
Position: Significant impact of numerical precision in scientific machine learning

Anonymous Authors¹

Abstract

The machine learning community has focused on computational efficiency, often leveraging reduced-precision formats such as those below the standard FP32. In contrast, little attention has been given to higher-precision formats, such as FP64, despite their critical role in scientific domains like materials science, where even small numerical differences can lead to significant inaccuracies in physicochemical properties. This need for high precision extends to the emerging field of *machine learning for scientific tasks*, yet it has not been thoroughly investigated. According to several studies and our toy experiment, models trained with FP32 exhibit insufficient accuracy compared to FP64, indicating that higher precision is also crucial in scientific machine learning, as it is in traditional scientific computing. Despite the potential of scientific machine learning, this precision issue often limits its adoption as replacements for traditional scientific computing in practical research. This position paper not only highlights these precision-related issues but also recommends reporting comparisons between FP32 and FP64 results while encouraging the release of FP64 models. We believe these efforts can enable machine learning to contribute meaningfully to the natural sciences, ensuring both scientific reliability and practical applicability.

1. Introduction

The rapid advancements in natural language processing (NLP) and computer vision (CV) in the machine learning (ML) field have accelerated the broad application across various domains (Litjens et al., 2017; Ozbayoglu et al., 2020; Ren et al., 2021; Lai et al., 2024; Raghu & Schmidt, 2020). Specifically, **ML for scientific tasks**—which has begun to re-

solve intellectually demanding problems in scientific fields—has been highlighted across disciplines, opening new possibilities for scientific breakthroughs. In recognition of these breakthroughs, the 2024 Nobel Prize in Chemistry honored the contributions of scientific ML, highlighting innovations such as AlphaFold and RoseTTAFold (Jumper et al., 2021; Baek et al., 2021; The Royal Swedish Academy of Sciences, 2024). These models transformed research by rapidly delivering results that once required significant resources and time-consuming experiments or simulations. Building on these successes, scientific ML not only addresses traditional labor-intensive workflows but also finds hidden patterns within complex data, thereby providing human researchers with direct insights into novel discoveries across natural sciences (Webb et al., 2018; Morgan & Jacobs, 2020; Karagiorgi et al., 2022; Yoo et al., 2024).

In the context of methodology, the development of scientific ML naturally follows the broader trends and paradigms of the ML research field. In the early stages of NLP and CV, most work focused on discriminative tasks (*e.g.*, named entity recognition and image classification) (Walker et al., 2006; Deng et al., 2009) before gradually shifting to generative tasks (*e.g.*, machine translation and text-to-image generation) (Bojar et al., 2014; Schuhmann et al., 2022). Further, generative approaches themselves have advanced sequentially, moving from variational autoencoders (VAEs) to generative adversarial networks (GANs), and more recently, to diffusion models (Kingma & Welling, 2014; Goodfellow et al., 2014; Ho et al., 2020; Song & Ermon, 2019). In a similar manner, numerous scientific domains have rapidly adopted the latest advances from the ML community. For example, among various areas of bioinformatics, research on DNA sequence data initially leveraged discriminative models such as DeepVariant (Poplin et al., 2018) and DeepSEA (Zhou & Troyanskaya, 2015), and over time, this expanded to generative models including ExpressionGAN (Zrimec et al., 2022) and Evo (Nguyen et al., 2024). Similarly, material structure prediction in the field of materials and drug discovery has followed this trend from VAEs (Sanchez-Lengeling & Aspuru-Guzik, 2018; Gómez-Bombarelli et al., 2018; Lim et al., 2018) and GANs (Prykhodko et al., 2019; Kim et al., 2020; Abasi et al., 2022) to diffusion models (Hooeboom et al., 2022; Peng et al., 2023; Zeni et al., 2025).

¹Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

In parallel with these advances, the most recent paradigm in ML research, often referred to as the *scaling law*, focuses on improving performance by progressively increasing the size of models, training datasets, and computational resources (Kaplan et al., 2020; Snell et al., 2024). Building upon this idea of continuously expanding scale, researchers have successfully validated the approach across diverse fields, including NLP, CV, reinforcement learning, and time-series forecasting (Zhai et al., 2022; Cherti et al., 2023; Hilton et al., 2023; Neumann & Gros, 2023; Shi et al., 2024). Accordingly, the scientific ML domain will also adopt this paradigm, and in fact, large models designed to address scientific tasks have already begun to emerge (Nguyen et al., 2024; Zhang et al., 2024).

As these models grow larger and more complex, they inevitably require massive computational power, which results in a significant challenge for both training and inference. To address this, the lower numerical precision, or quantization is a widely employed strategy, which helps reduce the computational expense (Zhu et al., 2024; Micikevicius et al., 2018). These approaches inevitably involve a trade-off between accuracy and computational budgets, resulting in an unwanted loss of accuracy. To minimize such losses, techniques such as mixed precision training (Micikevicius et al., 2018) or more advanced quantization methods (Banner et al., 2019; Dettmers et al., 2022; Liu et al., 2023; Xu et al., 2024) have been proposed, which allow researchers to conserve the original accuracy while achieving the advantages of reduced computational costs. Consequently, the ML community has accepted slight accuracy degradation as a natural trade-off for greater efficiency, thereby integrating these lower-precision techniques into real-world applications to balance computing cost and performance.

However, the tolerance for lower-precision techniques is highly problematic in the field of scientific computing. Scientific computing primarily aims to solve fundamental physics equations that are difficult to solve manually by simplifying or discretizing the inherently continuous and infinite real-world phenomena to make them computationally tractable. As a consequence, even tiny differences in numerical precision can lead to significant issues regarding the reliability of computational results. Specifically, our experimental results show that *single precision’s sensitivity to numerical deviations can substantially influence the accuracy of fundamental physical equations*. As a result of this high sensitivity, small precision differences can cause significant changes in physicochemical properties, such as absorption coefficient, defect energies, or reaction pathway predictions, thereby reducing the reliability of results, especially when accurate predictions are crucial for critical decisions. One critical aspect is that these challenges related to numerical precision are not confined to traditional computational science, as ML models are increasingly being

utilized in various studies to replace prevalent simulations. In other words, traditional computational science requires high precision, making it essential to verify whether FP32 produces valid results before using ML models, as numerical precision is key to maintaining reliability.

In this position paper, we argue for the significant role of numerical precision in scientific ML research, emphasizing the need for evaluating and analyzing its impact on results derived from varying precision levels. To this end, we first present real-world examples from previous computational simulations where numerical precision had a notable impact on their results. By introducing cases that reflect real-world scenarios from actual research fields, we aim to demonstrate the practical existence of numerical precision issues in scientific simulations. Subsequently, we explain that the importance of numerical precision is not confined to traditional scientific computing alone but is also deeply related to ML applications in scientific domains. Specifically, we provide examples involving ML potential models and physics-informed neural networks (PINNs), which are actively studied in both ML and science domains, demonstrating the critical role of numerical precision in these areas (Raissi et al., 2019; Kocer et al., 2022; Käser et al., 2023). Furthermore, we address the growing use of large language models (LLMs) in scientific ML and their implications for precision-related challenges.

In conclusion, we propose actionable recommendations for the ML community and potential research directions based on our earlier discussions. We then present alternative viewpoints, offer responses, and conclude. Since the main role of ML in scientific research is to deepen understanding in traditional domains, the issues we raise must be rigorously examined. When relatively simple actions by ML researchers can remove barriers that hinder natural scientists from applying ML models, these measures become essential, not optional. As scientific machine learning is still in its early stages, we hope that thorough debate will help minimize trial-and-error in future research.

2. Importance of numerical precision in scientific computing

In scientific computing, the main goal is solving complex physics equations through computational power, especially when manual solutions are impractical or non-existent. Many-body problems including multiple object interactions demonstrate the necessity of high-performance computing solutions. Various computational methods have emerged to solve fundamental physics equations: molecular dynamics for Newton’s Second Law, density functional theory (DFT) (Jones & Gunnarsson, 1989) for the Schrödinger equation, and the finite-difference time-domain (FDTD) (Yee, 1966) method for Maxwell’s equations.

Modern digital computers use discrete bit-based representations, creating an inherent challenge. The floating-point system limits direct solutions to physical equations that operate in continuous systems, including $F = ma$, Schrödinger equation, and Maxwell’s equations. Researchers have developed workarounds using approximated equations (e.g. the Kohn-Sham equation (Kohn & Sham, 1965)) or discretization methods with specific time units and grid systems (e.g. molecular dynamics). These solutions require high numerical precision to accurately represent physical phenomena, typically settling on double precision as a balance between computational cost and accuracy. For instance, Quantum ESPRESSO (Giannozzi et al., 2009), a leading open-source DFT implementation, strictly enforces double precision throughout its code.

To demonstrate the precision’s crucial role in scientific computing, we present examples showing how small numerical variations can significantly impact computational results, analyzing these effects in realistic research scenarios. Specifically, we illustrate the influence on realistic research scenarios, thereby analyzing the implications and identifying the precise numerical accuracy-related challenges.

2.1. Impact on density functional theory simulation

Quantum mechanics, beginning with Planck’s quantum hypothesis (Planck, 1900), revolutionized our understanding of microscopic phenomena. While exact calculations are only possible for simple systems like the hydrogen atom, the Kohn-Sham equation introduced density functional theory (DFT) as an efficient approach for many-body electron problems. Using Python-based Simulations of Chemistry Framework (PySCF), we performed geometry optimization calculations for water (H_2O) using both Hartree-Fock (HF) and DFT calculations with B3LYP functional and 6-311++G(d,p) basis set (Andersson & Uvdal, 2005; Tirado-Rives & Jorgensen, 2008; Yanai et al., 2004).

Figure 1 shows the results of geometry-optimized water molecules obtained from HF and DFT calculations under different numerical precision conditions: FP32 and FP64. When utilizing FP64, both HF and DFT calculations successfully converged within three optimization steps with satisfying the convergence criteria. Since DFT explicitly accounts for electron correlation effects (Becke, 1988), it is generally expected to provide more accurate results than HF, a trend that is also reflected in our findings. Comparing bond lengths, the reference (Bowen & Sutton, 1958) O-H bond length is 0.957 Å, while HF exhibits a deviation of 0.016 Å (1.7 % error), and DFT yields a smaller deviation of 0.005 Å (0.5 % error). Similarly, for the bond angle, HF deviates by 1.7 °(0.7 % error) from the reference value of 104.52 °, whereas DFT shows a smaller deviation of 0.55 °(0.5 % error). However, when using FP32, significant numerical

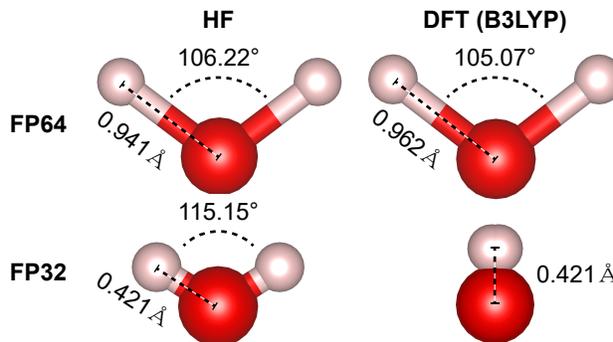


Figure 1. Water molecule geometry optimizations comparing FP64 (top) and FP32 (bottom) using HF (left) and DFT (right) methods. FP64 yields physically valid structures while FP32 produces unrealistic geometries.

Table 1. Atomic coordinates and total energy comparison of geometry-optimized H_2O at FP32 and FP64 precision using 6-311++G(d,p) basis set. (*) FP32 failed to converge in both HF and DFT methods, while FP64 showed consistent results.

	HF		DFT (B3LYP)		
	6-311++G(d,p)		6-311++G(d,p)		
	FP32	FP64	FP32	FP64	
Atomic coordinates (Å)	O_x	-0.000356*	0.000000	0.009524*	0.000000
	O_y	0.246311*	0.014028	0.578655*	0.000780
	O_z	0.000000*	0.000000	0.000000*	0.000000
	H_{1x}	0.453099*	0.752792	0.026584*	0.763642
	H_{1y}	0.534244*	0.578999	0.998814*	0.585902
	H_{1z}	0.000000*	0.000000	0.000000*	0.000000
	H_{2x}	-0.453725*	-0.752792	-0.024889*	-0.763642
	H_{2y}	0.534404*	0.578999	0.998054*	0.585902
H_{2z}	0.000000*	0.000000	0.000000*	0.000000	
Total energy (Ha)	-74.938*	-76.053	N/A*	-76.458	

*Not Converged

instabilities arise, preventing the convergence of optimization steps. In the case of HF calculations, the gradient of hydrogen atoms stagnates between 0.2–0.4 Ha/Bohr, which is significantly above the desired convergence threshold of 10^{-6} Ha/Bohr. For DFT calculations, the issue becomes even more pronounced, as the gradient values rapidly diverge beyond 10^5 Ha/Bohr, resulting in termination before reaching the maximum step. As a result, when using FP32, the HF calculation exhibits a substantial 50 % error, while the DFT calculation produces a molecular structure impossible to exist in reality, as illustrated in Figure 1.

A detailed examination of the atomic coordinates in Table 1 further highlights the differences. While the coordinates obtained from FP64 differ only by approximately 0.01 Å for oxygen and hydrogen atoms, FP32 results display considerable deviation. Notably, the FP32-calculated atomic positions deviate by up to 0.4 Å from those obtained using FP64, a significant difference considering that the O-H bond

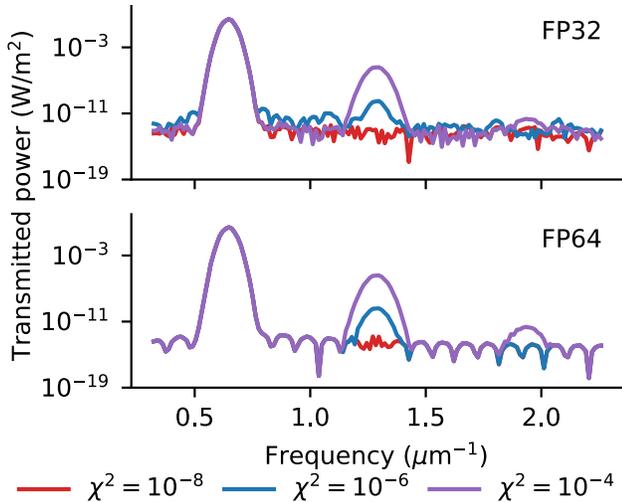


Figure 2. Transmittance spectra comparison between FP32 (top) and FP64 (bottom) in Kerr media, showing FP32’s failure to accurately model higher harmonics and low-power wave patterns below 10^{-10} W/m².

length itself is only 0.957 Å. In addition, the total energy difference between FP32 and FP64 calculations is approximately 1.1 Hartree (equivalent to 29.93 eV), which exceeds the formation energy of water (2.9 eV) by more than an order of magnitude. This clearly indicates that the FP32 result corresponds to a structure that is impossible to exist in reality. These results demonstrate that FP32 lacks the numerical precision necessary to achieve sufficient convergence tolerance in scientific computations. The failure of a simple molecular system such as water to reach an optimized structure under FP32 precision indicates its fundamental limitations in scientific calculations.

2.2. Impact on finite difference time domain simulation

Electromagnetism, established by Maxwell’s equations (Maxwell, 2010; 1865), provides the theoretical foundation for understanding electromagnetic waves. However, solving Maxwell’s equations for complex phenomena is computationally challenging. To address this, FDTD discretizes Maxwell’s equations in time and space. Using Meep (Oskooi et al., 2010), which an open-source FDTD software, we investigated numerical precision effects on electromagnetic simulations, comparing FP32 and FP64 in nonlinear Kerr media simulations. We simulated a Kerr medium (refractive index=1.65) excited by an electromagnetic wave source ($\lambda=1.55$ μm , $\Delta\lambda=0.15$ μm).

Figure 2 presents the transmission spectrum of the nonlinear Kerr medium under FP32 and FP64 precision settings. From left to right, the spectral peaks correspond to the fundamental generation induced by the source, the second

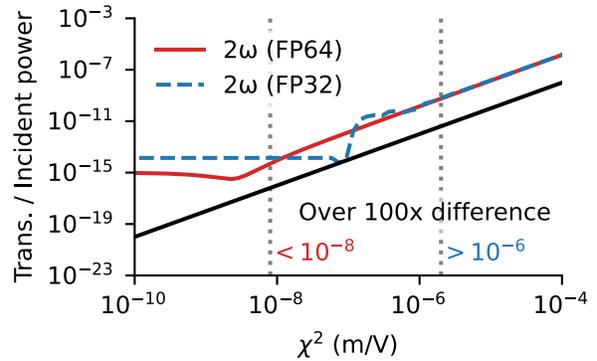


Figure 3. Computed second harmonic susceptibility shown in FP64 (red) and FP32 lines (blue dashed) compared to theoretical quadratic behavior (black). FP64 maintains accuracy to 10^{-8} m/V, while FP32 deviates above 10^{-6} m/V, rendering it unsuitable for typical nonlinear materials.

harmonic generation (SHG), and the third harmonic generation (THG). While the fundamental peak exhibits minimal differences between FP32 and FP64, notable discrepancies arise in the SHG and THG regions. Specifically, FP32 calculations display pronounced background signal instability and intensity variations in harmonic generation, which result from imprecise numerical computation. A particularly notable difference appears in the behavior of the background signal. In FP64 calculations, the background follows a well-defined periodic pattern governed by the electromagnetic wave, whereas in FP32, the background signal appears as unstructured Gaussian-like noise. This phenomenon indicates that the lack of numerical precision in FP32 significantly disrupts the accurate computation of low-intensity transmitted power, particularly for electromagnetic waves in the range of 10^{-11} W/m². These findings highlight the fundamental limitations of single precision (FP32) in reliably capturing weak electromagnetic signals and nonlinear optical effects.

To further analyze the impact of numerical precision, we examined the relationship between second-order nonlinear susceptibility (χ^2) and the transmittance-to-incident power ratio. As shown in Figure 3, the black upward-sloping line represents a quadratic line, serving as a reference line indicating the expected computational trend of transmittance over incident power ratio as nonlinear susceptibility varies. Ideally, the computationally simulated values should align with this reference trend, maintaining the same slope. Comparing the results obtained from FP64 (red solid line) and FP32 (blue dashed line), we observe that as nonlinear susceptibility decreases beyond a certain threshold, the ratio begins to saturate. This saturation point effectively defines the lower bound of computational precision achievable under each numerical setting.

Specifically, for values of χ^2 above 10^{-6} , both FP64 and FP32 provide reliable computational precision. However, for values below this threshold, FP32 results begin to exhibit saturation, rendering further calculations meaningless due to the loss of numerical resolution. In contrast, FP64 maintains simulation accuracy down to approximately 10^{-8} , demonstrating a computational precision that is at least two orders of magnitude higher than that of FP32. This result implies that for most nonlinear materials with χ^2 values below 10^{-6} , transmittance spectrum simulations using FP32 become inherently unreliable. These findings highlight the critical role of numerical precision in computational science, particularly in fields where small numerical deviations can lead to substantial errors. As demonstrated in both DFT and FDTD simulations, the limitations of single precision introduce significant inaccuracies, especially in cases involving highly sensitive physical properties. This also highlights the necessity of carefully selecting numerical precision levels when conducting computational simulations—particularly in scientific ML applications—where maintaining the reliability of results is essential.

3. Numerical precision issue in scientific ML

As demonstrated in the previous section, numerical precision can significantly affect the outcomes of traditional scientific simulations and potentially influence the results of scientific research. This naturally leads to an important question: **Do ML models designed for scientific tasks also suffer from similar precision-related issues?** To investigate this question, we survey various studies that apply ML to scientific research, searching for cases where the precision issue has been reported. We also conduct simple toy experiments to further assess the impact of numerical precision in ML-based scientific tasks. Through these analyses, we seek to determine whether the precision issue is a *significant challenge* or *just a theoretical concern*.

3.1. Impact on machine learning potential

The first example we present is an ML potential¹ (Kocer et al., 2022; Käser et al., 2023), which is closely related to Section 2.1. Fundamentally, ML potential models aim to compute potential energy and the associated forces for a given material structure, offering a much faster alternative to traditional quantum mechanical calculations. Due to their wide range of applications, ML potentials have been extensively studied not only in physics, materials science, chemistry, and biology but also within the ML community (Behler & Parrinello, 2007; Pukrittayakamee et al., 2009; Smith et al., 2017; Gilmer et al., 2017; Schütt et al., 2017). In addition, closely related topics, such as property

¹In other domains, the term *machine learning interatomic potential* (MLIP) is also used.

prediction and generation for material or drug discovery have also been actively explored, making ML potentials a familiar subject for ML researchers. In this position paper, we specifically focus on ML potentials based on neural networks, *i.e.*, neural network potentials. Since ML research often treats energy and force values in the same manner as other material properties, our discussion extends naturally to broader property prediction tasks.

A key challenge in ML potential studies lies in effectively representing and processing atomic information in three-dimensional space while ensuring rotational and translational equivariance or invariance. To tackle this, the field has evolved from vanilla graph neural networks (Scarselli et al., 2009) and transformers (Vaswani et al., 2017) to more specialized architectures that satisfy these constraints, achieving higher prediction accuracy (Schütt et al., 2017; Gasteiger et al., 2021; Satorras et al., 2021; Batzner et al., 2022; Batatia et al., 2022; Fuchs et al., 2020; Thölke & Fabritius, 2022; Liao & Smidt, 2023). As a result, many recent models are now integrated into widely used libraries or simulation software, such as the Atomic Simulation Environment (ASE) (Bahn & Jacobsen, 2002; Larsen et al., 2017) and LAMMPS (Plimpton, 1995; Thompson et al., 2022). This demonstrates that ML potential models are increasingly employed in practical research; thus, any numerical precision issues arising in these models could have significant implications for scientific discoveries.

Consequently, we aimed to investigate whether existing ML potential models suffer from precision issues. To this end, we surveyed the pretrained checkpoints of various ML potential models available in the ASE library to determine whether they support FP64 precision. Interestingly, among several models in ASE, only MACE (Batatia et al., 2022) provides pretrained checkpoints trained in FP64, while other models appear not to have considered FP64 training. Even before detailed analysis, this observation suggests that the ML potential community may not be fully aware of the potential significance of numerical precision.

To preliminarily understand the effect of precision, we conducted a toy experiment using MACE, the only model that provides FP64-trained parameters. We selected an ethanol molecule as a manageable small organic system containing multiple atom types (C, H, O), then moved one of its carbon atoms (specifically the one closest to oxygen) along a certain path (shown in the top of Figure 4) and observed changes in the potential energy and forces. We compared the results obtained in FP32 with those in FP64 by applying built-in type conversion in the MACE code to the FP64-trained checkpoint. The upper plot in Figure 4 illustrates the overall trends in energy and force, while the lower plot shows the direct differences between FP32 and FP64. Our results show that the differences remain within approximately 1

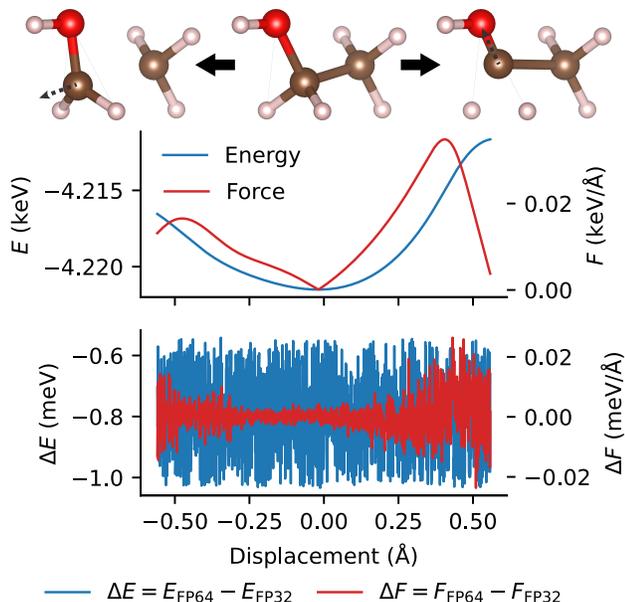


Figure 4. MACE model calculations showing energy (red) and force (blue) changes during carbon atom displacement in ethanol. FP32 versus FP64 precision reveals minimal deviations (1 meV energy, 0.02 meV/Å force), though broader testing needed.

meV for energy and 0.02 meV/Å for force, which is often considered acceptable in small-molecule simulations.

However, this experiment was deliberately simplistic, and the results should not be overinterpreted as evidence that FP32 models are generally reliable. Larger or more complex systems, such as polymers or proteins, could exhibit more substantial numerical errors. Additionally, the FP32 model tested here was originally trained in FP64 and then converted to FP32 for inference; a model trained entirely in FP32 from the outset might behave differently. In fact, Batatia et al. (2025) report that NequIP (Batzner et al., 2022) exhibits different numerical sensitivity when trained in FP32 versus FP64, and Maxson et al. (2024) also discuss similar issues. These observations highlight the importance of carefully assessing numerical precision in ML potential models and the need for systematic benchmarks regarding precision.

3.2. Impact on physics-informed neural network

Beyond the fundamental equations mentioned in the previous section, various subfields of natural science describe natural phenomena using differential equations. For example, in fluid dynamics, including weather prediction, Navier-Stokes, continuity, and heat transfer equations are commonly used (Tritton, 2012; Bauer et al., 2015). Moreover, differential equations such as the Black-Scholes equation (Black & Scholes, 1973) are also employed in fields beyond natural sciences, such as financial engineering. Many of these equations either lack general analytical solutions

or are too complex to be solved manually. As a result, numerical methods have been developed over time, leading to techniques such as the Euler method, Runge-Kutta methods, and Picard method (Butcher, 2016; Strauss, 2007). These techniques have also influenced modern approaches in ML, including diffusion models, NeuralODEs, and deep equilibrium models (Ho et al., 2020; Song & Ermon, 2019; Chen et al., 2018; Bai et al., 2019).

The concept of the PINNs (Raissi et al., 2019) leverages automatic differentiation (autograd), fundamental to back-propagation, to solve differential equations using neural networks. Due to its simple yet powerful approach, PINNs have been widely adopted in scientific domains that rely on numerical methods. This section explores whether numerical precision issues also arise in PINNs and investigates related challenges through a literature survey.

First, Nakamura et al. (2022) explicitly discussed the impact of numerical precision in scientific research, reporting that training PINNs with FP32 failed, whereas FP64 did not: *from a comprehensive standpoint, FP32 computation has a risk of failure for the present problem compared with FP64.* This work applies PINNs to a specific fluid dynamics problem involving surface tension modeling, which requires up to fourth-order derivatives, making it a specialized case of differential equations. Although this is a specific scenario, it is a real-world scientific study, demonstrating that precision issues can significantly impact the practical use of PINNs.

Meanwhile, Sharma & Shankar (2022) were well aware of precision issues and leveraged this understanding to improve the methodology of PINNs. The key idea of their work is to replace certain autograd operations in PINNs with a specialized FD method, reducing the computational cost associated with autograd. Here, to compensate for the loss of accuracy introduced by finite difference approximations, the authors proposed using high-precision (FP64) training. As a result, the reduction in computational cost from bypassing autograd exceeds the overhead introduced by FP64 operations, leading to an overall speedup that makes their approach faster than a vanilla PINN in FP32. Beyond the fields of PINNs and scientific ML, this study introduces a novel perspective on utilizing high-precision models in neural network research.

Thus, in the context of PINNs, a comprehensive study is needed to systematically assess the impact of numerical precision issues on scientific research. Fortunately, many fields share similar types of differential equations, e.g., Laplace equation in electrostatics and fluid dynamics, where it describes electric potential distribution and velocity potential in inviscid flow, respectively. By focusing on the precision challenges of commonly used differential equations and rigorously validating PINNs in this context, such research could have a substantial impact across multiple domains.

3.3. Challenges for large language models

The emergence of LLMs in scientific applications is accelerating, further raising concerns about numerical precision in such domains. To investigate these concerns, we examine both existing studies and empirical evidence that highlight precision-related challenges in LLM applications. The integration of LLMs in scientific domains follows two distinct approaches. The first involves direct inference without architectural modifications, where scientific data is transformed into natural language format for existing LLM architectures (Rubungo et al., 2024; Jacobs et al., 2024; Liu et al., 2024). The second approach develops specialized architectures that combine domain-specific encoders with fine-tuned language models, preserving the intrinsic properties of scientific data while leveraging LLM capabilities (Li et al., 2024; Park et al., 2024).

Regarding the first approach, unlike conventional scientific models, LLMs generate outputs based on tokens, which may compromise prediction accuracy. Numerous studies have demonstrated that LLMs struggle with symbolic tasks (Wei et al., 2022; Yao et al., 2023), similar to their difficulties in numerical predictions. For instance, these models often fail to accurately count the occurrences of specific characters within words (e.g., counting the letter ‘r’ in ‘strawberry’) or comparing the size of decimal numbers (e.g., determining whether 3.9 is larger than 3.11²). This limitation stems from their fundamental architecture, where words are processed as sequences of tokens rather than as individual alphabetic characters or numbers. Although various studies (Wei et al., 2022; Kojima et al., 2022; Yao et al., 2023; Besta et al., 2024) have been proposed to address these challenges, symbolic manipulation remains a significant obstacle for LLMs. Consequently, their application in scientific tasks requires careful consideration and validation.

Another critical consideration in LLM deployment is the continuous increase in model size. For instance, the open-source Llama series demonstrates this trend clearly: LLaMA (65B parameters) grew to Llama-2 (70B) and further to Llama-3.1 (405B) (Touvron et al., 2023a;b; Llama Team, 2024), and more recently, DeepSeek-v3 has pushed this expansion even further, reaching 671B (DeepSeek-AI, 2024). Such explosive growth in model sizes across LLMs has resulted in a substantial increase in computational costs for both training and inference. To mitigate the budget, researchers commonly employ parameter quantization techniques by reducing model precision to lower-bit formats (Liu et al., 2021; Dettmers et al., 2022; Liu et al., 2023), sometimes even 1-bit representations (Xu et al., 2024).

²In January 2025, GPT-4o incorrectly answers that 3.11 is larger than 3.9, due to tokenization: 3.9 as [‘3’, ‘.’, ‘9’] and 3.11 as [‘3’, ‘.’, ‘11’], leading to a direct comparison of 9 and 11.

However, these approaches directly contradict the high precision requirements of scientific computing, which we have emphasized throughout this discussion. This issue is particularly critical for the second approach, where domain-specific encoders—often derived from scientific ML models—serve as feature extractors. If quantization significantly reduces the precision of the extracted features, the LLM may fail to process them accurately, potentially degrading overall model performance. For example, Li et al. (2024) employed UniMol (Zhou et al., 2023), a model broadly categorized as an ML potential, as an encoder. Even if the encoder provides highly precise features, the LLM’s lower precision representations may obscure this information, ultimately leading to inaccurate final predictions. This inherent trade-off between computational efficiency and numerical precision highlights the importance of carefully designing LLM integration strategies in scientific applications to ensure both accuracy and practicality.

4. Suggestions for Advancing Scientific ML

Building on previous discussions, we outline key directions for the ML community to advance scientific tasks.

Exploring high-precision models and mixed high-precision training Most ML research has primarily explored lower-precision formats such as FP16, BF16, and INT8, whereas comparisons between FP32 and FP64 remain relatively limited. We argue that *researchers should explore the potential benefits of double precision beyond the commonly used FP16 and FP32*. Inspired by mixed-precision training, we propose extending this idea to high-precision training. Similar to conventional mixed-precision training, which employs lower precision for the majority of layers while preserving higher precision for numerically sensitive operations (e.g., batch normalization and softmax), we propose identifying sensitive layers and selectively training them in FP64. This direction is especially relevant from an energy efficiency perspective, as FP64 training inherently consumes more energy than FP32. While scientific ML is often considered advantageous over traditional scientific computing in terms of runtime, its energy consumption remains a critical concern. Investigating novel model architectures and training techniques that preserve high numerical precision while enhancing energy efficiency will be crucial for the widespread adoption of scientific ML.

Benchmarking and reporting FP32 vs. FP64 results Scientific ML typically demands higher numerical precision than general ML applications to ensure computational reliability. While predictive accuracy is the primary focus, other factors such as training and inference time remain significant, as well as energy efficiency. Consequently, researchers should explicitly report the numerical preci-

sion used in their studies, compare FP32- and FP64-trained models where applicable, and publicly release FP64-trained models to enhance reproducibility and facilitate further advancements. To support meaningful evaluations, standardized benchmarks that capture precision sensitivity across diverse scientific tasks are essential. Such benchmarks would provide a consistent framework for quantifying trade-offs between numerical precision, computational efficiency, and reproducibility in scientific ML research.

Collaboration with natural scientists Achieving meaningful progress in scientific ML requires close collaboration with natural scientists. This is not merely a conceptual argument but a practical requirement, as ML researchers often lack the domain-specific intuition to determine the appropriate level of numerical precision for a given scientific task. For instance, research on ML potential is published in both traditional scientific journals and ML conferences, yet the evaluation criteria and priorities differ significantly between these communities (Batatia et al., 2022; Kovács et al., 2023). Strengthening interdisciplinary collaboration will help bridge this gap, ensuring that precision requirements align with both scientific validity and practical usability.

Integrating ML into traditional computational methods Rather than solely focusing on developing high-precision ML models, an alternative approach is to integrate ML into traditional computational methods to achieve both accuracy and efficiency. One promising strategy is to first employ ML models while acknowledging their inherent numerical limitations and using them to generate an approximate solution (Arisaka & Li, 2023; Saverio, 2023; Napier, 2024). These ML-generated approximations can then serve as an initial guess for traditional computational methods, significantly accelerating convergence while preserving precision. This hybrid approach presents a compelling solution for scientific applications where both computational speed and numerical accuracy are critical.

5. Alternative Views

This section presents alternative views that challenge our position and provides responses to address these concerns.

Q1: Is the issue really about numerical precision, or could it be a capacity limitation of the model? An alternative view holds that the observed inaccuracies stem from fundamental limitations in network architecture or training methods, rather than numerical precision constraints. According to this viewpoint, neural networks may not yet be sufficiently expressive to solve the given task, regardless of numerical precision constraints. To distinguish numerical issues from capacity concerns, we can leverage numerical analysis tools, such as condition numbers and numerical

sensitivity analysis, to determine whether errors arise from numerical instability. Since modern neural networks rely heavily on matrix operations, existing research on matrix sensitivity provides a robust foundation for further analysis. These insights can help clarify the relationship between numerical stability and model expressivity.

Q2: If certain scientific computing tasks are not sensitive to numerical precision, does it matter? It is true that not all scientific tasks require high numerical precision because certain tasks can tolerate lower levels of precision. However, our focus should be on fields where high precision is essential, such as quantum chemistry, materials science, and nonlinear physics, where even slight inaccuracies can lead to significant deviations. Currently, there is still limited understanding of which tasks, models, and environments are most affected by numerical precision and what factors contribute to these sensitivities. A systematic analysis is necessary to identify precision-critical cases before making broad assumptions about acceptable precision levels. Until a clear understanding is established, a precision-aware approach should be prioritized, while relaxed conditions can be applied to tasks that do not require high precision.

Moreover, certain scientific tasks may not require explicit consideration of numerical precision. For example, in tasks where logical reasoning is more critical than numerical accuracy, such as those that rely on LLMs, precision constraints may be less significant. These include (1) Explaining or summarizing experimental results or literature (Xie et al., 2024), (2) generating new hypotheses for scientific research (Kumbhar et al., 2025; Lu et al., 2024), (3) providing guidance for tasks where the methodology is not clearly defined (*e.g.*, retrosynthesis), and (4) assisting scientific educations (Bewersdorff et al., 2025). In such cases, the role of ML extends beyond numerical fidelity, focusing instead on knowledge synthesis and interpretability.

6. Conclusions

Scientific ML has become a major field in modern ML research, with the goal of developing models that contribute to scientific discovery. This position paper highlights the impact of precision issues, which can affect the practical usability of scientific ML models but have been largely overlooked. The precision issues in scientific ML are closely tied to ethical concerns regarding the reliability and explainability of scientific findings. In summary, our contribution lies in a practical step toward making scientific ML models more reliable, reducing the risk of misleading scientific insights due to numerical inaccuracies. If our simple yet easily actionable proposal becomes widely adopted in scientific ML research, it can enhance the practicality of models and thereby accelerate scientific discovery.

References

- Abbasi, M., Santos, B. P., Pereira, T. C., Sofia, R., Monteiro, N. R. C., Simões, C. J. V., Brito, R. M. M., Ribeiro, B., Oliveira, J. L., and Arrais, J. P. Designing optimized drug candidates with generative adversarial network. *Journal of Cheminformatics*, 14(1):40, 2022. ISSN 1758-2946. doi: 10.1186/s13321-022-00623-6. (Cited on page 1)
- Andersson, M. P. and Uvdal, P. New scale factors for harmonic vibrational frequencies using the b3lyp density functional method with the triple- ζ basis set 6-311+g(d,p). *The Journal of Physical Chemistry A*, 109(12):2937–2941, 2005. doi: 10.1021/jp045733a. (Cited on page 3)
- Arisaka, S. and Li, Q. Principled acceleration of iterative numerical methods using machine learning. In *Proc. the International Conference on Machine Learning (ICML)*, 2023. (Cited on page 8)
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M. K., Dalwadi, U., Yip, C. K., Burke, J. E., Garcia, K. C., Grishin, N. V., Adams, P. D., Read, R. J., and Baker, D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021. doi: 10.1126/science.abj8754. (Cited on page 1)
- Bahn, S. R. and Jacobsen, K. W. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.*, 4(3):56–66, 2002. ISSN 1521-9615. doi: 10.1109/5992.998641. (Cited on page 5)
- Bai, S., Kolter, J. Z., and Koltun, V. Deep equilibrium models. In *Proc. the Advances in Neural Information Processing Systems (NeurIPS)*, 2019. (Cited on page 6)
- Banner, R., Nahshan, Y., and Soudry, D. Post training 4-bit quantization of convolutional networks for rapid-deployment. In *Proc. the Advances in Neural Information Processing Systems (NeurIPS)*, 2019. (Cited on page 2)
- Batatia, I., Kovacs, D. P., Simm, G., Ortner, C., and Csanyi, G. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Proc. the Advances in Neural Information Processing Systems (NeurIPS)*, 2022. (Cited on page 5, 8)
- Batatia, I., Batzner, S., Kovács, D. P., Musaelian, A., Simm, G. N. C., Drautz, R., Ortner, C., Kozinsky, B., and Csányi, G. The design space of e(3)-equivariant atom-centred interatomic potentials. *Nature Machine Intelligence*, 7(1):56–67, 2025. ISSN 2522-5839. doi: 10.1038/s42256-024-00956-x. (Cited on page 6)
- Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J. P., Kornbluth, M., Molinari, N., Smidt, T. E., and Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29939-5. (Cited on page 5, 6)
- Bauer, P., Thorpe, A., and Brunet, G. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015. ISSN 1476-4687. doi: 10.1038/nature14956. (Cited on page 6)
- Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A*, 38(6):3098, 1988. (Cited on page 3)
- Behler, J. and Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.*, 98:146401, 2007. doi: 10.1103/PhysRevLett.98.146401. (Cited on page 5)
- Besta, M., Blach, N., Kubicek, A., Gerstenberger, R., Podstawski, M., Gianinazzi, L., Gajda, J., Lehmann, T., Niewiadomski, H., Nyczyk, P., and Hoefler, T. Graph of thoughts: Solving elaborate problems with large language models. In *Proc. the AAAI Conference on Artificial Intelligence (AAAI)*, 2024. (Cited on page 7)
- Bewersdorff, A., Hartmann, C., Hornberger, M., Seßler, K., Bannert, M., Kasneci, E., Kasneci, G., Zhai, X., and Nerdel, C. Taking the next step with generative artificial intelligence: The transformative role of multimodal large language models in science education. *Learning and Individual Differences*, 118:102601, 2025. ISSN 1041-6080. doi: <https://doi.org/10.1016/j.lindif.2024.102601>. (Cited on page 8)
- Black, F. and Scholes, M. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3): 637–654, 1973. (Cited on page 6)
- Bojar, O., Buck, C., Federmann, C., Haddow, B., Koehn, P., Leveling, J., Monz, C., Pecina, P., Post, M., Saint-Amand, H., Soricut, R., Specia, L., and Tamchyna, A. Findings of the 2014 workshop on statistical machine translation. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, 2014. doi: 10.3115/v1/W14-3302. (Cited on page 1)
- Bowen, H. J. M. and Sutton, L. E. *Tables of interatomic distances and configuration in molecules and ions*. Number 11. Chemical Society, 1958. (Cited on page 3)

- 495 Butcher, J. C. *Numerical methods for ordinary differential*
496 *equations*. John Wiley & Sons, 2016. (Cited on page 6)
497
- 498 Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud,
499 D. K. Neural ordinary differential equations. In *Proc.*
500 *the Advances in Neural Information Processing Systems*
501 *(NeurIPS)*, 2018. (Cited on page 6)
502
- 503 Cherti, M., Beaumont, R., Wightman, R., Wortsman, M.,
504 Ilharco, G., Gordon, C., Schuhmann, C., Schmidt, L.,
505 and Jitsev, J. Reproducible scaling laws for contrastive
506 language-image learning. In *Proc. of the IEEE conference*
507 *on computer vision and pattern recognition (CVPR)*, pp.
508 2818–2829, 2023. (Cited on page 2)
509
- 510 DeepSeek-AI. Deepseek-v3 technical report, 2024. (Cited
511 on page 7)
512
- 513 Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei,
514 L. Imagenet: A large-scale hierarchical image database.
515 In *Proc. of the IEEE conference on computer vision and*
516 *pattern recognition (CVPR)*, 2009. doi: 10.1109/CVPR.
517 2009.5206848. (Cited on page 1)
518
- 519 Dettmers, T., Lewis, M., Belkada, Y., and Zettlemoyer, L.
520 Gpt3.int8(): 8-bit matrix multiplication for transformers
521 at scale. In *Proc. the Advances in Neural Information*
522 *Processing Systems (NeurIPS)*, pp. 30318–30332, 2022.
523 (Cited on page 2, 7)
524
- 525 Fuchs, F., Worrall, D., Fischer, V., and Welling, M. Se(3)-
526 transformers: 3d roto-translation equivariant attention
527 networks. In *Proc. the Advances in Neural Information*
528 *Processing Systems (NeurIPS)*, 2020. (Cited on page 5)
529
- 530 Gasteiger, J., Yeshwanth, C., and Günnemann, S. Direc-
531 tional message passing on molecular graphs via synthetic
532 coordinates. In *Proc. the Advances in Neural Information*
533 *Processing Systems (NeurIPS)*, pp. 15421–15433, 2021.
534 (Cited on page 5)
535
- 536 Giannozzi, P., Baroni, S., Bonini, N., Calandra, M., Car, R.,
537 Cavazzoni, C., Ceresoli, D., Chiarotti, G. L., Cococcioni,
538 M., Dabo, I., Dal Corso, A., de Gironcoli, S., Fabris, S.,
539 Fratesi, G., Gebauer, R., Gerstmann, U., Gougoussis, C.,
540 Kokalj, A., Lazzeri, M., Martin-Samos, L., Marzari, N.,
541 Mauri, F., Mazzarello, R., Paolini, S., Pasquarello, A.,
542 Paulatto, L., Sbraccia, C., Scandolo, S., Sclauzero, G.,
543 Seitsonen, A. P., Smogunov, A., Umari, P., and Wentz-
544 covitch, R. M. Quantum espresso: a modular and open-
545 source software project for quantum simulations of ma-
546 terials. *Journal of Physics: Condensed Matter*, 21(39):
547 395502, 2009. doi: 10.1088/0953-8984/21/39/395502.
548 (Cited on page 3)
549
- Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and
Dahl, G. E. Neural message passing for quantum chem-
istry. In *Proc. the International Conference on Machine*
Learning (ICML), 2017. (Cited on page 5)
- Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D.,
Hernández-Lobato, J. M., Sánchez-Lengeling, B., She-
berla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams,
R. P., and Aspuru-Guzik, A. Automatic chemical de-
sign using a data-driven continuous representation of
molecules. *ACS Central Science*, 4(2):268–276, 2018.
ISSN 2374-7943. doi: 10.1021/acscentsci.7b00572.
(Cited on page 1)
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B.,
Warde-Farley, D., Ozair, S., Courville, A., and Bengio,
Y. Generative adversarial nets. In *Proc. the Advances in*
Neural Information Processing Systems (NeurIPS), 2014.
(Cited on page 1)
- Hilton, J., Tang, J., and Schulman, J. Scaling laws for
single-agent reinforcement learning. *arXiv preprint*
arXiv:2301.13442, 2023. (Cited on page 2)
- Ho, J., Jain, A., and Abbeel, P. Denoising diffusion proba-
bilistic models. In *Proc. the Advances in Neural Informa-*
tion Processing Systems (NeurIPS), 2020. (Cited on page
1, 6)
- Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M.
Equivariant diffusion for molecule generation in 3D. In
Proc. the International Conference on Machine Learning
(ICML), pp. 8867–8887, 2022. (Cited on page 1)
- Jacobs, R., Polak, M. P., Schultz, L. E., Mahdavi, H.,
Honavar, V., and Morgan, D. Regression with large lan-
guage models for materials and molecular property pre-
diction. *arXiv preprint arXiv:2409.06080*, 2024. (Cited
on page 7)
- Jones, R. O. and Gunnarsson, O. The density functional
formalism, its applications and prospects. *Reviews of*
Modern Physics, 61(3):689, 1989. (Cited on page 2)
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnoy, M.,
Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek,
A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.
A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B.,
Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S.,
Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pa-
cholska, M., Berghammer, T., Bodenstern, S., Silver, D.,
Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P.,
and Hassabis, D. Highly accurate protein structure predic-
tion with alphafold. *Nature*, 596(7873):583–589, 2021.
ISSN 1476-4687. doi: 10.1038/s41586-021-03819-2.
(Cited on page 1)

- 550 Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B.,
551 Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and
552 Amodei, D. Scaling laws for neural language models.
553 *arXiv preprint arXiv:2001.08361*, 2020. (Cited on page
554 2)
- 555 Karagiorgi, G., Kasieczka, G., Kravitz, S., Nachman, B.,
556 and Shih, D. Machine learning in the search for new
557 fundamental physics. *Nature Reviews Physics*, 4(6):399–
558 412, 2022. (Cited on page 1)
- 559 Kim, S., Noh, J., Gu, G. H., Aspuru-Guzik, A., and Jung,
560 Y. Generative adversarial networks for crystal struc-
561 ture prediction. *ACS Central Science*, 6(8):1412–1420,
562 2020. ISSN 2374-7943. doi: 10.1021/acscentsci.0c00426.
563 (Cited on page 1)
- 564 Kingma, D. P. and Welling, M. Auto-encoding variational
565 bayes. In *Proc. the International Conference on Learning
566 Representations (ICLR)*, 2014. (Cited on page 1)
- 567 Kocer, E., Ko, T. W., and Behler, J. Neural network poten-
568 tials: A concise overview of methods. *Annual Review
569 of Physical Chemistry*, 73(Volume 73, 2022):163–186,
570 2022. ISSN 1545-1593. doi: [https://doi.org/10.1146/
571 annurev-physchem-082720-034254](https://doi.org/10.1146/annurev-physchem-082720-034254). (Cited on page 2, 5)
- 572 Kohn, W. and Sham, L. J. Self-consistent equations includ-
573 ing exchange and correlation effects. *Physical review*,
574 140(4A):A1133, 1965. (Cited on page 3)
- 575 Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., and Iwasawa, Y.
576 Large language models are zero-shot reasoners. In *Proc.
577 the Advances in Neural Information Processing Systems
578 (NeurIPS)*, 2022. (Cited on page 7)
- 579 Kovács, D. P., Batatia, I., Arany, E. S., and Csányi, G.
580 Evaluation of the mace force field architecture: From
581 medicinal chemistry to materials science. *The Journal
582 of Chemical Physics*, 159(4):044118, 2023. ISSN 0021-
583 9606. doi: 10.1063/5.0155322. (Cited on page 8)
- 584 Kumbhar, S., Mishra, V., Coutinho, K., Handa, D., Iquebal,
585 A., and Baral, C. Hypothesis generation for materials
586 discovery and design using goal-driven and constraint-
587 guided llm agents. *arXiv preprint arXiv:2501.13299*,
588 2025. (Cited on page 8)
- 589 Käser, S., Vazquez-Salazar, L. I., Meuwly, M., and Töpfer,
590 K. Neural network potentials for chemistry: concepts,
591 applications and prospects. *Digital Discovery*, 2:28–58,
592 2023. doi: 10.1039/D2DD00102K. (Cited on page 2, 5)
- 593 Lai, J., Gan, W., Wu, J., Qi, Z., and Yu, P. S. Large language
594 models in law: A survey. *AI Open*, 5:181–196, 2024.
595 ISSN 2666-6510. doi: [https://doi.org/10.1016/j.aiopen.
596 2024.09.002](https://doi.org/10.1016/j.aiopen.2024.09.002). (Cited on page 1)
- 597 Larsen, A. H., Mortensen, J. J., Blomqvist, J., Castelli,
598 I. E., Christensen, R., Duřak, M., Friis, J., Groves, M. N.,
599 Hammer, B., Hargus, C., Hermes, E. D., Jennings, P. C.,
600 Jensen, P. B., Kermode, J., Kitchin, J. R., Kolsbjerg, E. L.,
601 Kubal, J., Kaasbjerg, K., Lysgaard, S., Maronsson, J. B.,
602 Maxson, T., Olsen, T., Pastewka, L., Peterson, A., Rost-
603 gaard, C., Schiøtz, J., Schütt, O., Strange, M., Thygesen,
604 K. S., Vegge, T., Vilhelmsen, L., Walter, M., Zeng, Z.,
and Jacobsen, K. W. The atomic simulation environ-
ment—a python library for working with atoms. *Journal
of Physics: Condensed Matter*, 29(27):273002, 2017.
(Cited on page 5)
- Li, S., Liu, Z., Luo, Y., Wang, X., He, X., Kawaguchi, K.,
Chua, T.-S., and Tian, Q. Towards 3d molecule-text inter-
pretation in language models. In *Proc. the International
Conference on Learning Representations (ICLR)*, 2024.
(Cited on page 7)
- Liao, Y.-L. and Smidt, T. Equiformer: Equivariant graph
attention transformer for 3d atomistic graphs. In *Proc. the
International Conference on Learning Representations
(ICLR)*, 2023. (Cited on page 5)
- Lim, J., Ryu, S., Kim, J. W., and Kim, W. Y. Molecu-
lar generative model based on conditional variational
autoencoder for de novo molecular design. *Journal of
Cheminformatics*, 10(1):31, 2018. ISSN 1758-2946. doi:
10.1186/s13321-018-0286-7. (Cited on page 1)
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A.,
Ciompi, F., Ghafoorian, M., van der Laak, J. A., van
Ginneken, B., and Sánchez, C. I. A survey on deep
learning in medical image analysis. *Medical Image Anal-
ysis*, 42:60–88, 2017. ISSN 1361-8415. doi: <https://doi.org/10.1016/j.media.2017.07.005>. (Cited on page 1)
- Liu, F., Zhao, W., He, Z., Wang, Y., Wang, Z. W., Dai,
C., Liang, X., and Jiang, L. Improving neural network
efficiency via post-training quantization with adaptive
floating-point. In *Proc. of the IEEE international confer-
ence on computer vision (ICCV)*, 2021. (Cited on page
7)
- Liu, S., Wen, T., Ye, B., Li, Z., and Srolovitz, D. J. Large
language models for material property predictions: elastic
constant tensor prediction and materials design. *arXiv
preprint arXiv:2411.12280*, 2024. (Cited on page 7)
- Liu, S.-y., Liu, Z., Huang, X., Dong, P., and Cheng, K.-T.
LLM-FP4: 4-bit floating-point quantized transformers. In
*Proc. of the Conference on Empirical Methods in Natural
Language Processing (EMNLP)*, pp. 592–605, 2023. doi:
10.18653/v1/2023.emnlp-main.39. (Cited on page 2, 7)
- Llama Team, A. . M. The llama 3 herd of models, 2024.
(Cited on page 7)

- Lu, C., Lu, C., Lange, R. T., Foerster, J., Clune, J., and Ha, D. The ai scientist: Towards fully automated open-ended scientific discovery. *arXiv preprint arXiv:2408.06292*, 2024. (Cited on page 8)
- Maxson, T., Soyemi, A., Chen, B. W. J., and Szilvási, T. Enhancing the quality and reliability of machine learning interatomic potentials through better reporting practices. *The Journal of Physical Chemistry C*, 128(16): 6524–6537, 2024. ISSN 1932-7447. doi: 10.1021/acs.jpcc.4c00028. (Cited on page 6)
- Maxwell, J. C. Viii. a dynamical theory of the electromagnetic field. *Philosophical transactions of the Royal Society of London*, (155):459–512, 1865. (Cited on page 4)
- Maxwell, J. C. On physical lines of force. *Philosophical magazine*, 90(S1):11–23, 2010. (Cited on page 4)
- Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., Ginsburg, B., Houston, M., Kuchaiev, O., Venkatesh, G., and Wu, H. Mixed precision training. In *Proc. the International Conference on Learning Representations (ICLR)*, 2018. (Cited on page 2)
- Morgan, D. and Jacobs, R. Opportunities and challenges for machine learning in materials science. *Annual Review of Materials Research*, 50(1):71–103, 2020. (Cited on page 1)
- Nakamura, Y., Shiratori, S., Takagi, R., Sutoh, M., Sugihara, I., Nagano, H., and Shimano, K. Physics-informed neural network applied to surface-tension-driven liquid film flows. *International Journal for Numerical Methods in Fluids*, 94(9):1359–1378, 2022. doi: <https://doi.org/10.1002/fld.5093>. (Cited on page 6)
- Napier, K. J. Improved initial guesses for numerical solutions of kepler’s equation. *arXiv preprint arXiv:2411.15374*, 2024. (Cited on page 8)
- Neumann, O. and Gros, C. Scaling laws for a multi-agent reinforcement learning model. In *Proc. the International Conference on Learning Representations (ICLR)*, 2023. (Cited on page 2)
- Nguyen, E., Poli, M., Durrant, M. G., Kang, B., Katrekar, D., Li, D. B., Bartie, L. J., Thomas, A. W., King, S. H., Brixi, G., Sullivan, J., Ng, M. Y., Lewis, A., Lou, A., Ermon, S., Baccus, S. A., Hernandez-Boussard, T., Ré, C., Hsu, P. D., and Hie, B. L. Sequence modeling and design from molecular to genome scale with evo. *Science*, 386(6723): eado9336, 2024. doi: 10.1126/science.ado9336. (Cited on page 1, 2)
- Oskooi, A. F., Roundy, D., Ibanescu, M., Bermel, P., Joannopoulos, J., and Johnson, S. G. Meep: A flexible free-software package for electromagnetic simulations by the fdtd method. *Computer Physics Communications*, 181(3):687–702, 2010. ISSN 0010-4655. doi: <https://doi.org/10.1016/j.cpc.2009.11.008>. (Cited on page 4)
- Ozbayoglu, A. M., Gudelek, M. U., and Sezer, O. B. Deep learning for financial applications : A survey. *Applied Soft Computing*, 93:106384, 2020. ISSN 1568-4946. doi: <https://doi.org/10.1016/j.asoc.2020.106384>. (Cited on page 1)
- Park, J., Bae, M., Ko, D., and Kim, H. J. LLamo: Large language model-based molecular graph assistant. In *Proc. the Advances in Neural Information Processing Systems (NeurIPS)*, 2024. (Cited on page 7)
- Peng, X., Guan, J., Liu, Q., and Ma, J. MolDiff: Addressing the atom-bond inconsistency problem in 3D molecule diffusion generation. In *Proc. the International Conference on Machine Learning (ICML)*, pp. 27611–27629, 2023. (Cited on page 1)
- Planck, M. Zur theorie des gesetzes der energieverteilung im normalspektrum. *Berlin*, pp. 237–245, 1900. (Cited on page 3)
- Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117(1):1–19, 1995. ISSN 0021-9991. doi: <https://doi.org/10.1006/jcph.1995.1039>. (Cited on page 5)
- Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., Gross, S. S., Dorfman, L., McLean, C. Y., and DePristo, M. A. A universal snp and small-indel variant caller using deep neural networks. *Nature Biotechnology*, 36(10):983–987, 2018. ISSN 1546-1696. doi: 10.1038/nbt.4235. (Cited on page 1)
- Prykhodko, O., Johansson, S. V., Kotsias, P.-C., Arús-Pous, J., Bjerrum, E. J., Engkvist, O., and Chen, H. A de novo molecular generation method using latent vector based generative adversarial network. *Journal of Cheminformatics*, 11(1):74, 2019. ISSN 1758-2946. doi: 10.1186/s13321-019-0397-9. (Cited on page 1)
- Pukrittayakamee, A., Malshe, M., Hagan, M., Raff, L. M., Narulkar, R., Bukkapatnum, S., and Komanduri, R. Simultaneous fitting of a potential-energy surface and its corresponding force fields using feedforward neural networks. *The Journal of Chemical Physics*, 130(13):134101, 2009. ISSN 0021-9606. doi: 10.1063/1.3095491. (Cited on page 5)

- 660 Raghu, M. and Schmidt, E. A survey of deep learning for
661 scientific discovery. *arXiv preprint arXiv:2003.11755*,
662 2020. (Cited on page 1)
- 663 Raissi, M., Perdikaris, P., and Karniadakis, G. Physics-
664 informed neural networks: A deep learning framework for
665 solving forward and inverse problems involving nonlinear
666 partial differential equations. *Journal of Computational
667 Physics*, 378:686–707, 2019. ISSN 0021-9991. doi:
668 <https://doi.org/10.1016/j.jcp.2018.10.045>. (Cited on page
669 2, 6)
- 670 Ren, X., Li, X., Ren, K., Song, J., Xu, Z., Deng, K., and
671 Wang, X. Deep learning-based weather prediction: A
672 survey. *Big Data Research*, 23:100178, 2021. ISSN 2214-
673 5796. doi: <https://doi.org/10.1016/j.bdr.2020.100178>.
674 (Cited on page 1)
- 675 Rubungo, A. N., Li, K., Hatrnick-Simpers, J., and Dieng,
676 A. B. Llm4mat-bench: Benchmarking large language
677 models for materials property prediction. *arXiv preprint
678 arXiv:2411.00177*, 2024. (Cited on page 7)
- 679 Sanchez-Lengeling, B. and Aspuru-Guzik, A. Inverse
680 molecular design using machine learning: Generative
681 models for matter engineering. *Science*, 361(6400):360–
682 365, 2018. doi: 10.1126/science.aat2663. (Cited on page
683 1)
- 684 Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) equiv-
685 ariant graph neural networks. In *Proc. the International
686 Conference on Machine Learning (ICML)*, 2021. (Cited
687 on page 5)
- 688 Saverio, L. Accelerating convergence of linear iterative
689 solvers using machine learning. Master’s thesis, Politec-
690 nico di Milano, 2023. (Cited on page 8)
- 691 Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and
692 Monfardini, G. The graph neural network model. *IEEE
693 Transactions on Neural Networks*, 20(1):61–80, 2009.
694 doi: 10.1109/TNN.2008.2005605. (Cited on page 5)
- 695 Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C. W.,
696 Wightman, R., Cherti, M., Coombes, T., Katta, A., Mullis,
697 C., Wortsman, M., Schramowski, P., Kundurthy, S. R.,
698 Crowson, K., Schmidt, L., Kaczmarczyk, R., and Jitsev, J.
699 LAION-5b: An open large-scale dataset for training next
700 generation image-text models. In *Proc. the Advances in
701 Neural Information Processing Systems (NeurIPS)*, 2022.
702 (Cited on page 1)
- 703 Schütt, K., Kindermans, P.-J., Saucedo Felix, H. E., Chmiela,
704 S., Tkatchenko, A., and Müller, K.-R. Schnet: A
705 continuous-filter convolutional neural network for mod-
706 eling quantum interactions. In *Proc. the Advances in
707 Neural Information Processing Systems (NeurIPS)*, 2017.
708 (Cited on page 5)
- 709 Sharma, R. and Shankar, V. Accelerated training of physics-
710 informed neural networks (pinns) using meshless dis-
711 cretizations. In *Proc. the Advances in Neural Information
712 Processing Systems (NeurIPS)*, 2022. (Cited on page 6)
- 713 Shi, J., Ma, Q., Ma, H., and Li, L. Scaling law for time series
714 forecasting. In *Proc. the Advances in Neural Information
715 Processing Systems (NeurIPS)*, 2024. (Cited on page 2)
- 716 Smith, J. S., Isayev, O., and Roitberg, A. E. Ani-1: an
717 extensible neural network potential with dft accuracy at
718 force field computational cost. *Chem. Sci.*, 8:3192–3203,
719 2017. doi: 10.1039/C6SC05720A. (Cited on page 5)
- 720 Snell, C., Lee, J., Xu, K., and Kumar, A. Scaling llm test-
721 time compute optimally can be more effective than scal-
722 ing model parameters. *arXiv preprint arXiv:2408.03314*,
723 2024. (Cited on page 2)
- 724 Song, Y. and Ermon, S. Generative modeling by estimating
725 gradients of the data distribution. In *Advances in Neural
726 Information Processing Systems*, 2019. (Cited on page 1,
727 6)
- 728 Strauss, W. A. *Partial differential equations: An introduc-
729 tion*. John Wiley & Sons, 2007. (Cited on page 6)
- 730 The Royal Swedish Academy of Sciences. The
731 nobel prize in chemistry 2024, 2024. URL
732 [https://www.nobelprize.org/prizes/
733 chemistry/2024/press-release/](https://www.nobelprize.org/prizes/chemistry/2024/press-release/). (Cited on
734 page 1)
- 735 Thölke, P. and Fabritiis, G. D. Equivariant transformers for
736 neural network based molecular potentials. In *Proc. the
737 International Conference on Learning Representations
738 (ICLR)*, 2022. (Cited on page 5)
- 739 Thompson, A. P., Aktulga, H. M., Berger, R., Bolintineanu,
740 D. S., Brown, W. M., Crozier, P. S., in ’t Veld, P. J.,
741 Kohlmeyer, A., Moore, S. G., Nguyen, T. D., Shan, R.,
742 Stevens, M. J., Tranchida, J., Trott, C., and Plimpton, S. J.
743 LAMMPS - a flexible simulation tool for particle-based
744 materials modeling at the atomic, meso, and continuum
745 scales. *Computer Physics Communications*, 271:108171,
746 2022. ISSN 0010-4655. doi: [https://doi.org/10.1016/j.
747 cpc.2021.108171](https://doi.org/10.1016/j.cpc.2021.108171). (Cited on page 5)
- 748 Tirado-Rives, J. and Jorgensen, W. L. Performance of b3lyp
749 density functional methods for a large set of organic
750 molecules. *Journal of Chemical Theory and Computa-
751 tion*, 4(2):297–306, 2008. doi: 10.1021/ct700248k.
752 (Cited on page 3)
- 753 Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux,
754 M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E.,
755 Azhar, F., Rodriguez, A., Joulin, A., Grave, E., and Lam-
756 ple, G. Llama: Open and efficient foundation language
757 models, 2023a. (Cited on page 7)

- 715 Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi,
716 A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P.,
717 Bhosale, S., Bikel, D., Blecher, L., Ferrer, C. C., Chen,
718 M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W.,
719 Fuller, B., Gao, C., Goswami, V., Goyal, N., Hartshorn,
720 A., Hosseini, S., Hou, R., Inan, H., Kardas, M., Kerkez,
721 V., Khabsa, M., Kloumann, I., Korenev, A., Koura, P. S.,
722 Lachaux, M.-A., Lavril, T., Lee, J., Liskovich, D., Lu, Y.,
723 Mao, Y., Martinet, X., Mihaylov, T., Mishra, P., Molybog,
724 I., Nie, Y., Poulton, A., Reizenstein, J., Rungta, R., Saladi,
725 K., Schelten, A., Silva, R., Smith, E. M., Subramanian, R.,
726 Tan, X. E., Tang, B., Taylor, R., Williams, A., Kuan, J. X.,
727 Xu, P., Yan, Z., Zarov, I., Zhang, Y., Fan, A., Kambadur,
728 M., Narang, S., Rodriguez, A., Stojnic, R., Edunov, S.,
729 and Scialom, T. Llama 2: Open foundation and fine-tuned
730 chat models, 2023b. (Cited on page 7)
- 731 Tritton, D. J. *Physical fluid dynamics*. Springer Science &
732 Business Media, 2012. (Cited on page 6)
- 734 Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones,
735 L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. At-
736 tention is all you need. In *Proc. the Advances in Neural*
737 *Information Processing Systems (NeurIPS)*, 2017. (Cited
738 on page 5)
- 740 Walker, C., Strassel, S., Medero, J., and Maeda, K. Ace
741 2005 multilingual training corpus, 2006. (Cited on page
742 1)
- 744 Webb, S. et al. Deep learning for biology. *Nature*, 554
745 (7693):555–557, 2018. (Cited on page 1)
- 746 Wei, J., Wang, X., Schuurmans, D., Bosma, M., ichter, b.,
747 Xia, F., Chi, E., Le, Q. V., and Zhou, D. Chain-of-thought
748 prompting elicits reasoning in large language models. In
749 *Proc. the Advances in Neural Information Processing*
750 *Systems (NeurIPS)*, 2022. (Cited on page 7)
- 752 Xie, T., Wan, Y., Liu, Y., Zeng, Y., Zhang, W., Kit, C., Zhou,
753 D., and Hoex, B. Darwin 1.5: Large language models
754 as materials science adapted learners. *arXiv preprint*
755 *arXiv:2412.11970*, 2024. (Cited on page 8)
- 757 Xu, Y., Han, X., Yang, Z., Wang, S., Zhu, Q., Liu, Z., Liu,
758 W., and Che, W. Onebit: Towards extremely low-bit
759 large language models. In *Proc. the Advances in Neural*
760 *Information Processing Systems (NeurIPS)*, 2024. (Cited
761 on page 2, 7)
- 763 Yanai, T., Tew, D. P., and Handy, N. C. A new hy-
764 brid exchange–correlation functional using the coulomb-
765 attenuating method (cam-b3lyp). *Chemical Physics*
766 *Letters*, 393(1):51–57, 2004. ISSN 0009-2614. doi:
767 <https://doi.org/10.1016/j.cplett.2004.06.011>. (Cited on
768 page 3)
- 769 Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao,
Y., and Narasimhan, K. R. Tree of thoughts: Deliberate
problem solving with large language models. In *Proc.*
the Advances in Neural Information Processing Systems
(NeurIPS), 2023. (Cited on page 7)
- Yee, K. Numerical solution of initial boundary value prob-
lems involving maxwell’s equations in isotropic media.
IEEE Transactions on antennas and propagation, 14(3):
302–307, 1966. (Cited on page 2)
- Yoo, J., Cho, Y., Kim, D. H., Kim, J., Lee, T. G., Lee,
S. M., Choo, J., and Jeong, M. S. Unraveling the role
of raman modes in evaluating the degree of reduction
in graphene oxide via explainable artificial intelligence.
Nano Today, 57:102366, 2024. ISSN 1748-0132. doi:
<https://doi.org/10.1016/j.nantod.2024.102366>. (Cited on
page 1)
- Zeni, C., Pinsler, R., Zügner, D., Fowler, A., Horton, M., Fu,
X., Wang, Z., Shysheya, A., Crabbé, J., Ueda, S., Sordillo,
R., Sun, L., Smith, J., Nguyen, B., Schulz, H., Lewis, S.,
Huang, C.-W., Lu, Z., Zhou, Y., Yang, H., Hao, H., Li, J.,
Yang, C., Li, W., Tomioka, R., and Xie, T. A generative
model for inorganic materials design. *Nature*, Jan 2025.
ISSN 1476-4687. doi: 10.1038/s41586-025-08628-5.
(Cited on page 1)
- Zhai, X., Kolesnikov, A., Houlsby, N., and Beyer, L. Scaling
vision transformers. In *Proc. of the IEEE conference*
on computer vision and pattern recognition (CVPR), pp.
12104–12113, 2022. (Cited on page 2)
- Zhang, Q., Ding, K., Lv, T., Wang, X., Yin, Q., Zhang, Y.,
Yu, J., Wang, Y., Li, X., Xiang, Z., et al. Scientific large
language models: A survey on biological & chemical
domains. *ACM Computing Surveys*, 2024. (Cited on page
2)
- Zhou, G., Gao, Z., Ding, Q., Zheng, H., Xu, H., Wei, Z.,
Zhang, L., and Ke, G. Uni-mol: A universal 3d molecular
representation learning framework. In *Proc. the Interna-*
tional Conference on Learning Representations (ICLR),
2023. (Cited on page 7)
- Zhou, J. and Troyanskaya, O. G. Predicting effects of
noncoding variants with deep learning–based sequence
model. *Nature Methods*, 12(10):931–934, 2015. ISSN
1548-7105. doi: 10.1038/nmeth.3547. (Cited on page 1)
- Zhu, X., Li, J., Liu, Y., Ma, C., and Wang, W. A survey on
model compression for large language models. *Transac-*
tions of the Association for Computational Linguistics,
12:1556–1577, 2024. doi: 10.1162/tacl_a_00704. (Cited
on page 2)

770 Zrimec, J., Fu, X., Muhammad, A. S., Skrekas, C., Jau-
771 niskis, V., Speicher, N. K., Börlin, C. S., Verendel, V.,
772 Chehreghani, M. H., Dubhashi, D., Siewers, V., David,
773 F., Nielsen, J., and Zelezniak, A. Controlling gene ex-
774 pression with deep generative design of regulatory dna.
775 *Nature Communications*, 13(1):5099, 2022. ISSN 2041-
776 1723. doi: 10.1038/s41467-022-32818-8. (Cited on page
777 1)

778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824

A. Details in quantum mechanical calculation

Quantum mechanical calculations (also known as quantum chemistry calculations) were performed using PySCF (Python-based Simulations of Chemistry Framework, version 2.7.0). The input water molecule consists of a single oxygen atom at (0.000000, 0.000000, 0.000000) and two hydrogen atoms at (0.757000, 0.586000, 0.000000) and (-0.757000, 0.586000, 0.000000), respectively. All simulations were conducted using two nodes of an AMD EPYC 7543 32-Core Processor. To compare the geometry optimization result of a water molecule based on different exchange functionals, we performed both Hartree-Fock calculations and density functional theory calculations using the B3LYP functional. For both methods, we employed the 6-311++G(d,p) basis set. To evaluate the impact of numerical precision, we conducted the same calculations using both single precision (FP32) and double precision (FP64) by declaring `np.float32` and `np.float64`, respectively. The default convergence tolerances for structural stabilization were set as follows: $|\Delta E| < 1.00 \times 10^{-6}$, $\text{RMS-Grad} < 3.00 \times 10^{-4}$, $\text{Max-Grad} < 4.50 \times 10^{-4}$, $\text{RMS-Disp} < 1.20 \times 10^{-3}$, and $\text{Max-Disp} < 1.80 \times 10^{-3}$.

B. Finite-difference time-domain calculation

B.1. Nonlinear Material Properties

In this study, Kerr media were modeled with a second-order nonlinear susceptibility (χ^2) ranging from 10^{-12} to 10^{-2} and the refractive index was set to 1.65 to mimic the conventional nonlinear materials like beta barium borate. The nonlinear polarization of the material was expressed as:

$$P = \epsilon_0(\chi^{(1)}E + \chi^{(2)}E^2 + \chi^{(3)}E^3 + \dots) \quad (1)$$

And the second-order nonlinear polarization term is represented as: $P^{(2)} = \epsilon_0\chi^{(2)}E^2$. Meep incorporates such nonlinear polarization terms into Maxwell's equations to simulate interactions between electromagnetic waves and the material in the time domain

$$\nabla \times H = \epsilon_0 \frac{\partial E}{\partial t} + \frac{\partial P}{\partial t} \quad (2)$$

$$\nabla \times E = -\mu_0 \frac{\partial H}{\partial t} \quad (3)$$

B.2. Simulation Setup

The simulation domain consisted of a $100 \mu\text{m}$ medium, a $1 \mu\text{m}$ thick boundary layer, and $2 \mu\text{m}$ buffer regions at both ends. The spatial resolution was user-defined to capture fine electromagnetic field characteristics. Kerr media were placed at the center of the domain, with χ^2 explicitly defined. The calculations were conducted using both FP32 and FP64 precision on single core of AMD Ryzen5 8500G.

B.3. Source and Monitor Definition

The source was defined as a Gaussian plane wave with a central wavelength of $1.55 \mu\text{m}$ and a bandwidth of 0.15. Both the source and monitors were positioned $1 \mu\text{m}$ outside the nonlinear medium, with the electric field oscillating along the x-axis. Simulations were executed to allow sufficient decay of the fields after the source was turned off to confirm accurate measurements.

B.4. Harmonic Generation and Analysis

Using the Meep's add flux function, the optical flux outside the nonlinear medium was measured, and the transmitted power spectra of the fundamental frequency (ω) and harmonic components (2ω , 3ω) were calculated. The add flux function records the time-domain values of electric and magnetic fields at specific locations, then performs a Fourier transform to convert them into the frequency domain to compute flux. This process allows precise analysis of the intensity of each frequency component within the user-defined frequency range and intervals. The analysis frequency range extended from $\omega/2$ to 3.5ω , encompassing all relevant frequency bands of interest. Flux measurements were particularly useful for understanding the interaction between newly generated harmonic components and existing frequency components caused by the material's nonlinearity.

880 **B.5. Results and Reproducibility**

881 Simulation results demonstrated how the intensity and distribution of harmonic components varied with changes in χ^2 . The
882 nonlinear modeling capabilities of Meep enabled precise analysis of harmonic generation characteristics in nonlinear optical
883 materials.
884

885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934