

Solving layered rough contact via a physics-informed operator framework

Shaoqi Huang

Xi'an Jiaotong Univeristy

Junpeng Chen

Xi'an Jiaotong Univeristy

Haoyu Gu

Xi'an Jiaotong Univeristy

Shuwen Zhang

Xi'an Jiaotong Univeristy

Xiaoshen Zhang

Xi'an Aviation Brake Technology Co., Ltd

Chongpu zhai

zhaichongpu@xjtu.edu.cn

Xi'an Jiaotong Univeristy

Minglong Xu

Xi'an Jiaotong Univeristy

Research Article

Keywords: Layered rough contact, Physics-informed deep learning, Surface roughness, Contact deformation

Posted Date: August 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10674471/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Tribology Letters on October 12th, 2026.
See the published version at <https://doi.org/10.1007/s11249-026-02216-y>.

Solving layered rough contact via a physics-informed operator framework

Shaoqi Huang¹, Junpeng Chen¹, Haoyu Gu¹, Shuwen Zhang¹, Xiaoshen Zhang², Chongpu Zhai^{1*}, and Minglong Xu^{1**}

¹State Key Laboratory for Strength and Vibration of Mechanical Structure, School of Aerospace Engineering, Xi'an Jiaotong University, Xi'an 710049, China.

²Xi'an Aviation Brake Technology Co., Ltd, Xi'an, Shaanxi 710065, China.

*Corresponding author. Email: zhaichongpu@xjtu.edu.cn (C.Z.); mlxu@xjtu.edu.cn (M.X.)

Abstract

Predicting three-dimensional layered rough surface contact behavior is fundamental to modern tribological interface design and surface engineering. However, experimental measurements lack sub-surface resolution, theoretical models rely on oversimplified geometric assumptions, and high-fidelity numerical methods remain computationally prohibitive for multiscale interface problems. To address these limitations, we develop a physics-informed operator-decomposition framework that separates the unknown nonlinear contact-state mapping from the analytically known elastic compliance operator. This framework combines an attention-enhanced U-Net with a Fourier Neural Operator (FNO) to capture localized and long-range interactions while decomposing the predicted surface deformation into elastic and plastic components. By incorporating penalty terms derived from the Karush-Kuhn-Tucker (KKT) contact conditions into the loss function, the framework promotes physically admissible predictions with respect to nonpenetration, pressure nonnegativity, and contact complementarity. Furthermore, a multiscale feature-wise linear modulation (FiLM) module dynamically injects dimensionless physical parameters into the convolutional feature maps, facilitating generalization across broad ranges of coating properties and system scales. Validation against standard boundary element method (BEM) calculations demonstrates that the framework achieves high fidelity across diverse topographies while reducing full-field inference times from minutes to milliseconds. Overall, the proposed framework provides an efficient and physically interpretable surrogate for resolving full-field pressure and displacement responses in layered rough contact across broad geometric, material, and loading regimes.

Keywords: Layered rough contact; Physics-informed deep learning; Surface roughness; Contact deformation.

1. Introduction

In advanced manufacturing and structural design, surface engineering via thin films, coatings, and multi-layered material systems represents a primary paradigm for enhancing the tribological performance and longevity of critical components [1-4].

Across diverse applications, including aerospace propulsion systems [5], brake discs [6, 7], microelectronics [8], and soft robotics [9], these layered architectures must withstand severe mechanical, thermal, and contact loads. The performance of these systems is inherently governed by layered contact mechanics, which dictates the transmission of normal and tangential tractions across interfaces characterized by multiscale surface roughness [10]. Unlike homogeneous media, interfacial layers introduce sharp property discontinuities and pronounced elastoplastic stiffness mismatches between the coating and the underlying substrate [11, 12]. Under contact loading, discrete micro-asperity interactions drive complex elastoplastic deformation fields, generating severe pressure concentrations and critical subsurface shear stresses [13-15]. Consequently, the high-fidelity determination of elastoplastic contact fields on layered rough surfaces is essential for predicting structural reliability, mitigating macroscopic failure modes such as coating delamination, and guiding the rational design of advanced tribological interfaces [16-18].

The systematic investigation of layered contact mechanics initially progressed through analytical frameworks operating under the idealized assumption of geometrically smooth interfaces. Burmister [19] establishes the foundational mathematical solutions for displacements in layered half-spaces under axisymmetric loading. This smooth analytical paradigm achieves a critical milestone through the classic formulation by O'Sullivan and King [20, 21], who derived complete three-dimensional frequency response functions for coated media subjected to combined normal and tangential tractions, thereby defining the frequency-domain mapping between surface pressures and displacement fields. Subsequently, Plumet and Dubourg [22] extended this approach to arbitrary multi-layered configurations by systematically applying transfer matrix methods to enforce interfacial continuity conditions. Later, Gao et. al., [23] utilized modulus perturbation methods and energy conservation principles to derive load-displacement relationships for smooth indenters of arbitrary shape pressing into multi-layered media, providing a theoretical foundation for subsequent layered contact experiments and analyses. However, because real engineering topographies exhibit inherent stochastic fluctuations, theoretical research progressively shifted from smooth deterministic models toward statistical field representations of surface roughness. The classic single-scale Greenwood-Williamson model [24] demonstrated that, for nominally flat rough surfaces, treating independent elastic asperities statistically yields a linear proportional relationship between the real contact area and normal load. To capture the intrinsic multiscale and self-similar characteristics of rough surfaces in layered systems, Persson [25, 26] proposed a diffusion-type statistical theory wherein the evolution of the interfacial pressure probability density across magnification scales obeys a Fokker-Planck-type equation [27]. This framework enables quantitative predictions of real contact area, interfacial separation, and stress distributions for randomly rough coated surfaces.

Despite providing valuable physical insights, analytical theories rely on simplifying assumptions, motivating the development of numerical methods capable of resolving deterministic local contact fields [28]. Among these methods, the boundary element method (BEM) is particularly well suited to rough surface contact because it

confines discretization to the interface and represents the nonlocal pressure-displacement relation [29-31]. Nogi and Kato [32] utilize early semi-analytical approaches to reveal the critical competitive relationship between coating thickness and root-mean-square (RMS) roughness in modulating real contact area. As comprehensively reviewed by Bhushan and Peng [33], these semi-analytical frameworks are inherently constrained by predefined profile assumptions, matrix ill-conditioning, or heavy optimization scaling, respectively. Computational efficiency for multi-contact is improved when Liu et. al., [34, 35] pioneered discrete convolution and fast Fourier transform (FFT) acceleration to maximize boundary spectral efficiency. Extending these spectral formulations to layered elastoplastic configurations, however, requires explicit three-dimensional volumetric integration of plastic strain fields coupled with layered Green's functions [36-38]. This nonlinear coupling introduces prohibitive computational overhead in classical optimization solvers, thereby driving the contact mechanics community toward data-driven and machine-learning-augmented paradigms.

To bridge this computational gap, the contact mechanics community has increasingly turned to artificial intelligence paradigms, evolving through distinct stages of data-driven, physics-informed, and operator-learning frameworks [39-41]. Early methodologies predominantly relied on purely data-driven convolutional neural networks (CNNs) to bypass classical numerical solvers. Specifically, Jang et. al., [42] framed contact prediction as an image segmentation task, employing a U-Net architecture trained on extensive fractal surface datasets to efficiently predict contact distributions. Similarly, Bonagiri et. al., [43] applied an encoder-decoder convolutional network for image-to-image translation to accurately remove tip-convolution artifacts from nanoscale surface profiles, whereas Sahin et. al., [44] deployed diverse data-driven surrogate models to evaluate the effective contact area of rough interfaces. Concurrently, Suman and Prajapati [45] implemented a Bayesian-optimized deep neural network framework trained on boundary element datasets to estimate the real contact area of rough interfaces under varying loads. Recognizing the lack of physical interpretability in these black box models, researchers subsequently shifted toward physics-informed neural networks (PINNs) to embed fundamental governing laws directly into the learning process. A notable advancement is achieved by Zhou and Song [46, 47], who integrate Green's function molecular dynamics with physics-informed networks by incorporating the diffusion equation form of Persson's theory into the loss function to predict multiscale contact stresses. More recently, the research frontier has expanded toward neural operators capable of learning mesh-independent mappings in the frequency domain. Notably, Chen et. al., [48] pioneered AI-enabled high-throughput characterization by developing physics-informed Fourier Kolmogorov-Arnold networks to map indentation responses directly to continuous material property fields. Despite their exceptional inference speed, existing machine learning paradigms face critical mathematical and physical limitations in multiscale, layered elastoplastic rough contacts. Purely data-driven models suffer from unphysical boundary smoothing that degrades localized stress concentrations. Physics-informed frameworks commonly incorporate contact conditions through penalty terms, which may leave residual

violations of nonpenetration and complementarity [11]. Meanwhile, frequency-domain operators introduce spurious numerical oscillations across discontinuous multi-asperity boundaries under severe plastic deformation.

To resolve these long-standing computational trade-offs, this study develops a grey-box, physics-informed operator framework for layered elastoplastic rough contact. Rather than learning the entire contact solution as a black-box mapping, the proposed formulation decomposes the problem into an unresolved nonlinear contact-state mapping and an analytically known elastic compliance operator. The nonlinear contact-state mapping is learned using an attention-enhanced U-Net with a Fourier Neural Operator (FNO) bottleneck, whereas the nonlocal elastic response is explicitly represented by a differentiable FFT layer incorporating the layered Green's-function kernel. Penalty terms derived from the Karush-Kuhn-Tucker (KKT) contact conditions promote nonpenetration and contact complementarity, while a nonnegative output mapping enforces pressure nonnegativity. In addition, multiscale feature-wise linear modulation introduces dimensionless geometric, loading, and material parameters into the learned representations. The primary objective of this work is to overcome the computational bottleneck inherent in traditional layered elastoplastic solvers, reducing prediction times for multiscale rough contacts from minutes to milliseconds. The resulting framework provides a computationally efficient approach for resolving layered rough-contact fields while retaining the known nonlocal elastic response explicitly, thereby enabling systematic investigation of contact pressure, deformation, real contact area, and their dependence on surface and material parameters.

2. Theoretical and computational framework

2.1 Layered contact theory

This section establishes three-dimensional, frictionless, normal contact between a deformable rough indenter and a perfectly bonded single-layer coating on a semi-infinite substrate (Fig. 1). The film is completely adhered to the substrate without any relative sliding. Assuming that the indenter, the film, and the substrate are all explicitly idealized as elastic bodies, the solids on both sides of the contact interface undergo elastic deformation simultaneously. In the contact region, the interface interaction follows Newton's third law. By traction equilibrium, the two bodies experience equal and opposite normal contact pressures $p_n(x, y)$. Therefore, throughout the entire contact interface, the two interacting objects continuously experience a spatially synchronized contact pressure distribution field, i.e., $p_{d,f}(x, y) = p_{f,s}(x, y) = p_n(x, y)$. Here, $p_{d,f}(x, y)$ and $p_{f,s}(x, y)$ represent the pressure exerted by the indenter and the layer, as well as the pressure between the layer and the substrate, respectively. Furthermore, because the contacting structures jointly support the applied normal load, the total relative normal displacement $u_{tot}(x, y)$ at any spatial coordinate within the plane is necessarily equal to the geometric summation of the localized normal displacements individually produced by each contacting entity [11],

$$u_{tot}(x, y) = u_d(x, y) + u_{f,s}(x, y). \quad (1)$$

Under the classical assumptions of isotropic linear elasticity and infinitesimal strains, the localized normal displacement at any arbitrary spatial coordinate on the boundary

can be evaluated via the spatial convolution of the interfacial contact pressure field and the corresponding surface Green's function [21, 25, 33, 49]. Accordingly, the normal elastic displacements individually sustained by the indenter and the layer-substrate are explicitly formulated via two-dimensional continuous integrals over the contact domain Ω ,

$$u_d(x, y) = \iint_{\Omega} p_n(\xi, \eta) G_d(x - \xi, y - \eta) d\xi d\eta, \quad (2)$$

$$u_{f,s}(x, y) = \iint_{\Omega} G_{f,s}(x - \xi, y - \eta) p_n(\xi, \eta) d\xi d\eta, \quad (3)$$

where $G_d(x, y)$ represents the conventional Boussinesq half space Green's function for the homogeneous indenter, and $G_{f,s}(x, y)$ denotes the fundamental layer specific Green's function that uniquely incorporates the structural constraints and thickness characteristics of the thin film and substrate assembly.

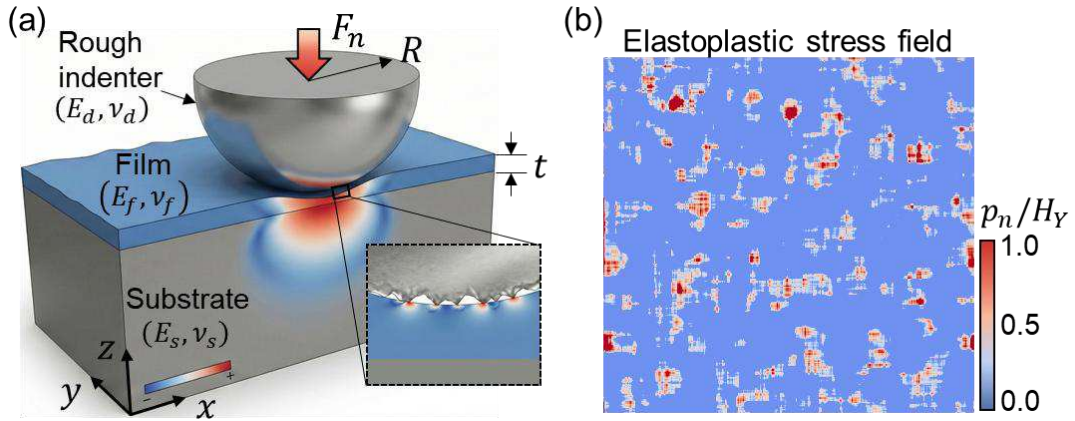


Fig. 1. Schematic representation of the multiscale contact modeling for a coated rough interface. (a) Macroscopic contact configuration of a spherical rough indenter with radius R is subjected to a normal force F_n and is loaded onto a film-substrate system with coating thickness t . The elastic properties of the indenter, coating film, and substrate are denoted by (E_d, ν_d) , (E_f, ν_f) , and (E_s, ν_s) , respectively. The inset provides a magnified view of local microcontacts, illustrating high-frequency stress concentrations at discrete microcontacts. (b) Two-dimensional full-field contact pressure distribution normalized by surface hardness (p_n/H_Y) across the rough interface.

Because direct spatial convolution presents substantial computational demands due to the non-local integration over the domain, the spatial governing equations are systematically projected into the spectral domain via a two-dimensional forward Fourier transform to efficiently circumvent this mathematical bottleneck. This integral transformation effectively converts the intricate spatial convolutions into straightforward algebraic multiplications within the wavenumber domain, i.e.,

$$\widehat{u_d}(q_x, q_y) = \widehat{G_d}(q_x, q_y) \widehat{p_n}(q_x, q_y) \quad \text{and} \quad \widehat{u_{f,s}}(q_x, q_y) = \widehat{G_{f,s}}(q_x, q_y) \widehat{p_n}(q_x, q_y).$$

Here, the circumflex symbol indicates the respective continuous Fourier spectrum, and $q = (q_x, q_y)$ defines the continuous wavenumber vector with a magnitude of $q =$

$\sqrt{q_x^2 + q_y^2}$. By substituting these spectral relations back into Eq. (1) of the frequency

domain, the total relative normal displacement spectrum is governed by a unified wavenumber dependent compliance relation,

$$\widehat{u}_{tot}(q_x, q_y) = [\widehat{G}_d(q_x, q_y) + \widehat{G}_{f,s}(q_x, q_y)]\widehat{p}_n(q_x, q_y), \quad (4)$$

where $\widehat{G}_d(q_x, q_y) = \frac{2(1-\nu_d^2)}{E_d} \frac{1}{q}$ and $\widehat{G}_{f,s}(q_x, q_y) = \kappa_{d,f} \widehat{G}_d \Psi(qt; \kappa_{f,s}; E_f, \nu_f, E_s, \nu_s)$.

Here, two non-dimensional parameters, $\kappa_{f,s} = \frac{E_f(1-\nu_s^2)}{E_s(1-\nu_f^2)}$ and $\kappa_{d,f} = \frac{E_d(1-\nu_f^2)}{E_f(1-\nu_d^2)}$, are introduced to characterize the elastic mismatch of the system. The $\kappa_{f,s}$ defines the relative stiffness ratio between the thin film layer and the underlying substrate, which dictates the attenuation degree of interfacial stress propagating from the film into the bulk substrate. The $\kappa_{d,f}$ represents the relative stiffness ratio between the indenter itself and the film, thereby governing the partitioning of displacement and deformation between the indenter and the layered structure. Additionally, $\widehat{G}_d$ corresponds to the homogeneous substrate spectrum, and the dimensionless function Ψ serves as the geometric correction factor for the finite thickness layer-based system. A detailed derivation is provided in Section S1 of the Supplementary Material.

In numerical implementations, this spectral relation establishes the mathematical foundation for the FFT-based BEM. By employing the discrete Fourier transform (DFT) to map the spatial fields, the non-local interfacial convolutions are resolved with a reduced computational complexity of $O(N \lg N)$ rather than $O(N^2)$, thereby enabling high throughput contact simulations across multiscale rough topographies [50]. Crucially, this analytical relation can be integrated into deep learning frameworks, serving as a physics-informed layer within FNO architectures. Incorporating this analytical relation into the learning framework provides a physics-based inductive bias by explicitly mapping the predicted pressure field to the corresponding elastic displacement response.

2.2 Numerical method

To numerically resolve the contact mechanics of rough interfaces, the continuous governing equations must be systematically translated into a discrete computational framework. The rectangular simulation domain is partitioned into a uniform mesh comprising $N_x \times N_y$ discrete grid cells, wherein the spatial dimensions of each individual cell are defined by Δx and Δy . Consequently, the continuous topography profiles, interfacial contact pressures, and localized normal displacements are represented as spatially discrete matrices. Because the DFT imposes periodic boundary conditions, direct execution of the spectral multiplication would introduce severe wrap around leakage and artificial boundary edge interference [51]. To suppress these periodic artifacts and simulate an isolated non periodic contact interface, a zero-padding technique is systematically implemented [34]. Additionally, a continuous to discrete shape factor correction is implemented within this extended wavenumber domain to eliminate high frequency aliasing noise. Because the contact pressure within each individual grid cell is treated as a piecewise constant distributed patch rather than an

idealized punctual delta force, the discrete transfer function spectrum incorporates a dual dimension sinc square filtering mechanism,

$$K_{FFT}(q_x, q_y) = [\widehat{G_d}(q_x, q_y) + \widehat{G_{l,s}}(q_x, q_y)] \left[\text{sinc}\left(\frac{q_x \Delta x}{2}\right) \text{sinc}\left(\frac{q_y \Delta y}{2}\right) \right]^2, \quad (5)$$

where $\text{sinc}(x) = \sin x / x$, and $K_{FFT}(q_x, q_y)$ represents the regularized discrete transfer function spectrum established across the $2N_x \times 2N_y$ computational domain. By taking advantage of this regularized spectrum, the localized elastic displacement field is updated via two-dimensional DFT operations. The non-local spatial matrix vector multiplication is bypassed by executing an element wise algebraic product in the frequency domain, followed by an inverse discrete transform,

$$u_{tot}(x, y) = \mathcal{F}^{-1} \left\{ K_{FFT}(q_x, q_y) \cdot \mathcal{F}\{p_{n,pad}(x, y)\} \right\}. \quad (6)$$

where $p_{n,pad}(x, y)$ denotes the zero padded pressure field, while $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent the forward and inverse two-dimensional FFT algorithms, respectively. This spectral acceleration diminishes the computational complexity to $O(N \lg N)$. The true contact status and the synchronized pressure distribution field are iteratively resolved by enforcing the KKT contact boundary conditions. These physical constraints constitute a nonlinear complementary problem, which mandates non-negative contact pressures within the real contact zone and zero pressures associated with positive local surface separations outside the contact area. To address the contact nonlinearity, a modified conjugate gradient (CG) method is employed to dynamically update the active contact set [52]. Furthermore, to capture elastoplastic behaviors, the local contact pressure is upper bounded by a regularized material hardness H_Y , wherein any localized pressure exceeding this physical limit triggers a local plastic flow correction ($p_n(x, y) \geq H_Y$), as shown in Fig. 1b. This elastoplastic solution procedure is executed through a multi-step incremental loading scheme to guarantee stable numerical convergence when simulating highly nonlinear contact interactions across multiscale topographies [53]. Further details are provided in Section S2 of the Supplementary Material.

Fig. 2 illustrates the contact response of a smooth spherical indenter on a coated substrate under varying normal loads and layer thickness ratios t/a_{EBM} . At a constant layer thickness ($t = 0.1$ mm), an increase in the normal load drives the contact pressure to deviate from classical Hertzian predictions (Fig. 2a). This behavior manifests a pronounced structural stiffening effect [26], as the expanding contact area enhances the deformation confinement imposed by the rigid substrate. In contrast, increasing the thickness ratio under a constant load mitigates substrate interference, causing the contact interface to undergo structural softening; consequently, the peak contact pressure diminishes and asymptotically converges to the semi-infinite Hertzian limit (Fig. 2b). Across all parameter combinations, the BEM predictions exhibit exceptional agreement with the reference LD numerical model [49], with relative errors remaining below 10^{-2} (Fig. S1). This agreement verifies the numerical implementation of the FFT-BEM reference solver for the considered smooth layered-contact cases.

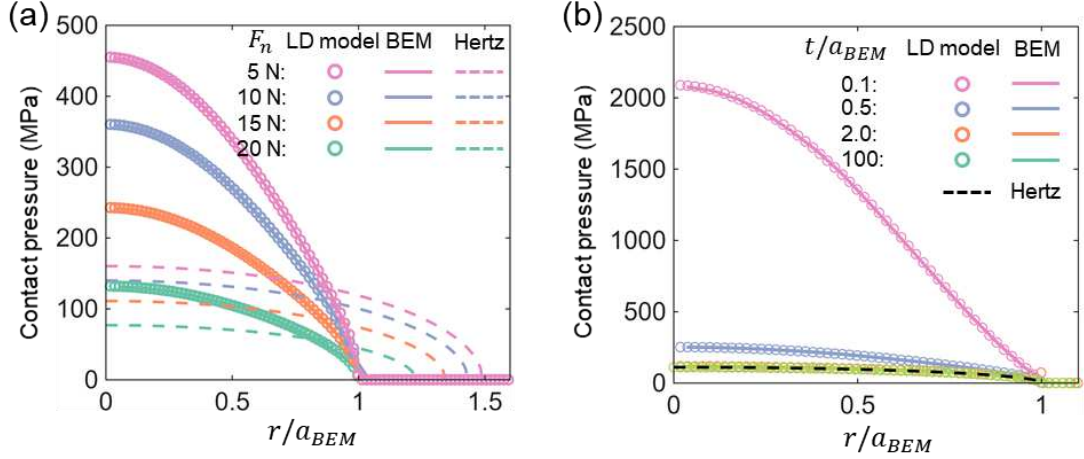


Fig. 2. Comparison of contact pressure distributions predicted by the LD model [49], BEM simulations, and classical Hertz theory for a rigid spherical indenter pressed into an elastic film on a nearly rigid substrate under purely elastic conditions. (a) Contact pressure profiles along the normalized radial position r/a_{BEM} under different normal loads ($F_n = 5, 10, 15$, and 20 N), where a_{BEM} denotes the contact radius computed from BEM simulations. (b) Influence of the normalized coating thickness t/a_{BEM} ($0.1, 0.5, 2.0$, and 100) on the contact pressure distribution, demonstrating convergence toward the Hertzian solution at large film thicknesses ($t/a_{BEM} = 100$). Here, the structural system consists of a rigid sphere with radius $R = 5$ mm, an elastic film ($E_d = 3$ GPa, $\nu_d = 0.38$, nominal thickness $t = 0.1$ mm), and a rigid substrate.

3. Physics-informed operator framework

To overcome the computational bottlenecks of iterative matrix solvers while maintaining mechanical fidelity, a operator framework for elastoplastic contact prediction is proposed. This framework integrates the flexible nonlinear feature mapping of deep neural networks with the rigorous kinematic compatibility conditions and boundary non-penetration constraints in contact mechanics. As illustrated in Fig. 3, the network architecture consists of four unified modules: a multidimensional spatial and parametric input layer, an attention-enhanced U-Net backbone equipped with FNO bottleneck, a differentiable physics-informed layer, and an elastoplastic field output layer. To construct a comprehensive training dataset that fully encompasses surface roughness characteristics, macro-to-micro geometric dimensions, and elastoplastic material properties, this study employs a Latin hypercube sampling (LHS) strategy governed by the maximin distance criterion. This space-filling strategy promotes broad and approximately uniform coverage of the high-dimensional parameter space [54]. As detailed in Table S1, thirteen design variables are systematically modulated to capture the highly coupled, multiscale mechanics of the layered contact system. To effectively accommodate variables spanning multiple orders of magnitude (such as elastic mismatch ratios and coating thicknesses), a logarithmic sampling scheme is applied to these multiscale parameters, whereas the remaining variables are sampled linearly. This hybrid sampling strategy ensures that the network encounters a balanced distribution across physical regimes, particularly near sensitive transition boundaries.

During the training period, the input to the network is partitioned into two primary classes to capture both localized spatial variations and global boundary constraints (Fig. 3i). The first class encompasses high dimensional interfacial field data, including the rough surface topography maps $z_{sphere}(x, y)$ and spatial coordinate grids. The BEM-derived pressure is used as supervisory targets in both optimization stages. The elastic-displacement field is additionally introduced during the physics-informed fine-tuning stage to supervise the displacement computed by the differentiable Green's-function-based FFT layer. These BEM-derived fields are used in the loss evaluation and are not supplied to the U-Net or FNO backbone as input features. To eliminate disparate numerical magnitudes across different physical domains and stabilize gradient dynamics during backpropagation, each spatial field is scaled by its corresponding analytical Hertzian solution. The second class consists of a multiscale scalar condition vector, designated as the feature-wise linear modulation (FiLM) vector $\mathbf{X}_{FiLM}$, which encodes the macroscopic loading boundaries and material constitutive relations of the layered system. To extract multiscale representations, the backbone network integrates convolutional layers, an adaptive spatial attention mechanism, and a FNO bottleneck (Fig. 3ii). The convolutional layers extract multiscale, high frequency geometric characteristics, such as localized roughness slopes and microstructural curvatures, using 3×3 pixel local receptive fields. Because rough contact interfaces exhibit spatial sparsity, where vast regions remain non-contacting, an adaptive attention gate (AG) is integrated. This module dynamically scales the feature maps with localized spatial weights [55], thereby guiding the network to prioritize active asperity regions where contact is highly probable. The attention-guided filtering process operates within the hierarchical downsampling path (Fig. S2), progressively compressing spatial dimensions from an initial resolution of 256×256 to intermediate levels of 128×128 and 64×64 . Simultaneously, the AG module intercepts skip connections at each corresponding resolution level, modulating the transmission of microscale geometric features prior to their fusion with the symmetric upsampling layers of the decoder. Upon reaching the bottleneck level at a resolution of 32×32 , the Fourier Neural Operator (FNO) maps these spatial features into the spectral domain to execute global convolutions. This facilitates long-range elastic coupling across the entire contact domain, effectively capturing far-field stress patterns that conventional localized convolutions fail to resolve [56]. To condition these structural features on macroscopic mechanical boundary conditions, scalar physical parameter vectors are fed into a multilayer perceptron (MLP). This MLP maps global parameters into channel-wise scaling and shifting coefficients, which adaptively modulate the decoded feature maps via FiLM layers, thereby enabling the feature representation to adapt to variations in loading, geometry, and material properties.

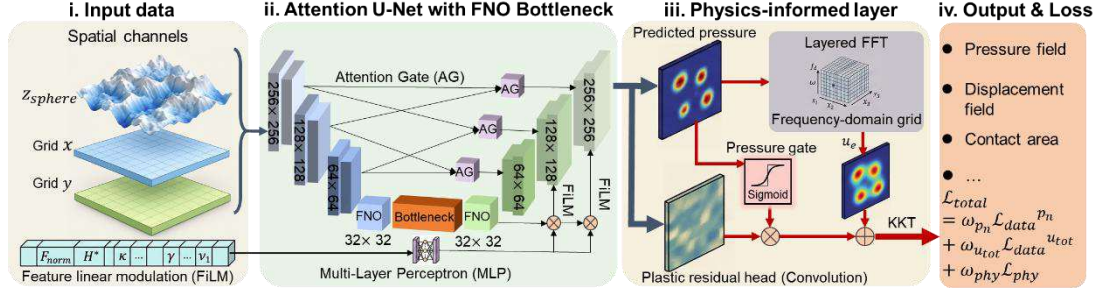


Fig. 3. Architecture of the proposed physics-informed operator framework for elastoplastic contact modeling of layered rough surfaces. (i) Input data: multi-channel spatial field inputs are processed alongside scalar material and loading parameters modulated via FiLM vector $\mathbf{X}_{\text{FiLM}}$, including rough surface topography $z_{\text{sphere}}(x, y)$ and spatial coordinate grids. (ii) Attention U-Net with FNO bottleneck: spatial attention gates (AG) are integrated into skip connections to capture localized high-frequency stress concentrations, while a FNO layer at the bottleneck captures long-range spatial dependencies. The MLPs adjust the feature representation of the decoder. (iii) Physics-informed layer: a differentiable layered FFT module in the frequency domain maps predicted surface pressure to elastic displacement fields, while the convolutional plastic residual head represents the plastic displacement, while the hardness penalty limits excessive contact pressure. (iv) Output & Loss: the framework outputs the total contact pressure, surface displacement, and true contact area through a composite loss function $\mathcal{L}_{\text{tot}}$ that incorporates KKT-based penalties for nonpenetration and contact complementarity.

Rather than relying on unconstrained data-driven regression to track interfacial contact deformation, the framework routes the predicted contact pressure $p_n(x, y)$ into a non-trainable physics-informed layer (Fig. 3iii). This layer encapsulates analytical layered frequency domain Green's functions and executes a parallelized FFT circular convolution to compute the continuous elastic displacement response $u_{\text{elastic}}(x, y)$ corresponding to the predicted pressure distribution, as shown in Eq. (6). Concurrently, the plastic deformation field $u_{\text{plastic}}(x, y)$ is generated by a residual neural network (ResNet) convolutional head. A differentiable pressure gate controlled by a sigmoid activation function modulates the plastic residual deformation to suppress plastic residual deformation in non-contacting regions. The total predicted interfacial displacement field $u_{\text{tot}}(x, y)$ is synthesized via a differentiable linear superposition of the continuous elastic response and the gated plastic residual, defined by $u_{\text{tot}}(x, y) = u_{\text{elastic}}(x, y) + u_{\text{plastic}}(x, y)$, which provides a physically interpretable decomposition of the total interfacial displacement. In particular, the frictionless contact conditions are incorporated into the training objective through differentiable KKT-based penalty terms. Specifically, the predicted contact pressure p_n and interfacial gap g are expected to satisfy pressure nonnegativity, nonpenetration, and contact complementarity, expressed as $p_n \geq 0$, $g \geq 0$, and $p_n \cdot g = 0$. The global network parameters are optimized through a composite objective function $\mathcal{L}_{\text{tot}}$, which simultaneously minimizes the data-driven error and the physical-based residuals,

$$\mathcal{L}_{\text{tot}} = \omega_{p_n} \mathcal{L}_{\text{data}}^{p_n} + \omega_{u_{\text{tot}}} \mathcal{L}_{\text{data}}^{u_{\text{tot}}} + \omega_{\text{phy}} \mathcal{L}_{\text{phy}}, \quad (7)$$

where ω_i represents the weight coefficient and the individual loss terms are mathematically formalized as,

$$\mathcal{L}_{data}^{p_n} = \frac{1}{N_x N_y} \sum_{x,y} [p_n^{pred}(x,y) - p_n^{true}(x,y)]^2, \quad (8)$$

$$\mathcal{L}_{data}^{u_{tot}} = \frac{1}{N_x N_y} \sum_{x,y} [u_{tot}^{pred}(x,y) - u_{tot}^{true}(x,y)]^2, \quad (9)$$

$$\begin{aligned} \mathcal{L}_{phy} = & \frac{1}{N_x N_y} \sum_{x,y} [\min(0, g_{pred}(x,y))]^2 + \\ & \lambda_{comp} \frac{1}{N_x N_y} \sum_{x,y} [p_n^{pred}(x,y) \cdot g_{pred}(x,y)]^2, \end{aligned} \quad (10)$$

Here, $g_{pred}(x,y) = z_{sphere}(x,y) + u_{tot}^{pred}(x,y) - \delta$ defines the predicted interfacial gap field for rough surface $z_{sphere}(x,y)$ under a given normal approach δ . The parameter λ_{comp} represents the penalty multiplier for KKT violations. To improve training stability, the model is optimized through a two-stage procedure. Adam is used for data-driven pretraining, while AdamW with weight decay, linear learning-rate warmup, and cosine decay is adopted for physics-informed fine-tuning. In both stages, the trainable network parameters are updated by minimizing the corresponding total loss function $\mathcal{L}_{tot}$. The first stage consists of a data-driven pretraining phase to capture the initial geometric contact boundaries and local pressure distributions from the baseline dataset. This initialization assists the network topology in establishing stable weight representations, which eliminates numerical divergence. The second stage follows with a physics constrained fine tuning phase. During this phase, the explicit differentiable physics-informed layer and the composite multitask loss function activate to govern the network optimization process. This sequential transition incorporates displacement consistency, KKT-related contact conditions, and the material-hardness limit during fine-tuning, while improving the stability of the optimization process. Further details are provided in Section S3 of the Supplementary Material.

Training of the physics-informed operator framework is evaluated on a hardware platform equipped with an NVIDIA Graphics Processing Unit (GeForce RTX 4070). Latin Hypercube Sampling generates 2100 unique rough surfaces for training and 300 independent surfaces for testing; each surface undergoes six sequential loading steps, yielding a total of 12600 training instances and 1800 validation test instances (Fig. S3). The contact loads within the dataset span nine orders of magnitude, with the test data extending one order of magnitude beyond the training set to validate model generalization capabilities, and the corresponding probability density function is displayed in Fig. S4. Optimization over this data space involves 600 epochs, with each epoch requiring four minutes of computation time under the given hardware configuration.

4. Results and Discussion

The convergence characteristics of the performance metrics during the model training process are presented in Fig. 4. Following the initial 200 epochs of pure data-driven training, the stabilization of the performance metrics indicates that the network establishes a primary mapping between localized asperities and contact zones. However, the purely data-driven stage reaches a performance plateau. As the physics constraints are employed, a systematic regularization is enforced across the network parameter space. This enables subsequent training to further reduce local variance and global integration errors, while improving consistency with the governing contact constraints. Across the two training stages, the mean intersection over union (IoU), serving as the primary metric for contact area prediction accuracy, monotonically increases from 0.75 during the data-driven phase and stabilizes at approximately 0.83 (Fig. 4(a)). At the same time, the mean absolute error (MAE) governing the total normal force F_n , localized pressure distributions p_n , and maximum peak pressures $p_{n,max}$ decrease below 8% (Fig. 4(b-d)), confirming model convergence toward a stable optimum consistent with mechanics principles.

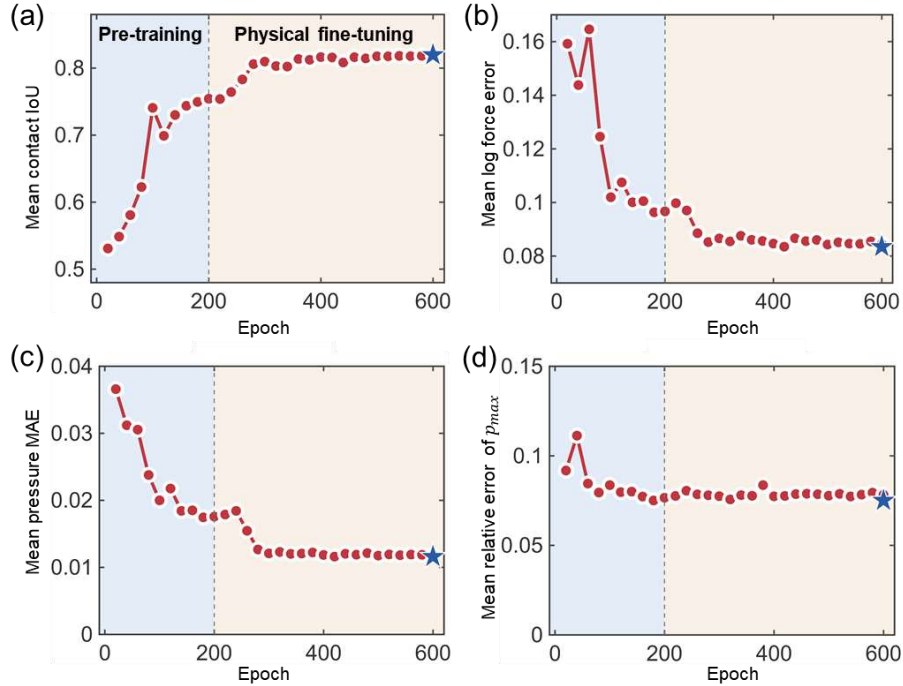


Fig. 4. Convergence dynamics and performance evolution of the model across the two-stage training scheme. (a) Mean contact Intersection over Union (IoU), quantifying the spatial prediction accuracy of active contact areas. (b) Mean logarithmic force error, evaluating global force equilibrium compliance. (c) Mean pressure MAE, measuring full-field pressure field discrepancy. (d) Mean relative error of peak contact pressure $p_{n,max}$, assessing local stress concentration estimation accuracy. The light blue and light orange shaded regions highlight the data-driven pre-training phase (epochs 0 to 200) and the physics-informed fine-tuning phase (epochs 200 to 600), respectively. Blue star symbols indicate the optimal metrics achieved at epoch 600.

To evaluate the predictive accuracy of the trained model, a comprehensive assessment is conducted between the true values and the predicted values for the contact area ratio A_r/A_{Hertz} , total normal force F_n , peak pressure $p_{n,max}$, and mean total displacement $\overline{u_{tot}}$, as shown in Fig. 5. Notably, although the test dataset spans a broader distribution range than the training dataset (Fig. 4), all four primary evaluation metrics exhibit a coefficient of determination R^2 exceeding 0.96. Specifically, the hybrid architecture demonstrates marked performance improvements over conventional unconstrained data-driven baselines in the contact area prediction [42, 45], retaining high fidelity even under complex multi-topography, multi-material, and highly nonlinear layered contact behavior. Predictive performance for geometric boundary features achieves a coefficient of determination R^2 of 0.986 (Fig. 5(a)). Importantly, within the extensive load range spanning nine orders of magnitude from 10^{-6} N to 10^3 N, the predicted total normal force exhibits a strong correlation with the true data, with R^2 reaching 0.997 (Fig. 5(b)). This wide-spectrum loading regime intensifies interaction effects among neighboring microcontacts, driving a continuous transition from isolated local contact mechanics to globally coupled interfacial behavior [57]. At lower load levels, discrete microcontact deform independently under local pressure peaks. As the normal load increases across multiple orders of magnitude, adjacent microcontacts expand, coalesce, and induce far-field elastic interactions throughout the layered substrate [58].

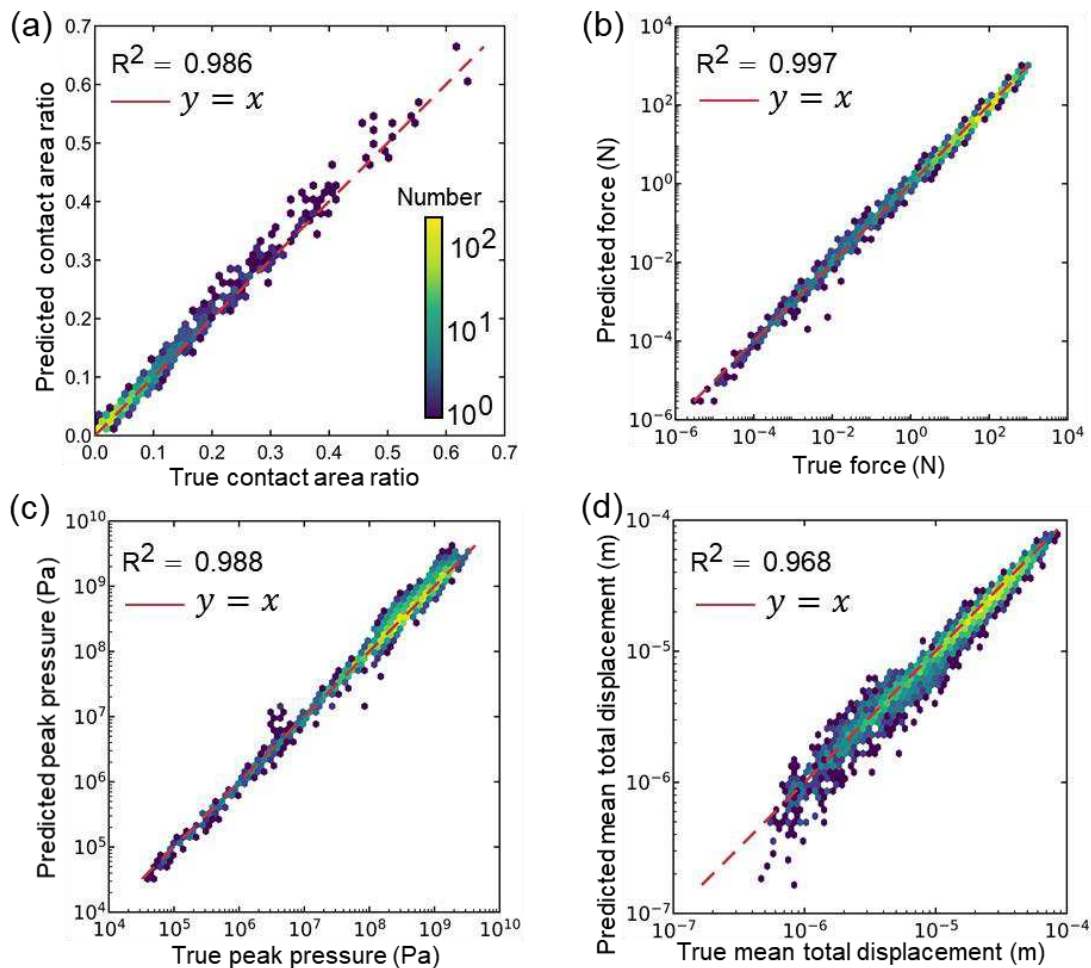


Fig. 5. Quantitative evaluation of model predictions against ground-truth values for key contact parameters across the test dataset. (a) Predicted versus true contact area ratio, A_r/A_{Hertz} , where A_{Hertz} represents the corresponding contact area for a smooth indenter while keeping all other parameters identical. (b) Predicted versus true total normal force F_n . (c) Predicted versus true peak contact pressure $p_{n,max}$. (d) Predicted versus true mean total surface displacement $\overline{u_{tot}}$. The red dashed line ($y = x$) denotes perfect agreement, and the color bar indicates local data point density.

By integrating a residual U-Net architecture augmented with an FNO bottleneck and spatial attention mechanisms, the hybrid framework simultaneously resolves localized stress concentrations governing peak pressures alongside long-range continuous displacement fields driven by multi-asperity interactions. Evaluation of localized stress fields in Fig. 5(c) demonstrates that the predicted peak pressure $p_{n,max}$ maintains excellent agreement across a broad stress domain reaching 10^{10} Pa, yielding a coefficient of determination R^2 of 0.988. Concurrently, the predicted mean total displacement $\overline{u_{tot}}$ exhibits high accuracy spanning spatial scales from the micrometer domain up to the near millimeter regime, achieving a coefficient of determination R^2 of 0.968 (Fig. 5(d)). The reliable prediction of these extreme local pressures highlights a distinct computational advantage over BEM [59], delivering high-precision outputs in millisecond execution times without suffering numerical instabilities.

To further assess the contribution of each architectural component, the proposed hybrid model was compared with a U-Net baseline, an FNO baseline, and a hybrid variant without spatial attention. All neural models were trained and evaluated using identical datasets, normalization procedures, physical inputs, optimization schedules, and test conditions. Table 1 summarizes their predictive accuracy and computational efficiency relative to the FFT-BEM reference. The U-Net baseline exhibits the largest errors across most evaluation metrics, indicating that localized convolutional features alone are insufficient to represent the multiscale and nonlocal behavior of layered rough contact. The FNO baseline improves the prediction of peak pressure and global displacement but remains less accurate in contact-zone identification and load consistency, suggesting that spectral representations alone cannot fully resolve localized contact boundaries. Removing the attention gates preserves the displacement accuracy but degrades the prediction of the real contact region and total normal force, confirming that spatial attention primarily improves the localization of sparse contact zones and the consistency of the reconstructed pressure field. Overall, the complete hybrid architecture provides the most balanced accuracy across contact area, load, peak pressure, and displacement. Importantly, our hybrid model can achieve millisecond-level inference and high fidelity, and is up to 4 orders of magnitude faster than iterative BEM solvers.

Table 1. Quantitative performance comparison and ablation evaluation for interfacial contact predictions across different model architectures. The proposed framework is compared with the FNO and U-Net baselines and a variant without the attention-gate mechanism. Predictive performance is evaluated using the relative errors in F_n , $p_{n,max}$, and $\overline{u_{tot}}$, together with the IoU

of the predicted contact area, all referenced to the BEM solutions. Computational efficiency is quantified by the average execution time per sample (milliseconds).

Model	$\lg(F_n)$	IoU	$\lg(p_{n,max})$	$\overline{u_{tot}}$	Time (ms)
BEM	\	\	\	\	$\sim 10^5$
This work	5.62%	0.83	5.89%	2.88%	~ 6.2
FNO	9.11%	0.79	5.93%	6.62%	~ 2.7
U-Net	16.12%	0.74	8.74%	8.08%	~ 3.6
Without AG	6.56%	0.79	6.23%	2.33%	~ 6.2

To establish parameter robustness and generalization boundaries for the hybrid architecture, predictive accuracy is evaluated across the comprehensive parameter space summarized in Table S1, which governs the mechanical response of layered contact systems. Specifically, the four independent scalar variables defined within the FiLM vector serve as generalized evaluation parameters: the normalized coating thickness t/a_{Hertz} , the normalized indent radius R/R_{rms} , and two stiffness ratios $\kappa_{f,s}$ and $\kappa_{d,f}$. Each scaling parameter spans multiple orders of magnitude to encompass a broad spectrum of macrogeometric variations and material pairings, ranging from compliant thin films under sharp indenters to rigid structural coatings under macroscopically flat contacts. As depicted in the box plots in Fig. 6, the model restricts the MAE of predicted contact pressure to a regime near 10^{-2} or below across all evaluation conditions. As the normalized layer thickness increases from 10^{-2} to 10^2 , the contact configuration approaches a semi-infinite continuum system, driving the median MAE for thick-film structures down to a magnitude near 10^{-4} (Fig. 6a). This reduction confirms the capability of the hybrid network to capture semi-infinite contact behavior, which exhibits a simplified mechanical response relative to layered systems. In thick-layered media, stress field reflections at the substrate interface are strongly attenuated, causing the frequency-domain transfer function to converge toward a homogeneous Green's function. In addition, the model accommodates multiscale geometric scaling, maintaining bounded error levels across four orders of magnitude of the normalized curvature radius R/R_{rms} from 10^3 to 10^6 (Fig. 6b). Although the MAE values exhibit a slight upward trend as R/R_{rms} increases, this behavior stems from contact area expansion toward computational domain boundaries rather than algorithmic failure. As the contact area approaches these domain edges, spatial truncation introduces localized numerical discrepancies [50].

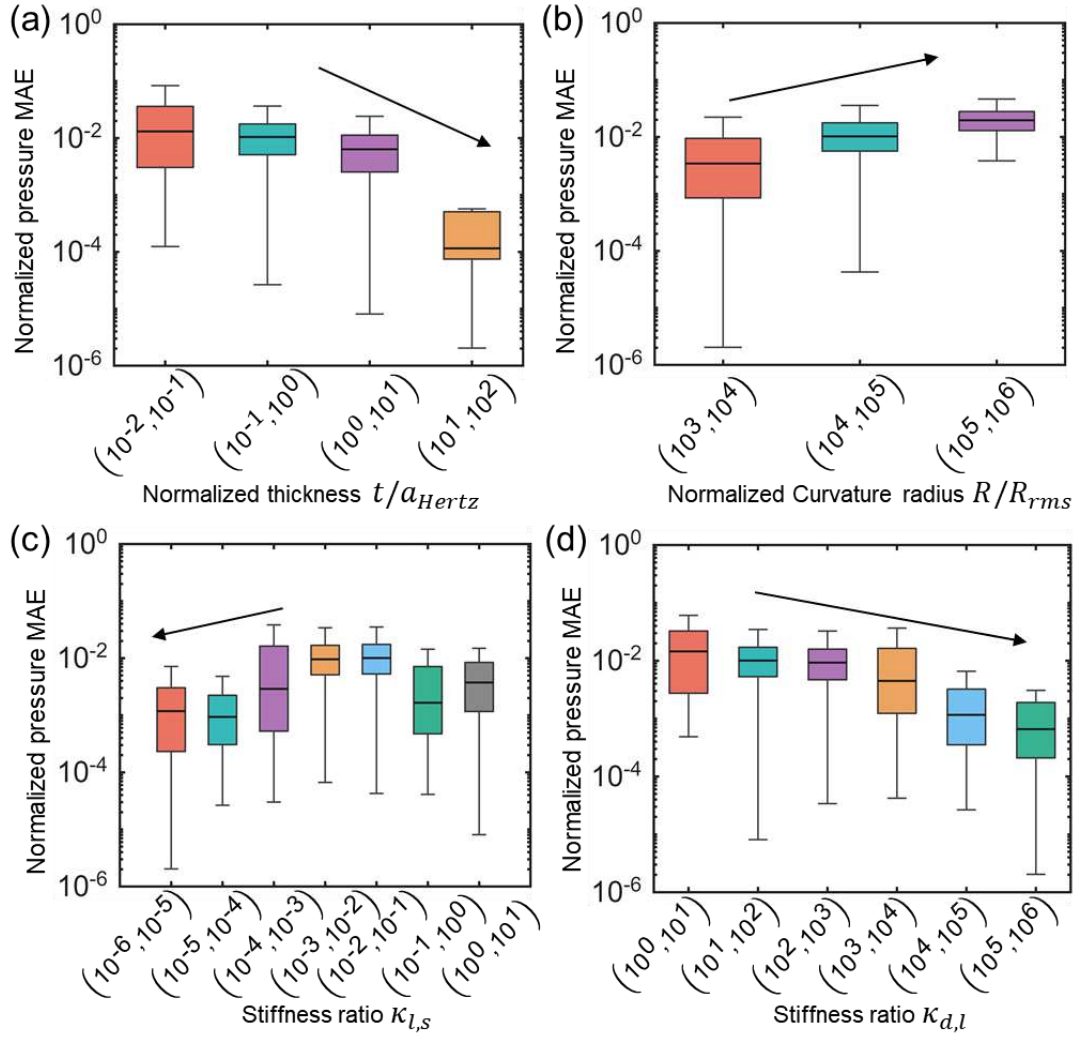


Fig. 6. Sensitivity analysis of contact pressure MAE normalized onto the range of 0 to 1 based on its minimum and maximum values across key dimensionless geometric and material parameter spaces. (a) Distribution of normalized pressure MAE across intervals of normalized coating thickness t/a_{Hertz} , showing a marked error reduction with increasing relative film thickness. Here, a_{Hertz} represents the corresponding contact radius for a smooth indenter while keeping all other parameters identical. (b) Influence of normalized curvature radius R/R_{rms} on prediction error. (c) Error variation across different regimes of the layer-to-substrate stiffness ratio $\kappa_{f,s}$. (d) Impact of the indenter-to-layer stiffness ratio $\kappa_{d,f}$ on the pressure error distribution. The box plots statistically summarize the error distributions across binned parameter ranges, and black arrows highlight the primary trends of error evolution.

We further investigated the stiffness ratios governing interfacial elastic mismatches that arise from material property variations across the indenter, coating layer, and substrate. As shown in Figs. 6(c) and Fig.6(d), the model maintains MAE values near or below 10^{-2} across seven orders of magnitude of $\kappa_{l,s}$ and six orders of magnitude of $\kappa_{d,l}$. Lower error occurs when the coating exhibits greater compliance relative to the substrate ($\kappa_{f,s} \rightarrow 0$) or indenter ($\kappa_{d,f} \rightarrow \infty$). This trend indicates that compliant surface layers absorb localized contact deformation, thereby smoothing

sharp stress gradients across interfacial zones. Softer layers effectively act as mechanical filters that suppress high-frequency spatial variations in the contact pressure field [26]. Because spectral convolution within the FNO reconstructs smooth field representations with higher numerical precision, reduced stress gradients diminish network prediction variance. As a result, the hybrid architecture demonstrates superior predictive stability across compliant surface regimes where mechanical compliance governs contact mechanics. In rough contact simulations, such extreme variations in geometric curvature, layer dimensions, and material stiffness introduce severe ill-conditioning into the governing boundary element matrices, which standard solvers must address through intensive multigrid refinement or specialized preconditioners to prevent convergence failure [60]. By contrast, the ability of the architecture to maintain uniform precision across these extreme regimes underscores its intrinsic capability to resolve and generalize across complex physical systems.

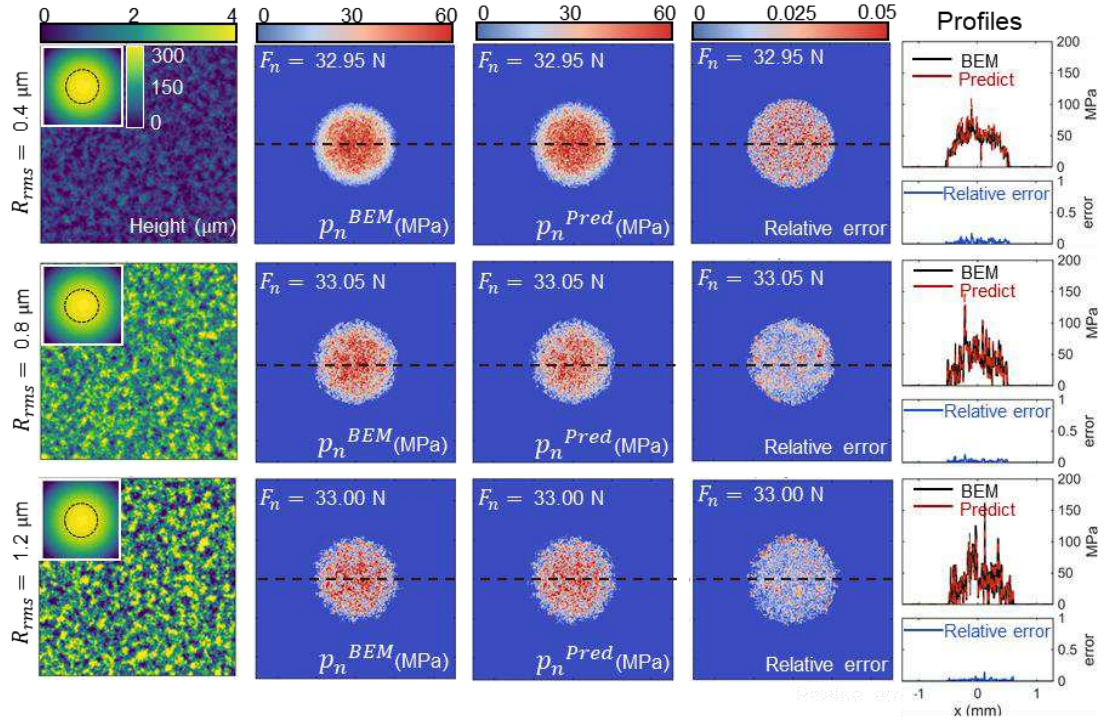


Fig. 7. Full-field contact pressure predictions compared with reference BEM simulations for a coated rough interface under varying surface topographies. The contact system comprises a spherical indenter ($R = 5$ mm, $E_d = 280$ GPa, $\nu_d = 0.33$), a compliant elastic film ($t = 0.52$ mm, $E_f = 0.3$ GPa, $\nu_f = 0.41$), and a stiff substrate ($E_s = 178$ GPa, $\nu_s = 0.43$) subjected to a normal force $F_n \approx 33$ N across a spatial domain of length 2.56 mm. Rows 1 to 3 represent three distinct surface roughness realizations under identical film thickness. Left column: surface height maps with macro-indenter profiles shown in insets. Second and third columns: full-field contact pressure distributions calculated by BEM (p_n^{BEM}) and predicted by the model (p_n^{Pred}), respectively. Fourth column: full-field relative error distributions. Rightmost column: cross-sectional pressure profiles extracted along the central horizontal dashed line, accompanied by local relative error distributions underneath.

To evaluate the spatial fidelity and local feature resolution of the model beyond scalar global metrics, we examine the predicted full-field stress distributions under representative contact configurations. Fig. 7 illustrates contact pressure fields across layered interfaces under equivalent normal loads ($F_n \approx 33$ N) for spherical contact configurations featuring identical macro curvature ($R = **$) alongside distinct surface roughness topographies (Fig. 7, Col. 1). Predicted normal pressure distributions $p_n^{Pred}(x, y)$ generated by the hybrid network (Fig. 7, Col. 3) display remarkable spatial congruence with reference boundary element method solutions $p_n^{BEM}(x, y)$ (Fig. 7, Col. 2). Full-field relative error distributions confirm that predictive deviations remain bounded within 5% across the entire contact domain (Fig. 7, Col. 4). The hybrid network resolves localized stress peaks and steep pressure gradients at discrete microcontact without introducing artificial spatial smoothing. Quantitative verification along central cross sectional dashed lines demonstrates that predicted pressure profiles overlap ground-truth BEM responses across the contact length, with local relative errors remaining below 5% (Fig. 7, Col. 5). In contrast to standard unconstrained deep learning models that often fail at high-frequency spatial variations and underpredict localized peak stress magnitudes [42], the integration of Fourier-domain spectral convolutions with physical boundary constraints preserves fine-scale surface features.

In addition to contact pressure, the predicted displacement fields $u_{tot}(x, y)$ confirm that the framework captures long-range elastic coupling across the layered medium (Fig. 8). Because the elastic displacement is obtained by convolving the predicted pressure with the layered Green's function (Eq. (6)), localized pressure variations are integrated into a smooth and spatially extended response. As shown in Fig. 8, the predicted displacement fields $u_{tot}^{Pred}(x, y)$ mirror reference BEM profiles $u_{tot}^{BEM}(x, y)$ across all rough topographies. The corresponding error maps (Fig. 8, Col. 4) nevertheless exhibit a roughness-dependent spatial distribution. For rougher surfaces, the discrepancies are mainly concentrated near the contact periphery and computational boundaries, where finite-domain truncation and the periodic representation used in the spectral operations may introduce boundary-related errors. For smoother surfaces, the broader contact region strengthens nonlocal elastic coupling, allowing small pressure prediction errors to accumulate over a larger spatial region through the Green's-function convolution. This phenomenon becomes more pronounced when the same rough surface is subjected to higher normal loads, as increased contact pressure enhances global elastic coupling and consequently elevates displacement errors in the central region (Fig. S6), whereas the stress field error remains relatively stable (Fig. S5). Despite these localized discrepancies, the cross-sectional profiles remain in close agreement with the BEM results, with the displacement relative error remaining below 5% across the computational domain (Fig. 8, Col. 5).

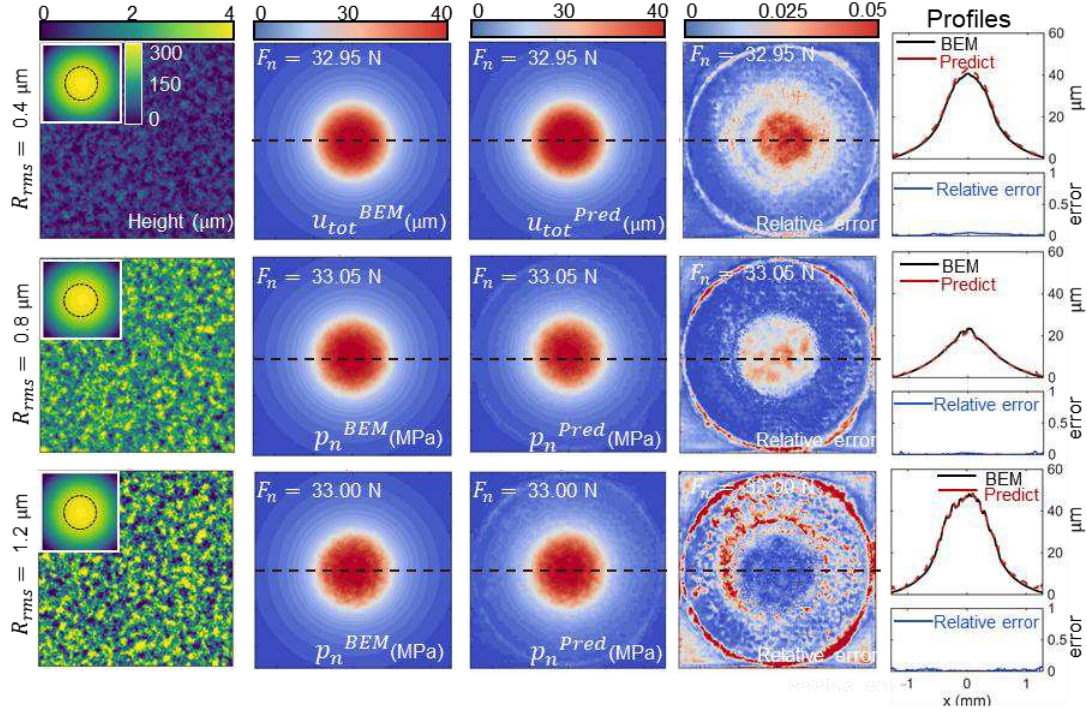


Fig. 8. Full-field predictions of total surface displacement compared with reference BEM simulations for a coated rough interface under varying surface topographies. Detailed material properties, geometric dimensions, and loading parameters are identical to those specified in Fig. 7.

To evaluate the zero-shot generalization capabilities of the hybrid network on unseen surface topographies, we subjected the model to non-fractal, deterministic geometries excluded from the training dataset, specifically unidirectional groove channels, orthogonal cross-grids, and periodic micro-asperity arrays (Figs. 9 and 10). These structured topographies represent a stringent out-of-distribution benchmark, as their abrupt geometric boundaries contrast sharply with the stochastic, self-affine nature of the fractal surfaces used during training. For contact pressure predictions on these engineered surfaces, the hybrid model accurately resolves the highly anisotropic stress transmission pathways and the formation of regular microcontact clusters (Fig. 9, Col. 3). Relative error distributions and centerline profiles demonstrate that high predictive fidelity is maintained across diverse geometries, successfully preserving sharp pressure peaks up to 100 MPa over the regular micro-asperity arrays (Fig. 9, Col. 5).

Similarly, the predicted displacement fields faithfully capture the long-range elastic interactions mediated through the substrate across the anisotropic channels, showing strong spatial agreement with reference BEM profiles (Fig. 10). Although predictive discrepancies are slightly elevated compared to those for fractal topographies (primarily due to the strong high-frequency spectral components induced by sharp geometric steps), the framework still demonstrates robust fidelity in capturing complex contact phenomena. The consistent performance across both structured and anisotropic topographies demonstrates promising out-of-distribution generalization to the tested structured topographies. This performance stems from the synergistic integration of the physics-informed loss function with the explicit embedding of analytical layered

Green's functions as a structural layer. These results indicate that the proposed framework can retain predictive accuracy for the tested structured topographies without parameter recalibration, despite their substantial geometric differences from the stochastic rough surfaces used during training. Furthermore, unlike traditional boundary element method (BEM) algorithms that rely on expensive iterative active-set and conjugate-gradient updates in the FFT-BEM solver, the proposed framework reconstructs full-field pressure and displacement distributions within sub-second inference times, as summarized in Table 1.

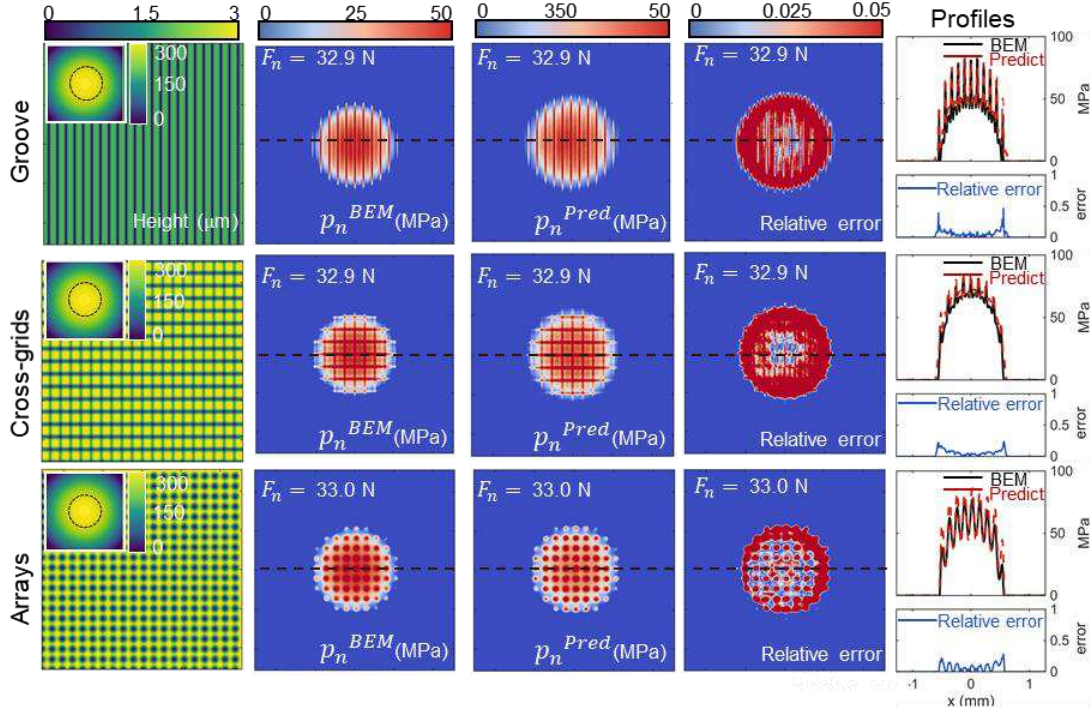


Fig. 9. Full-field contact pressure predictions compared with reference BEM simulations for a coated interface featuring deterministic micro-textured topographies. The system configuration, material parameters, and film thickness ($t = 0.52$ mm) are identical to those in Fig. 8, operating under dimensionless parameters $\kappa_{f,s} = 1.65 \times 10^{-3}$ and $\kappa_{d,s} = 870$ with normal load $F_n \approx 33$ N. Rows 1 to 3 correspond to three distinct regular topographies: unidirectional groove channels, orthogonal cross-grids, and periodic micro-asperity arrays, respectively. Left column: surface height maps with macro-indenter profiles shown in insets. Second and third columns: full-field contact pressure distributions calculated by BEM (p_n^{BEM}) and predicted by the model (p_n^{Pred}), respectively. Fourth column: full-field relative error distributions. Rightmost column: cross-sectional pressure profiles extracted along the central horizontal dashed line, accompanied by local relative error distributions underneath.

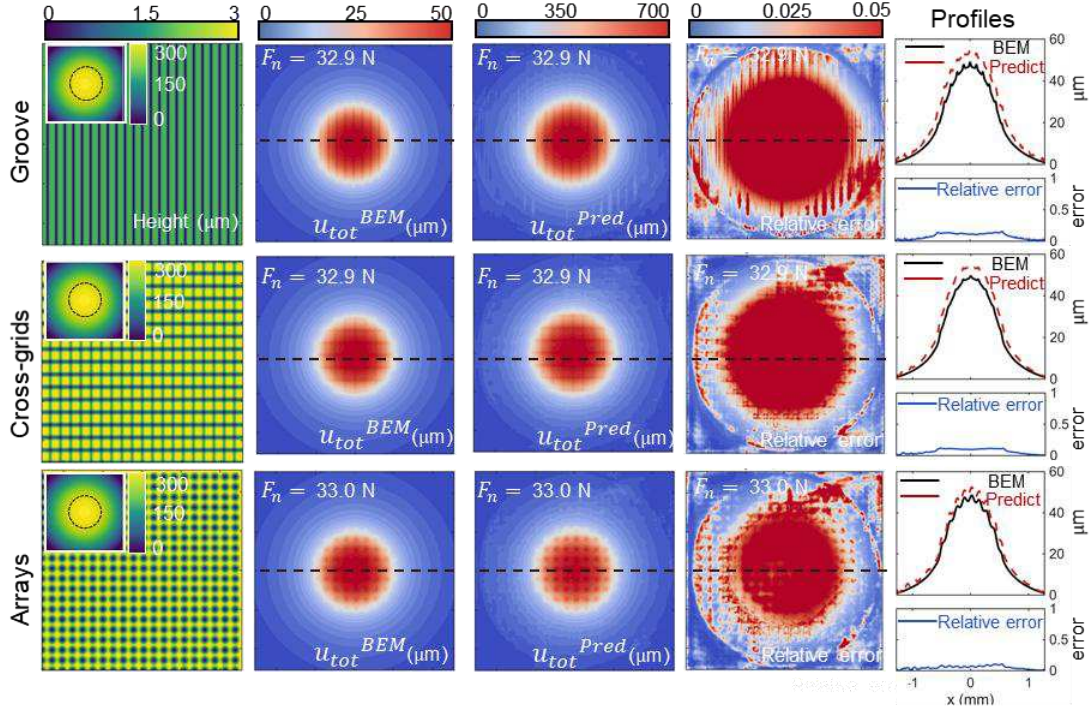


Fig. 10. Full-field predictions of total surface displacement compared with reference BEM simulations for a coated interface featuring deterministic micro-textured topographies. The system setup and parameter choices are identical to those described in Fig. 9.

While full-field distributions demonstrate excellent spatial alignment with numerical benchmarks, practical contact mechanics applications rely on macroscopic contact metrics. To evaluate whether the network preserves these macroscopic physical laws across the entire test dataset, we extract the total normal load $F_n = \iint_{\Omega} p_n(x, y) dx dy$ and real contact area $A_r = \iint_{\Omega} H(p_n(x, y)) dx dy$ by performing spatial integrations over the reconstructed local pressure fields $p_n(x, y)$. Here, $H(\cdot)$ denotes the Heaviside step function enforcing the active contact condition ($p_n(x, y) > 0$). Fig. 11 illustrates the non-dimensional scaling relationship between the normalized real contact area and the normalized normal load across the complete validation dataset,

$$\frac{A_r}{A_{Hertz}} = \left(\frac{F_n}{A_{Hertz} E_{d,l}^* R_{slope}} \right)^{\gamma}, \quad (11)$$

where $E_{d,f}^* = 1 / ((1 - \nu_d^2) / E_d + (1 - \nu_f^2) / E_f)$ is defined as the equivalent

elastic modulus of the indenter and film, and $R_{slope} = \sqrt{(\nabla z_{rough})^2}$ represents the RMS slope of the rough surface. Spanning nearly five orders of magnitude, the predicted median response (red curve) displays near-perfect alignment with the reference BEM calculations (black curve), exhibiting a clear log-linear trend. A power-law fit across all test samples (black dashed line) yields a characteristic scaling exponent of $\gamma = 0.91$. Notably, this power-law exponent is slightly lower than the

linear relation ($\gamma = 1$) predicted by classical asymptotic theories (such as Persson's contact mechanics theory [61, 62] and the BGT model [63]), which assume an ideal, unconstrained semi-infinite elastic half-space at small contact fractions [64]. This sub-linear scaling deviation ($\gamma < 1$) aligns well with previous studies on constrained or finite-thickness rough surface contact [65-67]. Physically, this phenomenon is attributed to two key mechanisms. Firstly, the finite thickness of the elastic layer introduces a structural stiffening effect under long-wavelength elastic deformations, which restricts further contact area growth as load increases. Secondly, as the load rises, adjacent micro-asperities undergo coalescence and nonlinear elastic interaction, reducing the incremental contact area generation per unit load compared to isolated asperity assumptions.

Additionally, color-mapping the test dataset by normalized thickness t/a_{Hertz} (Fig. 11(a)) and surface roughness R_{rms}/a_{Hertz} (Fig. 11(b)) clearly reveals distinct parameter-dependent spatial clustering on the scaling plot. Test data points corresponding to larger layer thicknesses systematically exhibit higher contact fractions (shifting upward) relative to thinner films, whereas topographies with elevated roughness yield smaller real contact areas (concentrating toward the lower-left regime). This parametric distribution reflects the interplay between structural compliance and microstructural steepness. On one hand, increasing t/a_{Hertz} mitigates rigid substrate confinement, raising the overall elastic compliance of the layer and thereby facilitating broader, long-wavelength deformation to generate larger contact areas under identical loading. On the other hand, higher R_{rms}/a_{Hertz} introduces taller and sharper micro-asperities that require steeper local stress gradients to flatten. Consequently, the normal force is borne by fewer, highly stressed microcontacts, which restricts further area growth and shifts the contact response toward the lower-left regime [26]. By applying a first-order linear approximation to the force-area relationship in Eq. (11), we define a characteristic shape factor, $\kappa = A_r E_{d,f}^* R_{slope} / F_n$, which represents the proportional slope of the curve [64]. Evaluating κ across various layer thicknesses reveals that it depends strongly on t/a_{Hertz} , asymptotically approaching the theoretical value predicted by classical semi-infinite contact theory as thickness increases (Fig. 11(a), inset). In contrast, after normalization by the RMS slope, the fitted shape factor κ exhibits limited sensitivity to RMS amplitude R_{rms}/a_{Hertz} over the tested range.

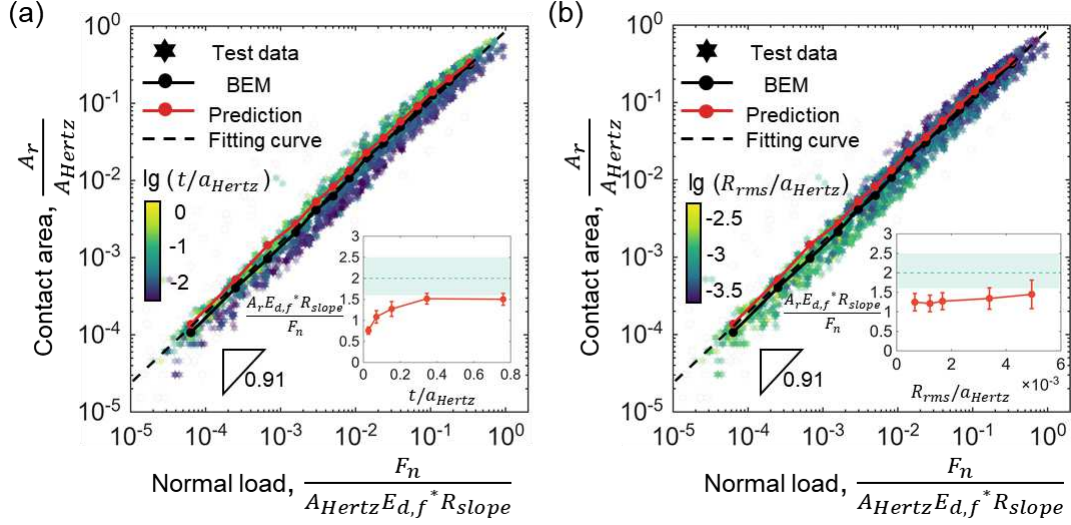


Fig. 11. Scaling behavior of normalized contact area as a function of normalized normal load under varying coating thickness and surface roughness. (a) Effect of normalized film thickness t/a_{Hertz} , color-coded on a logarithmic scale $\lg(t/a_{\text{Hertz}})$, with the inset depicting the variation of the dimensionless parameter $A_r E_{d,f} * R_{\text{slope}} / F_n$ against t/a_{Hertz} . (b) Influence of normalized RMS roughness $R_{\text{rms}}/a_{\text{Hertz}}$, color-coded by $\lg(R_{\text{rms}}/a_{\text{Hertz}})$, with the inset showing the evolution of $A_r E_{d,f} * R_{\text{slope}} / F_n$ across roughness levels. Scatter points represent test data across parameter spaces, while solid black and red curves denote reference BEM simulations and model predictions, respectively. The dashed line indicates the global power-law fitting curve with a scaling exponent of 0.91.

5. Conclusion

This study develops a physics-informed hybrid framework to address the trade-off between computational efficiency and physical fidelity in the elastoplastic contact analysis of coated and layered rough surfaces. The formulation separates the unresolved nonlinear contact-state mapping from the analytically known elastic compliance operator of the layered medium. An attention-enhanced U-Net with a FNO bottleneck learns the contact pressure and plastic-displacement fields, while a differentiable Green's-function-based FFT layer explicitly evaluates the nonlocal elastic response. Dimensionless geometric, loading, and material parameters are incorporated through multiscale feature modulation, and contact admissibility is promoted through KKT-related penalties, nonnegative pressure parameterization, and a material-hardness constraint. The main conclusions are summarized as follows:

- 1) The framework resolves both localized asperity-scale pressure concentrations and long-range elastic interactions. Ablation studies confirm the complementary roles of local convolutional features and global spectral representations in reconstructing complex rough-contact fields.
- 2) The explicit Green's-function elastic operator, together with KKT-related penalties and the material-hardness constraint, incorporates contact-mechanics information into the learned predictions and guides the reconstructed fields toward mechanically admissible solutions.

- 3) Beyond reproducing the reference FFT-BEM fields, the framework maintains high predictive fidelity for the tested unseen structured topographies and loads beyond the primary training range, while recovering the characteristic sublinear scaling of real contact area with normal load.

These results suggest that explicitly retaining the known elastic operator while restricting learning to the nonlinear contact state provides an effective balance between computational efficiency, physical interpretability, and generalization across the tested parameter regimes. The model provides a clear route for extending the framework to more general interface problems. Future developments may incorporate tangential traction and frictional slip, Green's function kernels for multilayer material systems, and time-dependent operators for viscoelastic contact. Overall, the proposed formulation provides an efficient and physically interpretable surrogate for multiscale layered rough-contact analysis and a basis for future extensions to frictional, multilayered, and time-dependent interface problems.

References

- [1] Okokpujie IP, Tartibu LK, Musa-Basheer HO, Adeoye A. Effect of coatings on mechanical, corrosion and tribological properties of industrial materials: a comprehensive review. *Journal of Bio- and Tribo-Corrosion*. 2024;10:2.
- [2] Ghosh G, Sidpara A, Bandyopadhyay P. Understanding the role of surface roughness on the tribological performance and corrosion resistance of WC-Co coating. *Surface and Coatings Technology*. 2019;378:125080.
- [3] Ren Y, Zhang L, Xie G, Li Z, Chen H, Gong H, et al. A review on tribology of polymer composite coatings. *Friction*. 2021;9:429-70.
- [4] Liu Y, Yu S, Shi Q, Ge X, Wang W. Multilayer coatings for tribology: A mini review. *Nanomaterials*. 2022;12:1388.
- [5] Guo L, He W, Chen W, Xue Z, He J, Guo Y, et al. Progress on high-temperature protective coatings for aero-engines. *Surface science and technology*. 2023;1:6.
- [6] Vasiljević S, Glišović J, Stojanović B, Vencl A. Review of the coatings used for brake discs regarding their wear resistance and environmental effect. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*. 2022;236:1932-49.
- [7] Sathyamoorthy G, Vijay R, Lenin Singaravelu D. Brake friction composite materials: a review on classifications and influences of friction materials in braking performance with characterizations. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*. 2022;236:1674-706.
- [8] Liu P, Huang L, Zheng C, Bao Y, Gao D, Zhou G. Spin coating in semiconductor lithography: Advances in modeling and future prospects. *Microelectron Eng*. 2025;298:112326.
- [9] Liu X, Li H, Tao M, Yu Y, Zhu Z, Wu D, et al. Organic flexible electronics for innovative applications in electronic skin. *Advanced Materials Technologies*. 2025;10:2400661.
- [10] Dalcin RL, de Menezes VM, da Silva Rocha A, de Moraes SP, Affonso LB, Cousseau T, et al. Correlation between roughness, film thickness, and friction coefficient with pitting wear resistance of spur gears. *The International Journal of Advanced Manufacturing Technology*. 2023;129:5473-92.
- [11] Johnson KL. *Contact mechanics*: Cambridge university press; 1987.
- [12] Dong Q, Li Y, Wei J, Lu F. Layered structures with rough surfaces and interfaces at contact loading. *International Journal of Mechanical Sciences*. 2020;178:105611.
- [13] Long H, Liu Y, Wei Y. Debonding characterization of stiff film/compliant substrate systems based on the bilinear cohesive zone model. *Engineering Fracture Mechanics*. 2022;265:108363.
- [14] Zhang H, Zhou Z, Chudnovsky A, Pham H. Time-dependent buckling delamination of thin plastic films and their conformability: Observations and modeling. *International Journal of Engineering Science*. 2020;150:103258.
- [15] Fu P, Zhao J, Zhang X, Miao H, Wen Z, Wang P, et al. Elasto-plastic partial slip contact modeling of graded layers. *International Journal of Mechanical Sciences*. 2024;264:108823.
- [16] Kim J-H, Kil H-J, Lee S, Park J, Park J-W. Interfacial delamination at multilayer thin films in semiconductor devices. *ACS omega*. 2022;7:25219-28.
- [17] Cheng A, Sun L, Menga N, Yang W, Zhang X. Interfacial performance evolution of ceramics-in-polymer composite electrolyte in solid-state lithium metal batteries. *International Journal of Engineering Science*. 2024;204:104137.
- [18] Juliá JM, Rodríguez-Tembleque L. Wear and subsurface stress evolution in tractive rolling contact. *International Journal of Mechanical Sciences*. 2025;294:110195.

- [19] Burmister DM. The general theory of stresses and displacements in layered systems. I. Journal of applied physics. 1945;16:89-94.
- [20] King R, O'sullivan T. Sliding contact stresses in a two-dimensional layered elastic half-space. International Journal of Solids and Structures. 1987;23:581-97.
- [21] O'sullivan T, King R. Sliding contact stress field due to a spherical indenter on a layered elastic half-space. 1988.
- [22] Plumet S, Dubourg M-C. A 3-d model for a multilayered body loaded normally and tangentially against a rigid body: application to specific coatings. 1998.
- [23] Huajian G, Cheng-Hsin C, Jin L. Elastic contact versus indentation modeling of multi-layered materials. International journal of Solids and Structures. 1992;29:2471-92.
- [24] Greenwood JA, Williamson JP. Contact of nominally flat surfaces. Proceedings of the royal society of London Series A Mathematical and physical sciences. 1966;295:300-19.
- [25] Persson B. Contact mechanics for layered materials with randomly rough surfaces. Journal of Physics: Condensed Matter. 2012;24:095008.
- [26] Persson B. Contact mechanics for layered materials: Rubber film on hard substrate. The Journal of Chemical Physics. 2025;162.
- [27] Xu Y, Li X, Chen Q, Zhou Y. Persson's theory of purely normal elastic rough surface contact: A tutorial based on stochastic process theory. International Journal of Solids and Structures. 2024;290:112684.
- [28] Müser MH, Dapp WB, Bugnicourt R, Sainsot P, Lesaffre N, Lubrecht TA, et al. Meeting the contact-mechanics challenge. Tribology Letters. 2017;65:1-18.
- [29] Sun L, Wang QJ, Zhao N, Zhang M. Discrete convolution and FFT modified with double influence-coefficient superpositions (DCSS-FFT) for contact of nominally flat heterogeneous materials involving elastoplasticity. Computational Mechanics. 2021;67:989-1007.
- [30] Li Q, Popov VL. Boundary element method for normal non-adhesive and adhesive contacts of power-law graded elastic materials. Computational Mechanics. 2018;61:319-29.
- [31] Bonari J, Marulli MR, Hagmeyer N, Mayr M, Popp A, Paggi M. A multi-scale FEM-BEM formulation for contact mechanics between rough surfaces. Computational Mechanics. 2020;65:731-49.
- [32] Nogi T, Kato T. Influence of a hard surface layer on the limit of elastic contact—Part II: Analysis using a modified GW model. J Trib. 2002;124:785-93.
- [33] Bhushan B, Peng W. Contact mechanics of multilayered rough surfaces. Appl Mech Rev. 2002;55:435-80.
- [34] Liu S, Wang Q, Liu G. A versatile method of discrete convolution and FFT (DC-FFT) for contact analyses. Wear. 2000;243:101-11.
- [35] Liu S, Wang Q. Studying contact stress fields caused by surface tractions with a discrete convolution and fast Fourier transform algorithm. J Trib. 2002;124:36-45.
- [36] Frérot L, Anciaux G, Molinari J-F. Crack nucleation in the adhesive wear of an elastic-plastic half-space. Journal of the Mechanics and Physics of Solids. 2020;145:104100.
- [37] Jacq C, Ne'lias D, Lormand G, Girodin D. Development of a three-dimensional semi-analytical elastic-plastic contact code. J Trib. 2002;124:653-67.
- [38] Falope FO, Lanzoni L, Radi E. 2D Green's functions for an elastic layer on a rigid support loaded by an internal point force. International Journal of Engineering Science. 2022;173:103652.

- [39] Yin N, Yang P, Liu S, Pan S, Zhang Z. AI for tribology: Present and future. *Friction*. 2024;12:1060-97.
- [40] Nath D, Ankit, Neog DR, Gautam SS. Application of Machine Learning and Deep Learning in Finite Element Analysis: A Comprehensive Review: D. Nath et al. *Archives of computational methods in engineering*. 2024;31:2945-84.
- [41] Pastewka L, Vakis AI, Eder SJ, Aghababaei R, Almqvist A, Carbone G, et al. Modeling in tribology: Recent advances, applications, and open questions. *Tribology International*. 2026;218:111326.
- [42] Jang I, Jang YH. PREDICTION OF CONTACT DISTRIBUTION ON ROUGH SURFACES USING DEEP LEARNING ALGORITHMS. *Facta Universitatis, Series: Mechanical Engineering*. 2025;23:757-85.
- [43] Bonagiri LKS, Wang Z, Zhou S, Zhang Y. Precise surface profiling at the nanoscale enabled by deep learning. *Nano letters*. 2024;24:2589-95.
- [44] Sahin T, Bonari J, Brandstaeter S, Popp A. Data-driven surrogate modeling techniques to predict the effective contact area of rough surface contact problems. *arXiv preprint arXiv:250417354*. 2025.
- [45] Suman S, Prajapati DK. Predictive modeling of real contact area on rough surfaces using deep artificial neural network. *Journal of Tribology*. 2025;147:111501.
- [46] Zhou Y, Song H. Data-driven enhanced rough contact mechanics: PINN estimation of gap distribution across length scales for partial contacts. *Tribology International*. 2025:111100.
- [47] Zhou Y, Song H. Physics-informed neural network approach to randomly rough surface contact mechanics. *Tribology Letters*. 2025;73:88.
- [48] Chen L, Zhu G, Chen Y, Lim C, Chen W. AI-enabled high-throughput characterization of material properties based on indentation response. *MechanoEngineering*. 2026;1.
- [49] Li H, Dempsey J. Axisymmetric contact of an elastic layer underlain by rigid base. *International journal for numerical methods in engineering*. 1990;29:57-72.
- [50] Liu S, Wang Q. Studying Contact Stress Fields Caused by Surface Traction With a Discrete Convolution and Fast Fourier Transform Algorithm. *Journal of Tribology*. 2001;124:36-45.
- [51] Wang W, Wang H, Liu Y, Hu Y, Zhu D. A comparative study of the methods for calculation of surface elastic deformation. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*. 2003;217:145-54.
- [52] Wang Z-J, Wang W-Z, Wang H, Zhu D, Hu Y-Z. Partial slip contact analysis on three-dimensional elastic layered half space. 2010.
- [53] Pérez-Ràfols F, Liang X-M, Ciavarella M. General results for loading and unloading rough elastoplastic interfaces. *European Journal of Mechanics-A/Solids*. 2025:105908.
- [54] van der Mensbrugghe D. A Latin hypercube sampling utility: With an application to an integrated assessment model. *Journal of Global Economic Analysis*. 2023;8.
- [55] Xiang Z, Zhou X, Bai P, Wei P, Zhao H, Sun Z, et al. Physics-Guided Deep Learning for Fast Prediction of Mechanics-Derived Frictional Engagement Descriptors on Rough Surfaces. *Tribology International*. 2026:112411.
- [56] Li Z, Huang DZ, Liu B, Anandkumar A. Fourier neural operator with learned deformations for pdes on general geometries. *Journal of Machine Learning Research*. 2023;24:1-26.
- [57] Afferrante L, Bottiglione F, Putignano C, Persson B, Carbone G. Elastic contact mechanics of randomly rough surfaces: an assessment of advanced asperity models and Persson's theory. *Tribology Letters*. 2018;66:75.

- [58] Taylor RI. Rough Surface Contact Modelling—A Review. *Lubricants*. 2022;10.
- [59] Wang D, De Boer G, Neville A, Ghanbarzadeh A. A new numerical model for investigating the effect of surface roughness on the stick and slip of contacting surfaces with identical materials. *Tribology International*. 2021;159:106947.
- [60] Wang QJ, Sun L, Zhang X, Liu S, Zhu D. FFT-Based Methods for Computational Contact Mechanics. *Frontiers in Mechanical Engineering*. 2020;6.
- [61] Persson BN. Theory of rubber friction and contact mechanics. *The Journal of Chemical Physics*. 2001;115:3840-61.
- [62] Persson BN. Contact mechanics for randomly rough surfaces. *Surface science reports*. 2006;61:201-27.
- [63] Bush A, Gibson R, Thomas T. The elastic contact of a rough surface. *Wear*. 1975;35:87-111.
- [64] Meng Q, Song H, Zhou Y, Liu X, Shi X. Unifying linear proportionality between real contact area and load in rough surface contact. *Journal of the Mechanics and Physics of Solids*. 2025;196:105975.
- [65] Yastrebov VA, Anciaux G, Molinari J-F. From infinitesimal to full contact between rough surfaces: Evolution of the contact area. *International Journal of Solids and Structures*. 2015;52:83-102.
- [66] Putignano C, Carbone G, Dini D. Mechanics of rough contacts in elastic and viscoelastic thin layers. *International Journal of Solids and Structures*. 2015;69:507-17.
- [67] Pastewka L, Prodanov N, Lorenz B, Müser MH, Robbins MO, Persson BN. Finite-size scaling in the interfacial stiffness of rough elastic contacts. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*. 2013;87:062809.

CRedit authorship contribution statement

Minglong Xu: Conceptualization, Supervision, Project administration. **Chongpu Zhai:** Conceptualization, Methodology, Supervision, Funding acquisition, Writing - review & editing. **Shaoqi Huang:** Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing - original draft. **Junpeng Chen:** Methodology, Software, Investigation, Writing - review & editing. **Haoyu Gu:** Investigation, Data curation, Validation, Writing - review & editing. **Shuwen Zhang:** Investigation, Data curation, Visualization, Writing - review & editing. **Xiaoshen Zhang:** Validation, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (NSFC) with grant No. of 12202342.

Data availability

The data and trained model supporting the findings of this study are publicly available in Figshare at <https://doi.org/10.6084/m9.figshare.33145292>.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial.docx](#)