Accelerating Thermochemical Equilibrium Calculations with Physics-Informed Neural Networks

Klara Meyer*§, Willem A. Roos†§, Johan H. Zietsman†§, Anna Sergeevna Bosman*‡§
*Department of Computer Science, University of Pretoria, Pretoria, South Africa
†Ex Mente Technologies, Pretoria, South Africa
‡African Institute for Data Science and Artificial Intelligence (AfriDSAI), University of Pretoria, Pretoria, South Africa
§E-mail: klara.meyer@tuks.co.za, willem.roos@ex-mente.co.za, johan.zietsman@ex-mente.co.za, anna.bosman@up.ac.za

***Abstract*—Estimating thermochemical equilibrium in multi-component systems is essential for materials science and process engineering, yet computational costs scale poorly with system complexity. This work presents a physics-informed neural network (PINN) approach for scalability from binary to quaternary oxide systems in the CaO-MgO-Al_2O_3-SiO_2 space. Systematic evaluation reveals that selective physics-guided feature engineering improves test R^2 by 8.3%, while embedding the Gibbs-Helmholtz equation as an architectural constraint reduces thermodynamic consistency error by over six orders of magnitude. The proposed two-stage hybrid architecture, comprising a Transformer encoder for phase estimation, and a multi-layer perceptron for property estimation, achieves $R^2 = 0.96$ on binary, $R^2 = 0.94$ on ternary, and $R^2 = 0.90$ on quaternary systems, with inference speeds of 72 μs/sample (14–140$\times$ faster than direct calculation). These results demonstrate that PINNs offer a scalable path toward rapid equilibrium estimation where traditional methods become computationally prohibitive.**

***Index Terms*—physics-informed neural networks, thermochemical equilibrium, multi-component systems, scalability, deep learning, oxide systems**

I. Introduction

Thermochemical equilibrium calculations underpin materials design, metallurgical processing, and geological modeling. For multi-component oxide systems, determining equilibrium phase assemblages requires minimizing the system's Gibbs energy. This is a computationally demanding optimization problem, where computational cost scales approximately cubically with the number of components [4]. While binary and ternary systems can be solved efficiently using traditional methods, industrial applications increasingly require estimates for quaternary, quinary, and higher-order systems, where traditional approaches become impractical for real-time use.

Traditional CALPHAD-based Gibbs energy minimization provides high accuracy, but computational cost grows rapidly with system complexity [1]. Roos et al. [2]–[4] achieved impressive acceleration (3–20$\times$ speedup for binary systems, up to 1000$\times$ for five-component systems) through geometric interpolation on in-situ populated equilibrium data. However, pre-computed grid-based approaches face storage requirements that scale exponentially with the number of components ($O(n^C)$ where n is grid resolution and C is the number of components), making them impractical beyond quaternary systems [14]. For a modest resolution of 100 points per dimension, a five-component system would require 10^{10} pre-computed points.

Physics-informed neural networks (PINNs) [5] incorporate domain knowledge through architecture, loss functions, or input features, offering $O(1)$ storage and inference scaling independent of system complexity. Recent developments have demonstrated integration of differential equations as soft constraints [6], adaptive activation functions for improved convergence [7], and applications to materials science including property estimation [8] and phase diagram generation [9]. The ThermoLearn framework [10] recently demonstrated thermodynamics-informed neural networks achieving high accuracy on metal oxide compounds by leveraging the Gibbs free energy equation as an architectural constraint. Additionally, a recent review by Shen [11] highlighted the growing synergy between machine learning and CALPHAD methodologies. However, systematic evaluation of physics incorporation strategies for thermochemical equilibrium estimation with explicit scalability validation from binary to higher-order systems remains absent from the literature. To the authors' knowledge, no prior machine learning studies have targeted equilibrium estimation in the specific oxide systems examined in this work, precluding direct baseline comparison.

The key question is not just whether neural networks can approximate equilibrium calculations for a fixed system, but whether an architecture developed on simple systems generalizes effectively to more complex ones while maintaining thermodynamic consistency. This work addresses scalability of PINNs through progressive validation from binary to quaternary systems. The key contributions include:

- Systematic evaluation of physics incorporation strategies: feature engineering (+8.3% R^2) and architectural constraints (six orders of magnitude consistency improvement).
- Novel two-stage hybrid architecture combining Transformer attention mechanisms with densely connected multilayer perceptron.
- Demonstration of consistent generalization from binary

($R^2 = 0.96$) through ternary to quaternary systems ($R^2 = 0.90$).
- Comparative framework positioning PINNs relative to geometric acceleration methods.

All code is available at https://github.com/KlaraMeyer0/TEC.

II. Methodology

A. Problem Formulation

For an oxide system with C components at temperature T and constant pressure P (1.00 bar throughout this work), the proposed method estimates the following quantities:

1) *Phase presence:* Binary indication of which of the $\hat{\phi}$ candidate phases are thermodynamically stable.
2) *Phase fractions:* Molar proportions x_i of each stable phase, constrained by $\sum_{i=1}^{\hat{\phi}} x_i = 1$.
3) *Thermodynamic properties:* Enthalpy H, entropy S, heat capacity C_p, and Gibbs energy G.

Physical constraints include non-negativity ($S \geq 0$, $C_p \geq 0$, $x_i \geq 0$), phase fraction normalization, and thermodynamic consistency via the Gibbs-Helmholtz equation:

$$G = H - TS. \tag{1}$$

The input dimensionality scales linearly as $C + 1$ (compositions plus T), while output dimensionality scales as $O(\hat{\phi})$, suggesting PINNs should handle increasing complexity more gracefully than exponentially-scaling geometric methods.

B. Dataset Construction

A comprehensive dataset spanning the quaternary space CaO-MgO-Al_2O_3-SiO_2 was constructed, generated using FactSage 8.2 with the FToxide database. Table I summarizes dataset characteristics.

TABLE I
Dataset Characteristics Across System Complexities

System	Samples	T (K)	P (bar)	Max $\hat{\phi}$
Binary (6 systems)	60,000	500–2500	1.00	2
Ternary (4 systems)	199,888	500–2500	1.00	3
Quaternary (1 system)	99,622	500–2500	1.00	4
Total				**359,510 samples**

Z-score outlier removal (threshold $= 4.0$) was applied per property per system, providing substantial performance gains (5–10% improvement in R^2) without compromising thermodynamic validity. Input features were normalized using z-score standardization fit exclusively on training data.

C. Two-Stage Hybrid Architecture

The architecture separates the inherently different tasks of phase estimation (classification/regression) and property estimation (regression), as illustrated in Fig. 1.

Stage 1: Phase Transformer. Input features (augmented with physics-guided features) are embedded to dimension 256 with positional encoding. A Transformer encoder with 3 layers and 8 attention heads processes these embeddings. Two output heads estimate phase presence (sigmoid activation) and phase fractions (softmax activation ensuring normalization). Phase fractions are only meaningful for phases predicted as present. During training, losses are masked to exclude absent phases, preventing undefined gradients (detailed in Section II-F).

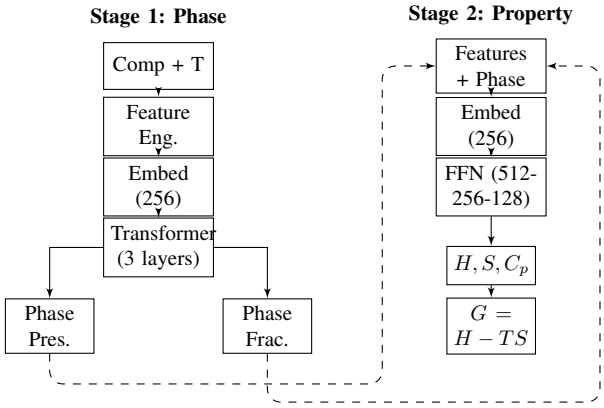

Fig. 1. Two-stage hybrid architecture. Stage 1 uses Transformer encoder for phase estimation. Stage 2 estimates H, S, C_p; G is computed via hard constraint.

Stage 2: Property Network. Original inputs are concatenated with *detached* Stage 1 outputs (gradients are not backpropagated to Stage 1, enabling independent optimization of each stage). A multilayer perceptron (MLP) with 3 hidden layers (512, 256, 128 neurons; see Section II-E for architecture selection) and ReLU activation estimates H, S, and C_p. Crucially, G is *computed* via Eq. (1) rather than estimated, embedding thermodynamic consistency as an architectural constraint.

D. Physics Incorporation Strategies

Three strategies for incorporating thermodynamic knowledge were systematically evaluated, with all experiments conducted on the CaO-SiO_2 binary system using 5 independent runs.

1) Physics-Guided Feature Engineering: Rather than relying on the neural network to discover thermodynamic relationships from raw inputs, domain knowledge can be explicitly encoded through engineered features. Three feature sets were evaluated: baseline (raw inputs only), exhaustive (all physically-motivated terms), and selective (top features by importance analysis).

Baseline features (7 total): Raw element compositions $[x_{Al}, x_{Ca}, x_{Mg}, x_O, x_{Si}]$, pressure P, and temperature T.

Exhaustive physics features (56 total) encode thermodynamic relationships: temperature-dependent terms ($1/T$, $\ln(T)$, $T\ln(T)$, T^2, T^3, $\sqrt{T}$) from heat capacity polynomials and Arrhenius kinetics; composition interaction terms including binary products $x_i \cdot x_j$ representing pairwise interaction energies, entropy of mixing $S_{mix} = -R\sum_i x_i \ln(x_i)$, and composition variance; temperature-composition cross-terms ($T \cdot S_{mix}$, $T^2 \cdot x_i$, $(1/T) \cdot \ln(x_i)$); and thermal energy scales (RT, RT^2, $R\ln(T)$).

TABLE II
FEATURE ENGINEERING COMPARISON (CaO-SiO_2, 5 RUNS)

Feature Set	Count	Test R^2	Δ vs Baseline
Baseline (raw inputs)	7	0.835	—
Exhaustive physics	56	0.811	-2.9%
Selective physics	**30**	**0.904**	$\mathbf{+8.3\%}$

TABLE III
HARD CONSTRAINT EVALUATION (CaO-SiO_2, 5 RUNS)

Model	Test R^2	G Consistency (MAE, J/mol)
Baseline (estimate all)	0.905	1494 ± 48
Calculate $G = H - TS$	**0.906**	**0.0004**
Derive $C_p = \partial H/\partial T$	0.594	0.0004

Selective physics features (30 total) were identified via permutation importance analysis. Top-ranked features were: binary composition products (highest), composition variance, S_{mix}, temperature terms (T^3, RT^2, T^2, $1/T$), and cross-terms ($T \cdot S_{mix}$, $P \cdot T$, $T^2 \cdot x_i$). Table II shows the R^2 comparison of the three approaches, confirming superiority of selective features.

The selective 30-feature set improved test R^2 by 8.3% compared to baseline. Notably, using all 56 exhaustive features *degraded* performance by 2.9%, demonstrating that indiscriminate feature addition introduces noise. The binary composition products emerged as the most important features, consistent with their role in capturing pairwise interaction energies that dominate non-ideal mixing behavior in oxide systems.

2) Soft Constraints: Thermodynamic relationships were encouraged through penalty terms:

$$\mathcal{L}_{total} = \mathcal{L}_{MSE} + \beta_1 \|G - (H - TS)\|^2 + \beta_2 \mathcal{L}_{mass} + \beta_3 \mathcal{L}_{pos} \tag{2}$$

where $\mathcal{L}_{mass} = \|\sum_i f_i - 1\|^2$ enforces phase fraction conservation and $\mathcal{L}_{pos} = \|\text{ReLU}(-C_p)\|^2$ penalizes thermodynamically forbidden negative heat capacities. Optimal weights $\beta_1 = 0.1$, $\beta_2 = 0.01$, $\beta_3 = 0.001$ were discovered via hyperparameter tuning, improving R^2 by 3.6%, which is inferior to feature engineering alone. Combining both strategies provided no additional benefit.

3) Hard Constraints: Instead of penalizing Gibbs-Helmholtz violations, Eq. (1) was embedded directly by computing $G = H - TS$ during the forward pass. Table III shows R^2 comparison of the approaches. Calculating G reduced consistency error by six orders of magnitude while maintaining accuracy. Deriving C_p via automatic differentiation failed ($R^2 = 0.59$), likely because neural networks struggle to learn smooth $H(T)$ functions near phase transitions.

4) Learning Rate Scheduling: Eight learning rate configurations were compared (5 runs each): fixed baseline, StepLR (30/50 epoch intervals, $\gamma = 0.5$), ExponentialLR (0.95/0.99 decay), CosineAnnealingLR, ReduceLROnPlateau, and OneCycleLR. ReduceLROnPlateau achieved best performance (+1.7% vs fixed, $p = 0.001$), adapting based on validation plateau detection (patience=10, factor=0.5). Aggressive strategies hurt performance: OneCycleLR (-2.1%) and fast exponential decay (-1.1%), suggesting thermochemical optimization requires conservative learning rate reduction rather than aggressive scheduling.

TABLE IV
SINGLE-STAGE VS. TWO-STAGE ARCHITECTURE COMPARISON

Architecture	Phase Acc.	Frac. MAE	Prop. R^2
Single-stage MLP	92.1%	0.320	0.847
Single-stage Transformer	94.4%	0.186	0.102
Two-stage MLP-MLP	99.7%	0.036	0.872
Two-stage Trans-Trans	99.7%	0.042	0.103
Two-stage Trans-MLP	**99.7%**	**0.036**	**0.952**

E. Architecture Selection

The two-stage hybrid architecture emerged from systematic exploration of multiple design choices. The key comparisons that informed these decisions are summarized below.

1) Single-Stage vs. Two-Stage Approaches: Initial experiments compared single-stage architectures (unified estimation of phases and properties) against two-stage approaches (sequential estimation). Table IV shows comparative results on the CaO-SiO_2 binary system.

The two-stage decomposition dramatically improved phase estimation (92–94% $\rightarrow$ 99.7% accuracy) by allowing each stage to specialize. It was observed that transformers [12] excelled at phase estimation through attention mechanisms capturing compositional interactions, while MLPs provided more stable property estimation. Transformers consistently struggled with property regression ($R^2 \approx 0.10$). The hybrid combination achieved best overall performance.

2) Hyperparameter Optimization: Given the architecture's sensitivity to hyperparameters, Optuna [13] was employed for Bayesian optimization across: learning rates (10^{-4}–10^{-2}), transformer dimensions (64–256), attention heads (2–8), layer counts (2–6), dropout rates (0–0.3), and loss weights. Key findings included:

- Learning rate scheduling: ReduceLROnPlateau (patience=10, factor=0.5) outperformed fixed rates by 1.7% (paired t-test, $p = 0.001$, $n = 5$ runs per condition).
- Stage-specific optimization: Stage 2 benefited from lower initial learning rates (5×10^{-4} vs 10^{-3}).
- Loss weighting: The Stage 2 loss function $\mathcal{L} = w_H\|H - \hat{H}\|^2 + w_S\|S - \hat{S}\|^2 + w_{C_p}\|C_p - \hat{C}_p\|^2 + w_G\|G - \hat{G}\|^2$ uses property-specific weights (w_H:5, w_S:1, w_{C_p}:10, w_G:1) weighted according to industrial importance: C_p is most critical for process control, H is secondary, while entropy S and Gibbs energy G are less important.

The final configuration (256-dim embeddings, 8 heads, 3 transformer layers, 512-256-128 MLP) achieved consistent performance across all system complexities.

TABLE V
BINARY SYSTEM PERFORMANCE (3 RUNS PER SYSTEM)

System	H	S	C_p	G
NRMSE (% of data range)				
Al_2O_3–SiO_2	7.58±0.58	4.86±0.34	1.62±0.05	6.75±0.48
CaO–Al_2O_3	6.44±0.76	4.68±0.66	2.71±0.26	5.94±0.73
CaO–MgO	0.60±0.10	0.83±0.09	0.85±0.11	0.59±0.09
CaO–SiO_2	6.03±0.43	5.18±0.52	3.70±0.33	5.73±0.49
MgO–Al_2O_3	5.48±0.11	3.81±0.16	1.74±0.17	5.07±0.14
MgO–SiO_2	5.69±0.64	3.65±0.11	2.72±0.27	4.90±0.40
Mean	**5.30**	**3.84**	**2.22**	**4.83**
Phase Estimation				
Mean Phase Acc.	**99.85±0.06%**			
Mean Frac. MAE	**3.18±0.38 pp**			

TABLE VI
TERNARY SYSTEM PERFORMANCE (3 RUNS PER SYSTEM)

System	H	S	C_p	G
NRMSE (% of data range)				
CaO–Al_2O_3–SiO_2	8.00±0.25	4.92±0.15	2.89±0.09	6.83±0.21
CaO–MgO–Al_2O_3	6.47±0.14	4.00±0.06	2.11±0.06	5.62±0.11
CaO–MgO–SiO_2	6.38±0.06	4.29±0.05	2.29±0.02	5.46±0.03
MgO–Al_2O_3–SiO_2	5.99±0.29	3.57±0.19	1.60±0.10	5.21±0.28
Mean	**6.71**	**4.20**	**2.22**	**5.78**
Phase Estimation				
Mean Phase Acc.	**99.50±0.19%**			
Mean Frac. MAE	**4.54±0.33 pp**			

TABLE VII
QUATERNARY SYSTEM (CAO-MGO-AL_2O_3-SIO_2, 3 RUNS)

Metric	R^2 Value	NRMSE (%)
Property Estimation		
Enthalpy (H)	0.869 ± 0.002	9.00 ± 0.18
Entropy (S)	0.953 ± 0.001	5.19 ± 0.06
Heat Capacity (C_p)	0.880 ± 0.023	1.39 ± 0.31
Gibbs Energy (G)	0.917 ± 0.001	7.41 ± 0.06
Overall R^2	$\mathbf{0.904 \pm 0.005}$	
Consistency & Phase Estimation		
G Consistency (MAE)	0.0047 ± 0.0012 J/mol	
Presence Accuracy	$98.88 \pm 0.11\%$	
Fraction MAE	6.41 ± 0.33 pp	
Computational Performance		
Inference Time	72 ± 3 μs/sample	
Training Time	40 ± 0.4 min	

F. Training Configuration

Training used the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with dropout (0.1) and gradient clipping (max norm = 1.0). Batch sizes were 64 for binary systems and 128 for ternary and quaternary systems. Training proceeded for a maximum of 150 epochs with early stopping (patience = 20) based on validation loss.

Data was partitioned into training (64%), validation (16%), and test (20%) sets. For multi-system training, stratified sampling by system identity ensured proportional representation of each complexity level. The validation set was used exclusively for learning rate scheduling and early stopping decisions; all reported metrics are computed on the held-out test set. To ensure reproducibility, all experiments used fixed random seeds, with results averaged over 3–5 independent runs as noted in each table.

The total model comprises approximately 3.2M parameters. Implementation used PyTorch 2.0 with mixed precision training on NVIDIA A100 GPUs.

Stage 1 used binary cross-entropy (BCE) for phase presence and masked MSE for fractions. Since fractions are undefined for thermodynamically absent phases, losses were computed only over phases with ground-truth presence $p_i = 1$, preventing spurious gradients. Stage 2 used weighted MSE (w_H:5, w_S:1, w_{C_p}:10, w_G:1) with analogous masking.

III. RESULTS AND DISCUSSION

A. Binary Systems: Method Development

The architecture was evaluated on all six binary subsystems within the CaO-MgO-Al_2O_3-SiO_2 space. Table V presents results averaged over 3 runs per system, including both R^2 and normalized RMSE (NRMSE) expressed as a percentage of each property's data range.

Binary systems achieved mean $R^2 = 0.962$, with entropy consistently performing best ($R^2 = 0.975$). When expressed as NRMSE, estimation errors represent only 2–6% of the property data ranges, with C_p showing the lowest relative error (mean NRMSE = 2.22%) and H the highest (mean NRMSE = 5.30%). The CaO–MgO system achieved near-perfect estimates ($R^2 = 0.999$, NRMSE $< 1\%$ for all properties), likely due to its simpler phase diagram with fewer competing phases. Phase presence accuracy exceeded 99.7% across all systems. These results establish the architecture's effectiveness on the simplest system complexity prior to scaling to ternary and quaternary systems.

B. Ternary Systems

Training on all four ternary subsystems demonstrated performance scaling with system complexity while maintaining useful accuracy: Table VI summarizes the results. Ternary systems achieved mean $R^2 = 0.945$, with S estimation remaining most accurate ($R^2 = 0.968$, NRMSE = 4.20%). The increase in NRMSE from binary to ternary systems was modest: H increased from 5.30% to 6.71%, while C_p maintained good relative accuracy at 2.22%. This shows controlled accuracy loss with increasing complexity. Phase presence accuracy of 99.5% matched binary performance, demonstrating that the Transformer attention mechanism effectively captures compositional interactions in the three-dimensional composition space.

C. Quaternary System

The full quaternary CaO-MgO-Al_2O_3-SiO_2 system represents the scalability target. Table VII shows that the model maintained strong performance despite increased complexity. The quaternary model achieved overall $R^2 = 0.904$. S continued to yield the best estimates ($R^2 = 0.953$). When contextualized as NRMSE, estimation errors remained practically useful: C_p errors represented only 1.39% of the data range, while H showed the largest relative error at

9.00%. The progression from binary (5.30%) through ternary (6.71%) to quaternary (9.00%) for H NRMSE demonstrates controlled degradation with increasing complexity rather than catastrophic failure. Phase presence accuracy remained high at 98.88% despite the four-dimensional composition space.

D. Scalability Analysis

Table VIII summarizes performance across all system complexities, highlighting the consistent relationship between R^2 and NRMSE metrics. The model demonstrated graceful degradation: R^2 decreased by approximately 0.03 per additional component (binary to ternary: -0.017; ternary to quaternary: -0.041), while NRMSE increased by roughly 1.5 percentage points per component for H. Critically, all NRMSE values remained below 10%, indicating that even quaternary system estimates capture over 90% of the property variation within their respective data ranges.

E. Generalization and Transfer Learning

A critical question for practical deployment is whether models trained on simple systems can transfer to complex ones, or vice versa. This section reports results on zero-shot transfer, demonstrates that physics-informed constraints can partially improve subsystem consistency, and shows that multi-system training offers a viable solution to system generalization.

1) Zero-Shot Transfer Limitations: Zero-shot transfer was systematically evaluated in both directions. Table IX shows that zero-shot transfer failed in both cases. Binary systems exhibit only 2 of 4 components present in the quaternary system. A model trained on quaternary data has likely learned representations for phases that do not exist in binary subsystems, and vice versa. This is a fundamental phase space coverage limitation, not an architecture deficiency.

The key insight from this experiment is that architecture generalizes, but weights do not: the same two-stage transformer-MLP architecture achieves strong performance ($R^2 > 0.90$) across binary, ternary, and quaternary systems when retrained, but zero-shot weight transfer between complexity levels fails catastrophically.

2) Subsystem Consistency: While upward transfer failed, downward transfer (evaluating higher-order models on lower-order subsystems with missing components set to zero) showed partial success. Table X summarizes results across all valid subsystem relationships. Transfer quality degraded with complexity gap: quaternary→ternary achieved $R^2 = 0.84$, but quaternary→binary dropped to $R^2 = 0.42$. Ternary→binary showed intermediate results ($R^2 = 0.67$), suggesting models transfer best to "adjacent" complexity levels. S showed strongest transfer across all tests ($R^2 = 0.71$–0.94), likely because configurational entropy contributions are more universal than enthalpy of mixing effects, which often yield negative R^2 for individual binary systems.

Given this degradation with increasing complexity gaps, the next investigation examined whether physics-informed soft constraints could improve subsystem consistency: Table XI presents the results. Several constraint types were implemented as auxiliary loss terms with tunable weights:

- Gibbs consistency: Enforces the fundamental relation $G = H - TS$ between estimated properties.
- Entropy integration: Requires temperature consistency via $S(T_2) - S(T_1) = \int_{T_1}^{T_2} C_p/T \, dT$.
- Non-negativity: Penalizes physically impossible negative values for heat capacity C_p and entropy S.
- Phase equilibrium: Encourages chemical potential equality across coexisting phases.

Entropy integration provided the largest improvement (+0.057 R^2, an 11.0% relative gain), followed by combining all constraints at moderate weights (+0.054 R^2) and phase equilibrium alone (+0.052 R^2). The improvement was more pronounced for binary subsystems (+0.077 R^2, a 23% relative gain) than ternary (+0.028 R^2, or 3.5%), suggesting physics constraints particularly help bridge larger complexity gaps where the model has less direct training signal. However, excessively strong constraint weights degraded performance (-0.084 R^2), and Maier-Kelley smoothness constraints were harmful for both in-distribution and generalization tasks, suggesting that enforcing specific functional forms conflicts with the flexibility needed for multi-phase systems. These results indicate that physics constraints reinforce thermodynamic relationships learned from data, improving within-domain generalization, but cannot compensate for fundamental phase space coverage gaps.

3) Multi-System Training: Rather than training separate models for each complexity level, training a single model on all systems simultaneously was evaluated. This approach

TABLE VIII
PERFORMANCE SCALING ACROSS SYSTEM COMPLEXITIES

System	NRMSE (%)				Phase	
	H	S	C_p	G	Acc.	MAE (pp)
Binary (6)	5.30	3.84	2.22	4.83	99.85%	3.18
Ternary (4)	6.71	4.20	2.22	5.78	99.50%	4.54
Quaternary	9.00	5.19	1.39	7.41	98.88%	6.41

TABLE IX
ZERO-SHOT TRANSFER RESULTS

Direction	Train → Test	R^2	Phase Presence
Upward	Binary (in-dist)	0.949	99.2%
	Binary → Ternary	-0.23	45.1%
	Binary → Quaternary	-0.89	22.3%
Downward	Quaternary → Ternary	0.37–0.67	65–78%
	Quaternary → Binary	-0.36	22.2%

TABLE X
HIGHER-ORDER MODELS APPLIED TO LOWER-ORDER DATA

Transfer	H	S	C_p	G	Overall
Quat → Ternary (4)	0.74	0.94	0.79	0.87	0.84
Quat → Binary (6)	0.00	0.71	0.53	0.43	0.42
Tern → Binary (12)	0.51	0.84	0.63	0.68	0.67

TABLE XI
EFFECT OF PHYSICS CONSTRAINTS ON SUBSYSTEM CONSISTENCY

Constraint (R^2)	**Binary**	**Ternary**	**Overall**	**Δ**
Baseline	0.328	0.809	0.521	—
Entropy integration	**0.405**	**0.837**	**0.578**	+0.057
All physics (moderate)	0.405	0.829	0.575	+0.054
Phase equilibrium	0.400	0.832	0.573	+0.052
Non-negativity	0.374	0.828	0.556	+0.035
Gibbs consistency	0.371	0.820	0.550	+0.029
Gibbs + entropy	0.364	0.814	0.544	+0.023
Strong Gibbs	0.336	0.822	0.530	+0.009
All physics (strong)	0.201	0.792	0.437	−0.084

TABLE XII
MULTI-SYSTEM TRAINING RESULTS

Configuration	**Quat R^2**	**Binary R^2**	**Phase Presence**
Quaternary only	**0.900**	-0.363^{a}	$22.2\%^{a}$
Multi-system (all)	0.871	0.926	98.8%
Multi + selective feat.	0.885	**0.934**	**99.1%**

[a]Zero-shot transfer (binary not in training data)

TABLE XIII
ERROR ANALYSIS BY PHASE REGION (QUATERNARY SYSTEM)

Region	**H MAE**	**S MAE**	**C_p MAE**	**G MAE**
Phase boundary	2,708	1.25	1.45	5,005
Phase interior	15,900	2.50	1.90	19,027
Ratio	**5.9×**	**2.0×**	**1.3×**	**3.8×**

succeeds because binary and ternary systems occupy compositional subspaces of the quaternary simplex. Training across all complexities exposes the model to regions that quaternary-only sampling covers sparsely, while shared weights act as a regularizer preventing overfitting to complexity-specific patterns. Table XII shows that multi-system training trades only 1.5% quaternary accuracy for dramatically improved subsystem consistency, recovering binary R^2 from -0.363 to 0.934. This approach is recommended when consistent estimates across complexity levels are required.

4) Deployment Implications: For applications requiring estimates across multiple system complexities, multi-system training provides a unified model with consistent behavior. For maximum accuracy on a single system complexity, dedicated models remain preferable.

F. Error Distribution Analysis

To understand where estimation errors concentrate, errors were analyzed by phase region. To distinguish phase boundary from interior regions, samples were classified based on phase fraction distribution: samples where a single phase constitutes $\geq$97% of the system were considered near phase boundaries, while samples with multiple competing phases (none exceeding 97%) were classified as phase region interiors. Table XIII and Fig. 2 present comparative performance results across these regions.

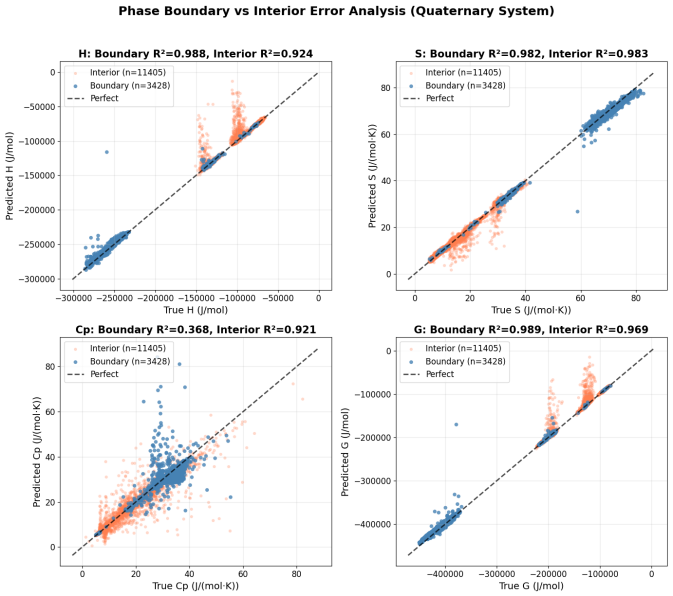

Fig. 2. Estimation accuracy at phase boundaries versus phase interiors (quaternary system). Boundary samples (blue, $n = 3428$) cluster tightly along the parity line, while interior samples (coral, $n = 11405$) show substantially higher scatter. The R^2 degradation is most severe for enthalpy ($0.988 \rightarrow 0.924$) and Gibbs energy ($0.989 \rightarrow 0.969$), while entropy remains consistent across regions (0.982 vs 0.983).

It is evident from Table XIII and Fig. 2 that errors concentrate in phase region interiors where multiple phases coexist, with phase boundary regions estimated at 4–6× higher accuracy for enthalpy and Gibbs energy. In single-phase regions, system properties are determined directly by the phase's equation of state, providing a smooth mapping from composition and temperature to thermodynamic properties. In contrast, multi-phase regions require estimating properties as weighted sums: $P_{total} = \sum_i x_i P_i$, where x_i are phase fractions and P_i are individual phase properties. Uncertainties in both phase fractions and individual phase properties compound, as fraction errors directly scale property errors through the weighted sum. The 5.9× higher enthalpy error in multi-phase interiors (Table XIII) reflects this propagation.

Heat capacity showed the smallest interior/boundary ratio (1.3×), suggesting it varies most smoothly across phase assemblages. Additionally, phase boundaries involve discontinuous changes in the derivative of Gibbs energy with respect to composition, creating sharp transitions that smooth neural network activation functions struggle to capture precisely. These results indicate that targeted improvements in multi-phase fraction estimation could substantially improve overall model performance.

G. Computational Efficiency and Comparison

Training times scaled sub-linearly with complexity: binary (1.7 min), ternary (15 min), quaternary (40 min) on A100 GPU. Inference required $\sim$72 μs per sample regardless of system complexity. For binary systems, where direct calculation requires 1–10 ms [4], this represents approximately 14–140×

speedup. Since direct calculation time increases with a third-order polynomial relationship with system complexity [4], speedups for quaternary and higher-order systems would be substantially greater. Model size remained constant at $\sim$1 MB (3.2M parameters).

Compared to geometric acceleration [4], which achieves up to 20$\times$ speedup for binary systems and up to 1000$\times$ for four- and five-component systems, PINNs offer comparable acceleration with critical scalability advantages. Uniform grid methods require storage scaling as $O(n^C)$ where n is grid resolution and C is the number of components estimated at 4 TB for a four-component system with 600 nodes per dimension [4]. PINNs require $\sim$1 MB regardless of complexity. This positions PINNs for rapid screening and exploratory design where pre-computation becomes prohibitive, while geometric methods remain preferable when highest accuracy is required for well-characterized systems. A rigorous quantitative comparison between the PINN and geometric acceleration approaches across matched systems and accuracy requirements is planned for future work.

H. Limitations and Future Work

The proposed model estimates phase presence and fractions, but not phase identity (e.g., specific polymorphs). Zero-shot weight transfer between complexity levels failed due to phase space coverage limitations, requiring either retraining for each system complexity, or multi-system training for universal models. Errors concentrated at phase interiors (3–4$\times$ higher than phase boundaries), with NRMSE reaching 9% for enthalpy in quaternary systems. Future work could investigate hybrid approaches using PINN estimates to initialize iterative Gibbs energy minimization solvers, potentially combining neural network speed with solver-level accuracy; ensemble methods for uncertainty quantification; extension to variable pressure conditions; and scaling to quinary systems where the primary challenge is training data generation rather than model capacity. Potential high-impact applications include high-throughput slag composition screening for alloy design and real-time equilibrium estimation for pyrometallurgical process control.

IV. Conclusion

This work presented a novel physics-informed neural network approach for thermochemical equilibrium estimation, validated progressively from binary to quaternary oxide systems. The proposed two-stage architecture, combining Transformer attention for phase estimation with MLP regression for properties, proved essential, with task separation improving phase accuracy from 92% to 99.7% versus single-stage alternatives. Systematic ablation studies identified optimal physics incorporation: selective feature engineering (+8.3% R^2) and hard Gibbs-Helmholtz constraints (six orders of magnitude consistency improvement). The architecture generalizes from binary ($R^2 = 0.96$, NRMSE $\approx$ 5%) through ternary ($R^2 = 0.95$, NRMSE $\approx$ 7%) to quaternary systems ($R^2 = 0.90$, NRMSE $\approx$ 9%) with inference requiring $\sim$72 μs/sample, which represents 14–140$\times$ speedup over direct calculation for binary systems, and greater acceleration for higher-order systems while requiring only $O(1)$ storage ($\sim$1 MB). However, zero-shot weight transfer between complexity levels fails due to phase space coverage limitations: the proposed architecture generalizes between systems, but weights do not. Multi-system training provides an effective alternative when subsystem consistency is required. Collectively, these findings establish both the capabilities and limitations of PINNs for thermochemical applications, offering a scalable complement to geometric acceleration for rapid equilibrium estimation in complex multi-component systems.

Acknowledgments

This work was supported by the University of Pretoria and Ex Mente Technologies. The authors thank Ex Mente Technologies for providing the FactSage simulation data used in this study. This paper was prepared with assistance from Claude (Anthropic) for manuscript organization and minor language editing; all experimental work and analysis are the authors' own.

References

[1] H. L. Lukas, S. G. Fries, and B. Sundman, *Computational Thermodynamics: The Calphad Method*. Cambridge University Press, 2007.

[2] W. A. Roos and J. H. Zietsman, "Geometric acceleration of complex chemical equilibrium calculations—algorithm and application to two- and three-component systems," *CALPHAD*, vol. 77, p. 102420, 2022.

[3] W. A. Roos, A. E. J. Bogaers, and J. H. Zietsman, "Geometric acceleration of complex chemical equilibrium calculations—performance in two- to five-component systems," *CALPHAD*, vol. 82, p. 102584, 2023.

[4] W. A. Roos, "Geometric acceleration of complex chemical equilibrium calculations for inclusion in multiphysics and process models," Ph.D. dissertation, University of Pretoria, 2023.

[5] 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.

[6] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.

[7] A. D. Jagtap, K. Kawaguchi, and G. E. Karniadakis, "Adaptive activation functions accelerate convergence in deep and physics-informed neural networks," *J. Comput. Phys.*, vol. 404, p. 109136, 2020.

[8] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, "Recent advances and applications of machine learning in solid-state materials science," *npj Computational Materials*, vol. 5, no. 1, pp. 1–36, 2019.

[9] G. Deffrennes, K. Terayama, T. Abe, and R. Tamura, "A machine learning–based classification approach for phase diagram prediction," *Materials & Design*, vol. 215, p. 110497, 2022.

[10] R. Hammad and S. Mondal, "Advancements in thermochemical predictions: a multi-output thermodynamics-informed neural network approach," Journal of Cheminformatics, vol. 17, Jun. 2025, doi: 10.1186/s13321-025-01033-0.

[11] C. Shen, "The synergy of machine learning and CALPHAD: Revitalizing traditional approaches," Computational Materials Science, vol. 258, p. 113970, Aug. 2025, doi: 10.1016/j.commatsci.2025.113970.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[13] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2019, pp. 2623–2631.

[14] A. ten Cate, B. Geurts, M. Muskulus, D. Köster, A. Muntean, J. van Opheusden, A. Peschansky, A. W. Vreman, and P. Zegeling, "Modeling and simulation of phase-transitions in multicomponent aluminum alloy casting," in *Proc. Sixty-Third European Study Group Mathematics With Industry*, Enschede, The Netherlands, 2008, pp. 117–139.