

PHYSICS-INFORMED NEURAL NETWORK SURROGATE FOR THERMO-METALLURGICAL-MECHANICAL QUENCHING MODEL

Felde I. , Annus N. , Wei S. 

Abstract A physics-guided conditional neural surrogate is developed for the transient thermal, microstructural, and stress response of a quenched plain-carbon-steel cylinder. The Robin boundary condition is governed by a heat-transfer-coefficient (HTC) field $h(T, z)$ whose boiling peaks vary along the axial coordinate. A family of 10,000 admissible HTC functions was generated for training, with 500 validation and 500 test functions. The conditional physics-informed neural network (PINN) receives normalized space-time coordinates and a 40-component HTC descriptor; separate fully connected heads predict temperature, phase fractions, and mechanical quantities. The thermal and metallurgical fields are coupled through transformation kinetics and latent heat, whereas the mechanical branch is driven by the thermal and phase histories without mechanical feedback to those fields. A conservative competing-hazard phase integrator refines the diffusional and martensitic outputs without FEM labels. The general six-case screening gives a mean temperature RMSE of 56.17 °C and identifies the early boiling transient as the principal limitation. For the out-of-family HTC03 case, FEM-assisted adaptation and targeted early-time correction reduce the mean full-field temperature RMSE over twelve instants from 6.77 to 3.50 °C. A final boundary specialist, assessed at 45 held-out times, gives a mean full-field RMSE of 2.20 °C and a maximum sampled nodal error of 16.62 °C. The FEM-calibrated mechanical comparison gives von Mises RMSE values of 8.09 MPa at 10 s and 1.73–2.76 MPa at 30–600 s. The results distinguish general conditional screening from case-specific numerical refinement.

Keywords: physics-informed neural network; machine learning; surrogate modeling; finite element method; finite element analysis; steel quenching; heat treatment; C45 steel; heat-transfer coefficient; phase transformation; kinetics; thermo-metallurgical-mechanical coupling; simulation.

AMS Mathematics Subject Classification: 65M60, 65M22, 68T07, 80M10.

DOI: 10.32523/2306-6172-2026-14-3-46-63

1 Introduction

Quenching of steel is a nonlinear multiphysics problem in which heat extraction, temperature-dependent properties, phase transformation, latent heat, transformation strain, plasticity, and mechanical equilibrium interact [1]. For a cylindrical specimen immersed in a liquid quenchant, the boundary heat flux is particularly difficult to model because film boiling, vapour-film collapse, nucleate boiling, and convection generate a strongly temperature-dependent heat-transfer coefficient [2]. Spatially non-uniform wetting further requires a field $h(T, z)$ rather than a single curve [3]. Standard treatments of quenching heat transfer and process modelling are given in [12]. Finite-element methods (FEM) remain the standard numerical tool for heat-treatment calculations, but repeated nonlinear transient solutions are costly in

*Corresponding Author.

optimisation, uncertainty quantification, inverse heat conduction, and digital-twin applications. This computational burden is especially important in two-dimensional inverse heat-conduction problems, for which massively parallel numerical strategies have previously been developed [19]. Physics-informed neural networks (PINNs) embed differential-equation and boundary residuals into the training objective and therefore provide a differentiable surrogate after an offline optimisation stage [3]. Their use is not automatically accurate: stiff kinetics, nonlinear Robin boundaries, broad time scales, and sharp early cooling fronts can produce spectral bias and imbalanced losses.

The central idea of a PINN is to represent an unknown field, such as temperature, displacement, or phase fraction, by a differentiable neural network and to train its parameters so that the predicted field simultaneously satisfies observations and the governing equations. Automatic differentiation evaluates the derivatives appearing in the residuals, while initial, boundary, conservation, and constitutive constraints enter the objective function or are built directly into the output transformation. In contrast to a conventional discretisation, the trained function is continuous in its coordinates and can also be conditioned on material or process parameters. PINNs are therefore attractive for forward prediction, inverse parameter identification, data assimilation, and reusable surrogate modelling. Their practical performance nevertheless depends strongly on scaling, sampling, loss balancing, network representation, and the stiffness of the underlying equations; satisfying a small composite loss is not by itself proof of an accurate physical solution [4]. In materials science, PINN development has progressed from relatively simple elasticity and diffusion benchmarks toward nonlinear and history-dependent material behaviour. Physics-informed formulations have been used for forward and inverse solid-mechanics problems, including identification of elastic or constitutive parameters from sparse full-field information [5]. Subsequent studies incorporated rate-independent finite-strain plasticity and internal-variable evolution into the network formulation [6]. Variational and energy-based PINNs have also been coupled with phase-field descriptions of fracture, where the neural approximation must resolve both displacement and a localised damage or phase field [7]. These developments are directly relevant to heat treatment because phase fractions, thermal strain, transformation strain, and plastic flow are path-dependent internal variables rather than independent regression outputs.

Thermal materials processing has become another important application area. Physics-informed models have been introduced for three-dimensional temperature prediction in laser metal deposition without labelled temperature fields [8], and related work increasingly targets additive manufacturing, welding, solidification, thermal management, and inverse boundary identification. Artificial-intelligence and biomimetic approaches have also been reviewed specifically in the context of heat-treatment-process support [18]. Beyond thermal processing, physics-informed deep learning has also been applied to improving lithium-ion battery performance [20]. The appeal is clear: high-fidelity FEM solutions may be accurate but expensive when the heat-source parameters, geometry, or boundary heat-transfer laws must be varied repeatedly. At the same time, materials-processing problems expose several weaknesses of baseline PINNs, including moving or sharply localised thermal fronts, strongly temperature-dependent coefficients, multiscale time behaviour, latent heat, and discontinuous changes in boundary regime. The present quenching problem combines many of these difficulties and adds axial variation of the nonlinear boiling HTC.

This work reports a transparent computational study rather than an experimentally certified heat-treatment model. Its contributions are: (i) a constrained family of 10,000 position- and temperature-dependent HTC functions; (ii) a 43-input conditional PINN with thermal, phase, and mechanical heads; (iii) phase-only refinement from conservative kinetics trajec-

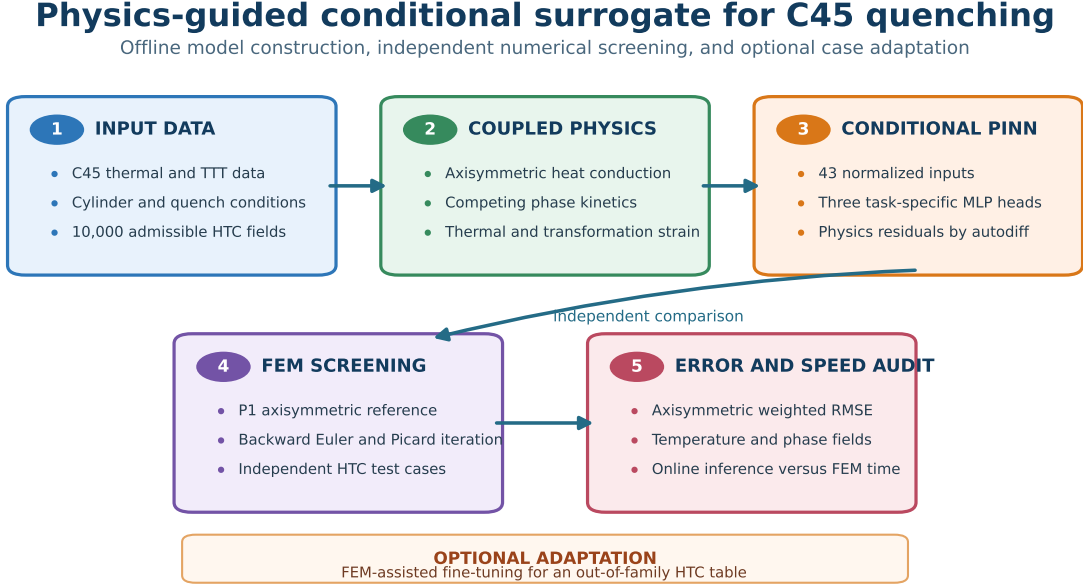


Figure 1: Computational concept. The general conditional PINN is trained from physics and phase-kinetics trajectories, screened against independent FEM solutions, and may subsequently be adapted to an out-of-family HTC table using FEM snapshots.

tories without FEM labels; (iv) independent numerical screening on development and fresh synthetic HTC fields; (v) case-specific adaptation to an out-of-family tabulated HTC with a separate finer FEM check; and (vi) targeted early-transient refinement followed by a tensor-level FEM comparison of radial, hoop, axial, shear, and von Mises stresses. Because the mechanical head is subsequently calibrated with FEM stresses, its reported errors are interpreted as a calibrated numerical comparison rather than as an independent generalisation test.

2 Physical and mathematical model

2.1 Geometry, initial state, and material data

The body is an axisymmetric C45 steel cylinder with radius $R = 10$ mm and height $H = 200$ mm. Initially, $T(\mathbf{x}, 0) = T_0 = 850^\circ\text{C}$ and the material is taken as austenitic [13, 17]. The quenchant temperature is $T_w = 20^\circ\text{C}$ and the simulated interval is $0 \leq t \leq 600$ s. The meridian domain is

$$\Omega = \{(r, z) : 0 \leq r \leq R, 0 \leq z \leq H\}.$$

The supplied material file defines the nominal chemical composition and the transformation and thermal data in Table 1. Linear interpolation is used for the tabulated conductivity and heat capacity. These sparse values are appropriate for a methodological study, but a production model would require heat-specific thermophysical and dilatometric data.

Table 1: Principal C45 input data used in the numerical model. Chemical composition is in mass percent.

Quantity	Value or tabulation
Composition	C 0.450, Si 0.250, Mn 0.650, P 0.018, S 0.018 Cr 0.190, Ni 0.130, Mo 0.042, V 0.030, Cu 0.150
A_1, A_3, B_s, M_s	727, 783, 628, 325 °C
Density ρ	7800 kg m ⁻³
Conductivity k	44 at 20 °C; 45 at 1000 °C [W m ⁻¹ K ⁻¹]
Specific heat c_p	450 at 20 °C; 456 at 1000 °C [J kg ⁻¹ K ⁻¹]
TTT information	start and finish times at 22 temperatures, 200–725 °C
Isohardness	22 values, 250–560 HV

2.2 Thermal field and boundary heat transfer

The axisymmetric heat equation is

$$\rho c_p(T, \mathbf{f}) \frac{\partial T}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} \left(r k(T, \mathbf{f}) \frac{\partial T}{\partial r} \right) + \frac{\partial}{\partial z} \left(k(T, \mathbf{f}) \frac{\partial T}{\partial z} \right) + \rho \sum_j L_j \dot{f}_j, \quad (1)$$

where $\mathbf{f}$ denotes phase fractions and L_j are effective latent heats. Symmetry gives $\partial T / \partial r = 0$ at $r = 0$. The water-exposed boundary comprises the lateral surface and both end faces. On these surfaces,

$$-k \nabla T \cdot \mathbf{n} = h(T, z)(T - T_w). \quad (2)$$

The HTC family is represented at axial anchors $z = 0, 80, 120$, and 200 mm. At each anchor, five temperature knots and five HTC ordinates define a shape-preserving cubic curve. Linear interpolation of the descriptors along z produces $h(T, z)$. Every generated curve contains one global maximum at 350–600 °C, with amplitude 3000–10 000 W m⁻² K⁻¹, and a smaller local maximum 30–60 °C above a Leidenfrost temperature sampled between 600 and 800 °C. The local peak amplitude is 1100–2400 W m⁻² K⁻¹. The axial construction favours the strongest global peak near the bottom, a lower peak near the top, and a minimum region around the specimen middle; a cap of 5000 W m⁻² K⁻¹ is retained at $z = 100$ mm. Figure 2 shows representative generated members.

2.3 Phase transformation and mechanical coupling

The phase vector contains austenite f_a , a lumped diffusional product f_d , and martensite f_m , subject to

$$f_a + f_d + f_m = 1, \quad f_j \geq 0. \quad (3)$$

Diffusional transformation rates are constructed from temperature-interpolated Johnson–Mehl–Avrami–Kolmogorov parameters fitted to the supplied TTT start and finish times [10, 11]. The treatment of continuously varying temperature is consistent with the non-linear extension of the Scheil–Cahn additivity rule proposed for non-isothermal diffusional transformations [21]. Related multiphase formulations have been applied directly to computer simulation of steel quenching [22], while generalized isothermal time–temperature parameters have also been extended to non-isothermal thermal histories and rapid tempering [23]. Martensitic evolution uses the Koistinen–Marburger relation [9, 14]. To distinguish the conditional Koistinen–Marburger fraction from the actual phase fraction in the competing

Representative members of the 10,000-function training family

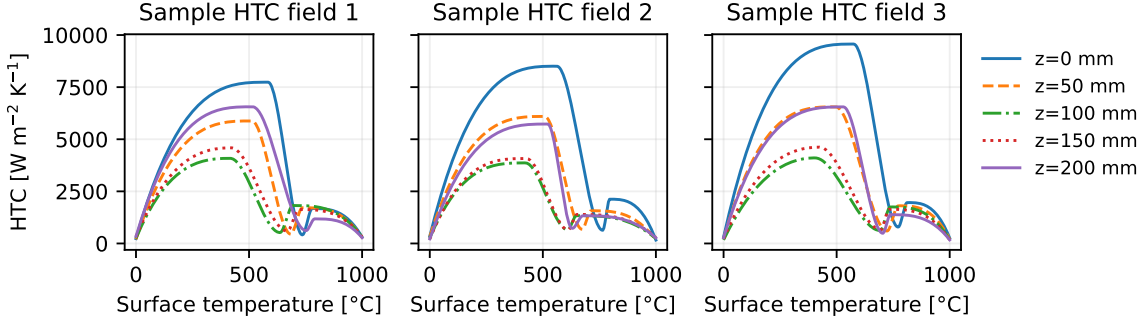


Figure 2: Representative HTC curves from the generated family at five axial positions. Each temperature slice has one global boiling maximum and one smaller high-temperature local maximum associated with the Leidenfrost regime.

transformation system, define

$$q_m(t) = 1 - \exp[-k_m \langle M_s - T_{\min}(t) \rangle_+], \quad T_{\min}(t) = \min_{0 \leq s \leq t} T(s), \quad (4)$$

where $\langle x \rangle_+ = \max(x, 0)$ and $k_m = 0.011 \text{ K}^{-1}$. The history variable $T_{\min}$ makes the martensitic driving fraction irreversible under reheating. During numerical refinement, increments derived from q_m and the diffusional hazard consume only the remaining austenite, so that simultaneous transformation modes preserve non-negativity and phase closure. Thus q_m is not identified directly with the total phase fraction f_m .

The coupled architecture also represents quasi-static axisymmetric equilibrium and a thermo-transformation strain split,

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^e + \boldsymbol{\varepsilon}^{th} + \boldsymbol{\varepsilon}^{tr} + \boldsymbol{\varepsilon}^p, \quad \nabla \cdot \boldsymbol{\sigma} = \mathbf{0}. \quad (5)$$

The mechanical branch uses temperature-dependent elasticity and thermal expansion. It also includes effective transformation strains and a regularised Perzyna evolution law for J_2 plasticity. The four independent axisymmetric stress components σ_{rr} , $\sigma_{\theta\theta}$, σ_{zz} , and τ_{rz} are predicted directly. The von Mises equivalent stress is calculated as

$$\sigma_{vm} = \left[\frac{(\sigma_{rr} - \sigma_{\theta\theta})^2 + (\sigma_{\theta\theta} - \sigma_{zz})^2 + (\sigma_{zz} - \sigma_{rr})^2}{2} + 3\tau_{rz}^2 \right]^{1/2}. \quad (6)$$

The tensor-level comparison therefore tests the individual components rather than only a scalar equivalent stress. Coupling to the mechanical branch is one-way in the present model: temperature and phase evolution drive the constitutive response, while stress-assisted kinetics and mechanical-dissipation feedback are neglected.

3 Numerical methods

3.1 Conditional PINN

The network input contains normalized (r, z, t) and a 40-component HTC descriptor. The latter stores, at four axial anchors, five normalized temperature knots and five normalized

HTC ordinates. Thus each of the thermal, phase, and mechanical heads receives 43 inputs. All heads are fully connected tanh multilayer perceptrons with four hidden layers of width 64. Their outputs are one thermal scalar, three phase logits, and eleven mechanical variables, respectively, for a total of 46,863 trainable parameters. Early cooling is emphasized through the non-uniform and partly log-uniform time sampling listed in Table 2; no additional time coordinate is appended to the 43-component input. Initial envelopes enforce $T = T_0$, $(f_a, f_d, f_m) = (1, 0, 0)$, and zero mechanical outputs exactly at $t = 0$; softmax phase outputs enforce positivity and closure.

The reported training objective is organised into the following weighted blocks:

$$\mathcal{L} = w_h \mathcal{L}_{heat} + w_b \mathcal{L}_{Robin} + w_e \mathcal{L}_{energy} + w_f \mathcal{L}_{phase} + w_p \mathcal{L}_{plastic} + \mathcal{L}_{sym} + \mathcal{L}_{IC}. \quad (7)$$

Automatic differentiation supplies spatial and temporal derivatives. The heat, Robin, phase, plastic, symmetry, and initial-condition blocks include the associated differential, constitutive, or boundary residuals after nondimensionalization. Complete residual definitions, constitutive tables, and normalization constants are implementation-defining information and should accompany the archived computational release. The HTC implementation used by the boundary residual was verified against SciPy’s PCHIP interpolation to a maximum discrepancy below $10^{-6} \text{ W m}^{-2} \text{ K}^{-1}$ during general training and $9.1 \times 10^{-13} \text{ W m}^{-2} \text{ K}^{-1}$ for the later HTC03 test.

Table 2: Principal conditional-PINN training parameters.

Item	Setting
HTC data split	10,000 train / 500 validation / 500 test functions
Coordinates and conditioning	$(r/R, z/H, t/t_{end}) + 40$ HTC descriptors
Architecture	three tanh MLP heads, 4 hidden layers, width 64
Parameter count	46,863
Precision and device	float64, CPU
Optimizer	Adam [16]; gradient clipping
Time sampling	50% in 0.05–30 s; 25% in 30–600 s; 25% log-uniform
Physics weights	heat 200; Robin 20; integral energy 100; phase 2; plastic 20
Phase refinement	20,000 trajectories, 16 time samples; 40 guidance epochs
Early case refinement	0.25–10 s oversampling; boundary-adaptive MSE
Mechanical refinement	full stress tensor plus von Mises target
FEM labels in general refinement	none

Training was staged. A physics-only conditional network was first optimised over the 10,000 HTC functions. An early-cooling ansatz and increased sampling within the first 30 s were then introduced. Finally, 20,000 phase trajectories were integrated from frozen PINN temperature histories using the conservative kinetics model, and the phase head was refined for 40 epochs. This last stage used no FEM labels and selected epoch 38. The step-halving check on 128 trajectories gave a maximum phase difference of 0.00332.

3.2 Finite-element screening

The numerical reference is a P1 triangular axisymmetric FEM discretisation with the radial weight r in the weak form [15]. Backward Euler time integration and Picard iteration treat the nonlinear Robin boundary. Phase fractions are updated nodally, and latent heat is included with a one-step lag. The general screening mesh contains 793 nodes and 1440 triangles; adaptive runs required 527–542 accepted steps, with 67–107 rejected steps and no non-converged

Picard step. Three development HTC cases and three fresh synthetic cases were screened at $t = 5, 15, 60, 180$, and 600 s.

Axisymmetric nodal RMSE is calculated as

$$\text{RMSE}_r(y) = \left[\frac{\sum_{i=1}^N r_i (y_i^{\text{PINN}} - y_i^{\text{FEM}})^2}{\sum_{i=1}^N r_i} \right]^{1/2}. \quad (8)$$

The weighting approximates a cylindrical volume measure but is not exact finite-element quadrature. Reported aggregate values are means of the snapshot RMSE values, not time-integrated norms.

3.3 Out-of-family case adaptation

An additional tabulated HTC field, denoted HTC03, does not satisfy the generated family: it lacks the prescribed Leidenfrost local maximum and exhibits a broad maximum plateau near the top. The zero-shot conditional PINN is therefore evaluated as an out-of-distribution test, followed by case-specific fine-tuning. The exact HTC03 table remains in the physics boundary loss. Coarse FEM data (793 nodes, 21 snapshots) provide 13,314 training and 3339 spatially held-out examples. Only the thermal and phase heads are adapted for 4000 Adam steps; mechanical weights remain unchanged. A distinct finer FEM calculation with 2121 nodes, 4000 triangles, one-quarter time step scale, and nine snapshot times is withheld from training and checkpoint selection. Four of its times (11, 40, 100, and 240 s) are also absent from the training snapshots.

3.4 Early-transient and tensor-level refinement

The case-specific study was extended to address the largest discrepancy, which occurred while the surface crossed the boiling regime during the first seconds. FEM snapshots were generated at 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 5, 6, 7.5, and 10 s. During thermal refinement, 75% of the sampled times were taken from this interval and 60% of the spatial points were drawn from water-exposed boundary bands. Direct mean-square error and error-adaptive sampling replaced the earlier robust loss because the latter under-penalised rare but large edge errors. Denote the case-adapted base temperature by T^{base} . It was augmented by a local residual head $\Delta T_e(r, z, t)$ to obtain the early-corrected field. Its multiplicative envelope is zero at $t = 0$ and decays rapidly after the initial transient,

$$T^{\text{early}}(r, z, t) = T^{\text{base}}(r, z, t) + \left(1 - e^{-t/(0.08 \text{ s})}\right) e^{-[t/(5 \text{ s})]^8} N_e(r, z, t), \quad (9)$$

where N_e is a four-layer tanh multilayer perceptron with width 80 and Fourier coordinate features. Consequently, the exact initial condition is retained and predictions from approximately 7.5 s onward revert to the original validated PINN. To impose an explicit engineering acceptance criterion on the remaining local error, a second boundary-specialist residual head was subsequently trained on top of the frozen conditional and early-transient networks. The temperature after both residual corrections is denoted T^{final} . This expert uses engineered space-time and Fourier features, five SiLU hidden layers of width 128, and an envelope that preserves the exact initial condition while concentrating its capacity on the first approximately 10 s. Its objective combines field MSE, a worst-5% tail MSE, and a smooth maximum-error penalty. Equal-time sampling prevents the densely sampled early interval from dominating the objective, while 60% of each update is drawn from the currently worst predicted nodes

and 25% from water-exposed boundary bands. The discrete acceptance criterion was fixed before validation as

$$\max_{t_k \in \mathcal{T}_{val}} \max_{i \in \mathcal{N}_{FEM}} |T^{\text{final}}(r_i, z_i, t_k) - T^{\text{FEM}}(r_i, z_i, t_k)| < 20^\circ\text{C}. \quad (10)$$

Here $\mathcal{T}_{val}$ contains 45 distinct intermediate and late validation times, and $\mathcal{N}_{FEM}$ contains the 4025 nodes of the 25×161 axisymmetric FEM grid. Training used 42 different snapshot times. Neither the validation times nor their fields entered optimisation; accordingly, the criterion bounds the sampled validation set rather than the continuous space–time domain. The mechanical head was then calibrated against FEM values of σ_{rr} , $\sigma_{\theta\theta}$, σ_{zz} , and τ_{rz} , with an additional von Mises loss. Boundary bands and 0.5–10 s states were over-sampled because the steepest thermo-transformation stress gradients occur there. Absolute MPa errors and full-field RMSE are reported because relative stress error is unstable where the FEM reference stress is close to zero. No independent mechanical hold-out split is claimed for these calibrated comparisons.

4 Results

4.1 General conditional model

Table 3 and Figure 3 summarise the six-case screening. Phase refinement reduced the mean diffusional-fraction RMSE by 85.7% and the martensite RMSE by 66.6%. Similar improvements on development and fresh cases indicate that the kinetics distillation did not merely memorise the three original checks. The temperature head was frozen and its aggregate RMSE remained 56.17°C.

Table 3: Mean snapshot RMSE before and after phase refinement. Fractions are dimensionless.

Case set	Temperature [°C]		f_d		f_m	
	Before	After	Before	After	Before	After
Development (3)	59.965	59.965	0.3379	0.0490	0.2922	0.1028
Fresh synthetic (3)	52.375	52.375	0.3361	0.0477	0.2947	0.0934
All six	56.170	56.170	0.3370	0.0483	0.2934	0.0981

The aggregate thermal number hides a strong time dependence. At 5 s, an unseen case exhibits substantial end and surface errors because the PINN under-resolves the rapid collapse of the nonlinear boiling boundary layer. By 60 s, temperature RMSE in individual cases is approximately 7°C, and it becomes negligible after the body approaches water temperature. Phase errors persist because final fractions integrate the earlier thermal path. The field comparisons in Figure 4 therefore represent both the capability and the current limitation of the general surrogate.

The six FEM runs required 34.0–46.9 s each on CPU, whereas evaluating five PINN snapshots required 0.033–0.039 s. These timings are not directly comparable as end-to-end workflows: FEM timing includes transient integration, whereas PINN timing covers inference only and excludes generation of the 10,000 boundary functions, optimisation, and kinetics refinement. They nevertheless quantify the workload-specific online advantage once a suitable surrogate exists.

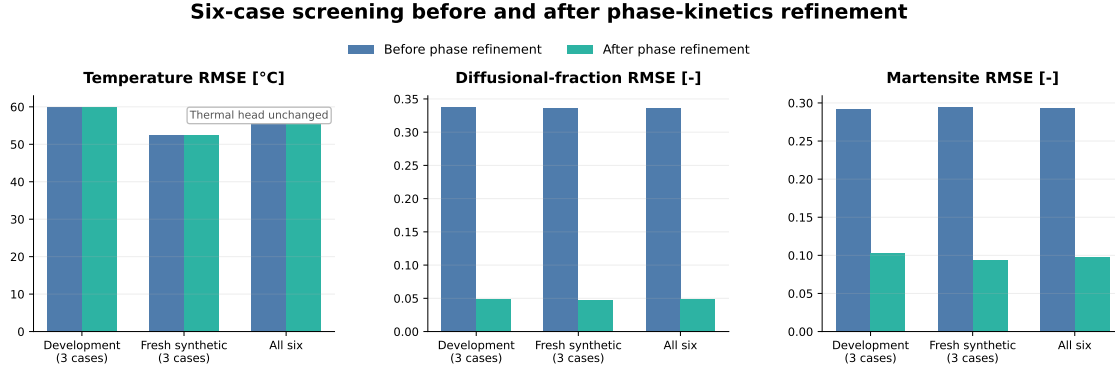


Figure 3: Mean snapshot RMSE in paired six-case FEM screening before and after phase refinement. The thermal head was not modified during this stage.

4.2 Case-specific HTC03 adaptation

The out-of-family HTC03 test provides a sharper assessment of model-domain limits. Zero-shot temperature and martensite errors are large. FEM-assisted adaptation lowers nine-snapshot temperature RMSE from 56.394 to 5.714 °C (89.9%) and martensite RMSE from 0.1065 to 0.0307 (71.1%). At the four unseen snapshot times, temperature RMSE is 4.989 °C and martensite RMSE is 0.0318. In contrast, the diffusional-fraction RMSE increases from 0.0215 to 0.0362 over all nine snapshots. This trade-off is important: minimising a combined loss can improve the dominant thermal and martensitic fields while degrading a smaller constituent.

The adapted 15 s field in Figure 6 reproduces the main axial thermal pattern and martensitic end zones. Remaining discrepancies are concentrated near the end faces and in narrow transformation bands. The first few seconds remain more difficult than the aggregate RMSE suggests. This observation motivated the dedicated early-transient and maximum-error refinement described in the following subsection. Figure 7 presents the final refined cooling curves. At the five $r = 9$ mm probes, the maximum absolute PINN-FEM difference over 0–100 s is 12.18 °C; the maximum sampled nodal error over 45 independent validation times is 16.62 °C. Both values satisfy the predefined discrete 20 °C acceptance threshold.

Table 4: Independent sampled validation of the maximum-error-constrained thermal PINN.

Metric	Value
Independent validation times	45
FEM nodes per field	4025
Mean full-field RMSE	2.20 °C
Maximum sampled nodal error	16.62 °C
Maximum five-probe cooling-curve error	12.18 °C
Predefined acceptance threshold	20.00 °C

Case adaptation required 46.38 s. Five-snapshot PINN evaluation required 0.0191 s; the finer FEM validation run required 363.88 s. The isolated inference-to-FEM ratio is therefore approximately 1.9×10^4 , but it must not be interpreted as an end-to-end speed-up. Including one coarse FEM training run (43.79 s) and adaptation yields an offline cost of about 90.2 s before inference, not counting the general model development. For one isolated trusted FEM

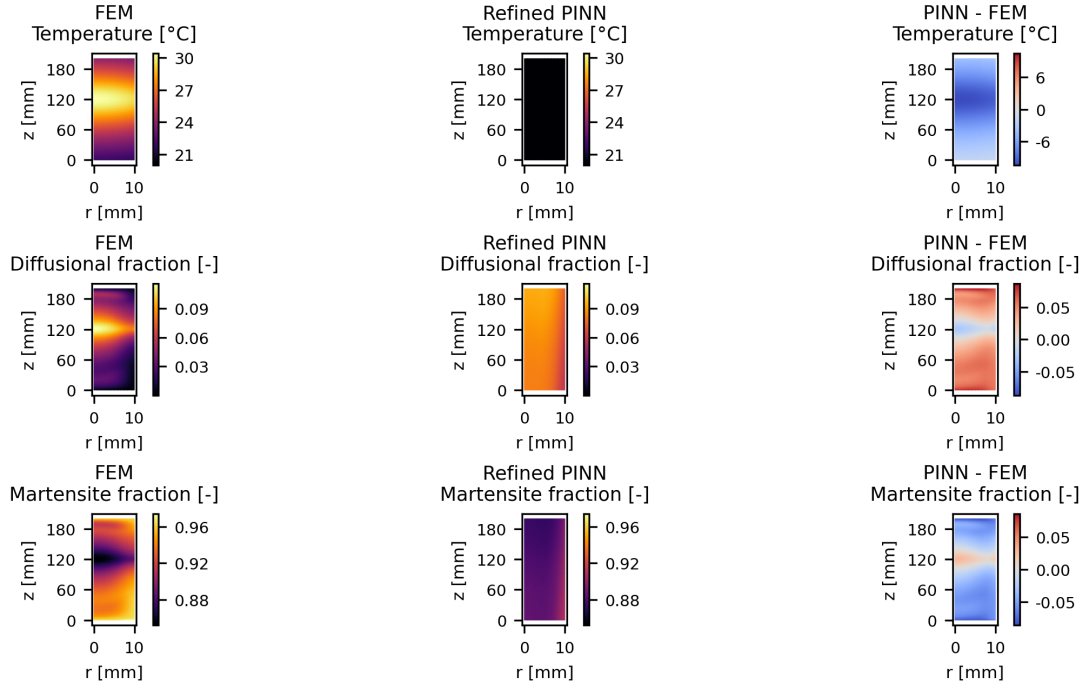
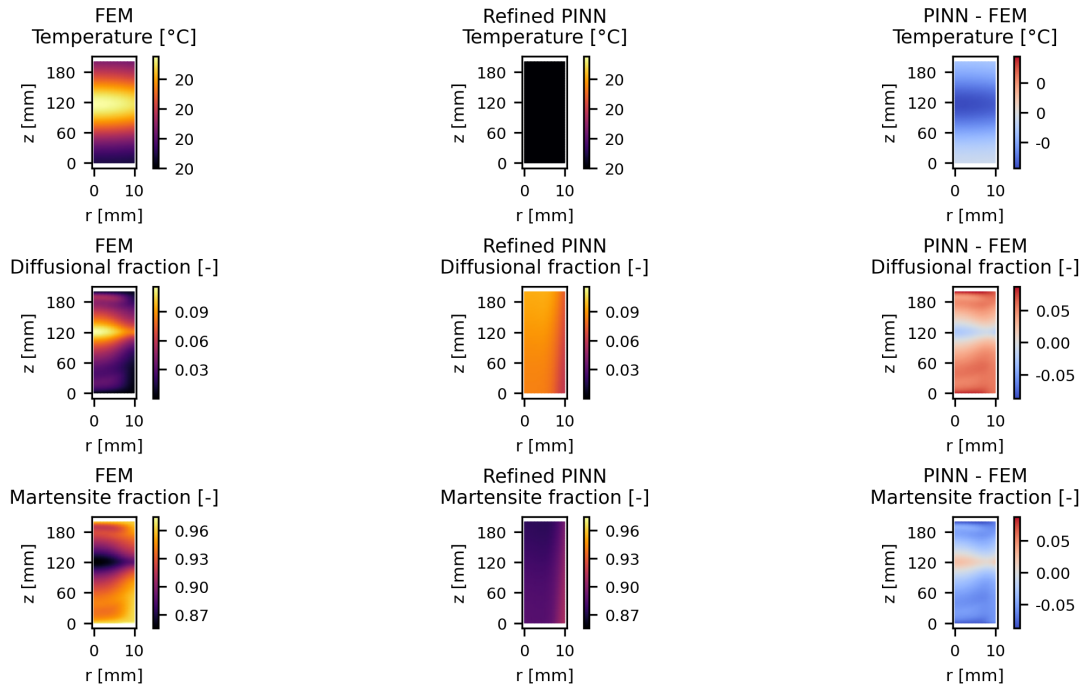
Fresh HTC case 10503 - $t = 60$ s - FEM screening**Fresh HTC case 10503 - $t = 180$ s - FEM screening**

Figure 4: Representative fresh HTC case 10503 at 60 s (upper block) and 180 s (lower block). Within each block, columns show FEM, refined conditional PINN, and the signed PINN–FEM difference; rows show temperature, lumped diffusional fraction, and martensite fraction. Temperature RMSE decreases from 6.72 °C at 60 s to 0.048 °C at 180 s; the corresponding phase-fraction RMSE values are approximately 0.046–0.047. The vertically stacked layout enlarges the field maps and keeps the comparison columns adjacent.

Out-of-family HTC03: finer-FEM evaluation before and after adaptation

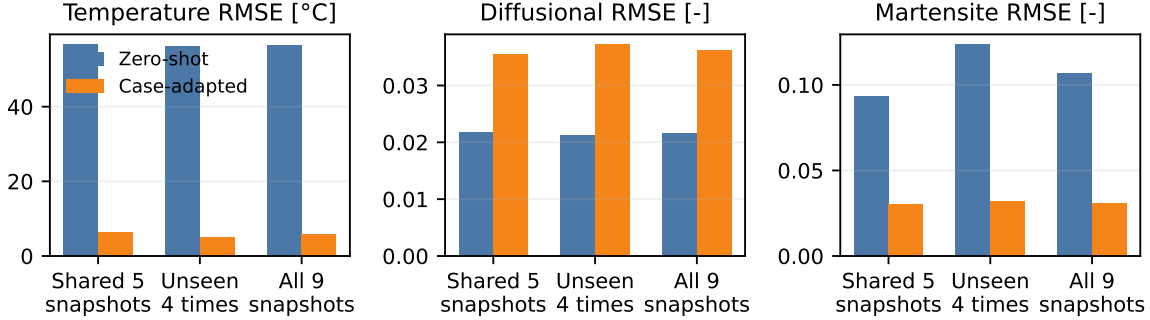


Figure 5: Finer-FEM HTC03 metrics before zero-shot and after case-specific adaptation. The separate bars for unseen snapshot times test temporal interpolation, not generalisation to a new HTC.

result, FEM remains the defensible choice; the adapted PINN becomes useful when many repeated evaluations of the same case are required.

4.3 Improved early transient and mechanical comparison

Table 5 and Figure 8 compare the original case-adapted thermal field with the early-corrected result. Averaged over the twelve evaluated instants between 0.25 and 10 s, full-field RMSE decreases from 6.77 to 3.50 °C, a reduction of 48.3%. Table 5 lists nine representative instants, while Figure 8 shows the complete twelve-time series. The largest improvements coincide with the rapid boundary-regime transition: RMSE decreases from 13.07 to 4.35 °C at 1 s, from 15.78 to 5.66 °C at 2 s, and from 8.49 to 4.24 °C at 3 s. The correction is deliberately local in time. At 5 s the original network is already accurate and the corrected RMSE is slightly larger; from 7.5 s onward the envelope in Eq. (9) restores the base solution.

Table 5: Full-field temperature RMSE at nine representative times before and after targeted early-transient refinement.

t [s]	0.25	0.50	0.75	1.0	1.5	2.0	3.0	5.0	10.0
Original [°C]	1.98	3.75	9.08	13.07	16.52	15.78	8.49	2.38	2.00
Corrected [°C]	2.85	2.92	3.32	4.35	5.68	5.66	4.24	2.95	2.00

The FEM-calibrated mechanical comparison was performed component by component rather than solely through the equivalent stress. Table 6 gives the full-field RMSE in MPa. The critical 5 s state remains the most difficult: radial and hoop stresses have RMSE values above 60 MPa, while the von Mises RMSE is 22.42 MPa. This discrepancy is concentrated in narrow end and edge zones where stress is highly sensitive to temperature and transformation gradients. The error decreases rapidly after the early transient. Von Mises RMSE falls to 8.09 MPa at 10 s, 2.76 MPa at 30 s, and 1.73 MPa at 180 s. Because the reported fields were used for calibration, these values quantify agreement with the FEM reference but do not establish independent mechanical generalisation.

Figure 9 shows the component-wise error decay. Figure 10 provides the 15 s spatial comparison, at which the main tensile and compressive zones are reproduced without the severe

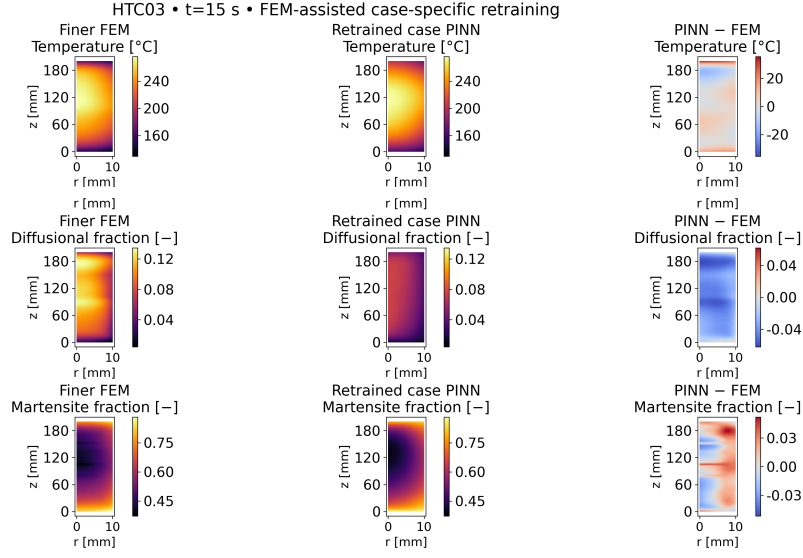


Figure 6: HTC03 at 15 s after case-specific FEM-assisted adaptation. The enlarged upper block shows temperature; the lower block shows diffusional and martensite fractions. In every row, the panels are finer FEM, adapted PINN, and signed PINN–FEM difference from left to right.

Table 6: FEM–PINN stress RMSE in MPa for the four axisymmetric tensor components and von Mises stress.

t [s]	σ_{rr}	$\sigma_{\theta\theta}$	σ_{zz}	τ_{rz}	σ_{vm}
5	61.85	62.75	42.50	21.48	22.42
10	21.94	32.71	27.63	10.22	8.09
15	5.79	12.50	10.90	5.32	5.08
30	4.61	5.14	5.94	2.41	2.76
60	3.17	4.09	3.83	1.21	2.19
180	2.40	2.84	5.15	1.62	1.73
600	2.42	3.15	5.22	1.90	1.90

concentration observed at 5 s. Columns show PINN, FEM, and their signed difference for all four tensor components and the von Mises field. This comparison provides the previously missing tensor-level evidence for the mechanical branch, while an independent hold-out study remains future work.

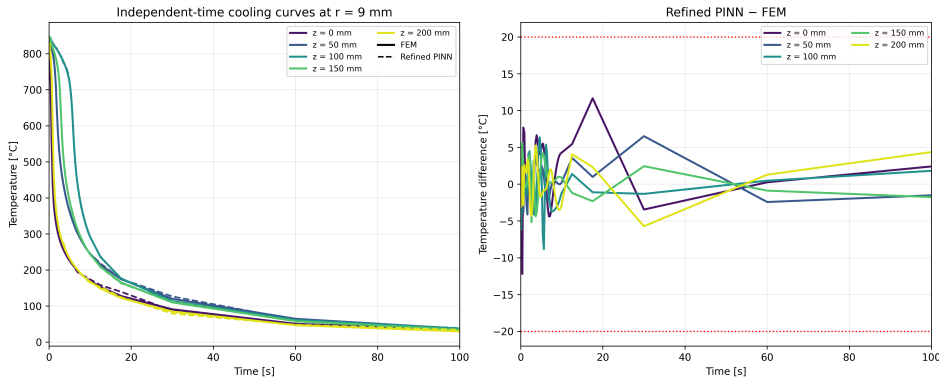


Figure 7: FEM and maximum-error-constrained PINN cooling curves at $r = 9$ mm and five axial positions over 0–100 s (left), with signed PINN–FEM temperature differences (right). The maximum absolute probe error is 12.18 °C; the dotted lines show the predefined ± 20 °C acceptance limits.

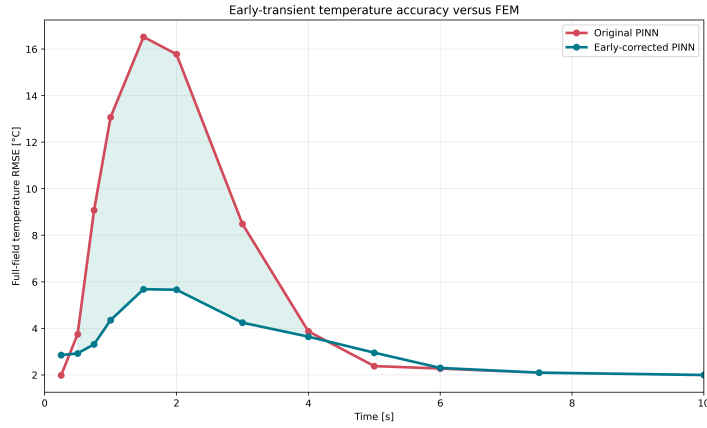


Figure 8: Full-field temperature RMSE relative to the FEM reference during the first 10 s. The residual head primarily removes the broad error maximum between 0.75 and 3 s and is switched off after the rapid initial transient.

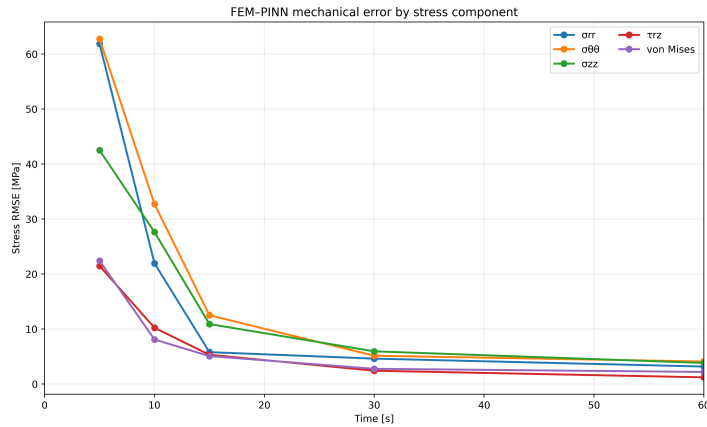


Figure 9: Component-wise mechanical RMSE during the most active 5–60 s interval. Absolute MPa error is used because relative error becomes unstable where the FEM stress is close to zero.

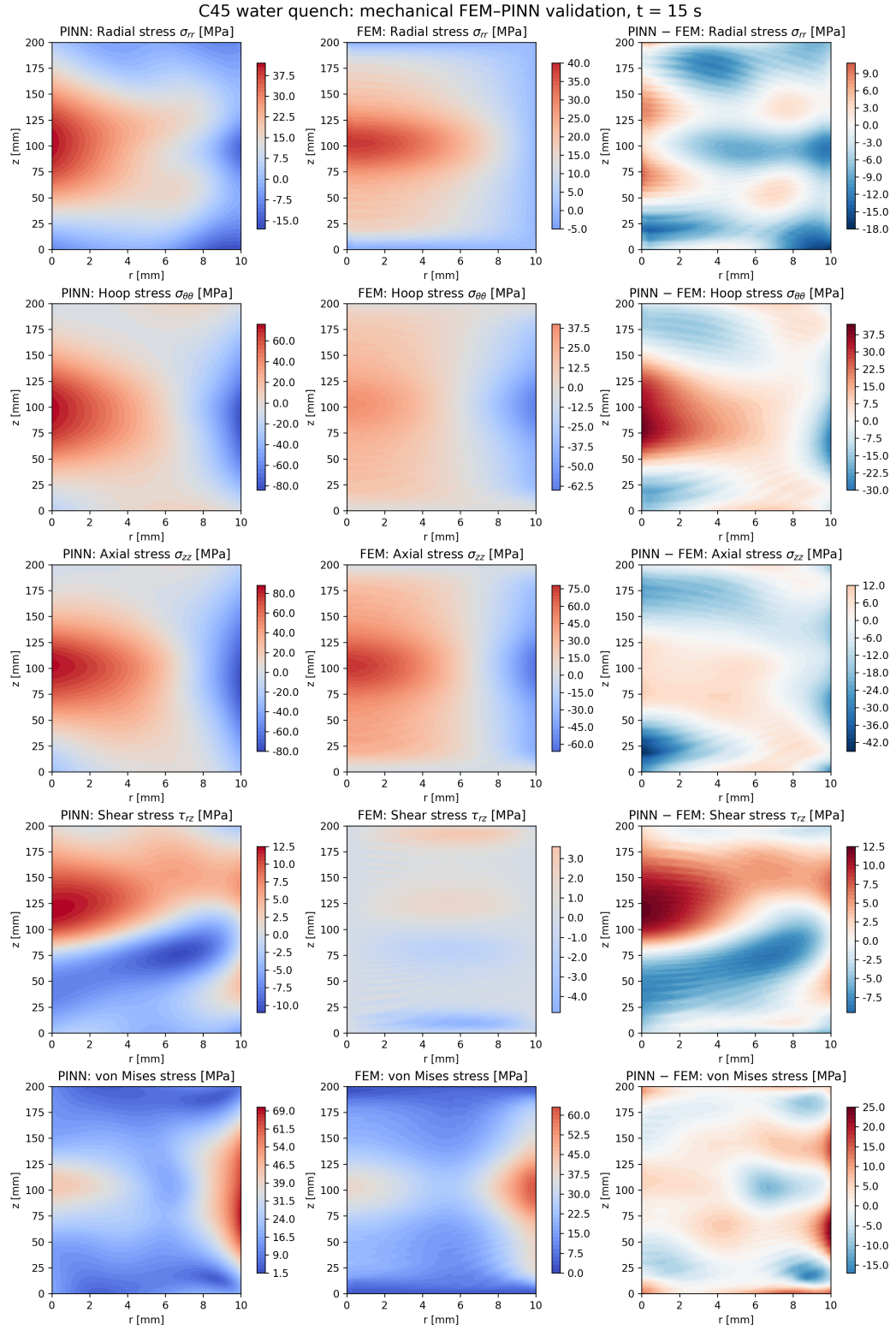


Figure 10: Mechanical FEM–PINN comparison at 15 s. Rows show σ_{rr} , $\sigma_{\theta\theta}$, σ_{zz} , τ_{rz} , and von Mises stress; columns show PINN, FEM, and signed PINN–FEM difference. All values are in MPa.

4.4 Computational-speed comparison

The principal deployment advantage of the conditional PINN is the reduction in repeated online calculation time. A separate batch benchmark was therefore performed for 1, 5, 15, and 50 model evaluations. The measured PINN inference times were 0.162, 0.808, 2.429, and 10.025 s, respectively, whereas the corresponding transient FEM times were 35.09, 175.47, 526.41, and 1754.69 s. As shown in Figure 11, the resulting online speed-up is approximately 217-fold for 1–15 evaluations and 175-fold for 50 evaluations. The nearly linear growth of both series makes the repeated-use benefit explicit.

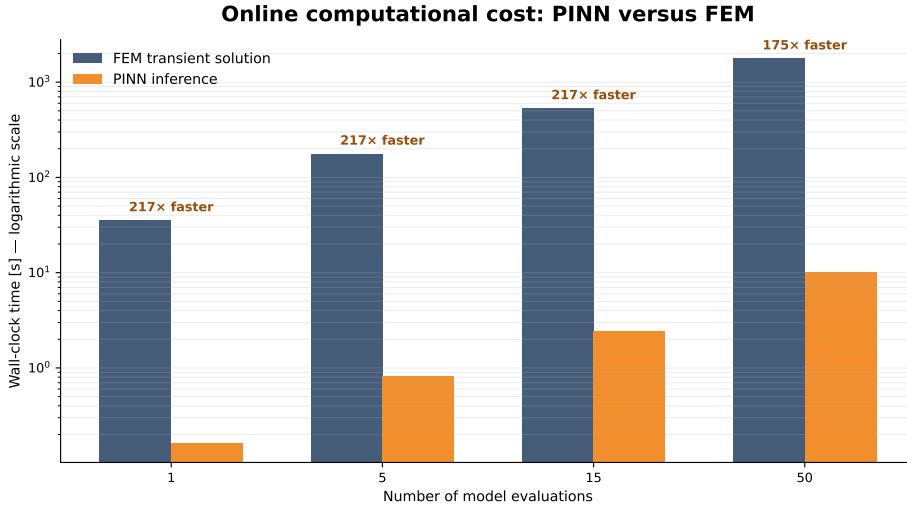


Figure 11: Measured wall-clock time for repeated model evaluations. The logarithmic ordinate preserves visibility of both methods. The reported 175–217× advantage is an online inference comparison: it excludes generation of the 10,000-function HTC training family, PINN optimisation, FEM reference-data generation, and any case-specific adaptation.

This distinction is essential for a fair assessment. The PINN is not faster for the first prediction if its full offline development cost is charged to that single case. It becomes computationally advantageous when the trained surrogate is reused for parameter sweeps, optimisation, uncertainty analysis, or monitoring inside its validated HTC domain. The case-specific HTC03 timing leads to the same conclusion: its much larger inference-to-FEM ratio applies only after the adaptation investment has been paid. The batch benchmark in this subsection uses a different workload from the earlier five-snapshot timings and should therefore be interpreted separately rather than combined into a single ratio.

5 Discussion

The study demonstrates three nested operating modes. The broad conditional PINN interpolates within a constrained HTC family and provides rapid screening. Case-specific HTC adaptation improves an out-of-family boundary law but sacrifices HTC generality. Finally, the local early-transient residual targets the short interval in which the nonlinear boundary layer is most difficult. This hierarchy is useful in practice: the least expensive model may be used first, while adaptation is reserved for boundary laws or time intervals that fail a predefined accuracy criterion. The dominant source of the early discrepancy is the combination of a nonlinear boundary flux and inadequate temporal/spatial representation near $t = 0$. In this

interval, the surface temperature crosses the boiling peaks rapidly, so a moderate temperature error changes $h(T, z)$ and feeds back into the boundary flux. The slender 20:1 height-to-radius geometry also lets different end and side histories overlap. Dense early-time sampling, boundary oversampling, Fourier features, and error-adaptive MSE first reduced the mean 0.25–10 s error by 48.3%, confirming this diagnosis. The subsequent boundary-specialist expert and worst-tail objective removed the isolated error peaks that were not controlled by RMSE alone: over the independent sampled 0–100 s validation set, the mean full-field RMSE is 2.20 °C and the maximum sampled nodal error is 16.62 °C. This result illustrates why a maximum-error acceptance test is needed in addition to an averaged norm [4]. The mechanical comparison also clarifies the propagation of thermal error through the coupled model. Stress depends on spatial derivatives of displacement and on thermal and transformation eigenstrains; it is therefore more sensitive than temperature to a narrow misplaced front. Tensor-level supervision is essential because matching only von Mises stress does not uniquely determine the individual normal and shear components. The 5 s tensor errors remain too large for local crack-risk assessment, whereas the 15–600 s von Mises RMSE of 1.73–5.08 MPa supports calibrated field-level screening for the studied case. Independent mechanical tests are still needed before claiming generalisation to new HTC fields. The FEM reference also has limitations. P1 spatial discretisation, nodal phase updates, approximate radial error weights, and one-step-lagged latent heat introduce numerical uncertainty. For HTC03, coarse-to-fine RMS differences are 3.34 °C at 5 s and 1.15 °C at 60 s; phase differences reach approximately 0.0365. These values provide context for the PINN error and show why a formal mesh/time-convergence study is necessary before a definitive accuracy claim. No experimental temperatures, dilatometry, hardness maps, distortion, or residual stresses are used here. C45 composition and TTT data describe a nominal steel rather than a particular heat. Ferrite and pearlite are lumped into one diffusional fraction, transformation plasticity is not independently calibrated, and stress-assisted kinetics and mechanical-dissipation feedback are absent. The residual correction is also case-specific and should not be interpreted as proof that the original 10,000-function conditional surrogate has acquired the same early-time accuracy over its complete HTC domain. Consequently, the model is best viewed as a documented numerical research platform and not as a qualified heat-treatment design tool.

6 Conclusions

The developed conditional PINN represents a coupled quenching model with thermal, metallurgical, and mechanical fields. It comprises transient heat conduction with a nonlinear temperature- and position-dependent HTC boundary condition, competing diffusional and martensitic transformation kinetics, and a downstream mechanical branch for thermal strain, transformation strain, plastic deformation, and stress prediction. Latent heat couples the thermal and metallurgical fields, while the mechanical response is driven by temperature and phase evolution without feedback through stress-assisted kinetics or mechanical dissipation. The results must be interpreted by operating mode. For the general six-case conditional screening, the mean temperature RMSE is 56.17 °C; phase-only refinement reduces the diffusional- and martensite-fraction RMSE values to 0.0483 and 0.0981, respectively. For the out-of-family HTC03 case, FEM-assisted adaptation and targeted early-time refinement reduce the twelve-time mean temperature RMSE over 0.25–10 s from 6.77 to 3.50 °C. The final boundary specialist, evaluated at 45 held-out times on a 4025-node FEM grid, achieves a mean full-field RMSE of 2.20 °C, a maximum sampled nodal error of 16.62 °C, and a maximum five-probe error of 12.18 °C, satisfying the predefined discrete 20 °C criterion. The FEM-calibrated me-

chanical comparison includes radial, hoop, axial, shear, and von Mises stresses. Von Mises RMSE decreases from 22.42 MPa at 5 s to 8.09 MPa at 10 s and 1.73 MPa at 180 s; independent mechanical hold-out validation remains necessary. Across the separate batch benchmark of 1–50 repeated evaluations, online PINN inference is approximately 175–217 times faster than the corresponding transient FEM calculations. This advantage excludes model development, FEM-data generation, and case-specific adaptation, and therefore applies only after the offline cost has been paid. The present evidence supports rapid numerical screening and repeated evaluation within the tested model domains. Optimisation, uncertainty analysis, digital-twin deployment, model-predictive control, and inverse-HTC reconstruction are promising future applications, but require dedicated inverse tests, uncertainty quantification, and experimental validation before industrial use.

Data and code availability

The computational package produced in this study contains Python source code, generated HTC examples, trained checkpoints, FEM fields, JSON/CSV metrics, and figure-generation scripts. A curated repository and persistent DOI will be added before submission. The archived release should include the residual definitions, constitutive parameter tables, normalization constants, data splits, and random seeds required to reproduce the reported calculations. Redistribution of the supplied C45 material file remains subject to its original licensing terms.

Acknowledgement

The authors acknowledge the support from project 2020-1.1.2-PIACI-KFI-2020-00129 and BR24993001 "Creating a Large Language Model (LLM) to support the Kazakh language and technological progress". This work was also supported by the National Key Research and Development Program of China (2013YFB3406801).

References

- [1] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019, doi: 10.1016/j.jcp.2018.10.045.
- [2] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nat. Rev. Phys.*, vol. 3, no. 6, pp. 422–440, 2021, doi: 10.1038/s42254-021-00314-5.
- [3] S. Cuomo, V. Schiano Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli, "Scientific machine learning through physics-informed neural networks: Where we are and what's next," *J. Sci. Comput.*, vol. 92, no. 3, Art. no. 88, 2022, doi: 10.1007/s10915-022-01939-z.
- [4] S. Wang, Y. Teng, and P. Perdikaris, "Understanding and mitigating gradient flow pathologies in physics-informed neural networks," *SIAM J. Sci. Comput.*, vol. 43, no. 5, pp. A3055–A3081, 2021, doi: 10.1137/20M1318043.
- [5] E. Haghighat, M. Raissi, A. Moure, H. Gomez, and R. Juanes, "A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics," *Comput. Methods Appl. Mech. Eng.*, vol. 379, Art. no. 113741, 2021, doi: 10.1016/j.cma.2021.113741.
- [6] S. Niu, E. Zhang, Y. Bazilevs, and V. Srivastava, "Modeling finite-strain plasticity using physics-informed neural network and assessment of the network performance," *J. Mech. Phys. Solids*, vol. 172, Art. no. 105177, 2023, doi: 10.1016/j.jmps.2022.105177.
- [7] H. Lian, P. Zhao, M. Zhang, P. Wang, and Y. Li, "Physics informed neural networks for phase field fracture modeling enhanced by length-scale decoupling degradation functions," *Front. Phys.*, vol. 11, Art. no. 1152811, 2023, doi: 10.3389/fphy.2023.1152811.

- [8] S. Li, G. Wang, Y. Di, L. Wang, H. Wang, and Q. Zhou, “A physics-informed neural network framework to predict 3D temperature field without labeled data in process of laser metal deposition,” *Eng. Appl. Artif. Intell.*, vol. 120, Art. no. 105908, 2023, doi: 10.1016/j.engappai.2023.105908.
- [9] D. P. Koistinen and R. E. Marburger, “A general equation prescribing the extent of the austenite–martensite transformation in pure iron-carbon alloys and plain carbon steels,” *Acta Metall.*, vol. 7, no. 1, pp. 59–60, 1959, doi: 10.1016/0001-6160(59)90170-1.
- [10] M. Avrami, “Kinetics of phase change. I: General theory,” *J. Chem. Phys.*, vol. 7, no. 12, pp. 1103–1112, 1939, doi: 10.1063/1.1750380.
- [11] J. W. Christian, *The Theory of Transformations in Metals and Alloys*, 3rd ed. Oxford, U.K.: Pergamon, 2002.
- [12] G. E. Totten, C. E. Bates, and N. A. Clinton, *Handbook of Quenchants and Quenching Technology*. Materials Park, OH, USA: ASM International, 1993.
- [13] B. Liščić, H. M. Tensi, L. C. F. Canale, and G. E. Totten, Eds., *Quenching Theory and Technology*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2010.
- [14] S. M. C. van Bohemen and J. Sietsma, “Kinetics of martensite formation in plain carbon steels: Critical assessment of possible influence of austenite grain boundaries and autocatalysis,” *Mater. Sci. Technol.*, vol. 31, no. 5, pp. 590–600, 2015, doi: 10.1179/1743284714Y.0000000532.
- [15] O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu, *The Finite Element Method: Its Basis and Fundamentals*, 7th ed. Oxford, U.K.: Butterworth-Heinemann, 2013.
- [16] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.
- [17] I. Felde, “Heat transfer during quenching processes,” in *ASM Handbook, Vol. 4F: Quenchants and Quenching Technology*, G. E. Totten, R. Simencio Otero, X. Luo, and L. C. F. Canale, Eds. Materials Park, OH, USA: ASM International, 2024, pp. 55–68, doi: 10.31399/asm.hb.v4F.a0006997.
- [18] I. Felde, “Artificial intelligence techniques and biomimetic methods supporting heat treatment processes,” *Acta Polytech. Hung.*, vol. 21, no. 10, pp. 317–329, 2024, doi: 10.12700/APH.21.10.2024.10.20.
- [19] S. Szénási and I. Felde, “Using multiple graphics accelerators to solve the two-dimensional inverse heat conduction problem,” *Comput. Methods Appl. Mech. Eng.*, vol. 336, pp. 286–303, 2018, doi: 10.1016/j.cma.2018.03.024.
- [20] M. Rezaei *et al.*, “Physics-informed deep learning for enhanced lithium-ion battery performance,” *J. Probl. Comput. Sci. Inf. Technol.*, vol. 4, no. 3, 2026.
- [21] T. Réti and I. Felde, “A non-linear extension of the additivity rule,” *Comput. Mater. Sci.*, vol. 15, no. 4, pp. 466–482, 1999.
- [22] T. Réti, Z. Fried, and I. Felde, “Computer simulation of steel quenching process using a multi-phase transformation model,” *Comput. Mater. Sci.*, vol. 22, no. 3–4, pp. 261–278, 2001.
- [23] T. Réti, I. Felde, J. Grum, R. Colás, G. Sánchez Sarmiento, and A. Moita de Deus, “Extension of isothermal time-temperature parameters to non-isothermal conditions: Application to the simulation of rapid tempering,” *Strojniški vestnik – J. Mech. Eng.*, vol. 56, no. 2, pp. 84–92, 2010.

Imre Felde,
Obuda University, Budapest, Hungary.
Selye Janos University, Slovakia.
email: felde.imre@uni-obuda.hu

Norbert Annus,
Selye Janos University, Slovakia.

Shi Wei,
Tsinghua University, China.