
Differentiable Interference Modeling for Cost-Effective Growth Estimation of Thin Films

Leonard Storcks

AstroAI Lab

Interdisciplinary Center for Scientific Computing
Heidelberg University

leonard.storcks@iwr.uni-heidelberg.de

Gunnar Ehlers

Chair for Composite Materials
Institute for Materials Science
Faculty of Engineering
Kiel University
gueh@tf.uni-kiel.de

Robin Janssen

HAWAII Lab

Institute of Computer Engineering
Heidelberg University

robin.janssen@stud.uni-heidelberg.de

Konrad Storcks

Chair for Composite Materials
Institute for Materials Science
Faculty of Engineering
Kiel University

Tayebbeh Ameri

Chair for Composite Materials
Institute for Materials Science
Faculty of Engineering
Kiel University
tam@tf.uni-kiel.de

Tobias Buck

AstroAI Lab
Interdisciplinary Center for Scientific Computing
Heidelberg University
tobias.buck@iwr.uni-heidelberg.de

Abstract

Accurate *in situ* monitoring of thin-film growth typically requires complex and expensive techniques such as ellipsometry or X-ray reflectivity. We introduce `reflax`, an open-source library to infer growth dynamics from simple, low-cost single-wavelength reflectance time series. `reflax` employs a differentiable physics-based simulator in which the film’s growth behavior is parameterized by a neural network, minimizing the discrepancy between simulated and experimental reflectance while incorporating physical priors such as growth monotonicity and smoothness for model robustness. To reduce the computational burden introduced by optimizing physical parameters in the simulator, we propose *neural operator initialized optimization*, where a neural operator provides a strong initial estimate of the growth function, which can be fine-tuned efficiently. First experimental tests show promising agreement with ellipsometry in the predicted final thickness, suggesting that accurate thickness monitoring can be achieved at much lower cost.

1 Introduction

»In ML, the considerations are almost all at the level of the data...«, note Hogg and Villar in [1]. In contrast, the natural sciences seek latent structure. We can bridge both worlds by embedding ML models in differentiable simulators, where learned terms gain physical meaning. Such approaches have been applied, e.g., to drag-force modeling in robotics [2].

We demonstrate our library `reflax` on interferometric monitoring of dielectric films grown by initiated chemical vapor deposition [3, 4, 5]. In this process, monomers polymerize on a cooled

substrate, and film thickness can be tracked *in situ* via interference of reflections at the film surface and the film–substrate interface [6]. `reflax` enables compact, low-cost thin-film monitoring. Existing methods face a trade-off: low-cost tools like quartz crystal microbalances suffer from large errors and limited lifetimes [7], while high-accuracy methods such as ellipsometry or X-ray reflectivity are expensive, bulky, and difficult to integrate [8, 9, 10, 11]. Interferometry addresses both issues but remains underutilized due to complex signal interpretation and the lack of ready-to-use systems.

Our thickness model is an »algorithmic model« in Breiman’s terminology [12], i.e., the neural network parameters lack a direct physical interpretation. However, the predicted thickness gains physical meaning through its placement in the simulator, and we incorporate inductive biases such as monotonicity (ensuring uniqueness) and smoothness (which leads to robustness against noise; see Section 3). Three challenges remain: (I) multiple learned quantities may be degenerate, which we address by carefully selecting measurements and model terms. (II) inaccurate simulators can yield wrong physical terms despite good fits, but in our case costly reference measurements allow verification and adaptation. (III) optimization through simulators can be computationally demanding, for which we introduce *neural operator initialized optimization* (Section 2).

2 Differentiable film growth models

Modelling task. Interferometry can be modeled either by ray propagation, summing individual ray contributions to estimate total reflectance, or by the 1D transfer matrix method (TMM), which tracks electric fields across stratified media. We implement (i) a single-layer Fabry–Pérot model [13] and (ii) TMM in Rumpf’s scattering-matrix formulation [14], which is computationally robust. Both are detailed in Appendix A.2. For fixed optical parameters, the reflectance r is a function of the thickness d , $r = f(d)$. The inverse problem—recovering $d(t)$ from $r(t)$ —is ill-posed unless $d(t)$ is constrained to be continuous and monotonic. We thus define an operator $\mathcal{R} : \mathcal{D} \rightarrow \mathcal{S}$ mapping admissible thickness functions \mathcal{D} to reflectance signals \mathcal{S} . Our task is to approximate \mathcal{R}^{-1} , with continuity and smoothness improving robustness against noise. Because only a few oscillations are typically recorded, we require models that incorporate the correct functional form of the signal; generic instantaneous frequency methods such as ChirpGP [15] are ill-suited without such priors.

Implementation. We provide both models in our open-source library `reflax`¹, which, being based on JAX, supports GPU scaling, forward/backward differentiation [16], and growth recovery tools tailored to thin films. While `tmmax` [17] implements TMM in JAX, `reflax` is specialized for growth inference and also includes the simpler single-layer model.

To obtain a unique solution, the thickness must be monotonic with known initial value ($d(0) = 0$). We model $d(t)$ with a neural network predicting growth rates, forced positive by a softplus activation and integrated via the trapezoidal rule. This yields a monotonic thickness profile from which the simulator generates a reflectance signal. The predicted and measured signals are compared with an MSE loss, optimized using Adam [18] (see Fig. 1). While numerical integration introduces small rate inaccuracies, thickness recovery remains reliable, and comparisons to ground truth are made on the predicted growth rate. Currently we assume a known refractive index. When amplitude as well as phase information is included (see Appendix A.2.1), both thickness and refractive index can be inferred.

Neural operator initialized optimization. Initialization of the growth network strongly influences optimization. We consider three strategies: (i) *random*, where network parameters are randomly initialized and scaled to yield reasonable growth outputs; (ii) *linear*, where the network is preset² or pre-trained to produce a linear thickness profile; and (iii) *neural operator initialized optimization*, where a neural operator is trained to approximate $\mathcal{R}^{-1} : \mathcal{S} \rightarrow \mathcal{D}$, mapping reflectance $r(t)$ to thickness $d(t)$. Its prediction serves as an informed initialization for the growth network, which is then refined by the differentiable simulator. Our baseline neural operator is a two-layer feed-forward network (512 neurons per layer); additional architectures (DEEPONET [19], hybrid) are compared in Appendix A.5, with hyperparameters tuned by OPTUNA [20]. This strategy combines the fast but

¹<https://github.com/leo1200/thin-films>

²The linear pre-set approach achieves a linear prediction by setting network weights to zero and choosing an appropriate bias term for the output layer, yielding a constant growth rate and hence linear growth.

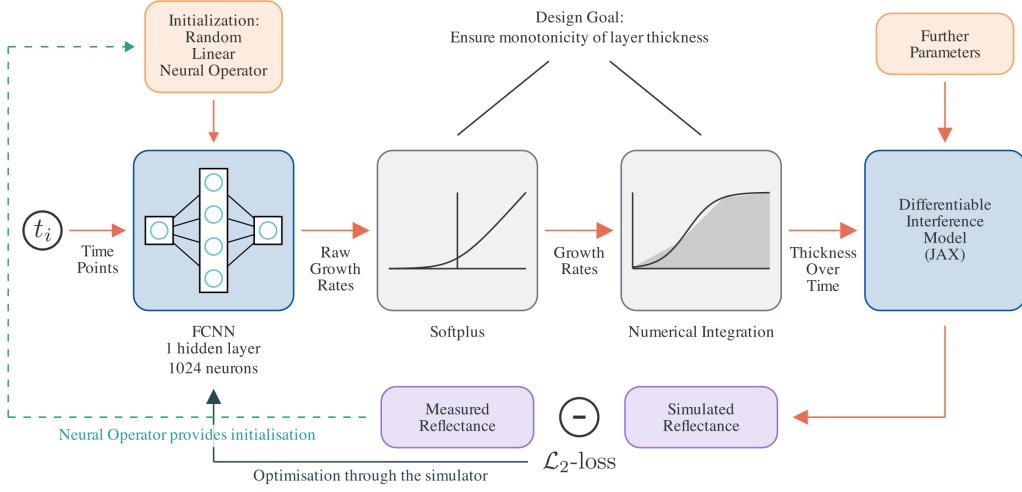


Figure 1: Optimization pipeline in `reflax` for thin-film growth inference. Growth rates predicted by a fully connected neural network (NN) are made positive via a softplus activation and integrated (trapezoidal rule) to yield a monotonic thickness profile. A differentiable interference model (JAX-based) simulates the corresponding reflectance. The NN parameters are optimized with Adam to minimize MSE between simulated and measured reflectance. Initialization of the NN (random, linear, or neural operator) precedes optimization.

uninformed guesses of neural operators with the physically grounded refinement of differentiable simulators, akin to warm-starting numerical solvers [21].

3 Experiments

Training and validation data. For thin-film depositions in our experimental setup on the one-hour timescale (parameters of the experimental setup can be found in Table 1), growth rates are non-decreasing, so we restrict to convex thickness functions.³ Training data is generated from smooth convex time series sampled via Gaussian Processes (GPs, see Section A.3), with an RBF kernel of length-scale $l = 0.4$. Each sample consists of $n_{\text{eval}} = 100$ points, scaled to a final thickness between $d_{\text{final,min}} = 800$ and $d_{\text{final,max}} = 1200$ nm. Kernel variance v is adapted to cover growth rates from 200–1800 nm/h. We generate 2000 such samples plus 100 constant-growth samples, and compute the corresponding reflectance signals using the single-layer model. Validation data consists of 200 convex samples with shorter RBF length-scale ($l = 0.1$), imitating more variable experimental conditions. Appendix Fig. 4 provides a visualisation of training and validation data.

Results All evaluations below use the validation data unless stated otherwise. The neural operator is trained for 20,000 epochs with Adam. Linear initializations assume a final thickness of 1000 nm, representing a plausible user guess within the 800–1200 nm range. Optimizations were run for 14,000 epochs and required about five minutes per run. Since no significant differences were observed between single-layer and transfer-matrix models, we report only the former.

Across the validation set, linear initialization frequently fails to converge to the true growth profile, with failure rates of about 28% for the pre-trained variant. Neural operator initialization, in contrast, consistently converges to accurate thickness estimates, eliminating these failure modes. As an illustration, Appendix Fig. 5 shows an example case where linear initialization fails while the neural operator initialization succeeds.

Fig. 2 summarizes quantitative comparisons. Kernel density estimates of reflectance, thickness, and growth-rate errors demonstrate that only neural operator initialization avoids a bimodal loss distribution with a high-loss mode, which is observed for all other methods. On average, the neural

³In later experiments this assumption did not hold up and we adapted the neural operator training accordingly. This should not impact the general results, but will be corrected in the future.

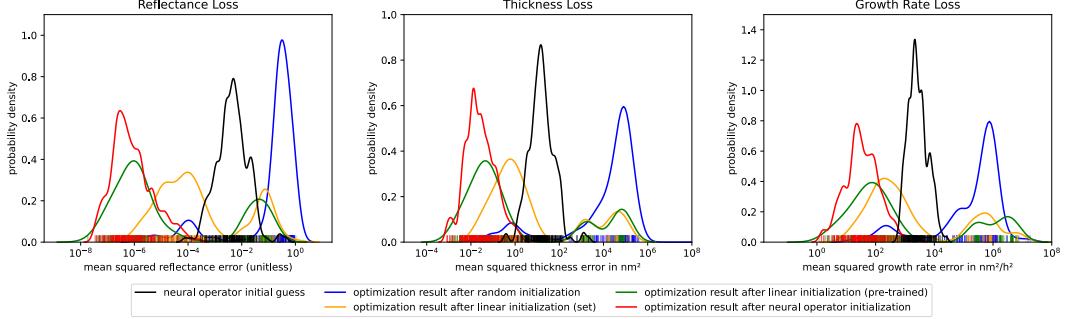


Figure 2: Kernel density estimates of reflectance, thickness, and growth rate MSEs across 200 validation samples. Neural operator initialization avoids the high-loss secondary peak seen for other methods and the subsequent optimization yields strong improvements over its own direct prediction.

operator’s initial guess is refined by 3–4 orders of magnitude through simulator-based optimization. Random initialization consistently performs worst, while linear pre-trained initialization is competitive when successful but unstable overall. In successful cases, growth rate errors remain minor (RMSE $\leq 6 \text{ nm/h}$, well below the 200–1800 nm/h scale).

Convergence analysis confirms these findings: optimizations initialized by the neural operator start closer to the correct solution and maintain a significant advantage throughout early epochs, while linear initialization often stalls at high reflectance loss. The loss trajectories corresponding to this observation are visualized in Appendix Fig. 6.

Neural operator initialization also improves robustness to measurement noise. Even when adding Gaussian noise corresponding to signal-to-noise ratios of ~ 37 , 23 , and 17 dB , thickness errors remain below 5 nm ($< 0.5\%$), with deviations mostly at the boundaries of the signal. An example is shown in Appendix Fig. 7. This robustness can be attributed to the spectral bias of ReLU networks toward smooth, low-frequency components [22], which aligns with the expected growth behavior.

Finally, we demonstrate a first application to laboratory data (Fig. 3). The recovered reflectance fits well (RMSE $< 3\%$ of signal amplitude), but discrepancies at early times suggest surface roughness effects not captured by our current model, leading to unphysical growth rates. Additionally, a first experimental comparison against ellipsometry further supports our approach: across several iCVD depositions with varying monomer shares, the final thickness inferred by our cost-effective setup deviates from ellipsometry by only 0.2–3.2% (cf. Appendix A.6, especially Table 3).

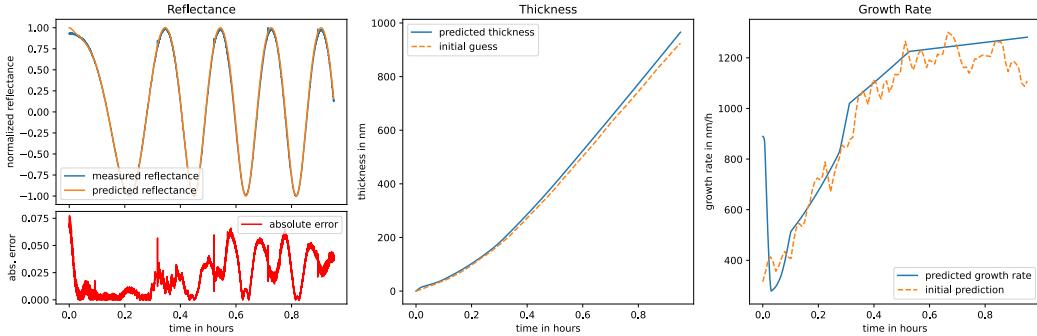


Figure 3: First application to laboratory data. Thickness profile and growth rate inferred from measured reflectance using neural operator initialization and simulator optimization. Early-time deviations point to missing surface roughness modeling.

4 Discussion

We presented `reflax`, an open-source library for inferring thin-film growth from low-cost single-wavelength reflectance time series. Our approach combines a differentiable physics-based simulator

with a neural network thickness model incorporating physical priors such as monotonicity and smoothness. A key contribution is *neural operator initialized optimization*, where a neural operator provides a robust initial guess subsequently refined by the simulator. This substantially improves convergence and stability compared to standard initializations, avoids common failure modes, and yields accurate growth recovery. We further demonstrated robustness to noise and a first successful application to real-world measurements.

Despite these advances, several limitations remain. Correct thickness recovery requires an accurate forward model; discrepancies observed in experiments suggest the need to model surface roughness. Current validation is limited to convex growth profiles and single-layer setups, chosen for simplicity, though we expect similar behavior with multilayer growth. Neural operator initialization assumes fixed measurement duration and must be retrained for each optical setup, restricting real-time monitoring, though training is fast (< 30 s on an A100).

Future work will focus on refining the physical model, extending the framework to variable-length time series for real-time applications, and validating against established *in situ* techniques such as ellipsometry. Beyond thin films, the paradigm of combining neural operators with differentiable simulators may benefit broader domains, from industrial monitoring to astrophysical modeling.

References

- [1] David W. Hogg and Soledad Villar. »Position: is machine learning good or bad for the natural sciences?« In: *Proceedings of the 41st International Conference on Machine Learning*. ICML'24. Vienna, Austria: JMLR.org, 2024.
- [2] Eric Heiden et al. *NeuralSim: Augmenting Differentiable Simulators with Neural Networks*. 2021. URL: <https://arxiv.org/abs/2011.04217>.
- [3] Stefan Schröder et al. »Nanoscale gradient copolymer films via single-step deposition from the vapor phase«. In: *Materials Today* 37 (2020), pages 35–42.
- [4] Wyatt E Tenhaeff and Karen K Gleason. »Initiated and oxidative chemical vapor deposition of polymeric thin films: iCVD and oCVD«. In: *Advanced Functional Materials* 18.7 (2008), pages 979–992.
- [5] Kenneth KS Lau and Karen K Gleason. »Initiated chemical vapor deposition (iCVD) of poly (alkyl acrylates): an experimental study«. In: *Macromolecules* 39.10 (2006), pages 3688–3694.
- [6] Brett Cruden et al. »Thermal decomposition of low dielectric constant pulsed plasma fluorocarbon films: I. effect of precursors and substrate temperature«. In: *Journal of The Electrochemical Society* 146.12 (1999), page 4590.
- [7] Sawit Na Songkhla and Takamichi Nakamoto. »Overview of quartz crystal microbalance behavior analysis and measurement«. In: *Chemosensors* 9.12 (2021), page 350.
- [8] G Jungk. »Possibilities and limitations of ellipsometry«. In: *Thin Solid Films* 234.1-2 (1993), pages 428–431.
- [9] Harland G Tompkins and James N Hilfiker. *Spectroscopic ellipsometry: practical application to thin film characterization*. Momentum Press, 2015.
- [10] Brian K Tanner. »Grazing incidence X-ray reflectivity and scattering«. In: *Handbook of Advanced Nondestructive Evaluation*. Springer Verlag, 2018.
- [11] David L Gil and Donald Windover. »Limitations of x-ray reflectometry in the presence of surface contamination«. In: *Journal of Physics D: Applied Physics* 45.23 (2012), page 235301.
- [12] Leo Breiman. »Statistical modeling: The two cultures (with comments and a rejoinder by the author)«. In: *Statistical science* 16.3 (2001), pages 199–231.
- [13] Charles Fabry. »Theorie et applications d'une nouvelle methods de spectroscopie interreferentielle«. In: *Ann. Chim. Ser. 7* 16 (1899), pages 115–144.
- [14] Raymond C Rumpf. »Improved formulation of scattering matrices for semi-analytical methods that is consistent with convention«. In: *Progress In Electromagnetics Research B* 35 (2011), pages 241–261.
- [15] Zheng Zhao et al. »Probabilistic Estimation of Instantaneous Frequencies of Chirp Signals«. In: *IEEE Transactions on Signal Processing* (2023). In press.
- [16] Atilim Gunes Baydin et al. »Automatic differentiation in machine learning: a survey«. In: *Journal of machine learning research* 18.153 (2018), pages 1–43.
- [17] Esra Zayim Bahrem Serhat Danis. *tmmax: transfer matrix method with jax*. Version 1.0.0. 2025. URL: <https://github.com/bahremsd/tmmax>.
- [18] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. URL: <https://arxiv.org/abs/1412.6980>.
- [19] Lu Lu et al. »Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators«. In: *Nature Machine Intelligence* 3.3 (Mar. 2021). Publisher: Nature Publishing Group, pages 218–229. DOI: 10.1038/s42256-021-00302-5.
- [20] Takuya Akiba et al. *Optuna: A Next-generation Hyperparameter Optimization Framework*. July 25, 2019. DOI: 10.48550/arXiv.1907.10902.
- [21] Xu-Hui Zhou et al. »Neural operator-based super-fidelity: A warm-start approach for accelerating steady-state simulations«. In: *Journal of Computational Physics* 529 (2025), page 113871. DOI: <https://doi.org/10.1016/j.jcp.2025.113871>. URL: <https://www.sciencedirect.com/science/article/pii/S0021999125001548>.
- [22] Nasim Rahaman et al. *On the Spectral Bias of Neural Networks*. 2019. URL: <https://arxiv.org/abs/1806.08734>.
- [23] Olaf Stenzel et al. *The physics of thin film optical spectra*. Volume 18. Springer, 2015.
- [24] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*. Volume 2. 3. MIT press Cambridge, MA, 2006.

- [25] Samuel Lanthaler, Siddhartha Mishra, and George Em Karniadakis. *Error estimates for DeepOnets: A deep learning framework in infinite dimensions*. Jan. 13, 2022. doi: 10.48550/arXiv.2102.09618.
- [26] Aaron Defazio et al. *The Road Less Scheduled*. Oct. 29, 2024. doi: 10.48550/arXiv.2405.15682.

A Technical Appendices and Supplementary Material

A.1 Supplementary figures

Fig. 4 illustrates the training and validation data used throughout our experiments, showing example traces and the increased variability in the validation set. Fig. 5 exemplifies how linear initialization can fail while neural operator initialization converges successfully. Full convergence behavior across all validation samples is shown in Fig. 6, highlighting the faster and more stable progress with neural operator initialization. Finally, Fig. 7 provides an additional example of robustness against measurement noise, where thickness errors remain below 5 nm even at low signal-to-noise ratios.

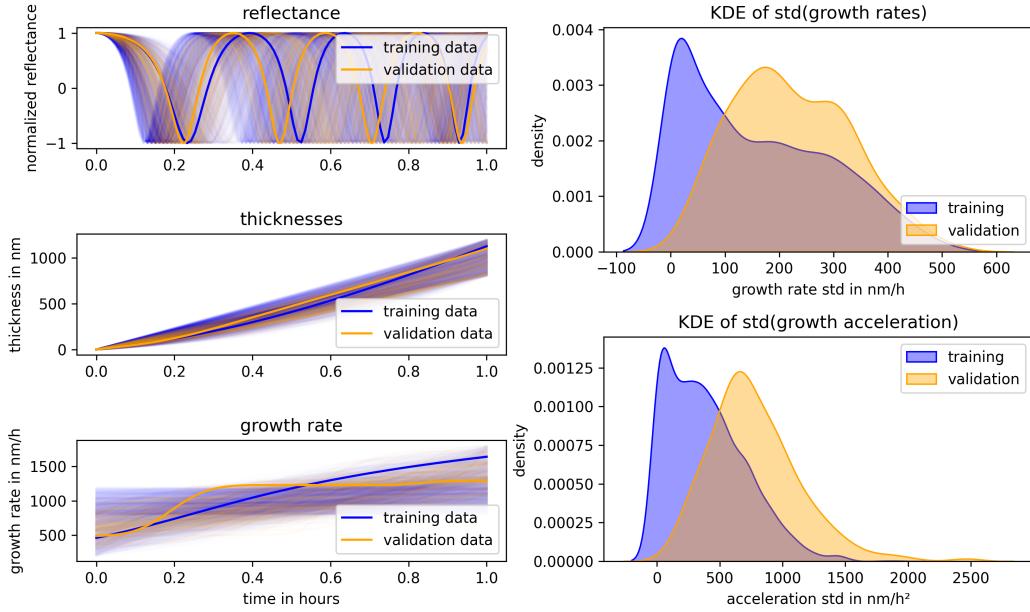


Figure 4: Training and validation data for neural operator initialization and testing of the thickness profile estimation pipeline. Left: example reflectance, thickness, and growth rate traces, generated by sampling convex thickness profiles and simulating reflectance with our model. Right: KDEs of per-sample growth-rate (top) and acceleration (bottom) standard deviations, used as proxies for time-series variability. Validation data shows higher variability due to shorter GP length-scale.

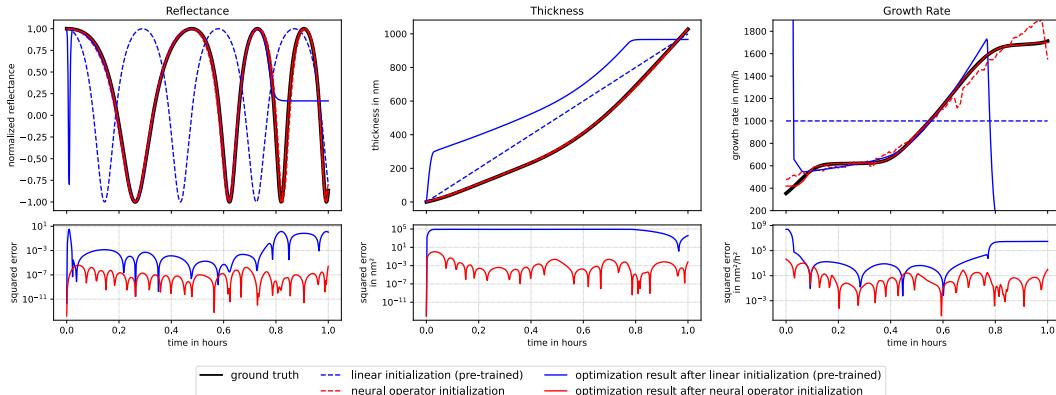


Figure 5: Example from the validation set. Left: reflectance fit. Center: recovered thickness. Right: predicted growth rate. Linear initialization (pre-trained) fails in this case, while neural operator initialization converges to the ground truth. Such failures occurred in $\sim 28\%$ of the 200 validation samples.

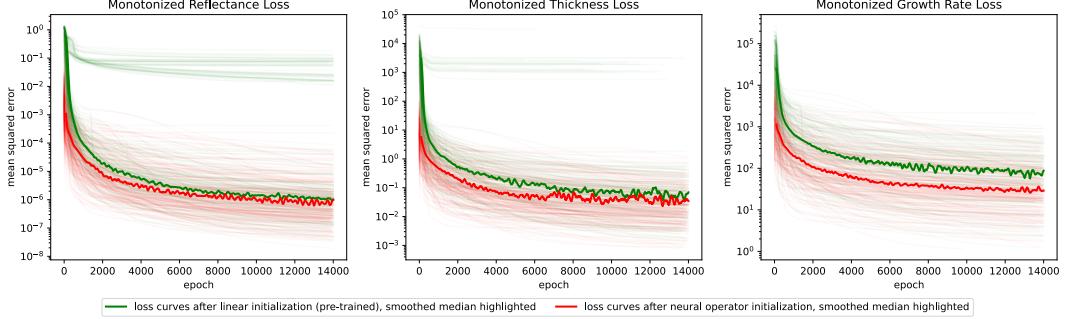


Figure 6: Reflectance, thickness, and growth rate losses over optimization epochs for linear (pre-trained) vs. neural operator initialization across all validation samples. Loss curves are monotonized for clarity; the smoothed median (100-epoch window) excludes failed runs (final thickness MSE > 100 nm²). Neural operator initialization provides faster convergence and avoids failure modes.

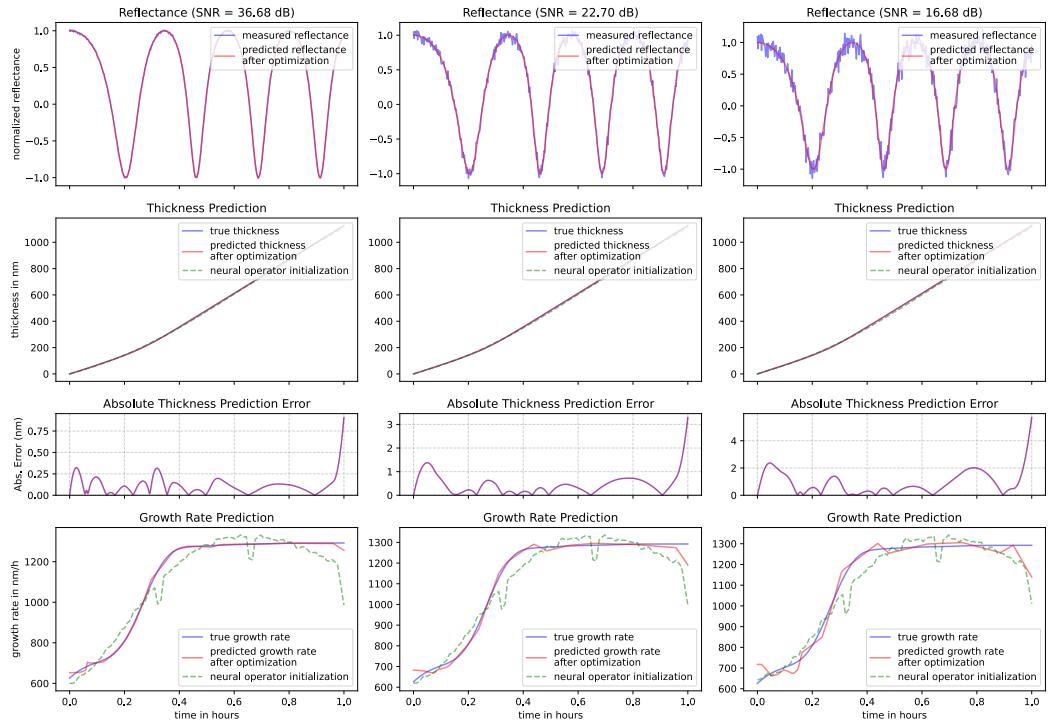


Figure 7: Noise robustness. Results after neural operator initialization and optimization for simulated reflectance signals with different SNRs ($\sim 37, 23, 17$ dB). Thickness errors remain < 5 nm even at lowest SNR.

A.2 Interference models

A.2.1 Single layer model

In its essence, the ray propagation approach used in this work for the single-layer model is based on a Fabry-Pérot etalon [13]. Using the geometric series summation, the condensed derivation of the

reflectance is [23]:

$$r_{012} = r_{01} + t_{01}e^{i\delta}r_{12}e^{i\delta}t_{10} + t_{01}e^{i\delta}r_{12}e^{i\delta}r_{10}e^{i\delta}r_{12}e^{i\delta}t_{10} + \dots \quad (1)$$

$$= \frac{r_{01} + r_{12}e^{2i\delta}}{1 + r_{01}r_{12}e^{2i\delta}} \quad (2)$$

with r_{ij} and t_{ij} as Fresnel reflection and transmission coefficients for the interface between materials i and j (0 = environment, 1 = growing layer, and 2 = substrate). The factor $e^{i\delta}$ constitutes the phase of the interferogram

$$\delta = \frac{2\pi d}{\lambda_0} \sqrt{n_1^2 - n_0^2 \sin^2(\theta)}, \quad (3)$$

which, for thin films with negligible attenuation, is the only part of the reflectance formula that depends on the film thickness d , whereas the amplitude is solely related to Fresnel reflection coefficients, hence independent of d :

$$A = |r_{\max}| - |r_{\min}| = \left| \frac{r_{01} + r_{12}}{1 + r_{01}r_{12}} \right| - \left| \frac{r_{01} - r_{12}}{1 - r_{01}r_{12}} \right|. \quad (4)$$

The reflectance takes the form of a complex Möbius transformation, mapping the phase factor $e^{2i\delta}$ onto a corresponding complex reflection coefficient, thereby producing the interference oscillations.

A.2.2 Multi-layer transfer matrix model

We also implemented the TMM based on R. Rumpf's improved formulation [14], modeling the propagation of complex electric field amplitudes through layers via scattering matrices (SMs) more efficiently and computationally robustly than other formulations of the TMM. It begins with solving Maxwell's curl equations with normalized magnetic fields $\tilde{H}_{x/y}$, leading to coupled differential equations for the transverse electric and magnetic field components

$$\frac{d}{dz'} \begin{bmatrix} E_x \\ E_y \end{bmatrix} = \mathbf{P} \begin{bmatrix} \tilde{H}_x \\ \tilde{H}_y \end{bmatrix}, \quad \frac{d}{dz'} \begin{bmatrix} \tilde{H}_x \\ \tilde{H}_y \end{bmatrix} = \mathbf{Q} \begin{bmatrix} \tilde{H}_x \\ \tilde{H}_y \end{bmatrix} \quad (5)$$

where \mathbf{P} and \mathbf{Q} are auxiliary matrices derived from material permittivities and permeabilities as well as wave vector components. From these, a second-order matrix wave equation for the electric field components is derived:

$$\frac{d^2}{dz'^2} \begin{bmatrix} E_x \\ E_y \end{bmatrix} = \underbrace{\mathbf{P}\mathbf{Q}}_{\Omega^2} \begin{bmatrix} E_x \\ E_y \end{bmatrix} \quad (6)$$

Fields in each layer are described by eigenmodes. At each interface, SMs—drawing analogies from network theory—are applied, whereby respective interfaces act as a 2-port, 4-terminal device:

$$\begin{bmatrix} \mathbf{c}_{i-1}^- \\ \mathbf{c}_{i+1}^+ \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{11}^{(i)} & \mathbf{S}_{12}^{(i)} \\ \mathbf{S}_{21}^{(i)} & \mathbf{S}_{22}^{(i)} \end{bmatrix} \begin{bmatrix} \mathbf{c}_{i-1}^+ \\ \mathbf{c}_{i+1}^- \end{bmatrix} \quad (7)$$

where the submatrices $\mathbf{S}_{11}^{(i)}$ (e.g., reflection at the interface between layer $i-1$ and i , i.e., propagation from port 1 to port 1), $\mathbf{S}_{12}^{(i)}$, $\mathbf{S}_{21}^{(i)}$ and $\mathbf{S}_{22}^{(i)}$ relate incident and transmitted wave amplitudes across interfaces. For multilayer systems, individual layer SMs are combined recursively using the Redheffer star product to yield a device SM:

$$\mathbf{S}^{(\text{global})} = \mathbf{S}^{(\text{ref})} \otimes \underbrace{\left[\mathbf{S}^{(1)} \otimes \mathbf{S}^{(2)} \otimes \dots \otimes \mathbf{S}^{(N)} \right]}_{\mathbf{S}^{(\text{device})}} \otimes \mathbf{S}^{(\text{trn})} \quad (8)$$

Once the global SM is computed, the reflected electric field vector \vec{E}_{ref} is obtained via the global reflection block $\mathbf{S}_{11}^{(\text{global})}$, acting on the normalized incident field \vec{P} :

$$\vec{E}_{\text{ref}} = \mathbf{S}_{11}^{(\text{global})} \vec{P} \quad (9)$$

In homogeneous isotropic media, field amplitudes directly correspond to modal coefficients. The longitudinal electric field component E_z is derived from E_x and E_y using Maxwell's divergence condition $\nabla \cdot \vec{E} = 0$:

$$E_z^{\text{ref}} = -\frac{\tilde{k}_x E_x^{\text{ref}} + \tilde{k}_y E_y^{\text{ref}}}{\tilde{k}_z^{\text{ref}}} \quad (10)$$

Accordingly, the reflectance r is defined as the squared norm of the total reflected electric field, normalized by that of the incident field:

$$r = \frac{|\vec{E}_{\text{ref}}|^2}{|\vec{E}_{\text{inc}}|^2} = |\vec{E}_{\text{ref}}|^2 \quad (11)$$

because \vec{E}_{inc} is normalized to be \vec{P} , with $|\vec{P}| = 1$.

A.3 Generation of monotonic and convex test thickness time series based on Gaussian Processes

We would like to generate smooth monotonic thickness time series to cover a wide range of test cases. We can sample such functions based on Gaussian Processes (GPs) [24]. All the necessary tools for doing so are implemented in our library `reflax`.

Given a covariance matrix K , e.g., given by pairwise application of a radial basis function (RBF) kernel:

$$k(x, x') = v \cdot \exp\left(-\frac{|x - x'|^2}{2 \cdot l^2}\right), \quad \text{two evaluation points } x, x' \quad (12)$$

on the evaluation points of interest, with the variance parameter v and length-scale parameter l , we can sample a smooth function with expected zero mean following the covariance structure K based on the following steps:

1. Find a matrix L such that $K = LL^T$ by Cholesky decomposition (where one adds a small jitter term to the diagonal of K for numerical stability).
2. Sample a vector z from a standard normal distribution.
3. Return the sample Lz .

To generate monotonic thicknesses, we apply a softplus function to the GP sample, numerically integrate, and then scale the result so that the final thickness lies in a predefined range. Additionally, we employ rejection sampling to only accept thickness profiles with growth rates in a limited, physically sensible range.

For our current thin-film depositions of interest (on the timescale of one hour), the growth rate itself is non-decreasing, so the thickness function is convex and monotonically increasing. To sample such a convex function, one samples the second derivative, applies the softplus, adds a uniformly distributed random number between the minimum and maximum desired growth rate, and then proceeds as before.

A.4 Experimental and simulation setup parameters

The experimental and simulation setup parameters are given in Table 1.

Table 1: Experimental parameters used for our experiments and simulations.

Quantity	Value
Polar angle θ	75° (simulation tests), 25° (experiment)
Polarization state	s-polarized
Wavelength λ	632.8 nm
Incident medium permittivity $\varepsilon_{r,\text{inc}}$	1.0
Transmission medium permittivity $\varepsilon_{r,\text{trans}}$	$(3.8827 + 0.019626i)^2$
Variable layer permittivity $\varepsilon_{r,\text{var}}$	1.457 ²
All permeabilities μ_r	1.0

A.5 Neural Operator Architectures

The goal of neural operators is to learn a mapping $\mathcal{N} : \mathcal{X} \rightarrow \mathcal{Y}$ between two function spaces. In our case, this means learning the mapping $\mathcal{R}^{-1} : \mathcal{S} \rightarrow \mathcal{D}$ from the continuously differentiable and bounded reflectance functions $r \in \mathcal{S} \subset C^1([0, 1], \mathbb{R})$ to the continuously differentiable and monotonically increasing thickness functions $d \in \mathcal{D} \subset C_{\text{mon}}^1([0, 1], \mathbb{R})$. A common first step in learning such mappings is to project from the infinite-dimensional function spaces into a finite-dimensional subspace (cf. [25]), which can be done simply by representing the functions at a set of discrete locations in the domain of interest. In the context of this work, the discretization arises naturally, as the measurements of the continuous growth process are inherently discrete. We choose to represent the functions $r(t)$ and $d(t)$ at $n_{\text{eval}} = 100$ equispaced time points t_i in the interval $[0, 1]$, corresponding to the growth time in hours.

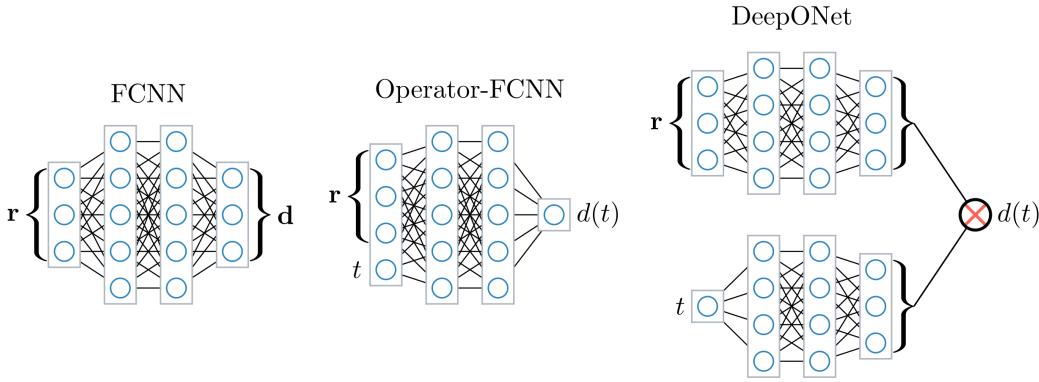


Figure 8: Schematic of the three investigated neural operator architectures FCNN, Operator-FCNN, and DeepONet. These are used to produce an »educated guess« for the mapping $\mathbf{r} \mapsto \mathbf{d}$, which is then refined using `reflax`'s differentiable pipeline.

It then remains to learn the mapping $\mathcal{N} : \mathbb{R}^{n_{\text{eval}}} \rightarrow \mathbb{R}^{n_{\text{eval}}}, \mathbf{r} \mapsto \mathbf{d}$. For this task, we compare three different architectures, schematics of which are shown in Fig. 8. The fully-connected neural network FCNN directly learns the mapping $\mathbf{r} \mapsto \mathbf{d}$ between the two vectors \mathbf{r} and \mathbf{d} representing the functions $r(t)$ and $d(t)$, hence the architecture has n_{eval} inputs and outputs. The second architecture, Operator-FCNN, aims to provide more flexibility in the output domain. It learns a mapping $\mathbb{R}^{n_{\text{eval}}} \times [0, 1] \rightarrow \mathbb{R}, (\mathbf{r}, t) \mapsto d(t)$, from which the desired mapping $\mathbf{r} \mapsto \mathbf{d}$ can then be obtained by evaluating the operator n_{eval} times at the required time points t_i . The third architecture, DeepONet following [19], learns the same mapping and is thus evaluated following the same procedure. It differs from Operator-FCNN in that it employs two separate neural networks, *branch* and *trunk*, to capture the different nature of the two inputs \mathbf{r} and t . The output vectors of branch and trunk are combined into the single value $d(t)$ by employing a scalar product. The structure of Operator-FCNN and DeepONet enables more flexibility in the output domain because they can be evaluated at any time $t \in \mathbb{R}$. In principle, this allows them to interpolate the thickness $d(t)$ between the time points t_i used for training, or even extrapolate out of the training domain.

Table 2: Optuna-tuned hyperparameters by architecture.

FCNN (tuned)	Operator-FCNN	DeepONet
hidden_size = 182	hidden_size = 291	hidden_width = 285
num_hidden_layers = 5	num_hidden_layers = 3	branch_layers = 4
learning_rate = 4.0×10^{-4}	learning_rate = 1.0×10^{-3}	trunk_layers = 2
reg_factor = 1.8×10^{-7}	reg_factor = 1.2×10^{-6}	latent_size = 89
activation = LeakyReLU	activation = LeakyReLU	learning_rate = 2.0×10^{-4}
		reg_factor = 1.2×10^{-6}
		activation = ReLU

For the results in Section 3, the FCNN architecture with a baseline configuration was utilised: two hidden layers with 512 neurons each. To investigate potential performance gains with optimised architectures, we performed hyperparameter tuning with OPTUNA [20], optimising for MSE on the test set. The resulting hyperparameter configurations are shown in Table 2. A noteworthy detail is that the trainings for all architectures utilised the AdamWScheduleFree optimiser first presented at NeurIPS 2024 [26]. In addition to architecture-specific parameters, the best activation function (chosen from ReLU, Tanh, LeakyReLU, Sigmoid, GELU, and Softplus) and the optimal weight for an L_2 -regularisation term (reg_factor) were determined. The training data used to train the architectures are the 2100 samples mentioned in Section 3, split in a 70/10/20 train/test/validation split. The additional 200 samples with shorter RBF length-scale were used as a second validation set, the results in Fig. 9 were obtained using this second validation set.

While hyperparameter tuning led to slight improvements in performance, as measured by MSE on the validation set (see Section 3), these gains remain modest, as illustrated in Fig. 9. Thus, although an optimized neural operator architecture can provide somewhat more accurate initial guesses than those shown in Fig. 2, true optimal performance still relies on the combined strength of a neural operator’s »educated guess« and the corrective influence of a physics-informed differentiable modeling pipeline.

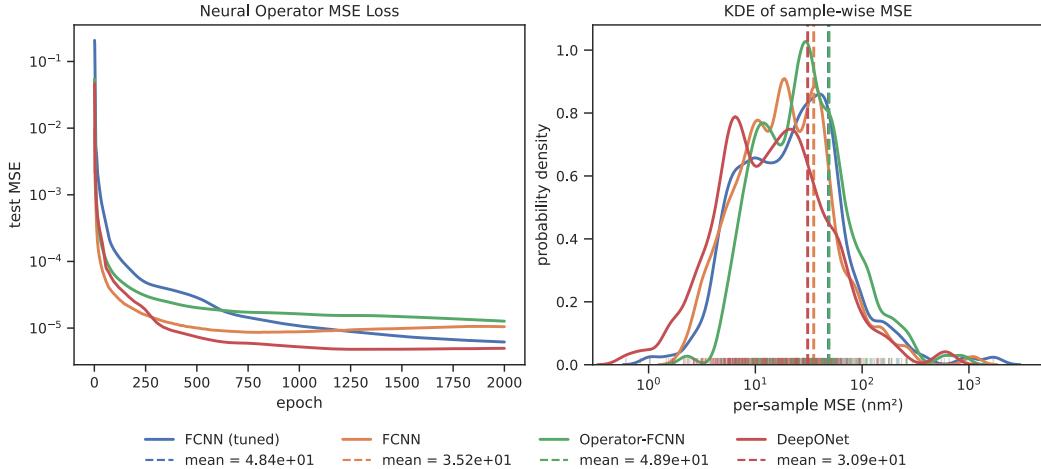


Figure 9: **Left:** Comparative plots of the loss trajectories of the four neural operator architectures FCNN (baseline and tuned), Operator-FCNN, and DeepONet. **Right:** Kernel density estimates (KDEs) of the distribution of the sample-wise MSE on the validation dataset (200 samples) for the four neural operator architectures. The dashed lines represent the overall MSE for each architecture; by sample-wise, we mean the MSE achieved while predicting a full thickness profile, which is one sample in the validation dataset.

A.6 Applications to experimental data

In order to examine the accuracy of the proposed method in an experimental laboratory setup, iCVD depositions are made and in-situ measured by laser interferometry and ex-situ ellipsometry. For the evaluation with *reflax*, the refractive index of the thin film is estimated from literature values. Results show a minimal, mostly positive deviation compared to the ellipsometry measurements, with a mean absolute error of 1.95. See table 3 and figures 10, 11, 12, 13, 14, 15.

Table 3: Final thickness values of iCVD depositions with varying monomer shares, measured by ellipsometry and by our low-cost laser interferometry setup analyzed with `reflax`. The maximal relative deviation between both methods is small, highlighting the accuracy of the proposed approach.

Nr.	Monomere share	Ell.[nm]	Laserint. [nm]	Ell. – Laserint. [nm (%)]
1	EGDMA (100)	646,16	665,46	+19,3 (3%)
2	HEMA + EGDMA (50/50)	860,84	888,22	+27,38 (3,2%)
3	HEMA + EGDMA (25/75)	914,0	929,27	+15,27 (1,7%)
4	HEMA + DVB (50/50)	741,9	761,44	+19,54 (2,6%)
5	HEMA + DVB (75/25)	909,22	910,55	+1,33 (0,2%)
6	HEMA + DVB (60/40)	865,1	858,9	-6,2 (1%)

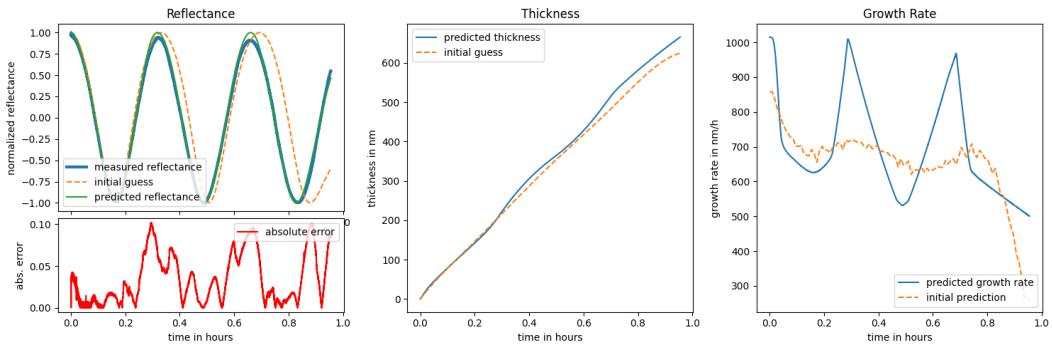


Figure 10: Nr. 1 EGDMA deposition.

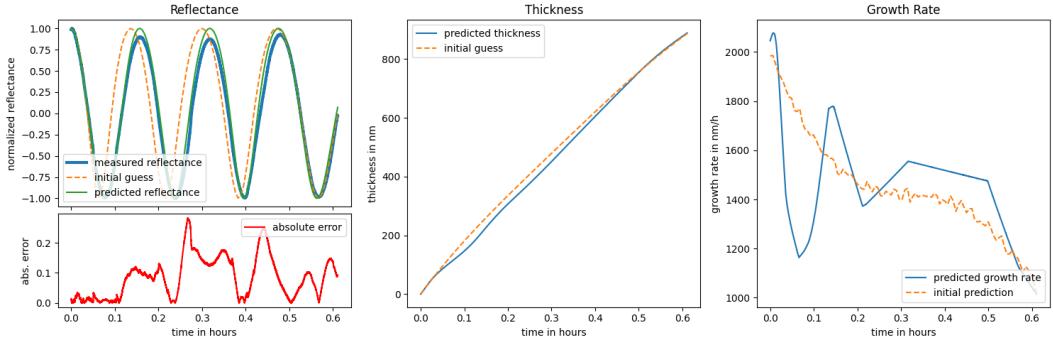


Figure 11: Nr. 2 HEMA/EGDMA (50/50)-deposition.

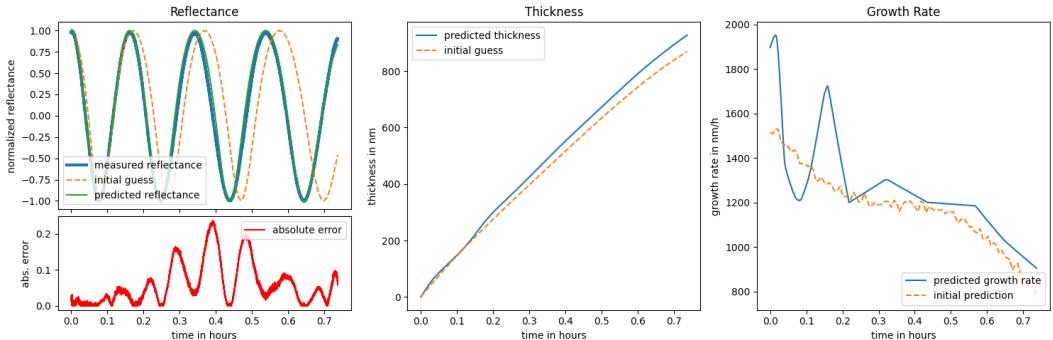


Figure 12: Nr. 3 HEMA/EGDMA (25/75)-deposition.

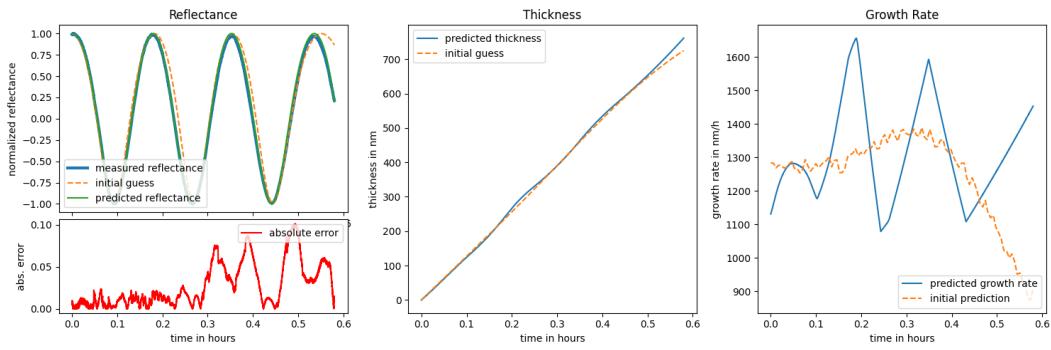


Figure 13: Nr. 4 DVB+HEMA-deposition (50/50).

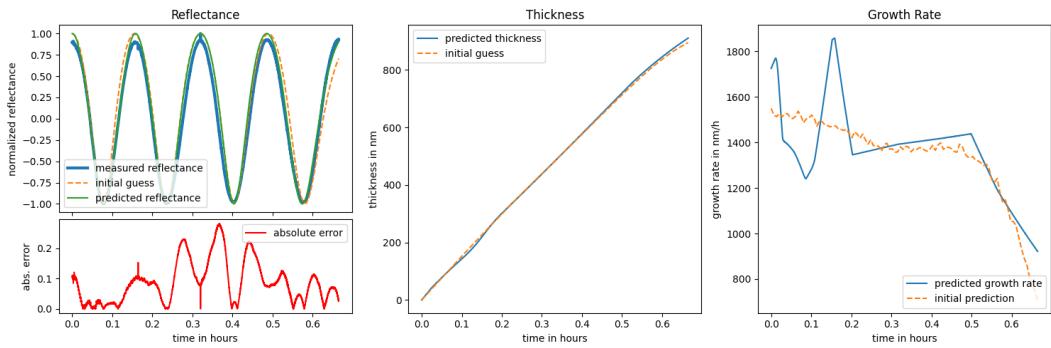


Figure 14: Nr. 5 DVB+HEMA-deposition (75/25).

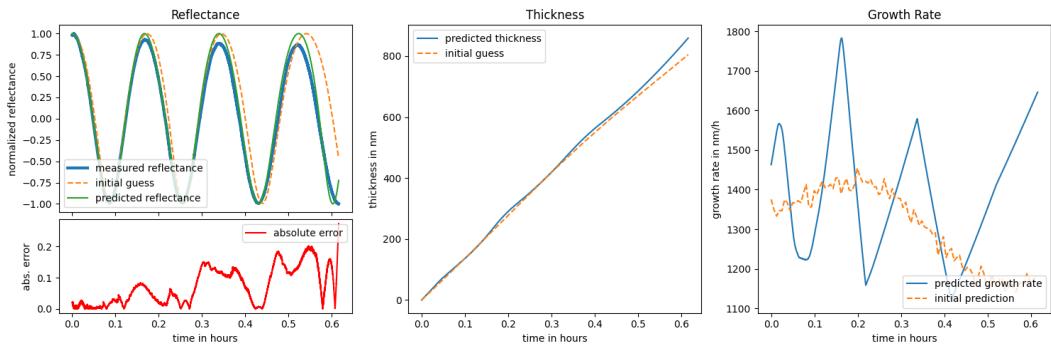


Figure 15: Nr.6 DVB+HEMA-deposition (60/40).