Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

HPC-Driven Modeling with ML-Based Surrogates for Magnon-Photon Dynamics in Hybrid Quantum Systems

Jialin Song
Simon Fraser University
Yingheng Tang
Lawrence Berkeley National Laboratory
Pu Ren
Lawrence Berkeley National Laboratory
Shintaro Takayoshi
Konan University
Saurabh Sawant
Lawrence Berkeley National Laboratory
Yujie Zhu
University of Wisconsin–Madison
Jiamian Hu
University of Wisconsin–Madison
Andy Nonaka
Lawrence Berkeley National Laboratory
Michael Mahoney
University of California, Berkeley
Benjamin Erichson
Lawrence Berkeley National Laboratory
Zhi (Jackie) Yao
jackie_zhiyao@lbl.gov

Lawrence Berkeley National Laboratory

Article

Keywords:

Posted Date: May 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9596588/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

HPC-Driven Modeling with ML-Based Surrogates for Magnon-Photon Dynamics in Hybrid Quantum Systems

Jialin Song[1,2], Yingheng Tang[3], Pu Ren[2], Shintaro Takayoshi[4], Saurabh Sawant[3], Yujie Zhu[5], Jia-Mian Hu[5], Andy Nonaka[3], Michael W. Mahoney[2,6,7], Benjamin Erichson[*2,6], and Zhi (Jackie) Yao[*3]

[1]*Simon Fraser University, 8888 University Dr W, Burnaby, BC V5A 1S6, Canada*
[2]*Scientific Data Division, Lawrence Berkeley National Laboratory, CA, 94720, USA*
[3]*Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory, CA, 94720, USA*
[4]*Konan University, Japan*
[5]*Department of Materials Science and Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA*
[6]*International Computer Science Institute, Berkeley, CA 94704, USA*
[7]*Department of Statistics, University of California, Berkeley, CA 94720, USA*

Simulating hybrid magnonic quantum systems remains a challenge due to the large disparity between the timescales of the two systems. We present a massively parallel GPU-based simulation framework that enables fully coupled, large-scale modeling of on-chip magnon-photon circuits. Our approach resolves the dynamic interaction between ferromagnetic and electromagnetic fields with high spatiotemporal fidelity. To accelerate design workflows, we develop a physics-informed machine learning surrogate trained on the simulation data, reducing computational cost while maintaining accuracy. This combined approach reveals real-time energy exchange dynamics and reproduces key phenomena such as anti-crossing behavior and the suppression of ferromagnetic resonance under strong electromagnetic fields. By addressing the multiscale and multiphysics challenges in magnon–photon modeling, our framework enables scalable simulation and rapid prototyping of next-generation quantum and spintronic devices.

1 Introduction

Hybrid quantum systems, which combine distinct physical platforms, are a promising route toward advanced quantum technologies, as they harness strong interactions that may not be readily achievable in a single platform [1, 2]. These systems take many forms, coupling any two (or more) quantum platforms — for example, superconducting qubits [3, 4], microwave resonators [5], single spins [6], spin ensembles [4, 7–9], or mechanical resonators [10–12] — to harness strong interactions. These heterogeneous systems leverage complementary advantages of each component, but their rich multi-physics interactions pose formidable modeling challenges. A prominent example is cavity magnonics, where collective spin excitations (magnons) couple with microwave photons in a resonant cavity to form hybrid magnon–polariton modes when tuned into resonance [13–15]. While an avoided crossing indicates hybridization, it is not sufficient for strong coupling; the system enters the strong-coupling regime only when the coupling rate g (half the on-resonance

*Email: jackie_zhiyao@lbl.gov, erichson@lbl.gov

splitting) exceeds the dissipation rates of the uncoupled magnon and photon modes, in which case energy oscillates coherently between them and the hybrid modes exhibit remarkable coherence owing to the high Q factors of microwave resonators and the narrow ferromagnetic resonance (FMR) linewidth [16]. These states are essential for quantum operations such as mode swapping [16, 17], quantum state storage [4, 18, 19], and dynamic control of energy exchange [19, 20]. The hallmark experimental signature of strong magnon–photon coupling is a pronounced avoided crossing (mode splitting) in the frequency spectrum, in agreement with theoretical predictions [21] and observed in many 3D [13, 22] and on-chip 2D [7, 8, 23] cavity based systems. In the time domain, the coupled system exhibits Rabi-like oscillations with very high energy exchange per cycle, reflecting the complete hybridization of the two excitations into a single quantum system [17]. Achieving practical quantum operations with magnon-photon systems requires precise dynamical control over energy transfer between the two hybrid magnon-polariton modes [24, 25]. Therefore, developing quantum devices that leverage these effects necessitates a detailed understanding of their time-domain dynamics.

However, accurately modeling the complex dynamics of such hybrid systems presents a significant computational challenge. Existing modeling approaches have been limited in scope. Most analytical treatments require inputting an assumed magnon–photon coupling strength, and the few numerical studies to date have often resorted to reduced dimensions (1D or 2D models) or macrospin approximations to mitigate computational cost [21]. Conventional micromagnetic modeling strategies rely on the magnetostatic approximation [26–28], which has provided valuable insights, but which typically neglects the dynamic nature of electromagnetic (EM) fields, including time-dependent displacement currents. The underlying physics of magnon-photon coupling spans multiple domains and scales, requiring a multiphysics simulation that solves Maxwell's equations for electromagnetic fields alongside the Landau–Lifshitz–Gilbert (LLG) equation for magnetization dynamics [17]. Solving these coupled equations self-consistently in three dimensions with realistic material parameters is computationally intensive, often necessitating high-performance computing (HPC) resources to achieve adequate spatial and temporal resolution [29]. The disparate time scales and length scales involved – from microwave photons to nanometer-scale spin precessions – further complicate direct simulation. Maxwell-LLG coupling is achieved by self-consistently updating the dynamic EM fields in sync with magnetization dynamics, thereby capturing the bidirectional interaction between microwave photons and magnons [17, 29–35]. The majority of such time-domain dynamical simulations employ either explicit schemes like finite-difference time-domain (FDTD) leapfrog integrators [17, 29, 32–34], or implicit schemes such as alternating-direction implicit (ADI) FDTD and Crank–Nicolson methods [30, 31]. While explicit FDTD is straightforward, it is only conditionally stable (restricted by the Courant limit), leading to significant oversampling of the EM wave in the time domain. To ease numerical constraints, researchers have explored strategies such as artificially increasing the medium permittivity—an approach that relaxes time-step oversampling [17]. However, this strategy might be difficult to extend to complex device structures with highly inhomogeneous local permittivity and conductivities, as in on-chip devices. On the other hand, implicit methods offer improved (often unconditional) stability, but they require solving more complex update steps. Additionally, implementing precise boundary conditions in implicit mathematical frameworks, such as perfectly matched layers

(PML), is challenging and limits scalability in real-world quantum devices [31]. HPC techniques, including GPU acceleration and parallel FDTD-LLG solvers combined with domain decomposition, significantly enhance simulation speed and scalability, enabling large-scale modeling on modern supercomputing architectures [29, 36]. Finite-element time-domain models that rely on commercial software have also been adopted, but they suffer from the overhead of assembling and solving large linear systems at each time step, leading to high CPU and memory usage [35]. Nonetheless, to the best of the authors' knowledge, dynamical time-domain simulation of practical sized, on-chip hybrid magnon–photon circuits has remained elusive [7, 8]. This underscores the need for more powerful computational methods to simulate hybrid magnon–photon systems in full detail.

The main numerical challenges arise from the need for large-scale three-dimensional (3D) simulations and fine meshes to accurately capture the detailed structures. In many practical scenarios, however, only a few spatial observation points are needed for circuit analysis tasks. This discrepancy underscores the potential for more efficient modeling strategies that focus resources on critical regions and time windows, reducing overall costs while preserving key physics. Machine learning (ML) has emerged as a powerful approach to tackle the computational challenges posed by high-dimensional and nonlinear dynamical systems [37, 38]. In particular, frameworks such as Physics-Informed Neural Networks (PINNs) [39, 40], Neural Operators (e.g., DeepONet [41] and Fourier Neural Operator [42]), and Neural Discrete Equilibrium (NeurDE) [43] have demonstrated strong capabilities in modeling partial differential equations (PDEs), by efficiently capturing their underlying spatial-temporal dynamics. Furthermore, recent advances in sequence modeling architectures, including Recurrent Neural Networks (RNNs) [44], state-space models (SSMs) [45, 46] and Transformers [47], have shown remarkable success in predicting complex time-evolving physical phenomena [48–55].

Despite significant advancements, accurately predicting long-horizon and multi-scale dynamics remains a challenge, particularly in coupled physical systems such as EM fields. Furthermore, classic data-driven ML approaches often require vast amounts of training data and struggle with generalization to out-of-distribution (OOD) scenarios, limiting their applicability in complex scientific domains.

In this work, we present a hybrid computational framework that combines HPC-driven numerical modeling with physics-informed ML-based surrogates to address these limitations. By combining short-duration high-fidelity numerical simulations with these data-driven models, one can learn the latent electric and magnetic field patterns, enabling reliable predictions of magnon-photon interactions across the entire spatial and temporal domain. This hybrid strategy drastically cuts down the need for computationally expensive numerical solvers to a limited time window, while using ML to extrapolate the remaining dynamics with embedded physics knowledge. Consequently, researchers can achieve significant speedups without sacrificing precision or scalability when exploring complex coupled phenomena. Specifically, our main contributions include the following.

- We present an HPC-enabled FDTD framework for large-scale 3D simulations, directly integrating Maxwell's equations with the LLG equation to dynamically model magnon-photon

interactions. This framework integrates short-duration simulation data with ML methods to enable fast and accurate surrogate modeling of long-term and multi-scale EM systems.

- We incorporate a domain-specific physics constraint into this hybrid ML framework to enhance training efficiency and ensure the preservation of fundamental physical principles, ultimately improving the OOD generalization capability.
- We have evaluated the performance of our proposed method across extensive numerical experiments. The results demonstrate the effectiveness of our hybrid method in terms of simulation accuracy and computational efficiency.

By incorporating a physics-informed ML network, our framework achieves highly accurate predictions of time-domain resonance responses using input sequences as short as 20% of their full length. This advancement enables up to $5\times$ acceleration, compared to standalone HPC simulations, representing a significant leap in both computational efficiency and predictive precision. Crucially, this hybrid modeling strategy combines first-principles physics-based simulation with opportunities for data-driven enhancement, aligning with the emerging paradigm of *scientific machine learning* in computational science. For example, the rich output data from our high-fidelity simulations could be used to train surrogate models or inform ML-based optimization routines, illustrating how data-driven methods can complement and accelerate traditional simulations. By uniting advanced multiphysics simulation with such data-driven techniques, our work showcases how innovations in computational science – from HPC-driven algorithms to modern ML – can broaden the exploration of complex quantum systems. Using the magnon–photon platform as a representative case, we demonstrate a cross-disciplinary approach that not only advances the modeling of a specific physical system but also exemplifies general strategies for tackling grand challenges in hybrid quantum computing and simulation.

2 Results

Magnon-Photon Hybrid Quantum Circuit Modeling Setup. The simulation setup is shown in Figure 1(a), where the core component is an on-chip coplanar waveguide (CPW) microwave resonator designed to generate and support photons. The metal conductivity is $1.0 \times 10^{11}\,\mathrm{S/m}$ to minimize resistive losses and approximate superconducting behavior, and the relative permittivity of the silicon substrate is $\epsilon_{\mathrm{Si}} = 11.4$. The metal layer is $2\,\mu\mathrm{m}$ thick, and the silicon substrate is $500\,\mu\mathrm{m}$ thick. The center line width is $50\,\mu\mathrm{m}$, and each lateral gap between the center line and the adjacent ground patches is $25\,\mu\mathrm{m}$. Such geometrical and material design results in a characteristic impedance of 50 Ω for the CPW transmission line. This CPW resonator is tuned to a fundamental mode at $f_{\mathrm{p}} = \omega_{\mathrm{p}}/2\pi = 15.2\,\mathrm{GHz}$, with a center conductor line length of $3.75\,\mathrm{mm}$. At both ends of the resonator (along $\pm y$), feeding metallic patches of $300\,\mu\mathrm{m}$ in length and $50\,\mu\mathrm{m}$ in width are implemented. Gap capacitors are formed by $200\,\mu\mathrm{m}$ gaps between the central conductor and the feeding conductors.

A magnetic thin film (red stripe in Figure 1(a)) is placed at the center of the conductor line. This film is $5\,\mu\mathrm{m}$ thick, $30\,\mu\mathrm{m}$ wide, occupying the center portion of the conductor, and $2.82\,\mathrm{mm}$ long, covering about three quarters of the resonator length. The saturation magnetization of the material is $4\pi M_S = 1.2 \times 10^4\,\mathrm{Gauss}$ (typical of Permalloy), and the Gilbert damping factor is

$\alpha = 0.003$, indicating low magnetic loss. The electrical conductivity of the magnetic material is set to zero to avoid internal electrical losses, confining losses to the CPW resonator. A DC magnetic bias field H_0 is applied in the longitunidal direction, to saturate the ferromagnet to its saturation magnetiation M_S. The value of H_0 ranges from 1500 Oersted to 2600 Oersted. All six boundaries of the simulation domain are set as PMLs. The top PML boundary in the z direction is placed $1\,\mathrm{mm}$ above the bottom, leaving $493\,\mu\mathrm{m}$ of vacuum above the magnetic thin film surface. The bottom PML boundary is applied to the lower surface of the silicon substrate, effectively extending the substrate thickness beyond its physical size. Similarly, the lateral metal ground planes (width $250\,\mu\mathrm{m}$) also have PML boundary conditions in the x direction, extending both the ground planes and the substrate. The longitudinal PML boundaries in the y direction also effectively extend the feeding patches and the associated substrate infinitely, eliminating reflections from both feeding patches. To excite magnon-photon polaritons in this structure, a voltage is applied at one end of the CPW from its center conductor to the lateral grounds (see Figure 1(a)). The voltage excitation follows a modified Gaussian pulse, given by $E_{\mathrm{in}} = \exp\left(-(t-3T_p)^2/2T_p^2\right)\cos(\omega_0 t)$, where $\omega_0 = 2\pi f_0$ centers the frequency at $16\,\mathrm{GHz}$, and $T_p = 0.0625\,\mathrm{ns}$ represents one period of the excitation. The simulation uses mesh sizes of $\Delta x \times \Delta y \times \Delta z = 5\,\mu\mathrm{m} \times 19\,\mu\mathrm{m} \times 1\,\mu\mathrm{m}$, and the domain dimensions are $x \times y \times z = 0.6\,\mathrm{mm} \times 4.75\,\mathrm{mm} \times 1\,\mathrm{mm}$, yielding a grid of $N_x \times N_y \times N_z = 120 \times 250 \times 1000$. A CFL factor of 0.9 sets the time step to $\Delta t = 2.94 \times 10^{-15}\,\mathrm{s}$.

In the fundamental mode, the CPW resonator has electric-field peaks at both ends of the center conductor and a magnetic-field peak in the middle (Figure 1(h)). Placing the ferromagnet on top disturbs the magnetic field **H**, creating an irregular vortex in regions where the ferromagnet is present (Figures 1(b)–(c)), while regions without ferromagnet retain the regular vortex (Figure 1(d)). Figures 1(f)–(g) further show that the **H** fields above and below the center metal point in opposite directions, consistent with Figures 1(b)–(d). In contrast, the ferromagnet minimally affects the electric field **E**, which primarily spans from the center conductor to the ground and forms a standard standing-wave pattern with peaks at the ends and a null at the center (Figure 1(e)).

Fully-Coupled Dynamic Numerical Framework and Physics-Informed ML Surrogates. We employ a fully coupled dynamic Maxwell–LLG solver, ARTEMIS, based on the approach in [29], using an explicit FDTD leap-frog scheme to update the electric field **E**, magnetic field **H**, and magnetization **M**. In nonmagnetic regions, only Maxwell's equations are solved (with **M** set to zero). In magnetic regions, Maxwell's equations and the LLG equation are solved self-consistently in a mutual feedback loop (see Numerical Model in Methods for details). All field variables computed by the solver—**E**, **H**, and **M**—achieve second-order accuracy, owing to the trapezoidal time-domain discretization of the LLG equation and the leap-frog time-marching scheme for Maxwell's equations. To handle large problem sizes, we implement both multicore and GPU parallelization for the spatial mesh, as illustrated in Figure 2(a). We use the AMReX [56] library to partition the computational domain into non-overlapping grids, each assigned to an MPI rank. AMReX follows an MPI+X parallelization model, where *X* can be OpenMP (for multicore CPUs) or CUDA (for GPUs).

The input data for the physics-informed ML model is collected from nine discrete spatial locations, shown in Figure 2(b). These points lie in the middle of the ferromagnet layer along

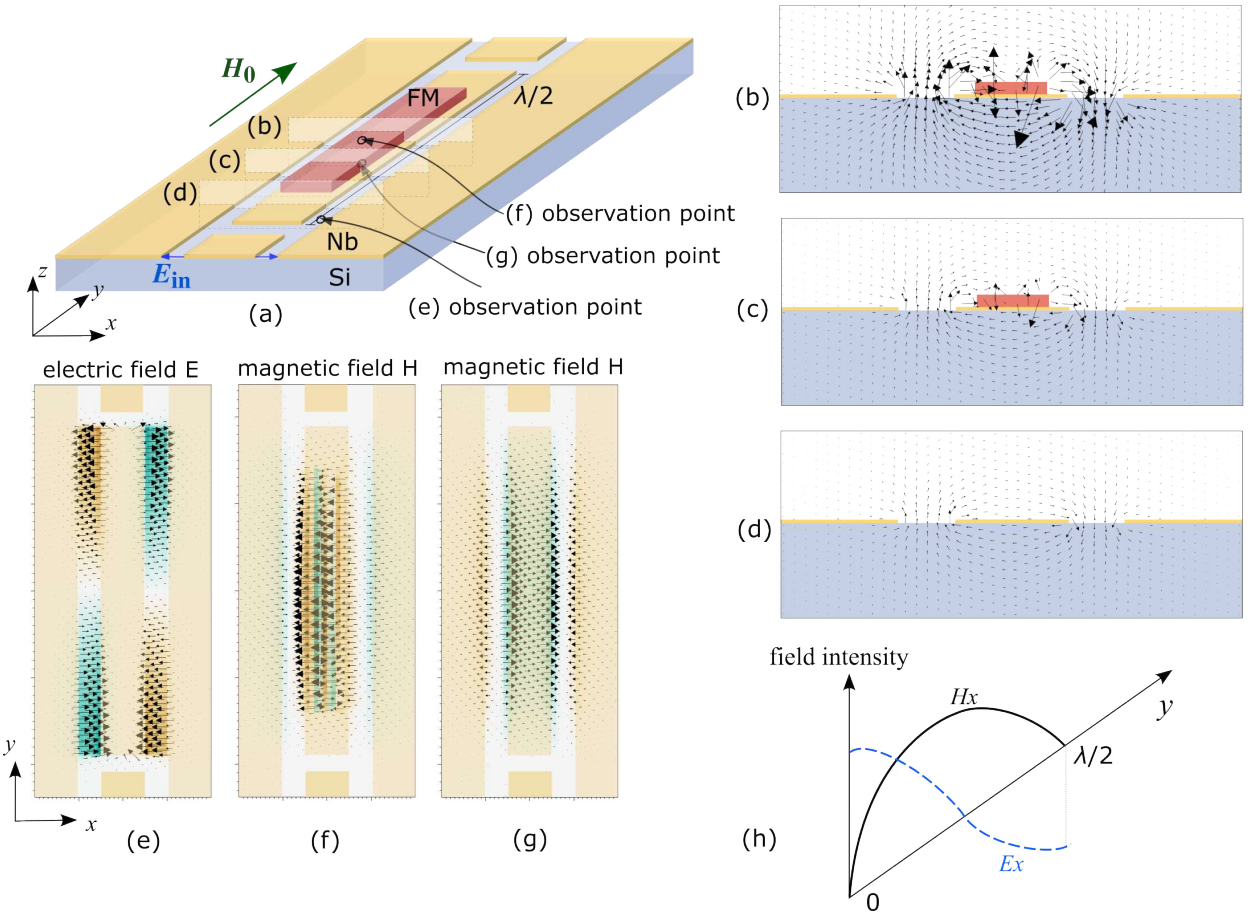


Figure 1: Magnon-photon hybrid quantum circuit modeling setup and simulated field profiles. (a) Perspective view of the magnon-photon coupling circuit, comprising a CPW resonator and a ferromagnetic thin film. The CPW resonator consists of a center conductor and lateral ground conductors (yellow) positioned on a silicon substrate (blue). The ferromagnet, shown as a red stripe, is placed on top of the center conductor. Material and geometrical parameters are detailed in the main text. (b)–(d) Cross-sectional views of the circuit, overlaid with the simulated magnetic field (**H**) distribution at different locations along the longitudinal direction of the CPW resonator, as indicated in (a). (e)–(g) Top views of the circuit, overlaid with the simulated electric and magnetic field distributions in the z-plane at different observation points, as marked in (a). (h) Theoretical field distributions along the longitudinal direction of the circuit, where the electric field E_x is concentrated in the lateral air gap, and the magnetic field H_x is observed above and beneath the center conductor.

its thickness and include the film's center, as well as the four corners and midpoints of the four edges in the lateral plane. More details can be found in the ML Training Implementation section of Methods. Figure 2(f) illustrates the general ML framework for sequential modeling of the magnetization. The model is based on the Long Expressive Memory (LEM) module [57], and it comprises an LEM encoder and decoder. The structure of the LEM cell is illustrated in Figure 2(d), which shows the computation flow across its internal states. At each time step, the input X_t, the previous hidden state H_{t-1} and latent state C_{t-1} are linearly projected and combined to generate two adaptive time-step parameters: Δt for updating the latent state C_t, and $\overline{\Delta t}$ for updating the hidden state H_t. These two time-step parameters act as gating coefficients, controlling the extent to which newly computed nonlinear activations (via $\tanh$) are blended into each state. The latent state C_t is updated first, capturing fast-evolving features from the input and hidden state. This updated C_t is then transformed and used to update the hidden state H_t, which serves as the main memory for downstream prediction. Additional implementation details and the mathematical formulation are provided in the LEM Cell of Methods. The LEM module processes the input sequence (200 temporal samples) using a stacked encoder composed of two LEM cells, which evolve coupled hidden H and latent states C with adaptive time constants. The final states from the encoder are used to initialize a two-stage decoder, which begins from a zero input and autoregressively generates a fixed-length output sequence. At each decoding step, the internal states are updated using the same LEM dynamics, and the top-layer hidden state is passed through a fully connected layer to produce the output. The model is trained using a composite loss function that integrates three component:

1. Reconstruction Loss: The Mean Squared Error (MSE) between the reconstructed input and the actual input sequence.
2. Prediction Loss: The MSE between the predicted and ground-truth output sequences.
3. Physics Loss: A term enforcing physical constraints, weighted by a coefficient λ (details provided in the Physics Loss section of Methods).

This combined approach enables the model to balance both data-driven objectives and physics-informed objectives.

Given the complexity of the hybrid quantum system, we incorporated Curriculum Learning [58] to improve generalization and help the model better capture long-term dynamics and the underlying physics. As shown in Figure 2(e), this approach enables the model to handle tasks of increasing complexity over time. During training, three key hyperparameters are dynamically adjusted:

Δt (dt): The sequence length for prediction.
lr: The model's learning rate.
λ: The weighting coefficient for the physics loss.

Training begins with simpler scenarios—shorter prediction sequences, no physics loss, and a higher learning rate—and gradually moves to more complex cases involving longer sequences, physics loss, and a lower learning rate. This staged approach ensures that the model generalizes effectively while capturing the intricate physics governing the field interactions. Details on how the training and testing datasets were curated are provided in the Datasets Curation section of

Methods.

Numerically Predicted Strong Magnon-Photon Coupling Phenomena. Our Maxwell-LLG dynamical modeling has predicted the dynamically evolving and interacting electric field and magnetization, shown in Figure 3. Figure 3(a) presents top-view snapshots of E_x (top row) and M_x (bottom row) at selected time instants, with an applied magnetic DC bias $H_0 = 2050$ Oersted. According to the Kittel's law, under such H_0, the FMR frequency is $f_{\mathrm{m}} = \omega_{\mathrm{m}}/2\pi = (\gamma/2\pi)\sqrt{H_0(H_0 + 4\pi M_S)} = 15.027$ GHz, detuned from the photon frequency f_{p} by about $\Delta = 200$ MHz. This small detune enables strong mode hybridization and, consequently, strong magnon-photon coupling. The E_x fields are observed at the same location described in Figure 1(e), i.e., along the vertical center plane passing through the CPW resonator's lateral air gap. The M_x fields are observed at the same location described in Figure 1(f), i.e., on the top surface of the ferromagnetic stripe. For additional details on these observation points, see Figure 1(a). At 0.0235 ns, the electric field appears only in the feeding patches, without exciting resonant modes (photons) in the CPW resonator, and no dynamic magnetization is observed in the ferromagnet. By 0.3528 ns, both E_x and M_x emerge in the resonator. Starting at 0.9819 ns, even though the input field excitation has vanished (the input signal completely decays by 0.5 ns, as shown in Figure 3(b)), oscillations of E_x and M_x persist within the CPW resonator. Notably, E_x peaks near the resonator's ends, whereas M_x peaks at the center. Both fields exhibit well-defined oscillations with only minor decay, indicating low damping—a result of minimal Gilbert damping, small metallic losses, and limited electromagnetic leakage into air or substrate. Furthermore, E_x and M_x alternate in reaching their standing-wave maxima: when E_x is at a peak, M_x is at a node, and vice versa. This phase alternation underscores strong energy exchange between the magnon and photon polaritons, more clearly seen in the transient response of Figure 3(c), and the yielded mode splitting in the Fourier transformed E_x spectra shown in Figure 3(d). Figure 3(e) visualizes the spectra of E_x magnitude under various H_0, revealing a clear anti-crossing behavior, indicating coherent coupling between magnons and photons [14]. The simulated spectrum matches theoretical analysis as shown in Extended Data Figure 1. The S-parameter ($|S_{12}|$) spectra shown in Figure 3(f) is converted from the E_x field following the workflow based on field calculation in [59], indicating the transmission rate from one end to the other of the microwave resonator, also showing an anti-crossing spectra. Both figures show mode splitting $g_{\mathrm{mp}} = 670$ MHz, where g_{mp} is the total magnon-photon coupling strength, defined as $g_{\mathrm{mp}} = g_0\sqrt{N}$, with g_0 defined as the coupling strength between photons and individual spins in the ferromagnet, and N is the number of spins in the ferromagnet [13, 21]. g_0 contributes to the total Hamiltonian of the system:

$$\hat{H} = \hbar\omega_{\mathrm{p}}\left(\hat{a}_p^{\dagger}\hat{a}_p + \tfrac{1}{2}\right) - \omega_{\mathrm{m}}(H_0)\,\hat{S}_y + g_0\left(\hat{S}_+\,\hat{a}_p^{\dagger} \;+\; \hat{S}_-\,\hat{a}_p\right), \tag{1}$$

where $\hat{a}_p^{\dagger}(\hat{a}_p)$ is the creation (annihilation) operator of microwave photon modes in the CPW resonator, and $\hat{\mathbf{S}} = -\hat{x}\frac{1}{2}(\hat{S}_+ + \hat{S}_-) + \hat{y}\hat{S}_y + \hat{z}\frac{1}{2}(\hat{S}_+ - \hat{S}_-)$ is the macrospin operator, with $\hat{S}_+(\hat{S}_-)$ raising (lowering) the y component of the macrospin. The eigen frequencies of hybrid system can be measured from Figure 3(e) and (f) as 14.95 GHz and 15.65 GHz, matching the theoretical

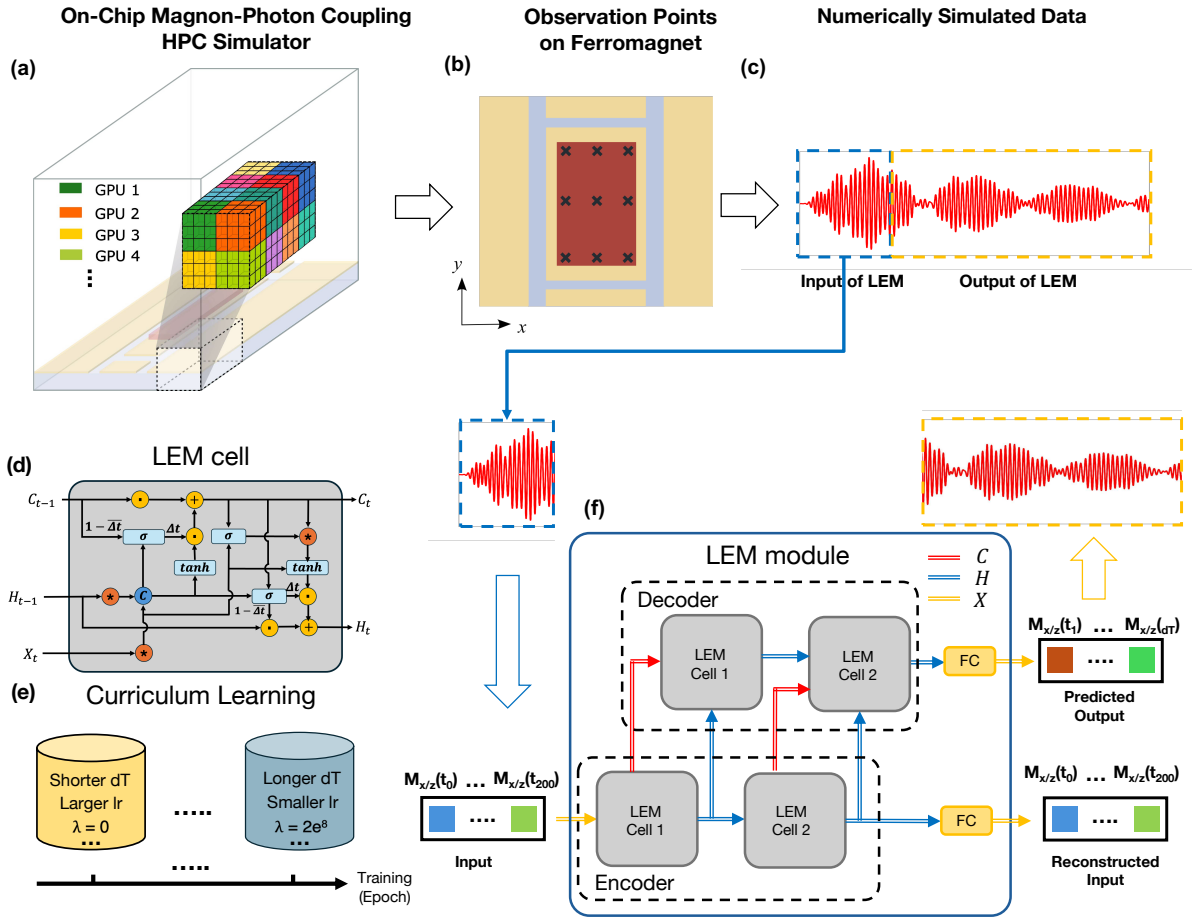


Figure 2: The HPC and ML frameworks for magnon-photon hybrid quantum modeling. (a) Simulation domain setup and GPU partitioning in the Maxwell–LLG coupled numerical model, ARTEMIS. The GPU parallelization leverages the grid-structured GPU library AMReX [56], achieving nearly flat weak-scaling performance. (b) Top view of the magnon-photon hybrid quantum circuit; the cross marks indicate probing points used for data collection. (c) Example time-series data of the magnetization **M** recorded at a single probing point. (d) Detailed architecture of the LEM cell. The cell updates a latent state (C_t) and a hidden state (H_t) using the input X_t, previous states, and learnable time constants (Δt). C_t captures short-term dynamics, while H_t integrates transformed latent features to maintain temporal context. (e) The curriculum learning strategy used to train the ML model, with a staged schedule with progressively adjusted learning rates, batch sizes, sequence lengths, and physics-based loss weights to stabilize training and enhance convergence. (f) Overall ML framework for time-series prediction, built out of the LEM-based architecture. The encoder consists of sequential LEM Cells that process the input sequence (X) into two evolving internal states: the latent state (C); and the hidden state (H). These states are used to initialize an auto-regressive decoder, which also consists of stacked LEM Cells. At each step, the decoder outputs a hidden state (H) that is passed through a shared fully connected (FC) layer to produce the predicted output.

prediction of [23]:

$$\omega_{\pm} = \omega_p + \frac{\Delta}{2} \pm \frac{\sqrt{\Delta^2 + 4\, g_{\mathrm{mp}}^2}}{2}. \tag{2}$$

One unique advantage of numerical simulation is the ability to generate signals on demand, including those with intensities that are experimentally unattainable. Extended Data Figure 2 displays the electric-field spectra as the excitation intensity is increased. It demonstrates that, in the absence of significant spin-spin interactions (e.g., exchange coupling), progressively stronger microwave magnetic fields suppress the magnon dynamics and leave only the cavity photon mode. Such behaviour arises because, at high powers, nonlinear spin-wave interactions—particularly three-magnon splitting—induce a Suhl instability that transfers energy from the uniform ferromagnetic resonance to non-uniform spin waves; this saturation of the Kittel mode causes the magnon–photon anticrossing gap to close and yields a single photon-like resonance, a phenomenon sometimes described as the high-power cavity magnon-polaritons regime [60]. For more theoretical analysis and validation under both linear and nonlinear regimes, refer to the **Supplementary Materials**. These results collectively demonstrate the model's capability in predicting the intricate hybrid quantum coupling, especially in complex, realistic integrated chip setup.

ML Surrogate Accuracy and Generalization. We evaluate the performance of our proposed ML model for modeling long-horizon dynamical fields under two scenarios: predictions at a single spatial location (Figure 4); and multiple probing locations (Figure 5). The model is trained using limited input data and physics constraint represented by a specific physical loss term (refer to Physics Loss in Methods for details). A representative example is shown in Figure 4(a), where the ML model is given only 200 temporal samples (0.6 ns) as input and predicts the long subsequent 1480 temporal samples (4.1 ns) under a DC magnetic bias of 2050 Oersted. The frequency spectrum in Figure 4(b) and Figure 4(d) shows that the ML surrogate reproduces the hybridized polaritons' characteristics, confirming it has learned the underlying physics rather than simply fitting short-term temporal trends. An ablation test (Fig. 4(c)) shows physics-informed training achieves lower long-horizon error than purely data-driven models, while requiring fewer samples and training steps. Using the ESPRIT algorithm [61], we extracted the two polariton frequencies and their Q-factors directly from the time traces in Figure 4(a); the results are plotted in panels (e) and (f). Frequencies match almost exactly, while small Q-factor offsets persist—typical for high-Q systems where tiny signal errors and hyper-parameter choices strongly affect the fitted decay rate. Even so, the physics-informed model reproduces both resonant modes with high fidelity, confirming its accuracy for precision-sensitive metrics.

To further test spatial generalization, we move from a single-spatial-probing model to a multi-spatial-probing network with a larger hidden layer. Training uses data from nine monitored points; held-out temporal samples from these same spatial points serve as in-distribution (ID) tests, while signals from 16 previously unseen spatial points provide out-of-distribution (OOD) tests. The nine color maps on the left "Prediction" panel of Figure 5(a) show the M_x spectra predicted at ID probing points, and they are visually indistinguishable from the numerical reference on the right "Truth" panel. The corresponding relative-error map in Figure 5(b) confirms that discrepancies stay negligible, particularly in the polariton resonance bands. We also applied the network

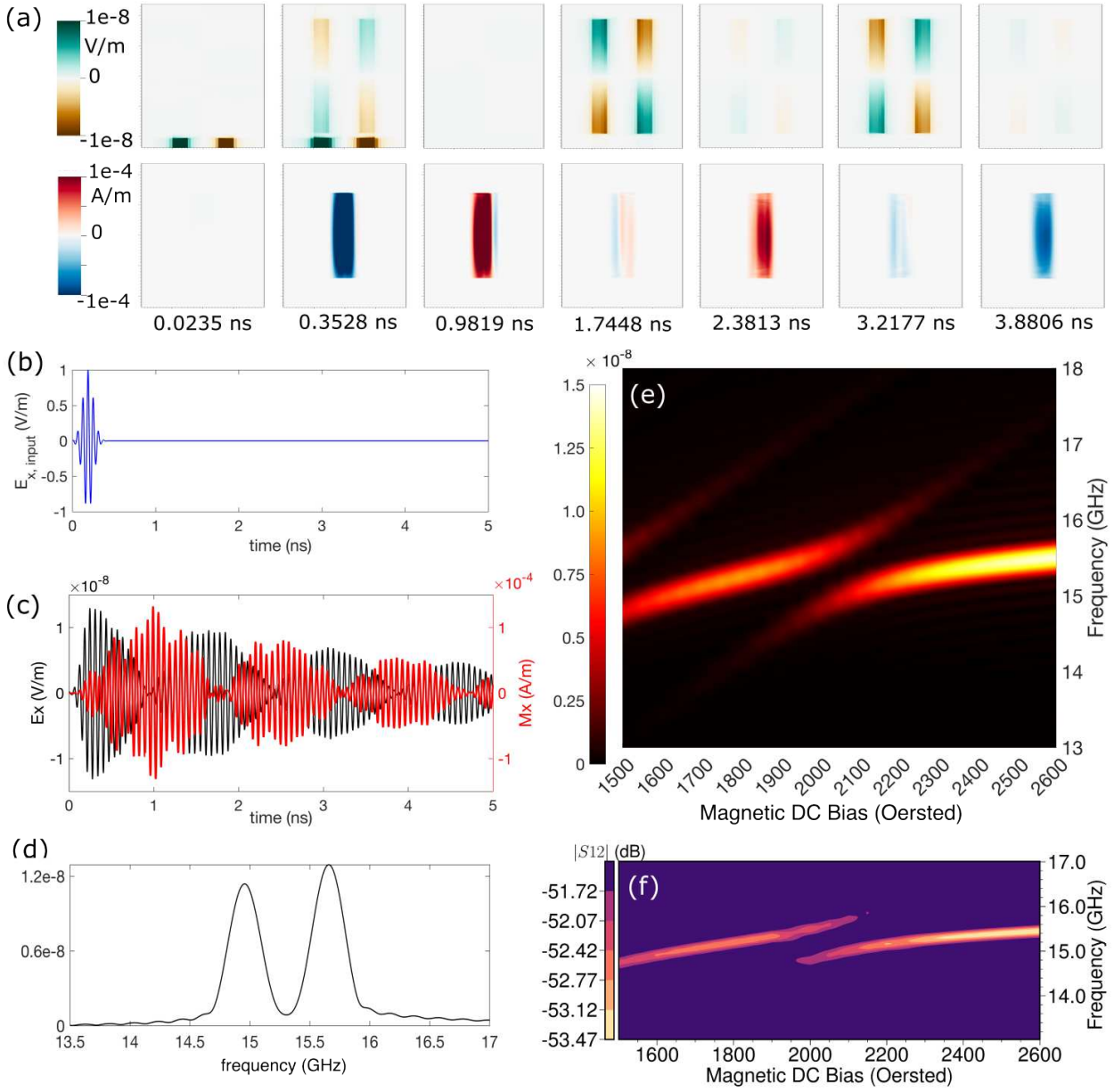


Figure 3: Numerical modeling results (a) Snapshots showing top-view distributions of E_x (top row) and M_x (bottom row) at selected time instants with an applied magnetic DC bias $H_0 = 2050$ Oersted. Here, the E_x fields are recorded at the same locations as Figure 1(e), i.e., the vertical center plane passing through the lateral air gap of the CPW resonator, while the M_x fields are recorded at the same locations as Figure 1(f), i.e., the top surface of the ferromagnetic stripe. See Figure 1(a) for details of the observation locations. Each snapshot helps illustrate how the electric field and magnetization evolve over time in their respective regions. (b) Input electric field in the form of modulated Gaussian pulse. The position of the input port is depicted in Figure 1(a). (c) Time sequences of E_x and M_x at specific spatial points. E_x is recorded at the CPW resonator's edge within its lateral air gaps (Figure 1), while M_x is recorded at the longitudinal center of the ferromagnet stripe on the top surface (Figure 1). (d) Spectral E_x field, obtained by Fourier-transforming the time sequence in sub-figure (c). (e) Magnitude spectra of E_x under varying magnetic DC bias H_0. (f) Magnitude spectra of $|S_{12}|$ parameter under the same sweeping magnetic DC bias H_0.

to 16 probing points that were withheld from training (OOD), and generated similar spectra as the ID data. Figure 5(c) plots the mean relative error for the first ("Peak 1") and second ("Peak 2") polariton modes; blue dotted lines mark the nine ID positions. Across all 25 probe points the physics-informed network keeps the resonance-frequency error below 2.3%, and the OOD sites perform on par with the ID ones. The uniformly low errors confirm true spatial generalization, enabled by the physics-based loss terms. Consequently, the model predicts field patterns and resonance spectra at unseen locations without any additional retraining.

3 Discussion

GPU Scalability. Figure 6 shows the weak-scaling results for our ARTEMIS multi-core/GPU solver, where each GPU advances a fixed mesh size. Performance is essentially ideal from 512 to 2048 GPUs, far beyond the scaling range of existing commercial or research codes that tackle Maxwell-LLG coupling, making full-chip modeling of hybrid quantum circuits and interconnects practical. For comparison, the same figure includes the cost of the physics-informed ML surrogate when inferring a single probe point on 1, 4, and 8 GPUs. The run time is normalized as $t_{\text{inf}} = TN_p/(N_tB)$, where T is the total wall-clock time, N_t the number of predicted steps, B the batch size, and N_p the number of spatial points, allowing direct comparison with the per-step cost of the numerical solver. Because the surrogate operates directly on time-series data and needs no full-3D mesh, its per-step cost is orders of magnitude lower — an advantage when only a handful of spatial points are of interest, as in most experimental measurements.

Future Directions. We demonstrated hybrid quantum simulation using a magnon–photon testbed, yet the framework is readily extensible to other hybrid magnonic, cavity electrodynamic, and phononic systems [9, 62–64]. Ongoing work therefore targets a generalized engine for quantum communication, storage, and computation that natively supports these additional interactions. Our current ML surrogate treats each mesh point as an independent one-dimensional time series, limiting its view of the device's collective field behavior. We therefore aim to develop spatiotemporal models that learn the coupled evolution of two- and three-dimensional fields, preserve spatial coherence, and predict the full device response with high fidelity.

4 Methods

Numerical Model. In the entire simulation domain, we solve the full-form of the dynamic Maxwell's equations, and Ampere and Faraday's laws, given as:

$$\nabla \times \mathbf{H} = \epsilon_0 \epsilon_r \frac{\partial \mathbf{E}}{\partial t} + \sigma \mathbf{E}, \tag{3}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{4}$$

where $\mathbf{E}$ is the electric field, $\mathbf{B}$ the magnetic flux density, $\mathbf{H}$ the magnetic-field intensity, and $\mathbf{J}_{\text{src}}$ an impressed current source. Material parameters are the conductivity σ, vacuum permittivity ϵ_0, relative permittivity ϵ_r, and vacuum permeability μ_0. The constitutive law linking $\mathbf{H}$ and $\mathbf{B}$ is $\mathbf{B} = \mu\mathbf{H} = \mu_0(\mathbf{H} + \mathbf{M})$, with $\mathbf{M}$ the magnetization density and $\mu = \mu_0\mu_r$ the permeability. This makes Equation 4 into $\nabla \times \mathbf{E} = -\partial\mathbf{H}/\partial t - \partial\mathbf{M}/\partial t$. For non-magnetic media, $\mu_r \approx 1$

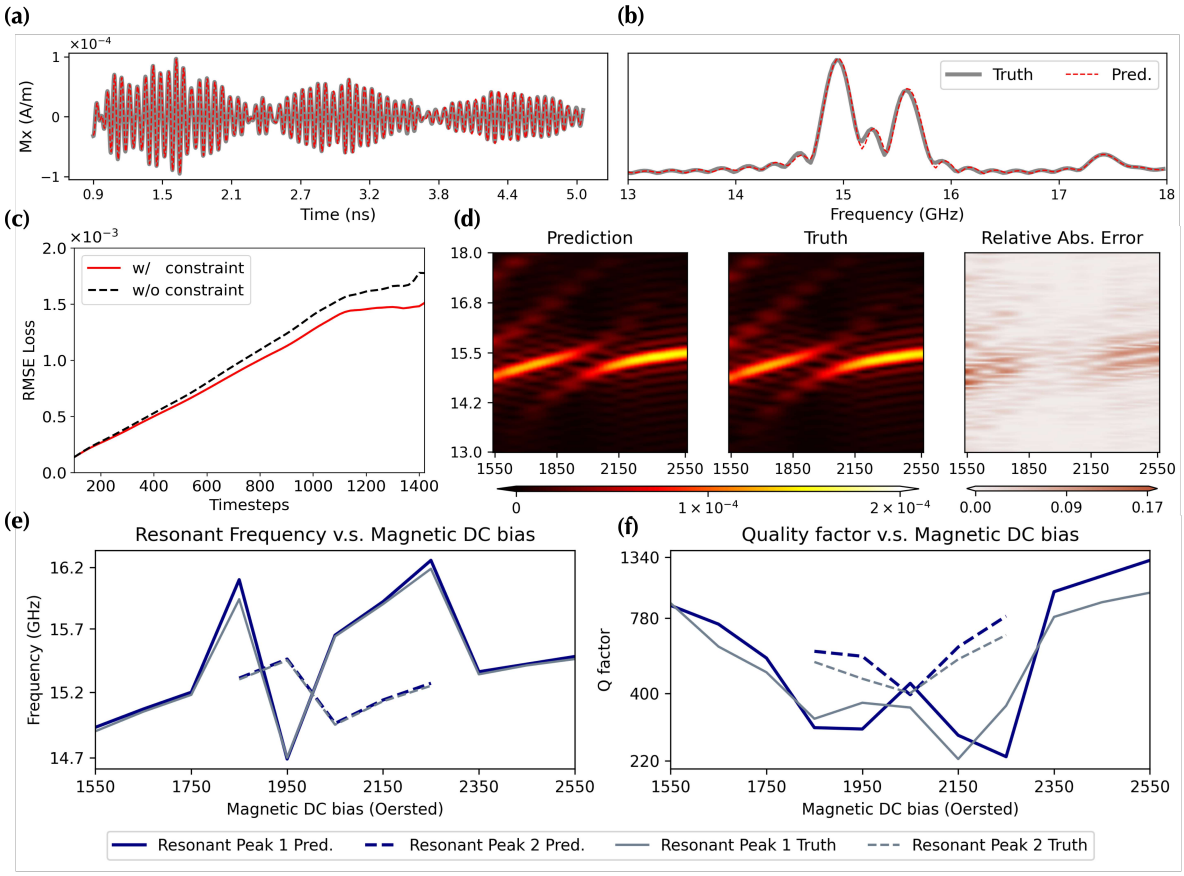


Figure 4: Physics-informed ML time sequence prediction at one single spatial probing point. (a) Time-series of the probed M_x field: ML prediction versus numerical ground truth. (b) Fourier spectra of the same traces as (a). (c) The root mean squared error (RMSE) as a function of forcast length, with and without the physical loss. (d) Magnitude spectra of M_x under varying magnetic DC bias H_0. The x-axis is the bias magnetic field in Oersted, and the y-axis is the frequency in GHz; color indicates field amplitude. (e) Hybrid-mode eigen frequecies extracted with ESPRIT [61]-calculated eigen frequencies of the hybrid modes. (f) Corresponding ESPRIT-calculated Q factors of the hybrid modes.

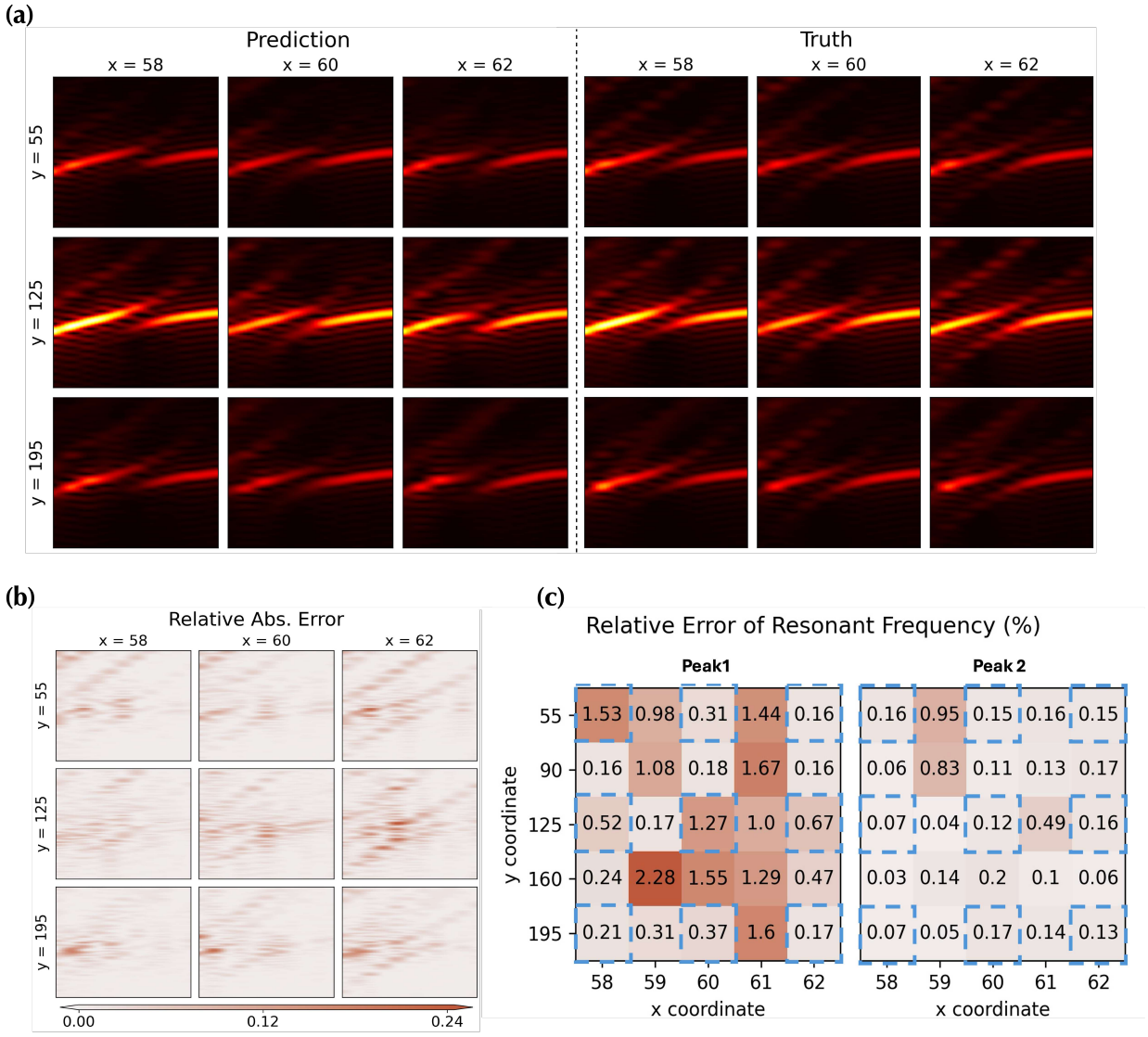


Figure 5: Physics-informed ML predictions at multiple probe points. (a) Magnitude spectra of the M_x field at the nine in-distribution (ID) locations: model output (left) and numerical reference (right). (b) Relative-error map between the spectra in (a). (c) Relative error of the two resonance frequencies extracted at each of the 25 probe points; blue dotted markers denote the nine ID points, the remaining points are out-of-distribution (OOD).

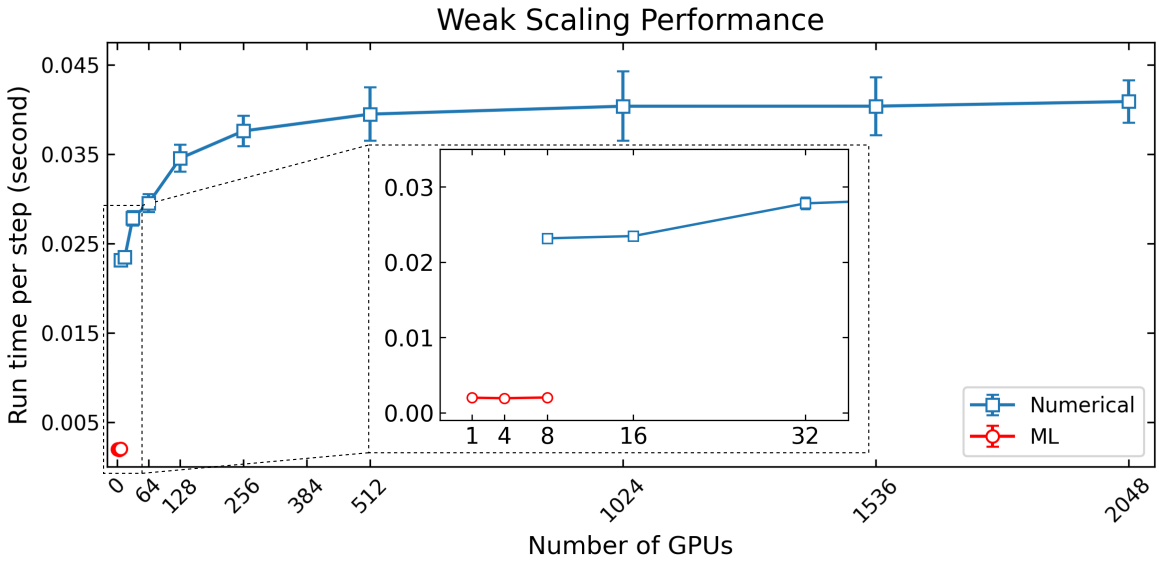


Figure 6: Weak scaling for simulations on the NERSC Perlmutter GPU (MPI+CUDA) system. Note that one Perlmutter node contains 4 GPUs. The weak scaling is nearly perfect for the Perlmutter simulations past the 16-node threshold up to 2048 GPUs (512 nodes). Details of numerical accuracy and performance tests can be found in the Code Scaling Performance section in Methods.

since $\mathbf{M} \approx 0$. In ferromagnets, $\mathbf{M}$ must be evolved (e.g., via the Landau–Lifshitz–Gilbert (LLG) equation) to capture magnon dynamics. The LLG equation is expressed as:

$$\frac{\partial \mathbf{M}}{\partial t} = \mu_0 \gamma (\mathbf{M} \times \mathbf{H}_{\text{eff}}) + \frac{\alpha}{|\mathbf{M}|}(\mathbf{M} \times \frac{\partial \mathbf{M}}{\partial t}), \tag{5}$$

where $\gamma = -1.759 \times 10^{11}$ C/kg is the gyromagnetic ratio and α is the dimensionless Gilbert damping constant (10^{-5} to 10^{-3} in typical ferromagnets). The effective field $\mathbf{H}_{\text{eff}}$ collects all magnetic interactions. Here we retain two terms essential for magnon-photon coupling: the time-dependent EM field $\mathbf{H}_{\text{EM}}$, and the static bias $\mathbf{H}_0$. Because the ferromagnet is driven close to saturation, the magnetization magnitude is taken constant, $|\mathbf{M}| = M_s$, where M_s is the (spatially dependent) saturation magnetization. We employ an iterative scheme that uses a trapezoidal discretization of the vector form of the LLG equation (5) described in [29], to solve for $\mathbf{M}^{n+1,r}$, where n denotes the time step and r is the counter index for the iterative procedure:

$$\mathbf{M}^{n+1,r} = \frac{\boldsymbol{b} + (\boldsymbol{a} \cdot \boldsymbol{b})\boldsymbol{a} - \boldsymbol{a} \times \boldsymbol{b}}{1 + |\boldsymbol{a}|^2}, \tag{6}$$

where

$$\boldsymbol{a} = -\left[\frac{\mu_0 |\gamma| \Delta t}{4} \mathbf{H}_{\text{eff}}^{n+1,r-1} + \frac{\alpha}{M_s} \mathbf{M}^n\right], \tag{7}$$

$$\boldsymbol{b} = \mathbf{M}^n - \frac{\mu_0 |\gamma| \Delta t}{4} (\mathbf{M}^n \times \mathbf{H}_{\text{eff}}^n). \tag{8}$$

We normalize $\mathbf{M}^{n+1,r} = \mathcal{N}(\mathbf{M}^{n+1,r})$ and then compute $\mathbf{H}^{n+1,r}$ by integrating Faraday's law following the standard FDTD procedure, with an updated approximation for the time-derivative of $\mathbf{M}$,

$$\mathbf{H}^{n+1,r} = \mathbf{H}^{n} + \mathbf{M}^{n} - \mathbf{M}^{n+1,r} - \frac{\Delta t}{2\mu_0}\nabla \times \mathbf{E}^{n}. \tag{9}$$

We continue to iterate over r using equations (6) and (9), until the maximum value of $(\mathbf{M}^{n+1,r} - \mathbf{M}^{n+1,r-1})/M_s$ over the entire domain is smaller than a user-defined tolerance that we choose to be 10^{-6}. After the iterations converge, we set $(\mathbf{H}, \mathbf{M})^{n+1} = (\mathbf{H}, \mathbf{M})^{n+1,r}$ and proceed to computing $\mathbf{E}^{n+1}$ by discretizing Ampere's law (3) using time-centered representations of the magnetic field and current,

$$\mathbf{E}^{n+1} = \left(\frac{\sigma}{2} + \frac{\epsilon}{\Delta t}\right)^{-1} \left[\nabla \times \mathbf{H}^{n+1} - \left(\frac{\sigma}{2} - \frac{\epsilon}{\Delta t}\right)\mathbf{E}^{n} - \mathbf{J}_{\text{src}}^{n+1}\right]. \tag{10}$$

Note that in regions where $M_s = 0$, we set $\mathbf{M} = 0$, skip this iterative scheme, and instead integrate $\mathbf{H}$ using equation (9), with μ_0.

Numerical Solver Implementation. We build our solver on AMReX [56], the DOE Exascale Computing Project's block-structured adaptive mesh refinement (AMR) library. AMReX provides structured-grid PDE infrastructure with performance portability: the domain is decomposed into non-overlapping boxes ("grids") that are distributed across MPI ranks, and execution follows an MPI+X model: X=OpenMP on multicore CPUs; or X=CUDA on NVIDIA GPUs. A single front-end code path uses AMReX loop abstractions, which select the appropriate back end at compile time. Each rank updates only the data it owns using triply nested cell loops. In the pure-MPI case, these are standard i-j-k loops; with MPI+OpenMP the loops are further tiled and assigned to threads; with MPI+CUDA AMReX launches device kernels that map cells to CUDA threads. AMReX manages data residency so arrays remain on the GPU whenever possible, minimizing host–device traffic; inter-rank communication and host/GPU transfers are therefore limited to ghost-cell exchanges performed a small number of times per time step. The efficiency and scalability of our code using MPI+CUDA simulations on NERSC Perlmutter systems is shown in Figure 6. Data from the simulation can be efficiently written using a user-defined number of MPI ranks, to prevent overwhelming the system with simultaneous writes. Visualization was performed with the publicly-available software package, VisIt [65].

Datasets Curation. The dataset for our ML model was generated using the ARTEMIS simulator introduced in sections Numerical Solver Implementation and Numerical Model. We monitored the x and z component of magnetization (M_x and M_z) at discrete spatial points across a defined domain, creating a dataset with sparse spatial sampling. The x-indices range from 58 to 62 with index interval of 1, corresponding to the x location of 290 μm to 310 μm. The y-indices spanned 55 to 195 with a sampling interval of 35, corresponding to 1045 μm and 3705 μm. The z-coordinate was fixed at 505 μm (and thus the z-index is also 505), corresponding to the center of the NiFe magnetic strip along the z-axis, resulting in a total of 25 spatial sampling points. Simulations were conducted under 20 distinct external DC magnetic bias conditions, ranging from 1550 Oersted to 2550 Oersted, sampled uniformly at intervals of 50 Oersted. For the single-location scenario, we used the temporal field data M_x and M_z from a specific spatial location $(60, 125, 505)$, located at the center of the NiFe magnetic material strip. Four trajectories corresponding to external

magnetic biases of 2050, 2250, 2350, and 2550 Oersted were reserved for testing, while the remaining trajectories were used for training. For multiple-location secario, the spatial locations include 9 monitoring points: $[x, y, 505]$, where $x \in \{58, 60, 62\}$ μm and $y \in \{55, 125, 195\}$ μm. For each spatial point, the training trajectories are chosen from the magnetic field H_0 across the range $[1500, 1600, \ldots, 2600]$ Oersted, using a step size of 100. Each trajectory is preprocessed by truncating the first 160 steps, which is the ramping-up stage of the input signal, and downsampling by a factor of 1000. The data is normalized using min-max normalization.

LEM Cell. We use the Long Expressive Memory (LEM) model [57] as a robust approach for modeling complex temporal dynamics. LEM is specifically designed to capture multi-scale and long-term behaviors in systems characterized by intricate temporal structures. Its architecture offers a mathematical foundation for handling sequences with varying temporal scales, while ensuring gradient stability and high expressivity. EM fields, which exhibit a combination of rapid oscillations (e.g., high-frequency wave propagation) and slower dynamics (e.g., energy dissipation), present a challenging time-series modeling task. The LEM framework is well-suited for our scientific task due to the capability to effectively learn and predict diverse and interdependent behaviors within EM dynamics. The LEM model is rooted in multi-scale Ordinary Differential Equations (ODEs) that can be used to capture fast and slow dynamics. Let us consider a general formulation of a two-scale ODE system [57], given as

$$\begin{aligned} \frac{d\mathbf{y}}{dt} &= \tau_y \left(\sigma(\mathbf{W}_y \mathbf{z} + \mathbf{V}_y \mathbf{u} + \mathbf{b}_y) - \mathbf{y}\right), \\ \frac{d\mathbf{z}}{dt} &= \tau_z \left(\sigma(\mathbf{W}_z \mathbf{y} + \mathbf{V}_z \mathbf{u} + \mathbf{b}_z) - \mathbf{z}\right). \end{aligned} \tag{11}$$

Here, $\mathbf{y}(t) \in \mathbb{R}^d$ and $\mathbf{z}(t) \in \mathbb{R}^d$ represent the slow and fast variables, $\mathbf{u}(t) \in \mathbb{R}^m$ is the input signal, and $\{\tau_y, \tau_z\}$ are time-scale parameters ($0 < \tau_y \leq \tau_z \leq 1$). We define $\{\mathbf{W}_y, \mathbf{W}_z, \mathbf{V}_y, \mathbf{V}_z\}$ as weight matrices and $\{\mathbf{b}_y, \mathbf{b}_z\}$ as bias vectors. $\sigma(\cdot)$ denotes the nonlinear activation function (e.g., hyperbolic tangent function). This formulation can be further generalized to represent multiple temporal scales by incorporating additional gating functions. These gating mechanisms, based on adaptive weights $\{\mathbf{W}_1, \mathbf{W}_2, \mathbf{V}_1, \mathbf{V}_2\}$ and biases $\{\mathbf{b}_1, \mathbf{b}_2\}$, allow the model to dynamically learn and modulate across various time scales. Therefore, the LEM model is capable of capturing more intricate dependencies and temporal patterns. To ensure practical applicability, LEM discretizes the continuous multi-scale ODE system using an implicit-explicit (IMEX) time-stepping scheme, which can be formulated as

$$\begin{aligned} \Delta\mathbf{t}_n &= \Delta t \, \hat{\sigma}(\mathbf{W}_1 \mathbf{y}_{n-1} + \mathbf{V}_1 \mathbf{u}_n + \mathbf{b}_1), \\ \overline{\Delta\mathbf{t}}_n &= \Delta t \, \hat{\sigma}(\mathbf{W}_2 \mathbf{y}_{n-1} + \mathbf{V}_2 \mathbf{u}_n + \mathbf{b}_2), \\ \mathbf{z}_n &= (1 - \Delta\mathbf{t}_n) \odot \mathbf{z}_{n-1} + \Delta\mathbf{t}_n \odot \sigma(\mathbf{W}_z \mathbf{y}_{n-1} + \mathbf{V}_z \mathbf{u}_n + \mathbf{b}_z), \\ \mathbf{y}_n &= (1 - \overline{\Delta\mathbf{t}}_n) \odot \mathbf{y}_{n-1} + \overline{\Delta\mathbf{t}}_n \odot \sigma(\mathbf{W}_y \mathbf{z}_n + \mathbf{V}_y \mathbf{u}_n + \mathbf{b}_y). \end{aligned} \tag{12}$$

Here, $\odot$ denotes element-wise multiplication, $\hat{\sigma}(\cdot)$ is the sigmoid activation function, $\Delta t > 0$ is a tunable parameter, and n represents the discrete time step within $[1, N]$. $\{\mathbf{y}_n, \mathbf{z}_n\} \in \mathbb{R}^d$ and $\mathbf{u}_n \in \mathbb{R}^m$ denote two hidden states and input variable, respectively. The final output state $\mathbf{o}_n \in \mathbb{R}^o$ is augmented with a linear operation $\mathcal{W}_y \in \mathbb{R}^{o \times d}$, where $\mathbf{o}_n = \mathcal{W}_y \mathbf{y}_n$.

ML Training Implementation. We implement curriculum training to enhance model performance. This approach is widely adopted in scientific ML since it allows the ML model to gradually adapt to increasing levels of complexity and facilitates the effective learning of long-term temporal dynamics. For the single probing point training, the model is trained for a total of 1250 epochs with the hidden unit number of the LEM cell set to 64. Throughout the training, the learning rate is gradually decreased from 0.001 to 1.25×10^{-4}, the predicted sequence length is adjusted from 50 to 400, the batch size increases from 64 to 512, and the regularization parameter for the physics loss is progressively raised from 0 to 2.5×10^{6}. Additionally, the training sequence size alternates between 100 and 200 at different stages to optimize learning dynamics. The training protocol is generally set the same for the multiple probing point training, except that the predicted sequence length is adjusted from 50 to 500.

Physics Loss. Physics-informed ML has been shown to effectively preserve physical properties in various domains [39]. Physics-informed loss function is adopted into the LEM framework to further improve the model's ability to learn the underlying physics. We adopted the small-signal approximation (the dynamic magnetization in the bias direction is negligible, i.e. $M_y \approx 0$) in the LLG equation when deriving the physics loss term, which is

$$\mathcal{L}_{\text{phy}} = \|\text{Residual}_{M_x}\|^2 + \|\text{Residual}_{M_z}\|^2, \tag{13}$$

where

$$\text{Residual}_{M_x} = M_x^{n+1}(\text{LLG}) - M_x^{n+1}(\text{ML}), \tag{14}$$

$$\text{Residual}_{M_z} = M_z^{n+1}(\text{LLG}) - M_z^{n+1}(\text{ML}). \tag{15}$$

The analytical derivatives $M_x(\text{LLG})$ and $M_z(\text{LLG})$ are the x- and z-components derived from the trapezoidal discretization of the LLG equation, as given in Eqs. (6)–(8). Given a sequence of network predictions for $n = 0, \ldots, T$, these residuals are evaluated for $n = 0, \ldots, T-1$.

From the ML side, we estimate the time derivatives from the predicted sequence using a finite forward–Euler discretization:

$$\frac{\partial M_x^n}{\partial t}(\text{ML}) = \frac{M_x^{n+1} - M_x^n}{\Delta t_{\text{ML}}}, \tag{16}$$

$$\frac{\partial M_z^n}{\partial t}(\text{ML}) = \frac{M_z^{n+1} - M_z^n}{\Delta t_{\text{ML}}}. \tag{17}$$

Given these derivatives, the ML one-step *next values* are

$$M_x^{n+1}(\text{ML}) = M_x^n + \Delta t_{\text{ML}} \frac{\partial M_x^n}{\partial t}, \qquad M_z^{n+1}(\text{ML}) = M_z^n + \Delta t_{\text{ML}} \frac{\partial M_z^n}{\partial t}. \tag{18}$$

Here Δt_{ML} denotes the sampling interval of the training data, which may differ from the Δt utilized in the numerical solver, the magnetization values $M_{x,z}^n$ and $M_{x,z}^{n+1}$ are the neural network predictions at two consecutive timesteps.

The total training objective is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{data}} + \lambda \, \mathcal{L}_{\text{phy}}.$$

The loss term $\mathcal{L}_{\text{phy}}$ penalizes deviations from the LLG dynamics, ensuring that the predicted magnetization components M_x and M_z closely follow the physical laws governing their evolution. By minimizing this loss during training, the model learns to approximate the time evolution of magnetization accurately, while respecting the underlying physics.

Code Scaling Performance. We assess the performance and scalability of our code on Perlmutter, the NERSC flagship supercomputer. Perlmutter comprises 1,792 GPU-accelerated nodes, each with 4 NVIDIA A100 (40 or 80 GB) GPUs. We perform weak-scaling tests using an MPI+CUDA strategy with one MPI rank per GPU. To ensure proper load balancing, the weak-scaling cases use a continuous CPW line and ferromagnetic strip along y, so no cavity resonances form; instead, propagating waves traverse the CPW. This yields uniform compute and memory load per GPU. Our baseline domain is $N_x \times N_y \times N_z = 120 \times 250 \times 250$ (7.5M cells) on 8 GPUs. For weak scaling, we increase N_y proportionally to the GPU count to keep the cells per GPU fixed; e.g., 32 GPUs for $120 \times 1000 \times 250$ (30M cells), matching the main-manuscript problem size. Weak-scaling results (average simulation time per time step vs. number of nodes) are shown in Figure 6. Initialization time is excluded (one-time cost). We tested up to 512 nodes (2,048 GPUs) and observe near-ideal scaling from 128 to 512 nodes. This behavior reflects that with a fixed per-GPU subdomain, both computation and halo-exchange volumes per GPU remain constant; thus inter-node communication overhead stays small relative to compute for the problem sizes considered. Details of the machine-learning surrogate scaling are provided in Section 3.

5 Data availability

The datasets generated during and/or analysed during the current study will be available in the zenodo repository upon publication.

6 Code availability

All codes are publicly available on the open-source github platform. The AMReX library can be obtained at AMReX, hash `2ee8ae270` and this code can be obtained at ARTEMIS, hash `151b5a32f`. The ML code will be available online upon publication.

7 Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator (Award No. DE-SCL0000121). Additional support is acknowledged from the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research Program, Microelectronics Science Research Center, under contract No. DE-AC02-05-CH11231. This work was also supported in part by previous breakthroughs obtained through the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231. The work at the University of Wisconsin-Madison was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award Number DE-SC0020145 as part of the Computational Materials Sciences Program (Y.Z. and J.-M.H.). This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-

AC02-05CH11231.

8 Author contributions

Z.Y. conceived the study and supervised the research. Z.Y. and Y.T. led the manuscript writing. Z.Y. and A.N. developed, configured, and executed the numerical model, generated the datasets, and performed post-processing. B.E. designed and implemented the machine learning model. J.S. and P.R. carried out model training and inference. J.S. organized and uploaded all data and codes to Zenodo. S.S. performed S-parameter extraction and analysis. S.T. performed the Floquet nonlinear analysis. Y.T. helped with the training and inference. Y.Z. and J.-M.H. performed the 1D theoretical analysis and the supplementary material writing. All authors discussed the results, contributed to writing and revising the manuscript, and approved the final version of the paper.

9 Competing interests

The authors declare no competing interests.

10 References

1. Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

2. Gershon Kurizki, Patrice Bertet, Yuimaru Kubo, Klaus Mølmer, David Petrosyan, Peter Rabl, and Jörg Schmiedmayer. Quantum technologies with hybrid systems. *Proceedings of the National Academy of Sciences*, 112(13):3866–3873, 2015.

3. P Scarlino, D J van Woerkom, U C Mendes, J V Koski, A J Landig, C K Andersen, S Gasparinetti, C Reichl, W Wegscheider, K Ensslin, T Ihn, A Blais, and A Wallraff. Coherent microwave-photon-mediated coupling between a semiconductor and a superconducting qubit. *Nature Communications*, 10(1):3011, 2019.

4. Yutaka Tabuchi, Saiko Ishino, Atsushi Noguchi, Toyofumi Ishikawa, Ryo Yamazaki, Koji Usami, and Yasunobu Nakamura. Coherent coupling between a ferromagnetic magnon and a superconducting qubit. *Science*, 349(6246):405–408, 2015.

5. A A Clerk, K W Lehnert, P Bertet, J R Petta, and Y Nakamura. Hybrid quantum systems with circuit quantum electrodynamics. *Nature Physics*, 16(3):257–267, 2020.

6. J. J. Viennot, M. C. Dartiailh, A. Cottet, and T. Kontos. Coherent coupling of a single spin to microwave cavity photons. *Science*, 349(6246):408–411, 2015.

7. Justin T Hou and Luqiao Liu. Strong Coupling between Microwave Photons and Nanomagnet Magnons. *Physical Review Letters*, 123(10):107702, sep 2019.

8. Yi Li, Tomas Polakovic, Yong-Lei Wang, Jing Xu, Sergi Lendinez, Zhizhi Zhang, Junjia Ding, Trupti Khaire, Hilal Saglam, Ralu Divan, John Pearson, Wai-Kwong Kwok, Zhili Xiao, Valentine Novosad, Axel Hoffmann, and Wei Zhang. Strong Coupling between Magnons and Mi-

crowave Photons in On-Chip Ferromagnet-Superconductor Thin-Film Devices. *Physical Review Letters*, 123(10):107701, sep 2019.

9. Jilei Chen, Chuanpu Liu, Tao Liu, Yang Xiao, Ke Xia, Gerrit E W Bauer, Mingzhong Wu, and Haiming Yu. Strong interlayer magnon-magnon coupling in magnetic metal-insulator hybrid nanostructures. *Physical review letters*, 120(21):217202, 2018.

10. Aaron D O'Connell, Max Hofheinz, Markus Ansmann, Radoslaw C Bialczak, Mike Lenander, Erik Lucero, Matthew Neeley, Daniel Sank, H Wang, Martin Weides, et al. Quantum ground state and single-phonon control of a mechanical resonator. *Nature*, 464(7289):697–703, 2010.

11. RW Andrews, AP Reed, K Cicak, JD Teufel, and KW Lehnert. Quantum-enabled temporal and spectral mode conversion of microwave signals. *Nature communications*, 6(1):10021, 2015.

12. Riccardo Manenti, Anton F Kockum, Andrew Patterson, Tanja Behrle, Joseph Rahamim, Giovanna Tancredi, Franco Nori, and Peter J Leek. Circuit quantum acoustodynamics with surface acoustic waves. *Nature communications*, 8(1):975, 2017.

13. Xufeng Zhang, Chang-Ling Zou, Liang Jiang, and Hong X Tang. Strongly Coupled Magnons and Cavity Microwave Photons. *Physical Review Letters*, 113(15):156401, oct 2014.

14. Babak Zare Rameshti, Silvia Viola Kusminskiy, James A Haigh, Koji Usami, Dany Lachance-Quirion, Yasunobu Nakamura, Can-Ming Hu, Hong X Tang, Gerrit E W Bauer, and Yaroslav M Blanter. Cavity magnonics. *Physics Reports*, 979:1–61, 2022.

15. D D Awschalom, C R Du, R He, F J Heremans, A Hoffmann, J Hou, H Kurebayashi, Y Li, L Liu, V Novosad, J Sklenar, S E Sullivan, D Sun, H Tang, V Tyberkevych, C Trevillian, A W Tsen, L R Weiss, W Zhang, X Zhang, L Zhao, and C W Zollitsch. Quantum Engineering With Hybrid Magnonic Systems and Materials (Invited Paper). *IEEE Transactions on Quantum Engineering*, 2:1–36, 2021.

16. Christophe Match, Michael Harder, Lihui Bai, Paul Hyde, and Can-Ming Hu. Transient response of the cavity-magnon-polariton. *Physical Review B*, 99(13):134445, 2019.

17. Shihao Zhuang, Yujie Zhu, Changchun Zhong, Liang Jiang, Xufeng Zhang, and Jia-Mian Hu. Dynamical phase-field model of cavity electromagnonic systems. *npj Computational Materials*, 10(1):191, 2024.

18. Xufeng Zhang, Chang-Ling Zou, Na Zhu, Florian Marquardt, Liang Jiang, and Hong X. Tang. Magnon dark modes and gradient memory. *Nature Communications*, 6:8914, 2015.

19. Amin Pishehvar, Zhaoyou Wang, Yujie Zhu, Yu Jiang, Zixin Yan, Fangxin Li, Josep M. Jornet, Jia-Mian Hu, Liang Jiang, and Xufeng Zhang. On-demand magnon resonance isolation in cavity magnonics. *arXiv preprint arXiv:2412.15600*, 2024.

20. Jing Xu, Changchun Zhong, Xu Han, Dong Jin, Liang Jiang, and Xufeng Zhang. Floquet cavity electromagnonics. *Physical Review Letters*, 125(23):237201, 2020.

21. Ö. O. Soykal and M. E. Flatté. Strong field interactions between a nanomagnet and a photonic cavity. *Phys. Rev. Lett.*, 104:077202, Feb 2010.

22. Maxim Goryachev, Warrick G. Farr, Daniel L. Creedon, Yaohui Fan, Mikhail Kostylev, and Michael E. Tobar. High-cooperativity cavity qed with magnons at microwave frequencies. *Phys. Rev. Appl.*, 2:054002, Nov 2014.

23. Hans Huebl, Christoph W. Zollitsch, Johannes Lotze, Fredrik Hocke, Moritz Greifenstein, Achim Marx, Rudolf Gross, and Sebastian T. B. Goennenwein. High cooperativity in coupled microwave resonator ferrimagnetic insulator hybrids. *Phys. Rev. Lett.*, 111:127003, Sep 2013.

24. Isabella Boventer, Christine Dörflinger, Tim Wolz, Rair Macêdo, Romain Lebrun, Mathias Kläui, and Martin Weides. Control of the coupling strength and linewidth of a cavity magnon-polariton. *Physical Review Research*, 2(1):13154, feb 2020.

25. Igor A. Golovchanskiy, Nikolay N. Abramov, Vasily S. Stolyarov, Martin Weides, Valery V. Ryazanov, Alexander A. Golubov, Alexey V. Ustinov, and Mikhail Yu. Kupriyanov. Ultra-strong photon-to-magnon coupling in multilayered heterostructures involving superconducting coherence via ferromagnetic layers. *Science Advances*, 7(25), June 2021.

26. Michael Joseph Donahue and Donald G Porter. Oommf user's guide: Version 1.0, 1999.

27. Arne Vansteenkiste, Jonathan Leliaert, Mykola Dvornik, Felipe García-Sánchez, and Bartel Van Waeyenberge. The design and verification of mumax3. *AIP Advances*, 4(10):107133, 2014.

28. Rasmus Bjørk, Emil Blaabjerg Poulsen, Kaspar Kirstein Nielsen, and Andrea Roberto Insinga. Magtense: A micromagnetic framework using the analytical demagnetization tensor. *Journal of Magnetism and Magnetic Materials*, 535:168057, 2021.

29. Zhi Yao, Revathi Jambunathan, Yadong Zeng, and Andrew Nonaka. A massively parallel time-domain coupled electrodynamics–micromagnetics solver. *The International Journal of High Performance Computing Applications*, 36(2):167–181, 2022.

30. Zhi Yao, Sidhant Tiwari, Joseph Schneider, Robert N. Candler, Gregory P. Carman, and Yuanxun Ethan Wang. Enhanced planar antenna efficiency through magnetic thin-films. *IEEE Journal on Multiscale and Multiphysics Computational Techniques*, 6:249–258, 2021.

31. Zhi Yao, Rüştü Umut Tok, Tatsuo Itoh, and Yuanxun Ethan Wang. A multiscale unconditionally stable time-domain (must) solver unifying electrodynamics and micromagnetics. *IEEE Transactions on Microwave Theory and Techniques*, 66(6):2683–2696, 2018.

32. Jing Xu, Changchun Zhong, Shihao Zhuang, Chen Qian, Yu Jiang, Amin Pishehvar, Xu Han, Dafei Jin, Josep M Jornet, Bo Zhen, et al. Slow-wave hybrid magnonics. *Physical Review Letters*, 132(11):116701, 2024.

33. Amin Pishehvar, Zhaoyou Wang, Yujie Zhu, Yu Jiang, Zixin Yan, Fangxin Li, Josep M Jornet, Jia-Mian Hu, Liang Jiang, and Xufeng Zhang. On-demand magnon resonance isolation in cavity magnonics. *Physical Review Applied*, 23(2):024053, 2025.

34. Ivan S Maksymov, Jessica Hutomo, Donghee Nam, and Mikhail Kostylev. Rigorous numerical study of strong microwave photon-magnon coupling in all-dielectric magnetic multilayers. *Journal of Applied Physics*, 117(19), 2015.

35. Weichao Yu, Tao Yu, and Gerrit EW Bauer. Circulating cavity magnon polaritons. *Physical Review B*, 102(6):064416, 2020.

36. Sidi Fu, Weilong Cui, Matthew Hu, Ruinan Chang, Michael J Donahue, and Vitaliy Lomakin. Finite-difference micromagnetic solvers with the object-oriented micromagnetic framework on graphics processing units. *IEEE Transactions on Magnetics*, 52(4):1–9, 2015.

37. Steven L Brunton, Bernd R Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual review of fluid mechanics*, 52(1):477–508, 2020.

38. Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023.

39. George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

40. Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

41. Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

42. Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

43. J. A. L. Benitez, J. Guo, K. Hegazy, I. Dokmanic, M. W. Mahoney, and M. V. de Hoop. Neural equilibria for long-term prediction of nonlinear conservation laws. Technical Report Preprint: arXiv:2501.06933, 2025.

44. S Hochreiter. Long short-term memory. *Neural Computation MIT-Press*, 1997.

45. Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *ICLR*. OpenReview.net, 2022.

46. A. Yu, D. Lyu, S. H. Lim, M. W. Mahoney, and N. B. Erichson. Tuning frequency bias of state space models. Technical Report Preprint: arXiv:2410.02035, 2024.

47. A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

48. Yingheng Tang, Jichao Fan, Xinwei Li, Jianzhu Ma, Minghao Qi, Cunxi Yu, and Weilu Gao. Physics-informed recurrent neural network for time dynamics in optical resonances. *Nature Computational Science*, 2(3):169–178, March 2022.

49. Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden, Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall Robinson, Ellen Clancy, Alberto Arribas, and Shakir Mohamed. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878):672–677, September 2021.

50. Chengping Rao, Pu Ren, Qi Wang, Oral Buyukozturk, Hao Sun, and Yang Liu. Encoding physics to learn reaction–diffusion processes. *Nature Machine Intelligence*, 5(7):765–779, 2023.

51. Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619(7970):533–538, 2023.

52. Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, et al. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, 2023.

53. Jialin Song, Zezheng Song, Pu Ren, N Benjamin Erichson, Michael W Mahoney, and Xiaoye S Li. Forecasting high-dimensional spatio-temporal systems from sparse measurements. *Machine Learning: Science and Technology*, 5(4):045067, 2024.

54. Dongwei Lyu, Rie Nakata, Pu Ren, Michael W Mahoney, Arben Pitarka, Nori Nakata, and N Benjamin Erichson. Wavecastnet: An ai-enabled wavefield forecasting framework for earthquake early warning. *arXiv preprint arXiv:2405.20516*, 2024.

55. D. Hansen, D. C. Maddix, S. Alizadeh, G. Gupta, and M. W. Mahoney. Learning physical models that can respect conservation laws. Technical Report Preprint: arXiv:2302.11002, 2023.

56. Weiqun Zhang, Ann Almgren, Vince Beckner, John Bell, Johannes Blaschke, Cy Chan, Marcus Day, Brian Friesen, Kevin Gott, Daniel Graves, Max Katz, Andrew Myers, Tan Nguyen,

Andrew Nonaka, Michele Rosso, Samuel Williams, and Michael Zingale. AMReX: a framework for block-structured adaptive mesh refinement. *Journal of Open Source Software*, 4(37):1370, May 2019.

57. T. Konstantin Rusch, Siddhartha Mishra, N. Benjamin Erichson, and Michael W. Mahoney. Long expressive memory for sequence modeling. In *ICLR*. OpenReview.net, 2022.

58. Petru Soviany, Radu Tudor Ionescu, Paolo Rota, and Nicu Sebe. Curriculum learning: A survey, 2021.

59. Saurabh S. Sawant, Zhi Yao, Revathi Jambunathan, and Andrew Nonaka. Characterization of transmission lines in microelectronic circuits using the artemis solver. *IEEE Journal on Multiscale and Multiphysics Computational Techniques*, 8:31–39, 2023.

60. Oscar Lee, Kei Yamamoto, Maki Umeda, Christoph W. Zollitsch, Mehrdad Elyasi, Takashi Kikkawa, Eiji Saitoh, Gerrit E. W. Bauer, and Hidekazu Kurebayashi. Nonlinear magnon polaritons. *Phys. Rev. Lett.*, 130:046703, Jan 2023.

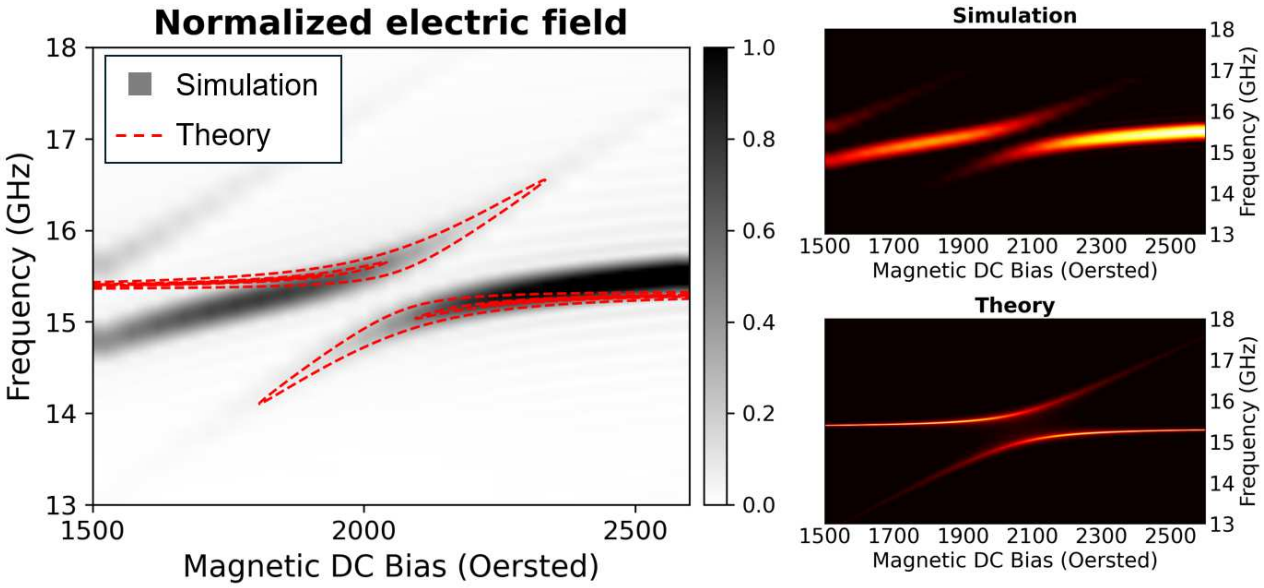
61. R. Roy and T. Kailath. Esprit-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(7):984–995, 1989.

62. D Bossini, M Pancaldi, L Soumah, M Basini, F Mertens, M Cinchetti, T Satoh, O Gomonay, and S Bonetti. Ultrafast Amplification and Nonlinear Magnetoelastic Coupling of Coherent Magnon Modes in an Antiferromagnet. *Physical Review Letters*, 127(7):77202, aug 2021.

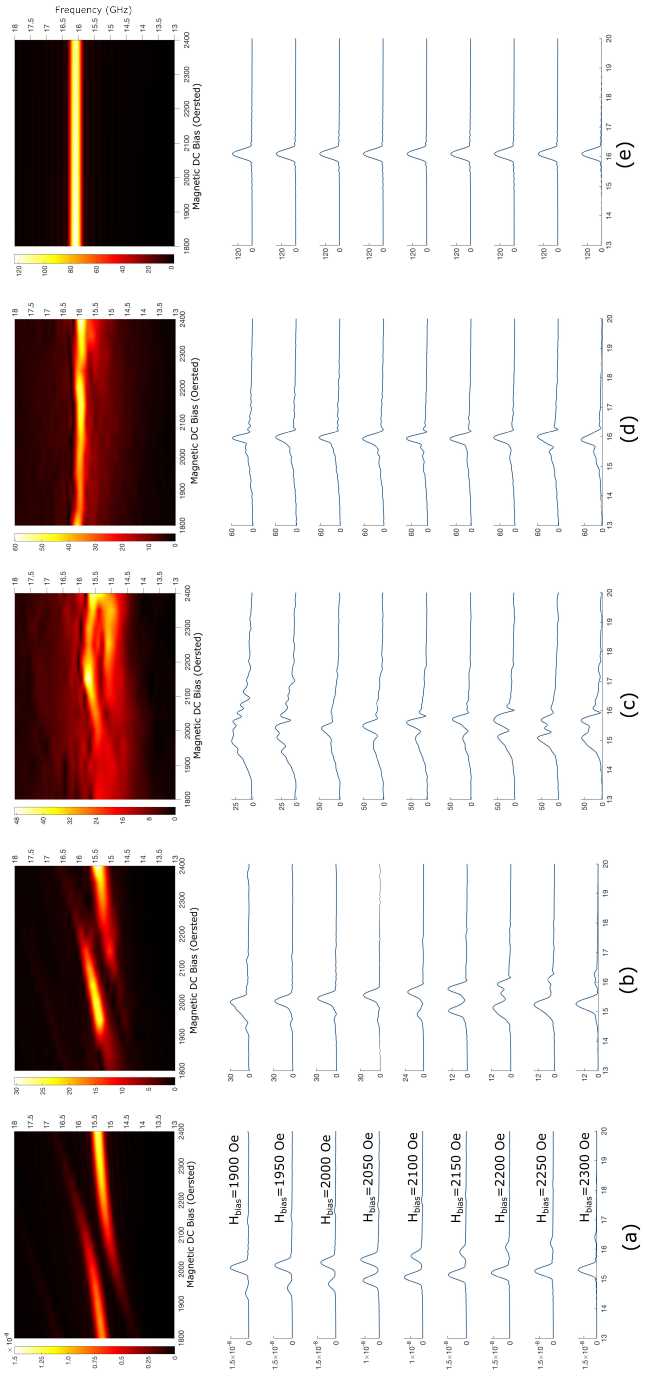
63. Yunyoung Hwang, Jorge Puebla, Kouta Kondou, Carlos Gonzalez-Ballestero, Hironari Isshiki, Carlos Sánchez Muñoz, Liyang Liao, Fa Chen, Wei Luo, Sadamichi Maekawa, and Yoshichika Otani. Strongly Coupled Spin Waves and Surface Acoustic Waves at Room Temperature. *Physical Review Letters*, 132(5):56704, 2024.

64. Markus Aspelmeyer, Tobias J Kippenberg, and Florian Marquardt. Cavity optomechanics. *Reviews of Modern Physics*, 86(4):1391–1452, dec 2014.

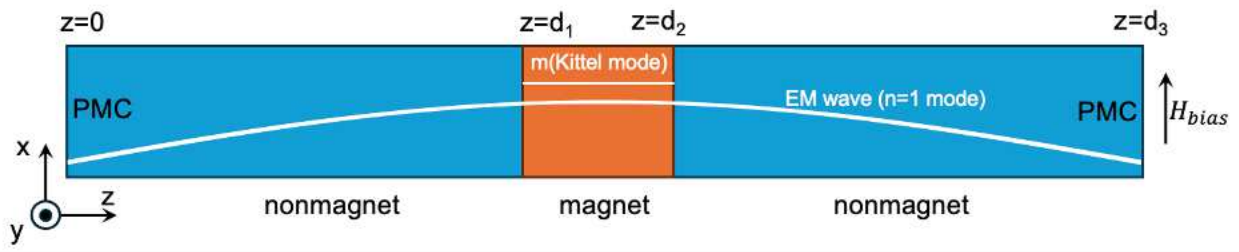
65. Hank Childs, Eric Brugger, Brad Whitlock, Jeremy Meredith, Sean Ahern, David Pugmire, Kathleen Biagas, Mark C. Miller, Cyrus Harrison, Gunther H. Weber, Hari Krishnan, Thomas Fogal, Allen Sanderson, Christoph Garth, E. Wes Bethel, David Camp, Oliver Rubel, Marc Durant, Jean M. Favre, and Paul Navratil. High Performance Visualization–Enabling Extreme-Scale Scientific Insight, 2012.



Extended Data Figure 1. The theoretical analysis of the frequencies and linewidths of the hybrid polaritons agrees well with the numerical simulations. The right-hand figures show the color map of the simulated electric field (top) and the theoretically calculated electric field spectrum (bottom). In the left figure, the theoretical results are plotted as contour lines (red dashed) for clearer visualization. See **Supplementary Information** for details.



Extended Data Figure 2. Electric-field spectra as the excitation intensity is increased. Stronger microwave magnetic fields suppress the magnon dynamics and leave only the cavity photon mode. Such behaviour arises because, at high powers, nonlinear spin-wave interactions – particularly three-magnon splitting – induce a Suhl instability that transfers energy from the uniform ferromagnetic resonance to non-uniform spin waves; this saturation of the Kittel mode causes the magnon–photon anticrossing gap to close and yields a single photon-like resonance, a phenomenon sometimes described as the high-power cavity magnon-polaritons regime [60]. See **Supplementary Information** for details.



Supplementary Figure 1. 1D hybrid magnon-photon cavity.

Supplementary Information:

HPC-Driven Modeling with ML-Based Surrogates for Magnon-Photon Dynamics in Hybrid Quantum System

1 Analytical model of a 1D hybrid magnon–photon cavity

Consider a 1D hybrid magnon-photon cavity where all physical quantities vary only along the z axis; see Supplementary Fig. 1.

A bias magnetic field $\mathbf{H}^{\rm bias} = (H_x^{\rm bias},\, 0,\, 0)$ is applied along $+x$, stabilizing a spatially uniform magnetization, $\mathbf{m}(t{=}0) = \mathbf{m}^0 = (1, 0, 0)$, along the same direction in the permalloy. Perfect magnetic conductor (PMC) boundary conditions, i.e., $\mathbf{n}\times\mathbf{H} = 0$ and $\mathbf{n}\cdot\mathbf{E} = 0$ with $\mathbf{n} \parallel \hat{\mathbf{z}}$, confine a half-wavelength standing EM wave of angular frequency $\omega_{\rm ph}$. The photon frequency $\omega_{\rm ph}$ is set by the cavity length and does not vary with $H^{\rm bias}$.

Because the EM wavelength is much larger than the magnet thickness $d_2 - d_1$, the EM fields are approximately uniform inside the magnet ($k_{\rm ph} \approx 0$). A microwave magnetic drive

$$\mathbf{h}(t) = \left(h_x^0 e^{-{\rm i}\omega t},\; h_y^0 e^{-{\rm i}\omega t},\; h_z^0 e^{-{\rm i}\omega t}\right)$$

of angular frequency ω is applied to the magnet region and excites a spatially uniform (Kittel) magnon with zero wavevector $k_m{=}0$. When ω equals the Kittel resonance ω_m (FMR), the magnon amplitude peaks and the absorbed microwave power $P_{\rm abs}$ is maximized. The resonance ω_m is tunable with $H^{\rm bias}$. When ω_m is tuned close to $\omega_{\rm ph}$, the Kittel magnon and cavity photon hybridize into a magnon-polariton whose mode splitting is observable in the map $P_{\rm abs}(\omega, H^{\rm bias})$. Below we derive $P_{\rm abs}(\omega, H_x^{\rm bias})$ for fixed field direction ($+x$) and fixed drive amplitude:

$$P_{\rm abs}(\omega, H^{\rm bias}) \;\propto\; {\rm Im}\left(\mathbf{h}^{*\mathsf{T}}\Delta\mathbf{m}\right) = {\rm Im}\left(\mathbf{h}^{*\mathsf{T}}\boldsymbol{\chi}\,\mathbf{h}\right) = {\rm Im}\left(h_i^0\,\chi_{ij}\,h_j^0\right), \quad i,j \in \{x,y,z\}, \tag{S1}$$

where the (complex) magnetic susceptibility $\boldsymbol{\chi}$ is defined by $\Delta\mathbf{m} = \boldsymbol{\chi}\,\mathbf{h}$ with $h_i = h_i^0 e^{-{\rm i}\omega t}$ and $h_i^* = h_i^0 e^{{\rm i}\omega t}$. The tensor $\boldsymbol{\chi}$ follows from the linearized Landau–Lifshitz–Gilbert (LLG) equation and Maxwell's equations. The LLG equation reads

$$\frac{\partial\mathbf{m}}{\partial t} = -\gamma\,\mathbf{m}\times\left(\mathbf{H}^{\rm eff} + \mathbf{h}\right) + \alpha\,\mathbf{m}\times\frac{\partial\mathbf{m}}{\partial t}. \tag{S2}$$

For an isotropic permalloy supporting only the Kittel mode, and neglecting magnon–phonon coupling, the effective field is

$$\mathbf{H}^{\text{eff}} = \mathbf{H}^{\text{bias}} + \mathbf{H}^{\text{EM}} + \mathbf{H}^{\text{d}} + \mathbf{h}, \qquad \mathbf{H}^{\text{EM}} = (H_x^{\text{EM}},\, H_y^{\text{EM}},\, 0), \qquad \mathbf{H}^{\text{d}} = (0,\, 0,\, -M_s m_z).$$

In this 1D geometry $H_z^{\text{EM}}(z,t) = 0$; the time-varying $m_z(t)$ produces only a dynamic demagnetizing field $\Delta H_z^{\text{d}}(t) = -M_s m_z(t)$. Let $\mathbf{m} = \mathbf{m}^0 + \Delta\mathbf{m}(z,t)$, with

$$\mathbf{m} = \left(m_x^0 + \Delta m_x^0 e^{-\mathrm{i}\omega t},\ \Delta m_y^0 e^{-\mathrm{i}\omega t},\ \Delta m_z^0 e^{-\mathrm{i}\omega t}\right).$$

Keeping only linear terms yields

$$\begin{pmatrix} \mathrm{i}\omega & 0 & 0 \\ 0 & \mathrm{i}\omega & A_{23} \\ 0 & A_{32} & \mathrm{i}\omega \end{pmatrix} \begin{pmatrix} \Delta m_x^0 e^{-\mathrm{i}\omega t} \\ \Delta m_y^0 e^{-\mathrm{i}\omega t} \\ \Delta m_z^0 e^{-\mathrm{i}\omega t} \end{pmatrix} = \gamma \begin{pmatrix} 0 \\ -m_x^0\, h_z^0 e^{-\mathrm{i}\omega t} \\ m_x^0\, h_y^0 e^{-\mathrm{i}\omega t} \end{pmatrix} + \begin{pmatrix} \Omega_x(z,t) \\ \Omega_y(z,t) \\ \Omega_z(z,t) \end{pmatrix}, \tag{S3}$$

with

$$A_{23} = \mathrm{i}\alpha\omega - \gamma H_x^{\text{bias}} - \gamma M_s, A_{32} = -\mathrm{i}\alpha\omega + \gamma H_x^{\text{bias}}, \tag{S4a}$$

$$\Omega_x(z,t) = 0, \quad \Omega_y(z,t) = 0, \quad \Omega_z(z,t) = \gamma\, H_y^{\text{EM}}(z,t). \tag{S4b}$$

The quantities Ω_i describe the bidirectional magnon–photon coupling (units of Hz). Introduce an auxiliary coupling matrix ϑ_{ij} via

$$\Omega_i(z,t) = \vartheta_{ij}(z)\, \Delta m_j(t) = \vartheta_{ij}(z)\, \Delta m_j^0 e^{-\mathrm{i}\omega t}.$$

Then (S3) and (S4b) lead to

$$\begin{pmatrix} \Delta m_x^0 e^{-\mathrm{i}\omega t} \\ \Delta m_y^0 e^{-\mathrm{i}\omega t} \\ \Delta m_z^0 e^{-\mathrm{i}\omega t} \end{pmatrix} = \boldsymbol{\chi} \begin{pmatrix} h_x^0 e^{-\mathrm{i}\omega t} \\ h_y^0 e^{-\mathrm{i}\omega t} \\ h_z^0 e^{-\mathrm{i}\omega t} \end{pmatrix}, \tag{5a}$$

$$\boldsymbol{\chi} = \gamma \begin{pmatrix} \mathrm{i}\omega & 0 & 0 \\ 0 & \mathrm{i}\omega & A_{23} \\ -\vartheta_{zx} & A_{32} - \vartheta_{zy} & \mathrm{i}\omega - \vartheta_{zz} \end{pmatrix}^{-1} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{5b}$$

The unknowns in $\boldsymbol{\chi}$ are ϑ_{zi}, which follow from

$$\Omega_z(z,t) = \gamma H_y^{\text{EM}}(z,t) = \vartheta_{zj}(z)\, \Delta m_j(t),$$

once $H_y^{\text{EM}}(z,t)$ is obtained from Maxwell's equations under PMC boundaries.

EM field solution (1D). With $\mathbf{H}^{\mathrm{EM}} = (H_x, H_y, 0)$ and uniform $\mathbf{M}(t) = (M_x, M_y, M_z)$, Maxwell–Ampère's law $\nabla \times \mathbf{H} = \mathbf{J} + \partial \mathbf{D}/\partial t$ with $\mathbf{J} = \sigma \mathbf{E}$ and $\mathbf{D} = \epsilon_0 \epsilon_r \mathbf{E}$ yields

$$\nabla(\nabla \cdot \mathbf{H}) - \nabla^2 \mathbf{H} = \nabla \times \left(\sigma \mathbf{E} + \frac{\partial \mathbf{D}}{\partial t} \right). \tag{6}$$

Using $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$, Faraday's law $\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t$, and assuming plane-wave time dependence $e^{-\mathrm{i}\omega t}$, the x, y components satisfy

$$\frac{\partial^2 H_i(z)}{\partial z^2} + k^2 H_i(z) = -k^2 M_i, \qquad k^2 = \epsilon_0 \epsilon_r \mu_0 \omega^2 + \mathrm{i}\sigma \mu_0 \omega, \qquad i \in \{x, y\}. \tag{S7}$$

In the magnet (m) and in the nonmagnetic cavity region (c), the 1D solutions are

$$H_i^m(z) = H_i^{(m+)} e^{\mathrm{i}k^m z} + H_i^{(m-)} e^{-\mathrm{i}k^m z} - M_i, \tag{S8a}$$

$$H_i^c(z) = H_i^{(c+)} e^{\mathrm{i}k^c z} + H_i^{(c-)} e^{-\mathrm{i}k^c z}, \tag{S8b}$$

where $H_i^{(m\pm)}$ and $H_i^{(c\pm)}$ are the forward/backward wave amplitudes in magnet and cavity, respectively. From $\nabla \times \mathbf{H} = (\sigma - \mathrm{i}\epsilon_0 \epsilon_r \omega)\mathbf{E}$, one obtains

$$E_i^m(z) = \left(\sigma^m - \mathrm{i}\epsilon_0 \epsilon_r^m \omega\right)^{-1} \left(-\frac{\partial H_j^m(z)}{\partial z} \right), \qquad i \neq j,\ i, j \in \{x, y\}, \tag{S9a}$$

$$E_i^c(z) = \left(\sigma^c - \mathrm{i}\epsilon_0 \epsilon_r^c \omega\right)^{-1} \left(-\frac{\partial H_j^c(z)}{\partial z} \right), \tag{S9b}$$

with PMC boundaries $H_i^c(z) = 0$ at $z = 0, d_3$ and EM continuity $H_i^c = H_i^m$, $E_i^c = E_i^m$ at $z = d_1, d_2$ ($i = x, y$). These conditions determine $H_i^{(c\pm)}$ and $H_i^{(m\pm)}$ in terms of M_i.

Consequently, $H_y^{\mathrm{EM}}(z,t) \equiv H_y^m(z) e^{-\mathrm{i}\omega t}$. For the parameter set in Table 1, at resonance $\omega/2\pi = \omega_m/2\pi = 14.17\,\mathrm{GHz}$ with $H_x^{\mathrm{bias}} = 1800\,\mathrm{Oe}$, one finds

$$H_y^{\mathrm{EM}}(z,t) = \left[(424551 - 257055\,\mathrm{i})\, e^{\mathrm{i}k^m z} + (424224 + 256900\,\mathrm{i})\, e^{-\mathrm{i}k^m z} \right] \Delta m_y^0\, e^{-\mathrm{i}\omega t}, \quad d_1 < z < d_2, \tag{S10}$$

with $k^m = (297.033 + 0.188\,\mathrm{i})\ \mathrm{m}^{-1}$, whose imaginary part encodes ohmic loss. From the definition of ϑ_{ij},

$$H_y^{\mathrm{EM}}(z,t) = \frac{1}{\gamma} \left(\vartheta_{zx}(z)\, \Delta m_x^0 + \vartheta_{zy}(z)\, \Delta m_y^0 + \vartheta_{zz}(z)\, \Delta m_z^0 \right) e^{-\mathrm{i}\omega t}. \tag{S11}$$

Comparing (S10) with (S11) shows $\vartheta_{zx} = \vartheta_{zz} = 0$ and

$$\vartheta_{zy}(z) = (424551 - 257055\,\mathrm{i})\, e^{\mathrm{i}k^m z} + (424224 + 256900\,\mathrm{i})\, e^{-\mathrm{i}k^m z}, \qquad d_1 < z < d_2. \tag{S12}$$

Because the EM wave is nearly uniform in the magnet,

$$H_y^{\mathrm{EM}}(t) \approx \langle H_y^m(z) \rangle e^{-\mathrm{i}\omega t} = \frac{\langle \vartheta_{zy}(z) \rangle}{\gamma}\, \Delta m_y^0\, e^{-\mathrm{i}\omega t} = \frac{1}{\gamma(d_2 - d_1)} \left(\int_{d_1}^{d_2} \vartheta_{zy}(z)\, dz \right) \Delta m_y^0 e^{-\mathrm{i}\omega t}. \tag{S13}$$

Table 1: Parameters used in the analytical example.

Parameter	Value
σ^m	$10^{-3}\ \mathrm{S/m}$
σ^c	$10^{-3}\ \mathrm{S/m}$
ε_r^m	7.1
ε_r^c	1
M_s	$9.7 \times 10^5\ \mathrm{A/m}$
γ	$2.23 \times 10^5\ \mathrm{Hz\,m/A}$
α	0.003
$d_2 - d_1$	$7 \times 10^{-7}\ \mathrm{m}$
d_3	$3.67 \times 10^{-3}\ \mathrm{m}$
h_x^0	0
h_y^0	$10^3\ \mathrm{A/m}$
h_z^0	0

For $H_x^{\text{bias}} = 1800\,\text{Oe}$ and $\omega = \omega_m$, we obtain $\langle \vartheta_{zy} \rangle = 4.966 \times 10^9 + 5.859 \times 10^6\,\text{i}$ Hz, giving $\Omega_z(t) = (-2.222 \times 10^8 + 1.349 \times 10^7\,\text{i})\, e^{-\text{i}\omega_m t}$ Hz for $\Delta m_y^0 \approx -4.473 \times 10^{-2} + 2.7684 \times 10^{-3}\,\text{i}$ (the imaginary parts capture EM loss and phase lag). Using $\langle \vartheta_{zy} \rangle$ in (5b) yields the full $\boldsymbol{\chi}$ and, via (S1), P_{abs}.

2 Nonlinear Effect Analysis - Floquet Engineering.

For the nonlinear effect analysis, let us ignore the damping term in the nonlinear analysis since the α value is set to be very small ($\alpha = 0.003$ in our setup). At the center of the ferrite thin film, the EM magnetic field primarily involves H_x and H_z components, as shown in Figure 1 in the main manuscript. H_y is considered to be minimal since the fundamental TEM mode is predicted in the simulation. Therefore, at the center of the ferrite, $\mathbf{H}_{\text{eff}}$ can be decomposed into three components:

$$\mathbf{H}_{\text{eff}} = \hat{x} H_x + \hat{y} H_0 + \hat{z} H_z, \tag{21}$$

where H_0 is the constant magnetic bias and $H_{x,z}$ are the time-dynamic EM field components. In a stand-alone CPW resonator without any ferrite thin films, the resonant magnetic field can be simply expressed as $H_x = H_{x0} \cos(2\pi f_0 t) \sin(\pi y / y_0)$, where H_{x0} is a constant magnitude. With the ferrite thin film, the fundamental mode EM fields get distorted and hence excites an extra H_z component, which shares similar spatial and temporal distributions as H_x, giving an expression of $H_z = H_{z0} \cos(2\pi f_0 t) \sin(\pi y / y_0)$, where H_{z0} is a constant magnitude. Note that the EM H_y component is negligible, as shown by the simulation result, which agrees with the fundamental mode property. In our setup, $y_0 = \lambda/2$, thus in the center of the ferrite, where $y = \lambda/4$, $\sin(\pi y / y_0) = 1$, which simplifies the expression of H_x and H_z into

$$H_x = H_{x0} \cos(2\pi f_0 t), \quad H_z = H_{z0} \cos(2\pi f_0 t). \tag{22}$$

The simulated H_x and H_z fields contain two resonant peaks, i.e.,

$$H_x = H_{x1} \cos(2\pi f_1 t) + H_{x2} \cos(2\pi f_2 t), \quad H_z = H_{z1} \cos(2\pi f_1 t) + H_{z2} \cos(2\pi f_2 t). \tag{23}$$

For simplicity, let's start with just one resonance as indicated in Eq. (22).

To perform Floquet analysis, let us scalarize the LLG equation:

$$\frac{\partial M_x}{\partial t} = \mu_0 \gamma [M_y H_z - M_z (H_y + H_0)], \tag{24}$$

$$\frac{\partial M_y}{\partial t} = \mu_0 \gamma [M_z H_x - M_x H_z], \tag{25}$$

$$\frac{\partial M_z}{\partial t} = \mu_0 \gamma [M_x (H_y + H_0) - M_y H_x], \tag{26}$$

where $M_x^2 + M_x^2 + M_x^2 = M_S^2$. The terms H_x and H_z are shown in Eq. (22); and $H_y = 0$.

Substituting Eq. (22) into Eqs. (24), (25) and (26) reduces them to

$$\frac{\partial M_x}{\partial t} = \mu_0 \gamma [M_y H_{z0} \cos(2\pi f_0 t) - M_z H_0], \tag{27}$$

$$\frac{\partial M_y}{\partial t} = \mu_0\gamma[M_z H_{x0}\cos(2\pi f_0 t) - M_x H_{z0}\cos(2\pi f_0 t)], \tag{28}$$

$$\frac{\partial M_z}{\partial t} = \mu_0\gamma[M_x H_0 - M_y H_{x0}\cos(2\pi f_0 t)]. \tag{29}$$

In matrix form, the above equations can be rewritten as:

$$\frac{\partial}{\partial t}\begin{bmatrix} M_x \\ M_y \\ Mz \end{bmatrix} = \begin{bmatrix} 0 & H_{z0}\cos(2\pi f_0 t) & -H_0 \\ -H_{z0}\cos(2\pi f_0 t) & 0 & H_{x0}\cos(2\pi f_0 t) \\ H_0 & -H_{x0}\cos(2\pi f_0 t) & 0 \end{bmatrix}\begin{bmatrix} M_x \\ M_y \\ Mz \end{bmatrix}. \tag{30}$$

Note that H_{x0}, H_{z0} and H_0 are all constants, and H_{x0} and H_{z0} increase when we increase the input EM power.

Alternatively, the LLG equation can be written in spherical coordinates (note that the y-direction is the polar axis): $M_x = M_S\sin(\theta)\cos(\phi)$, $M_y = M_S\cos(\theta)$, and $M_z = M_S\sin(\theta)\sin(\phi)$. It follows that:

$$\frac{\partial[\sin(\theta)\cos(\phi)]}{\partial t} = \mu_0\gamma[\cos(\theta)H_{z0}\cos(2\pi f_0 t) - \sin(\theta)\sin(\phi)H_0], \tag{31}$$

$$\frac{\partial[\cos(\theta)]}{\partial t} = \mu_0\gamma[\sin(\theta)\sin(\phi)H_{x0}\cos(2\pi f_0 t) - \sin(\theta)\cos(\phi)H_{z0}\cos(2\pi f_0 t)], \tag{32}$$

$$\frac{\partial[\sin(\theta)\sin(\phi)]}{\partial t} = \mu_0\gamma[\sin(\theta)\cos(\phi)H_0 - \cos(\theta)H_{x0}\cos(2\pi f_0 t)]. \tag{33}$$

Further simplifying this, we obtain:

$$-\sin(\theta)\sin(\phi)\frac{\partial\phi}{\partial t} + \cos(\theta)\cos(\phi)\frac{\partial\theta}{\partial t} = \mu_0\gamma[\cos(\theta)H_{z0}\cos(2\pi f_0 t) - \sin(\theta)\sin(\phi)H_0], \tag{34}$$

$$-\sin(\theta)\frac{\partial\theta}{\partial t} = \mu_0\gamma[\sin(\theta)\sin(\phi)H_{x0}\cos(2\pi f_0 t) - \sin(\theta)\cos(\phi)H_{z0}\cos(2\pi f_0 t)], \tag{35}$$

$$\sin(\theta)\cos(\phi)\frac{\partial\phi}{\partial t} + \cos(\theta)\sin(\phi)\frac{\partial\theta}{\partial t} = \mu_0\gamma[\sin(\theta)\cos(\phi)H_0 - \cos(\theta)H_{x0}\cos(2\pi f_0 t)]. \tag{36}$$