



Physics-informed recurrent neural network for time dynamics in optical resonances

Yingheng Tang^{1,2}, Jichao Fan¹, Xinwei Li³, Jianzhu Ma^{4,5}, Minghao Qi², Cunxi Yu¹✉ and Weilu Gao¹✉

Resonance structures and features are ubiquitous in optical science. However, capturing their time dynamics in real-world scenarios suffers from long data acquisition time and low analysis accuracy due to slow convergence and limited time windows. We report a physics-informed recurrent neural network to forecast the time-domain response of optical resonances and infer corresponding resonance frequencies by acquiring a fraction of the sequence as input. The model is trained in a two-step multi-fidelity framework for high-accuracy forecast, using first a large amount of low-fidelity physical-model-generated synthetic data and then a small set of high-fidelity application-specific data. Through simulations and experiments, we demonstrate that the model is applicable to a wide range of resonances, including dielectric metasurfaces, graphene plasmonics and ultra-strongly coupled Landau polaritons, where our model captures small signal features and learns physical quantities. The demonstrated machine-learning algorithm can help to accelerate the exploration of physical phenomena and device design under resonance-enhanced light-matter interaction.

Forecasting time dynamics is central in many contexts of scientific research and commercial decision-making such as climate modelling¹, medical science² and finance³. Classical parametric models such as the autoregressive integrated moving average⁴ enjoy merits of model simplicity and easy solutions but suffer from large prediction errors when using strongly nonlinear and stochastic data. Machine-learning approaches provide the opportunity to learn time dynamics and the underlying complex representation in a data-driven manner without fixed parameters or structures⁵. In particular, given the success of recurrent neural networks (RNNs) in natural language processing and the logical interpretation of time series as sequences, various RNN architectures, such as long short-term memory and the gated recurrent unit (GRU), have been employed for time-series forecast applications^{6–8}. However, the problems of overfitting and local minima often lead to low model efficacy⁹. Moreover, these models are limited to forecasting short sequences, while high-accuracy long-sequence forecasting is more desirable¹⁰. In addition, conventional machine-learning approaches generally require a substantial amount of training data^{11,12}, which is infeasible in many practical contexts, and they are also unable to extract interpretable knowledge from the dataset¹³. The fusion of machine-learning algorithms with empirical models thus emerges as an efficient learning philosophy to address these limitations^{14,15}, especially physics-informed machine-learning methods in physical science^{16,17}.

Optical cavities with resonance-enhanced light-matter interactions enable broad innovations including optical quantum technologies^{18–24}. One specific example showing the importance of capturing and understanding time dynamics is terahertz time-domain spectroscopy (THz-TDS) of THz resonances, which is stimulating ongoing interest in diverse disciplines such as health²⁵, sensing and imaging^{26,27}, security²⁸, computing^{29,30} and communication³¹, in addition to quantum applications^{32,33}. In contrast to continuous-wave spectroscopy, which only captures intensity spectra,

the uniqueness of THz-TDS includes the simultaneous acquisition of intensity and phase information, broad spectral coverage and time-resolved capability^{34–36}. However, the time-domain signals of resonance features decay slowly, and long data acquisition times are needed for accurate analysis in numerical simulations and experimental measurements. Fast captured short time-domain signals contain an incorrect corresponding frequency-domain representation, which prevents conventional harmonic analysis. Thus, there is a trade-off between the data acquisition time and the analysis accuracy. Moreover, the time window of experimental THz-TDS is limited to prevent undesired effects such as multiple reflections, which introduces an additional constraint on accurate data analysis³⁶.

Here, we describe a high-performance model of physics-informed cascaded GRU networks to forecast long time-domain signals using short input signals obtained from finite-difference time-domain (FDTD) simulations and THz-TDS experiments. This model breaks the trade-off between the data acquisition time and the analysis accuracy by using fast captured short input sequences and highly accurately predicted long output sequences. Instead of directly training the model using a substantial amount of high-fidelity data obtained from time-consuming electromagnetic calculations or experiments, we employ a two-step multi-fidelity training approach. A large number of low-fidelity free induction decay (FID)-model-generated synthetic data is first used to pre-train the model, which reduces the search space of the network parameters for fast and efficient learning, alleviates the problem of local minima for high-accuracy forecast and generalizes the applicability of the model. Through transfer learning using a small set (at least one order of magnitude smaller than that of the low-fidelity data) of high-fidelity application-specific data, the pre-trained model is tailored to a broad range of resonance features, including resonant dielectric metasurfaces³⁷, graphene plasmonics³⁸ and ultra-strongly coupled electron cyclotron resonance in a Landau-quantized high-mobility two-dimensional electron gas (2DEG) with photons in a

¹Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT, USA. ²Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. ³Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, CA, USA.

⁴Institute of Artificial Intelligence, Peking University, Beijing, China. ⁵Beijing Institute of General Artificial Intelligence, Beijing, China.

✉e-mail: cunxi.yu@utah.edu; weilu.gao@utah.edu

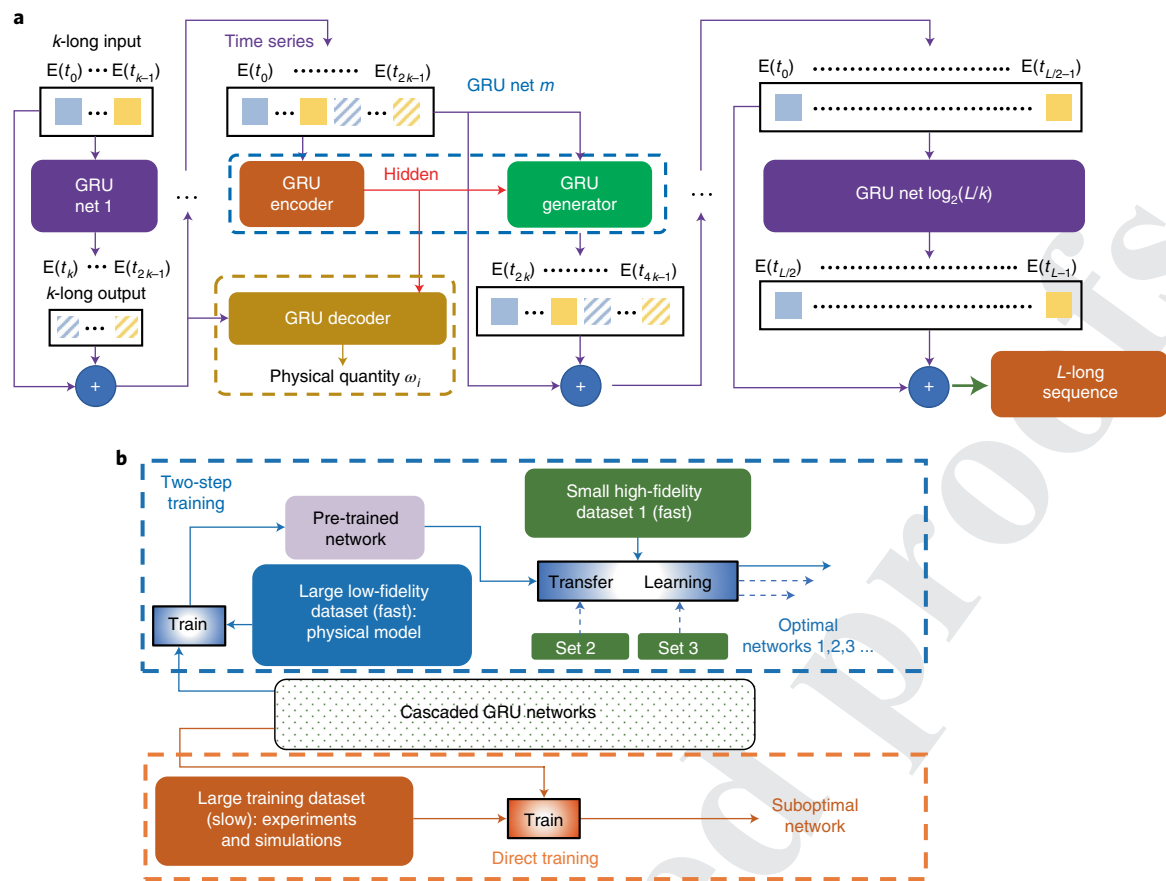


Fig. 1 | Cascaded GRU networks and two-step multi-fidelity training approach. a, Forecast of an L -long time-domain signal with a k -long time-domain signal using cascaded GRU networks. The output sequence of the GRU network at each stage is combined with the input sequence to serve as the input for the next-stage GRU. In the experiment, a branch GRU network is connected to a GRU encoder in the middle stage of the cascaded GRU networks that are for time-series forecast. **b**, The slow and suboptimal conventional training approach (orange path) and our fast, broadly applicable and optimal two-step multi-fidelity training approach (blue path).

high-quality-factor cavity³³. The cascaded GRU networks enable precise long-sequence forecasting, and for dielectric metasurfaces the input sequence is 12.5% of the full sequence, suggesting an eight-fold speed-up of data acquisition. Furthermore, this model accurately captures signal features that account for only 0.01% of the total signal energy in experimental data of Landau polaritons and simultaneously learns resonance frequencies in spectra. The polariton dispersions obtained from experimentally measured and forecast time-domain signals, as well as model-learned quantities, all agree well. The values obtained for the coupling rates and cooperativity from the forecast spectra (147.9 GHz and 3,663) and the model inference (145.6 GHz and 3,550) match well with the experimental values (150.1 GHz and 3,513)³³. The variations of the predicted coupling rates from those obtained from the forecast spectra and model inference are 1.5% and 3%. For cooperativity, these two values are 4% and 1%.

Results

Cascaded GRU networks and multi-fidelity training. Figure 1a illustrates the model of cascaded GRU networks taking short input sequences (length k) and forecasting long output sequences (length L). An input sequence is used in the first GRU network and then combined with the forecast output sequence for the input to the next-stage network. The GRU network at each stage has the same sequence length for both the input and the predicted output. In the case where $L/k = 2^M$ (where M is a non-negative integer), the minimum required number of cascaded stages of the GRU networks

is $\log_2(L/k)$, corresponding to the recursive bisection of an L -long sequence. In one middle stage of the cascaded networks, a GRU decoder is branched out for the simultaneously learning of resonance frequencies associated with time signals (Fig. 1a, dashed gold rectangle). The hidden states from the GRU encoder and the time-domain data from the previous-stage GRU network form the input for both the branch GRU decoder and the forecast GRU generator that produces the predicted time-domain output of this GRU stage. The branch GRU decoder suggests that the hidden states obtained from time-signal-forecast GRU networks are related to interpretable physical observables, and it is also experimentally beneficial for accelerated acquisition and understanding of time dynamics.

Figure 1b displays the training process of the cascaded GRU networks. A conventional training approach (Fig. 1b, orange path) requires a large training dataset, which generally takes a long time to generate through FDTD simulations or experiments and frequently leads to suboptimally trained networks because of issues related to local minima¹⁶. In contrast, we utilize a physics-informed two-step multi-fidelity training approach (Fig. 1b, blue path), where the networks with randomly initialized weights are first pre-trained by using a large number of low-fidelity synthetic data that are generated instantaneously from analytical physical models based on domain expertise. In the transfer learning process, the pre-trained networks are then fine trained with a small high-fidelity application-specific dataset obtained from either FDTD simulations or experiments. The fully trained network is optimal and has superior performance over the one trained using the conventional approach.

We first demonstrate our model and training methodology for predicting time-domain signals in resonant dielectric metasurfaces, which consist of a periodic array of unit cells with four dielectric cylindrical pillars of varying diameter (d) (ref. ³⁷) (Fig. 2a). We generate high-fidelity training data by numerically simulating the electrical field time response of the structures with the diameter in a range to have THz resonances (see Methods section for details). In addition to bright photonic modes, the coupling between neighbouring pillars can generate sharp Fano resonances. The bottom trace in Fig. 2b shows the electric field norm of a representative time-domain signal on a logarithmic scale. The purple section is the input short sequence, while the remaining cyan section is the long sequence to be predicted. Although the standard GRU architecture has been employed in a variety of time-series forecast applications^{6–8}, they are limited to forecasting short sequences using long sequences, yet architectural concepts are similar to sequence-to-sequence (seq2seq) models³⁹ that can be extended to long sequence forecasts. We trained our cascaded GRU networks and a standard seq2seq model using high-fidelity FDTD simulation training data. The black dashed line and the blue dashed line with cross markers in Fig. 2c show the training mean squared error (MSE) loss of the seq2seq model constructed using the GRU architecture and our model trained in a conventional manner, respectively (see Methods section for detailed descriptions of the MSE loss and seq2seq model and Supplementary Fig. 1b for an illustration of the seq2seq model). The clear contrasts of the training MSE loss shown in Fig. 2c and the test loss shown in Supplementary Fig. 1c indicate that the standard seq2seq model does not perform well.

To improve the forecast accuracy, we employ a multi-fidelity training framework¹⁶. Most physical resonance features and structures appear as a sum of damped oscillations in time-domain signals originating from the FID in two-level atomic systems⁴⁰. They follow the general mathematical form $\sum_i A_i e^{-\alpha_i t} \sin(\omega_i t)$, where A_i is the amplitude factor, $e^{-\alpha_i t}$ describes the decay envelope of the harmonic where α_i is connected to a lifetime in the dephasing process, $\sin(\omega_i t)$ is an oscillating carrier with a resonance frequency of ω_i , corresponding to the energy level in a two-level system and i enumerates all the resonance features. We employed this analytical model to generate 40,000 low-fidelity synthetic data (see Methods section and Supplementary Fig. 2 for details). Instead of random initialization, all the GRU networks went through the physics-informed initialization process, where they were pre-trained and their weights were informed by the synthetic data generated by the FID model. The transfer learning of the pre-trained model was achieved by fine-tuning the model parameters with 4,000 high-fidelity time-domain data of dielectric metasurfaces obtained through FDTD simulations (see Methods section for details). We trained the model with input sequences of various lengths (k) and evaluated its performance (see Methods section and Supplementary Fig. 3 for details). The ratio of the full sequence length to the input sequence length, L/k , is chosen as ~ 8 for the shortest input sequence with high forecast accuracy. How the time-domain signals are segmented, and thus the length of the GRU network at each stage, has little influence on the model forecast accuracy (see Methods section and Supplementary Fig. 1c,d).

Figure 2d and the upper traces in Fig. 2b display the part of and the full predicted time-domain signals with and without the pre-training process and the target time-domain signal. As also shown in Fig. 2c and Supplementary Fig. 1c, the cascaded GRU networks trained using the multi-fidelity approach clearly outperform those trained using the conventional approach in terms of forecast accuracy. Note that there is no overfitting in our model since the training and test losses show a negligible difference. Unless otherwise stated, all the time-domain signals and corresponding frequency-domain spectra are from test datasets and the MSE loss refers to the test loss (see Methods section for details of the MSE loss and Supplementary Fig. 1a for a comparison of the training and test losses).

The better forecast accuracy of the model trained using the multi-fidelity approach suggests that the physics-informed pre-training initialization process utilizing synthetic data generated by the FID model can help with the escape from local minima¹⁶. Furthermore, we obtained three types of full-length time-domain signals by combining input sequences with the forecast time-domain signals obtained using the models trained by the conventional and multi-fidelity approaches, and directly by zero-padding input sequences. We then calculated the sample Fourier transformation of the obtained signals as well as the reference Fourier transformation of a time-domain signal from a bare silicon substrate without any structures (Methods). There is no need to forecast the reference time-domain signal since it was taken only once. Figure 2e displays the transmittance spectra as the squared norm of the ratio of the Fourier transformation of the sample over that of the reference. Figure 2f displays the corresponding phase spectra as the angle of this complex-valued ratio. Both the transmittance and phase spectra calculated from the time-domain signals generated from the model trained using the multi-fidelity approach show good agreement with the target spectra calculated based on the data obtained from FDTD simulations. The spectra calculated from the time-domain signals generated from the model trained using the conventional approach show clear deviation, while the resonance feature nearly disappears in those obtained from zero-padded input sequences.

Model generalization to different optical resonances. Our method can be generalized to other resonance structures and features. The similar damped oscillation signature of the time-domain signals in most resonance features originating from the FID process physically guarantees the feasibility of such generalization. We demonstrate the generalization of the approach to two physically distinct resonance features, viz. active graphene plasmonics³⁸ and ultra-strongly coupled Landau polaritons^{32,33}, in addition to dielectric metasurfaces. Specifically, periodically patterned monolayer graphene ribbons that can support localized THz plasmonic resonance from bounded carriers are simulated to obtain time-domain signals (Fig. 3a). The width and Fermi level of the graphene ribbon are selected to obtain a THz resonance frequency (see Methods section for details). The model was pre-trained using 8,000 synthetic data and fine-trained using 800 high-fidelity FDTD simulation data. The input sequence length is chosen with $L/k \sim 4$ (Methods and Supplementary Fig. 3). Both the predicted time-domain signals and the corresponding frequency-domain spectra show agreement with the target time signal (Fig. 3b) and spectra (Fig. 3c) obtained from FDTD simulations. The transmission spectra were calculated in the same manner as used for the dielectric metasurfaces, and the reference time-domain signal was taken from a bare silicon oxide substrate without graphene structures.

Moreover, an ultra-high-mobility 2DEG inside a high-quality-factor one-dimensional THz photonic-crystal cavity under a magnetic field displayed ultra-narrow Landau polaritons^{32,33} (Fig. 3d). Their spectra were measured experimentally by standard THz-TDS under high magnetic fields as reported in ref. ³³, where the reference time-domain signal was measured from air transmission. There are 71 spectra in total, under various magnetic fields from 0 to 4.5 T. We use 24 spectra to fine-train the pre-trained model trained using 3,000 synthetic data, and the remaining 47 spectra as test data. These 24 spectra are from measurements taken under magnetic fields either uniformly or randomly distributed between 0 and 4.5 T. The test loss displays small variance when the training spectra are shuffled for different magnetic fields (Supplementary Fig. 4). The length of the input sequence for the 47 test time-domain signals is chosen with $L/k \sim 7$ (Methods and Supplementary Fig. 3). In contrast to simulation data, experimental data contain measurement noise, for example, from lasers and electronic components. In addition, the number of high-fidelity time-domain data (71) is

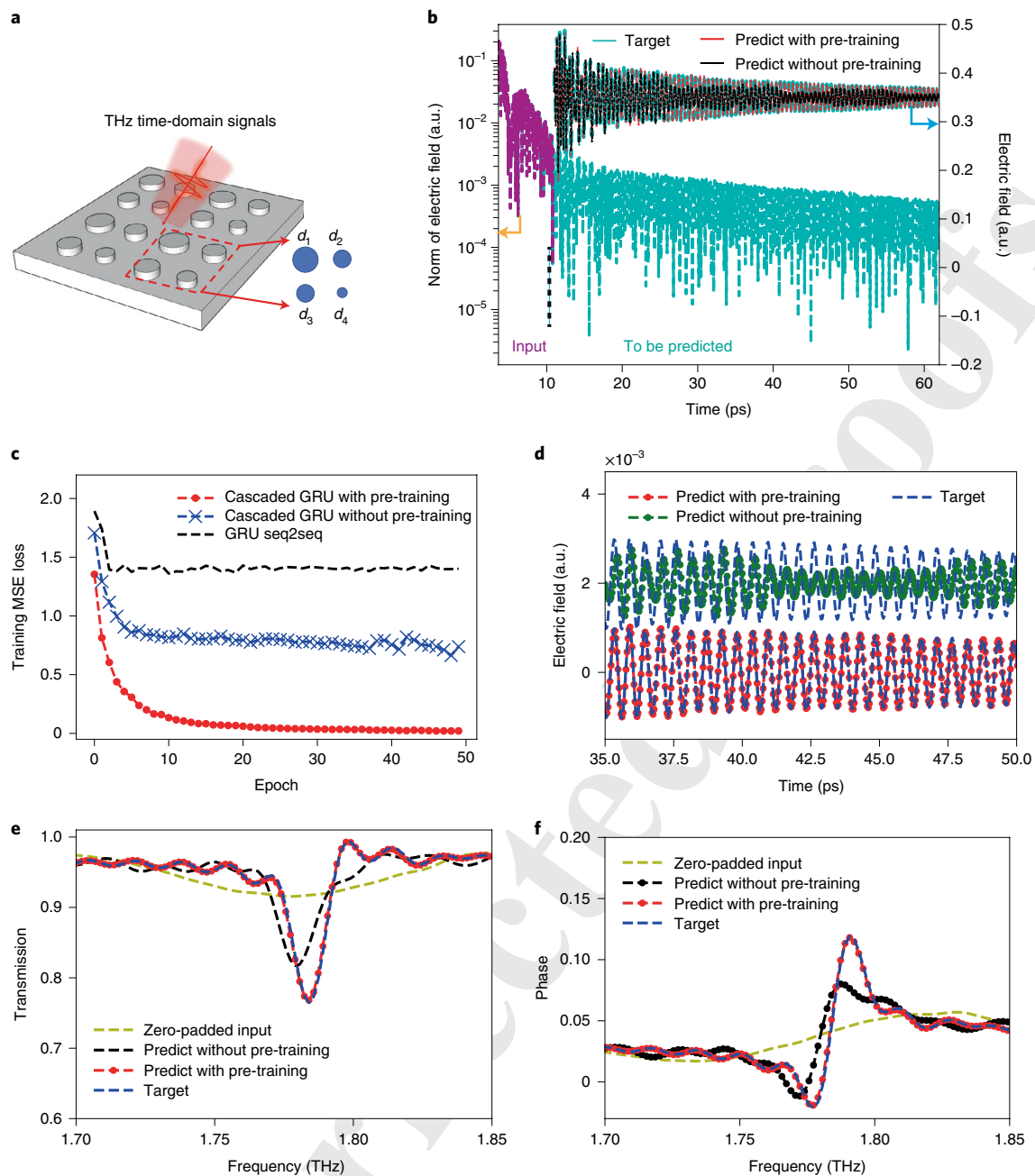


Fig. 2 | Model demonstration for dielectric metasurfaces. **a**, A dielectric metasurface with a unit cell consisting of four cylindrical pillars of different dimensions. **b**, Lower trace: the electric field norm of a representative time-domain signal illustrating the input sequence (purple section) and the sequence to be predicted (cyan section) on a logarithmic scale. The orange arrow indicates that the lower trace uses the left y axis. Upper traces: the time-domain signal corresponding to the sequence to be predicted (that is, target) (cyan line), the predicted signal from the model trained through the conventional approach (black line, without pre-training) and the predicted signal from the model trained using the two-step approach (red line, with pre-training). The signals are multiplied by 35 and offset vertically for easy visualization. The blue arrow indicates that the upper traces use the right y axis. **c**, The training MSE loss as a function of the training epoch for a seq2seq model constructed using the GRU architecture, our cascaded GRU networks without pre-training and our cascaded GRU networks with pre-training. **d**, Predicted time signals using the conventional training approach and the two-step multi-fidelity approach. Clearly better forecast performance is observed when using the two-step approach. **e, f**, THz transmission (**e**) and phase spectra (**f**) produced from the Fourier transformation of time-domain signals obtained from zero-padded input signals, predicted signals without pre-training, predicted signals with pre-training and target time-domain signals from numerical simulations.

much smaller than those used for the graphene plasmonics (1,000) or dielectric metasurfaces (5,000). Moreover, our Landau polariton features are located inside the defect mode of the photonic-crystal cavity stop band, and the signal energy of interest only accounts for 0.01% of the total signal energy (see Supplementary Fig. 5 for a full spectrum). All these factors make it challenging to forecast using

the conventional direct training approach. However, the two-step multi-fidelity approach substantially improves the prediction power and accuracy, thus small but important resonance features can be captured. Figure 3e,f displays one representative time-domain signal and the corresponding transmission spectra, both aligning well with target experimental results.

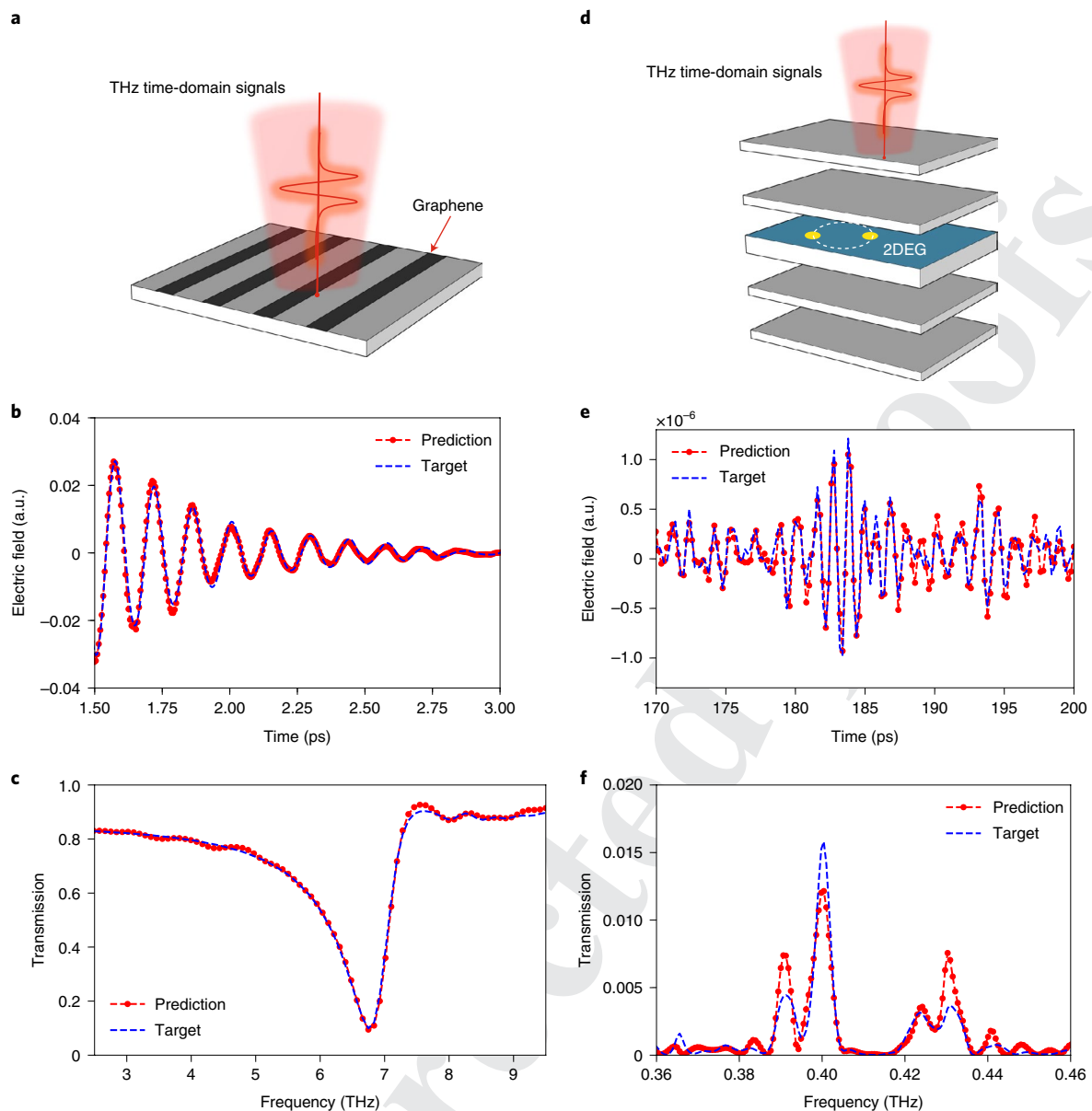


Fig. 3 | Model generalization to graphene plasmonics and Landau polaritons. **a–c**, Graphene plasmonics, showing schematics of graphene ribbons supporting localized plasmonic resonance (**a**) and time-domain (**b**) and corresponding frequency-domain response (**c**) for the predicted time signal and the full target time signal. **d–f**, Landau polaritons from strongly coupled photons in a one-dimensional photonic-crystal cavity with electron cyclotron resonance in a high-mobility two-dimensional electron gas (**d**), showing the time-domain (**e**) and corresponding frequency-domain response (**f**) under a specific magnetic field for the predicted time signal and the full target time signal.

For Landau polariton experiments, a branch GRU network is added to a middle stage of the cascaded GRU networks used for time-series forecast to infer the resonance frequencies associated with the time-domain signals (Fig. 1a). The cascaded GRU networks were first trained for time-series forecast, then the branch GRU network was trained with the hidden states and intermediate sequences from time-series-forecast GRU networks as the input (see Methods section for details). After training, the short time signals from the test data are input into the full model, simultaneously generating both the forecast time-domain signals and the corresponding resonance frequencies.

Model analysis in Landau polariton experiments. Figure 4a,b displays all 47 experimental spectra and the spectra obtained by Fourier transformation of the corresponding predicted time-domain

signals. All the essential physical features expected in linearly polarized transmission spectra are well reproduced, including cyclotron-resonance-active lower polaritons (CRA-LPs), CRA upper polaritons (CRA-UPs) and CR-inactive (CRI) modes³³ (see Supplementary Figs. 6 and 7 for more data). In stark contrast, the spectra obtained directly from short input sequences with zero paddings (Fig. 4c) display completely random patterns with all features lost. The distinct difference seen among Fig. 4a–c highlights the necessity for long data acquisition to capture essential THz features if no prediction is employed, as well as the high prediction power and accuracy of our cascaded GRU networks and the two-step multi-fidelity training approach.

Furthermore, yellow stars in Fig. 4d, orange circles in Fig. 4e and gold triangles in Fig. 4f show peak positions extracted from experimentally measured spectra obtained as the Fourier transformation

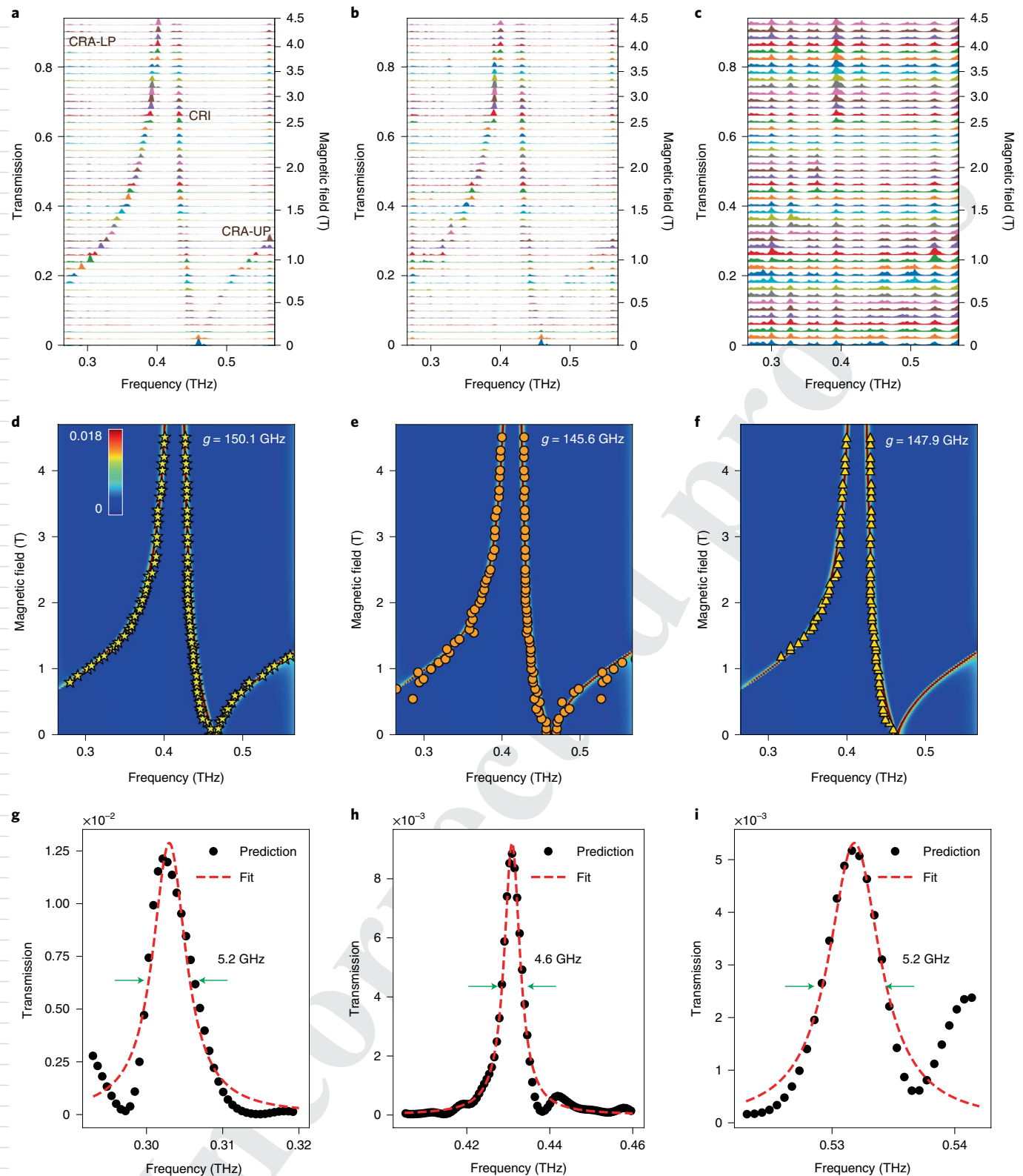


Fig. 4 | Experimental verification of the GRU model in Landau polaritons. **a–c**, Linearly polarized THz transmittance spectra under various magnetic fields obtained by Fourier transformation of experimentally acquired time-domain signals (**a**), predicted time-domain signals combined with much shorter input experimental time-domain signals (**b**) and short input experimental time-domain signals with zero padding (**c**). The spectra are offset vertically for easy visualization. **d–f**, Simulated transmittance colour contour maps obtained using the transfer matrix method to fit the experimental spectra shown in **a** (**d**), the model-learned resonance frequencies (**e**) and the predicted spectra shown in **b** (**f**). The colour bar is the same and shown in **d**, representing the transmittance. Yellow stars in **d**, orange circles in **e** and gold triangles in **f** mark the peak positions extracted from experimental spectra, learned from the branch GRU network and extracted from predicted spectra. The extracted coupling rates g are 150.1, 145.6 and 147.9 GHz, respectively. **g–i**, Lorentzian fits (dashed red line) of the predicted spectra for the UP peak at zero detuning (1 T) (**g**), the LP peak at zero detuning (**h**) and the CRI mode at 3 T (**i**). The obtained full-width at half-maximum values are indicated by green arrows.

of time-domain signals, the model-learned resonance frequencies and the peaks extracted from the predicted spectra, respectively. We utilized the transfer matrix method (see Methods section and ref.³³ for details) to calculate the transmission spectra for samples under various magnetic fields, through which the coupling ratio g can be extracted. Figure 4d–f also displays the transmittance colour contour map calculated to match the extracted and learned peaks. Note that the CRA-UP branch is not clearly predicted (Fig. 4b), so the corresponding peaks are missing from Fig. 4f. These missing features again reflect the challenging aspects of predicting small signals in experimental Landau polariton spectra. Despite such missing peaks, unique fitting is still possible for the CRA-LP and CRI branches. In contrast, the branch GRU decoder successfully captures all the resonance frequencies. The coupling ratio extracted from the experimental spectra, the branch GRU decoder that generates the resonance frequencies and the predicted spectra is 150.1, 145.6 and 147.9 GHz, respectively. This agreement confirms not only the precise forecast of the cascaded GRU networks but also that the generated hidden states are related to interpretable physical observables, which are resonance frequencies. Moreover, the visualization of the hidden states (see Methods section and Supplementary Fig. 8a for details) features a clustering of the time-domain data in the test dataset under different magnetic fields into three categories: one for the data in the range from 0 to 1.5 T, one for the data in the range from 1.6 to 2.6 T and the rest for high fields from 2.7 to 4.5 T. As discussed in ref.³³ and shown in Supplementary Fig. 8b, when the magnetic field increases to 1.5 T, the CRA mode starts to move into the transmission band of the photonic-crystal cavity, and the CRA mode starts to move out of the transmission band to the second stop band of the photonic-crystal cavity when the magnetic field increases to 2.6 T. The interaction of the strongly absorptive CRA mode with the large signals in the transmission band can lead to the substantial change of the time-domain signals. The clear clustering of the hidden states is fully consistent with the interaction of the CRA mode with the transmission band, and their connection with interpretable physical observables suggests that our cascaded GRU networks successfully capture the damped oscillatory nature of time-domain signals in optical resonances.

With the obtained predicted dispersion, the specific magnetic fields corresponding to the conditions at zero detuning and far away from zero detuning can be determined. As shown in Fig. 4g–i, we predicted the spectra of the CRI mode at 3 T and the CRA-UP and CRA-LP peaks at zero detuning by using the final stage of the trained GRU networks and the corresponding input sequences. The spectral line width determined from the Lorentzian fitting is 4.6, 5.2 and 5.2 GHz, respectively. Thus, the values obtained for the cooperativity from the forecast spectra (3,663) and from model inference (3,550) are close to the reported experimental value (3,513) (ref.³³), with variation of 4% and 1%, respectively. The long-sequence forecast capability of our cascaded GRU networks with short input sequences to accelerate experimental data acquisition can be impacted by additional experimental noise and the change of the time-domain sampling rate. We artificially added white noise to the original experimental time-domain signals and evaluated the prediction performance. When the signal-to-noise (SNR) ratio drops below ~40% of the SNR of the current experimental data, the prediction performance of our model starts to degrade substantially (see Methods section and Supplementary Fig. 9 for mode details). From the perspective of experimentalists, it is helpful to average more measurement results to reduce random noise when employing our model. Furthermore, the increase of the time sampling rate gradually increases the prediction loss, possibly due to the increasing number of data points. Thus, it seems beneficial to choose a time sampling rate close to the Nyquist rate. However, this choice is more dependent on the purpose of the experiments (see Methods section and Supplementary Fig. 10 for more details).

Discussion

In comparison with the demonstrations of the resonant dielectric metasurfaces and active graphene plasmonics based on numerical FDTD simulations, the forecast performance of our cascaded GRU networks in experimental data of Landau polaritons is less optimal. This is largely due to the measurement noise associated with signals, the small energy of the signals of interest and the small number of high-fidelity training data. Future improvements could be achieved by introducing more advanced RNN models such as transformers, and more diverse input features such as experiment parameters and cavity structures in addition to time sequences. The high-fidelity training dataset could be augmented through generative networks such as generative adversarial networks^{41,42} and variational auto-encoders⁴³ for better performance. Furthermore, an additional frequency-domain term in the frequency range of interest could be added to the loss function to better capture small signals. The applicability of cascaded GRU networks could also be extended to infer more physical observables and equations and to the discovery of new phenomena in complex systems through analogies with more accessible optical systems⁴⁴ and the exploration of device functionalities under resonant light–matter interaction.

Methods

High-fidelity training data generation. FDTD time-domain simulations were performed using commercial Ansys Lumerical software and applied to generate the high-fidelity training data for the dielectric metasurface and graphene plasmonics examples. For the dielectric metasurfaces, we use four silicon cylindrical pillars as the base structure with periodic boundary conditions. The time-domain data of the electric field along one polarization were obtained by simulating structures with randomly selected radii for the four cylindrical pillars. The radius ranged from 39.5 to 44.5 μm in steps of 0.25 μm . The pillar height was set as 30 μm . We generated a total number of 5,000 samples, using 4,000 of them as the training set and the remaining 1,000 as the test set. The reference was simulated once, being the time-domain signal of a bare silicon substrate without any structures on top. For the graphene plasmonics, the graphene monolayer is modelled as a two-dimensional (2D) rectangular conducting sheet from the Ansys Lumerical material library, including both inter- and intraband contributions. The Fermi level and scattering rate are two parameters used to calculate the dielectric constants used by the software. The dataset was generated by randomly sweeping the width and Fermi level of the graphene ribbon. The width ranged between 3.8 and 13.8 μm in steps of 0.2 μm . The Fermi level ranged from 0.18 to 0.41 eV in steps of 0.01 eV. The scattering rate was set as 0.00099 eV. We generated a total number of 1,000 samples, using 800 of them as the training set and the remaining 200 as the test set. The reference was taken once, being the time-domain signal of a bare silicon oxide substrate without graphene structures. In both examples, the structural parameters of the cylindrical pillars and graphene material properties were chosen such that the resonance features were located in the THz range (0.1–10 THz). The detailed physical mechanism governing the resonance frequency can be found in refs.³⁷ and ³⁸, respectively.

Loss function. In all the dielectric metasurface, graphene plasmonics and Landau polaritons examples with all types of models, we use the following MSE loss function:

$$\text{MSELoss} = \sum_{i=1}^n (E_i - E'_i)^2, \quad (1)$$

where E_i is the electric field value forecast at the certain time index i and E'_i is the target electric field value at the same time index. Both E_i and E'_i are real values. We use the electric field value along only one polarization direction, which is also consistent with standard THz-TDS measurements. n is the total number of time steps in the sequences to be predicted. The MSE loss calculated on a training dataset is called the training MSE loss, while the MSE loss calculated on a test dataset is called the test MSE loss. Except for Fig. 2c, which is presented in terms of the training MSE loss, all the time-domain signals and the corresponding frequency-domain spectra are within test datasets, and the comparison of different models and model parameters is made in terms of the test MSE loss. Supplementary Fig. 1a compares the training loss and the test loss of one network stage in the cascaded GRU networks trained using the multi-fidelity approach. The negligible difference between these two losses confirms that our model is not overfitting⁷.

Seq2seq model. The standard seq2seq model that we implemented using the GRU architecture is illustrated in Supplementary Fig. 1b. The encoder block takes the input sequence and extracts the encoded hidden states. The extracted hidden states

along with the last input sequence x_i , are then fed into the decoder block to predict the next value x_{i+1} . The decoding process can propagate infinitely since it only requires the previous input value and the previous hidden state. The test MSE loss is shown in Supplementary Fig. 1c.

Low-fidelity synthetic pre-training data generation. Generally, optical resonance features follow the mathematical representation of a sum of multiple damped oscillations because of FID. Thus, we created synthetic data for the pre-training step in the two-step multi-fidelity training approach using the equation

$$f(t) = A_1 e^{-\alpha_1 t} \sin(\omega_1 t) + A_2 e^{-\alpha_2 t} \sin(\omega_2 t) + \dots + A_n e^{-\alpha_n t} \sin(\omega_n t). \quad (2)$$

For the pre-training data for the dielectric metasurfaces, the number of sinusoidal signals used was 2. The sweep range for α_1 and α_2 was from 0.002 to 0.01 and from 0.05 to 0.2, respectively. The sweep range for ω_1 and ω_2 was from 0.4π to 2π . For A_1 and A_2 , the ranges were from 1.72 to 3.5 and from 150 to 220, respectively. The total number of synthetic time-domain signals generated was 40,000. For the pre-training data of graphene plasmonics, we used one sinusoidal signal. The sweep range for α was from 0.03 to 0.07. The sweep range for ω was from 0.09π to 0.16π . A was fixed as 1. The total number of synthetic data generated was 8,000. For the pre-training data for the Landau polaritons, the number of sinusoidal signals used was 2. The sweep ranges for α_1 and α_2 were from 0.0001 to 0.0002 and from 0.055 to 0.065, respectively. The sweep range for ω_1 and ω_2 was from 1.8π to 2.2π . The variables A_1 and A_2 were set to be 1. The total number was 3,000. Supplementary Fig. 2 shows two examples of synthetic time signals generated by using the equation given above for dielectric metasurfaces. When we pre-train our cascaded GRU networks, pre-training data are divided into segments (Data pre-processing) to first train the model.

Training of RNNs with GRU. The multi-fidelity training consists of two steps: pre-training and fine training (transfer learning). During the pre-training stage, the synthetic data are divided into segments and fed into the model to first tune the random weights. In the fine-training process, we first use the pre-trained model as a starting model, which is from the physics-informed initialization process. All the weights of the pre-trained networks are trainable in the fine-training process, and the high-fidelity data from FDTD simulations or experimental measurements are used to further update weights without constraints. The detailed hyperparameters for both the pre-training and fine-training processes of all the demonstrations are summarized in Supplementary Tables 1–3.

In the example of Landau polaritons, once the model for the time dynamics forecast is fully trained, a branch GRU decoder is connected to the trained encoder of the second GRU to receive the encoded hidden states and the processed time signal outputs from the previous stage. We trained this branch GRU network with resonance frequency labels associated with the time-domain signals, while we kept the weights of the other cascaded GRU networks unchanged and only updated this branch GRU decoder. The decoder is a GRU network with four hidden layers, similar to the decoders used in the time-domain signal forecasting networks. This branch GRU network was trained using the Adam optimizer with a learning rate of 5×10^{-4} . The total number of epochs for the training was 300, with a batch size of 1. The learning rate decayed every 100 epochs with a decay rate of 0.2.

Data pre-processing. In the example of dielectric metasurfaces, the full time sequence with a total of 1,600 data points was divided into six segments: [0, 200] (input), [200, 400], [400, 600], [600, 800], [800, 1200] and [1200, 1600]. We also tried another segmentation method that divided the signal into four segments: [0, 200] (input), [200, 400], [400, 800] and [800, 1600]. These two segmentation approaches are shown in Supplementary Fig. 1d. The test MSE loss was calculated for both segmentation approaches (Supplementary Fig. 1c). The MSE loss for the test data was 0.0135 and 0.0169 for the first and second segmentation method, respectively. This suggests that the model performance shows little dependence on the signal segmentation approach. We also evaluated the model prediction performance as a function of the input sequence length for all the demonstrations, so that we can select the shortest input length while maintaining high prediction accuracy (that is, low MSE loss). The results are summarized in Supplementary Fig. 3. Furthermore, since the electric field values obtained from the FDTD simulations and experiments are small, a simple scaling of the signals is performed before training. In other words, the original signals are multiplied by a scaling factor. For the example of the dielectric metasurfaces, the simulated time-domain signals were pre-processed by multiplying by a scaling factor of 1,000. For the graphene plasmonics, the full time sequence with a total of 600 data points was divided into three segments: [0, 150] (input), [150, 300] and [300, 600]. The scaling factor used in this example was 100. For the Landau polaritons, the full time sequence with a total of 950 data points was divided into four segments: [0, 125] (input), [125, 250], [250, 500] and [500, 950]. The scaling factor was 1×10^6 .

Transfer matrix method. We used the same transfer matrix method as described in ref. ³³ to fit the experimental and predicted spectra. Since our one-dimensional (1D) photonic-crystal cavity integrated with the multiple quantum well (QW) structure had translational symmetry within the x - y plane of the sample (where z is along the Bragg mirror stacking direction, that is, the multiple QW

growth direction), we were able to use the transfer matrix method to reproduce experimental transmission spectra. For an electromagnetic wave normally incident onto an isotropic multilayer structure, the complex transmission coefficient t and reflection coefficient r satisfy

$$\begin{pmatrix} t \\ 0 \end{pmatrix} = Q \begin{pmatrix} 1 \\ r \end{pmatrix}. \quad (3)$$

Here, Q is the 2×2 transfer matrix calculated from cascading multiplications of the matrices of the different layers

$$Q = M_{N,N-1} \cdot P_{N-1} (d_N - 1) \cdot M_{N-1,N-2} \dots M_{2,1} \cdot P_1 (d_1) \cdot M_{1,0}, \quad (4)$$

where M and P represent an interface matrix and a propagation matrix, respectively, d is the layer thickness and the subscripts are layer indexes that range from 0 to N . t and r can be calculated as $t = Q_{11} - (Q_{12} Q_{21} / Q_{22})$ and $r = -Q_{21} / Q_{22}$, and the power transmittance and reflectance are $T = |t|^2$ and $R = |r|^2$, respectively.

Material parameters enter equation (4) through the refractive index of the N th layer n_N in the M and P matrices:

$$M_{N,N-1} = \frac{1}{2} \begin{pmatrix} 1 + n_{N-1}/n_N & 1 - n_{N-1}/n_N \\ 1 - n_{N-1}/n_N & 1 + n_{N-1}/n_N \end{pmatrix}, \quad (5)$$

$$P_N = \begin{pmatrix} e^{i n_N \frac{\omega}{c} d_N} & 0 \\ 0 & e^{-i n_N \frac{\omega}{c} d_N} \end{pmatrix}. \quad (6)$$

We used $n_{\text{Si}} = 3.4$ for silicon and $n_0 = 1$ for the vacuum spacings. For the 2DEG layer, we first calculated the direct-current surface conductivity from the expression $\sigma_{\text{DC}} = ne\mu$, where n is the total surface electron density and $\mu = e\tau/m^*$ is the electron mobility. The elements of the Drude conductivity tensor of the 2DEG in a perpendicular magnetic field are given by

$$\sigma_{xx} = \frac{\sigma_{\text{DC}}(1 - i\omega\tau)}{(1 - i\omega\tau)^2 + (\omega_c\tau)^2}, \quad \sigma_{xy} = -\frac{\sigma_{\text{DC}}\omega_c\tau}{(1 - i\omega\tau)^2 + (\omega_c\tau)^2}. \quad (7)$$

In the circular polarization basis, the conductivity eigenvalues for the CR-active and CR-inactive polarization modes, respectively, are expressed as

$$\sigma_{\text{CRA}} = \sigma_{xx} + i\sigma_{xy} = \frac{\sigma_{\text{DC}}}{1 - i(\omega - \omega_c)\tau}, \quad \sigma_{\text{CRI}} = \sigma_{xx} - i\sigma_{xy} = \frac{\sigma_{\text{DC}}}{1 - i(\omega + \omega_c)\tau}. \quad (8)$$

The bulk dielectric permittivity and refractive index of the 2DEG layer for the CRA and CRI polarization modes are then calculated as

$$\epsilon_{\text{CRA}} = \epsilon_{\text{bg}} + i\sigma_{\text{CRA}}/(\epsilon_0\omega d_{\text{QW}}), \quad \epsilon_{\text{CRI}} = \epsilon_{\text{bg}} + i\sigma_{\text{CRI}}/(\epsilon_0\omega d_{\text{QW}}). \quad (9)$$

$$n_{\text{CRA}} = (\epsilon_{\text{CRA}})^{1/2}, \quad n_{\text{CRI}} = (\epsilon_{\text{CRI}})^{1/2}, \quad (10)$$

where we chose the background dielectric permittivity $\epsilon_{\text{bg}} = 3.6^2$ to be the same as that of GaAs, and d_{QW} is the total thickness of the multiple QW membrane.

The above material parameters, combined with experimental cavity structure parameters such as the layer thicknesses and separations, allowed us to calculate transmission spectra as a function of the magnetic field. We can then extract the coupling rate g by following the supplementary information of ref. ³³.

Visualization of hidden states in cascaded GRU networks. We utilized t -distributed stochastic neighbour embedding to visualize the hidden states in the middle stage of a GRU network, here reducing the hidden states from 200 to 2 dimensions. The perplexity was set as 5, and the learning rate as 100. Supplementary Fig. 8a clearly highlights three types of time-domain data in the test dataset: the data in the range of 0–1.5 T, the data in the range of 1.6–2.6 T and the data in the range of 2.7–4.5 T. As shown in Supplementary Fig. 8b adapted from ref. ³³, when the magnetic field increases to around 1.5 T, the CRA mode starts to move out of the stop band (black area) into the transmission band (yellow white area). When the magnetic field further increases to around 2.6 T, the CRA mode starts to move out of the transmission band of the 1D photonic-crystal cavity (yellow white area) into the second stop band (black area). The signal is strong in the transmission band, so that the strongly absorptive CRA feature in the transmission band would lead to a substantial influence on the time-domain signals. The clustering of the hidden states in the GRU networks, which is fully consistent with the interaction of the CRA mode with the transmission band, strongly suggests that our GRU networks can learn the features of time-domain signals following the strong-coupling physics in Landau polaritons. In addition, the connection of hidden states with physical observables (that is, resonance frequencies) further suggests that the GRU networks learn about the nature of oscillating time-domain signals.

Signal noise in Landau polaritons. A white noise was intentionally added on top of the original time-domain signal as $\hat{a}(t) = a(t) + n(t)$, where $n(t)$ is the noise signal, modelled based on the original signal $a(t)$ as $n(t) = A \times N(0, \text{std}(a(t)))$. Here, $\text{std}(a(t))$ stands for the standard deviation of the original signal and $N(0, \sigma)$ is a normal distribution with mean of zero and standard deviation of σ . A is the noise strength and was swept from 0 to 0.2. The transmission spectra calculated from the corresponding time-domain signals for various values of A are shown in Supplementary Fig. 9a. To better connect with experimental settings, we define a SNR ratio calculated from the spectra of Landau polaritons. Specifically, as shown in Supplementary Fig. 9a, the spectra under magnetic fields between 2.5 and 3.0 T in the range from 0.41 to 0.55 THz (shaded area) were selected to evaluate the SNR. Within this frequency range, there is only the CRI peak and the signal above 0.45 THz should be zero if there is no noise. Thus, we define the peak intensity as the signal strength and the average intensity from 0.45 to 0.55 THz as the noise strength. The SNR is defined as the ratio of the defined signal strength over the defined noise strength, and we take the average SNR for all the spectra between 2.5 and 3.0 T. We first calculated the signal strength for the case with $A = 0$, then use the same signal strength for all other values of A . The noise strength is evaluated for each value of A . For the dispersion with different values of A shown in Supplementary Fig. 9a, $A = 0$ corresponds to an SNR of 60.45, $A = 0.05$ corresponds to an SNR of 22.49, $A = 0.1$ corresponds to an SNR of 6.86 and $A = 0.2$ corresponds to an SNR of 1.85.

We used the test MSE loss to quantify the influence of the added noise on the model performance. Here, we explored two methods. The first method was to train the model using signals without noise and do the prediction with the input sequence with noise (that is, to train with $a(t)$ but infer with $\hat{a}(t)$). This corresponds to the red dots in Supplementary Fig. 9b. The second method is to train the model using the signals with additional artificial noise ($n(t)$ term) and do the prediction (that is, to both train and infer with $\hat{a}(t)$). This corresponds to the blue dots in Supplementary Fig. 9b. In both cases, when the strength of the added noise was $A \geq 0.05$ (SNR ≤ 22.49), we observed that the MSE loss became noticeably increasing and the spectra became noticeably noisy, as is clear from Supplementary Fig. 9a. In an experimental setting, $A = 0$ corresponds to a SNR in a specific standard THz-TDS setup. If the SNR drops below $\sim 40\%$ of the value under normal operation, the prediction performance of the cascaded GRU models could start to degrade substantially.

Signal sampling rate in Landau polaritons. The sampling rate of the experimental measurement of Landau polaritons is already at the Nyquist rate. The total THz bandwidth is 2.5 THz (Supplementary Fig. 5a), and the time sampling rate is 0.2 ps. From the experimental perspective, further downsampling below the Nyquist rate is generally already avoided by experimentalists to prevent aliasing. However, time-domain signals can be further downsampled by picking data points every n points, where n is an integer referred to as the downsampling factor. These signals are also upsampled by linearly interpolating m points between neighbouring points. We both trained and tested cascaded GRU networks with these re-sampled data. For downsampling, we swept n from 2 to 20. For upsampling, we chose $m = 1$, thus the length of the sequence doubles (that is, the downsampling factor is 0.5). A few downsampled time-domain signals are shown in Supplementary Fig. 10a. Supplementary Figure 10b summarizes the test MSE loss for various downsampling factors. When n lies in the range of 2–10, the MSE loss is reduced compared with the original sequence ($n = 1$). This could be because of the reduction of the total amount of data, which could be beneficial for the training process. A further increase of n (to 20) leads to a situation in which the damped oscillatory nature of the signals cannot be accurately captured and the MSE loss increases substantially. On the other hand, for upsampled signals ($n = 0.5$), the MSE loss increases as well. This can be attributed to the increasing difficulty of training the model with a growing number of parameters. From this analysis and experimental considerations, a sampling rate close to the Nyquist rate is desirable for both experiments and the prediction performance of the cascaded models.

Data availability

The source data of all figures in both main text and Supplementary Information are available at https://github.com/GaoUtahLab/Cascaded_GRU_Networks. The Zenodo version is available at ref. ⁴⁵. Source data for Figs. 2–4 is available with this manuscript.

Code availability

The code for the models in all three demonstrations of optical resonances and that support the plots within this paper and other findings of this study is available at https://github.com/GaoUtahLab/Cascaded_GRU_Networks. The Zenodo version is available at ref. ⁴⁵.

Received: 6 September 2021; Accepted: 16 February 2022;

References

- Mudelsee, M. Trend analysis of climate time series: a review of methods. *Earth Sci. Rev.* **190**, 310–322 (2019).

- Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).
- Lim, B. & Zohren, S. Time-series forecasting with deep learning: a survey. *Philos. Trans. R. Soc. A* **379**, 20200209 (2021).
- Box, G. E., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. *Time Series Analysis: Forecasting and Control* (Wiley, 2015).
- Goodfellow, I., Bengio, Y., Courville, A. & Bengio, Y. *Deep Learning*, vol. 1 (MIT Press, 2016).
- Fu, R., Zhang, Z. & Li, L. in *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)* 324–328 (IEEE, 2016).
- Kong, W. et al. Short-term residential load forecasting based on lstm recurrent neural network. *IEEE Trans. Smart Grid* **10**, 841–851 (2017).
- Nelson, D. M., Pereira, A. C. & de Oliveira, R. A. in *2017 International Joint Conference on Neural Networks (IJCNN)* 1419–1426 (IEEE, 2017).
- Hyndman, R. J. A brief history of forecasting competitions. *Int. J. Forecast.* **36**, 7–14 (2020).
- Zhou, H. et al. Informer: beyond efficient transformer for long sequence time-series forecasting. Preprint at <https://arxiv.org/abs/2012.07436> (2020).
- Jiang, J., Chen, M. & Fan, J. A. Deep neural networks for the evaluation and design of photonic devices. *Nat. Rev. Mater.* 1–22 (2020).
- Ma, W. et al. Deep learning for the design of photonic structures. *Nat. Photonics* 1–14 (2020).
- Xie, N., Ras, G., van Gerven, M. & Doran, D. Explainable deep learning: a field guide for the uninitiated. Preprint at <https://arxiv.org/abs/2004.14545> (2020).
- Rangapuram, S. S. et al. in *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS 2018)* 7796–7805 (2018).
- Salinas, D., Flunkert, V., Gasthaus, J. & Januschowski, T. DeepAR: probabilistic forecasting with autoregressive recurrent networks. *Int. J. Forecast.* **36**, 1181–1191 (2020).
- Willard, J., Jia, X., Xu, S., Steinbach, M. & Kumar, V. Integrating scientific knowledge with machine learning for engineering and environmental systems. Preprint at <https://arxiv.org/abs/2003.04919> (2020).
- Karniadakis, G. E. et al. Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440 (2021).
- Pellizzari, T., Gardiner, S. A., Cirac, J. I. & Zoller, P. Decoherence, continuous observation, and quantum computing: a cavity qed model. *Phys. Rev. Lett.* **75**, 3788 (1995).
- Cirac, J. I., Zoller, P., Kimble, H. J. & Mabuchi, H. Quantum state transfer and entanglement distribution among distant nodes in a quantum network. *Phys. Rev. Lett.* **78**, 3221 (1997).
- Duan, L.-M., Lukin, M. D., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001).
- Mabuchi, H. & Doherty, A. Cavity quantum electrodynamics: coherence in context. *Science* **298**, 1372–1377 (2002).
- Englund, D. et al. Controlling cavity reflectivity with a single quantum dot. *Nature* **450**, 857–861 (2007).
- O'Brien, J. L., Furusawa, A. & Vučković, J. Photonic quantum technologies. *Nat. Photonics* **3**, 687–695 (2009).
- Forn-Díaz, P., Lamata, L., Rico, E., Kono, J. & Solano, E. Ultrastrong coupling regimes of light–matter interaction. *Rev. Mod. Phys.* **91**, 025005 (2019).
- Siegel, P. H. Terahertz technology in biology and medicine. *IEEE Trans. Microw. Theory Techn.* **52**, 2438–2447 (2004).
- Tonouchi, M. Cutting-edge terahertz technology. *Nat. Photonics* **1**, 97–105 (2007).
- Mittleman, D. M. Twenty years of terahertz imaging. *Opt. Express* **26**, 9417–9431 (2018).
- Liu, H.-B., Zhong, H., Karpowicz, N., Chen, Y. & Zhang, X.-C. Terahertz spectroscopy and imaging for defense and security applications. *Proc. IEEE* **95**, 1514–1527 (2007).
- Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
- Li, Y., Chen, R., Sensale-Rodriguez, B., Gao, W. & Yu, C. Real-time multi-task diffractive deep neural networks via hardware-software co-design. *Sci. Rep.* **11**, 11013 (2021).
- Kleine-Ostmann, T. & Nagatsuma, T. A review on terahertz communications research. *J. Infrared Millim. Terahertz Waves* **32**, 143–171 (2011).
- Zhang, Q. et al. Collective non-perturbative coupling of 2D electrons with high-quality-factor terahertz cavity photons. *Nat. Phys.* **12**, 1005–1011 (2016).
- Li, X. et al. Vacuum Bloch–Siegert shift in Landau polaritons with ultra-high cooperativity. *Nat. Photonics* **12**, 324–329 (2018).
- Ulbricht, R., Hendry, E., Shan, J., Heinz, T. F. & Bonn, M. Carrier dynamics in semiconductors studied with time-resolved terahertz spectroscopy. *Rev. Mod. Phys.* **83**, 543 (2011).
- Jepsen, P. U., Cooke, D. G. & Koch, M. Terahertz spectroscopy and imaging—modern techniques and applications. *Laser Photonics Rev.* **5**, 124–166 (2011).
- Neu, J. & Schmittenmaier, C. A. Tutorial: an introduction to terahertz time domain spectroscopy (THz-TDS). *J. Appl. Phys.* **124**, 231101 (2018).

37. Nadell, C. C., Huang, B., Malof, J. M. & Padilla, W. J. Deep learning for accelerated all-dielectric metasurface design. *Opt. Express* **27**, 27523–27535 (2019).
38. Ju, L. et al. Graphene plasmonics for tunable terahertz metamaterials. *Nat. Nanotechnol.* **6**, 630–634 (2011).
39. Sutskever, I., Vinyals, O. & Le, Q. V. Sequence to sequence learning with neural networks. *Adv. Neural Inform. Process. Syst.* 3104–3112 (2014).
40. Allen, L. & Eberly, J. H. *Optical Resonance and Two-Level Atoms*, vol. 28 (Courier Corporation, 1987).
41. Liu, Z., Zhu, D., Rodrigues, S. P., Lee, K.-T. & Cai, W. Generative model for the inverse design of metasurfaces. *Nano Lett.* **18**, 6570–6576 (2018).
42. Jiang, J. et al. Free-form diffractive metagrating design based on generative adversarial networks. *ACS Nano* **13**, 8872–8878 (2019).
43. Tang, Y. et al. Generative deep learning model for inverse design of integrated nanophotonic devices. *Laser Photonics Rev.* **14**, 2000287 (2020).
44. Li, X. et al. Observation of Dicke cooperativity in magnetic interactions. *Science* **361**, 794–797 (2018).
45. Tang, Y. et al. Physics-informed recurrent neural network for time dynamics in optical resonances (code). <https://doi.org/10.5281/zenodo.6058054>

Acknowledgements

J.F. and W.G. thank the University of Utah start-up fund for support. X.L. acknowledges support from the Caltech Postdoctoral Prize Fellowship and the IQIM. C.Y. acknowledges support from grants NSF-2047176 and NSF-2008144.

Author contributions

C.Y. and W.G. conceived the idea and designed the project. Y.T. performed the modelling and calculations with the help of J.F., J.M., M.Q., C.Y. and W.G. X.L. helped with the analysis of the Landau polariton data. Y.T. and W.G. wrote the manuscript. All authors discussed the manuscript and provided feedback.

Competing interests

The authors declare that they have no competing financial interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43588-022-00215-2>.

Correspondence and requests for materials should be addressed to Cunxi Yu or Weilu Gao.

Peer review information *Nature Computational Science* thanks Andrey Baydin, Bowen Zheng and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Kaitlin McCardle, in collaboration with the *Nature Computational Science* team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022