

# An integrated microwave neural network for broadband computation and communication

BAL GOVIND

[bg373@cornell.edu](mailto:bg373@cornell.edu)

Cornell University <https://orcid.org/0000-0003-3499-2071>

**Maxwell Anderson**

Cornell University

**Fan Wu**

Cornell University <https://orcid.org/0000-0002-9470-7677>

**Peter McMahon**

Cornell University <https://orcid.org/0000-0002-1177-9887>

**Alyssa Apsel**

Cornell University <https://orcid.org/0000-0001-9199-2292>

---

## Article

### Keywords:

**Posted Date:** January 10th, 2025

**DOI:** <https://doi.org/10.21203/rs.3.rs-5494383/v1>

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

[Read Full License](#)

**Additional Declarations:** Yes there is potential Competing Interest. The authors have filed a provisional US patent application based on the neural processor presented in this article.

---

# An integrated microwave neural network for broadband computation and communication

Bal Govind<sup>\*, 1, a</sup>, Maxwell G. Anderson<sup>\*, 2, a</sup>, Fan O. Wu<sup>2</sup>,  
Peter L. McMahon<sup>†, 2, 3, a</sup> and Alyssa Apsel<sup>†, 1, a</sup>

<sup>1</sup> School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA

<sup>2</sup> School of Applied and Engineering Physics, Cornell University, Ithaca, NY, USA

<sup>3</sup> Kavli Institute at Cornell for Nanoscale Science, Cornell University, Ithaca, NY, USA

High-bandwidth applications, from multi-gigabit communication<sup>1–3</sup> and high-performance computing<sup>4,5</sup> to radar signal processing<sup>6,7</sup>, demand ever-increasing processing speeds. However, they face limitations in signal sampling and computation due to hardware and power constraints<sup>8–10</sup>. In the microwave regime, where operating frequencies exceed the fastest clock rates<sup>11</sup>, direct sampling becomes difficult, prompting interest in neuromorphic analog computing systems<sup>12,13</sup>. We present the first demonstration of direct broadband frequency-domain computing using an integrated circuit that replaces traditional analog and digital interfaces. This features a Microwave Neural Network (MNN) that operates on signals spanning tens of gigahertz, yet reprogrammed with slow, 150 MBit/sec control bitstreams. By leveraging significant nonlinearity in coupled microwave oscillators, features learned from a wide bandwidth are encoded in a comb-like spectrum spanning only a few gigahertz, enabling easy inference. We find that the MNN can search for bit sequences in arbitrary, ultra-broadband 10 GBit/sec digital data, demonstrating suitability for high-speed wireline communication. Notably, it can emulate high-level digital functions without custom on-chip circuits, potentially replacing power-hungry sequential logic architectures. Its ability to track frequency changes over long capture times also allows for determining flight trajectories from radar returns. Furthermore, it serves as an accelerator for radio-frequency machine learning<sup>14</sup>, capable of accurately classifying various encoding schemes used in wireless communication. The MNN achieves true, reconfigurable broadband computation, which has not yet

---

<sup>\*</sup>These authors contributed equally to this work.

<sup>†</sup>These authors co-supervised this work.

<sup>a</sup>Authors to whom correspondence should be addressed:

bg373@cornell.edu, mga58@cornell.edu, pmcmahon@cornell.edu, aba25@cornell.edu.

been demonstrated by classical analog modalities<sup>12,15</sup>, quantum reservoir computers using superconducting circuits<sup>16</sup>, or photonic tensor cores<sup>17</sup>, and avoids the inefficiencies of electro-optic transduction<sup>18,19</sup>. Its sub-wavelength footprint in a Complementary Metal-Oxide-Semiconductor process and sub-200 milliwatt power consumption enable seamless integration as a general-purpose analog neural processor in microwave and digital signal processing chips.

Managing communication and computation at hundreds of gigabits per second<sup>1-3</sup> for high-performance computing<sup>4,5</sup> is increasingly computationally expensive. It requires sampling and processing at clock speeds constrained by semiconductor physics and power limitations<sup>8-11</sup>, where higher speeds lead to increased power consumption and heat dissipation.

Consider, for example, the typical chain of electronic signal processing used by data centers, as shown in Fig. 1a. The first data center transmits high-speed data over an electrical or fiber-optic cable. During transmission, the bits are distorted and must be restored to their original integrity. This is done by boosting the signal and using complex synchronization circuits to reconstruct the transmission<sup>20</sup>. The deciphered bits are then sent to compute and memory blocks. Subsequently, newly computed bits are serialized and transmitted to the second data center. A key concern in this process is ensuring that signals are accurately timed and sampled after passing through lossy media, which requires power-hungry parallel processing.

We propose addressing the simultaneous challenges of data relay and computation from a new perspective. We recognize that fast digital data at tens of gigabits per second consists of microwave signals with an ultra-broadband spectrum. This realization enables us to operate directly in the frequency domain, without concern for how the bits were distorted in the time domain. We achieve this using an integrated circuit called a Microwave Neural Network (MNN). As shown in Fig. 1a, the MNN processes spectral components by capturing input data features that are informationally sparse but span a broad bandwidth. The key advantage lies in its ability to programmatically manipulate gigahertz-speed signals using only megahertz-speed controls. Its output encodes computations in a comb-like spectrum covering only a few gigahertz, enabling easy electronic readout. This can then be mapped to a binary output using a cheap linear regression model in post-processing.

One might also consider the potential impact of the MNN on wireless communication and sensing. As shown in Fig. 1b, a typical wideband radar receiver consists of multiple signal chains, each handling a narrow frequency band. Each chain includes a filter, a mixer to down-convert the microwave signal to a low-frequency baseband, and an analog-to-digital converter. The targets' positions and speeds are inferred with a backend digital computer. However, scaling such an architecture to cover tens of gigahertz of bandwidth for imaging complex target environments is hardware-intensive. We propose replacing this conventional system with the MNN, which monitors targets' movement over long capture periods as changes in carrier frequency, and manifests these inflections within a narrower output spectrum. A trained backend model then directly maps the spectral features to the targets' trajectories.

Notably, prior efforts to combine analog computing modalities with deep learning have not demonstrated the ability to perform fast, reconfigurable computations on wide bandwidth signals. These include systems like memristor crossbar arrays<sup>21</sup>, photodiode arrays<sup>22</sup> and photonic tensor cores<sup>17</sup>. Machine-learning tasks have been of the low-bandwidth variety like image, speech, or gesture recognition<sup>12,13,15</sup>. Recent attempts to operate at radio frequencies, using quantum superconducting circuits<sup>16</sup>, surface plasmon resonance structures<sup>23</sup> and spintronics with magnetic tunnel junctions<sup>24</sup>, only process a few megahertz of bandwidth. Some even involve bulky printed circuit boards<sup>25</sup> to form multi-level perceptrons. Also, recent microwave photonics chips<sup>18</sup>, while capable of broadband computation, are limited to a few, immutable math functions and are bulky and power-inefficient (peak electro-optic transduction efficiencies reach a few percent<sup>26</sup>). Similarly, Ising machines based on networks of coupled oscillators for combinatorial optimization problems have so far only been used in the megahertz range<sup>27,28</sup>.

In contrast, the MNN distinguishes itself as a high-speed, reconfigurable general-purpose analog computer. It is fabricated with standard Complementary Metal-Oxide-Semiconductor (CMOS) technology and as shown in Fig. 1c, occupies a compact, sub-wavelength footprint of only 0.088 mm<sup>2</sup> on chip. This would support seamless integration in microwave transceivers and Neural Processing Units<sup>29</sup>.

## Mechanism of the Microwave Neural Network

The integrated MNN we have designed is a nonlinear system that produces a comb-like, input-sensitive spectrum for computation. Fig. 2a illustrates the electromagnetic structure that generates this. It consists of one nonlinear waveguide (A) and three linear waveguides (B, C and D). The nonlinear waveguide's frequency modes are strongly influenced by the incoming microwave drive signals' amplitude and phase, while the linear waveguides' modes are largely unaffected by them. Ground-Signal-Ground-Signal-Ground (GSGSG) waveguides inject gigahertz-speed signals into the system. Miniature quadrature hybrid couplers, built on two overlapping metal layers, divide the power from these incoming signals and direct them to the waveguides. These smaller portions of drive signals then bounce off of the waveguides and add up at the couplers' output ports, to be extracted through another set of GSGSG waveguides.

The main source of input-sensitivity is a cascade of coupled nonlinear resonators within waveguide A (Fig 2b.i). These resonators comprise combinations of inductive segments paired with nonlinear capacitors. We utilize anti-parallel diodes<sup>30</sup> (Fig. 2c.i) that can generate a capacitance with polynomial linearity, where the degree of nonlinearity depends on the bias voltage applied to them and strength of the microwave signal. See Extended Data Fig 1 for the pretzel-shaped layout of the nonlinear waveguide. This contrasts with the structure of the linear waveguides, which are adjustable-length transmission lines (Fig. 2b.ii). Switches, installed periodically along their length, provide options to lengthen or shorten the return path<sup>31</sup> of the microwave signal to a DC power supply, without introducing distortion. Extended Data Fig. 2 shows the implementation details of this waveguide, including the

metal stack used.

More importantly, parametric (time-varying) coupling is established by turning on and off a pair of switches ( $S_{\text{par}}$ ) connected between pairs of waveguides. These switches are N-type Metal-Oxide-Semiconductor (NMOS) transistors (Fig. 2c.ii) that are driven by a low-speed 150 MBit/sec bitstream fed through a third Ground-Signal-Ground (GSG) waveguide. The sequence of on-off parametric coupling is key to programming the neural network for various computing tasks. Lastly, to sustain the nonlinearity induced by high-amplitude microwave transmission in the circuit at tens of gigahertz, despite the attenuation from cables and probes used in the experiment, regenerative gain is provided by cross-coupled transistor pairs<sup>32</sup> of thin-gate-oxide power-amplifier-class NMOS transistors (Fig. 2c.iii). The shape of the MNN's inherent spectral response, even without drives or parasitic switching, is highly dependent on factors such as the oscillators' core supply voltage and the biases applied to the nonlinear capacitors. A couple of these measured spectra are shown in Fig 2d.i and characterized in more detail in Extended Data Figs. 7 and 8. This is, to our knowledge, the first demonstration of a microwave comb generated by actively coupled nonlinear resonances in integrated CMOS circuits.

Upon receiving the 12 Gbit/sec drive, which is inherently ultra-broadband, the spectral response becomes complex, as shown in Fig. 2e.i. The nonlinearities in the MNN translate features of the full input spectrum to the range where its response is most prominent. Applying low-frequency parametric modulation further distorts the MNN's response, as Fig. 2e.ii shows. This reprogrammable, feature-rich spectrum is well-suited for machine learning inference tasks.

This design stands apart from traditional CMOS oscillators<sup>33</sup>, which rely on symmetry for stable, single-tone oscillation. It also differs from complex pulse-sharpening circuits used to generate weak harmonic combs<sup>34</sup> for spectroscopy, as well as from narrowband combs created by passively coupling linear and nonlinear resonators with high quality factors (over 2000), which are limited to simple tasks like low-frequency range finding<sup>35</sup>. Since we aim to implement this analog computer using a commercial CMOS process, designing electromagnetic structures with quality factors exceeding 40 is impractical. Furthermore, unlike stable optical frequency sources such as Kerr-combs<sup>36</sup> and electro-optic frequency combs<sup>37</sup>, which are well isolated from messy external drive signals, our design intentionally exposes the coupled waveguides to incoming microwaves. It is this intentional exposure to broadband inputs that enables near instantaneous computations facilitated by the resonators' internal nonlinearities and asymmetry.

It may be helpful to reduce the circuit to its most essential components, as shown in Extended Data Fig. 3. In the experiment, we reduced the number of physical circuit parameters by keeping the linear waveguides (B, C and D) highly detuned from the nominal oscillation frequency of waveguide A. By applying generalized Coupled-Mode Theory<sup>38</sup> as formulated in Note 3 of the Methods Section, the system can be described as a group of connected nonlinear modes, with the first nonlinear mode linked to a linear mode through slow parametric coupling and a fixed phase delay. The system is powered by saturable gain, and the parametric oscillations are modulated by fast microwave drive signals, which dynamically reconfigure the impedances of the resonators and shape the steady-state spectral

response of the system.

Extended Fig. 4 shows how the system’s dynamics are influenced by the initial conditions of the individual nonlinear modes without parametric coupling, while Extended Fig. 5 illustrates its dynamics under the influence of the parametric bitstream. Interestingly, the MNN’s dynamics bear similarities to neural network models of memory formation, particularly through “attractor networks”<sup>39</sup>. In neural systems, memories are stored in stable, repeating patterns of information exchange, forming within interconnected nodes, or neurons. These attractor networks emerge in response to external inputs, evolving weights between nodes and producing persistent patterns. Over time, however, these patterns shift from stable, structured behavior to more chaotic states, allowing multiple networks to coexist within the system. As the networks transition from fixed-point attractors to chaotic attractors, the system becomes increasingly random, leading to memory decay and eventual loss. Extended Figs. 4 and 5 reveal that in this circuit, similar transitions from structured to more chaotic states occur. The slightly-chaotic states, as shown by the Poincaré maps<sup>40</sup> (as in Extended Data Fig. 5 b.iv and c.iii), could evolve into highly chaotic patterns (Extended Data Fig. 5 c.iv) and fully-chaotic states (Extended Fig 4 c.iv). This evolution reflects the presence and degradation of memory in the system, arising from specific interactions between the microwave drive signals and the frequency-modulated microwave parametric oscillations.

## Emulating high-speed digital functions with a microwave circuit

Recognizing that gigabit-speed digital signals, composed of square-wave signals, are actually analog signals with spectral content spanning tens of gigahertz (as seen in Fig. 2e.i), suggests that the MNN can perform computations directly in the frequency domain using microwave-circuit behavior. This is in contrast to traditional digital hardware, which operates in the time domain. The MNN manipulates signals and expresses its output more prominently as oscillatory modes within a narrower band of a few gigahertz, centered around its comb-like spectrum. This approach bypasses the need for strict signal integrity in bit-level, time-domain calculations. It resembles compressed spectrum sensing, as features from an incoming signal’s ultra-wide bandwidth are captured and strongly manifested in the MNN’s narrower nominal frequency range. Consequently, fewer “compressed” features can be used to train a single-layer digital neural network in post-processing. See Methods Section 5 for training details.

Figure 3 presents the first attempt to emulate ultra-high-speed digital operations without relying on fixed-functionality digital CMOS circuits. 32-bit bitstreams are fed in slowly and repeatedly at a 150 MBit/sec rate. The nonlinear resonances react instantaneously and the MNN’s output is recorded over several cycles, which the spectrum analyzer averages to ensure a reliable Fourier Transform. The experimental setup for feeding multi-gigahertz digital and RF data into the MNN, along with slow parametric bits, is detailed in Methods section 1 and in Extended Data Fig. 6. During inference, the effects of different combinations of parametric bits are observed. By reprogramming these MBit/sec parameters, various

GBit/sec digital operations can be emulated, and the system’s accuracy can be optimized for specific tasks. Here, we focus on the computed features observed within the 10 to 14 gigahertz range, corresponding to the *X*-band and *Ku*-band frequencies used in satellite communications.

First, Fig. 3b.i shows the measured accuracy in emulating primitive digital gate behavior at 10 GBit/sec, without error correction. For example, a representative bitstream could be [01000110000111110111110101010000]. Our focus is to evaluate the MNN’s accuracy in predicting the bitwise NAND of the inner eight bits of the first sixteen bits,  $A = [00011111]$ , and the inner eight bits of the second sixteen bits,  $B = [01111101]$ . The correct result is  $A \text{ NAND } B = [11100010]$ . We found that by adjusting the content of the 150 MBit/sec 32-bit parametric bitstream and extracting features within only 4 GHz of the output spectrum, there exists a set that produces the correct NAND operation (and similarly for NOR), irrespective of the incoming fast 32-bit ultra-broadband digital signal. In other words, a “golden” parametric bitstream can frequency-modulate the parametric oscillations to perform computations that, in the abstract spectral domain, effectively emulate the behavior of an 8-bit NAND operation. Also, we must emphasize that, to demonstrate the MNN’s behavior as a deep, nonlinear neural network, we implemented the digital backend using a simple linear regression model to map the output spectra to the predicted bits. This ensures that the accuracy in emulating gate operations is entirely due to the MNN’s inherent nonlinear behavior. After validating the output on 20% of a dataset derived from 10,000 sample inputs and upon action of 200 parametric bitstreams, with the linear layer trained for mapping in each case, the best measured accuracy was around 85%, while the worst was 81%.

One might expect that if mimicking a low-level gate operation yields less than 100% accuracy, emulating a more complex digital operation involving many cascaded gates would be significantly worse. However, our method achieves accuracy independent of the circuit hardware required for a specific operation on a digital computer, regardless of the number of gates involved. To investigate this, we considered an architecture with multiple hierarchical layers of binary adders, consisting of hundreds of gates that form a population counter—a circuit that counts the number of 1’s in an input bitstream (shown in Fig. 3b.ii). For the input bitstream in the example above, the correct result would be 16. Despite the apparent increased complexity, our approach to emulating this behavior with a parametric bitstream, with outputs mapped with a linear layer, achieves an accuracy of 81% on the validation dataset. This could potentially remove the need for large application-specific circuits to be physically present on-chip. Curiously, we observe that the accuracy of predicting the correct final output is significantly different for various parametric bitstreams (a variation of 65%), for this high-level hardware operation. This highlights that only certain 150 MBit/sec parameter sequences can properly harness the nonlinearity from frequency-modulated parametric oscillations to enable clockless emulation of otherwise time-consuming, sequential bitwise digital functions on multi-gigabit/sec data.

Importantly, we find that our approach to emulating high-level operations is well-suited for application in high-speed wireline communication. Traditional modems rely on complex clock-and-data recovery circuits to decipher which bits were transmitted through a cable between data centers. They help compensate the clock frequency mismatches be-

tween a transmitter and a receiver. To handle frequency-dependent losses in interconnects and inter-symbol interference, gain-restoring circuits like continuous-time linear equalizers<sup>41</sup> are required at both ends of a channel (as Fig 1a indicates). At speeds greater than 10 GBit/second, inter-symbol interference becomes more challenging, and a technique called Maximum Likelihood Sequence Detection<sup>42</sup> is necessary to infer which bit patterns were transmitted. This technique involves several time-interleaved, high-resolution and power-hungry analog-to-digital converters. We take a different approach by treating this problem as a “linear search” for bit sequences in the incoming data, aware that a single bit flip can significantly alter the MNN’s spectral response. Once again, we toggle low-frequency 150 MBit/sec parametric bits, in conjunction with a simple linear backend. In this instance, the system classifies if the output spectra contains features of the queried bits. This process was validated for both 10 GBit/sec (Fig. 3c.i) and 5 GBit/sec (Fig. 3c.ii) input data and for numerous parametric bitstreams to find the slowly fed MBit/sec parameters that optimize performance.

The measured search accuracy, with no error correction applied, for queried bitstreams ranging from three to eight bits in length within the incoming 32-bit data, is shown in Fig. 3c. Our hypothesis is that shorter queried words (for 3 or 4 bits long) might generate more identifiable features in the spectrum, leading to a high search accuracy of 90%, as their bits appear more frequently. Conversely, longer words (8 bits or more) have distinct features that appear less frequently, but they are still easily identifiable and found with high accuracy. For intermediate-length sequences (5 or 6 bits), we observe a slightly higher error rate, likely due to imperfect memory in the system, i.e., this method recognizes that such sequences exhibit higher entropy. This makes their outcomes less predictable and requires extracting more information from the spectrum to achieve higher search accuracy. This scheme could provide a cost-effective alternative to the Maximum Likelihood Sequence Detection technique for enhancing the signal-to-noise ratio. Unlike that method, which requires recovering the time-domain signal with high resolution and large bandwidth to accurately match a distorted waveform to a reference bit sequence, our scheme uses only a small fraction of the input signals’ bandwidth. Supplementary Fig. 3a shows the MNN’s advantage in detecting bitstreams compared to using a linear layer alone.

Finally, to demonstrate the MNN’s ability to perform complex conditional logic, we test its capability to emulate both bit sequence detection and population counting within a single block, as Fig. 3d shows. This would mimic a simple data-center-link-plus-compute operation. Specifically, the algorithm searches for a particular bit sequence if the count of 1’s falls below a threshold and searches for an alternative bitstream if it exceeds that threshold. For example, using the [010001100001111011110101010000] bitstream mentioned above, since there are 16 ones, the system searches for [10101] (not [11100]). As [10101] is present, the result is “yes”. Achieving such bit-counting and detection at 10 GBit/sec is typically infeasible without clocked operations on Application-Specific Integrated Circuits or Field-Programmable Gate Arrays with RF transceivers operating over 20 GHz, which consume multiple watts of power. In contrast, our method manipulates only the slow 150 MBit/sec parametric bitstream. Even without optimizing the physical circuit parameters, the MNN combined with a single linear layer achieves 75% accuracy in predicting the algorithm’s final output on the first attempt. Importantly, the system consumes no more than the nominal

160 mW required by the coupled microwave oscillators.

## Detecting radar targets and classifying wireless signal encoding schemes

The MNN’s ability to detect subtle frequency changes indicates that it is well-suited for broadband radar<sup>43</sup> applications. We are particularly interested in observing its response to variations in the distances of flying targets from a radar tower, which are reflected in the time-varying frequency of the carrier wave. Fig. 4a shows a simulated airspace we created (see Methods section 1 for details), featuring multiple aircrafts following distinct polygonal trajectories. These trajectories vary in orientation, radius, and origin. The targets move through waypoints along these paths at unique velocities. Their flight patterns cause instantaneous changes in the radar signal reflected off of them, which, when downconverted to low baseband frequencies, alter the analog voltage waveforms received by the radar tower. These waveforms are recorded over a 100  $\mu$ s time base (“fast-time”) as shown in Fig 4b.i and once every sixty milliseconds (“slow-time”). We then modulate the center frequency of a square wave based on the instantaneous analog voltage, ranging from 100 MHz to 2.1 GHz, as illustrated in Fig. 4b.ii. This frequency-modulated signal is fed to the MNN, from which we extract the output response averaged across many cycles in the spectral range of 8 to 10 GHz. Aspects of the flight trajectories are then inferred by a digital neural network backend (see Methods section 6 for its architecture).

After simulating five hundred flight scenarios, each with one thousand fast-time captures, we characterize the targets’ flight patterns using the truncated output spectra from the MNN. Typically, this task would require multiple smaller-band RF receivers, including mixers and ADCs, along with digital inference through CPUs or GPUs. However, we find these may be unnecessary, as the MNN can learn flight patterns by forming distinct responses to frequency changes over long capture times. Supplementary Notes 1 and 2 discuss how this analog processing scheme could be integrated in wideband radar receivers.

Figure 4a.v demonstrates the MNN’s ability to predict the number of dynamic targets in simulated flight scenarios. Supplementary Fig. 3b highlights its advantage, showing lower classification accuracy when using only the backend neural network with identical training parameters. Notably, the MNN can isolate specific targets’ movements in the airspace. For example, in Fig. 4a.vi, the MNN-plus-digital-backend combination accurately estimates the speeds of the fastest targets in the validation set. Additionally, as shown in Fig. 4a.vii, the system detects various polygonal flight trajectories—such as triangular, square, or pentagonal—and achieves high F1-scores (harmonic mean of precision and recall) across different numbers of aircraft. These insights help infer whether unidentified aircraft are potential adversaries or allies.

While processing very high-frequency signals is crucial, testing the lowest frequency the MNN can handle is equally valuable. This enables the expansion of data from smaller bandwidths to larger ones while adding features based on its learned patterns. To explore this

capability, we investigated whether the MNN can identify the encoding schemes for wireless communication, which inherently utilizes narrowband techniques. For this, we used the RadioML2016.10a<sup>44–46</sup> dataset, a standard training set used in RF machine learning studies. It includes eleven modulation classes—nine digital and two analog. In our experiment (Fig. 4b.i), various baseband signals modulate a 50 MHz carrier wave, which is fed to the MNN with a signal-to-noise ratio of 18 dB. The MNN’s sensitivity enables it to transform transient changes in the low-frequency drive signal into observable features across several gigahertz, even when significantly detuned from the nominal operating range around 12 GHz. These features were then only extracted between 8 and 8.5 GHz to train the backend linear layer.

Fig. 4b.ii shows that some parameters lead to very high modulation classification accuracy, while others perform poorly (Fig. 4b.i). This reinforces the importance of stimulating the coupled waveguides’ modes with the appropriate time sequence of coupling. Again, the simple linear backend is used solely to prove that the accuracy of the MNN’s nonlinear mapping, of around 88%, can match that of digital neural networks<sup>44–46</sup>. Moreover, this accuracy suggests that the MNN could serve as a deep learning accelerator, potentially replacing a full ResNet<sup>47</sup> neural network while significantly reducing model size, as the comparison in Supplementary Table 1 shows. This capability is ideal for edge computing, demonstrating its role beyond merely acting as a smart-sensing front end.

## Discussion and outlook

The integrated MNN, unlike digital processors and RF transceivers, is able to perform broadband computations using a slow control mechanism, without the need for application-specific circuits on chip. Its low power consumption across a broad range of frequencies and miniature footprint in a standard CMOS process allow for integration into production-scale microwave electronics and data conversion systems.

While the current integrated circuit, based on parametrically coupled waveguides, demonstrates capabilities such as input-signal sensitivity and parametric reprogrammability, the physical parameters—such as nonlinear capacitor biases, resonator frequencies, and saturable gain—were fixed during the experiment. Dynamically adjusting these parameters and employing parametric coupling in a more analog fashion, rather than relying solely on binary-bitstream-driven switches, could enhance classification accuracy. Additionally, the component count on chip could be reduced by using a single linear waveguide instead of three. Moreover, a compact form-factor could also enable the creation of arrays of interconnected combs to generate richer output spectra, providing more features within a compressed bandwidth to enhance training data. This smaller band could be read using on-chip mixers and baseband circuitry (see Supplementary Note 2 for an example implementation) instead of spectrum analyzers. Also, it should be noted that this coupled-waveguide structure, with gain, is only one method for achieving complex input-output relations. Among other alternatives, one could leverage 3D stacking of RF metal layers in CMOS to create compact, multi-level perceptron-based diffractive-type neural networks which might be suited for specific tasks requiring optimization for expressivity, power and circuit area.

In the method of inference presented in this article, our system’s low-cost digital backend maps output spectra to bits (for digital emulation) and RF signal attributes, iterating to find the most accurate parametric bitstreams. The system’s accuracy during radar target tracking suggests that further optimization of slow control bitstreams could be achieved by collecting additional field deployment data. Regardless, this is only a representative approach to MNN programming. Future work could explore advanced techniques such as improved search algorithms, gradient-based optimization of time-varying parameters, and end-to-end joint training of both the MNN and backend network using reinforcement learning. Exhausting the design space could ultimately lead to a band-agnostic neural processor capable of decoding complex ultra-high-speed digital data and mm-wave signals spanning hundreds of gigahertz.

## Methods

### 1. Experimental setup and training data generation

Extended Data Fig. 6 illustrates the experimental setup for testing the Microwave Neural Network. The CMOS chip is wire-bonded to a printed circuit board that connects it to power supplies, which provide the necessary voltages for the oscillator cores, drivers, and the on-chip Serial-to-Parallel Interface (SPI). The SPI controls the static voltage drivers that operate the switches on the linear resonators. Additionally, six bias voltages are supplied to the nonlinear capacitors embedded in waveguide *A*, along with a bias voltage to set the center-frequency of the broadband couplers.

Slow 150 MBit/sec bitstreams, which modulate the parametric coupling between pairs of waveguides, are generated by a HP 8133A pulse pattern generator. These bitstreams are delivered to the gates of the parametric switches via short wirebonds and on-chip Ground-Signal-Ground (GSG) waveguides. The fast multi-Gigabit-speed input data, used for digital operation emulation and frequency-modulated radar, is generated by Anritsu MP1758A and MP1763C (optional) pulse pattern generators. These produce input signals with a maximum speed of 12.5 GBit/s and an amplitude of 2 V peak-to-peak, which are directly applied to the input ports feeding waveguides *A* and *B* through a PicoProbe GSGSG probe. The output microwave spectra are measured using Rohde & Schwarz FSU26 and Agilent 8564EC spectrum analyzers, which are connected through a separate PicoProbe<sup>TM</sup> GSGSG waveguide and probe. During the emulation of digital operations shown in Fig 3, each 32-bit, 5 GBit/sec (or 10 GBit/sec) input bitstream and each 32-bit slow, 150 MBit/sec bitstream were continuously cycled. The averaged spectra were recorded over a span of around 4 GHz, ranging from 10 to 14 GHz.

A similar setup was used for the radar target identification task shown in Fig. 4a. MATLAB’s built-in `radarScenario` subroutine from its Radar Toolbox was employed to simulate a 3D environment with moving targets and a radar tower, as shown in Fig. 4a.i. The targets’ trajectories were defined using programmable waypoints. To capture the radar returns, `radarDataGenerator` and `radarTransceiver` subroutines were utilized. These subroutines convert the radar echoes’ RF waveforms to baseband analog voltages over a ‘fast-time’ interval of 0.1 ms, at the radar tower. Sample returns are collected every few millisecond, resulting in a thousand ‘fast-time’ samples per minute of the simulated scenario. An Anritsu MP1763C pulse pattern generator was then used to generate the frequency-modulated square wave with 2 GHz of swept bandwidth, following the scheme depicted in Fig. 4c. This signal was fed to the MNN and the computed spectra are read out on the spectrum analyzers, as before.

For the signal modulation classification task shown in Fig. 4b, the waveforms from the RadioML2016.10a dataset, modulating a 50 MHz carrier, were generated using a Tektronix AFG3102C

Arbitrary Function Generator. This replaced the Anritsu pulse pattern generators, to produce the analog signals required for the task. Each signal had a peak-to-peak amplitude of 2 V and was delivered to the input port feeding the nonlinear waveguide *A*.

All RF data acquisition and parameter sweeps were fully automated using custom code developed in C++, Python and BASIC, and executed from a single workstation.

### 3. Modeling the nonlinear dynamics in the Microwave Neural Network

The integrated MNN circuit may be distilled to its core elements, allowing us to simplify its analysis using generalized Coupled-Mode Theory. Refer to Extended Data Fig. 3 for the steps involved in reducing the circuit to its coupled-mode model. Extended Fig. 3a illustrates the on-chip implementation, which consists of three interconnected linear resonators and one nonlinear waveguide. In each linear resonator, the length of the transmission line comprising a series of sub-segments,  $L_{\text{lin}}$ , is adjustable by leaving open one or more switches ( $[S_1, S_2, \dots, S_6]$ ) and shorting all subsequent switches to ground, i.e., in a thermometer-encoded scheme like [011111] or [011111] to engage only a chosen length of waveguide to complete the return path of the microwave. The nonlinear waveguide, on the other hand, consists of a transmission line embedded with polynomially nonlinear capacitors. Incoming ultra-broadband drive signals are delivered to both types of waveguides via quadrature hybrid couplers. The left coupler directs signals to two linear resonators, while the right coupler channels them to a linear resonator and the nonlinear waveguide. A pair of capacitor banks provides slight tunability to the modes supported by the waveguides. Losses incurred by signals circulating through these passive electromagnetic structures are compensated by saturable gain elements, implemented as cross-coupled transistor pairs, which connect the waveguides on opposite sides of the circuit. We define the gain of saturable gain element as a function of the mode amplitude. For transistors in a common-source configuration, for small signals, the gain is typically  $G_0 = g_m \cdot r_{\text{out}}$ . Here,  $g_m$  is the transconductance of the transistor and  $r_{\text{out}}$  is its output resistance. For large signals, however, this gain compresses as

$$G(v) \approx \frac{G_0}{1 + \left(\frac{v}{v_{\text{sat}}}\right)^2} \quad (1)$$

Here,  $v$  is the amplitude of the signal at the drain of the transistor and  $v_{\text{sat}}$  represents the amplitude at which the gain saturates. Importantly, there exists the time-varying coupling between the circuit's upper and lower halves, established by a pair of slow-bitstream driven switches.

Extended Fig. 3b shows that the linear resonators, each supporting a single natural frequency, can be simplified as LC tank circuits. The symmetry in the lower half allows the capacitor banks to be approximated as two evenly split capacitors ( $C_{\text{CB}}/2$ ), contributing to the capacitance of the tank circuits. Therefore, their natural frequencies are  $\omega_{\text{lin}} = \frac{1}{\sqrt{L_{\text{lin}}(C_{\text{lin}} + C_{\text{CB}}/2)}}$ . However, due to the asymmetry in the upper half, this simplification is not possible.

To simplify experimental permutations of physical circuit parameters, we deliberately kept the linear waveguides (*B*, *C* and *D*) highly detuned from the nominal oscillation frequency of waveguide *A*. Therefore, as shown in Extended Fig. 3c, waveguides *A* and *D* are sufficient to explain the dynamics of the parametrically driven comb. These waveguides interact through a passive phase shift of  $\pi/2$  through the coupler's inductive path ( $\omega L_{\text{coup}}$ ) and a short-circuit through the coupler's capacitance between overlapping turns,  $C_{\text{coup}}$ . The source of regenerative gain through the cross-coupled pairs is retained. Also, we represent the parametrically driven switch as a tunable capacitor, which can be toggled between a very small value (open circuit) and a very large value (short circuit). While the flow of large amplitude microwave signals between oscillators causes rapid variation of the switches' conductivity due to modulation of their gate-source voltages and, subsequently, induces

harmonic distortion, we will exclude such second-order effects in our analysis.

To generalize the operation of this reduced circuit, we represent its nonlinear dynamics through an ensemble of coupled modes - a cascade of nonlinear resonators (in waveguide  $A$ ) connected to a linear resonator ( $D$ ) through a parametrically varied switched coupling and the fixed phase delay via the coupler. This is shown in Extended Fig. 3d. The nonlinear modes have amplitudes  $v_j$  ( $j = 1, 2, \dots, 6$ ) while the linear resonator's mode amplitude is  $v_{\text{lin}}$ . The nonlinear modes interact with each other through couplings  $\beta_{\text{NL}}$  and with the linear mode through the parametric-switch coupling  $\beta_{\text{par}}(t)$  and through the passive phase-shifted coupling,  $\beta_{\text{passive}}e^{\frac{i\pi}{2}}$ . Internal losses, due to finite on-chip Quality-Factors are  $\gamma_{\text{NL},j}^{\text{int}}$  and  $\gamma_{\text{lin}}^{\text{int}}$ . Since the nonlinear resonators' capacitances go as  $C_{\text{NL},j} \approx C_{\text{NL},0} \left(1 - \kappa_1 v_j + \kappa_2 v_j^2 - \kappa_3 v_j^3 \dots\right)$ , their instantaneous resonant frequencies can be Taylor-expanded to be

$$\omega_{\text{NL},j} \approx \omega_{\text{NL},0} \left(1 - \epsilon_1 v_j + \epsilon_2 v_j^2 + \dots\right)^{\frac{1}{2}} \quad (2)$$

In particular, the first nonlinear resonance (with a mode amplitude  $v_1$ ) interacts with the incoming drive (fast gigabit/sec microwave-speed data),  $B_1$ . The saturable gain,  $G$  that re-energizes the system can be assumed to feed directly to this mode as well. Its dynamics can then be written as

$$v_1 = i\beta_{\text{NL}}v_2 + i\beta_{\text{pass}}e^{\frac{i\pi}{2}}v_{\text{lin}} + \beta_{\text{par}}(t)v_{\text{lin}} + i\omega_{\text{NL}}(v_1)v_1 - \gamma_{\text{NL},1}^{\text{int}}v_1 + \frac{G_0}{1 + |v_1|}v_1 + B_1 \quad (3)$$

Nonlinear modes further down the cascade have amplitudes that vary with time as

$$v_j = i(\beta_{\text{NL}}v_{j+1} + \beta_{\text{NL}}v_{j-1}) + i\omega_{\text{NL}}(v_j)v_j - \gamma_{\text{NL},j}^{\text{int}}v_j \quad \text{with} \quad (j \geq 2) \quad (4)$$

The linear mode has dynamics that, when considering the parametric and reciprocal couplings with  $v_1$ , can be described by

$$v_{\text{lin}} = i\beta_{\text{pass}}e^{-\frac{i\pi}{2}}v_1 + i\beta_{\text{par}}(t)v_1 + i\omega_{\text{lin}}v_1 - \gamma_{\text{lin}}^{\text{int}}v_{\text{lin}} \quad (5)$$

These nonlinear dynamics are highly influenced by the initial conditions set by the nonlinear bias voltages (without parametric coupling), as analyzed in Extended Data Fig. 4. When affected by both microwave drives and slow parametric bitstreams, these oscillations exhibit behaviors ranging from simple harmonic to quasi-stable, locally chaotic, and fully chaotic, as shown in the Poincaré maps in Extended Data Fig. 5.

#### 4. Circuit simulation and layout

The CMOS chip, featuring mm-wave and digital interface circuits, was designed and simulated using the Cadence Virtuoso environment, employing transistor models from the GlobalFoundries 45 nm RFSOI, process. Parasitic resistances and capacitances were extracted using Siemens' Calibre tool. The layout of the waveguides, couplers, and interconnecting transmission lines was simulated with the 2.5D EMX electromagnetic tool to accurately model high-frequency performance.

## 5. Backend training for digital emulation

The MNN outputs spectral data that contains information-rich features from the raw input, rather than direct digital outputs. For task-specific outcomes, this information must be extracted from the spectra, aiming to reduce digital computation. We use a linear regression model on the 625 measured frequencies within reduced bandwidths to map these features to final outputs. This linear model cannot capture the nonlinearity typical of neural networks, ensuring the MNN itself performs the complex transformations needed for high accuracy on challenging tasks. The following methods outline task-specific optimizations and evaluations:

- **Linear Search and Conditional Algorithm Emulation:** Dataset of 500 randomly-generated 32-bit sequences per parametric bitstream, split 90/10 for training/validation in 10-fold cross-validation. We use a linear Support Vector Machine from the `sklearn` software package<sup>48</sup> with 5000 max iterations, squared hinge loss, and regularization parameter  $C = 0.02$ , tested on 40 parametric bitstreams.
- **Bit Count:** Similar to Linear Search but with 10,000 max iterations and a hyperparameter sweep on  $C$  from 0.02 to 0.22. Formulated as a 32-class classification task, with labels derived from the Linear Search dataset.
- **Primitive Bit Operations (NAND, XOR, NOR):** A linear model fit via stochastic gradient descent (SGD) with logistic loss and  $l1$  regularization strength of 0.3, on a dataset of 500 randomly-generated 32-bit sequences with 16 fixed tag bits. Using 10-fold cross-validation, the task is a multi-label classification for each output bit, tested on 120 parametric bitstreams.
- **Encoding Classification (RadioML2016.10a):** Using a subset of the RadioML2016.10a<sup>44</sup> dataset, we perform an 80/20 training/validation split. A linear model (single layer) is trained with cross-entropy loss in the PyTorch framework<sup>49</sup>, optimized over 150 epochs with AdamW<sup>50</sup> (learning rate 0.05, weight decay 0.03, minibatch size 128, and decay factor 0.98). Data is augmented with Gaussian noise (standard deviation 0.01) during training, tested on 13 parametric bitstreams.

## 6. MNN evaluation for radar tasks

To predict the targets' flight patterns using the entire set of "fast-time" captures at the radar tower, a digital neural network backend was employed. Each capture provides a spectrum with 2 GHz of bandwidth, resulting in an input shape of  $(L, S)$  for each scene, where  $L=1000$  captures (covering the total time of the scene) and  $S$  is the spectrum size. Unlike previous tasks that used only one spectrum per sample, we utilized a deep ResNet<sup>47</sup> architecture with convolutional layers applied to each frame's spectrum data, with convolutional kernels sweeping across frames over time. Although this backend is more computationally demanding than a linear model, it operates directly on the MNN spectra, allowing the MNN to potentially replace several layers that would otherwise need to process raw, high-bandwidth signals, which are challenging and costly to digitize. Additionally, integrating spectra from different fast-time captures is essential to create a larger context for target movement over time, making a single linear layer insufficient.

The neural network architecture is as follows: the ResNet consists of a stack of blocks, each containing a 2x downsampling pooling layer and a residual branch with two convolutional layers (kernel size 3), with batch normalization, ReLU activation, and dropout regularization between the convolutions. This block structure is repeated 7 times, following an initial input convolution that expands the spectra to a width of 768. After the final residual block, an average pooling layer

outputs features, which are then passed to task-specific two-layer MLPs (one MLP "head" per task, such as target counting or flight pattern detection). The block widths are 256 for the first four blocks and 512 for the final three blocks, including the last residual block.

To train the model, we used a method combining bitstream search with training the backend neural network on experimental outputs to generate the desired classification results. Due to the time-intensive nature of data collection, we passed FM-modulated square-wave signals (as discussed in the main article) through the MNN for 50 flight scenarios across 14 randomly selected parametric bitstreams. The bitstream yielding the highest accuracy on the target counting task was selected, and experimental data for 500 flight scenarios were collected to train the final model, whose performance is reported here.

To train the backend, we applied Cross Entropy loss (for multi-class classification) to classify the number of targets in each scene, binary cross-entropy for flight path detection (as multiple possibilities exist per scene, necessitating binary cross-entropy for multi-label classification), and mean-squared error (MSE) loss for predicting the fastest target's speed. These losses were equally weighted and summed during training. The neural network was optimized using the AdamW optimizer for 5000 epochs with a learning rate of 0.0002, weight decay of 0.2,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and dropout regularization probability of 0.3. We used an exponentially decaying learning rate schedule, reducing the learning rate by a factor of 0.9996 per epoch. Reported predictions and accuracies reflect the average validation set performance across 10 randomly-sampled 80-20 train-validation splits of the full dataset.

Due to the small size of the dataset for training a neural network, we employed data augmentation in order to boost generalization performance. We used the following methods to modify the input spectra during training:

- **Random Shift.** The input sequence of fast-time captures/spectra is shifted forward by a random amount between 1 and 256.
- **Random Bias.** A random amount selected uniformly from [-0.1, -0.075, -0.05, -0.01, 0.001, 0.004, 0.007, 0.008, 0.01, 0.05, 0.075, 0.1, 0.12, 0.15, 0.2] is added to every point on the input spectrum. This represents multiplicative noise (since the spectrum is logarithmically scaled).
- **Random Noise.** A random, normally distributed sample of noise is added to each spectrum. The standard deviation of the noise is selected uniformly at random from [0.001, 0.003, 0.005, 0.007, 0.01, 0.015, 0.02, 0.04, 0.05, 0.07, 0.1, 0.11, 0.14, 0.18].
- **Random Masking.** 1/6th of the input spectra are set to zero. The section is selected at random, and this is done either once or twice with equal probability.

All augmentations are applied with 20% probability on each sample.

## Data and code availability

Measured training data and code for backend training, baseband radar data generation and coupled-mode simulations that support the conclusions in this article are available in the Zenodo repository at <https://zenodo.org/records/14188322>

## References

- [1] P. Wang et al. “Multi-gigabit millimeter wave wireless communications for 5G: from fixed access to cellular networks”. In: *IEEE Communications Magazine* 53.1 (Jan. 2015), pp. 168–178. DOI: 10.1109/MCOM.2015.7010531.
- [2] K. Li et al. “An integrated CMOS–silicon photonics transmitter with a 112 giga-baud transmission and picojoule per bit energy efficiency”. In: *Nat Electron* 6 (2023), pp. 910–921. DOI: 10.1038/s41928-023-01048-1.
- [3] Z. Zhou et al. “Communications with guaranteed bandwidth and low latency using frequency-referenced multiplexing”. In: *Nat Electron* 6 (2023), pp. 694–702. DOI: 10.1038/s41928-023-01022-x.
- [4] A. Li et al. “Evaluating Modern GPU Interconnect: PCIe, NVLink, NV-SLI, NVSwitch and GPUDirect”. In: *IEEE Transactions on Parallel and Distributed Systems* 31.1 (Jan. 2020), pp. 94–110. DOI: 10.1109/TPDS.2019.2928289.
- [5] W. Gomes et al. “Ponte Vecchio: A Multi-Tile 3D Stacked Processor for Exascale Computing”. In: *2022 IEEE International Solid-State Circuits Conference (ISSCC)*. San Francisco, CA, USA, 2022, pp. 42–44. DOI: 10.1109/ISSCC42614.2022.9731673.
- [6] I. Gresham et al. “Ultra-wideband radar sensors for short-range vehicular applications”. In: *IEEE Transactions on Microwave Theory and Techniques* 52.9 (Sept. 2004), pp. 2105–2122. DOI: 10.1109/TMTT.2004.834185.
- [7] S. Skaria et al. “Deep-Learning Methods for Hand-Gesture Recognition Using Ultra-Wideband Radar”. In: *IEEE Access* 8 (2020), pp. 203580–203590. DOI: 10.1109/ACCESS.2020.3037062.
- [8] H. Esmaeilzadeh et al. “Dark silicon and the end of multicore scaling”. In: *2011 38th Annual International Symposium on Computer Architecture (ISCA)*. IEEE. San Jose, CA, USA, 2011, pp. 365–376.
- [9] U. Gupta et al. “Chasing Carbon: The Elusive Environmental Footprint of Computing”. In: *2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*. IEEE. Seoul, Korea (South), 2021, pp. 854–867. DOI: 10.1109/HPCA51647.2021.00076.
- [10] R. Desislavov et al. “Trends in AI inference energy consumption: Beyond the performance-vs-parameter laws of deep learning”. In: *Sustainable Computing: Informatics and Systems* 38 (2023), p. 100857.
- [11] K. Rupp. *Microprocessor Trend Data*. Accessed: October 17, 2024. 2021. URL: <https://github.com/karlrupp/microprocessor-trend-data/tree/master/50yrs>.
- [12] L. G. Wright et al. “Deep physical neural networks trained with backpropagation”. In: *Nature* 601 (2022), pp. 549–555. DOI: 10.1038/s41586-021-04223-6.
- [13] A. Sludds et al. “Delocalized photonic deep learning on the internet’s edge”. In: *Science* 378 (2022), pp. 270–276. DOI: 10.1126/science.abq8271.

[14] J. Zhang et al. “Radio Frequency Fingerprint Identification for Narrowband Systems, Modelling and Classification”. In: *IEEE Transactions on Information Forensics and Security* 16 (2021), pp. 3974–3987. DOI: 10.1109/TIFS.2021.3088008.

[15] H. Cai et al. “Brain organoid reservoir computing for artificial intelligence”. In: *Nature Electronics* 6 (2023), pp. 1032–1039. DOI: 10.1038/s41928-023-01069-w.

[16] A. Senanian et al. “Microwave signal processing using an analog quantum reservoir computer”. In: *Nature Communications* 15 (2024), p. 7490. DOI: 10.1038/s41467-024-06320-7.

[17] B. Dong et al. “Higher-dimensional processing using a photonic tensor core with continuous-time data”. In: *Nature Photonics* 17 (2023), pp. 1080–1088. DOI: 10.1038/s41566-023-01313-x.

[18] H. Feng et al. “Integrated lithium niobate microwave photonic processing engine”. In: *Nature* 627 (2024), pp. 80–87. DOI: 10.1038/s41586-024-07078-9.

[19] R. Davis III et al. *Frequency-Encoded Deep Learning with Speed-of-Light Dominated Latency*. Preprint at <https://arxiv.org/abs/2207.06883>. 2022.

[20] K. A. Clark et al. “Synchronous subnanosecond clock and data recovery for optically switched data centres using clock phase caching”. In: *Nature Electronics* 3 (2020), pp. 426–433. DOI: 10.1038/s41928-020-0423-y.

[21] W. Zhang et al. “Edge learning using a fully integrated neuro-inspired memristor chip”. In: *Science* 381 (2023), pp. 1205–1211. DOI: 10.1126/science.ade3483.

[22] Y. Zhou et al. “Computational event-driven vision sensors for in-sensor spiking neural networks”. In: *Nature Electronics* 6 (2023), pp. 870–878. DOI: 10.1038/s41928-023-01010-w.

[23] X. Gao et al. “Programmable surface plasmonic neural networks for microwave detection and processing”. In: *Nature Electronics* 6 (2023), pp. 319–328. DOI: 10.1038/s41928-023-00944-7.

[24] A. Ross et al. “Multilayer spintronic neural networks with radiofrequency connections”. In: *Nature Nanotechnology* 18 (2023), pp. 1273–1280. DOI: 10.1038/s41565-023-01452-w.

[25] M. Zhu et al. “A reconfigurable linear RF analog processor for realizing microwave artificial neural network”. In: *IEEE Transactions on Microwave Theory and Techniques* (2023).

[26] Y. Xu et al. “Bidirectional interconversion of microwave and light with thin-film lithium niobate”. In: *Nature Communications* 12 (2021), p. 4453. DOI: 10.1038/s41467-021-24809-y.

[27] W. Moy et al. “A 1,968-node coupled ring oscillator circuit for combinatorial optimization problem solving”. In: *Nature Electronics* 5 (2022), pp. 310–317. DOI: 10.1038/s41928-022-00742-8.

[28] S. Dutta et al. “An Ising Hamiltonian solver based on coupled stochastic phase-transition nano-oscillators”. In: *Nature Electronics* 4 (2021), pp. 502–512. DOI: 10.1038/s41928-021-00602-6.

[29] N. Jouppi et al. “In-Datacenter Performance Analysis of a Tensor Processing Unit”. In: *Proceedings of the 44th Annual International Symposium on Computer Architecture (ISCA)*. Association for Computing Machinery, 2017, pp. 1–12. DOI: 10.1145/3079856.3080246.

[30] D. L. Sounas et al. “Broadband passive isolators based on coupled nonlinear resonances”. In: *Nature Electronics* 1 (2018), pp. 113–119. DOI: 10.1038/s41928-018-0025-0.

[31] B. Govind et al. “Ultra-compact quasi-true time delay for boosting wireless channel capacity”. In: *Nature* 627 (2024), pp. 88–94. DOI: 10.1038/s41586-024-07075-y.

[32] B. Razavi. “The Cross-Coupled Pair - Part I [A Circuit for All Seasons]”. In: *IEEE Solid-State Circuits Magazine* 6.3 (Summer 2014), pp. 7–10. DOI: 10.1109/MSSC.2014.2329234.

[33] M. Babaie et al. “A Class-F CMOS Oscillator”. In: *IEEE Journal of Solid-State Circuits* 48.12 (Dec. 2013), pp. 3120–3133. DOI: 10.1109/JSSC.2013.2273823.

[34] B. Jamali et al. “A Fully Integrated 50–280-GHz Frequency Comb Detector for Coherent Broadband Sensing”. In: *IEEE Transactions on Terahertz Science and Technology* 9.6 (Nov. 2019), pp. 613–623. DOI: 10.1109/THZ.2019.2944129.

[35] H. M. E. Hussein et al. “Passive frequency comb generation at radiofrequency for ranging applications”. In: *Nature Communications* 15 (2024), p. 2844. DOI: 10.1038/s41467-024-02844-1.

[36] H. Shu et al. “Microcomb-driven silicon photonic systems”. In: *Nature* 605 (2022), pp. 457–463. DOI: 10.1038/s41586-022-04579-3.

[37] M. Zhang et al. “Broadband electro-optic frequency comb generation in a lithium niobate microring resonator”. In: *Nature* 568 (2019), pp. 373–377. DOI: 10.1038/s41586-019-1008-7.

[38] H. A. Haus et al. “Coupled-mode theory”. In: *Proceedings of the IEEE* 79.10 (Oct. 1991), pp. 1505–1518. DOI: 10.1109/5.104225.

[39] U. Pereira et al. “Attractor dynamics in networks with learning rules inferred from in vivo data”. In: *Neuron* 99.1 (2018), 227–238.e4. DOI: 10.1016/j.neuron.2018.05.038.

[40] T. S. Parker et al. *Practical Numerical Algorithms for Chaotic Systems*. New York: Springer, 1989. ISBN: 978-0387968366.

[41] Y. Frans et al. “A 56-Gb/s PAM4 Wireline Transceiver Using a 32-Way Time-Interleaved SAR ADC in 16-nm FinFET”. In: *IEEE Journal of Solid-State Circuits* 52.4 (Apr. 2017), pp. 1101–1110.

[42] O. E. Agazzi et al. “A 90 nm CMOS DSP MLSD Transceiver With Integrated AFE for Electronic Dispersion Compensation of Multimode Optical Fibers at 10 Gb/s”. In: *IEEE Journal of Solid-State Circuits* 43.12 (Dec. 2008), pp. 2939–2957. DOI: 10.1109/JSSC.2008.2006232.

- [43] J. Al-Eryani et al. “Fully Integrated Single-Chip 305–375-GHz Transceiver With On-Chip Antennas in SiGe BiCMOS”. In: *IEEE Transactions on Terahertz Science and Technology* 8.3 (May 2018), pp. 329–339. DOI: [10.1109/TTHZ.2018.2823202](https://doi.org/10.1109/TTHZ.2018.2823202).
- [44] Y. Lin et al. “An Improved Neural Network Pruning Technology for Automatic Modulation Classification in Edge Devices”. In: *IEEE Transactions on Vehicular Technology* 69.5 (May 2020), pp. 5703–5706. DOI: [10.1109/TVT.2020.2983143](https://doi.org/10.1109/TVT.2020.2983143).
- [45] Y. Wang et al. “MsmcNet: A Modular Few-Shot Learning Framework for Signal Modulation Classification”. In: *IEEE Transactions on Signal Processing* 70 (2022), pp. 3789–3801. DOI: [10.1109/TSP.2022.3191783](https://doi.org/10.1109/TSP.2022.3191783).
- [46] H. Zhang et al. “A Data Preprocessing Method for Automatic Modulation Classification Based on CNN”. In: *IEEE Communications Letters* 25.4 (Apr. 2021), pp. 1206–1210. DOI: [10.1109/LCOMM.2020.3044755](https://doi.org/10.1109/LCOMM.2020.3044755).
- [47] K. He et al. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [48] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [49] A. Paszke et al. “PyTorch: an imperative style, high-performance deep learning library”. In: *Advances in Neural Information Processing Systems 32 (NeurIPS)*. 2019, pp. 8024–8035. URL: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- [50] I. Loshchilov et al. “Decoupled Weight Decay Regularization”. In: 2019. URL: <https://openreview.net/forum?id=Bkg6RiCqY7>.

## Acknowledgements

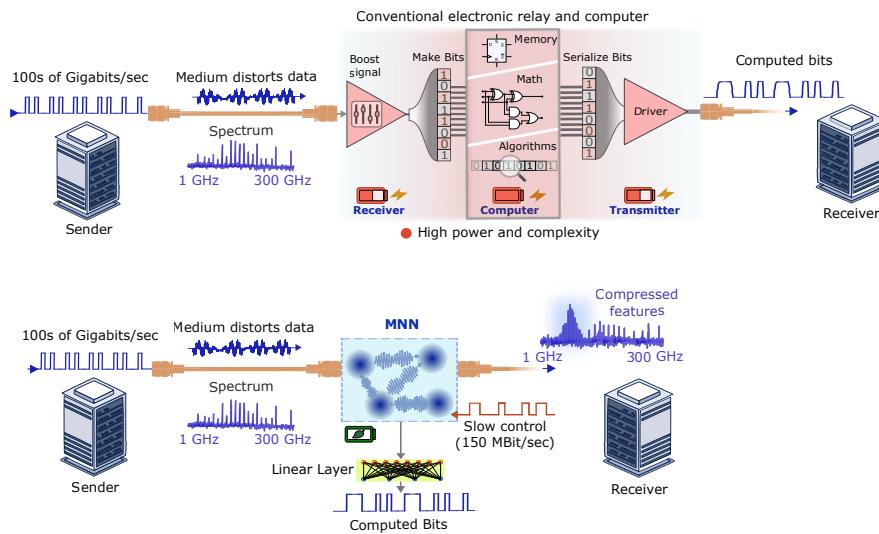
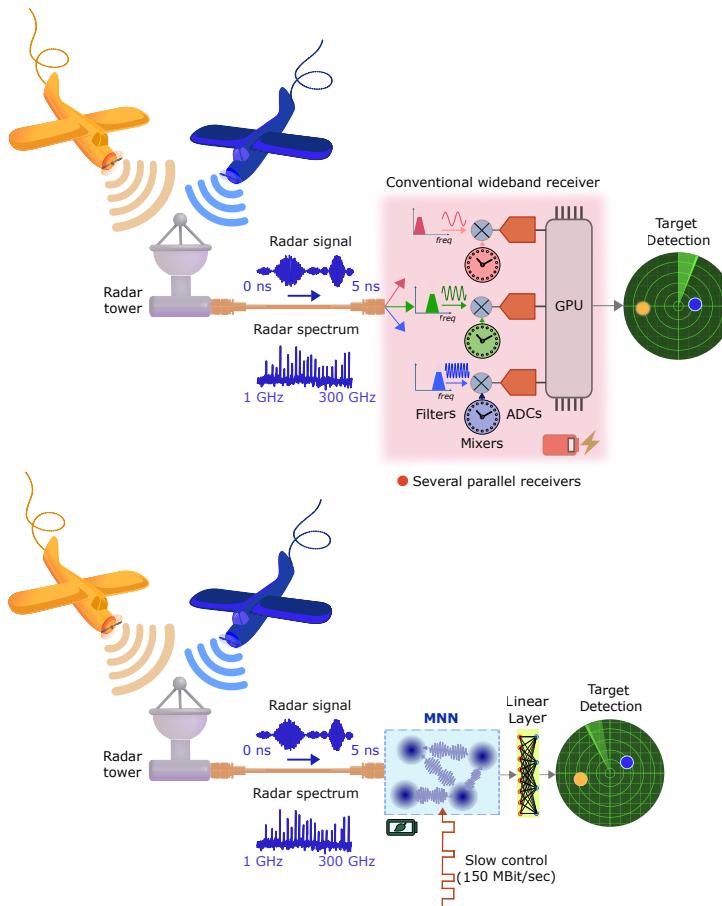
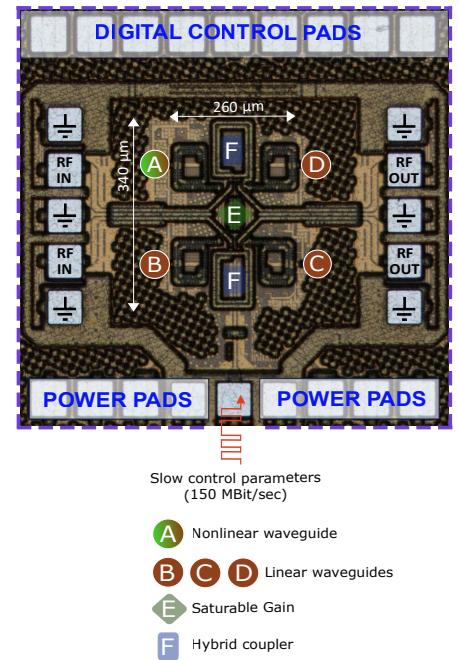
The authors acknowledge the Defense Advanced Research Projects Agency (DARPA), through its Wideband Adaptive RF Protection (WARP) program, and GlobalFoundries for providing chip fabrication. They thank the Cornell NanoScale Facility, a member of the National Nanotechnology Coordinated Infrastructure (NNCI), which is supported by the National Science Foundation (Grant NNCI-2025233) and where the work was done in part. They also thank A. Senanian, S. Prabhu, and V. Kremenski for discussions about physical neural networks. B.G thanks A. Aksaray for assistance with PCB design and S. Huang, T. Tapen and F. Monticone for theoretical discussions. M.G.A. and P.L.M. additionally acknowledge funding from the National Science Foundation (Award CBET-2123862). F.O.W. acknowledges support from the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a program of Schmidt Sciences, LLC.

## Competing interests

The authors have filed a provisional US patent application based on the neural processor presented in this article.

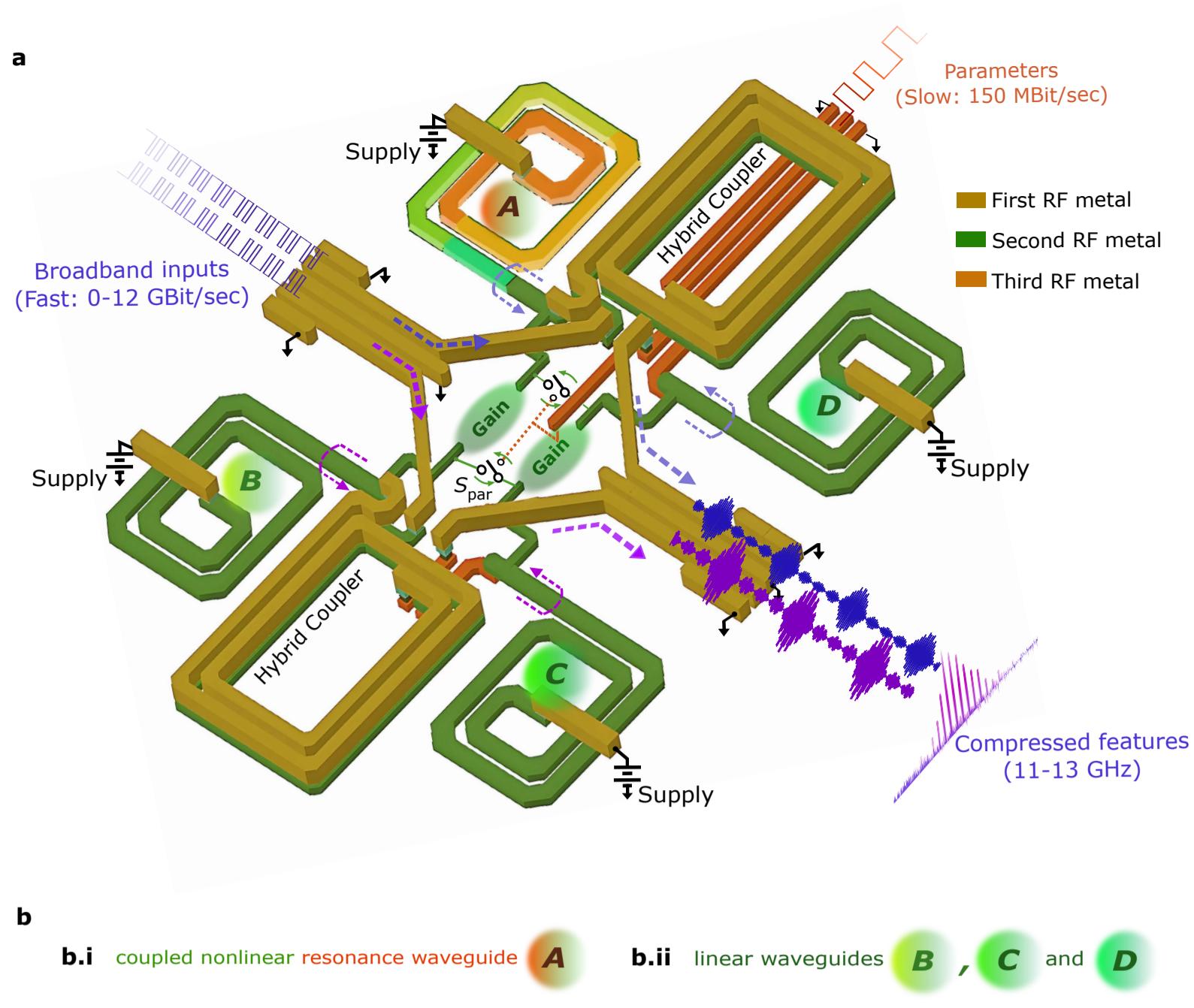
### **Author Contributions**

B.G. and M.G.A contributed equally. B.G. designed the CMOS chip and was involved in all aspects of the experiment. M.G.A. automated the experimental setup and performed the machine-learning experiments and data analysis. B.G. and F.O.W. formulated the coupled model of the frequency-modulated parametrically driven oscillators, with F.O.W. performing numerical simulations. P.L.M., B.G. and A.A. conceived the project. A.A. and P.L.M. supervised the project. B.G. and A.A. wrote the manuscript with input from all the authors.

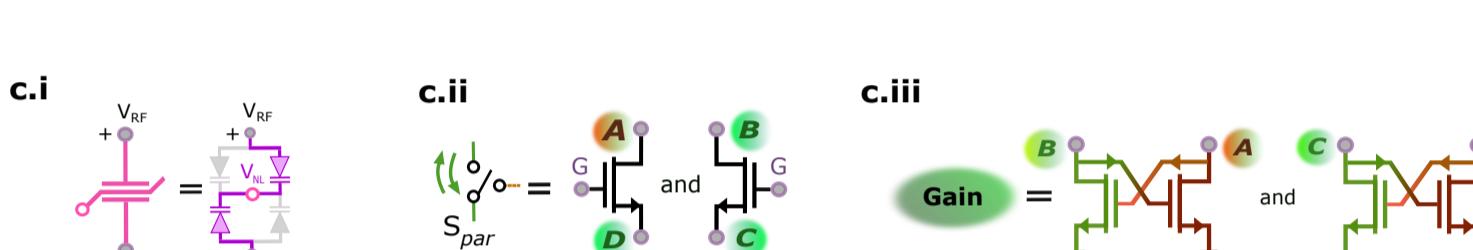
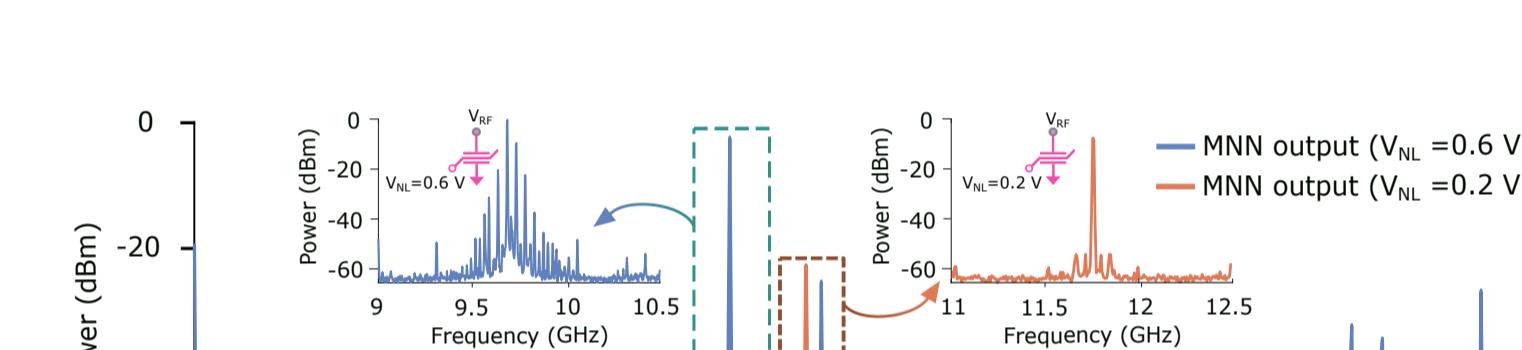
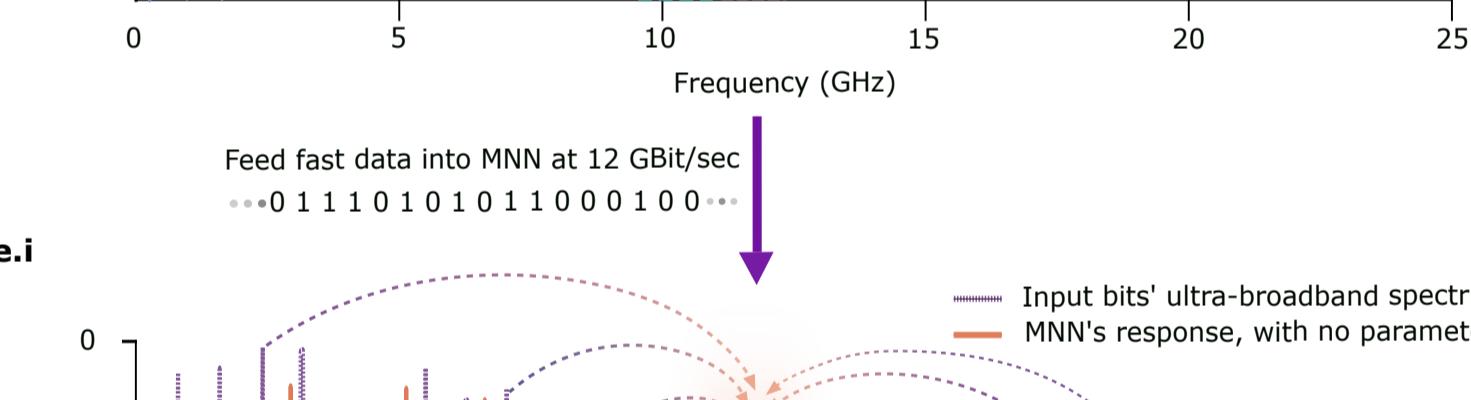
**a****b****c**

**Figure 1. A solution to ultra-wideband computation and communication signal processing problems.**

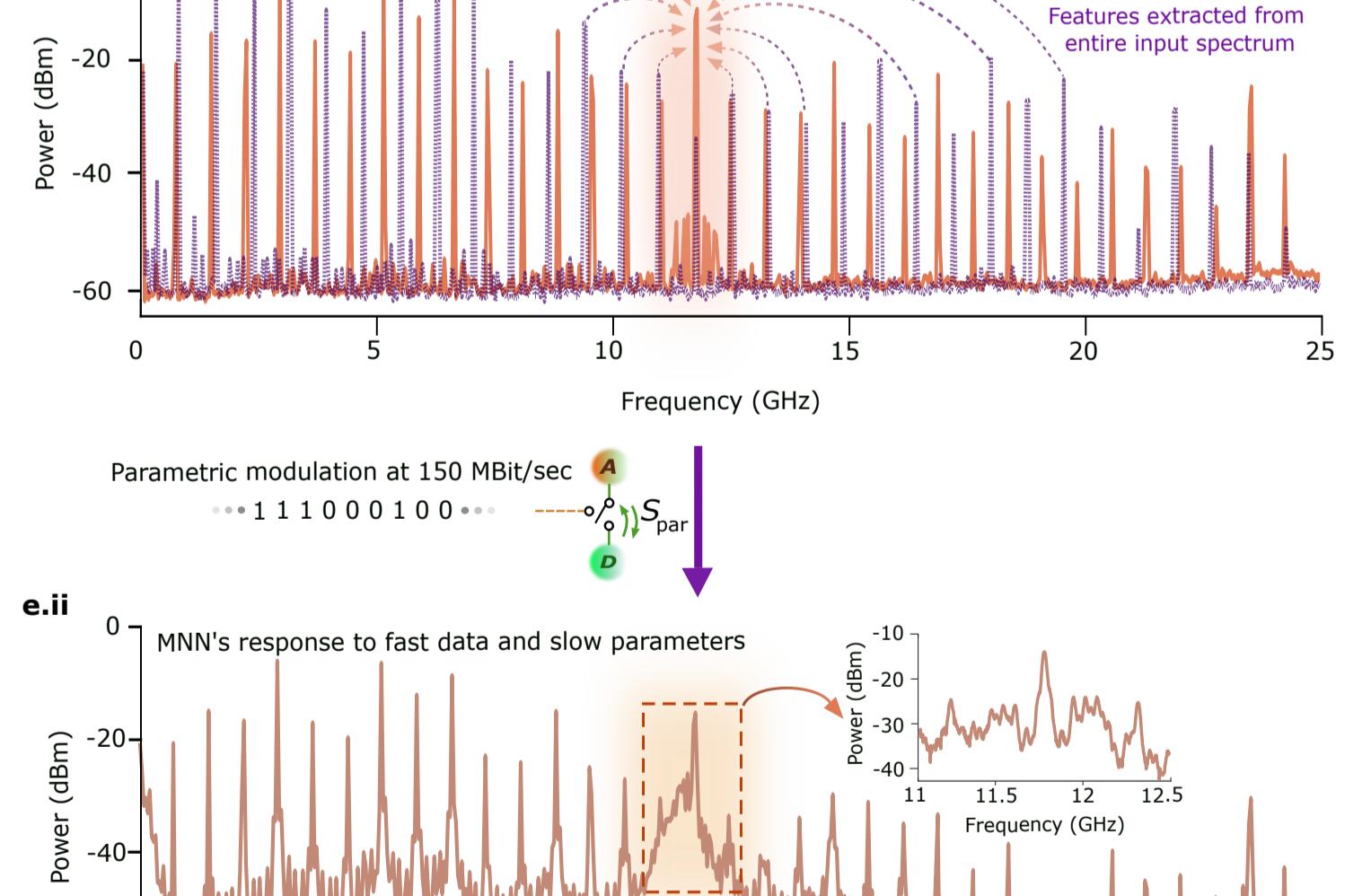
**a**, High data rate electronics transmit bits that get distorted over long distances due to lossy, dispersive channels, such as those in data centers. Complex amplifiers (equalizers) and clock/data recovery circuits process and restore signal integrity in the time-domain. The data is then stored in memory, digitally processed, and serialized for further transmission, through more lossy channels, to the client. This signal chain has historically been bottlenecked by clock speed and consumes several kilowatts of power. In contrast, the integrated Microwave Neural Network (MNN) consumes under two hundred milliwatts and processes data at microwave speed in the frequency domain, disregarding the shape of distorted bits. It computes upon and manifests the features of the input spectrum in a compressed comb of frequencies spanning a few gigahertz. Its computational functionality over tens of gigahertz can be dynamically reprogrammed with parameters fed in at only a few Megabits/second, disregarding the equivalent digital circuit's complexity. **b**, Broadband radar typically employs several parallel circuits to process narrow bands using filters, local oscillators, mixers, and analog-to-digital data converters (ADCs). Target positions are deciphered via digital signal processing on graphics processing units (GPUs). The MNN, on the other hand, learns the features of incoming radar returns through inherent nonlinearities, drastically reducing complexity. The resulting spectra are fed to a linear layer that predicts flight patterns of the targets. **c**, The integrated MNN is fabricated in a 45 nm RF CMOS process, occupying a sub-wavelength footprint of  $0.088 \text{ mm}^2$ . It comprises an ensemble of linear and nonlinear waveguides, connected by gain elements and power-splitting couplers. The system is sensitive to ultra-fast incoming bits or radar signals. Its parametric oscillations are reprogrammed through slow control bitstreams in the megahertz range, which drive switches installed on lower metal layers (not visible in the photograph).

**a****b**

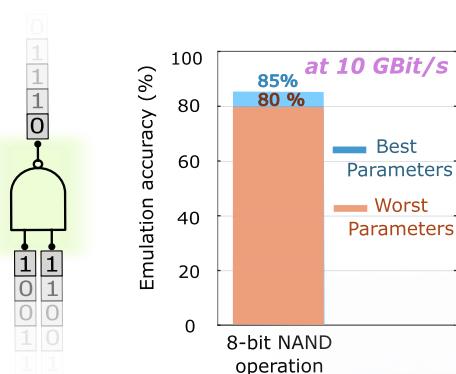
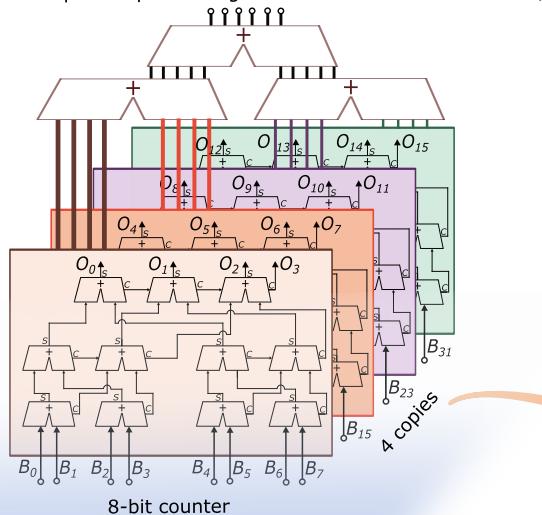
**b.i** coupled nonlinear resonance waveguide **A**      **b.ii** linear waveguides **B**, **C** and **D**

**c****d****e**

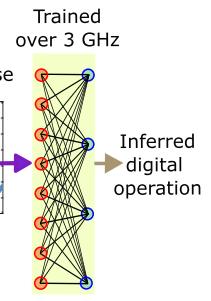
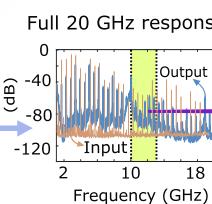
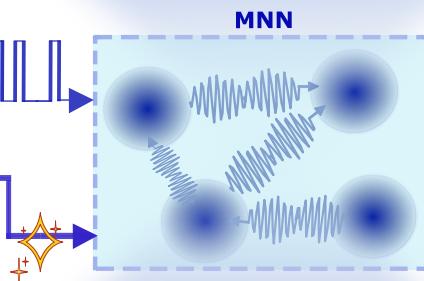
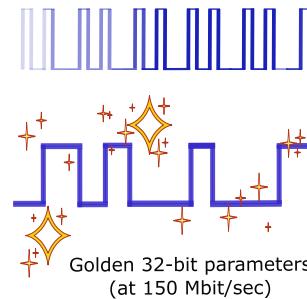
Feed fast data into MNN at 12 GBit/sec  
...0 1 1 1 0 1 0 1 0 1 1 0 0 0 1 0 0 ...



**Figure 2. Distributed microwave oscillators form an ultra-broadband neural network.** **a**, The MNN features CMOS microwave waveguide resonators, both linear (B,C and D) and nonlinear (A), coupled by multiple mechanisms: (i) slow parametric switching through  $S_{par}$ , (ii) saturable gain, and (iii) distributed nonlinear capacitances in resonator D. Ultra-broadband analog signals or digital bitstreams enter through a Ground-Signal-Ground-Signal-Ground (GSGSG) waveguide, are split by hybrid couplers, and fed to the resonators. The resonators' modes are read through another GSGSG waveguide. **b.i**, The nonlinear waveguide has coupled resonances formed of fixed inductive segments interspersed with tunable nonlinear capacitors. **b.ii**, A linear, tunable transmission line, with a variable-length return path that supports various frequencies is used for resonators B,C and D. **c**, CMOS Silicon-on-Insulator devices used in the circuit. **c.i**, Antiparallel diodes effecting polynomially nonlinear capacitance. **c.ii**, Parametric switches between pairs of resonators that shape the inherent comb-like response of the MNN. **c.iii**, Cross-coupled transistor pairs that provide regenerative, saturable gain. **d**, The measured comb-like spectra generated by the MNN for both low and high bias voltages applied to the nonlinear capacitors, in the absence of external drives or parametric switching. **e**, The impact of low-speed switching between the nonlinear and linear resonators on the MNN's output spectrum. **e.i**, The Fourier transform of the cyclically fed 12 GBit/sec digital bitstream (in purple), sweeping across the microwave spectrum, and the MNN's measured output spectrum when this bitstream is fed to its ports (in orange). Here, the input data here is [1110101011000100]. **e.ii**, The distortion in the MNN's spectrum is a result of frequency-modulated parametric oscillations which are simultaneously influenced by the incoming high-speed digital data at 12 GBit/sec, and the slow parametric bitstream, cyclically fed at 150 MBit/sec. Here, the parametric bitstream is [1110001000100010000110010111010].

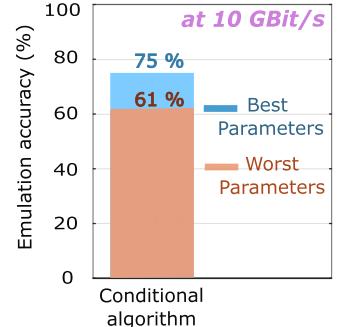
**b****b.i** 8-bit bitwise NAND**b.ii** Complex sequential logic - count of '1's in the 32-bit input**a**

Fast data at 10 GBit/sec



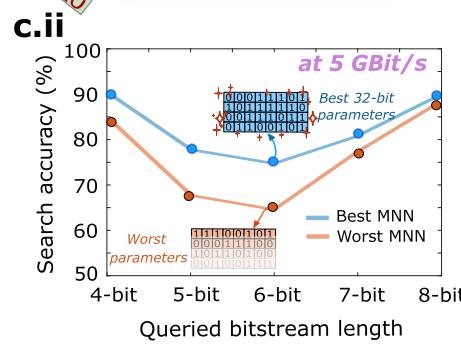
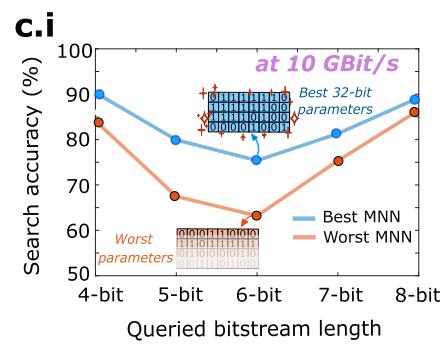
**d Emulation of conditional algorithms at Gigahertz-speed**

```
if count < 16
  result = Linear_Search([1, 0, 1, 0, 1])
else
  result = Linear_Search([1, 1, 0, 0, 1])
```

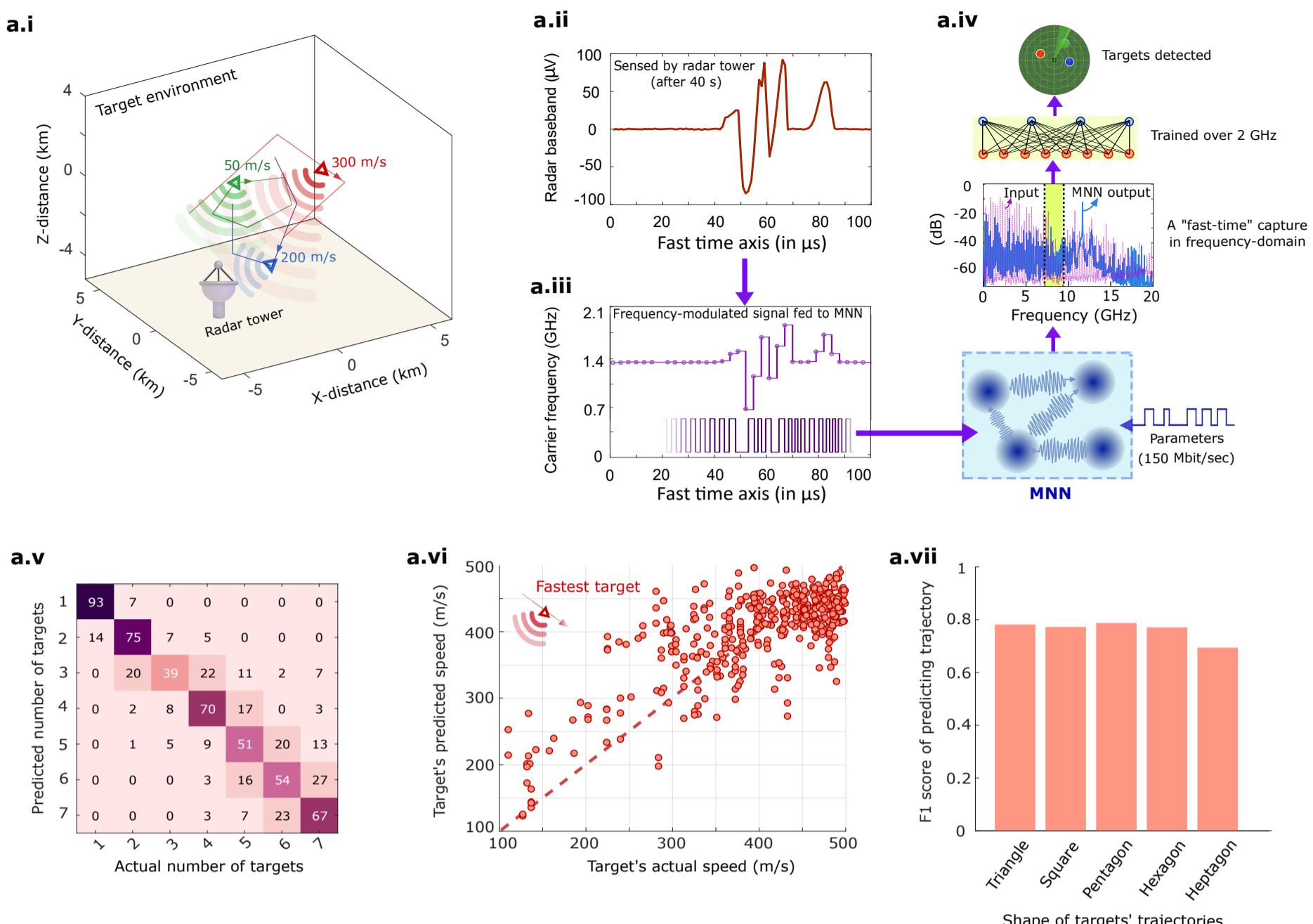
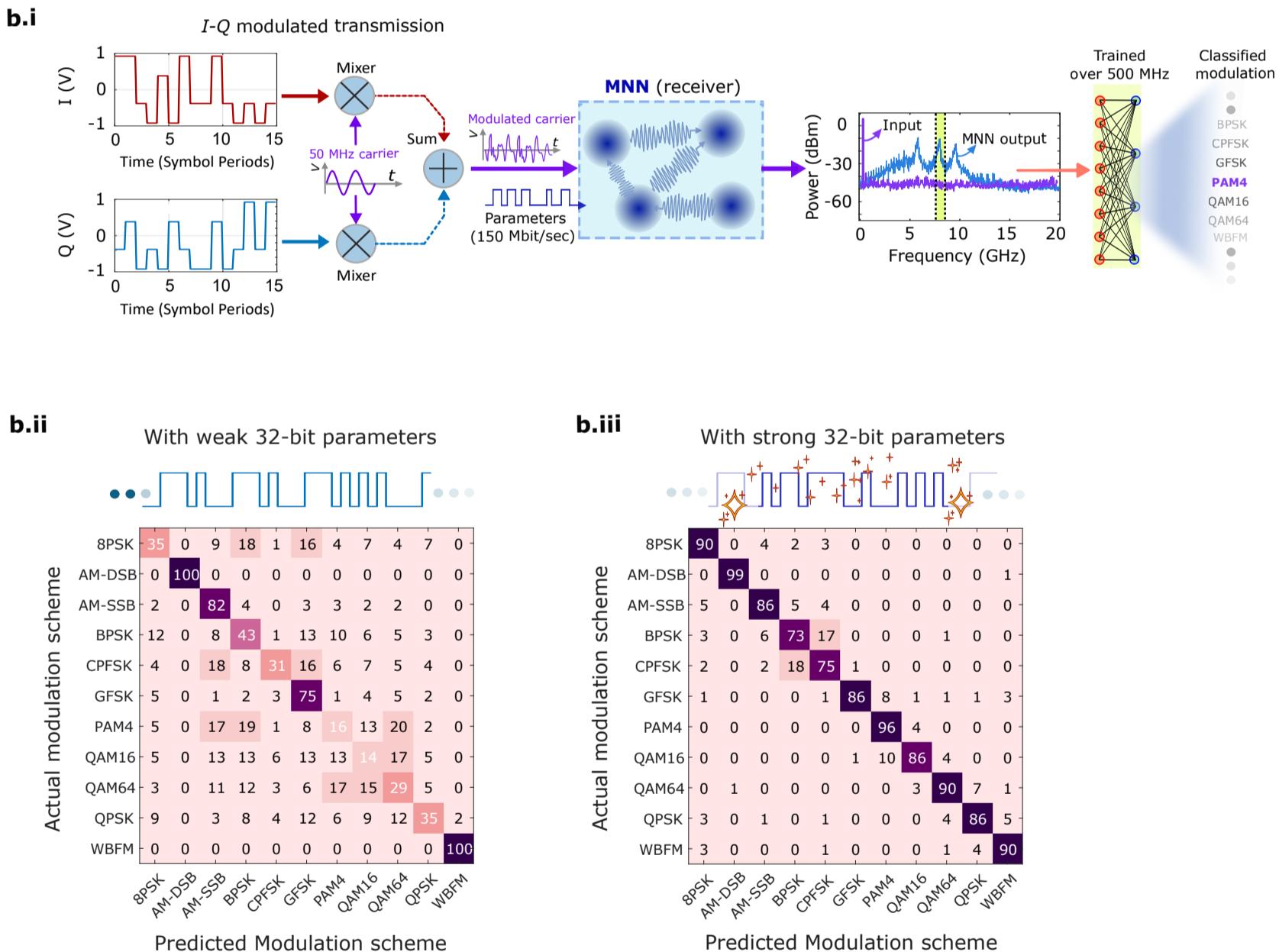
**c Emulation of linear bit search at Gigahertz speed**

...010110011101010101...

11010

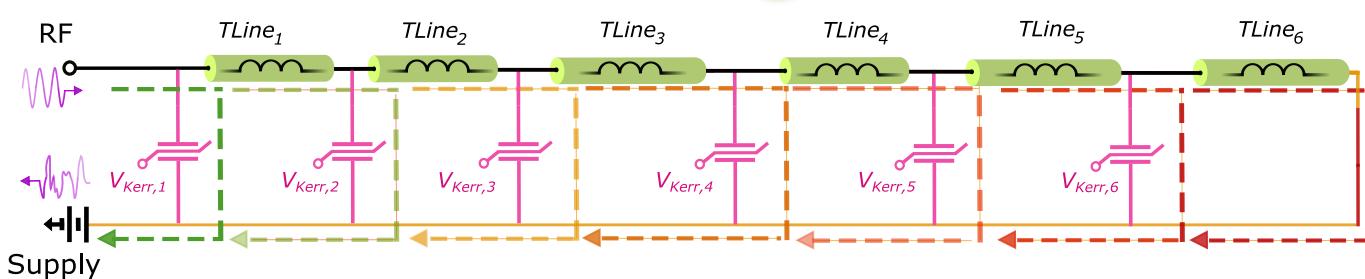
**Figure 3. CMOS Integrated Microwave Neural Network mimics an ultra-high-speed digital computer without the need for a clock.**

**a**, 32-bit sequences are fed into the MNN at 10 GBit/sec and 5 GBit/sec, cycling to produce a periodic output. The output (shown as a Fourier Transform), produced by a spectrum analyzer, focuses on a compressed frequency band between 11 and 14 GHz. In post-processing, this compressed spectrum is used to train a linear layer for classification and validation of the mimicked digital operation. **b.i**, Simple circuits like primitive gate operations and **b.ii**, complex, time-consuming sequential logic, such as population counters for counting 1's in 32-bit data, can be emulated instantaneously without clock bottlenecks. Both achieve approximately equal accuracy, when using well-chosen parametric bitstreams. **c**, The MNN, through mode-coupling in the frequency domain, also aids in emulating high-level computational tasks, such as detecting bit patterns in high-speed data streams at tens of gigahertz. For both 5 GBit/sec (**c.i**) and 10 GBit/sec data streams (**c.ii**), it provides high search accuracy regardless of the length of the queried bit pattern (ranging from 4 to 8 bits) in a 32-bit input data. This scheme offers an alternative to the Maximum Likelihood Sequence Estimation (MLSE) scheme used in wireline modems. Unlike MLSE, it does not require Serializer-Deserializer receivers or backend analog-to-digital converters and digital signal processing, which rely on complex time-domain signal processing circuits. **d**, The emulated bit-search and population-counting behaviors can be combined to execute larger algorithms. Digital computers would need several clock cycles to process these sequential portions, store data, and handle conditional statements at a couple of gigahertz with several watts of power consumption. In contrast, the MNN operates at tens of gigahertz while maintaining sub-200 milliwatt power consumption.

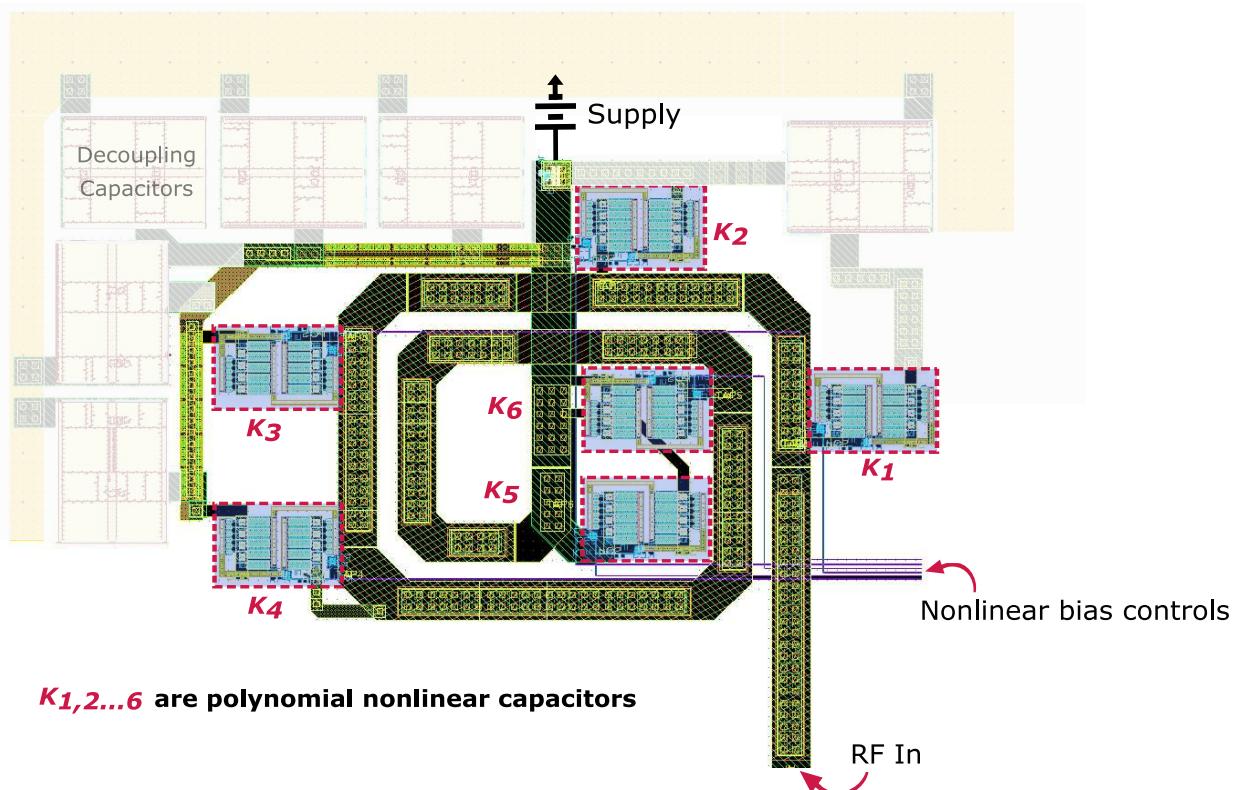
**a****b**

**Figure 4. Simplified radar signal processing and communication decoding through the integrated MNN.** **a.**, The MNN is repurposed to process radar waveforms for detecting the trajectories of dynamic targets. **a.i**, A simulated 3D environment for target detection, illustrating multiple aircraft moving along distinct polygonal flight paths. The radar tower scans the airspace and receives reflected RF signals from the targets. **a.ii**, An example radar baseband signal at the 40-second mark shows characteristic fluctuations due to targets' reflections. **a.iii**, The analog voltage is converted into a frequency modulated square-wave signal, which is processed by the MNN. **a.iv**, The output spectrum contains several narrower comb-like spectra resulting from the nonlinear interactions between the frequency-modulated drive and the parameterized oscillations. These complex features are read out over a 2 GHz bandwidth and utilized by the trained backend neural network to decipher the targets' flight patterns. **a.v**, The MNN is able to infer the number of targets in each flight scenario **a.vi**, The MNN is capable of honing in on a single target, specifically the fastest one. It closely tracks the speed of this target based solely on minute changes in carrier frequency, for each flight scenario in the validation set. **a.vii**, The system can also identify the presence of polygonal trajectories with confidence, using the validation set comprising 450 flight scenarios. Classification is evaluated using the F1-score, the harmonic mean of precision (true positives divided by the sum of true positives and false positives) and recall (true positives divided by the sum of true positives and false negatives). **b.**, Modulation Scheme Classification using the integrated MNN. **b.i**, Schematic of I-Q modulated transmission. **b.ii**, Confusion matrix for modulation classification with weak 32-bit parameters. **b.iii**, Confusion matrix for classification with strong 32-bit parameters. Within the RadioML 2016.10A RF signal classification benchmarking dataset, the following are digital modulation schemes: 8PSK (8-Phase Shift Keying), BPSK (Binary Phase Shift Keying), CPFSK (Continuous Phase Frequency Shift Keying), GFSK (Gaussian Frequency Shift Keying), PAM4 (Pulse Amplitude Modulation, 4-level), QAM16 (Quadrature Amplitude Modulation, 16-level), QAM64 (Quadrature Amplitude Modulation, 64-level) and QPSK (Quadrature Phase Shift Keying). The following are analog modulation schemes: AM-DSB (Amplitude Modulation - Double Sideband), AM-SSB (Amplitude Modulation - Single Sideband) and WBFM (Wideband Frequency Modulation).

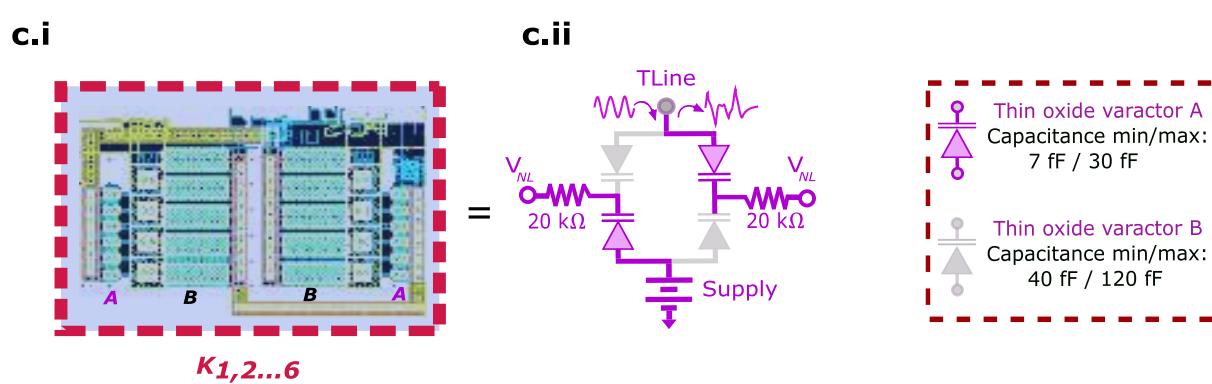
### a Tunable, distributed nonlinear resonator



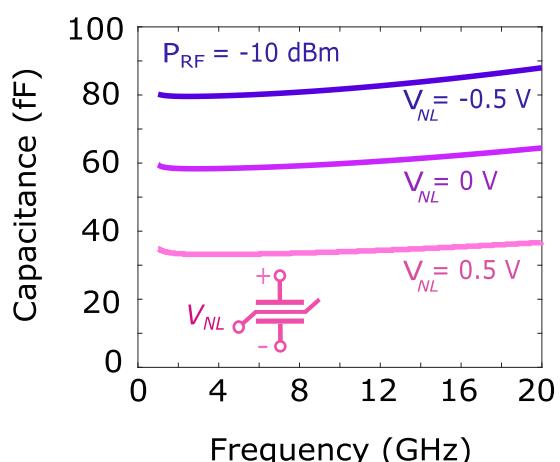
### b Top View of Layout



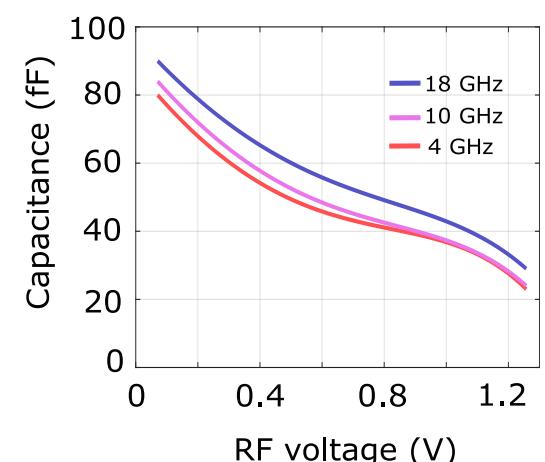
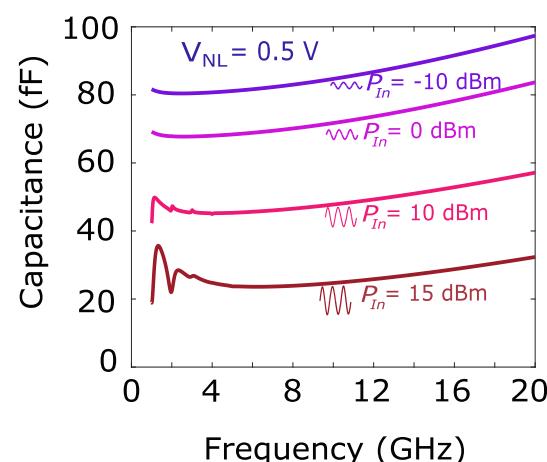
### c Polynomial nonlinear capacitor



#### c.iii

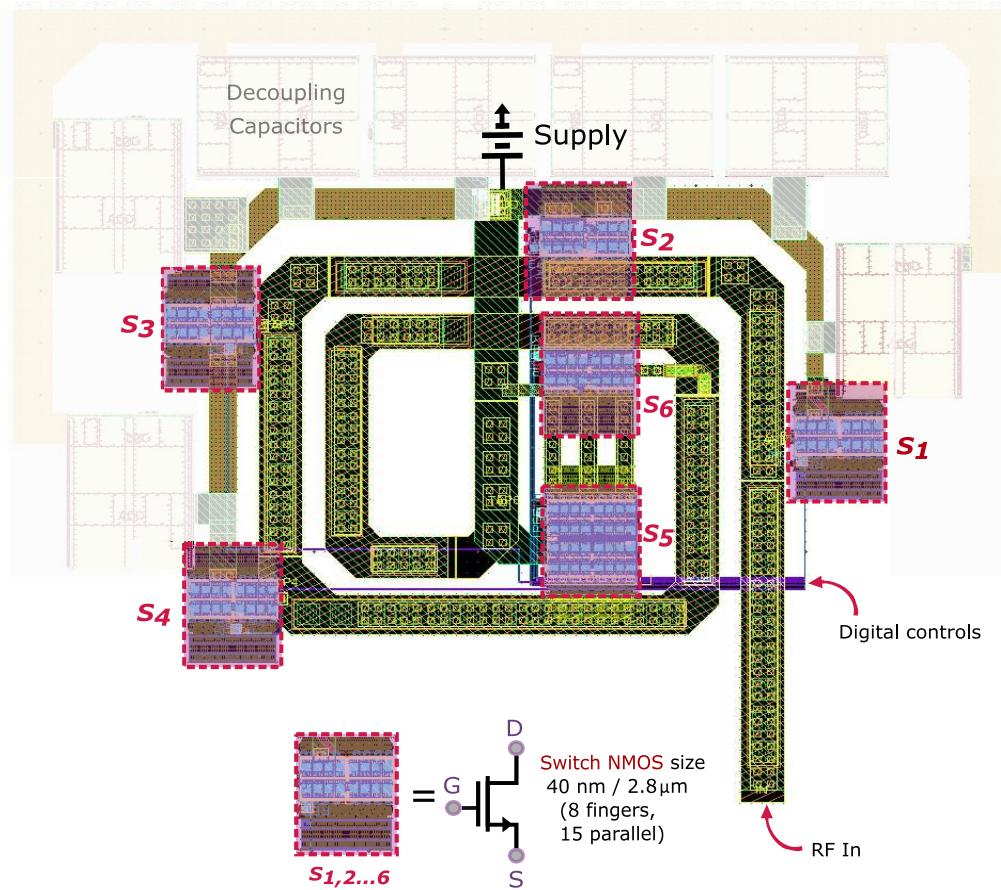


#### c.iv

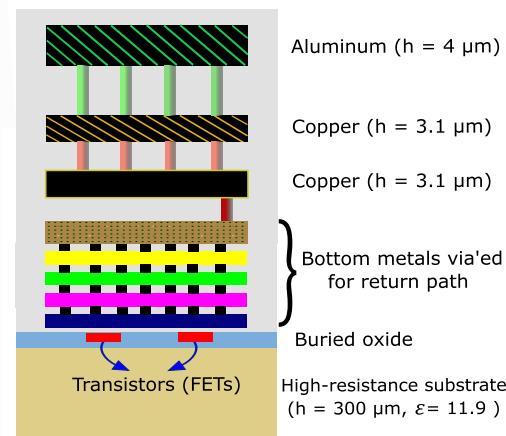


**Extended Data Fig 1 | Structure of the distributed nonlinear waveguide resonator.** **a**, The nonlinear transmission line comprises cascaded  $\pi$  sections of fixed inductive waveguide segments and drive-sensitive nonlinear capacitors. Each resonance absorbs part of the incoming microwave signal and transmits the rest to subsequent segments. The overall nonlinear response results from cumulative reflections back to  $\text{RF}_{\text{in}}$  port. **b**, Top view of the layout of the nonlinear waveguide. Nonlinear capacitors are inserted periodically along the pretzel-shaped trace. Their nominal bias points are set through analog voltages. **c**, In the Silicon-on-Insulator process used here, a single polynomial nonlinear capacitor consists of two pairs of antiparallel diodes. **c.i**, Compact layout of this component. **c.ii** The schematic, wherein the diode pairs are biased at their mid-point. **c.iii**, SPICE / Spectre-model simulated characterization of the nonlinear capacitor across variation in bias voltages and input RF power shows that by injecting a constant RF input power of -10 dBm, the effective capacitance decreases with increasing bias voltage. **c.iv**, The effective capacitance is highly sensitive to the input RF drive's power. The capacitance reduces with increasing field strength. **c.v**, The same nonlinearity, when considering its variance for different frequencies. Here, the capacitance varies with RF voltage as  $C_{\text{eff}}(V_{\text{NL}}, V_{\text{in}}) = a - b V_{\text{in}} + c V_{\text{in}}^2 - d V_{\text{in}}^3 + \dots$  fF which is well-suited to generating expansive functions for a neural pre-processor.

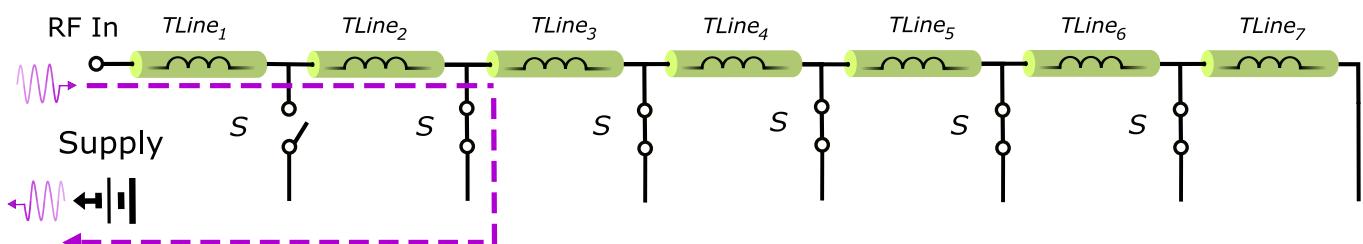
### a Top view of waveguide layout



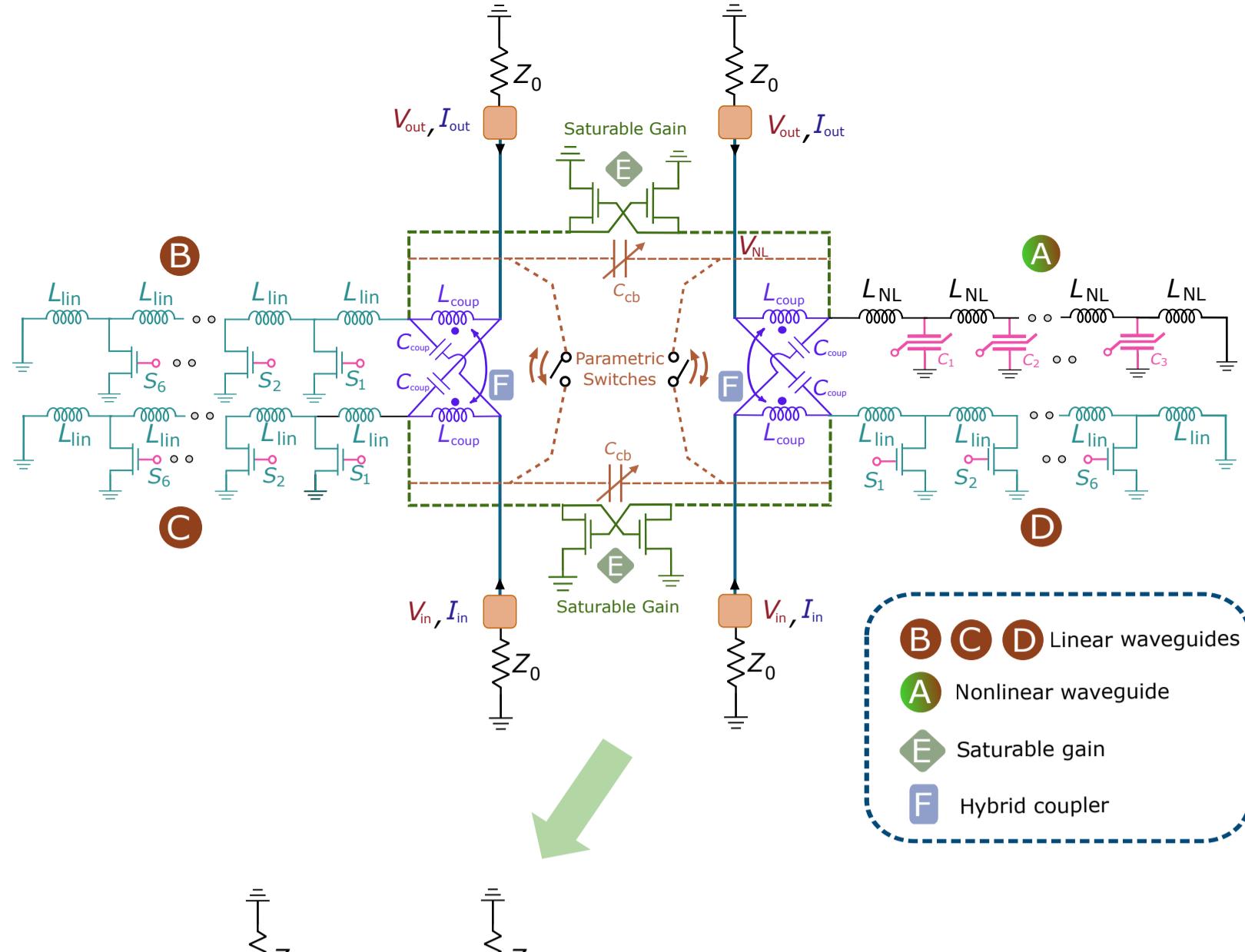
### b Cross section of CMOS stackup



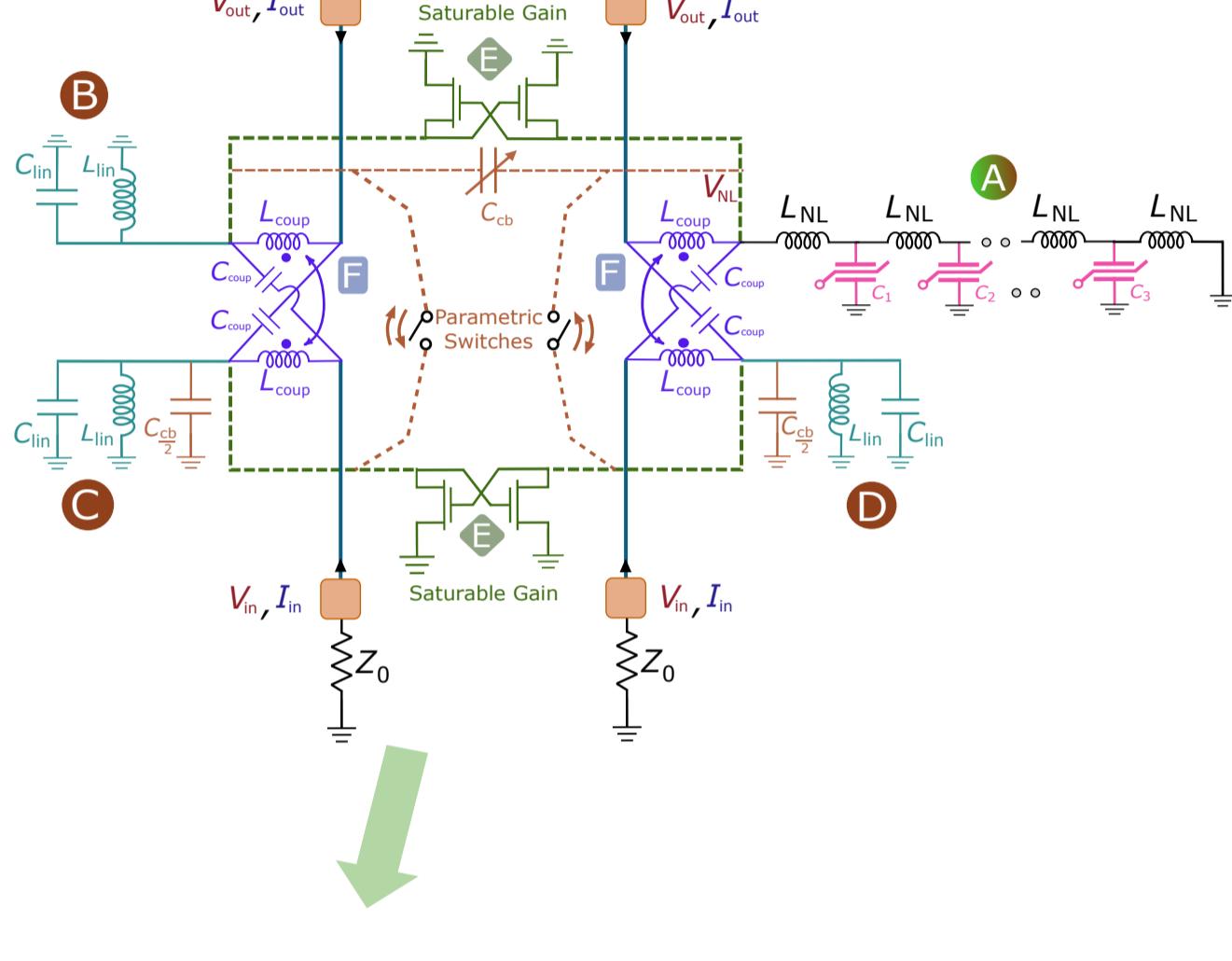
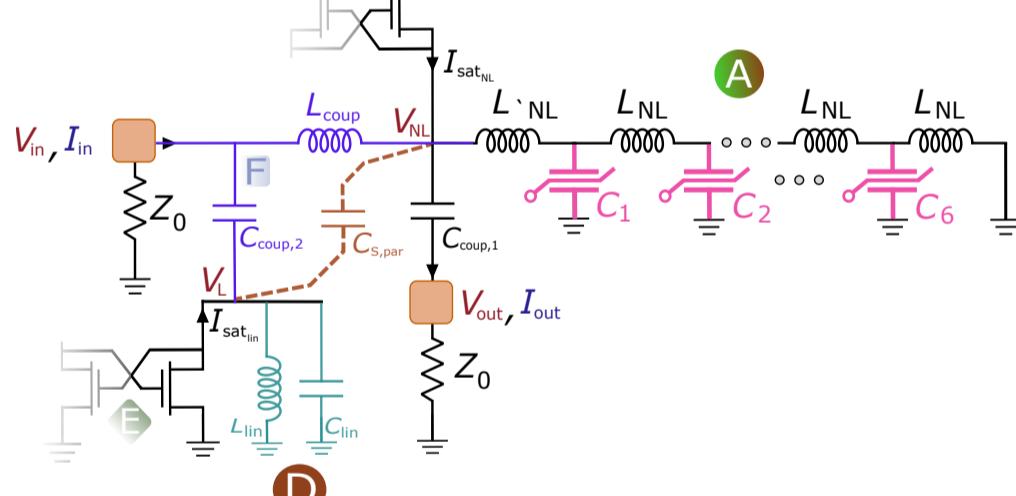
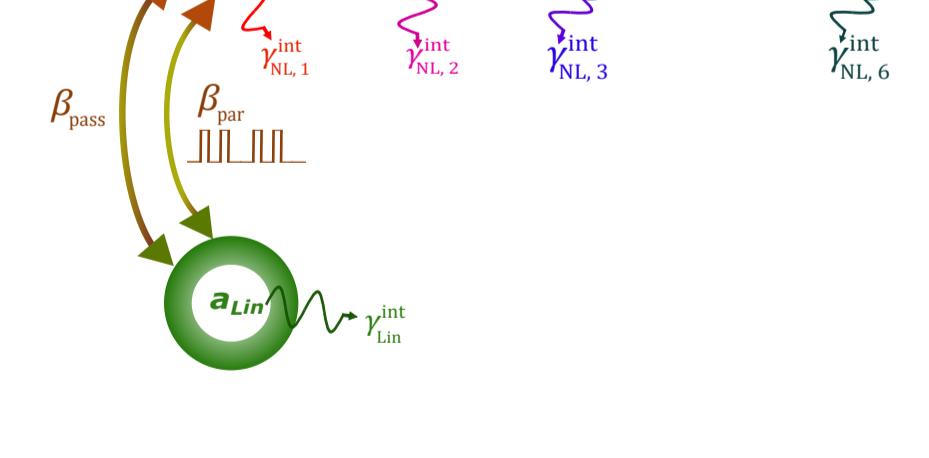
### c Tunable-length linear waveguides $B$ , $C$ and $D$



**Extended Data Fig 2 | Structure of the distributed linear waveguide** **a**, Top view of the layout of the tunable waveguide that produces linear modes  $B$ ,  $C$  and  $D$ . Switches  $S_{1,2,\dots,6}$  are inserted periodically along the length the transmission line and makes it tunable in length, with subsegments  $TLine_{1,2,\dots,6}$ . Shorting the line through these switches alters the effective length of the waveguide to support different fundamental frequencies. **b**, Cross-section of RF-optimized and digital metal layers in the 45 nm Silicon-on-Insulator CMOS metal stack. Here, the top three layers are via'ed together for low-loss transmission. The five metal layers below are used for routing control signals from a Serial-to-Parallel Interface, to the switches, and also via'ed together to form a low-loss return path to the power supply. **c**, A schematic of the linear waveguide resonators, with options to lengthen or shorten the return path of the microwave signal. In the experiments, however, for simplicity, only the shortest path (configuration with all switches turned on) was used for training in the machine learning tasks.

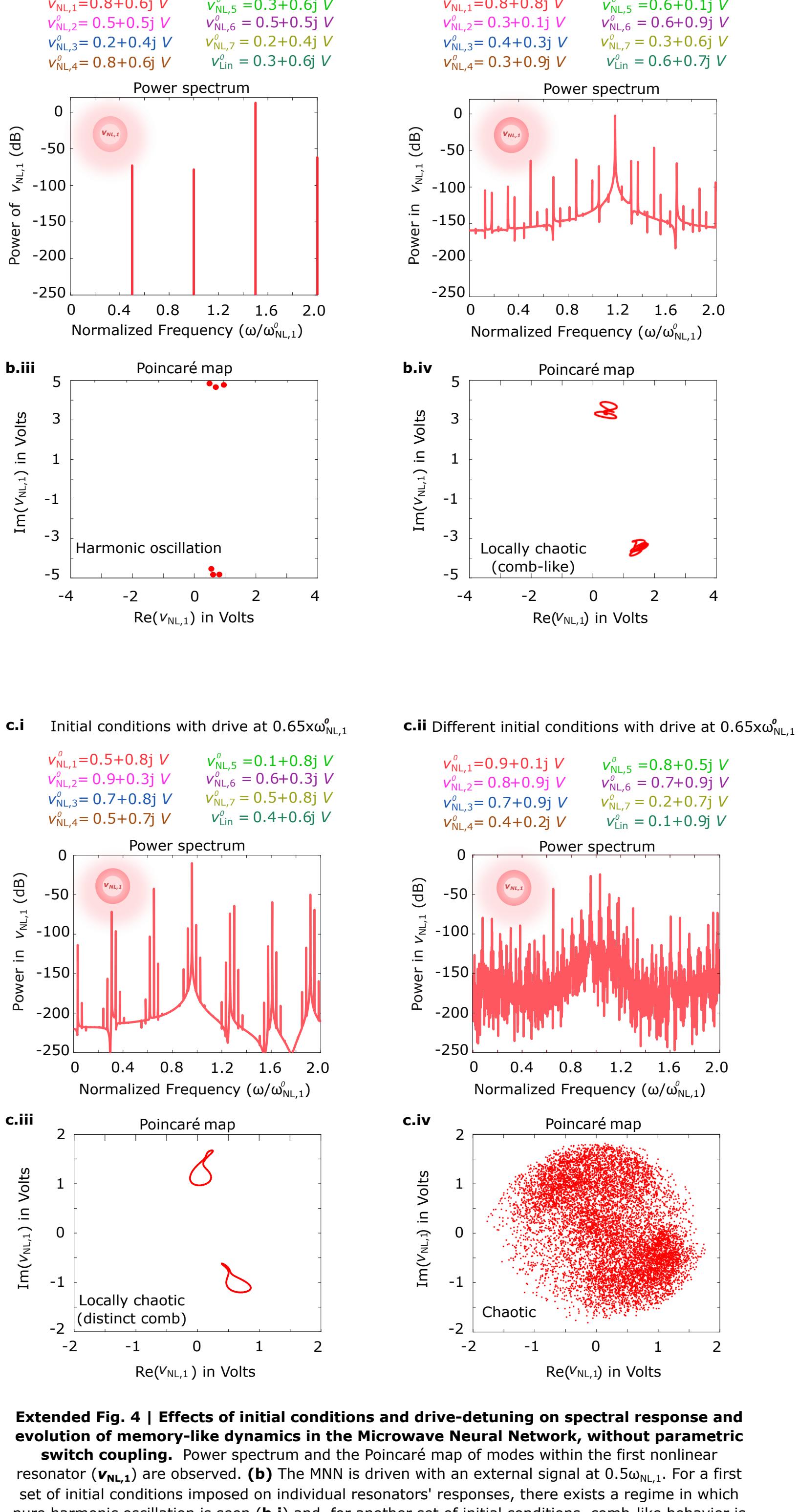
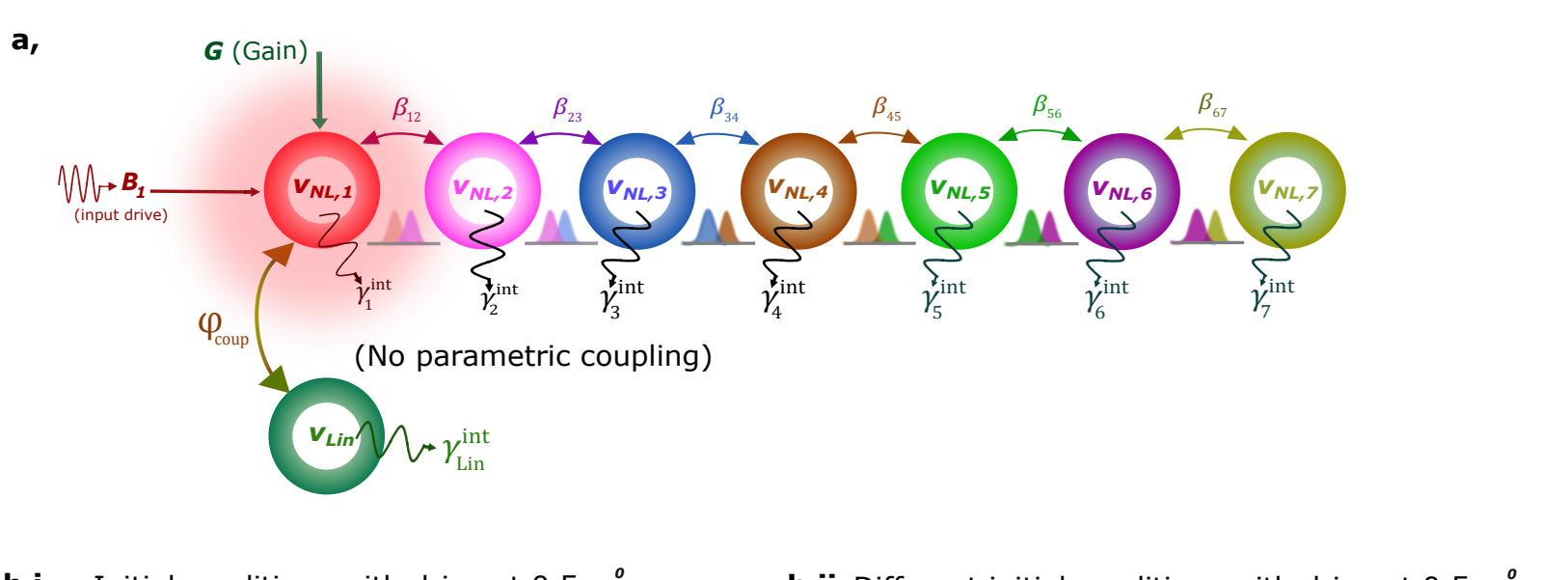
**a**

**B** **C** **D** Linear waveguides  
**A** Nonlinear waveguide  
**E** Saturable gain  
**F** Hybrid coupler

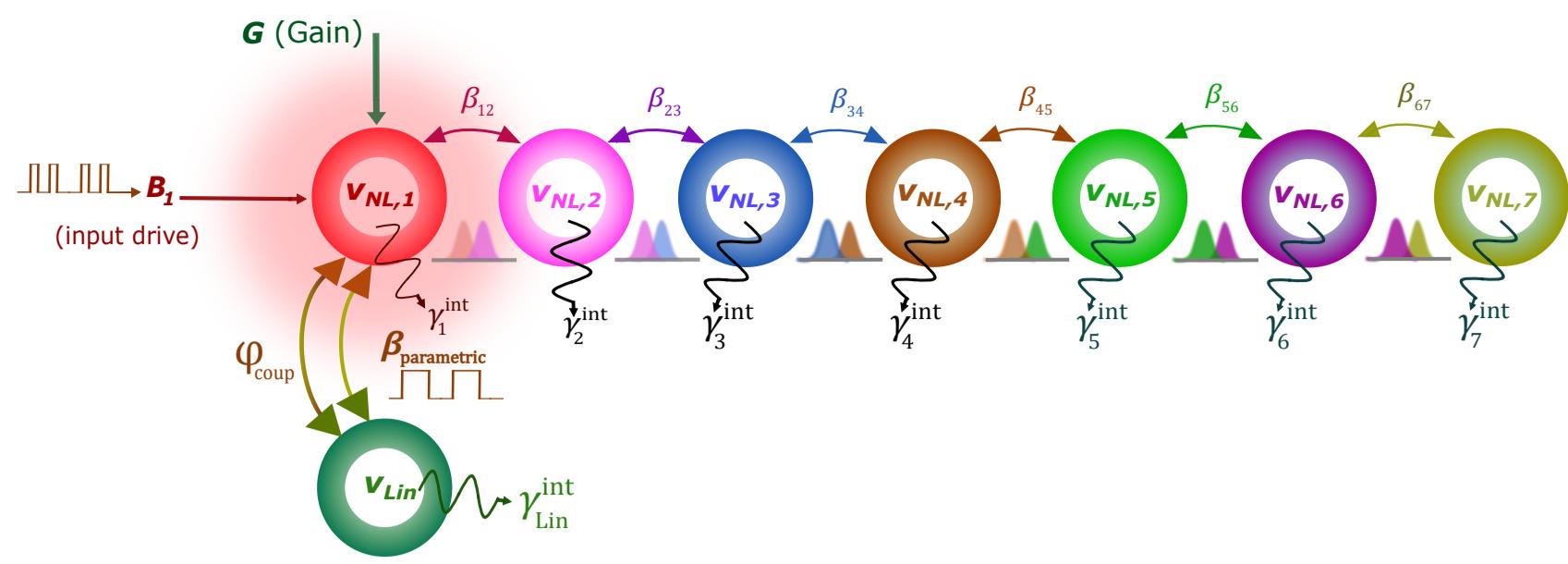
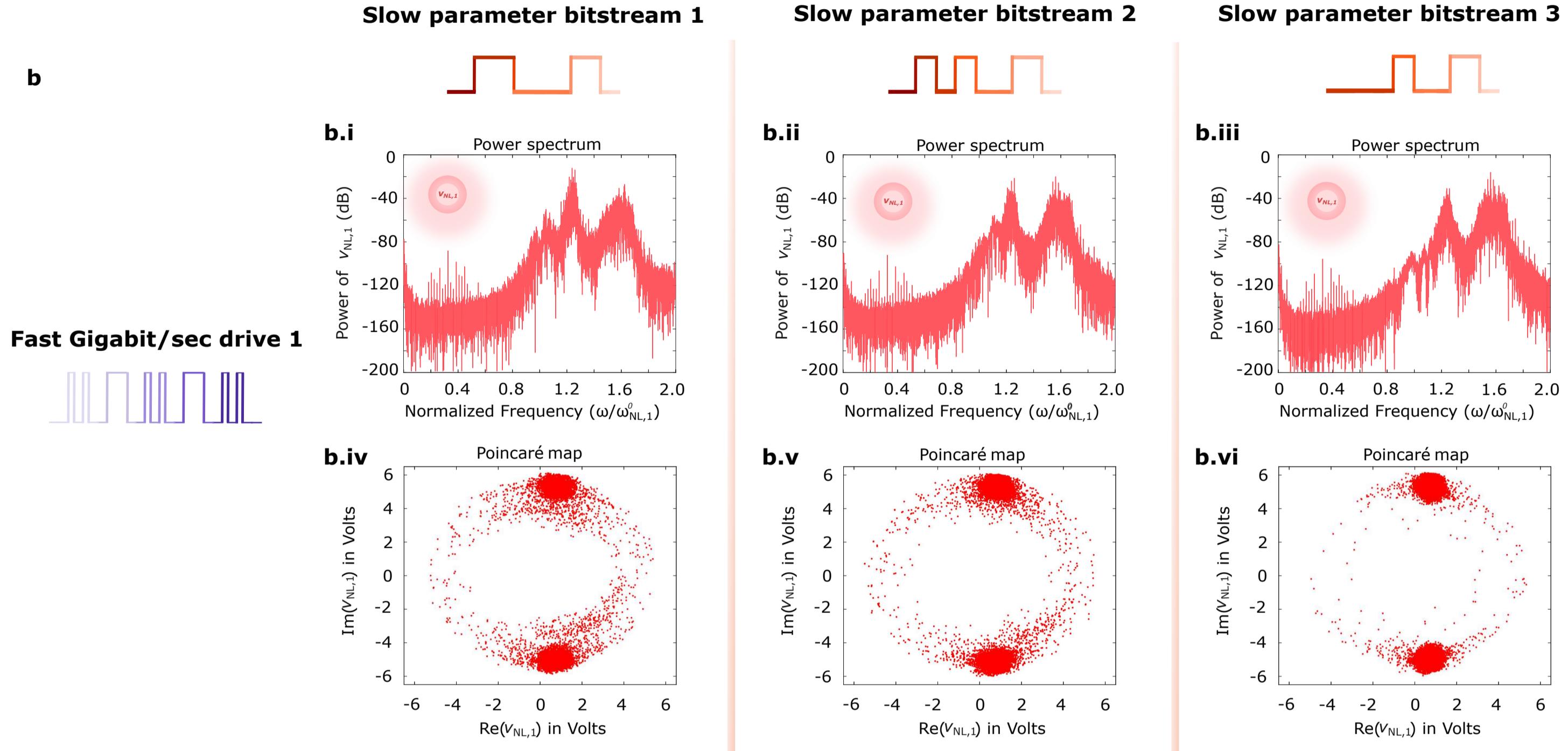
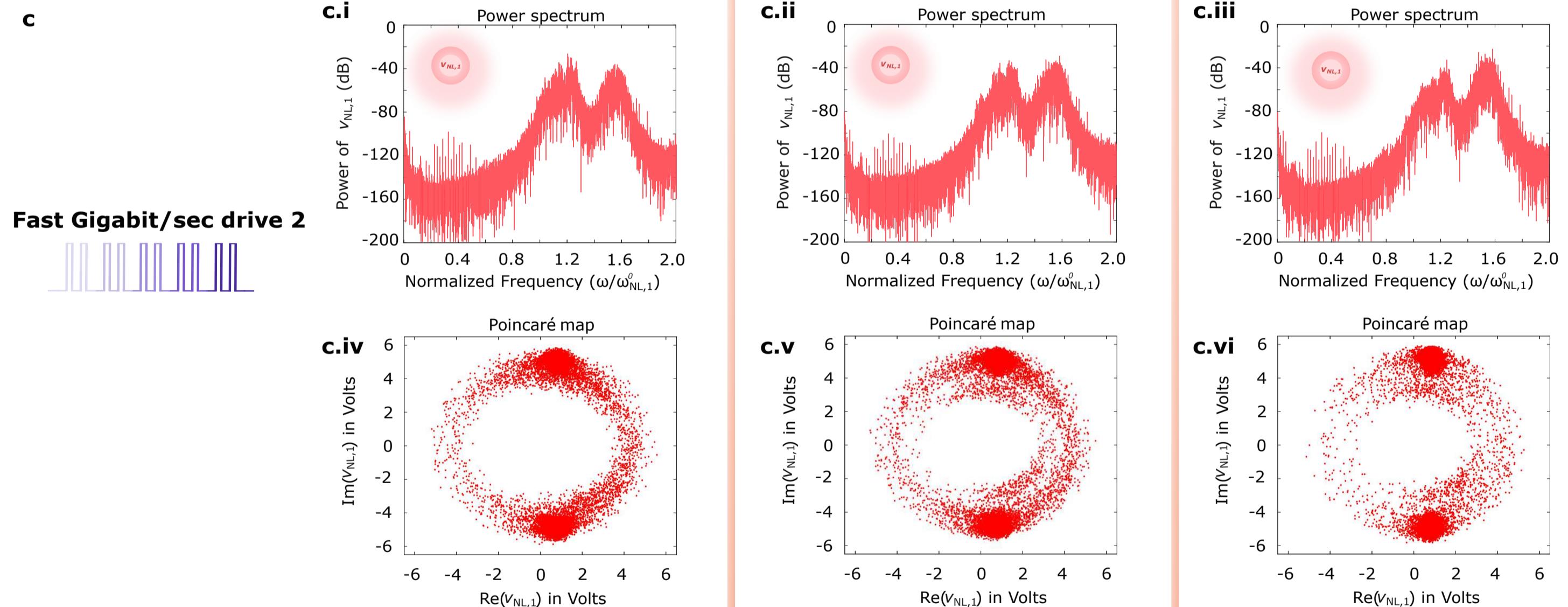
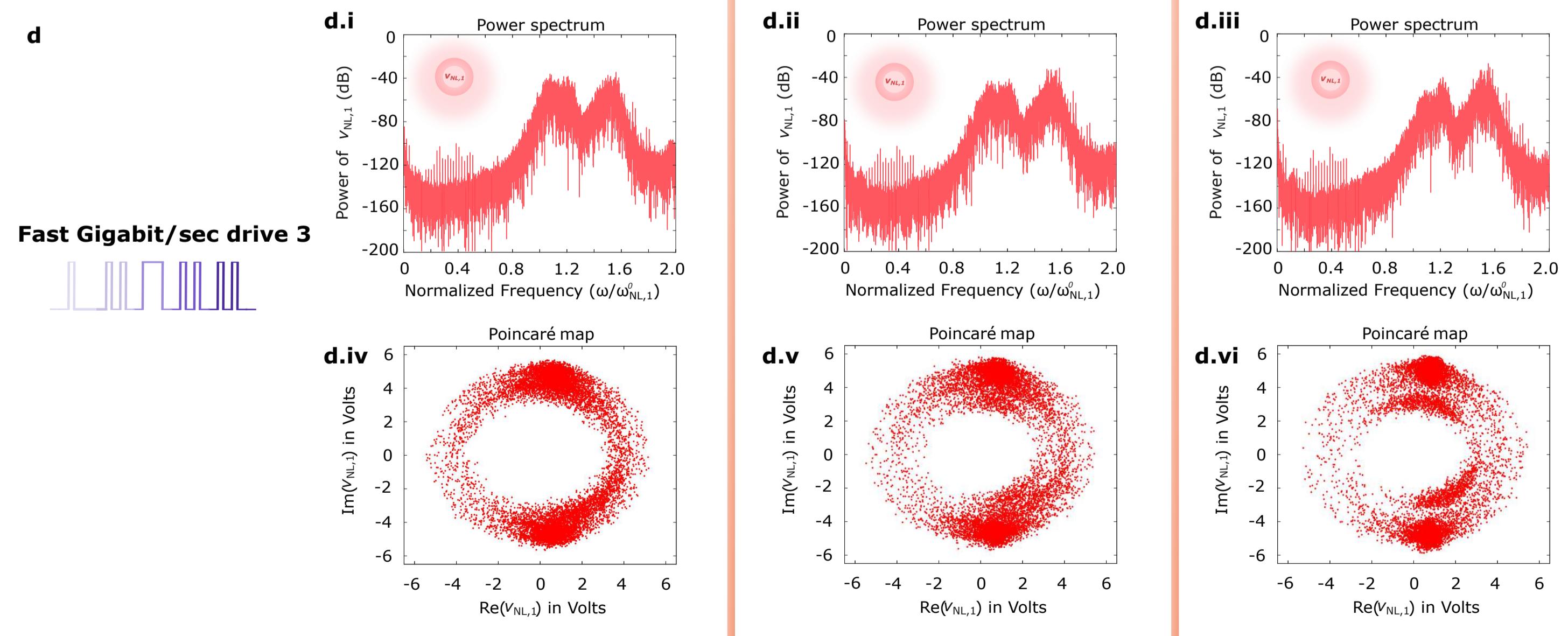
**b****c****d**

### Extended Data Fig 3. Reduction of the CMOS circuit to a generalized coupled mode model. a,

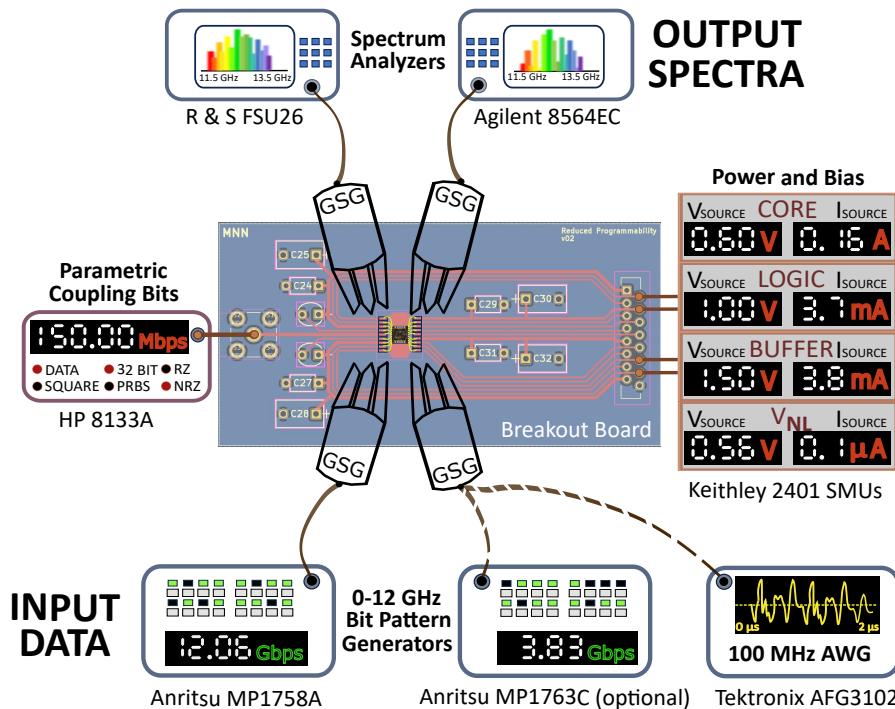
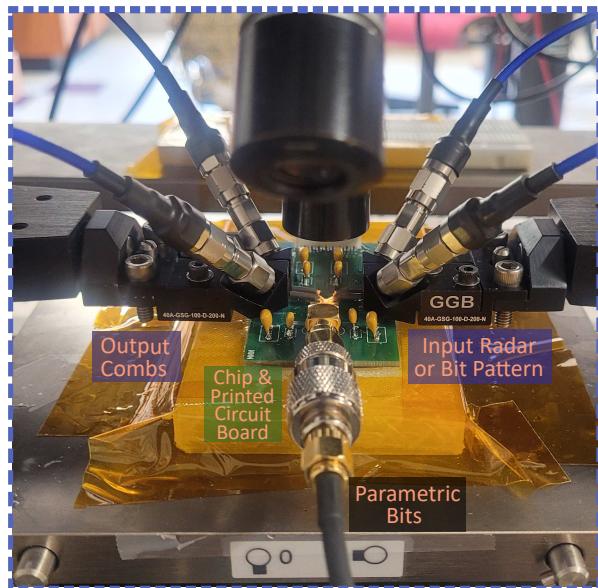
The integrated microwave neural network consists of interconnected linear and nonlinear resonators. The linear resonator is designed as a single waveguide with an adjustable length, implemented through a cascade of sub-segments, each referred to as  $L_{lin}$ . These sub-segments can be grounded via switches ( $S_1, S_2, \dots, S_6$ ), which immediately terminate the microwave signal's return path at the first switch that is shorted to ground. In contrast, the nonlinear resonator features a transmission line loaded with polynomially nonlinear capacitors. These capacitors form  $C-L-C$   $\Pi$  sections that are coupled by delays. Microwave power from input pads is distributed to these resonators through symmetrically arranged couplers (whose equivalent circuits are marked in purple). The left coupler divides power into two linear waveguides, while the right coupler feeds into a linear waveguide and into a nonlinear waveguide. Saturable gain elements, implemented as cross-coupled transistor pairs, connect the waveguides on opposite sides of the circuit, compensating for losses within the electromagnetic structures. Additionally, a pair of capacitor banks provides a small degree of tunability to the modes supported by the waveguides. Critically, there is parametric coupling between the circuit's upper and lower halves through a pair of slow bitstream-driven switches. **b**, To simplify the complex circuit, we recognize that the linear resonators support only a single natural frequency, they can be represented as tank circuits composed of  $L_{lin}$  and  $C_{lin}$ . The symmetry in the bottom half of the circuit allows us to approximate the capacitor banks as two evenly split capacitors, contributing to the overall capacitance of the tank circuits. However, the asymmetry in the configuration of the resonators in the upper half does not permit such a simplification. **c**, To focus on the primary mechanism by which the system's sensitivity to incoming signals is enhanced, we can largely ignore the left half of the circuit and concentrate on the interaction between the nonlinear distributed resonances and the linear resonator on the right half. These components interact only through the inductive path via a coupler and a coupling capacitor between the turns of the coupler. The source of regenerative gain through the cross-coupled pair is retained. For analysis, we represent the parametrically driven switch as a tunable capacitor, which can be toggled between a very small value (open circuit) and a very large value (short circuit). **d**, The reduced circuit can be represented as an ensemble of coupled modes—a cascade of nonlinear resonators connected to a linear resonator via a parametrically varied switched coupling and a fixed phase delay (through the coupler). These modes interact with the incoming drive (radar or fast Gigabit/sec digital data) with internal losses being compensated by saturable gain. Here,  $\beta_{NL}$  is the coupling coefficient between nonlinear modes,  $\gamma_{NL}^{int}$  is the internal decay rate of these modes,  $\beta_{par}$  is the parametric coupling rate,  $\beta_{pass}$  is the passive coupling between waveguides,  $G$  is the saturable gain and  $B_1$  is the external drive (fast digital signals or radar waveforms, for instance).



**Extended Fig. 4 | Effects of initial conditions and drive-detuning on spectral response and evolution of memory-like dynamics in the Microwave Neural Network, without parametric switch coupling.** Power spectrum and the Poincaré map of modes within the first nonlinear resonator ( $v_{NL,1}$ ) are observed. **(b)** The MNN is driven with an external signal at  $0.5\omega_{NL,1}$ . For a first set of initial conditions imposed on individual resonators' responses, there exists a regime in which pure harmonic oscillation is seen (**b.i**) and for another set of initial conditions, comb-like behavior is produced (**b.ii**). These correspond to situations where the Poincaré map shows sparse and organized points. This suggests a quasi-stable dynamic state (**b.iii**) and another where the two islands reflect more unstable, dynamic responses (**b.iv**), indicating locally chaotic solutions. **(c)** If, instead, the drive was fed to the MNN at  $0.65\omega_{NL,1}$ , different working regimes are triggered by different initial conditions. True comb-like behavior (**c.i**) can be produced, as evidenced by isolated, longer-memory, coherent solutions on the Poincaré map (**c.iii**). In another instance, divergent chaotic behavior shown by dense, scattered points (indicative of chaotic dynamics) (**c.iv**) manifests itself by a less structured spectrum emitted by the nonlinear resonator(s) (**c.ii**). For all simulations, nonlinear coupling coefficients, gain and decay rates are normalized with respect to  $\omega_{NL,1}$ . Here,  $\beta_{i,i+1} = 0.02$  and  $\gamma_i^{\text{int}} = 0.03$ , with  $i = 1, 2, \dots, 7$  and saturable gain,  $\mathbf{G}$ , equals 0.2. Since parametric coupling is absent,  $\beta_{\text{par}} = 0$ .

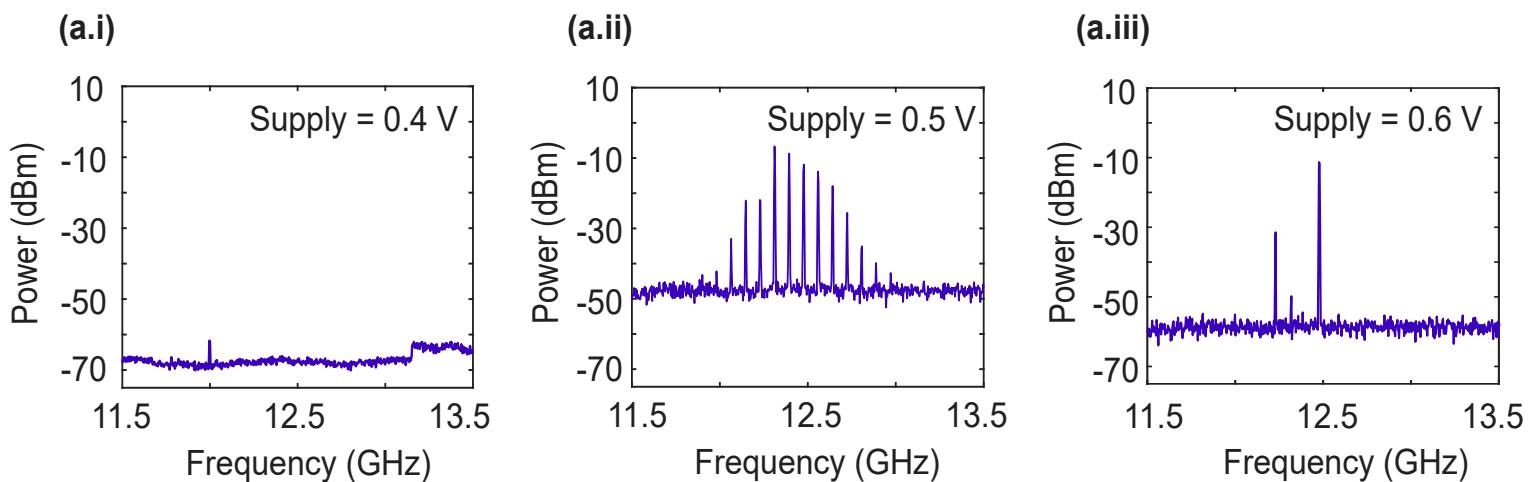
**a****b****c****d****Extended Data Fig. 5 | Manipulation of the comb-like spectra emitted by the Microwave Neural Network, with microwave-speed signals and slow parametric bitstreams.**

**a**, In the experiments involving Gigabit/sec data and radar returns, the accessible phase space of the MNN, in terms of the real and imaginary components of the resonators' modes, is constrained by the configurations of incoming drive signals and the parametric modulation applied on the switch between the first non-linear mode and a linear mode. **b** and **c**, The output spectra and Poincaré maps. For Drive 1, the Poincaré map for slow parameter bitstreams 1 and 2 shows clustered points with a few points that are more spread out (**b.iv** and **b.v**), suggesting quasi-periodic dynamics while Bitstream 3 gives tightly clustered points (**b.vi**), indicating stable, periodic behavior. For Drive 2, under action of all three parametric bitstreams, the formation of ring-like structures indicates quasi-periodic behavior with non-linear dynamics (**c.iv**, **c.v** and **c.vi**). Finally, under Drive 3, the Poincaré maps, parametric bitstreams 1 and 2 produce a combination of ring-like patterns and clustered points (**d.iv** and **d.v**), revealing a tendency towards chaotic behavior, while Bitstream 3 exhibits multiple, structured rings, indicating quasi-periodic behavior with complex and unstable dynamics (**d.vi**). For all simulations, nonlinear coupling coefficients, gain and decay rates are normalized with respect to  $\omega_{NL,1}$ . Here,  $\beta_{i,i+1} = 0.02$ ,  $\gamma_{int}^i = 0.03$ , with  $i=1,2\dots7$  and saturable gain,  $G$ , equals 0.2.

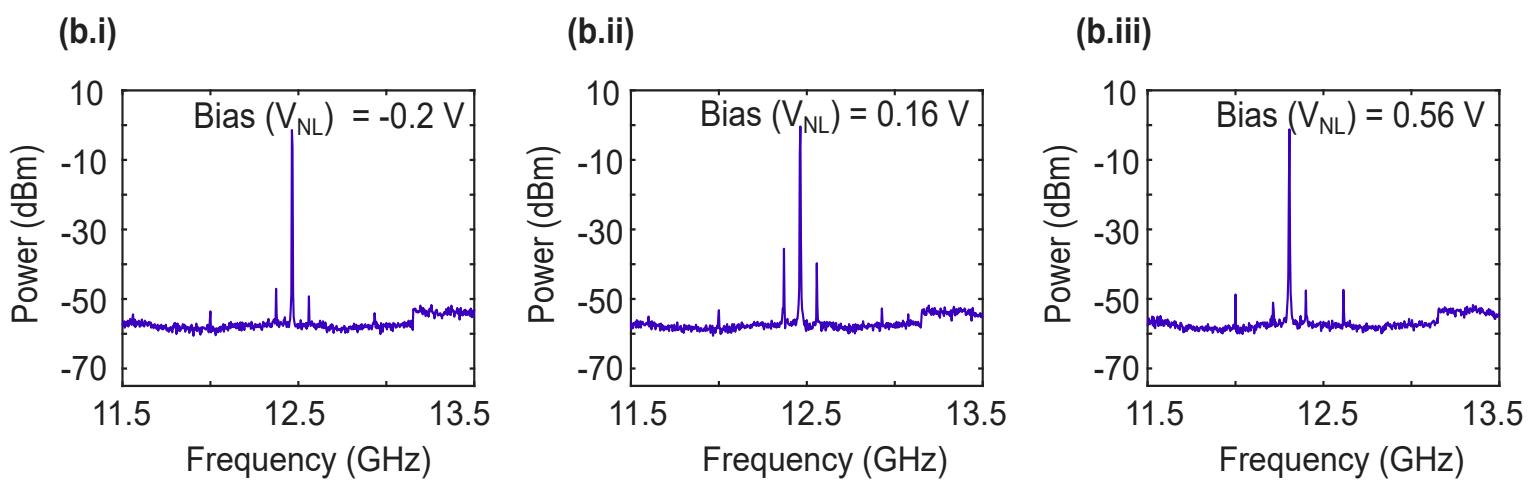
**a****b**

**Extended Data Figure 6. Experimental setup to record the MNN's response to microwave drives and parametric bitstreams.** The CMOS MNN chip is wire-bonded to a Printed Circuit Board, which connects it to external power supplies and bias voltages for the oscillators' core, logic and drivers and nonlinear capacitors' bias. **a**, A low-speed parametric bit pattern (at 150 Mbit/sec) drives switches that establish parametric coupling between linear and nonlinear oscillators. A first pair of probes, forming a Ground-Signal-Ground-Signal-Ground (GSGSG) configuration cyclically transfers high bandwidth 0-12 Gbit/sec bitstreams into the chip to interact with the default comb-like response. The resulting microwave computations are manifested as new comb-like spectra. These output spectra, from two ports, are read off spectrum analyzers in smaller bands of about two or three gigahertz, through a second pair of probes and Ground-Signal-Ground-Signal-Ground waveguides. **b**, The probe-station assembly consists of the CMOS die attached to a breakout PCB, interfaced with millimeter wave probes for input and output data and power supplies. It also includes a low-speed BNC cable interface for feeding in the parametric bitstream.

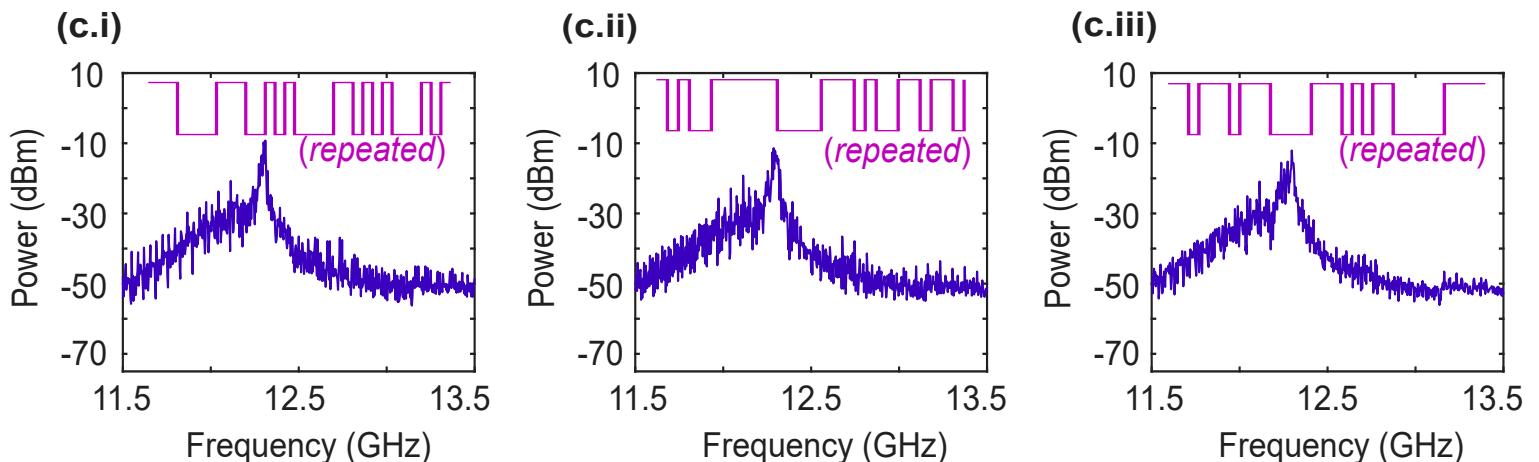
**a.** Spectra formed by variations in saturable gain supply voltage, with a constant nonlinear bias ( $V_{NL}$ ) = 0.2 V



**b.** Spectra formed by variations in nonlinear bias ( $V_{NL}$ ) with a constant supply voltage = 0.5 V

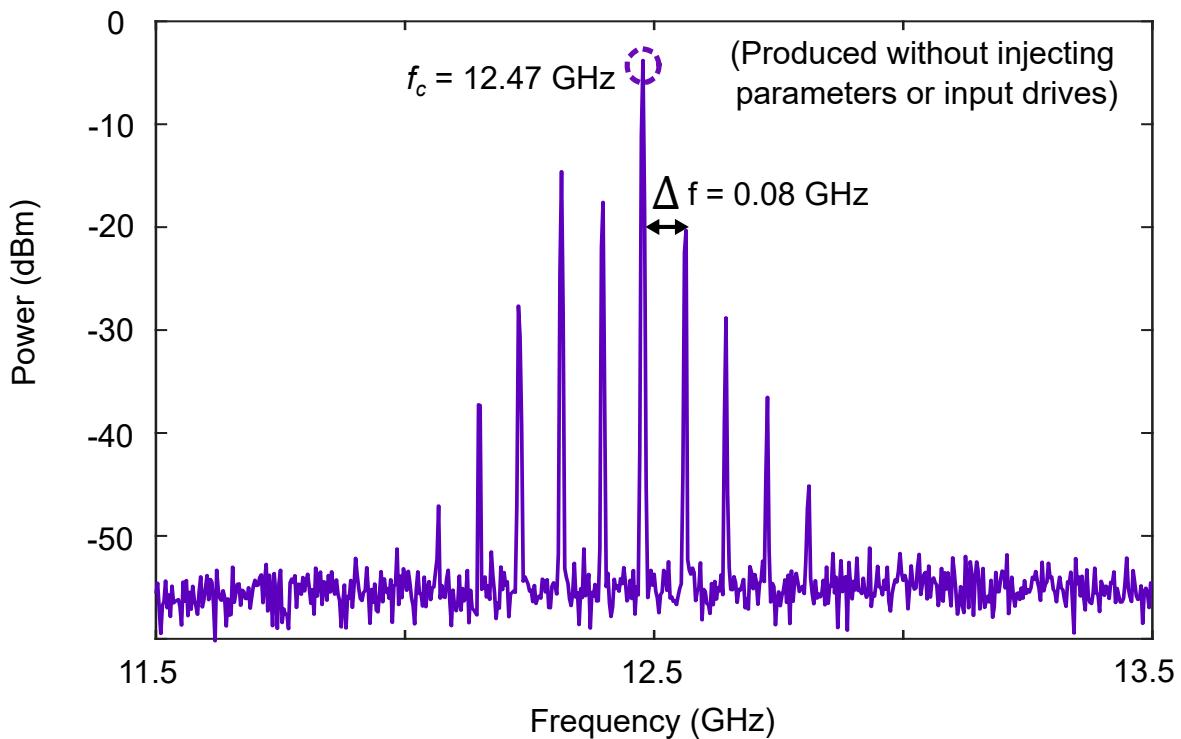
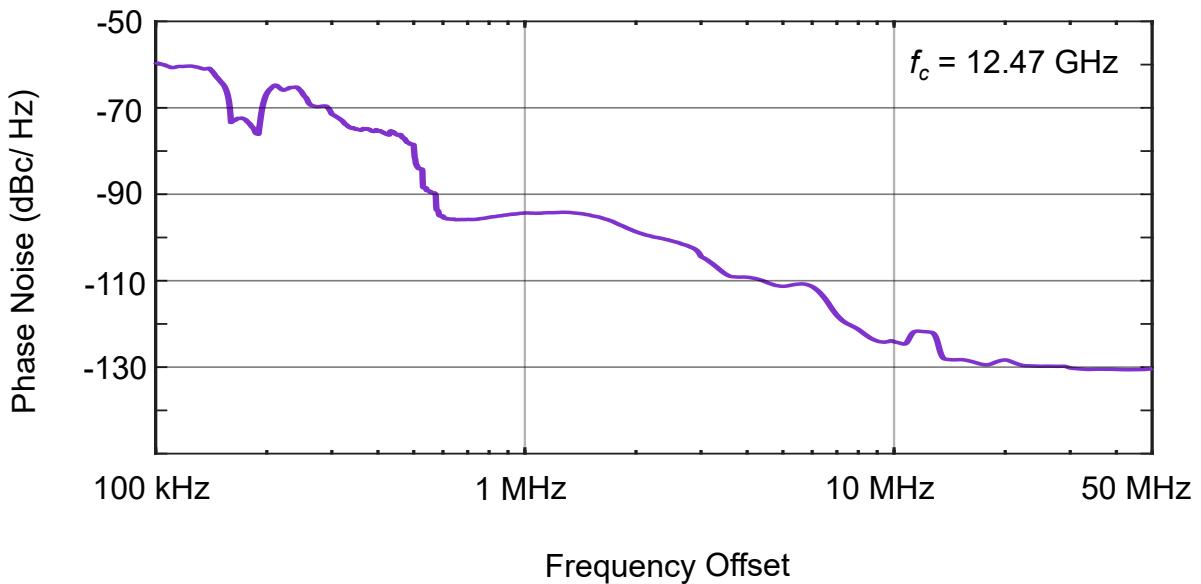


**c.** Comb-like spectra reshaped by variations in slow (50 MHz) parametric bits, at  $V_{NL}$  = 0.56 V and supply = 0.5 V



**Extended Data Figure 7.** Measured frequency-comb-like spectra generated by the coupled nonlinear and linear resonators. **a**, The supply voltage on the saturable gain element alters the transconductance of the cross-coupled pair's transistors and, thereby, the stability of the comb.

It has a nominal value of 0.5 V, below which it collapses. **b**, The polynomial nonlinearity is tuned using bias voltages that set the sensitivity of the transmission line's capacitances to incoming microwaves. **c**, The inherent response can be altered by feeding parameters through a 32-bit sequence, run cyclically at slow speeds under 150 MHz. In the experiment, this parameterized comb is exposed to incoming drive signals and performs computation on them.

**a Measured power spectrum of undisturbed comb****b Measured Phase Noise**

**Extended Data Figure 8. Characterization of the comb generated by coupling of nonlinear modes on chip.** **a**, Under nominal biasing conditions (power supply of 0.6 V and  $V_{NL} = 0.6$  V) and without the influence of parametric bits and incoming microwave signals, the measured frequency comb is centered at 12.47 GHz and has a constant line spacing of 80 MHz. **b**, The phase noise is measured for various offsets from the central component. While it has slightly higher close-in noise than conventional CMOS oscillators, microwave neurons built from this modality are stable and highly sensitive to drive signals.

# Supplementary Information Guide

## TABLE OF CONTENTS

<b>Supplementary Notes' Titles</b>	<b>2</b>
<b>Supplementary Figures' Legends</b>	<b>2</b>
<b>Supplementary Tables' Legends</b>	<b>3</b>
<b>Additional Supplementary References</b>	<b>4</b>

## SUPPLEMENTARY NOTES' TITLES

**Supplementary Note 1.** Proposed application of the Microwave Neural Network in a wideband receiver chain

**Supplementary Note 2.** Proposed readout for compressed spectral features without a spectrum analyzer

**Supplementary Note 3.** Comparison of predicted accuracy with and without the Microwave Neural Network

## SUPPLEMENTARY FIGURES' CAPTIONS

**Fig S1.** A single receiver can be used to produce a baseband signal that frequency-modulates a square wave that is fed to the MNN. Its feature-rich output could be used by a cheap backend neural network, for inference (instead of a power-hungry Graphics Processing Unit).

**Fig S2.** The integration of a low-bandwidth receiver on-chip involves using 4-phase passive mixers to down-convert the outputs of the coupled oscillators to low-frequency baseband signals (<50 MHz). These signals should contain compressed features from the MNN's computations on high-bandwidth data and can replace the off-chip spectrum analyzer previously used for readout. The remainder of the integrated circuit consisting of coupled oscillators, couplers and gain remains the same.

**Fig S3.** Comparison of MNN-assisted and backend-neural-network-only accuracy. (a) Bit sequence search accuracy at 10 Gbit/sec for various queried bitstream lengths. The MNN (blue) consistently achieves higher accuracy than the optimized linear layer (orange) across all queried bitstream lengths. (b) Using the MNN enhances the accuracy of counting the number of flying targets, particularly in complex cases with 3 to 6 aircraft, where the backend neural network struggles to resolve ambiguity.

## SUPPLEMENTARY TABLE'S TITLE

**Supplementary Table 1.** Comparison of the accuracy and complexity of state-of-the-art digital neural networks and the MNN for the standard task of wireless signal encoding classification using the RadioML2016 dataset.

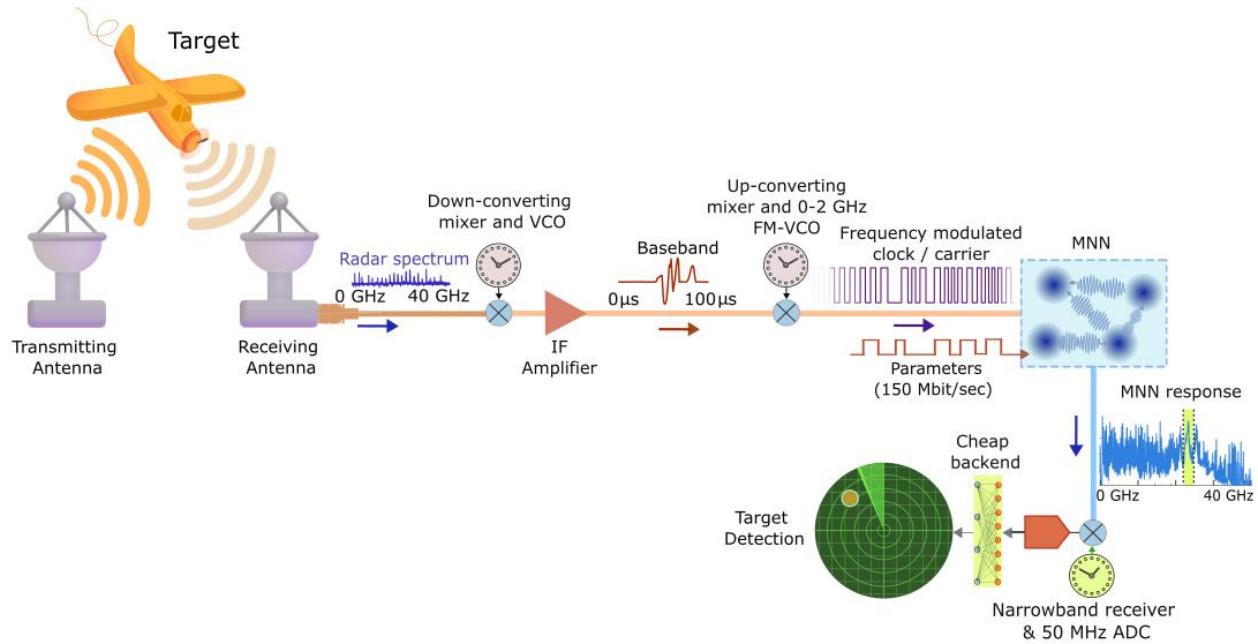
## SUPPLEMENTARY INFORMATION REFERENCES

- S1. Yi, X., Wang, C., Chen, X., Wang, J., Grajal, J., & Han, R. (2021). "A 220-to-320-GHz FMCW Radar in 65-nm CMOS Using a Frequency-Comb Architecture." *IEEE Journal of Solid-State Circuits*, 56(2), 327–339. <https://doi.org/10.1109/JSSC.2020.3020291>
- S2. J. Al-Eryani et al. "Fully Integrated Single-Chip 305–375-GHz Transceiver With On-Chip Antennas in SiGe BiCMOS". In: *IEEE Transactions on Terahertz Science and Technology* 8.3 (May 2018), pp. 329–339. doi: 10.1109/TTHZ.2018.2823202.
- S3. A. Ghaffari, E. A. M. Klumperink, M. C. M. Soer and B. Nauta, "Tunable High-Q N-Path Band-Pass Filters: Modeling and Verification", *Solid-State Circuits IEEE Journal of*, vol. 46, pp. 998-1010, 2011.
- S4. O'Shea, Timothy J., Corgan, Johnathan, and Clancy, T. Charles. "Convolutional Radio Modulation Recognition Networks." *arXiv preprint arXiv:1602.04105*, 2016.
- S5. Courtat, Thomas, and Hélion du Mas des Bourboux. "A light neural network for modulation detection under impairments." *arXiv preprint arXiv:2003.12260*, 2020.
- S6. Ramjee, Sharan, Ju, Shengtai, Yang, Diyu, Liu, Xiaoyu, El Gamal, Aly, and Eldar, Yonina C. "Fast Deep Learning for Automatic Modulation Classification." *arXiv preprint arXiv:1901.05850*, 2019.

## 1. Proposed application of the Microwave Neural Network in a wideband receiver chain

In radar systems, microwave signals are transmitted to a target, and the reflected signals are captured by receiving antennas. Supplementary Ref [1] highlights that a wideband receiver using parallelized signal chains that process smaller frequency bands (e.g., 0-2 GHz, 2-4 GHz, ..., 18-20 GHz) can provide consistent Effective Isotropic Radiation Efficiency (EIRP) and Noise Figure across the full spectrum. Each signal chain consists of dedicated mixers, voltage-controlled oscillators (VCOs) filters, with each Intermediate Frequency (IF) signals digitized by an Analog-to-Digital Converter. They would be sent to a GPU for target detection. However, this approach is power-intensive and the receiver does not add learned features that could ease machine-learning inference for environments with multiple dynamic targets.

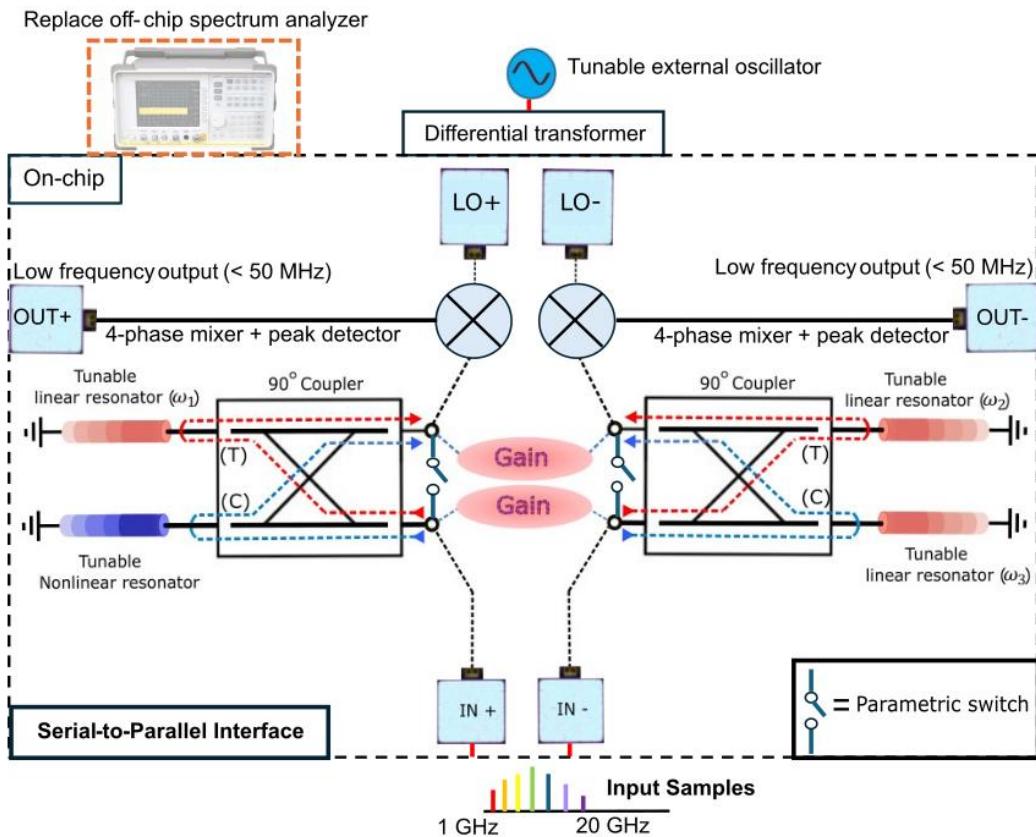
The Microwave Neural Network (MNN) presents a simpler alternative. It can be integrated within a single receiver, such as that in Supplementary Ref [2]. Supplementary Fig. 1 shows its proposed position within the signal chain. The MNN, owing to its extreme sensitivity to input signals, adds learned features through its nonlinear expansion functions, even when paired with a suboptimal radar front-end. As explained in Fig. 4a.i of the main article, the baseband signal frequency-modulates a carrier (FM-VCO), and slow parameter bits (150 Mbit/sec) enable the MNN to perform computation. A single down-conversion mixer (see Supplementary Note 2) and sub-50 MHz ADC can then manage readout. A lightweight neural network then maps compressed features to target trajectories without the need for power-hungry GPUs.



**Supplementary Fig 1.** A single receiver can be used to produce a baseband signal that frequency-modulates a square wave that is fed to the MNN. Its feature-rich output could be used by a cheap backend neural network, for inference (instead of a power-hungry Graphics Processing Unit).

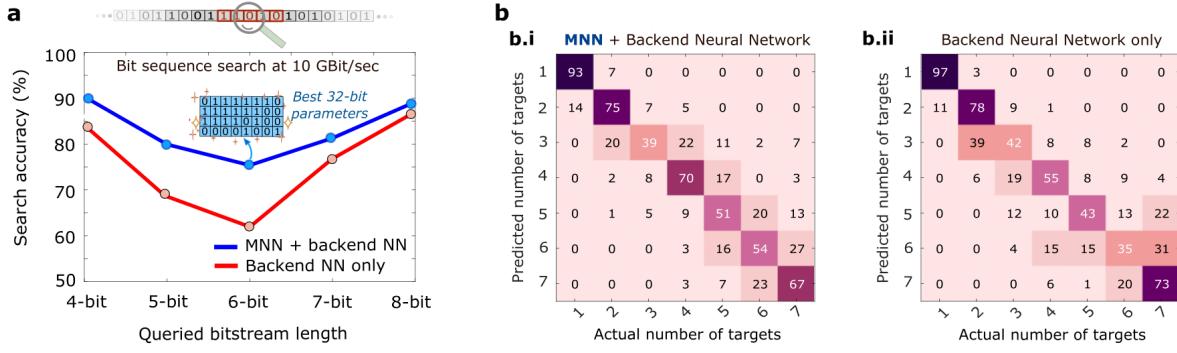
## 2. Proposed readout for compressed spectral features without a spectrum analyzer.

The Microwave Neural Network presented in the main article utilized off-chip digitization with a spectrum analyzer that read off a narrow band of frequencies. In a fully integrated version we have fabricated (not discussed in the main article), we have proposed replacing this with an on-chip solution, whose schematic is shown in Supplementary Fig. 2. Here, incoming microwave drive signals interact with the MNN's parametric oscillations to produce distinct comb-like spectra. Instead of feeding directly to the output pads, processed signals from the couplers' outputs are directed to two mixers that down-convert portions of these spectra based on the frequency of a tunable external oscillator. This employs an 'N-path' topology (Supplementary Ref [3]), functioning as a tunable bandpass filter at RF frequencies, translating signals in that band to differential, sub-50 MHz outputs. The quality factor of these filters determines the bandwidth of the received baseband signal. This signal can then be digitized with a low-bandwidth, sub-50 MHz analog-to-digital converter, either on- or off-chip, and the features used to train a linear backend neural network. For even simpler readout, on-chip peak detectors can instead be used to record the power at the desired frequency.



**Supplementary Fig. 2.** The integration of a low-bandwidth receiver on-chip involves using 4-phase passive mixers to down-convert the outputs of the coupled oscillators to low-frequency baseband signals (<50 MHz). These signals should contain compressed features from the MNN's computations on high-bandwidth data and can replace the off-chip spectrum analyzer previously used for readout. The remainder of the integrated circuit consisting of coupled oscillators, couplers and gain remains the same.

**3. Comparison of predicted accuracy with and without the Microwave Neural Network**  
Supplementary Fig. 3 compares performance with and without the MNN in the path of incoming signals. Results with the MNN present are repeated from the main article. For the bit-sequence detection task, for the case of without using MNN, the linear layer is directly trained with on time-domain sequences. For the radar tracking task, the frequency modulated square waves (as in the main article) are fed directly to the spectrum analyzer, and the output spectra are truncated to the reduced frequency ranges used by the MNN. The backend neural networks discussed in Methods Section 5 are reused here for fair comparison. In the representative results shown below, for both digital emulation and RF communication, tasks, the inclusion of the MNN improves accuracy. For bit sequence detection, the MNN achieves higher search accuracy than the optimized linear-layer backend across all queried bit-sequence lengths, exceeding it by 10% in the worst case of 6-bit queries. When analyzing flight patterns in simulated airspace, the MNN shows a clear improvement in accurately predicting the number of flying targets, especially in the worst cases with 3 to 6 aircraft.



**Supplementary Fig. 3.** Comparison of MNN-assisted and backend-neural-network-only accuracy. **(a)** Bit sequence search accuracy at 10 Gbit/sec for various queried bitstream lengths. The MNN (blue) consistently achieves higher accuracy than the optimized linear layer (orange) across all queried bitstream lengths. **(b)** Using the MNN enhances the accuracy of counting the number of flying targets, particularly in complex cases with 3 to 6 aircraft, where the backend neural network struggles to resolve ambiguity.

**Supplementary Table 1.** Comparison of the accuracy and complexity of state-of-the-art digital neural networks and the MNN for the standard task of wireless signal encoding classification using the RadioML2016 dataset.

Model	RadioML2016.10a Val. Accuracy at 18dB SNR	Number of Parameters
RML-ResNet [S4]	90%	240K
Mod-LRCNN [S5]	91%	100K
CLDNN [S6]	88%	6.5M
MNN + Lin. Backend (ours)	87.4%	<7K (linear layer) + 32 bits

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformationIntegratedMicrowaveNeuralNetwork.pdf](#)
- [Extendedfig.pdf](#)