equation [24].

$$\hat{H}\psi(x) = E\psi(x) \tag{16}$$

Here, a Physics-Informed Neural Network is trained to approximate the wavefunction $\psi_\theta(x)$, with both the wavefunction and energy eigenvalue E being the outputs of the learning process. The loss function is constructed to enforce the underlying physics without labeled data, and typically includes the following:

- A PDE residual term:

$$L_{PDE} = \left|\left|H\widehat{\psi_\theta}(x) - E\psi_\theta(x)\right|\right|^2 \quad (17)$$

to ensure the Schrödinger equation is satisfied.

- Boundary condition term: For example,

$$L_{BC} = \left|\left|\psi_\theta\left(x_{boundary}\right)\right|\right|^2 \quad (18)$$

for a particle in a box where the wavefunction must be zero at the boundaries.

- A normalization constraint:

$$L_{norm} = (\int |\psi_\theta(x)|^2 dx - 1)^2 \quad (19)$$

to enforce the probabilistic nature of the wavefunction.

By minimizing the total loss

$$L_{total} = L_{PDE} + L_{BC} + L_{norm} \quad (20)$$

PINN discovers the physically valid eigenstate and its corresponding energy. This approach has also been extended to complex PDEs in astrophysics and plasma physics [25].

3.3.8 Limitations and Future Prospects

Despite notable successes, pure ML-based PDE solvers still face significant challenges. Issues include high computational costs during training and difficulty in capturing high-frequency or boundary-layer phenomena. This is largely due to "spectral bias," the inherent tendency of neural networks optimized with gradient descent to learn smooth, low-frequency functions far more easily than high-frequency ones, causing them to struggle with sharp gradients or oscillatory solutions. Moreover, there is a risk of overfitting when data or physical constraints are limited [26]. In practice, many of the most successful applications to date are hybrid in nature, where a classical framework handles core numerical tasks, and a neural network is used selectively to approximate complex substructures or as a surrogate for large-scale parameter sweeps.

Ongoing work aims to improve scalability, automate architecture selection, and develop adaptive collocation or sampling schemes that refine training where PDE residuals are large. Incorporating more advanced physics priors, such as conservation laws, symmetries, or known invariants, should further enhance the model stability and interpretability. As research continues, these developments may pave the way for robust, widely used ML physics solvers that accelerate discovery and design in fields ranging from fluid mechanics and materials science to quantum information. A comparative overview of these key approaches is presented in Table 5.

4. Physics-Informed Neural Networks (PINNs) and Related Approaches

Although traditional machine learning is a powerful tool for data analysis, Physics-Informed Neural Networks (PINNs) are the quintessential embodiment of the paradigm of fusing data with physical laws. They construct the required bridge by embedding the governing equations directly into the learning process.

4.1. Core Concept: Why Physics-Informed Neural Networks?

Advances in data-centric methods have transformed the landscape of simulation and modeling; however, purely data-driven strategies often struggle when measurements are sparse or parameter spaces are large. Physics-Informed Neural Networks (PINNs) present an alternative by embedding governing physical laws, expressed as partial differential equations (PDEs) or integral constraints, directly into the training process of a neural network. In classical machine-learning approaches, one trains a model solely on input–output data, with no built-in guarantee that the predictions will adhere to known laws of physics. By contrast, PINNs incorporate PDE residuals, boundary conditions, and other domain knowledge into the loss

function, penalizing solutions that deviate from fundamental conservation or constitutive laws [25, 27, 28]. One can think of it in this way: a standard neural network is like a student who learns only from a sparse set of correct answers (the training data). A PINN is like a student who has the same answer sheet but also a "physics coach, " who constantly checks their work everywhere else, penalizing any reasoning that violates fundamental laws (the PDE residual). This analogy highlights the transformation of the workflow. A classical solver is a student who can solve only the exact problems they are given. The PINN, guided by its 'coach,' learns the underlying principles, allowing it to generate a continuous, physically valid solution across the entire domain—effectively turning a sparse data problem into a well-posed physics problem.

This principle becomes especially relevant for nanoscale transport problems, which often exhibit unique or extreme regimes (e.g., ballistic–diffusive conduction or strong thermal non-equilibrium). At such small scales, experiments may be expensive or challenging, yielding only partial or localized data—like temperature profiles at selected points or times. Moreover, classical continuum models (e.g., Fourier's law in heat conduction) can fail, and direct Boltzmann-based or kinetic models are computationally prohibitive for multi-dimensional, time-dependent problems [13, 38]. By fusing small experimental datasets with first-principles PDE constraints, PINNs offer a path toward physically consistent surrogate modeling that does not require dense measurements.

Several other factors underscore the appeal of PINNs in nanoscale transport:

Sparse or Partial Data: The ability to handle incomplete or noisy measurements is critical. PINNs can "fill in the gaps" by enforcing a PDE that governs the entire domain, effectively interpolating between known sensor points in a manner consistent with physical principles [42].

Complex PDEs: Submicron systems often require Boltzmann Transport Equation or fractional PDEs to capture memory effects and ballistic transport [13, 33, 38]. Traditional solvers in high-dimensional momentum or frequency spaces can be intractably large. PINNs, on the other hand, allow a mesh-free approach with the direct integration of boundary terms or scattering laws through the loss function.

Inverse/hybrid scenarios: In nanoscale research, unknown boundary conditions, scattering coefficients, or doping profiles may be as critical as the solution itself. PINNs can simultaneously solve these unknowns by treating them as learnable parameters under PDE constraints and effectively performing PDE-constrained optimization with minimal data [28, 29].

Hence, the "physics-informed" paradigm helps to avoid unphysical overfitting. The model is guided by fundamental transport equations in addition to standard data-driven loss terms, leading to solutions that remain faithful to the continuity, momentum, energy conservation, or scattering rules.

4.2. Architecture, Loss Function, and Training

4.2.1. General PINN Architecture

A typical PINN for solving PDE-based problems approximates the unknown solution $u_\theta(x,t)$ — or a more elaborate multi-variable function if the velocity, frequency, or polarization spaces are includedthrough a feed-forward neural network [20, 25]. The inputs to the neural network are the coordinate or state variables (e.g., space, time, and wavevector), and the outputs are physical fields, such as temperature, concentration, velocity potential, or mode-resolved distribution functions. Rather than performing finite-difference or finite-element schemes to approximate derivatives, one uses automatic differentiation on the neural network itself to compute the partial derivatives that appear in the PDE residual [25, 27, 40].

For example, when modeling phonon transport in a microdevice, one might feed (x,t,k) into a network that outputs the phonon distribution $f_\theta(x,t,k)$. The PDE residual includes the

streaming (advection in k-space) and collision terms. Because neural networks are universal function approximators, in principle, they can represent the underlying solution. Importantly, the topology of the network and the choice of activation functions (e.g., hyperbolic tangent vs. sine-based) can significantly impact the training speed and accuracy, especially when sharp gradients or wave-like solutions are involved [35, 39].

4.2.2. Loss Function Construction

The construction of this total loss function is the foundation of the PIML paradigm. Fusion of sparse, varied information sources—governing laws (L_{PDE}), boundary constraints ($L_{\text{BC/IC}}$), and sparse physical measurements (L_{data}) — occurs. The training process does not merely fit data; it searches for a solution that simultaneously satisfies all these constraints, making it a powerful tool for both forward prediction and inverse inference.PDE Residual: Let $\mathcal{N}$ denote the PDE operator, for instance:

$$\mathcal{N}[u_\theta(x_i)] = 0, \tag{21}$$

for collocation points $\{x_i\}_{i=1}^{N_{\text{col}}}$. The PDE residual loss is typically

$$L_{\text{PDE}} = \frac{1}{N_{\text{col}}} \sum_{i=1}^{N_{\text{col}}} (\mathcal{N}[u_\theta(x_i)])^2. \tag{22}$$

For the Boltzmann equation, $\mathcal{N}$ includes streaming (advection in the phase space) and collision integrals [13, 38]. In fractional PDEs, $\mathcal{N}$ represents operators with noninteger derivatives [33].

Boundary and Initial Conditions: Known boundaries or initial data can be applied via "soft constraints" (penalty terms) or "hard constraints" (analytic embedding). In the soft-constraint approach, the loss function has the following additional terms.

$$L_{\text{BC/IC}} = \frac{1}{N_{\text{BC}}} \sum_{j=1}^{N_{\text{BC}}} \left(u_\theta(x_j) - g_{\text{BC}}(x_j)\right)^2, \tag{23}$$

to ensure that the neural solution matches the measured or prescribed values g_{BC} at the boundary points $\{x_j\}_{j=1}^{N_{\text{BC}}}$ [25, 28]. Alternatively, a "hard constraint" technique analytically imposes the PDE constraints by rewriting the neural network output to satisfy the boundary conditions [28].

Experimental/observational data: In many nanoscale contexts, partial temperature or flux measurements are available using advanced metrology techniques (e.g., thermoreflectance). These data points can be integrated into a

$$L_{\text{data}} = \frac{1}{N_{\text{data}}} \sum_{k=1}^{N_{\text{data}}} \left(u_\theta(x_k) - u_{\text{obs}}(x_k)\right)^2, \tag{24}$$

ensuring that the solution is faithful to real observations [28, 29, 42].

The total loss L_{total} is typically a weighted sum of these terms:

$$L_{\text{total}} = \alpha_{\text{PDE}} L_{\text{PDE}} + \alpha_{\text{BC}} L_{\text{BC/IC}} + \alpha_{\text{data}} L_{\text{data}}, \tag{25}$$

where the weights α's are chosen heuristically or using adaptive techniques [29, 30].

4.2.3. Training and Optimization

Once the loss function is defined, training proceeds via gradient-based optimizers (e.g., Adam and RMSProp) to update the network parameters θ [27].

The core of the training process is to iteratively update the network parameters θ to minimize the total loss function L_total. This is achieved through gradient-based optimization, where the basic update rule is

$$\theta_{k+1} = \theta_k - \eta \nabla_\theta L_{\text{total}} \tag{26}$$

where θ_k represents the parameters at iteration k, η is the learning rate, and $\nabla_\theta L_{\text{total}}$ is the gradient of the loss with respect to the parameters computed via backpropagation.

Often, one follows with a second-stage optimizer (e.g., L-BFGS or quasi-Newton method) to refine the convergence. Domain decomposition methods such as conservative

PINNs (cPINNs) or extended PINNs (XPINNs) can further alleviate the difficulties posed by steep gradients or high-dimensional PDE spaces [30].

The multiterm loss can be unbalanced if the PDE residual is orders of magnitude different from the boundary/data residual, potentially causing optimization stagnation or "spectral bias." Techniques such as dynamic reweighting of loss terms or progressive training ("curriculum learning") help maintain equilibrium during optimization [29]. Despite these complexities, the synergy between PDE constraints and data typically yields solutions that track essential physics, even with small training sets.

4.3 Variants and Improvements of PINNs

As the PINN framework has matured, researchers have proposed numerous extensions to handle specialized PDE forms, multiresolution domains, or more advanced operator-learning scenarios:

4.3.1 Fractional PINNs

Fractional-order PDEs emerge in anomalous diffusion or in systems with non-local memory effects, as observed in nanoscale mass/charge transport in porous media [33]. For instance, fractional advection–diffusion equations can capture heavy-tailed probability distributions of random walks. As standard automatic differentiation cannot directly compute fractional derivatives, fractional PINNs rely on numerical quadratures or finite-difference stencils to approximate these operators. The PDE residual is then included in the loss function, allowing the model to learn solutions with long-range correlations or fractal scaling behaviors [33].

Because standard automatic differentiation cannot compute fractional derivatives, the fPINNs approximate them numerically. A common definition is the Grünwald-Letnikov fractional derivative of the order α:

$$D^{\alpha}f(x) = \lim_{h\to 0} \frac{1}{h^{\alpha}} \sum_{k=0}^{\lfloor x/h \rfloor} (-1)^k \binom{\alpha}{k} f(x - kh) \tag{27}$$

where $\binom{\alpha}{k}$ is the generalized binomial coefficient. This formulation shows that the derivative at point x depends on all past values of the function, capturing the memory effects inherent in many anomalous transport phenomena.

4.3.2 Physics-Based Activation Functions

A crucial challenge in PINNs is capturing oscillatory or steep-gradient solutions with standard activation functions (such as $tanh$). Incorporating domain knowledge into activation layers, for example, using sine functions for wave-dominated PDEs or exponential forms for decay processes, can improve the accuracy and reduce training epochs [35, 39]. Physical Activation Functions (PAFs) can directly embed known solution motifs into a neural network, enabling more efficient coverage of the solution space.

For example, to capture a wave-like solution, the network output $u_\theta(x,t)$ can be constructed as a composite function:

$$u_\theta(x,t) = N_\theta(x,t) \cdot \sin(\omega x - kt) + \text{mean} \tag{28}$$

In this structure, a standard neural network$N_\theta(x,t)$ learns the wave's amplitude modulation, while the `sin` function explicitly imposes the underlying oscillatory physics.

4.3.3 Neural Operators

Distinguishing the goal of a standard PINN from that of a neural operator is crucial. A standard PINN was trained to solve a single instance of a PDE with fixed parameters and boundary conditions. The network must be retrained if the boundary conditions change. By contrast, a neural operator learns the entire solution operator for a family of PDEs.

Rather than learning a single PDE solution, neural operators (Fourier Neural Operator, DeepONet) aim to learn the mapping from input functions (e.g., boundary conditions or material parameters) to solution fields [19, 20, 34]. This approach is suitable for repeated PDE solutions, such as parametric sweeps over doping concentrations in semiconductors or geometric changes in microdevices. Once trained, a neural operator can instantly infer the solution for new input functions, offering near-real-time PDE simulations for design or optimization loops.

This makes them exceptionally powerful for parametric design sweeps or uncertainty quantification, in which the governing PDE must be solved thousands of times with varying inputs. The tradeoff is that training a neural operator is significantly more data-hungry and computationally intensive than training a standard PINN for a single case.

4.3.4 Hybrid or Mechanistic–AI Approaches

Sometimes, domain decomposition or subphysics models can be coupled with a PINN. For instance, a partial solver can handle ballistic regimes of transport, whereas PINN manages diffusive subdomains with uncertain boundary fluxes [30, 36]. Alternatively, a "relaxation network" can approximate the complicated Boltzmann collision operator in momentum space, preserving conservation laws while speeding up full PDE solutions [42]. These hybrid frameworks allow large-scale or multiphysics problems, such as ballistic–diffusive phonon flows or multiionic reactive transport, to be addressed with more tractability.

4.3.5 Extended or Conservative PINNs

XPINNs and cPINNs partition the domain into subregions with separate networks, enforcing the continuity of the flux or solution across subdomain boundaries [30]. Such local networks can capture steep gradients or shock layers better by focusing on smaller pieces of the domain. In particular, cPINNs maintain global conservation properties by matching fluxes across interfaces, which is vital for correct mass and energy budgets [30].

These methods enforce physical constraints at the interface Γ_{ij} between the two subdomains Ω_i and Ω_j. For neural network solutions u_i and $u_j$in in each subdomain, the following conditions are enforced via the loss function:

- Continuity of the solution:

$$u_i(x) - u_j(x) = 0, \quad \text{for } x \in \Gamma_{ij} \tag{29}$$

- Continuity of the flux (for a diffusion problem):

$$\nabla u_i(x) \cdot n - \nabla u_j(x) \cdot n = 0, \quad \text{for } x \in \Gamma_{ij} \tag{30}$$

where $\boldsymbol{n}$ is the vector normal to the interface. This ensures that the global solution is physically consistent and conserves quantities, such as mass or energy, across the entire domain.

A summary comparison of these primary PINN variants is provided in Table 4.

Table 4
Таблиця 4
Comparison of Physics-Informed Neural Network Variants.
Порівняння варіантів фізико-інформованих нейронних мереж.

Variant	Core Idea	Best Use Case	Key Limitation
Standard PINN	Solves a single PDE instance using a PDE residual in the loss.	Inverse problems with sparse data; fixed geometry.	Must be retrained for each

			new problem instance.
fPINN	Incorporates fractional derivatives into the loss function.	Problems with anomalous diffusion or non-local effects.	Calculation of fractional derivatives can be complex.
XPINN / cPINN	Uses domain decomposition with separate NNs for subdomains.	Problems with sharp gradients, shocks, or complex geometries.	Enforcing continuity/flux conservation at interfaces adds complexity.
Neural Operator	Learns the solution operator for a family of PDEs.	Rapid parametric studies, optimization, UQ.	Requires a large dataset of solved PDE instances for training.

4.4 PINNs for Nanoscale Transport: Boltzmann Transport Equation

The primary motivation for applying PINNs to the Boltzmann Transport Equation was to combat the "curse of dimensionality." In its full form, the phonon BTE exists in 7-dimensional phase space (three spatial dimensions, three momentum/wavevector dimensions, and one time dimension). Traditional mesh-based solvers become computationally intractable in such high-dimensional spaces, whereas the mesh-free nature of PINNs offers a viable forward path.

4.4.1 Motivation and Scope

As device sizes drop to the nanometer regime, standard diffusion-based PDEs increasingly fail to accurately describe thermal or charge-carrier transport. The Boltzmann Transport Equation (BTE) provides a kinetic description of these carriers (phonons and electrons), capturing ballistic effects, mode-dependent scattering, and intricate boundary interactions. However, BTE, with high-dimensional phase spaces (spatial, momentum, frequency, polarization), can be computationally overwhelming for classical solvers [13].

Here, PINNs hold promise because:

- They can incorporate partial data from advanced nanoscale metrologies, for example, time-domain thermoreflectance for local flux, while still obeying ballistic–diffusive physics [38].
- They bypassed extensive mesh generation in the phase space by embedding PDE constraints into the loss.
- They allow parametric or inverse analyses of unknown scattering coefficients or boundary conditions within a single training workflow [38].

4.4.2 Illustrative Approaches

High ΔT Boltzmann PDE: Large temperature differences in ultrathin films or microdevices can induce strong phonon non-equilibrium, far from standard Fourier conduction.

A PINN approach can solve the stationary or transient BTE by penalizing collision integrals in the PDE residual, whereas boundary flux data from experiments or classical continuum approximations anchor the solution [13]. The network systematically adjusts the scattering parameters if they are left as free variables, thereby matching the PDE solution to the measured thermal flux.

Time-dependent Mode-Resolved BTE: In time-domain thermoreflectance experiments, short-pulsed laser heating and subsequent relaxation are monitored. Capturing this dynamic requires mode-resolved BTE in space and time [38]. Each phonon mode has unique relaxation times, velocities, and dispersion relationships. A PINN can approximate the distribution function $f_\theta(x, t, k)$, with the PDE residual enforcing streaming + collision. Sparse data from local temperature sensors or transient reflectance signals can help calibrate unknown boundary reflectivities or scattering rates.

Implementing physical boundary conditions: Because nanoscale transport is dominated by boundary scattering, the correct implementation of physical boundary conditions is essential for the accuracy of the BTE-PINN. Different physical scenarios, such as the specular versus diffuse reflection of phonons at an interface, must be encoded into the loss function. For specular reflection, the distribution of outgoing particles is a direct mirror of the incoming particles, whereas for diffuse reflection, the particles are re-emitted according to an equilibrium distribution. Enforcing these distinct mathematical forms within the PINN framework is a key area of research and is critical for accurately modeling phenomena, such as thermal boundary resistance.

Multi-Ionic or Multi-Carrier Transport: In doping or multi-ionic contexts (e.g., Li-ion transport and multi-species doping diffusion), partial differential equations become coupled across species, and local scattering or reaction rates are uncertain [36]. A PINN can unify these PDEs by sharing domain knowledge regarding flux continuity or electroneutrality. Instead of individually calibrating each PDE with expensive iterative solvers, the network simultaneously fits all species, enforcing PDE coupling constraints and partial data.

Structure-preserving Boltzmann gates: A primary challenge in formulating a BTE-PINN is the collision operator, $\partial f/\partial t_{coll}$. This term is not a simple derivative, but a high-dimensional integral operator that accounts for all possible scattering events, making its direct evaluation within the loss function computationally prohibitive. To tackle this, recent efforts propose training a neural network surrogate, such as a 'RelaxNet,' for the collision operator itself. [42] This is a non-trivial task, as the surrogate must be constrained to preserve physical invariants such as the conservation of mass, momentum, and energy. When a structure-preserving surrogate is integrated into the PDE residual, the computational overhead is drastically reduced by replacing the expensive integral with a learned mapping that still yields physically meaningful solutions.

While PINNs are 'mesh-free,' training a neural network in the BTE's 7-dimensional phase space (three spatial, three momentum/wavevector, one time) remains a monumental challenge. To overcome this 'curse of dimensionality' without resorting to costly discretizations, recent work has introduced the Monte Carlo Physics-Informed Neural Network (MC-PINN). The core innovation of this approach is a two-step sampling strategy in which points are first randomly sampled in the temporal-spatial domain and then separately in the solid angular domain. The final training points are constructed from the tensor product of these two sets, making the framework entirely mesh-free and avoiding the need for *a priori* angular discretization. This strategy has proven to be highly effective for multiscale heat conduction, successfully modeling the transport from the ballistic to the diffusive regime within a unified framework. Furthermore, it is remarkably memory-efficient; for a 3D ballistic transport problem, the MC-PINN requires only 16% of the memory of a state-of-the-art deterministic BTE solver [53].

4.4.3 Advantages and Current Challenges

Failure Mode and Mitigation. PINNs can deal with stiff collision operators, boundary layers, and high-frequency content (spectral bias). These difficulties are often rooted in the fundamentally complex and rugged loss landscape of physics-informed problems, which are fraught with spurious local minima that can trap standard gradient-based optimizers.

The spectral bias is detrimental. This well-documented challenge refers to the inherent tendency of standard neural networks optimized with gradient descent to learn smooth, low-frequency functions far more easily than high-frequency functions [52]. This limitation is critically relevant for nanoscale transport, a field rife with high-frequency phenomena, such as sharp temperature gradients at interfaces, ballistic phonon effects, and shock layers. When a PINN attempts to learn such a solution, the spectral bias can cause it to converge on an overly smooth approximation, failing to capture the physics that defines the nanoscale regime.

Several state-of-the-art mitigation strategies have been developed to address this critical mode of failure. These include:

- Multiscale or multi-grade network architectures that decompose the problem by frequency allow different parts of the network to specialize in learning different components of the solution.
- The use of adaptive activation functions can adjust the slope during training to better approximate steep gradients.
- Fourier feature mapping transforms the input coordinates into a higher-dimensional space where high-frequency components are represented by lower frequencies, making it significantly easier for the network to learn.

Remedies include domain decomposition (XPINNs/cPINNs), physics-aware activations (sine/exponential), curriculum schedules with loss reweighting, and structure-preserving surrogates for the collision terms. Where parametric reuse is required, consider training a neural operator against a family of BTE instances, and use a small PINN to enforce constraints locally.

By directly enforcing Boltzmann-based PDE constraints, PINNs for nanoscale conduction or electron transport are promising.

- Mesh-Free Handling: Sideskipping the need for a fine mesh in high-dimensional momentum or frequency space.
- Direct Data Integration: Surpassing purely numerical PDE solvers by unifying experimental data with ballistic–diffusive PDE constraints.
- Parametric or Inverse Solutions: Simultaneously learning unknown scattering parameters, boundary conditions, or doping profiles within the BTE framework.

However, major hurdles remain to be overcome. High dimensionality, stiff collision terms, or steep ballistic boundary layers can hamper the training convergence. In addition, guaranteeing global conservation (e.g., net energy and momentum) in strongly anisotropic or non-local domains may require specialized network architectures (e.g., cPINNs or structure-preserving surrogates). Hyperparameter tuning and domain decomposition remain active research areas for achieving robust accuracy at scale [38]. Nonetheless, successes in handling subcontinuum conduction highlight that PINNs aligned with domain-specific physical constraints can significantly advance nanoscale transport modeling [42].

Table 5
Таблиця 5
ML-focused summary: method family, short essence, advantages, typical limitations, and primary use cases.
ML-огляд: сімейство методів, короткий виклад, переваги, типові обмеження та основні варіанти використання.

Method (family)	**Short Essence**	**Advantages**	**Disadvantages / Limits**	**Typical Use**
PINNs	Solve single PDE instance via PDE-residual + data loss minimization	Data-efficient; inverse problems; mesh-free	Training instability; may struggle with stiff/high-freq regimes; retrain per-instance	Sparse/noisy data; fixed geometry inverse problems
Neural Operators (FNO, DeepONet)	Learn solution operator for PDE family	Orders-of-magnitude faster inference; generalizes to new inputs	Data-hungry; heavy training; sometimes grid/BC constraints	Design/optimization surrogates; UQ
Equation Discovery (SINDy)	Sparse regression to reveal governing equations	Interpretable, parsimonious models	Sensitive to noise; needs good library/derivatives	Discovering missing physics; model reduction
Symmetry-preserving NNs (HNN/LNN)	Encode conservation/symmetry via Hamiltonian/Lagrangian forms	Physical consistency and long-term stability	Limited for dissipative/forced systems (needs extensions)	Conservative dynamics, long-horizon trajectories

4.5 Concluding Remarks

Physics-Informed Neural Networks represent a critical shift in computational modeling; rather than relying solely on data or classical discretizations, PINNs couple both routes to produce solutions that adhere to the underlying physical laws. Their mesh-free nature and capacity to incorporate partial data make them particularly appealing for nanoscale transport, where Boltzmann-based or fractional PDE descriptions are mandatory; however, direct numerical solvers can be too costly. Furthermore, the recent surge in specialized variants, such as fractional PINNs, operator networks, and structure-preserving surrogates, illustrates that the method remains highly adaptable across diverse transport regimes.

Looking ahead, challenges such as optimizing loss weighting, handling multiple stiff PDE operators, and ensuring long-term stability in time-dependent problems require sophisticated training strategies. Continued innovation in domain decomposition (XPINNs and cPINNs) and physically aware network designs (custom activation functions and coupling classical solvers) can mitigate these issues. Research efforts have demonstrated that PINNs can unify partial measurements and sophisticated PDE operators in a single trainable pipeline, thereby opening new directions in nanoscale heat transfer, multi-ionic membrane transport, and quantum-scale electronics. As computational power increases and neural frameworks mature, PINNs promise to become a cornerstone in bridging experimental data and first-principles modeling for advanced nanoscale applications.

5. Data Challenges and Resources in Nanomaterial Research

Nanomaterial research is critically dependent on reliable data; however, obtaining high-fidelity datasets at the nanoscale remains a major bottleneck. A key distinction exists between computationally generated data derived from first-principles modeling and experimental databases that compile the measured results. Although both are essential, they present different challenges. High-accuracy simulations (e.g., molecular dynamics or Monte Carlo) are computationally expensive, limiting their scope to smaller systems or shorter timescales [14]. On the experimental side, measurements can exhibit substantial noise and variability due to sample inconsistencies or environmental fluctuations. This data landscape requires sophisticated strategies to bridge the gap between theory and reality.

5.1 The Landscape of Open Data Initiatives

To address these challenges, the materials science community has developed several crucial open-data initiatives. These platforms provide programmatic access to vast amounts of data, enabling high-throughput screening and training of machine learning models. They can be broadly categorized into foundational databases that provide simulation inputs, specialized databases with calculated transport properties, and curated experimental datasets for validation.

5.1.1. Foundational Computational Databases

These repositories serve as bedrock for most computational transport studies, providing the necessary inputs for BTE solvers.

Materials Project (MP) is a massive open-access database containing DFT-calculated properties of over 140,000 inorganic compounds. This is the de facto starting point for many studies, providing fundamental data such as crystal structures and electronic band structures, all accessible via a powerful API. Similarly, phonon properties, which are crucial for calculating thermal conductivity, can be found in specialized collections, such as the Figshare dataset by Petretto et al. [57] or calculated from MP structures.

Automatic FLOW for Materials Discovery (AFLOW) is another large DFT database with over 3.5 million entries. It serves a similar role to MP, but uses a different set of conventions and includes some directly calculated thermal properties using the AGL model for a subset of materials.

Specialized 2D Materials Databases, such as C2DB and 2DMatPedia, focus specifically on two-dimensional materials. They are invaluable for researching nanoscale transport in low-dimensional systems, providing not only structures and band structures but also key parameters for BTE calculations, such as deformation potentials and effective masses.

5.1.2. Specialized Transport Property Databases

These databases go a step further by performing computationally intensive BTE calculations and obtaining the resulting transport coefficients. Crucially, these high-fidelity

BTE solvers rely on the foundational computational databases mentioned previously, such as the Materials Project (MP) and AFLOW, as the primary sources of their required ab initio inputs (e.g., crystal structures and phonon properties). The large-scale datasets generated by these BTE solvers, in turn, become essential training data for data-hungry surrogate models, such as PINNs or Neural Operators.

The Dryad dataset by Ricci et al. [54] is a landmark resource that contains the electronic transport properties (conductivity and Seebeck coefficient) for nearly 48,000 materials from the Materials Project. These properties were calculated using BTE under the Constant Relaxation Time (CRT) approximation. This dataset is ideal for training ML models, but it is crucial to recognize its underlying physical simplification, as the CRT approximation neglects the energy-dependent scattering that is crucial in many nanoscale systems.

5.1.3. Key Experimental Datasets

Experimental databases are essential for validating computational models and grounding them in physical reality.

The Zenodo Interfacial Thermal Resistance (ITR) dataset [55] is a curated collection of experimentally measured ITR values from the scientific literature covering approximately 300 materials. Because interfacial scattering is a dominant mechanism in nanostructures, this database is a critical resource for validating models of thermal transport across interfaces. This collection serves as a critical benchmark for validating PIML models that aim to solve the inverse problem of ITR prediction. Furthermore, such curated experimental compilations have been successfully used to train classical ML models (e.g., Support Vector Machines, Gaussian Process Regression) to predict ITR, providing a useful baseline against which more advanced physics-informed approaches can be compared.

The OBELiX dataset by Hargreaves et al. [56] provides expertly curated experimental data on the ionic conductivities of over 800 solid-state lithium-ion conductors. While focusing on a specific application, it serves as the gold standard for validating models of ionic transport and demonstrates the value of carefully collected experimental data for machine learning.

5.2. Strategies to Handle Data Limitations

Even with these resources, researchers must employ robust strategies to address data limitations. Multifidelity data approaches, which creatively combine large but approximate computational datasets with sparse but high-accuracy experimental measurements, are particularly promising. Other key strategies include synthetic data generation to expand limited experimental datasets [14, 16], discrepancy modeling to learn the missing physics from experimental data [43], transfer learning to leverage knowledge from data-rich domains [45], and ensemble methods to enhance noise resilience [44].

Physics-informed methods such as PINNs are uniquely suited to this "small data" regime. The inclusion of governing equations in the loss function acts as a powerful regularizer, enabling physically consistent predictions, even when direct data are unavailable. The continued growth of open databases, combined with advanced, physics-aware ML frameworks, paves the way for the accelerated discovery and design of nanoscale transport.

Ultimately, this integration reframes the landscape of data resources as a strategic guide for researchers. The choice of PIML method is fundamentally linked to the available data regime. If a researcher has access to thousands of simulations across a parameter space—a 'big data' scenario—a Neural Operator becomes a viable and powerful tool for creating a fast surrogate model. Conversely, if the researcher has only a single, noisy experimental trajectory but a well-established governing PDE—a 'small data' scenario—a PINN is the more appropriate choice, using physics as a regularizer. Discussing this data-model symbiosis provides readers with a practical framework for selecting an appropriate tool for their specific research context.

6. Current Gaps and Future Research Directions

6.1. Bridging Data Scarcity and Multi-Fidelity Requirements

The chronic scarcity of high-fidelity experimental and simulation data for nanomaterials has long been a major bottleneck for traditional modeling. However, this is precisely an environment in which the PIML paradigm offers a transformative solution. PIML is uniquely designed to thrive on sparse data by embedding physical laws directly into the learning process, turning a perceived weakness into a tractable one. This opens the door for multifidelity approaches that strategically combine large, low-cost datasets with minimal high-accuracy information [4].

Multi-fidelity approaches promise to exploit both sparse high-accuracy datasets and inexpensive but approximate models. Coarse-grid continuum solvers or simplified ballistic–diffusive formalisms can map broad parameter spaces, whereas specialized BTE or Monte Carlo (MC) solvers refine local regions in which the system exhibits sharp gradients or ballistic transport [2, 9]. In machine learning terms, a single neural network may initially learn from large amounts of "low-fidelity" data to establish an approximate solution manifold, and then incorporate smaller amounts of high-fidelity data to correct local discrepancies.

Active learning is a key supporting technology, in which the model identifies which uncharted or poorly predicted regions would provide the greatest improvement if measured or computed at higher fidelity. For example, local sensitivities to boundary conditions or doping concentrations may be used to prioritize new simulations using a full BTE solver. By injecting only the most informative data points into the training, the total simulation cost can be substantially reduced. This is particularly helpful in multi-dimensional parameter spaces where naive random sampling becomes infeasible. Physical constraints, such as global energy conservation or phonon population balances, also help ensure that model predictions, even in regions with limited data, do not drift into non-physical regimes [13, 30]. Representative examples of nanomaterials are summarized in Table 6.

As HPC resources expand, multifidelity schemes can be applied iteratively at scale, each time refining a neural network surrogate to capture the elusive features of nanoscale transport. In practice, dynamic reweighting or "curriculum learning" can present simpler tasks, such as uniform geometries or moderate temperature gradients, and gradually incorporate more extreme or localized conditions. Over time, this synergy between HPC-based active learning and physics constraints should enable robust PIML models to operate with fewer data demands.

Table 6
Таблиця 6

Recent examples from nanoscale ML: system, phenomenon, methodology, key quantitative finding, and rationale.

Нещодавні приклади машинного навчання у нанорозмірних системах: система, явище, методологія, ключовий кількісний висновок та обґрунтування.

Nanomaterial system	Transport phenomenon	ML methodology	Key finding / Quantitative result	Additional details and rationale
Nanofluid Al2O3–	Convective heat transfer; reconstruction of thermal and	Continuous PINN (PDE solver)	Reconstructed fields with error <2%; outperform	PINNs reconstruct full thermal and hydrodynamic fields (pressure,

Cu/water [46]	hydrodynamic fields		classical surrogate models for Nusselt number and Fanning friction factor	temperature, velocity) in microchannel nanofluid convection; better prediction of efficiency metrics
Hybrid nanofluid Al2O3–Cu/water [46]	Heat transfer and flow; uncertainty handling with sparse data	Bayesian PINN (BIPINN)	Improved accuracy and generalization under sparse data	BIPINNs predict hybrid nanofluid performance while effectively accounting for uncertainty and limited data
Graphene / h-BN interface [47]	Interfacial thermal conductance; optimal defect structure design	Molecular Dynamics (MD) + CNN	Optimal defect configuration increases ITC by ≈97%	Hybrid MD+CNN explores millions of defect configurations to optimize heat transfer across the heterointerface
Graphene nanoribbon [48]	Thermal conductivity as a function of porosity	Molecular Dynamics (MD)	Porosity reduces thermal conductivity by ~90%	Porosity fundamentally alters phonon transport and thermal conductivity in graphene
Electrochemical systems (electrodes, electrolytes) [49]	Charge/ion transport, polarizability; potential energy surface modeling	E(3)-equivariant neural potential (PiNN package)	State-of-the-art performance for polarized electrodes; excellent results for liquid electrolytes	PiNN fits potential energy surfaces respecting physical symmetries; predicts quantum properties like dipole moments and charges
General nanostructures (e.g., Si, InAs,	Lattice thermal conductivity, phonon transport;	ShengBTE (iterative phonon BTE) / PINN for phonon BTE	ShengBTE: computes lattice κ and related quantities; PINN: successful	ShengBTE solves the linearized phonon BTE; PINNs have been applied to TD-BTE

lonsdaleite) [10]	phonon scattering		solution of time-dependent phonon BTE	
Kinetic theory of gases [42]	Approximation of the Boltzmann collision operator	RelaxNet (structure-preserving NN)	Solutions equivalent to reference Boltzmann results	Solution-dependent equilibrium state and relaxation frequency; trained on fast spectral data with case-specific fine-tuning
Interfacial thermal resistance (ITR) [51]	Predicting ITR across materials	SVM, GPR, LSBoost	Good agreement with experimental data	Classical ML regressors used to estimate ITR, a key parameter for thermal management in nanomaterials
Silicon (phonon MFP spectroscopy) [50]	Reconstruction of phonon mean free path (MFP) spectra	MFP spectroscopy technique	Recovered MFP spectra agree with first-principles calculations	Reconstructs MFP distributions from experimental data without parameter fitting, enabling nanoscale heat transport insight

6.2. Handling High-Dimensional PDEs with Reduced Complexity

When dealing with sub-100 nm transport, phonon scattering becomes highly dependent on the frequency, polarization, and direction, leading to a multi-dimensional phase space [2, 9]. Even after simplifying the BTE, fully resolving the ballistic–diffusive transition may require fine discretization in both real and momentum spaces. Similarly, quantum electron transport in extremely confined geometries entails discretizing wave functions across multiple dimensions. Such PDEs quickly become intractable to standard solvers.

In recent years, operator-based neural networks, such as Fourier Neural Operators (FNO) or Deep Operator Networks, have been developed to tackle this scenario precisely [19, 32]. Instead of learning a function that maps one discrete input (e.g., boundary condition fields) to a corresponding discrete output, these models learn mapping from function spaces to function spaces. In effect, the operator viewpoint implies that once the network is trained, it can produce solutions for different boundary conditions or parameters without having to solve the PDE from scratch. This “resolution-agnostic” property is especially appealing for multiscale transport problems, where a single fine-scale grid is often insufficient.

Furthermore, innovative sampling strategies, such as the two-step Monte Carlo approach for phonon BTE, can significantly reduce the memory and computational burden of high-

dimensional kinetic equations without sacrificing accuracy across ballistic-to-diffusive regimes [53].

Implementation details are important. Large 3D or 4D PDE domains are not trivially learned using a single neural operator. Domain decomposition: splitting the problem into sub-domains, each handled by smaller operator networks, can be combined with HPC parallelization. The decomposition approach also helps to enforce boundary conditions consistently and ensures that each subdomain model is not overwhelmed by high-dimensional complexity [12, 30]. For instance, a user may apply a specialized operator network near subcontinuum interfaces, where ballistic scattering dominates, and a more generic network in the device interior.

Realizing operator-based surrogates for BTE-like equations holds promise for rapid parametric design and real-time solutions. Important technical challenges include ensuring stable training on diverse geometries, capturing sharp discontinuities in the boundary layers, and performing spectral domain manipulations (e.g., Fourier transforms) for irregular device shapes. Extensions of the existing methods to unstructured meshes or adaptive meshes would further broaden their applicability. Once these methods have matured, they can dramatically accelerate the design cycles for next-generation transistors and nanophotonic devices by replacing repeated PDE solutions with near-instant surrogates.

6.3. Coupling Ballistic–Diffusive Frameworks and Physics-Informed ML

Modern devices with feature sizes below ~100 nm frequently exhibit neither purely diffusive nor ballistic transport. For example, in a thin silicon layer, phonons may behave ballistically near boundaries, but scatter sufficiently in the bulk to appear diffusive. Conventional continuum models (e.g., Fourier's law) fail to capture the boundary scattering, whereas purely ballistic BTE solutions are prohibitively expensive if extended to the entire domain [2, 11]. Hybrid frameworks that solve BTE near boundaries but revert to simpler continuum equations in the interior have emerged to address this complexity [12].

Physics-Informed neural networks (PINNs) allow PDE residuals to be enforced in the training loss, thereby allowing the neural network to approximate the solution directly [13, 38]. In a ballistic–diffusive scenario, the network may have to satisfy the BTE in near-boundary elements and the diffusion equation in the interior, with boundary conditions at the interface ensuring the continuity of temperature or flux. Such domain decomposition is well-suited for PINNs: one sub-network can handle the ballistic boundary layer, incorporating a partial phonon distribution or specialized flux boundary conditions, while another sub-neutralizing solution that deviates from fundamental conservation or constitutive lawspenalizing solutions that deviate from fundamental conservation or constitutive lawswork manages the diffusive interior region [30].

A significant challenge is frequency-dependent (or wavevector) scattering. Real materials have phonon modes with different mean free paths and scattering intensities, which vary with temperature or doping [4, 36]. In principle, PINNs can track these dependencies if the input dimension includes frequency bins. However, training such high-dimensional distributions can result in stiffness and instability. Specialized activation functions or local weighting in the loss function (to emphasize steep gradients near the boundaries) may be helpful. Another approach is to embed partial domain knowledge about scattering selection rules, thereby constraining the learned distributions of the network to remain physically meaningful [38, 42].

Longer term, robust ballistic–diffusive PINNs or operator networks can unify multiple transport pathways, for instance, ballistic phonons coexisting with diffusive electronswithin the same computational framework. Practical breakthroughs would reduce the time required to converge on solutions for complex device architectures such as gate-all-around nanowire transistors or 2D/3D heterogeneous stacks. As manufacturing processes push dimensions

deeper into the sub-10 nm realm, these hybrid methods could become a mainstay in thermal design workflows.

A particularly complex and promising direction is to quantitatively model the impact of size and geometry on transport by embedding physical boundary scattering models directly into the PDE residual of PINN. For instance, instead of merely enforcing a diffusive or specular boundary, a future PINN-BTE framework could incorporate terms representing surface-roughness scattering. This allows for the direct prediction of the thermal conductivity (κ) or electron mobility (μ) scales with a nanowire diameter or width of a 2D nanoribbon. Successful training of such a model would require not only solving the BTE, but also accurately representing the complex physics of carrier-boundary interactions within the loss function, which is a significant challenge at the frontier of physics-informed learning.

6.4. Parameter Estimation, Interfaces, and Uncertainty Quantification

In nanoscale systems, many parameters, from doping profiles to interface resistances to grain boundary scattering lengths, are difficult to measure directly [2, 4]. Even for simpler geometries, the presence of doping fluctuations or unknown oxide–semiconductor boundary conditions can drastically affect the device performance. Traditional inverse modeling often involves repeated PDE solving, each time the guesses for the unknown parameters are adjusted until the simulated outputs match the measured data. However, high-dimensional parameter spaces and expensive solvers can result in impractically high computational loads.

One emerging solution is to integrate Bayesian inference with physics-informed networks, rather than producing a single “best-fit” parameter set. These methods generate a posterior distribution that quantifies uncertainty [30, 36]. Physically, this means that the user gains not only a point estimate of, say, an interface thermal conductance, but also a statistical confidence interval. This is invaluable in design scenarios where tight tolerance margins matter, for example, ensuring that a local hotspot remains below a certain critical temperature.

Similarly, ensemble methods that build multiple neural approximations (Ensemble-PINNs or ensemble-SINDy) can evaluate the consistency of each candidate solution with both PDE constraints and measurement data [44]. Discrepancy modeling offers another perspective: a known PDE-based model is corrected by a learned term that captures unmodeled physics, such as extra scattering or boundary friction [43]. This discrepancy approach reduces the risk of overfitting by focusing on the learning capacity of the mismatch rather than re-deriving all known physics from scratch.

Bayesian or ensemble-based strategies can be computationally expensive, particularly for large-scale 3D devices. However, advanced Markov Chain Monte Carlo (MCMC) techniques—possibly combined with low-rank or operator-based surrogates—can mitigate these costs [19, 32]. In practice, a robust pipeline might start with a lower-dimensional version of the model, estimate parameters coarsely, and iteratively refine to higher fidelity while maintaining strict PDE-consistent constraints in the learning process. Such pipelines would yield more interpretable and reliable estimates than purely black box methods.

A significant future challenge lies in using PINNs to model the impact of defects and interfaces on transport based on first principles. The goal is to predict interfacial thermal resistance (ITR) not as a fitted parameter but as an emergent property. This would require a multi-level approach: first, constructing atomistic models of interfaces (e.g., in heterostructures) or defects; second, calculating their specific phononic or electronic properties; and finally, using a PINN-BTE solver, where the loss function explicitly includes physical interface scattering formalisms, such as the Acoustic Mismatch or Diffuse Mismatch models. The model predictions were then validated against curated experimental databases of the ITR values. This represents a true multiscale challenge, wedding quantum-level structural inputs to meso-scale transport phenomena through a physics-informed deep learning framework.

6.5. Real-Time Digital Twins and Control of Nanoscale Systems

The ultimate application of a fast, data-driven, and physically consistent modeling paradigm is the creation of real-time digital twins, a goal completely out of reach for traditional, computationally intensive solvers. The PIML workflow, particularly trained neural operators, makes this possible. By replacing the slow PDE with near-instantaneous surrogate models, a digital twin can ingest live sensor data from a physical device, infer the complete thermal and electrical state in real time, and inform a control loop to optimize performance and prevent failure. This represents the final step in the paradigm shift, moving from offline analysis to online adaptive control.

This is especially true for ballistic–diffusive phenomena that require multiscale PDE solutions [2, 11]. The concept of a digital twin, that is, a virtual model that mirrors the physical device in real time, demands a predictive engine that can ingest sensor data on the fly and update temperature or flux fields with minimal latency. A critical step toward this goal is the development of generalizable models that can adapt to new data or conditions without complete retraining. Meta-learning via neuroevolution provides a powerful framework. For instance, the "Baldwinian-PINN" approach uses an evolutionary algorithm to discover a model initialization that can be rapidly adapted to solve new PDE instances, such as those with different boundary conditions or material parameters, with a single near-instant update step. The ability to quickly specialize a pre-trained, physics-aware model is a key enabler for the practical deployment of real-time digital twins in nanoscale applications.[52]

Trained neural surrogates, particularly operator-based models, can produce full-field solutions in microseconds once properly trained [19, 41]. With integrated sensors providing boundary conditions or partial state measurements, these surrogates could be re-tuned “online” to correct for drift or unforeseen disturbances. In the simplest scenario, a PIML model runs in parallel with the device, receiving measured data (e.g., local transistor temperatures or a set of thermal sensor readings) and rapidly inferring a global heat map. This map then feeds a control loop that adjusts the operating voltages, fan speeds, or gating patterns.

Sensor Integration: The device must be instrumented with appropriate thermal or electrical sensors, whose data can be fed into the surrogate network in real time.

Model order reduction: An operator-based model can be too large if the domain is complex. Domain-decomposition strategies or multifidelity subnetwork approaches help maintain a low real-time inference overhead.

Robustness Under Uncertainty: Real sensor data can be noisy or partial. Probabilistic frameworks, data assimilation, and regularization can maintain the stability of the model and prevent erratic updates.

In advanced chip designs, dynamic thermal management might involve partially “throttling” certain areas, redistributing tasks to cooler regions, or reconfiguring doping as the operating conditions evolve. These capabilities require the synergy of HPC-grade offline training, physics-informed constraints, and real-time streaming sensor data. Success in this area would significantly reduce the risk of thermal overstress in nanoscale devices and open new frontiers in adaptive or self-optimizing nanoelectronics.

6.6. Toward Unified Hybrid Methods and Interoperability

Historically, multiscale modeling workflows patch together separate tools: a quantum solver for active regions, a semiclassical BTE solver for mid-range scattering, and a continuum solver for heat conduction in the substrate. Each submodel may use different numerical schemes and require bridging variables or boundary conditions at the interfaces [2, 11]. This fragmentation can introduce inconsistencies or numerical instabilities, particularly if the sub-model codes are proprietary or not designed for straightforward coupling.

A unified PIML framework incorporates multiple levels of physical approximations in a single trainable environment. For instance, one module might handle ballistic phonon boundary scattering via partial distribution functions, another might approximate continuum conduction in the device interior, and yet another might incorporate quantum corrections near the contact region. Coupling occurs at shared boundaries and is enforced as part of a joint loss function that ensures flux continuity, phase coherence, or energy conservation [30],[38]. Open-data and open-source standards for domain decomposition would allow HPC codes to communicate seamlessly with specialized neural modules [32, 44].

Cross-Module Consistency: Different domains (e.g., ballistic vs. diffusive) can adopt separate coordinate systems or finite-element representations. Ensuring a stable handoff of boundary conditions is non-trivial. Operator-based networks that handle geometric transformations or mesh invariance may be useful.

Scalability: Large-scale HPC resources are still required to train complex multimodule networks, especially if each submodule includes thousands of trainable parameters. Techniques, such as asynchronous parallel updates or distributed training across HPC clusters, may mitigate the overhead.

Community Adoption: Achieving standardization for model definitions, boundary condition formats, and data handoffs is a social-technical challenge. Collaboration among academic research groups, software developers, and industry stakeholders is required.

If realized, an interoperable PIML ecosystem would accelerate fundamental research on electron–phonon coupling, interface scattering, and ballistic–diffusive transitions. It also translates quickly into industrial design workflows, where reliability and performance are paramount. In the best-case scenario, researchers and engineers can mix and match specialized modules, such as a quantum tight-binding solver, phonon BTE surrogate, or continuum PDE solver, without rewriting entire codebases, ultimately speeding up both fundamental discoveries and practical device development.

6.7 Advanced Optimization and Generalization with Neuroevolution

Although improvements in network architecture and loss weighting are crucial, a more fundamental challenge lies in the optimization process itself. The limitations of gradient-based methods for navigating the complex loss landscapes of PINNs have motivated the development toward alternative strategies [52].

Physics-Informed Neuroevolution (PINE), which employs gradient-free, population-based evolutionary algorithms (EAs) for PINN training, has emerged as a promising future direction. Unlike point-based gradient descent, EAs perform a global search that is less susceptible to trapping in poor local minima [52].

This paradigm offers several advantages. First, multi-objective EAs can naturally handle the competing terms in the PINN loss function (e.g., PDE residual vs. boundary conditions) without requiring heuristic weight tuning, instead identifying the entire Pareto front of the optimal trade-off solutions. Second, neuroevolution can be used for automated [52].

Neural Architecture Search (NAS) was used to discover bespoke network topologies and custom activation functions tailored to the physics of a specific transport problem. Finally, these methods show great promise for creating generalizable models, as discussed below [52].

7. Conclusion

Modeling nanoscale transport has long presented a stark choice: the fidelity of first-principles solvers at the cost of prohibitive computation or the speed of continuum models at the cost of physical accuracy. This review has charted the emergence of a new approach, Physics-Informed Machine Learning, that resolves this dilemma not with a better tool, but with a fundamentally transformed scientific workflow—a synergistic bridge that fuses the rigor of

physical law with the flexibility of machine learning. By embedding fundamental governing equations, such as the Boltzmann Transport Equation, directly into the training loss of a neural network, PIML fundamentally changes the modeling paradigm. This approach enables the fusion of sparse and noisy experimental data with physical laws, thereby allowing the creation of surrogate models that are both accurate and computationally efficient. As demonstrated, these methods are uniquely equipped to handle complex geometries, couple multiphysics phenomena across ballistic and diffusive regimes, and solve challenging inverse problems to uncover unknown material parameters from limited observations.

However, the ultimate potential of PIML extends far beyond the acceleration of the existing simulations. This paradigm shift paves the way for true inverse design, a long-sought-after goal in material science and device engineering. Instead of merely analyzing the transport properties of a given nanostructure, we can design novel materials and devices with precisely tailored thermal and electrical characteristics. Operator-learning networks, which learn the entire solution mapping for a family of PDEs, are a critical enabler of this vision, allowing for near-instantaneous parameter sweeps and design optimization that would be intractable with traditional solvers. This capability forms the bedrock for creating real-time digital twins of nanodevices, which could enable adaptive thermal management and on-the-fly performance optimization in next-generation electronics.

Despite its transformative potential, the path to widespread adoption is not without obstacles. Significant challenges remain in handling the high dimensionality and stiffness of kinetic equations, ensuring robust uncertainty quantification for mission-critical applications, and developing models that can be generalized to new problems without costly retraining. Therefore, the grand challenge for the next decade will be to move beyond bespoke solutions and create unified, interoperable PIML frameworks capable of seamlessly integrating quantum, atomistic, and continuum physics within a single environment. Realizing this vision will demand unprecedented interdisciplinary collaboration among physicists, material scientists, computational experts, and engineers. By successfully uniting first-principles physics with scalable machine learning architectures, these integrated frameworks will be indispensable in engineering next-generation nanotechnologies. By fully realizing this integration, PIML will move beyond being a tool for analysis and become the primary engine for inverse design, enabling scientists and engineers to specify the desired transport properties and generate novel materials and device architectures that can achieve them. This is the ultimate goal of this new paradigm.

References

1. David G. Cahill, Paul V. Braun, Gang Chen, David R. Clarke, Shanhui Fan, Kenneth E. Goodson, Pawel Keblinski, William P. King, Gerald D. Mahan, Arun Majumdar, Humphrey J. Maris, Simon R. Phillpot, Eric Pop, Li Shi; Nanoscale thermal transport. II. 2003–2012. *Appl. Phys. Rev.* 1 March 2014; 1 (1): 011305. Retrieved from https://doi.org/10.1063/1.4832615
2. Pop, E., Sinha, S., & Goodson, K. E. (2006). Heat Generation and Transport in Nanometer-Scale Transistors. *Proceedings of the IEEE, 94*, 1587–1601. Retrieved from https://doi.org/10.1109/JPROC.2006.879794
3. Ideue, T., & Iwasa, Y. (2021). Nonlinear Electrical Transport in Van der Waals Nanomaterials. *Annual Review of Condensed Matter Physics, 12*, 201–223. Retrieved from https://doi.org/10.1146/annurev-conmatphys-060220-100347
4. Zeng, L., Collins, K. C., Hu, Y., Luckyanova, M. N., Maznev, A. A., et al. (2015). Measuring phonon mean free path distributions by probing quasiballistic phonon transport in grating nanostructures. *Scientific Reports, 5*, 17131. Retrieved from https://doi.org/10.1038/srep17131

5. Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics, 3*, 218–229. Retrieved from https://doi.org/10.1038/s42254-021-00314-5
6. Wang, R., & Yu, R. (2025). *Physics-Guided Deep Learning for Dynamical Systems: A Survey*. *ACM Computing Surveys*. Retrieved from https://doi.org/10.1145/3766887
7. Hao, Z., Liu, S., Zhang, Y., et al. (2023). Physics-Informed Machine Learning: A Survey on Problems, Methods and Applications. *Tsinghua University Survey* (preprint). Retrieved from https://doi.org/10.48550/arXiv.2211.08064
8. Farea, A., Yli-Harja, O., & Emmert-Streib, F. (2024). Understanding physics-informed neural networks: Techniques, applications, trends, and challenges. *AI, 5*(3), 1534–1557. Retrieved from https://doi.org/10.3390/ai5030074
9. Lukes, J. R., & Zhong, H. (2007). Thermal conductivity of individual single-wall carbon nanotubes. *ASME J. Heat Transf., 129*(705-715). Retrieved from https://doi.org/10.1115/1.2717242
10. Li, W., Carrete, J., Katcho, N. A., & Mingo, N. (2014). ShengBTE: A solver of the Boltzmann transport equation for phonons. *Computer Physics Communications, 185*(1747-1758). Retrieved from https://doi.org/10.1016/j.cpc.2014.02.015
11. Majumdar, A. (1993). Microscale heat conduction in dielectric thin films. *Journal of Heat Transfer, 115*(7). Retrieved from https://doi.org/10.1115/1.2910673
12. Li, H.-L., Hua, Y.-C., & Cao, B.-Y. (2018). A hybrid phonon Monte Carlo-diffusion method for ballistic-diffusive heat conduction in nano- and micro-structures. *International Journal of Heat and Mass Transfer, 127*(1014-1022). Retrieved from https://doi.org/10.1016/j.ijheatmasstransfer.2018.06.080
13. Li, R., Wang, J.-X., Lee, E., & Luo, T. (2022). Physics-informed deep learning for solving phonon Boltzmann transport equation with large temperature non-equilibrium. *npj Computational Materials, 8*(29). Retrieved from https://doi.org/10.1038/s41524-022-00712-y
14. Badar, M. S., Shamsi, S., Ahmed, J., & Alam, A. (2020). Molecular dynamics simulations: Concept, methods, and applications. In *Advances in Biological Science*. Springer. Retrieved from https://doi.org/10.1007/978-3-030-94651-7_7
15. Bock, F. E., Aydin, R. C., Cyron, C. J., Huber, N., Kalidindi, S. R., & Klusemann, B. (2019). A review of the application of machine learning and data mining approaches in continuum materials mechanics. *Frontiers in Materials, 6*, 110. Retrieved from https://doi.org/10.3389/fmats.2019.00110
16. Bishara, D., Xie, Y., Liu, W. K., & Li, S. (2022). A state-of-the-art review on machine learning-based multiscale modeling, simulation, homogenization, and design of materials. *Archives of Computational Methods in Engineering*. Retrieved from https://doi.org/10.1007/s11831-022-09795-8
17. Greydanus, S., Dzamba, M., & Yosinski, J. (2019). Hamiltonian Neural Networks. *ArXiv*. Retrieved from https://doi.org/10.48550/arXiv.1906.01563
18. Cranmer, M., Greydanus, S., Hoyer, S., Battaglia, P., Spergel, D., & Ho, S. (2020). Lagrangian neural networks. *arXiv*. Retrieved from https://doi.org/10.48550/arXiv.2003.04630
19. Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2021). Fourier neural operator for parametric partial differential equations. *Proceedings of ICLR 2021*. Retrieved from https://doi.org/10.48550/arXiv.2010.08895
20. Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators.

Nature Machine Intelligence, 3(3), 218–229. Retrieved from https://doi.org/10.1038/s42256-021-00302-5

21. Champion, K., Lusch, B., Kutz, J. N., & Brunton, S. L. (2019). Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences, 116*(45), 22445–22451. Retrieved from https://doi.org/10.1073/pnas.1906995116
22. Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. (2018). Neural ordinary differential equations. In *Advances in Neural Information Processing Systems 31* (pp. 6572–6583). Curran Associates. Retrieved from https://papers.nips.cc/paper/7892-neural-ordinary-differential-equations
23. Baddoo, P. J., Herrmann, B., McKeon, B. J., Kutz, J. N., & Brunton, S. L. (2023). Physics-informed dynamic mode decomposition. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 479*(2271), 20220576. Retrieved from https://doi.org/10.1098/rspa.2022.0576
24. Jin, H., Mattheakis, M., & Protopapas, P. (2022). Physics-Informed Neural Networks for Quantum Eigenvalue Problems. *2022 International Joint Conference on Neural Networks (IJCNN)*, 1-8. Retrieved from https://doi.org/10.48550/arXiv.2203.00451
25. Baty, H. (2024). A hands-on introduction to Physics-Informed Neural Networks for solving partial differential equations with benchmark tests taken from astrophysics and plasma physics *[Preprint]*. *HAL*. Retrieved from https://hal.science/hal-04491808v1
26. Markidis, S. (2021). The old and the new: Can physics-informed deep-learning replace traditional linear solvers? *Frontiers in Big Data, 4*(669097). Retrieved from https://doi.org/10.3389/fdata.2021.669097
27. Costabal, F. S., Pezzuto, S., & Perdikaris, P. (2023). Δ-PINNs: Physics-informed neural networks on complex geometries. *Engineering Applications of Artificial Intelligence, 118*, 107324. Retrieved from https://doi.org/10.1016/j.engappai.2023.107324
28. Chen, S., Liu, Z., Zhang, W., & Yang, J. (2024). A hard-constraint wide-body physics-informed neural network model for solving multiple cases in forward problems for partial differential equations. *Applied Sciences, 14*(1), 189. Retrieved from https://doi.org/10.3390/app14010189
29. Krishnapriyan, A., Gholami, A., Zhe, S., Kirby, R., & Mahoney, M. W. (2021). Characterizing possible failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems 34* (NeurIPS 2021). Retrieved from https://proceedings.neurips.cc/paper/2021/file/df438e5206f31600e6ae4af72f2725f1-Paper.pdf
30. Jagtap, A. D., Kharazmi, E., & Karniadakis, G. E. (2020). Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems. *Computer Methods in Applied Mechanics and Engineering, 365*, 113028. Retrieved from https://doi.org/10.1016/j.cma.2020.113028
31. Rohrhofer, F. M., Posch, S., Gößnitzer, C., & Geiger, B. C. (2023). Data vs. Physics: The Apparent Pareto Front of Physics-Informed Neural Networks. *IEEE Access, 11*, 86252–86261. Retrieved from https://doi.org/10.1109/ACCESS.2023.3302892
32. Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2023). Neural Operator: Learning maps between function spaces. *Journal of Machine Learning Research, 24*, 4061-4157. Retrieved from https://doi.org/10.5555/3648699.3648788
33. Pang, G., Lu, L., & Karniadakis, G. E. (2019). fPINNs: fractional physics-informed neural networks. *SIAM Journal on Scientific Computing***,** 41(4), A2603–A2626. Retrieved from https://doi.org/10.1137/18M1229845
34. He, J., Pal, D., Najafi, A., Abueidda, D., Koric, S., & Jasiuk, I. (2024). Material-Response-Informed DeepONet and its application to polycrystal stress–strain prediction

in crystal plasticity. *JOM, 76*(10), 5744–5754. Retrieved from https://doi.org/10.1007/s11837-024-06681-5

35. Abbasi, J., & Andersen, P. Ø. (2024). Physical activation functions (PAFs): An approach for more efficient induction of physics into physics-informed neural networks (PINNs). *Neurocomputing, 608*, 128352. Retrieved from https://doi.org/10.1016/j.neucom.2024.128352
36. Rehman, D., & Lienhard, J. H. (2023). Physics-constrained neural differential equations for learning multi-ionic transport. *ICLR 2023 Workshop on Physics for Machine Learning*. Retrieved from https://doi.org/10.48550/arXiv.2303.04594
37. Cai, S., Wang, Z., Wang, S., Perdikaris, P., & Karniadakis, G. E. (2021). Physics-informed neural networks (PINNs) for heat transfer problems. *ASME Journal of Heat Transfer, 143*(6). Retrieved from https://doi.org/10.1115/1.4050542
38. hou, J., Li, R., & Luo, T. (2023). Physics-informed neural networks for solving time-dependent mode-resolved phonon Boltzmann transport equation. *npj Computational Materials, 9*(212). Retrieved from https://doi.org/10.1038/s41524-023-01165-7
39. Faroughi, S. A., Soltanmohammadi, R., Datta, P., Mahjour, S. K., & Faroughi, S. (2024). Physics-Informed Neural Networks with Periodic Activation Functions for Solute Transport in Heterogeneous Porous Media. *Mathematics, 12*(63). Retrieved from https://doi.org/10.3390/math12010063
40. Raissi, M., Perdikaris, P., & 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*(16), 686–707. Retrieved from https://doi.org/10.1016/j.jcp.2018.10.045
41. Dave, A. J., & Vilim, R. B. (2024). Physics-informed state-space neural networks for transport phenomena. *Engineering Applications of Artificial Intelligence, 133*(Part C), 108245. Retrieved from https://doi.org/10.1016/j.engappai.2024.108245
42. Xiao, T., & Frank, M. (2023). RelaxNet: A structure-preserving neural network to approximate the Boltzmann collision operator. *Journal of Computational Physics, 490*, 112317. Retrieved from https://doi.org/10.1016/j.jcp.2023.112317
43. Kaheman, K., Kaiser, E., Strom, B., Kutz, J. N., & Brunton, S. L. (2019). Learning discrepancy models from experimental data. *arXiv*. Retrieved from https://doi.org/10.48550/arXiv.1909.08574
44. Fasel, U., Kutz, J. N., Brunton, B. W., & Brunton, S. L. (2022). Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proceedings of the Royal Society A, 478*(20210904). Retrieved from https://doi.org/10.1098/rspa.2021.0904
45. Xu, C., Cao, B. T., Yuan, Y., & Meschke, G. (2023). Transfer learning based physics-informed neural networks for solving inverse problems in engineering structures under different loading scenarios. *Computer Methods in Applied Mechanics and Engineering, 405*, 115852. Retrieved from https://doi.org/10.1016/j.cma.2022.115852
46. Li, Y., Liu, T., & Xie, Y. (2022). Thermal fluid fields reconstruction for nanofluids convection based on physics-informed deep learning. *Scientific Reports, 12*, Article 12567. Retrieved from https://doi.org/10.1038/s41598-022-16463-1
47. Yang, H., Li, L., & Yang, P. (2025). Deep learning-based molecular dynamics simulation reconfiguration of efficient heat energy transport of Gra/h-BN heterointerface. *Applied Physics Letters, 127*(2), 021601. Retrieved from https://doi.org/10.1063/5.0270700
48. Gomes, D. S., Félix, I. M., Radel, W. F., Dias, A. C., Ribeiro Jr., L. A., & Pereira Jr., M. L. (2025). Computational characterization of the recently synthesized pristine and porous 12-atom-wide armchair graphene nanoribbon. *Nano Letters, 25*(21), 8596–8603. Retrieved from https://doi.org/10.1021/acs.nanolett.5c01319

49. Li, J., Knijff, L., Zhang, Z.-Y., Andersson, L., & Zhang, C. (2025). PiNN: Equivariant neural network suite for modeling electrochemical systems. *Journal of Chemical Theory and Computation, 21*(3), 1382–1395. Retrieved from https://doi.org/10.1021/acs.jctc.4c01570
50. Zeng, L., Collins, K. C., Hu, Y., Luckyanova, M. N., Maznev, A. A., Huberman, S., Chiloyan, V., Zhou, J., Huang, X., Nelson, K. A., & Chen, G. (2015). Measuring phonon mean free path distributions by probing quasiballistic phonon transport in grating nanostructures. *Scientific Reports, 5*, 17131. Retrieved from https://doi.org/10.1038/srep17131
51. Wu, Y.-J., Fang, L., & Xu, Y. (2019). Predicting interfacial thermal resistance by machine learning. *npj Computational Materials, 5*, Article 56. Retrieved from https://doi.org/10.1038/s41524-019-0193-0
52. Wong, J. C., Gupta, A., Ooi, C. C., Chiu, P.-H., Liu, J., & Ong, Y.-S. (2025). Evolutionary optimization of physics-informed neural networks: Evo-PINN frontiers and opportunities. *arXiv*. Retrieved from https://doi.org/10.48550/arXiv.2501.06572
53. Lin, Q., Zhang, C., Meng, X., & Guo, Z. (2025). Monte Carlo physics-informed neural networks for multiscale heat conduction via phonon Boltzmann transport equation. *Journal of Computational Physics, 542*, 114364. Retrieved from https://doi.org/10.1016/j.jcp.2025.114364
54. Ricci, F., Chen, W., Aydemir, U. *et al.* (2017). An *ab initio* electronic transport database for inorganic materials. *Sci Data* 4, 170085. Retrieved from https://doi.org/10.1038/sdata.2017.85
55. Wu, Y.-J., Zhan, T., Hou, Z., Fang, L., & Xu, Y. (2020). Physical and chemical descriptors for predicting interfacial thermal resistance. *Scientific Data, 7*, 36. Retrieved from https://doi.org/10.1038/s41597-020-0373-2
56. Hargreaves, C.J., Gaultois, M.W., Daniels, L.M. *et al.* (2023). A database of experimentally measured lithium solid electrolyte conductivities evaluated with machine learning. *npj Comput Mater* **9**, 9. Retrieved from https://doi.org/10.1038/s41524-022-00951-z
57. Petretto, G., Dwaraknath, S., Miranda, H. P. C., Winston, D., Giantomassi, M., van Setten, M. J., Gonze, X., Persson, K. A., Hautier, G., & Rignanese, G.-M. (2018). High-throughput density-functional perturbation theory phonons for inorganic materials. Scientific Data, 5, Article 180065. Retrieved from https://doi.org/10.1038/sdata.2018.65

Список використаної літератури:

1. Cahill D. G. Nanoscale thermal transport. II. 2003–2012 / D. G. Cahill, P. V. Braun, G. Chen, D. R. Clarke, S. Fan, K. E. Goodson, P. Keblinski, W. P. King, G. D. Mahan, A. Majumdar, H. J. Maris, S. R. Phillpot, E. Pop, L. Shi // Applied Physics Reviews. – 2014. – Vol. 1, № 1: 011305. – Режим доступу: https://doi.org/10.1063/1.4832615
2. Pop E. Heat Generation and Transport in Nanometer-Scale Transistors / E. Pop, S. Sinha, K. E. Goodson // Proceedings of the IEEE. – 2006. – Vol. 94. – P. 1587–1601. – Режим доступу: https://doi.org/10.1109/JPROC.2006.879794
3. Ideue T. Nonlinear Electrical Transport in Van der Waals Nanomaterials / T. Ideue, Y. Iwasa // Annual Review of Condensed Matter Physics. – 2021. – Vol. 12. – P. 201–223. – Режим доступу: https://doi.org/10.1146/annurev-conmatphys-060220-100347
4. Zeng L. Measuring phonon mean free path distributions by probing quasiballistic phonon transport in grating nanostructures / L. Zeng, K. C. Collins, Y. Hu, M. N. Luckyanova, A. A. Maznev, та ін. // Scientific Reports. – 2015. – Vol. 5: 17131. – Режим доступу: https://doi.org/10.1038/srep17131

5. Karniadakis G. E. Physics-informed machine learning / G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang // Nature Reviews Physics. – 2021. – Vol. 3. – P. 218–229. – Режим доступу: https://doi.org/10.1038/s42254-021-00314-5
6. Wang R. Physics-Guided Deep Learning for Dynamical Systems: A Survey / R. Wang, R. Yu // ACM Computing Surveys. – 2025. – Режим доступу: https://doi.org/10.1145/3766887
7. Hao Z. Physics-Informed Machine Learning: A Survey on Problems, Methods and Applications / Z. Hao, S. Liu, Y. Zhang, та ін. // Tsinghua University Survey (preprint). – 2023. – Режим доступу: https://doi.org/10.48550/arXiv.2211.08064
8. Farea A. Understanding physics-informed neural networks: Techniques, applications, trends, and challenges / A. Farea, O. Yli-Harja, F. Emmert-Streib // AI. – 2024. – Vol. 5, № 3. – P. 1534–1557. – Режим доступу: https://doi.org/10.3390/ai5030074
9. Lukes J. R. Thermal conductivity of individual single-wall carbon nanotubes / J. R. Lukes, H. Zhong // ASME Journal of Heat Transfer. – 2007. – Vol. 129. – P. 705–715. – Режим доступу: https://doi.org/10.1115/1.2717242
10. Li W. ShengBTE: A solver of the Boltzmann transport equation for phonons / W. Li, J. Carrete, N. A. Katcho, N. Mingo // Computer Physics Communications. – 2014. – Vol. 185. – P. 1747–1758. – Режим доступу: https://doi.org/10.1016/j.cpc.2014.02.015
11. Majumdar A. Microscale heat conduction in dielectric thin films / A. Majumdar // Journal of Heat Transfer. – 1993. – Vol. 115. – P. 7. – Режим доступу: https://doi.org/10.1115/1.2910673
12. Li H.-L. A hybrid phonon Monte Carlo-diffusion method for ballistic-diffusive heat conduction in nano- and micro-structures / H.-L. Li, Y.-C. Hua, B.-Y. Cao // International Journal of Heat and Mass Transfer. – 2018. – Vol. 127. – P. 1014–1022. – Режим доступу: https://doi.org/10.1016/j.ijheatmasstransfer.2018.06.080
13. Li R. Physics-informed deep learning for solving phonon Boltzmann transport equation with large temperature non-equilibrium / R. Li, J.-X. Wang, E. Lee, T. Luo // npj Computational Materials. – 2022. – Vol. 8: 29. – Режим доступу: https://doi.org/10.1038/s41524-022-00712-y
14. Badar M. S. Molecular dynamics simulations: Concept, methods, and applications / M. S. Badar, S. Shamsi, J. Ahmed, A. Alam // In: Advances in Biological Science. – Springer, 2020. – Режим доступу: https://doi.org/10.1007/978-3-030-94651-7_7
15. Bock F. E. A review of the application of machine learning and data mining approaches in continuum materials mechanics / F. E. Bock, R. C. Aydin, C. J. Cyron, N. Huber, S. R. Kalidindi, B. Klusemann // Frontiers in Materials. – 2019. – Vol. 6: 110. – Режим доступу: https://doi.org/10.3389/fmats.2019.00110
16. Bishara D. A state-of-the-art review on machine learning-based multiscale modeling, simulation, homogenization, and design of materials / D. Bishara, Y. Xie, W. K. Liu, S. Li // Archives of Computational Methods in Engineering. – 2022. – Режим доступу: https://doi.org/10.1007/s11831-022-09795-8
17. Greydanus S. Hamiltonian Neural Networks / S. Greydanus, M. Dzamba, J. Yosinski // arXiv. – 2019. – Режим доступу: https://doi.org/10.48550/arXiv.1906.01563
18. Cranmer M. Lagrangian neural networks / M. Cranmer, S. Greydanus, S. Hoyer, P. Battaglia, D. Spergel, S. Ho // arXiv. – 2020. – Режим доступу: https://doi.org/10.48550/arXiv.2003.04630
19. Li Z. Fourier neural operator for parametric partial differential equations / Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, A. Anandkumar // ICLR 2021 Proceedings. – 2021. – Режим доступу: https://doi.org/10.48550/arXiv.2010.08895

20. Lu L. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators / L. Lu, P. Jin, G. Pang, Z. Zhang, G. E. Karniadakis // Nature Machine Intelligence. – 2021. – Vol. 3, № 3. – P. 218–229. – Режим доступу: https://doi.org/10.1038/s42256-021-00302-5
21. Champion K. Data-driven discovery of coordinates and governing equations / K. Champion, B. Lusch, J. N. Kutz, S. L. Brunton // Proceedings of the National Academy of Sciences. – 2019. – Vol. 116, № 45. – P. 22445–22451. – Режим доступу: https://doi.org/10.1073/pnas.1906995116
22. Chen R. T. Q. Neural ordinary differential equations / R. T. Q. Chen, Y. Rubanova, J. Bettencourt, D. Duvenaud // Advances in Neural Information Processing Systems 31. – 2018. – P. 6572–6583. – Режим доступу: https://papers.nips.cc/paper/7892-neural-ordinary-differential-equations
23. Baddoo P. J. Physics-informed dynamic mode decomposition / P. J. Baddoo, B. Herrmann, B. J. McKeon, J. N. Kutz, S. L. Brunton // Proceedings of the Royal Society A. – 2023. – Vol. 479(2271): 20220576. – Режим доступу: https://doi.org/10.1098/rspa.2022.0576
24. Jin H. Physics-Informed Neural Networks for Quantum Eigenvalue Problems / H. Jin, M. Mattheakis, P. Protopapas // 2022 International Joint Conference on Neural Networks (IJCNN). – 2022. – P. 1–8. – Режим доступу: https://doi.org/10.48550/arXiv.2203.00451
25. Baty H. A hands-on introduction to Physics-Informed Neural Networks for solving partial differential equations with benchmark tests taken from astrophysics and plasma physics / H. Baty // HAL (Preprint). – 2024. – Режим доступу: https://hal.science/hal-04491808v1
26. Markidis S. The old and the new: Can physics-informed deep-learning replace traditional linear solvers? / S. Markidis // Frontiers in Big Data. – 2021. – Vol. 4: 669097. – Режим доступу: https://doi.org/10.3389/fdata.2021.669097
27. Costabal F. S. Δ-PINNs: Physics-informed neural networks on complex geometries / F. S. Costabal, S. Pezzuto, P. Perdikaris // Engineering Applications of Artificial Intelligence. – 2023. – Vol. 118: 107324. – Режим доступу: https://doi.org/10.1016/j.engappai.2023.107324
28. Chen S. A hard-constraint wide-body physics-informed neural network model for solving multiple cases in forward problems for partial differential equations / S. Chen, Z. Liu, W. Zhang, J. Yang // Applied Sciences. – 2024. – Vol. 14, № 1: 189. – Режим доступу: https://doi.org/10.3390/app14010189
29. Krishnapriyan A. Characterizing possible failure modes in physics-informed neural networks / A. Krishnapriyan, A. Gholami, S. Zhe, R. Kirby, M. W. Mahoney // Advances in Neural Information Processing Systems 34 (NeurIPS 2021). – 2021. – Режим доступу: https://proceedings.neurips.cc/paper/2021/file/df438e5206f31600e6ae4af72f2725f1-Paper.pdf
30. Jagtap A. D. Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems / A. D. Jagtap, E. Kharazmi, G. E. Karniadakis // Computer Methods in Applied Mechanics and Engineering. – 2020. – Vol. 365: 113028. – Режим доступу: https://doi.org/10.1016/j.cma.2020.113028
31. Rohrhofer F. M. Data vs. Physics: The Apparent Pareto Front of Physics-Informed Neural Networks / F. M. Rohrhofer, S. Posch, C. Gößnitzer, B. C. Geiger // IEEE Access. – 2023. – Vol. 11. – P. 86252–86261. – Режим доступу: https://doi.org/10.1109/ACCESS.2023.3302892

32. Kovachki N. Neural Operator: Learning maps between function spaces / N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, A. Anandkumar // Journal of Machine Learning Research. – 2023. – Vol. 24. – P. 4061–4157. – Режим доступу: https://doi.org/10.5555/3648699.3648788
33. Pang G. fPINNs: fractional physics-informed neural networks / G. Pang, L. Lu, G. E. Karniadakis // SIAM Journal on Scientific Computing. – 2019. – Vol. 41, № 4. – P. A2603–A2626. – Режим доступу: https://doi.org/10.1137/18M1229845
34. He J. Material-Response-Informed DeepONet and its application to polycrystal stress–strain prediction in crystal plasticity / J. He, D. Pal, A. Najafi, D. Abueidda, S. Koric, I. Jasiuk // JOM. – 2024. – Vol. 76, № 10. – P. 5744–5754. – Режим доступу: https://doi.org/10.1007/s11837-024-06681-5
35. Abbasi J. Physical activation functions (PAFs): An approach for more efficient induction of physics into physics-informed neural networks (PINNs) / J. Abbasi, P. Ø. Andersen // Neurocomputing. – 2024. – Vol. 608: 128352. – Режим доступу: https://doi.org/10.1016/j.neucom.2024.128352
36. Rehman D. Physics-constrained neural differential equations for learning multi-ionic transport / D. Rehman, J. H. Lienhard // ICLR 2023 Workshop on Physics for Machine Learning. – 2023. – Режим доступу: https://doi.org/10.48550/arXiv.2303.04594
37. Cai S. Physics-informed neural networks (PINNs) for heat transfer problems / S. Cai, Z. Wang, S. Wang, P. Perdikaris, G. E. Karniadakis // ASME Journal of Heat Transfer. – 2021. – Vol. 143, № 6. – Режим доступу: https://doi.org/10.1115/1.4050542
38. Hou J. Physics-informed neural networks for solving time-dependent mode-resolved phonon Boltzmann transport equation / J. Hou, R. Li, T. Luo // npj Computational Materials. – 2023. – Vol. 9: 212. – Режим доступу: https://doi.org/10.1038/s41524-023-01165-7
39. Faroughi S. A. Physics-Informed Neural Networks with Periodic Activation Functions for Solute Transport in Heterogeneous Porous Media / S. A. Faroughi, R. Soltanmohammadi, P. Datta, S. K. Mahjour, S. Faroughi // Mathematics. – 2024. – Vol. 12: 63. – Режим доступу: https://doi.org/10.3390/math12010063
40. Raissi M. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations / M. Raissi, P. Perdikaris, G. E. Karniadakis // Journal of Computational Physics. – 2019. – Vol. 378. – P. 686–707. – Режим доступу: https://doi.org/10.1016/j.jcp.2018.10.045
41. Dave A. J. Physics-informed state-space neural networks for transport phenomena / A. J. Dave, R. B. Vilim // Engineering Applications of Artificial Intelligence. – 2024. – Vol. 133 (Part C): 108245. – Режим доступу: https://doi.org/10.1016/j.engappai.2024.108245
42. Xiao T. RelaxNet: A structure-preserving neural network to approximate the Boltzmann collision operator / T. Xiao, M. Frank // Journal of Computational Physics. – 2023. – Vol. 490: 112317. – Режим доступу: https://doi.org/10.1016/j.jcp.2023.112317
43. Kaheman K. Learning discrepancy models from experimental data / K. Kaheman, E. Kaiser, B. Strom, J. N. Kutz, S. L. Brunton // arXiv. – 2019. – Режим доступу: https://doi.org/10.48550/arXiv.1909.08574
44. Fasel U. Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control / U. Fasel, J. N. Kutz, B. W. Brunton, S. L. Brunton // Proceedings of the Royal Society A. – 2022. – Vol. 478: 20210904. – Режим доступу: https://doi.org/10.1098/rspa.2021.0904
45. Xu C. Transfer learning based physics-informed neural networks for solving inverse problems in engineering structures under different loading scenarios / C. Xu, B. T. Cao, Y. Yuan, G. Meschke // Computer Methods in Applied Mechanics and Engineering. – 2023. – Vol. 405: 115852. – Режим доступу: https://doi.org/10.1016/j.cma.2022.115852

46. Li Y. Thermal fluid fields reconstruction for nanofluids convection based on physics-informed deep learning / Y. Li, T. Liu, Y. Xie // Scientific Reports. – 2022. – Vol. 12: 12567. – Режим доступу: https://doi.org/10.1038/s41598-022-16463-1
47. Yang H. Deep learning-based molecular dynamics simulation reconfiguration of efficient heat energy transport of Gra/h-BN heterointerface / H. Yang, L. Li, P. Yang // Applied Physics Letters. – 2025. – Vol. 127, № 2: 021601. – Режим доступу: https://doi.org/10.1063/5.0270700
48. Gomes D. S. Computational characterization of the recently synthesized pristine and porous 12-atom-wide armchair graphene nanoribbon / D. S. Gomes, I. M. Félix, W. F. Radel, A. C. Dias, L. A. Ribeiro Jr., M. L. Pereira Jr. // Nano Letters. – 2025. – Vol. 25, № 21. – P. 8596–8603. – Режим доступу: https://doi.org/10.1021/acs.nanolett.5c01319
49. Li J. PiNN: Equivariant neural network suite for modeling electrochemical systems / J. Li, L. Knijff, Z.-Y. Zhang, L. Andersson, C. Zhang // Journal of Chemical Theory and Computation. – 2025. – Vol. 21, № 3. – P. 1382–1395. – Режим доступу: https://doi.org/10.1021/acs.jctc.4c01570
50. Zeng L. Measuring phonon mean free path distributions by probing quasiballistic phonon transport in grating nanostructures / L. Zeng, K. C. Collins, Y. Hu, M. N. Luckyanova, A. A. Maznev, S. Huberman, V. Chiloyan, J. Zhou, X. Huang, K. A. Nelson, G. Chen // Scientific Reports. – 2015. – Vol. 5: 17131. – Режим доступу: https://doi.org/10.1038/srep17131
51. Wu Y.-J. Predicting interfacial thermal resistance by machine learning / Y.-J. Wu, L. Fang, Y. Xu // npj Computational Materials. – 2019. – Vol. 5: 56. – Режим доступу: https://doi.org/10.1038/s41524-019-0193-0
52. Wong J. C. Evolutionary optimization of physics-informed neural networks: Evo-PINN frontiers and opportunities / J. C. Wong, A. Gupta, C. C. Ooi, P.-H. Chiu, J. Liu, Y.-S. Ong // arXiv. – 2025. – Режим доступу: https://doi.org/10.48550/arXiv.2501.06572
53. Lin Q. Monte Carlo physics-informed neural networks for multiscale heat conduction via phonon Boltzmann transport equation / Q. Lin, C. Zhang, X. Meng, Z. Guo // Journal of Computational Physics. – 2025. – Vol. 542: 114364. – Режим доступу: https://doi.org/10.1016/j.jcp.2025.114364
54. Ricci F. An ab initio electronic transport database for inorganic materials / F. Ricci, W. Chen, U. Aydemir, та ін. // Scientific Data. – 2017. – Vol. 4: 170085. – Режим доступу: https://doi.org/10.1038/sdata.2017.85
55. Wu Y.-J. Physical and chemical descriptors for predicting interfacial thermal resistance / Y.-J. Wu, T. Zhan, Z. Hou, L. Fang, Y. Xu // Scientific Data. – 2020. – Vol. 7: 36. – Режим доступу: https://doi.org/10.1038/s41597-020-0373-2
56. Hargreaves C. J. A database of experimentally measured lithium solid electrolyte conductivities evaluated with machine learning / C. J. Hargreaves, M. W. Gaultois, L. M. Daniels, та ін. // npj Computational Materials. – 2023. – Vol. 9: 9. – Режим доступу: https://doi.org/10.1038/s41524-022-00951-z
57. Petretto G. High-throughput density-functional perturbation theory phonons for inorganic materials / G. Petretto, S. Dwaraknath, H. P. C. Miranda, D. Winston, M. Giantomassi, M. J. van Setten, X. Gonze, K. A. Persson, G. Hautier, G.-M. Rignanese // Scientific Data. – 2018. – Vol. 5: 180065. – Режим доступу: https://doi.org/10.1038/sdata.2018.65

М. В. ГЕРКАЛЮК
аспірант, Фізико-технічний інститут, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

m.herkaliuk@gmail.com

А. В. ГІЛЬЧУК
кандидат фіз.-мат. наук, доцент, доцент кафедри прикладної фізики, Фізико-технічний інститут, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна
a.v.gilchuk@gmail.com

ФІЗИЧНО-ІНФОРМОВАНЕ МАШИННЕ НАВЧАННЯ ДЛЯ МОДЕЛЮВАННЯ ЯВИЩ ПЕРЕНОСУ НАНОРОЗМІРНИХ СИСТЕМ:ВИКЛИКИБ ПІДХОДИ ТА ПЕРСПЕКТИВИ

DOI: 10.31651/2076-5851-2025-11-57 PACS: 05.20.Dd;66.70.-f; 81.07.-b; 02.60.Cb;07.05.Mh

Моделювання явищ переносу в нанорозмірних системах є критично складним завданням, в якому класичні рівняння континууму виявляються неефективними, а високоточні обчислювальні системи є надто дорогими. Фізично-інформоване машинне навчання (PIML) стало революційним підходом до вирішення цієї дилеми шляхом синергетичного поєднання розрізнених експериментальних даних з основними законами транспортних моделей першого принципу. Цей огляд надає вичерпну інформацію про те, як PIML — особливо фізично-інформовані нейронні мережі (PINN), методи навчання операторів та підхід пожднання моделей різної точності — прискорює аналіз нанорозмірного транспорту від фононного транспорту на основі BTE до балістично-дифузійного теплоперенесення та ефектів випромінювання в ближньому полі. Ми розглядаємо постійні проблеми з даними в дослідженнях наноматеріалів, включаючи шумні вимірювання та формулювання високорозмірних диференціальних рівнянь з частинними похідними (PDE), і представляємо передові стратегії, такі як декомпозиція доменів та гібридні механістичні методи машинного навчання (ML), щоб підвищити гнучкість та масштабованість цих нових підходів. Нарешті, ми окреслюємо поточні прогалини в цій галузі, від кількісної оцінки невизначеності до розробки цифрових двійників у реальному часі, та окреслюємо майбутні напрямки досліджень, спрямовані на об'єднання квантових симуляцій, експериментальної метрології та глибокого навчання. Вбудовуючи фізичні обмеження безпосередньо в робочий процес навчання, ці фізично обґрунтовані методи пропонують трансформаційний шлях для оптимізації нанорозмірного транспорту, та сприятиме вдосконаленню методів дослідження наноматеріалів. Додатково ми формуємо практичну «дорожню карту» інтеграції PIML з високопродуктивними та диференційовними солверами (BTE/MC/FEM) для швидких параметричних досліджень і калібрування інтерфейсної теплопровідності. Запропоновано базові метрики, протоколи валідації та бенчмарки (TDTR, SThM, грейтингові структури) для відтворюваного порівняння PINN/операторних моделей із першим-принципним еталоном. Окрему увагу приділено стратегіям UQ (баєсівські

PINN, ансамблі, багатовірогідні втрати), що є ключем до надійного впровадження цифрових двійників у реальному часі в нанотеплотехніці та наноелектроніці.

Ключові слова: явища переносу нанорозмірних систем, наноматеріали, машинне навчання, фізико-інформовані нейронні мережі, глибоке навчання.

Одержано редакцією 22.09.2025
Прийнято до друку 02.10.2025