Bulk Thermal Conductivity Model for Undoped Semiconductor and Dielectric Materials

Brandon Ferraro[1*]

[1]Department of Mechanical Engineering, Stanford University, CA, United States.

Corresponding author(s). E-mail(s): bjf97@stanford.edu;

Abstract

The thermal conductivity of semiconductor/dielectric materials is governed by phonon transport. In pure, bulk semiconductor and dielectric materials, phonons scatter solely on themselves which dictates the rate of heat transport. Thermal conductivity in these materials scale with temperature as $\boldsymbol{T^3}$ and $\boldsymbol{T^{-1}}$ in the regimes where $\boldsymbol{T << \theta_D}$ and $\boldsymbol{T >> \theta_D}$, respectively. In the intermediate region where $\boldsymbol{T \approx \theta_D}$, competing scattering parameters cause high nonlinearity that isn't easily quantifiable with respect to temperature. In this study, a parametric model for $\boldsymbol{k(T)}$ is sought to quantify the thermal conductivity of pure single-crystal semiconductor/dielectric materials over the entire temperature distribution. While we expect the learned model for different materials to have terms that consistently model the physics, it should be noted that the learning model weights will inherently differ based on the stoichiometry of the crystalline materials. Nevertheless, the end goal is to learn a universal form for relating temperature to thermal conductivity for all specified materials, recovering the known non-monotonic scaling behavior and discovering an interpolation for the nonlinear regime.

My analysis pipeline and trained models are freely available at https://github.com/bjf97stanford/ME233Final

Keywords: Semiconductors, Dielectrics, Phonon Scattering, Conduction, Constitutive Neural Network, Non-monotonicity, Single-crystals, Nonlinearity

Nomenclature

Variable	Definition
T	temperature
k	isotropic thermal conductivity
v	mean phonon velocity
θ_D	debye temperature
C_s	heat capacity per unit volume
x_ω	dimensionless integration parameter
τ	relaxation time
$\hbar$	Planck's constant divided by 2π
ω	angular wave frequency
k_B	Boltzmann's constant
n_a	atomic number density

1 Motivation

The thermal conductivity of different solids can vary greatly depending on the microstructure of the material and the energy carriers that facilitate the movement of heat. Single-crystal semiconductor and dielectric materials conduct heat through the transport of phonons which are quantized displacement waves of atoms in their crystalline lattice [1]. While semiconductors, like metals, can conduct heat through the movement of electrons, the contributions of electrons are negligible for low temperatures, specifically in undoped semiconductors [2]. Doping semiconductors decreases the band-gap for these materials, enabling electronic contributions when desired for engineering applications [3]. The band-gap for dielectric materials is so large that electrons are bound to their lattice site at nearly all temperatures. As a result, phonon transport is the dominant energy carrier in pure single-crystal semiconductor/dielectric materials, and analysis of the scattering mechanisms of phonons leads to certain scaling behavior of thermal conductivity with temperature [4, 5].

To determine the relationship between thermal conductivity and temperature for these phonon-dominant semiconductor and dielectric materials, the following phonon thermal conductivity integral is implemented to account for all phonon momentum space (wave-vector space) [2].

$$k(T) = \frac{1}{3} v^2 \int_0^{\theta_D/T} C_s(x_\omega, T)\tau(x_\omega, T) dx_\omega, \quad \text{where } x_\omega = \frac{\hbar\omega}{k_B T} \tag{1}$$

We find that the integral requires known functions for phonon heat capacity per unit volume, C_s, and phonon relaxation time, τ. C_s takes the following form, varying with the third power of temperature for temperatures much smaller than the Debye temperature of the material. For temperatures much greater than the Debye temperature, the heat capacity per unit volume approaches a constant as the cut-off wavenumber is reached when the phonon wavelength approaches the interatomic spacing of the crystalline lattice.

$$C_s = 9 n_a k_B \left(\frac{T}{\theta_D}\right)^3 \frac{x_\omega^4 \exp(x_\omega)}{[\exp(x_\omega) - 1]^2} \tag{2}$$

Lastly, the relaxation time, τ, describes the average time between phonon scattering events [2]. Phonons (and other energy carriers) scatter on other phonons, free electrons, lattice imperfections, grain boundaries, and impurities. By assuming the semiconductors/dielectrics are pure (i.e. undoped), single-crystal materials that are not limited by size/boundary effects, the relaxation time of phonons is limited by phonon-phonon interactions which can be empirically expressed by the following relation.

$$\tau^{-1}(x_\omega, T) = \tau_{phonon-phonon}^{-1} + \cancel{\tau_{phonon-impurity}^{-1}} + \cancel{\tau_{phonon-boundary}^{-1}} \tag{3}$$

$$= A T^\gamma \omega^n \exp(-B x_\omega) \tag{4}$$

Note that the constants A, B, and γ are still a topic of debate among solid-state physicists.

Through implementation of Equations 1, 2, & 4 in conjunction with material property measurements, we observe $k \propto T^3$ for $T >> \theta_D$ and $k\ T^{-1}$ for $T << \theta_D$. Despite these known scaling laws in limiting cases, a global relationship for $k(T)$ is still not precisely known due to the complexity of the integral expression and ambiguity of the relaxation time relationship. These are the reasons that motivate the discovery of a universal model for the thermal conductivity of these materials.

Objectives of this study.

1. Foremost, I am to develop a constitutive neural network that can learn the known physics scaling relationships as well as interpolate the intermediate, nonlinear regime. This step requires the testing of different model architectures, loss functions, and other model parameters.
2. Next, this model will be evaluated using the defined model to determine optimal hyperparameter tuning (e.g. regularization parameters).
3. Lastly, we can compare the tuned model to our known physics and observed data to understand the generalizability and interpretability of the learned model.

2 Methods

2.1 Problem Overview

The goal of this project is to find a general model architecture for modeling the bulk thermal conductivity of pure single-crystal semiconductor/dielectric materials as a function of temperature. There exist physically defined scaling laws for thermal conductivity as a function of temperature for these materials (e.g. T^3 for $T << \theta_D$, T^{-1} for $T >> \theta_D$); this insight leads me to believe that a composition of these scaling behaviors could produce a global thermal conductivity distribution $k(T)$. While the weighting factors of these models will differ according to elemental/molecular composition of the crystal, I expect similar model functions to be learned. Assuming a consistent model is discovered across different semiconductor/dielectric materials, I can then determine if the result is consistent with the underlying physics of phonon transport.

2.2 Assumptions

The semiconductor and dielectrics of interest are restricted to pure (un-doped) single-crystal materials with negligible size effects (bulk). These assumptions reduce phonon scattering to only phonon-phonon interactions with no influence by impurities, defects, and boundaries. This also implies that electronic contributions to the thermal conductivity, specifically for semiconductor materials, are not considered in the underlying physics. Additionally, materials with isotropic thermal conductivity behaviors were chosen due to access to data. As a result, models will be mapping a scalar temperature distribution to a scalar thermal conductivity value; this removes the geometric physical constraints of typical material stiffness models referenced in ME 233.

2.3 Model Architecture

This model will be implemented in Python via Google Colab. The NumPy and TensorFlow libraries will be used for computation functions and machine learning functions, respectively.

The following neural network maps the log of temperature to the log of thermal conductivity. The log-log pre-processing of the data points will be justified in Section 2.5 below. The architecture consists two hidden layers $H_1 = \{\sigma(w_{1,1}((\cdot) - w_{1,2})), (\cdot)\}$ and $H_2 = \{\exp(\cdot), (\cdot), \tan(\cdot)\}$. The following figure shows the graph network which represents the implemented architecture.

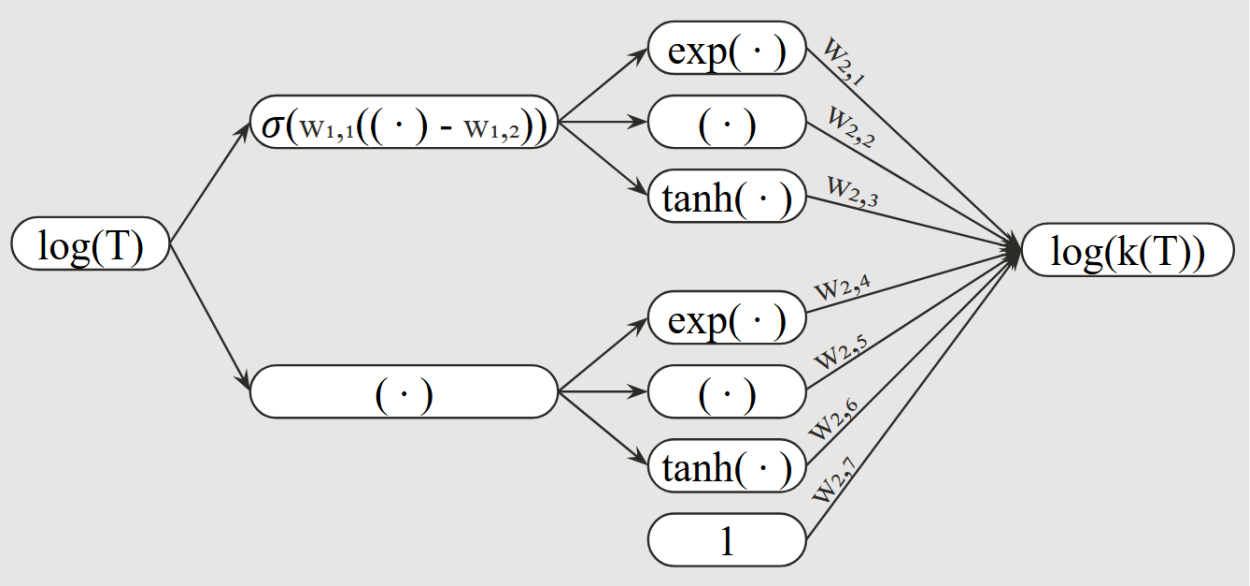


Fig. 1: Constitutive neural network architecture.

This final architecture was reached after many design iterations. Before deciding on log-log preprocessing, 3 hidden layers were used with an intermediate layer for temperature power-law scaling (e.g. $H = \{(\cdot)^{-1}, (\cdot), (\cdot)^2, (\cdot)^3, ...\}$) in order to learn the T^3 and T^{-1} scaling explicitly. However, the log-log preprocessing was implemented for efficiency and accuracy of model training (discussed further in Section 2.5) in addition to increasing the interpretability of model performance by reducing the number of terms from 48 to 9. Additionally, note that the T^3 and T^{-1} scaling behavior will be compared to select learned weights (e.g. $\log(k(T)) = w_{2,5} \log(T) \Rightarrow k(T) = T^{w_{2,5}}$).

The first hidden layer was self-theorized as a method for potentially partitioning the two scaling known scaling regions. Piecewise functions such as Heaviside and ramp functions were considered due to their semi-infinite activation behavior. As a result, I unconventionally implemented a smooth Heaviside function (i.e. sigmoid function) in order to explicitly operate as an activation function which may "learn" the approximate regime where temperature scaling changes. Note that there are two learnable weights for the smooth Heaviside function: the first weight, $w_{1,1}$, controls the steepness of the sigmoid slope and the second weight, $w_{1,2}$, controls the location of the sigmoid function center. A typical Heaviside function was not used because functions must contain non-zero gradients in order to "learn" via gradient descent.

Lastly, the final hidden layer consists of nonlinear functions that ideally help to interpolate the nonlinear $k(T)$ region. Exponential functions were selected according to the types of functions present in Equations 1-4 (note, $\tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$). A bias term, $w_{2,7}$ is also present because the distribution of temperatures does not include $k(T) = 0$ at absolute zero. The finite offset (i.e. intercept) can thus be learned by this bias.

The final form of the nonlinear equation for the full CANN model is detailed below in Equation 5.

$$\begin{aligned}\log(k(T)) = {} & w_{2,1}(\sigma(w_{1,1}(\log(T) - w_{1,2}))) + w_{2,2} \tanh(\sigma(w_{1,1}(\log(T) - w_{1,2}))) \\ & + w_{2,3} \log(\log(T)) + w_{2,4} \exp(\log(T)) \\ & + w_{2,5}(\log(T)) + w_{2,6} \tanh(\log(T)) + w_{2,7}\end{aligned} \tag{5}$$

2.4 Loss function

For my loss function, I am implementing a regression loss due to data as well as an L1 regularization loss to promote sparsity in the model. Ideally, the learned model is expected to have minimal terms (i.e. approximately 2-3) given the known 3 temperature regimes. The hyperparameter α will need to be tuned to

best represent the data and physics. As a result, I get the following loss function.

$$L(\theta; log(T)) = ||\widehat{log(k)} - log(k)(\theta; T)||_2^2 + \alpha||\theta||_1 \tag{6}$$

Here, θ represents the model parameter weights.

2.5 Data

Distributions of thermal conductivity with temperature are reported for pure single-crystal semiconductor/dielectric materials found in current literature [3, 6]. Note, the specific diamond type was chosen for its optimal crystal microstructure.

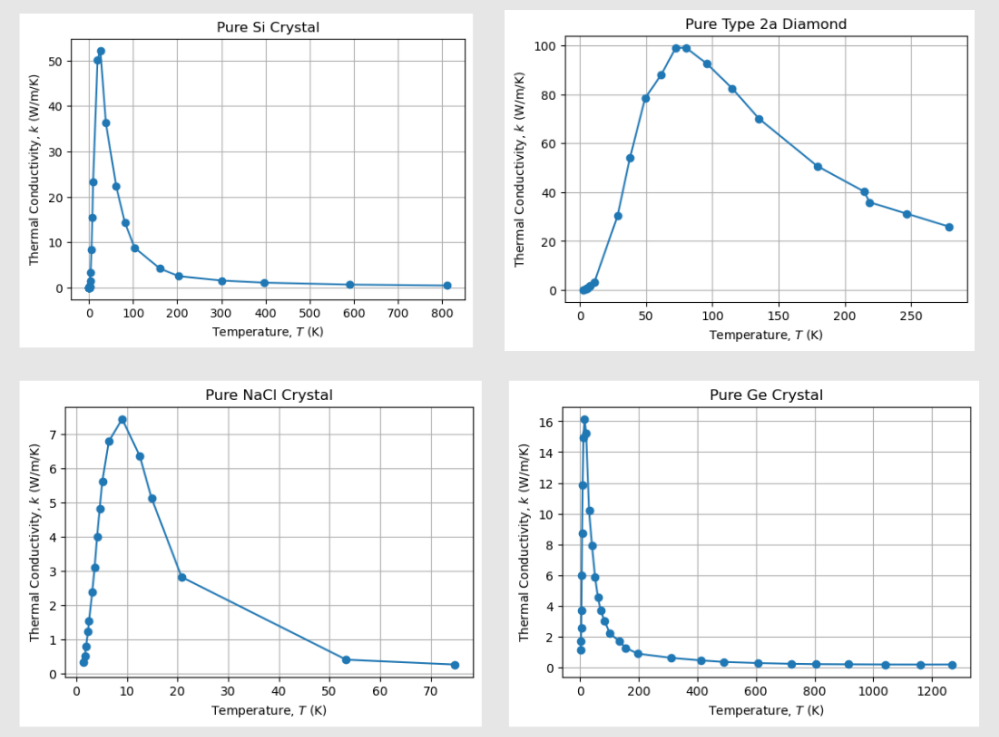


Fig. 2: Raw $k(T)$ data for Si, Diamond, NaCl, and Ge crystals.

As evident from the plots shown above, $k(T)$ distributions are often best visualized on log-log axes due to large differences in the magnitude of the thermal conductivity at different temperatures. This is additionally relevant for model training because scaling the data accordingly improves both numerical stability and computational runtime. After log-log transformation, both of these factors noticeably improved as computational runtime was significantly cut down and R^2 metrics dramatically improved. As a result, the log-log scaling of the data was performed before training the model, and the model architecture was updated as a result of this decision.

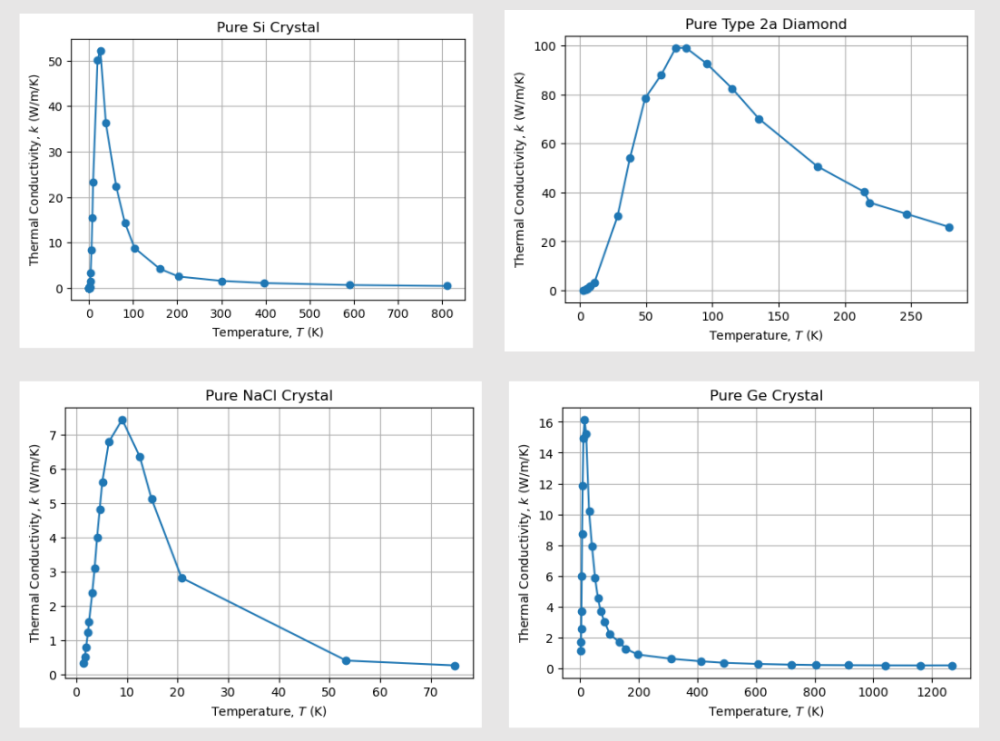


Fig. 3: Log-log transformed $k(T)$ data for Si, Diamond, NaCl, and Ge crystals.

2.6 Analyses

A study for the regularization hyperparameter will need to be done first to determine how to control the sparsity of the learned model. This will be done on the basis of models derived from individual material datasets in addition to a global learned model across all material datasets. The goal for this parameter study is to determine the optimal regularization penalty and to determine if it varies significantly with material. Specifically, the model will be trained for each material and L1 penalty for 10000 maximum epochs with a batch size of 10 and L1 regularization penalties of $\alpha \in \{0, 0.001, 0.01, 0.1\}$.

2.7 Evaluation

Since the derived models will be empirically parameterized by the input data such that element weights may vary from material to material, the R^2 metric would best explain the variance of the thermal conductivity in temperature.

$$R^2 = 1 - \frac{||\log(k) - \widehat{\log(k)}||_2^2}{||\log(k) - \overline{\log(k)}||_2^2} \tag{7}$$

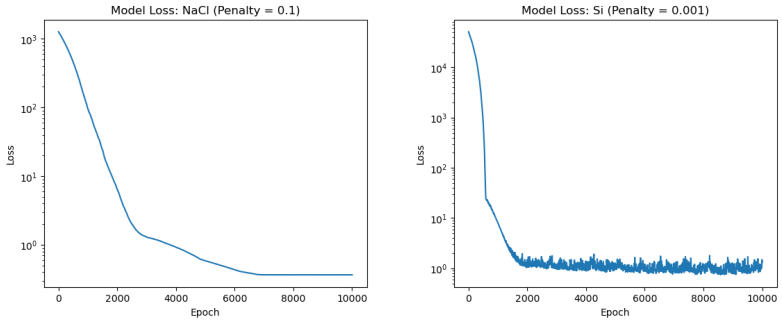
Note, however, the mean-squared error metric is used in the loss function for model optimization.

$$MSE = \frac{1}{n}||\log(k) - \widehat{\log(k)}||_2^2 \tag{8}$$

Furthermore, regarding model optimization parameters, an Adam optimizer from the tensorflow.keras.optimizers library is used with an exponentially decaying learning rate where the initial learning rate is $5e^{-4}$, decay rate is 0.95, and number of steps to decay is 5000. The exponentially decaying learning rate was to ensure numerical stability and prevent oscillatory loss behavior at large epochs.

3 Results

Shown below are two loss profiles among the 16 separately trained models (i.e. 4 materials, 4 hyperparameter values) to show the bounds in numerical stability. On the left is an example of steady minimization of the loss function for NaCl and an L1 regularization penalty of $\alpha = 0.1$. The loss profile for Si and $\alpha = 0.001$, by comparison, shows large oscillation and numerical instability for epoch counts greater than 2000. Data that showed this oscillatory behavior was the major limiting factor for numerical performance when network architecture, optimization rates, etc. were changed.



(a) Training Loss Profile for NaCl data, $\alpha = 0.1$.

(b) Training Loss Profile for si data, $\alpha = 0.001$.

Fig. 4: Example loss profiles.

With the approximate upper and lower bounds of loss performance given by the figures above, respectively, we can now inspect the nonlinear regression performance by looking at all 16 model tests in the figure below.

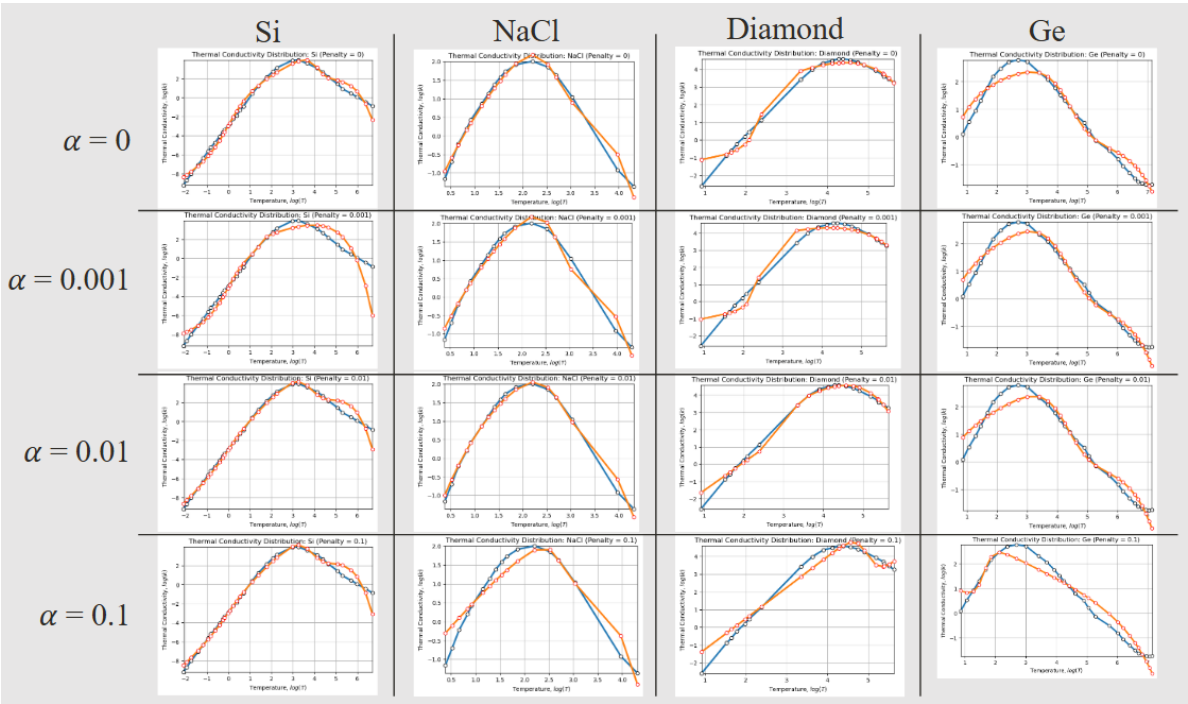


Fig. 5: Model performance for varying material and L1 regularization penalty.

Though the model performance does not vary too strongly with material or L1 regularization, we can observe that the NaCl data performs better fitting among all datasets. Additionally, by slim margins, an L1 regularization parameter of $\alpha = 0.01$ performs optimally given these 16 trials, largely due to the best performance on the diamond dataset. That being said, the model appears to struggle to model specific regions of certain datasets (e.g. Si for large T, Diamond for small T, & Ge for small T). Unexpected contours consistently appear in these notable regions for almost all L1 regularization values. This indicates a potential issue likely with the nonlinear functions in the network function space (e.g. smooth Heaviside function).

	Si	NaCl	Diamond	Ge	Si	NaCl	Diamond	Ge	Si	NaCl	Diamond	Ge	Si	NaCl	Diamond	Ge
	a = 0	a = 0	a = 0	a = 0	a = 0.001	a = 0.001	a = 0.001	a = 0.001	a = 0.01	a = 0.01	a = 0.01	a = 0.01	a = 0.1	a = 0.1	a = 0.1	a = 0.1
w_1,1	3.9384	4.4522	6.0648	2.3931	8.3909	5.6488	6.8676	2.0785	2.9153	4.0872	7.9840	2.3680	2.9267	7.4892	6.4518	6.4599
w_1,2	3.9880	2.6474	2.4755	4.2583	1.5422	2.7429	2.4713	4.2351	3.7523	2.7257	3.2696	4.3753	3.6933	2.6672	4.8974	1.5550
w_2,1	-0.0097	-0.0835	-0.0125	-0.0021	-0.0154	-0.0753	-0.0086	-0.0027	-0.0131	-0.0686	-0.0188	-0.0028	-0.0133	-0.0846	0.0004	-0.0017
w_2,2	1.3637	1.8962	1.0744	0.4097	0.7934	1.6742	0.5762	0.6919	2.1926	1.3862	1.8005	0.6242	2.1732	1.5746	1.7462	-0.5773
w_2,3	2.6462	0.9600	-1.8159	3.0462	3.2878	1.0606	-0.2216	1.9121	1.1258	1.9800	-0.3588	1.1452	1.0472	0.0001	-0.0566	0.0017
w_2,4	-2.0959	-0.6929	-0.5344	-1.5165	-0.2970	-0.8010	-0.3715	-1.9132	-1.7944	-0.9790	-1.1451	-0.7153	-2.9272	-0.7843	-1.8975	1.3959
w_2,5	-0.9297	-1.2446	1.2860	-0.9124	0.3526	0.1753	2.6601	-0.2082	-1.3702	-0.3130	2.7478	-0.7597	-0.0004	-0.0001	-0.0002	0.0019
w_2,6	1.4883	-0.2919	3.7637	-0.0395	1.3911	-1.6817	2.6933	-1.4594	-0.9271	-0.3783	0.9724	-2.5921	-0.0004	-0.0001	0.0003	0.0024
w_2,7	-0.8124	-1.2067	-0.2163	-0.2427	-2.7414	-0.9596	-1.0129	0.6889	-1.1565	-1.1543	-1.8633	0.2745	-0.0003	0.0001	-1.0590	0.0016
R^2	0.9836	0.9804	0.9672	0.9580	0.9097	0.9740	0.9578	0.9628	0.9803	0.9876	0.9864	0.9492	0.9789	0.9036	0.9604	0.9351

Fig. 6: Learned weights and R^2 scores for each trained model.

Furthermore, more information about the model performance can be determined by directly looking at the learned weights and R^2 values detailed in the figure above.

Foremost, let's look at the smooth Heaviside function weights and the weights highlighted in pink. From scaling theory, we expect $w_{2,5} = 3$ and $w_{2,2} = -4$ with the center of the smooth Heaviside function being approximated by $w_{1,2}$. The weights highlighted in pink do not perform how we expect as the values of $w_{2,5}$ and $w_{2,2}$ show high variation and inconsistent sign behavior. The smooth Heaviside center performs slightly more consistently at finding the center of the nonlinear region, yet fails in the three temperature regions noted in the material datasets listed above. Many of the contour inaccuracies align with the location of the largest sigmoid function gradient when comparing the plot figures with the values noted in the table. This error is associated to combined effects of many nonlinear functions in the two hidden layers which are competing to lower the loss residual.

Additionally, it must be noted that model weights are allowed to have negative values due to the non-monotonic correlation of $k(T)$. This complexity, which is not present in the traditional ME 233 bio-material stiffness models, also must be accounted for when analyzing this data. Non-monotonicity introduces many difficulties in the learning process due to (1) potentially more local minima in loss function landscape, (2) negative weights, and (3) reduced interpretability. Terms cannot be visualized as adding together monotonically with rainbow color codes (as is traditionally done in ME 233 models) because there are negative weights. While this presents visualization and interpretation difficulties, this also affects the convexity of the loss landscape, making the model struggle to learn optimal weights for the many nonlinear parameters. This is a likely source of the model error in light of the poor physics interpretability of the model weights.

Moreover, it is important to look at the R^2 values to evaluate the model performance regardless of the physics explainability. For all 16 models, the R^2 value is greater than 0.9 which naively indicates decent performance of the learned model. However, there does not appear to be any reliable correlation between R^2 value and material or L1 regularization penalty. As a result, the stochastic nature of the R^2 metric and poor contour matching for many models indicates unreliable prediction of the general universal form of $k(T)$. Weights in the hidden layers show large unexplained variation that makes the selection of a global model difficult despite all R^2 metrics indicating decent predictive performance.

Lastly, it should be noted that the L1 regularization struggled to induce sparsity into the model. We only observe near 0 weights for $\alpha = 0.1$, resulting in the learned model shown below.

$$\log\left(k(T)\right) = w_{2,2}\tanh\left(\sigma(w_{1,1}(\log\left(T\right) - w_{1,2}))\right) + w_{2,4}\exp\left(\log\left(T\right)\right) \tag{9}$$

This model is derived by pruning weights that have at least 3/4 near 0 weight among all 4 materials. This model results in a thermal conductivity which varies with temperature according to $k(T) \propto T + \exp\left(-T\right)$.

Because this model fails to draw interpretation from phonon transport theory in addition to the model errors listed above, I fail to conclude that this model effectively captures the nonlinear relationship between the bulk thermal conductivity of undoped semiconductor/dielectric materials and temperature.

4 Discussion

Overall, I believe this model failed to meet most of the original objectives due to numerous compounding factors. Despite decently high R^2 metrics, I observed unpredictable contours due to competing nonlinear terms and likely non-convexity in the loss function due to non-monotonicity of the underlying relationship. Additionally, analysis of the model weights showed more unexplained variability in magnitude and sign. For this reason, the physical interpretability is model is poor, especially in spare form as seen in Equation 9 when $\alpha = 0.1$. Due to these many competing factors, there are many suggestions I would make for further work. Additionally, there are many unique challenges for non-monotonic functions that would be interesting to further explore by future students.

One primary source of error lies in the datasets themselves. Some datasets have limited datapoints in either the high or lower temperature regimes. A large majority of the datapoints for each material lie in the nonlinear region which most definitely hindered the recovery of the known physical scaling behaviors in the low/high temperature regimes. Additionally, the ability of the smooth Heaviside function to distinguish these two linear regions was likely impeded by insufficient datapoints at the extrema of temperature ranges.

Another large source of error is the function space selection for the hidden layers after log-log data preprocessing. One factor I did not consider when designing the model architecture was how the interpretability of the nonlinear functions in the second hidden layer would be affected by the log-log data transformation. While the log-log operation was successful improving model training efficiency/stability, the composition of log() functions with additional nonlinear functions severely affected the interpretability of the learned model in equation form. This issue coupled with weights not being restricted to positive domains made the model difficult to modify during testing while also being difficult to interpret the results of. One step I wish I took prior to implementation of the hidden layers would have been to perform more research on a more complete derivation of $k(T)$, if existent, or derivation using approximation methods (e.g. series approximations). Getting a more clear idea of the function space, especially with a log-log transformation, would help in redesigning the model architecture.

Lastly, I think a more complete investigation of L1 regularization could be completed with better datasets. Testing L1 regularization values of $\alpha = 0.5$ may provide a more complete understanding of how sparsity can be implemented to achieve more interpretable results. The numerical stability of the model will need to be addressed, as increasing the L1 norm affects the loss convergence. With more time for model training, using training over more epochs with smaller learning rates may also improve numerical stability. Hopefully we these modifications, we can recover the known physics scaling laws by inducing sparsity gradually for a long period. For a proper value of α, overfitting should not be a concern with this approach.

Acknowledgments

Authors' contributions Aside from initial papers and direction provided by Dr. Kenneth Goodson, all work was completed by Brandon Ferraro. Lectures and coursework in ME 233 (Automated Model Discovery) & ME 352B (Heat Conduction) directly inspired this work.

References

[1] K. E. Goodson, M. I. Flik, L. T. Su, and D. A. Antoniadis, "Prediction and Measurement of the Thermal Conductivity of Amorphous Dielectric Layers," Journal of Heat Transfer, vol. 116, no. 2, pp. 317–324, May 1994, doi: https://doi.org/10.1115/1.2911402.

[2] K. Goodson, Heat Conduction Physics and Properties. LAD Custom Publishing, Inc., 2025.

[3] Vandersande, J. W., and C. Wood. 'The Thermal Conductivity of Insulators and Semiconductors'. Contemporary Physics, vol. 27, no. 2, Informa UK Limited, Mar. 1986, pp. 117–144, https://doi.org/10.1080/00107518608211003.

[4] D. G. Cahill et al., "Nanoscale thermal transport," Journal of Applied Physics, vol. 93, no. 2, pp. 793–818, Jan. 2003, doi: https://doi.org/10.1063/1.1524305.

[5] M. Asheghi, Y. K. Leung, S. S. Wong, and K. E. Goodson, "Phonon-boundary scattering in thin silicon layers," Applied Physics Letters, vol. 71, no. 13, pp. 1798–1800, Sep. 1997, doi: https://doi.org/10.1063/1.119402.

[6] J. E. Graebner, "Thermal Conductivity of Diamond," Diamond: Electronic Properties and Applications, pp. 285–318, 1995, doi: https://doi.org/10.1007/978-1-4615-2257-7_7.