Resonant Machine Learning Based on Complex Growth Transform Dynamical Systems

Oindrila Chatterjee, *Student Member, IEEE*, and Shantanu Chakrabartty, *Senior Member, IEEE*

***Abstract*— Traditional energy-based learning models associate a single energy metric to each configuration of variables involved in the underlying optimization process. Such models associate the lowest energy state with the optimal configuration of variables under consideration and are thus inherently dissipative. In this article, we propose an energy-efficient learning framework that exploits structural and functional similarities between a machine-learning network and a general electrical network satisfying Tellegen's theorem. In contrast to the standard energy-based models, the proposed formulation associates two energy components, namely, active and reactive energy with the network. The formulation ensures that the network's active power is dissipated only during the process of learning, whereas the reactive power is maintained to be zero at all times. As a result, in steady state, the learned parameters are stored and self-sustained by electrical resonance determined by the network's nodal inductances and capacitances. Based on this approach, this article introduces three novel concepts: 1) a learning framework where the network's active-power dissipation is used as a regularization for a learning objective function that is subjected to zero total reactive-power constraint; 2) a dynamical system based on complex-domain, continuous-time growth transforms that optimizes the learning objective function and drives the network toward electrical resonance under steady-state operation; and 3) an annealing procedure that controls the tradeoff between active-power dissipation and the speed of convergence. As a representative example, we show how the proposed framework can be used for designing resonant support vector machines (SVMs), where the support vectors correspond to an *LC* network with self-sustained oscillations. We also show that this resonant network dissipates less active power compared with its non-resonant counterpart.**

***Index Terms*— Complex-domain machine learning, coupled oscillators, electrical resonance, energy-based learning models, energy-efficient learning models, resonant networks, support vector machines (SVMs), Tellegen's theorem.**

I. Introduction

FROM an energy point of view, the dynamics of a machine-learning framework is analogous to an electrical network since both evolve over a conservation manifold to attain a low-energy state. In the literature, this analogy has served as the basis for energy-based learning models, where the learning objective function is mapped to an equivalent network energy [1], [2]. The network variables then evolve according to some physical principles subject to network constraints to seek out an energy optimal state. However, even if the solution reached by such learning models may be optimal from the objective function point of view, it may not be the most energy-efficient solution when considering an actual physical implementation. This is because most of these formulations assume that the energy being minimized in the network is dissipative in nature.Whereas, in a physical electrical network, the total power S_N (also known as the apparent power) comprises not only of the dissipative component (also referred to as the active power) but also a latent or stored nondissipative component (also referred to as the reactive power) [3], [4]. This is shown in Fig. 1(a) and can be mathematically expressed as

$$\text{Total Network Power } S_N = \text{Active Power } P_N + j \times \text{ Reactive Power } Q_N \quad (1)$$

where $j = \sqrt{-1}$ denotes the imaginary component. While the active power P_N represents the rate-of-energy loss in the network, the reactive power Q_N represents the rate of change in stored energy in the network's electric and magnetic fields (typically modeled as lumped capacitive and inductive elements). In the design of electrical networks, reactive power is generally considered to be a nuisance since it represents the latent power that does not perform any useful work [5], [6]. However, from the point of view of learning, the reactive

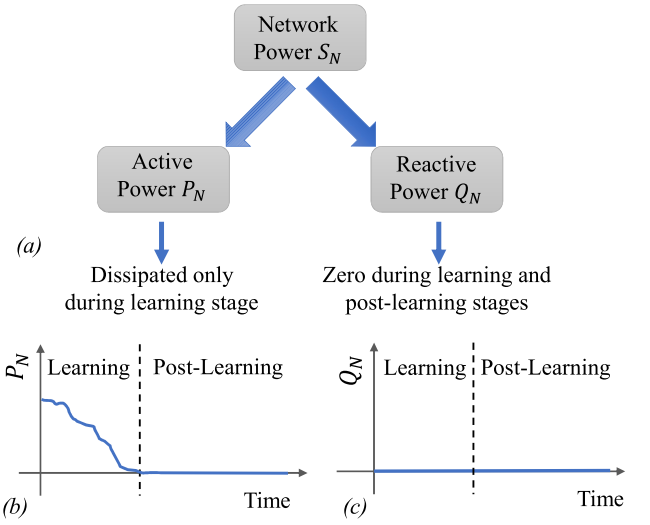

Fig. 1. (a) Total or apparent power S_N of the electrical network comprising of the active power P_N or the dissipated power and the reactive power Q_N or the power used for energy storage. The goal of the proposed resonant learning framework: (b) minimizing active power P_N during learning and ensuring $P_N = 0$, post-learning or steady state, and (c) maintaining $Q_N = 0$ in learning and post-learning phases.

Manuscript received August 14, 2019; revised January 13, 2020; accepted March 25, 2020. This work was supported in part by the National Science Foundation under Grant ECCS:1550096 and Grant ECCS: 1935073. *(Corresponding author: Shantanu Chakrabartty.)*

The authors are with the Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail: shantanu@wustl.edu).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2020.2984267

power could be useful not only for storing the learned parameters of the network but could also improve the dynamics of the network during learning. In this article, we propose a framework which exploits both active and reactive network powers for learning and memory. The objective will be to achieve a network power profile as shown in Fig. 1 (b) and (c). During learning, the network will optimize its active power (P_N), and under steady-state condition or post-learning, ensure $P_N = 0$. The reactive power (Q_N), on the other hand, will always be maintained at zero. This implies that the stored energy (mathematically—the time integral of the reactive power) is conserved across the learning and post-learning phases. Thus, during the post-learning phase or in steady state, the network will not dissipate any power, and the reactive energy is used to maintain its current network state or memory.

This steady-state condition corresponds to a state of electrical resonance, and in Section III, we generalize this concept to a framework of resonant machine learning. To reach this steady state, in Section IV, we present a dynamical system based on complex-domain continuous-time growth transforms, which extends our previous work on growth transform networks using real variables [7]. The complex-domain formulation allows manipulation of the relative phase between the voltage and current variables associated with the network nodes and thus is used to optimize the active-power dissipation during learning. While the approach could be applied to different learning networks, in Section V, we use this framework for designing resonant one-class support vector machines (SVMs) [8]. In this context, we also compare the performance of the resonant SVM model with its non-resonant variant on both synthetic and real datasets. Finally, Section VI concludes this article with a brief discussion on the implication of the proposed model when applied to other network-based models that do not involve learning, for instance, the coupled oscillator networks [9].

The key contributions of this article can be summarized as follows.

1) Contrary to standard energy-based models (EBMs) that optimize for a single energy metric, we map a learning problem to an electrical network with two energy components: dissipative(active) and nondissipative(reactive) energy.
2) Active power dissipation is introduced as a regularizer in the original learning problem while enforcing zero total reactive power by driving the network into electrical resonance. This ensures zero active power dissipation in the post-learning stage.
3) We present a novel growth transform-based complex dynamical system for optimizing the cost function that ensures a resonant condition in steady state.
4) We propose an annealing schedule that can trade off the speed of convergence with active-power dissipation for different applications.

A. Notations

For the rest of this article, we will follow the mathematical notations summarized in Table I.

TABLE I
NOTATIONS

Variable	Definition
$\mathbb{R}_+$	One-dimensional positive real vector space
$\mathbb{R}^N$	N-dimensional real vector space
$\mathbb{C}$	One-dimensional complex vector space
$\|z\|$	magnitude of a complex variable z
$\mathrm{Re}(z)$	real part of a complex variable z
$\mathrm{Im}(z)$	imaginary part of a complex variable z
z^*	complex conjugate of a complex variable z
$z(t)$	a continuous-time complex variable at time t
z_n	a discrete-time complex variable at the n^{th} time step

II. Background and Related Works

A. Energy-Based Learning Models

The principle of minimum energy states that all self-organizing physical systems evolve dynamically over an intrinsic conservation manifold to attain their minimum energy configuration in the steady state [10]. Energy-based machine-learning models follow a similar cue, where the goal is to find the optimal configuration of a predetermined energy landscape determined by the learning problem at hand [1], [11]. Some of the most notable examples of EBMs are those based on the Ising model of statistical mechanics, such as the Hopfield network [2], and its stochastic counterpart the Boltzmann machine [12] and its variants [13]. In these models, the network converges to the local minimum of a Lyapunov function in steady state. Another class of EBMs [1] proposes a unified learning framework by mapping the learning problem to a global scalar "energy function." The algorithm minimizes a loss functional to seek out an optimal energy landscape that associates the lowest energy to the observed configuration of variables involved in the underlying optimization process. These methods are essentially non-probabilistic and often involve intractable integrals, which require MCMC methods for normalization when applied to probabilistic frameworks. Different classes of optimization algorithms such as contrastive divergence, contrastive Hebbian learning, and equilibrium propagation [14], have been proposed for a variety of supervised, unsupervised, and reinforcement learning applications [15]–[18] in the context of EBMs. However, all these approaches consider an energy metric that solely depends on the learning task under consideration. This implies that in an analog implementation of the model, the optimal solution may not be the most energy-efficient one.

B. Complex-Domain Machine-Learning Models

A variety of complex-domain learning models have been proposed in the literature for different learning algorithms, e.g., complex neural networks, complex SVMs, and complex

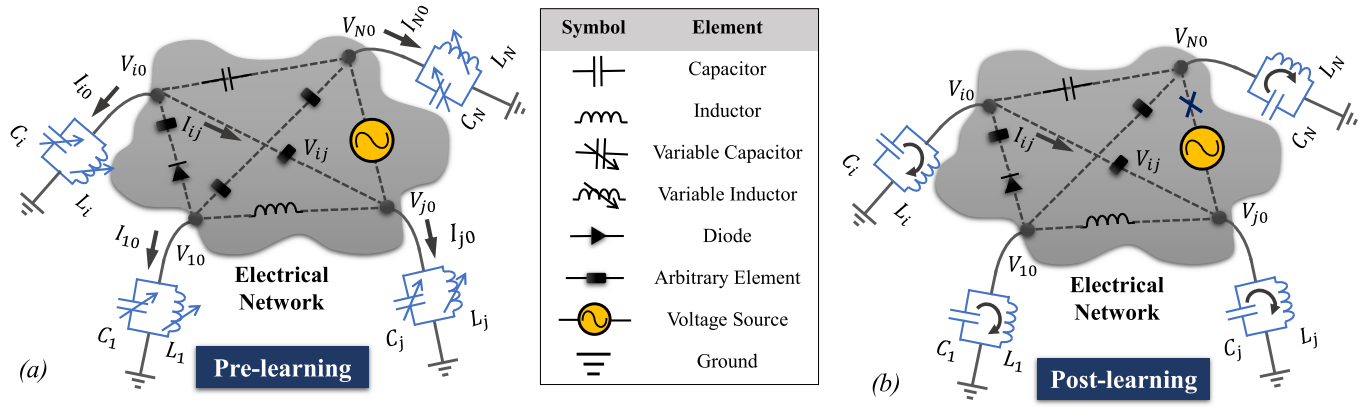

Fig. 2. Equivalent network model comprising of N electrical nodes, with an inductive and a capacitive element associated with each of the nodes. (a) Learning is equivalent to changing the values of inductive and capacitive elements. (b) In steady state, the network is driven into electrical resonance.

deep networks [19]–[23]. In addition to providing the phase variables that allow for an additional degree of freedom, other advantages of the complex-domain operation have been shown in these works. It has been demonstrated that complex learning models lead to a richer set of dynamics, noise robustness, and better convergence properties in the case of classification problems [20], [24], [25]. Moreover, phase information might provide additional insights in the context of many complex-valued physical signals (or complex-domain transforms of physical signals, e.g., Fourier or wavelet transforms) [20], [26]. However, most of these complex learning models either treat the real and imaginary components of the signal separately, or do not utilize the entire phase space when operating in the complex domain.

III. Optimization and Electrical Resonance

Consider an electrical network as shown in Fig. 2(a), comprising of N internal nodes. The voltage difference between the ith and jth nodes is denoted by V_{ij}, with V_{i0} being the ith nodal voltage with respect to ground terminal (referred to as the 0th node). Similarly, the current flowing between the ith and jth nodes is given by I_{ij}, and I_{i0} denotes the current flowing out of the ith node into the ground terminal. Then, according to Tellegen's theorem [27]

$$\sum_i V_{ij} I_{ij}^* = 0 \tag{2}$$

which states that the total complex electrical power or apparent power is zero. Isolating the apparent power flowing from the nodes to the ground terminal from that flowing between the internal nodes, we have

$$\begin{aligned}
&\sum_{i\neq j,0;\, j\neq 0} (V_{ij}) I_{ij}^* + \sum_i (V_{i0}) I_{i0}^* = 0 \\
&\qquad \Longrightarrow \sum_{i=1}^{N} V_{i0} I_{i0}^* = -\sum_{i,j=1}^{N} V_{ij} I_{ij}^* \\
&\qquad \Longrightarrow S_T = -S_N \\
&\qquad \Longrightarrow P_T + jQ_T = -P_N - jQ_N \\
&\qquad \Longrightarrow |P_T| = |P_N| \\
&\qquad \Longrightarrow |Q_T| = |Q_N|
\end{aligned} \tag{3}$$

where $S_T = \sum_i V_{i0} I_{i0}^*$ is the nodal apparent power, and $P_T = \sum_i \mathrm{Re}\{V_{i0} I_{i0}^*\}$ and $Q_T = \sum_i \mathrm{Im}\{V_{i0} I_{i0}^*\}$ are the total active and reactive power consumed at the terminal nodes. Similarly, S_N, P_N, and Q_N represent the apparent, active, and reactive power consumed due to current flow between the network nodes (other than the ground terminal). Note that this result holds even if active-power sources are embedded in the network, as shown in Fig. 2(a). Thus, (3) implies that if we minimize the active power at the nodes of the network P_T subject to the constraint that the nodal reactive power $Q_T = 0$, then the formulation would be equivalent to minimizing the network active power P_N while ensuring that the network reactive power $Q_N = 0$. This result can be equivalently expressed as

$$\begin{aligned}
&\min_{\{V_i, I_i \in \mathbb{C}\}} \quad \mathcal{D} = \sum_{i=1}^{N} |\mathrm{Re}\{V_i I_i^*\}|^2 \\
&\text{s.t.} \quad \sum_i \mathrm{Im}\{V_i I_i^*\} = 0
\end{aligned} \tag{4}$$

where we have used the notations $V_{i0} = V_i$ and $I_{i0} = I_i$ for the sake of brevity. If we assume that the ith node is associated with a lumped capacitance C_i and a lumped inductance L_i, ensuring zero nodal reactive power implies

$$\sum_{i=1}^{N} V_i \left(C_i \frac{dV_i}{dt} \right)^* + \left(L_i \frac{dI_i}{dt} \right) I_i^* = 0 \tag{5}$$

where $\left(C_i \frac{dV_i}{dt}\right)$ and $\left(L_i \frac{dI_i}{dt}\right)$ represent the current flowing through C_i and the voltage across L_i, respectively. Equation (5) is equivalent to

$$\sum_{i=1}^{N} \left(\frac{1}{2} C_i |V_i|^2 + \frac{1}{2} L_i |I_i|^2 \right) = E_0 \tag{6}$$

where $|V_i|^2 = V_i V_i^*$ and $|I_i|^2 = I_i I_i^*$, implying that the total network reactive energy is conserved to be equal to some constant value E_0. Satisfying this constraint is equivalent to sustaining a condition of electrical resonance. The optimization problem (4) can be transformed as

$$\min_{\{|V_i|, |I_i| \in \mathbb{R}_+, \phi_i \in \mathbb{R}\}} \quad \mathcal{D} = \sum_{i=1}^{N} |V_i|^2 |I_i|^2 \cos^2 \phi_i \tag{7}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} \left(\frac{1}{2} C_i |V_i|^2 + \frac{1}{2} L_i |I_i|^2 \right) = E_0, \;\; |\phi_i| \leq \pi \;\; \forall i \tag{8}$$

where ϕ_i denotes the phase angle between the voltage and current phasors at the ith node. Note that the optimization in (7) admits only three types of solutions in steady state: 1) $(|I_i| \neq 0, |V_i| \neq 0, |\phi_i| = \pi/2)$ which corresponds to a resonant LC tank; 2) $(|I_i| = 0, |V_i| \neq 0)$ which corresponds to an electrically isolated or floating node; and 3) $(|I_i| \neq 0, |V_i| = 0)$, which corresponds to a short-circuit. In Appendix A, we illustrate steady-state resonance conditions using a simple LC tank. Note that in all cases, active power is dissipated only during the learning phase, where C_i and L_i adapt to change the relative magnitude of the voltage and current variables. Fig. 2(b) shows this resonant condition of the network in the post-learning stage, whereby active power sources in the network get electrically isolated from the rest of the network. The lumped capacitance and inductance associated with the LC network at the terminal nodes adapt such that the resonant frequency condition in maintained in steady state (see Appendix A). This implies that in steady state, the learned network parameters are stored and sustained by the resonant electric and magnetic fields of the LC tanks.

The constraint in (8) can be simplified by normalizing with respect to E_0 such that

$$\sum_{i=1}^{N}(|V_i|^2 + |I_i|^2) = 1 \tag{9}$$

where $V_i \leftarrow \left(\frac{C_i}{2E_0}\right)^{1/2} V_i$ and $I_i \leftarrow \left(\frac{L_i}{2E_0}\right)^{1/2} I_i$ represent the dimension-less voltages and currents. Henceforth, unless stated otherwise, we will use dimension-less quantities in our derivations.

We now extend the optimization framework in (7) to include a general optimization function $\mathcal{H}$ as

$$\begin{aligned} \min_{\{|V_i|,|I_i|,\phi_i\}} \quad & \mathcal{L}(\{|V_i|, |I_i|, \phi_i\}) = \mathcal{H}(\{|V_i|, |I_i|\}) + \beta\mathcal{D} \\ \text{s.t.} \quad & \sum_{i=1}^{N}(|V_i|^2 + |I_i|^2) = 1, \quad |\phi_i| \leq \pi \ \ \forall i. \end{aligned} \tag{10}$$

In this formulation, the active-power dissipation $\mathcal{D}$ in (10) acts as a regularization function with $\beta \geq 0$ being a hyper-parameter. Note that the objective function $\mathcal{H}(\{|V_i|, |I_i|\})$ is only a function of the magnitudes of the voltages and currents and is independent of the phase angle ϕ_i. This ensures independent control of the magnitudes and the phases to achieve the desired objective of optimizing the active-power dissipation. This is shown in Fig. 3, where controlling the phase allows different paths from the initial to the final state, whereas evolution over the real domain allows only one possible path. The complex-domain approach thus results in steady-state limit cycle oscillations that encode the final solution. Compared with other complex-domain machine-learning frameworks [20], [23], the proposed formulation avoids nonlinear processing of phase/frequency that produces unwanted higher order harmonics. This would have made it difficult to maintain the network in the state of resonance (at a specific set of frequencies) under steady state.

The two important properties of the optimization framework in (10) are as follows.

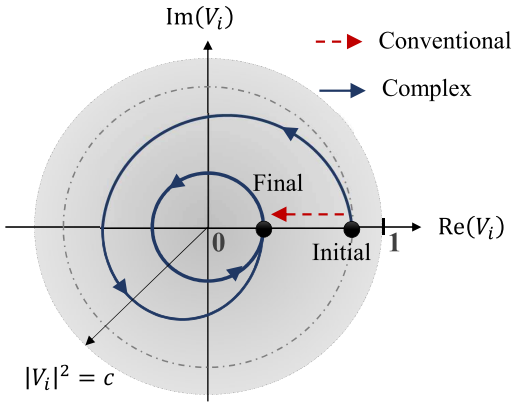

Fig. 3. Illustration showing that operating in the complex domain allows different possible learning trajectories from an initial state to the final steady state. Regularization with respect to the phase factor could then be used to select the trajectory with an optimal active-power dissipation profile and results in limit cycle oscillations in steady state. The circles indicate the constant magnitude loci.

1) For a convex cost function H, we have

$$\min_{\{|V_i|,|I_i|,\phi_i\}} \mathcal{L}(\{|V_i|, |I_i|, \phi_i\}) = \min_{\{|V_i|,|I_i|\}} \mathcal{H}(\{|V_i|, |I_i|\}). \tag{11}$$

This result follows from the three possible solutions of optimization problem (7).

2) If β is slowly annealed to a sufficiently high value, then $\phi_i \to \pi/2$ under steady state for $i : |V_i||I_i| \neq 0$. This implies that network achieves zero active power dissipation in steady state. Note that the method holds for nonconvex objective functions as well. In this case, however, the network might show resonant behavior at a locally optimal solution.

Example 1: Consider a single-variable quadratic optimization problem of the form $\mathcal{H}_1(x) = x^2$, subject to the constraint $|x| \leq 1$, $x \in \mathbb{R}$. Substituting $x = |V|^2 - |I|^2$, the problem can be mapped (please see Appendix B for more details) into the form equivalent to (10) as

$$\min_{\{|V|,|I|,\phi\}} \quad \mathcal{L}_1 = (|V|^2 - |I|^2)^2 + \beta|V|^2|I|^2\cos^2\phi \tag{12}$$

$$\text{s.t.} \quad |V|^2 + |I|^2 = 1, \quad |\phi| \leq \pi. \tag{13}$$

Fig. 4(a)–(c) plots $\mathcal{L}_1$ for different values of β. As shown in Fig. 4(a) and as expected for $\beta = 0$, the cost function has several minima (or attractors), whereas for $\beta > 0$, the minima corresponds to $\phi = \pm\pi/2$, for which the active-power dissipation is zero. Fig. 4(b) and (c) shows that controlling β will control the optimization landscape (without changing the location of the attractors) and will determine the attractor trajectory. This feature has been exploited in Sections IV and V to optimize the active-power dissipation profile during the learning phase.

IV. Complex Growth Transforms

The problem in (10) involves complex phasors and hence entails the use of learning models operating in the complex domain for reaching the optimal solution. To this end, in this section, we propose a dynamical system that can be used to

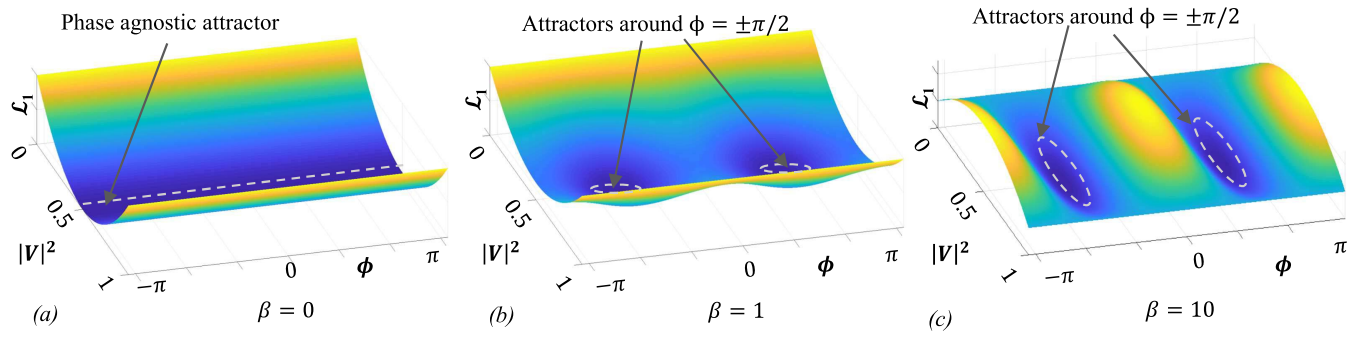

Fig. 4. Cost function $\mathcal{L}_1$ plotted for different values of the hyperparameter β. (a) $\beta = 0$. (b) $\beta = 1$. (c) $\beta = 10$.

Algorithm 1: Complex Growth Transform Dynamical System (Proof in Appendix C)

• For an optimization problem of the form:

$$\min_{\{|V_i|,|I_i|,\phi_i\}} \mathcal{L}(\{|V_i|,|I_i|,\phi_i\}) = \mathcal{H}(\{|V_i|,|I_i|\}) + \beta \sum_{i=1}^{N} |V_i|^2 |I_i|^2 \cos^2\phi_i$$

$$s.t. \quad \sum_{i=1}^{N} \left(|V_i|^2 + |I_i|^2\right) = 1, \;\; |\phi_i| \leq \pi \;\; \forall i = 1,\ldots,N, \;\; \beta \geq 0 \tag{14}$$

• If $\mathcal{H}(\{|V_i|,|I_i|\})$ is Lipschitz continuous in the domain $D = \{|V_i|,|I_i| : \sum_{i=1}^{N}\left(|V_i|^2 + |I_i|^2\right) = 1\}$, the following system of nonlinear dynamical equations

$$\frac{\partial V_i(t)}{\partial t} = j\omega\sigma_{V_i}(t)V_i(t) - \Delta\sigma_{V_i}(t)V_i(t), \tag{15}$$

$$\frac{\partial I_i(t)}{\partial t} = j(\omega + \omega_{\phi_i})\sigma_{I_i}(t)I_i(t) - \Delta\sigma_{I_i}(t)I_i(t), \tag{16}$$

$$\text{and } \tau_i\omega_{\phi_i} + \phi_i(t) = g_{\phi_i}(t) \;\; \forall i = 1,\ldots,N \tag{17}$$

$$\text{ensures that } \frac{\partial \mathcal{L}}{\partial t} \leq 0, \tag{18}$$

where $\sigma_{V_i}(t) = \sqrt{\frac{1}{V_i^*\eta}\left(-\frac{\partial \mathcal{L}}{\partial V_i} + \lambda V_i^*\right)}$, $\sigma_{I_i}(t) = \sqrt{\frac{1}{I_i^*\eta}\left(-\frac{\partial \mathcal{L}}{\partial I_i} + \lambda I_i^*\right)}$, $\omega_{\phi_i} = \frac{d\phi_i(t)}{dt}$, $\Delta\sigma_{V_i}(t) = 1 - \sigma_{V_i}(t)$, $\Delta\sigma_{I_i}(t) = 1 - \sigma_{I_i}(t)$ and $g_{\phi_i}(t) = \pi\dfrac{\lambda\phi_i - \pi\dfrac{\partial \mathcal{L}}{\partial \phi_i}}{-\phi_i\dfrac{\partial \mathcal{L}}{\partial \phi_i} + \lambda\pi}$, with $\eta = \sum_{k=1}^{N}\left(V_k\left[-\frac{\partial \mathcal{L}}{\partial V_k} + \lambda V_k^*\right] + I_k\left[-\frac{\partial \mathcal{L}}{\partial I_k} + \lambda I_k^*\right]\right)$, ω is an arbitrary angular frequency, and τ_i is the time-constant associated with the evolution of ϕ_i.

solve this optimization problem. The main result is summarized in Algorithm 1, and the details of the proof is given in Appendix C.

Theorem 1: *The system of nonlinear dynamical equations given by (15)–(18) in Algorithm 1 converge to the optimal point of (14) in the steady state, with zero energy dissipation, i.e.,* $\sum_{n=1}^{N}|V_i||I_i|\cos\phi_i = 0$ *(proof given in Appendix C).*

Consider an optimization problem that is an multivariable extension of Example 1, given by (12) and (13):

$$\min_{\{|V_i|,|I_i|,\phi_i\}} \mathcal{L}_N = \sum_{i=1}^{N}(|V_i|^2 - |I_i|^2)^2 + \beta\sum_{i=1}^{N}|V_i|^2|I_i|^2\cos^2\phi_i \tag{19}$$

$$\text{s.t.} \quad \sum_{i=1}^{N}(|V_i|^2 + |I_i|^2) = 1, \quad |\phi_i| \leq \pi \;\; \forall i. \tag{20}$$

The optimal solution is reached when $\phi_i = \pm(\pi/2)\;\forall i$, which implies $\mathcal{L}_N = 0$. For the sake of comparison, we will consider two variants: 1) the non-resonant model ($\mathcal{M}_{nr}$) where $\beta = 0$, and 2) the resonant model ($\mathcal{M}_r$) where $\beta \neq 0$. Fig. 5(a) and (b) shows a comparison of the cost $\mathcal{H}_N$ and the regularization metric $\mathcal{D}_N$ ($\sum_{i=1}^{N}|V_i|^2|I_i|^2$ for $\mathcal{M}_{nr}$ and $\sum_{i=1}^{N}|V_i|^2|I_i|^2\cos^2\phi_i$ for $\mathcal{M}_r$), for $N = 5$ and $\omega = \pi/10$. In the case of $\mathcal{M}_{nr}$, $\mathcal{L}_N = \mathcal{H}_N$ and in the case of $\mathcal{M}_r$, $\mathcal{L}_N = \mathcal{H}_N + \beta\mathcal{D}_N$, with $\beta = 1$. The solid line indicates the mean response across ten trials with random initializations of V_i, I_i, and ϕ_i. The shaded region shows the standard deviation across the trials in each case. Also shown in Fig. 5(b) are the true active-power dissipation profiles ($\sum_{n=1}^{N}|V_i||I_i|$ for $\mathcal{M}_{nr}$ and $\sum_{n=1}^{N}|V_i||I_i|\cos\phi_i$ for $\mathcal{M}_r$) over the ten trials.

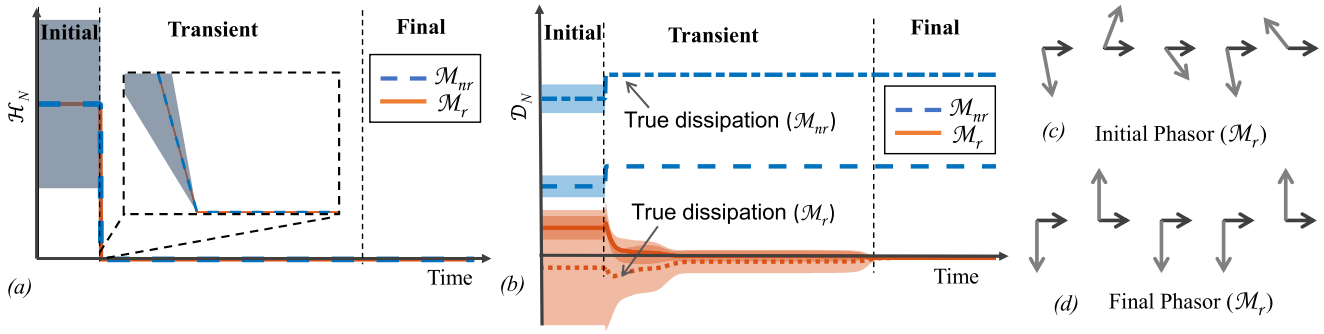

Fig. 5. Comparison of the resonant optimization model ($\mathcal{M}_r$) with its non-resonant variant ($\mathcal{M}_{nr}$) for a quadratic objective function $\mathcal{L}_N$ shown in (19). For this experiment, $N = 5$, input frequency $\omega = \pi/10$, and the mean response is estimated over ten trials with random initializations of $V_i, I_i, \phi_i \ \forall i = 1, \ldots, 5$. (a) Comparison of the time evolution of $\mathcal{H}_N$ for $\mathcal{M}_{nr}$ and $\mathcal{M}_r$. (b) Comparison of the time evolution of $\mathcal{D}_N$ ($\sum_{i=1}^{N} |V_i|^2 |I_i|^2$ for $\mathcal{M}_{nr}$ and $\sum_{i=1}^{N} |V_i|^2 |I_i|^2 \cos^2 \phi_i$ for $\mathcal{M}_r$). For all the curves, the line indicates the mean response, while the shaded region shows the standard deviation about the mean across the ten trials. The true dissipation ($\sum_{n=1}^{N} |V_i||I_i|$ for $\mathcal{M}_{nr}$ and $\sum_{n=1}^{N} |V_i||I_i| \cos \phi_i$ for $\mathcal{M}_r$) over the ten trials is also shown. Phasor representations of the LC tank voltages and currents for a single trial. (c) Initial configuration and (d) final configuration for $\mathcal{M}_r$. For $\mathcal{M}_r$, $\beta = 1$ for all the trials.

It can be observed that under steady-state conditions, the model $\mathcal{M}_{nr}$ dissipates power. However, for the model $\mathcal{M}_r$, the steady-state active power goes to zero. This is shown in Fig. 5(b). Fig. 5(c) shows the initial phasor configuration for the currents and voltages at each node of the network for a single trial for $\mathcal{M}_r$. Here, we assume the voltage phasor V_i to be aligned with the horizontal axis and the current phasor I_i to be at an angle ϕ_i with respect to the horizontal $\forall i$. Fig. 5(d) shows the steady-state phasor configuration for $\mathcal{M}_r$ for the same trial. The voltage and current phasors are orthogonal to each other for all the nodes, thus verifying the zero active-power dissipation.

In the next set of experiments, we annealed the hyper-parameter β and evaluated its impact on the active-power dissipation metric $\mathcal{D}_N$ and the convergence of the object function $\mathcal{H}_N$ for the model $\mathcal{M}_r$. Fig. 6 shows a comparison of the performance of $\mathcal{M}_r$ for different choices of annealing schedule for β, with the angular frequency $\omega = \pi/10$ as before. Fig. 6(a) shows the time evolution of the objective function $\mathcal{H}_N$, Fig. 6(b) shows the time evolution of the dissipation metric $\mathcal{D}_N$, and Fig. 6(c) shows the annealing schedules adopted for β. In all the cases, the optimization process starts after time $t = 0.1$ $a.u.$ from the onset of the simulation. The curves corresponding to $\beta = 1$ denote the case when β takes a constant value from $t = 0.1$ $a.u.$; $\beta =$ logistic corresponds to the case when β is slowly increased from $\beta_{\min} = 0$ following a logistic curve of the form $\beta(t) = \beta_{\min} + (\beta_{\max} - \beta_{\min})/(1 + \exp(-k(t + t_0)))$ from $t = 0.1$ $a.u.$ and takes on a maximum value of $\beta_{\max} = 1$ (k and t_0 are hyperparameters determining the steepness and midpoint of the sigmoid, respectively); $\beta =$ switching corresponds to the case when β switches from a minimum value ($\beta_{\min} = 0$) to a maximum value ($\beta_{\max} = 1$) at $t = 0.3$ $a.u.$, after the system has converged to the optimal solution. We can observe that in all the cases, the model converges to the optimal solution, irrespective of the choice of β. However, different annealing schedules for β lead to different active-power dissipation profiles. For example, a constant value of β throughout the duration of the experiment would lead to faster minimization of the active-power dissipation metric but at the cost of slower convergence. The opposite trend can be seen when β is slowly annealed to a sufficiently high value throughout the course of the optimization. The annealing schedule thus acts as a tradeoff between the speed of convergence and the rate of minimization of the active power.

A. Model Properties and Extensions

The dynamical system represented by (15)–(18) and the resonant optimization framework also exhibit the following properties and extensions.

1) *Energy Constraints Can Be Imposed Over Subgroups of Nodes in the Network:* We can have the reactive energy conservation constraint between subgroups of nodes, instead of on the network as a whole, i.e., $\sum_{i=1}^{N_k} (|V_{ik}|^2 + |I_{ik}|^2) = 1 \ \forall k = 1, \ldots, M$, where M is the number of subgroups and N_k = is the number of nodes in the kth subgroup. The update equations in this case are given by

$$\begin{aligned} \frac{\partial V_{ik}(t)}{\partial t} &= j\omega_k \sigma_{V_{ik}}(t) V_{ik}(t) - \Delta\sigma_{V_{ik}}(t) V_{ik}(t) \\ \frac{\partial I_{ik}(t)}{\partial t} &= j(\omega_k + \omega_{\phi_{ik}}) \sigma_{I_{ik}}(t) I_{ik}(t) \\ &\quad - \Delta\sigma_{I_{ik}}(t) I_{ik}(t) \\ \tau_{ik}\omega_{\phi_{ik}} + \phi_{ik}(t) &= g_{\phi_{ik}}(t) \quad \forall i, k \end{aligned} \tag{21}$$

where ω_k is the constant angular frequency of the kth subgroup of nodes, and $\omega_{\phi_{ik}} = \frac{d\phi_{ik}(t)}{dt}$.

2) *System Dynamics and Reactive-Energy Constraints Remain Invariant Under the Imposition of a Global Phase:* The network dynamics remain invariant to the imposition of a global phase component on all the network variables, and the conservation constraint is also satisfied in this case. The governing equations are given by

$$\begin{aligned} \frac{\partial V_i(t)}{\partial t} &= j(\omega + \omega_{\phi_g}) \sigma_{V_i}(t) V_i(t) - \Delta\sigma_{V_i}(t) V_i(t), \\ \frac{\partial I_i(t)}{\partial t} &= j(\omega + \omega_{\phi_g} + \omega_{\phi_i}) \sigma_{I_i}(t) I_i(t) - \Delta\sigma_{I_i}(t) I_i(t) \end{aligned} \tag{22}$$

where ϕ_g is the global phase and $\omega_{\phi_g} = \frac{d\phi_g}{dt}$.

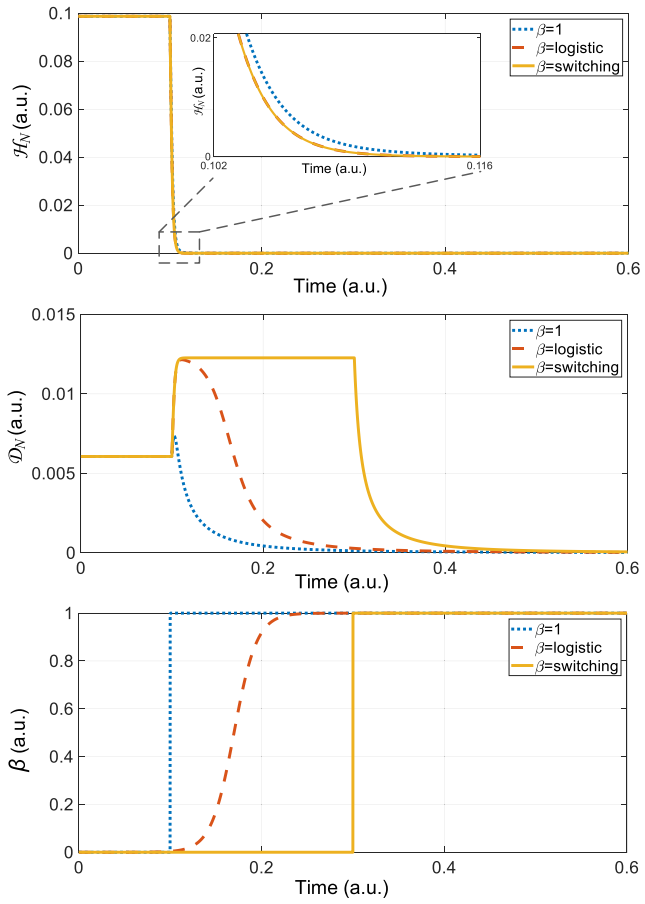

Fig. 6. Comparison of the performance of the resonant model $\mathcal{M}_r$ for different choices of annealing schedule for β ($N = 5, \omega = \pi/10$). (a) Time evolution of $\mathcal{H}_N$ (inset shows a zoomed-in view of the cost evolution in the transient phase). (b) Time evolution of $\mathcal{D}_N$. (c) Time evolution of β. In all the cases, the optimization process starts after 0.1 *a.u.* from the onset of the simulation. The curves corresponding to $\beta = 1$ denotes the case when β takes a constant value from $t = 0.1$ *a.u.*, $\beta =$ logistic corresponds to the cases when β is slowly increased following a logistic curve from $t = 0.1$ *a.u.* and takes on a maximum value of $\beta = 1$, and $\beta =$ switching corresponds to the case when β switches from a minimum value(= 0) to a maximum value(= 1) at $t = 0.3$ *a.u.*, after the system has converged to the optimal solution.

3) *Reactive-Energy Constraints Remain Invariant With Varying Network Dynamics Under the Imposition of a Relative Phase:* The conservation constraints are satisfied on the introduction of a relative phase component between the voltage and current phasors of each node, even though the overall network dynamics change. The governing equations are given by

$$\begin{aligned}\frac{\partial V_i(t)}{\partial t} &= j(\omega+\omega_{\phi_{V_i}})\sigma_{V_i}(t)V_i(t) - \Delta\sigma_{V_i}(t)V_i(t),\\ \frac{\partial I_i(t)}{\partial t} &= j(\omega+\omega_{\phi_{I_i}}+\omega_{\phi_i})\sigma_{I_i}(t)I_i(t) - \Delta\sigma_{I_i}(t)I_i(t)\end{aligned} \tag{23}$$

where $\phi_i = \phi_{I_i} - \phi_{V_i}$ is the relative external phase shift applied between the voltage and current phasors of the ith node, $\omega_{\phi_{V_i}} = \frac{d\phi_{V_i}}{dt}$, and $\omega_{\phi_{I_i}} = \frac{d\phi_{I_i}}{dt}$.

4) *Model Is Dissipative and Converges to Limit Cycle Oscillations in Steady State:* The second-order time derivatives of (15) and (16) lead to the following:

$$\begin{aligned}\frac{\partial^2 V_i}{\partial t^2} &= j\omega\sigma_{V_i}\dot{V}_i + j\omega V_i\dot{\sigma}_{V_i} - \Delta\sigma_{V_i}\dot{V}_i + \dot{\sigma}_{V_i}V_i\\ &= \underbrace{-\omega^2\sigma_{V_i}^2 V_i}_{\text{limit cycles}}\\ &\quad + \underbrace{[1+j\omega]\dot{\sigma}_{V_i}V_i - 2j\omega\sigma_{V_i}\Delta\sigma_{V_i}V_i + (\Delta\sigma_{V_i})^2 V_i}_{\text{dissipation}}\end{aligned} \tag{24}$$

$$\begin{aligned}\frac{\partial^2 I_i}{\partial t^2} &= j\omega'_{\phi_i}\sigma_{I_i}\dot{I}_i + j\omega'_{\phi_i}I_i\dot{\sigma}_{I_i} - \Delta\sigma_{I_i}\dot{I}_i + \dot{\sigma}_{I_i}I_i\\ &= \underbrace{-\omega'^2_{\phi_i}\sigma_{I_i}^2 I_i}_{\text{limit cycles}}\\ &\quad + \underbrace{[1+j\omega'_{\phi_i}]\dot{\sigma}_{I_i}I_i - 2j\omega'_{\phi_i}\sigma_{I_i}\Delta\sigma_{I_i}I_i + (\Delta\sigma_{I_i})^2 I_i}_{\text{dissipation}}\end{aligned} \tag{25}$$

where $\omega'_{\phi_i} = \omega + \omega_{\phi_i}$. The first terms in the right-hand side of (24) and (25) correspond to stable limit cycle oscillations of all the phasors, whereas the other terms correspond to the dissipative effects in the network. This demonstrates that the network as a whole is essentially a coupled dissipative system that is capable of self-sustained oscillations under steady state. Each individual state variable describing the network thus returns to the same position in its respective limit cycle at regular intervals of time, even when subjected to small perturbations.

V. Resonant Machine-Learning Framework

In this section, we show how the framework introduced in Section III can be applied for constructing resonant machine-learning networks. In general, the framework can be applied to any learning network that optimizes a cost function defined over a set of learning variables α_i as

$$\begin{aligned}&\min_{\{\alpha_i\}} \quad \mathcal{H}(\{\alpha_i\}) + h\Psi(\{\alpha_i\})\\ &\text{s.t.} \quad \sum_{i=1}^{N}\alpha_i = 1, \quad \alpha_i \geq 0 \quad \forall i = 1,\ldots,N\end{aligned} \tag{26}$$

where $\mathcal{H}(\{\alpha_i\})$ represents a loss function [28], which depends on the learning problem (e.g., supervised, unsupervised, or semisupervised) and the dataset under consideration (e.g., training data). The second term $\Psi(\cdot)$ in the objective function is any linear or nonlinear function which represents either a regularization function or a penalty function used to satisfy optimization constraints. h is a hyperparameter that acts as a tradeoff between $\mathcal{H}(\cdot)$ and $\Psi(\cdot)$. Because α_i could be viewed as probability measure, the optimization framework in (26) naturally lends itself to probabilistic learning models [29]–[31].

The problem above can be mapped to the resonant learning framework in Section III by substituting

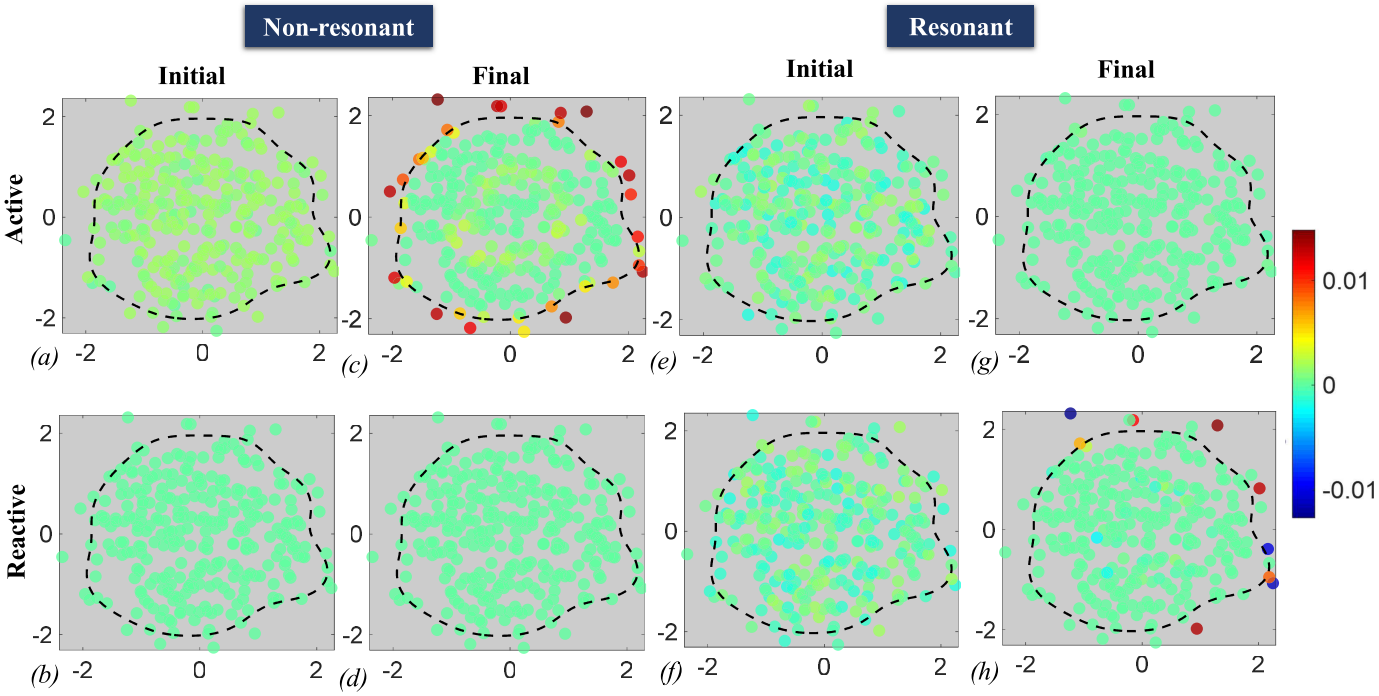

Fig. 7. Comparison of the active and reactive power dissipated at each node for the non-resonant model $\mathcal{M}_{nr}$ and the resonant model $\mathcal{M}_r$ for the synthetic one-class SVM problem on a two-dimensional dataset (Dataset I), with $N = 300$, $\nu = 0.1$, $\omega = \pi/4$, and random initializations for V_i, I_i, ϕ_i $\forall i = 1, \ldots, N$. (a) and (c) Values of the active power metric ($= |V_i||I_i|$) at each node in the initial and final stages, respectively, for $\mathcal{M}_{nr}$. (b) and (d) Values of the reactive power metric ($= 0$) at each node in the initial and final stages, respectively, for $\mathcal{M}_{nr}$. (e) and (g) Values of the active power metric ($= |V_i||I_i|\cos\phi_i$) at each node in the initial and final stages, respectively, for $\mathcal{M}_r$. (f) and (h) Values of the reactive power metric ($= |V_i||I_i|\sin\phi_i$) at each node in the initial and final stages, respectively, for $\mathcal{M}_r$. For both models, $K(\cdot,\cdot)$ was chosen to be a Gaussian kernel with kernel parameter $\sigma = 1$, and $\beta = 1$ throughout the optimization process for $\mathcal{M}_r$.

$\alpha_i = |V_i|^2 + |I_i|^2$, to arrive at the following problem:

$$\begin{aligned}\min_{\{|V_i|,|I_i|,\phi_i\}} \quad & \mathcal{H}(\{|V_i|,|I_i|\}) + h\Psi(\{|V_i|,|I_i|\}) \\ & + \beta\sum_{i=1}^{N}|V_i|^2|I_i|^2\cos^2\phi_i \\ \text{s.t.} \quad & \sum_{i=1}^{N}(|V_i|^2+|I_i|^2)=1, \quad |\phi| \le \pi \quad \forall i=1,\ldots,N. \end{aligned} \tag{27}$$

Note that non-probabilistic learning problems can also be mapped to the probabilistic framework by imposing an additional constraint, as discussed in Appendix A.

A. One-Class Resonant SVM

We now show how the framework in (27) can be used to design a resonant one-class support vector machine.

The solution of a generic one-class SVM is obtained by solving the following optimization problem [8], [32], [33]:

$$\begin{aligned}\min_{\{\alpha_i\}} \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i K(\mathbf{x_i},\mathbf{x_j})\alpha_j \\ \text{s.t.} \quad & \sum_{i=1}^{N}\alpha_i = 1, \quad 0 < \alpha_i < \frac{1}{\nu N} \quad \forall i = 1,\ldots,N \end{aligned} \tag{28}$$

where $\mathbf{X} = [\mathbf{x_1}, \ldots, \mathbf{x_i}, \ldots, \mathbf{x_N}] \in \mathbb{R}^{N\times D}$ is the D-dimensional input dataset of size N, $\nu \in \{0, 1\}$ is a parameter that controls the size of the decision surface, $K(\cdot,\cdot)$ is a positive-definite kernel function satisfying Mercer's conditions, and α_i values are the learning parameters. The optimization problem above can be reformulated by replacing the inequality constraint with a smooth penalty or loss function $\Psi(\cdot)$ such as the logarithmic barrier, e.g., $\Psi(\alpha_i, \nu) = -\log\left(\frac{1}{\nu N} - \alpha_i\right)$

$$\begin{aligned}\min_{\{\alpha_i\}} \quad & \mathcal{H} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i K(\mathbf{x_i},\mathbf{x_j})\alpha_j + h\sum_{i=1}^{N}\Psi(\alpha_i,\nu) \\ \text{s.t.} \quad & \sum_{i=1}^{N}\alpha_i = 1. \end{aligned} \tag{29}$$

The parameter h determines the steepness of the penalty function, where a lower value of h implies an almost-accurate inequality constraint.

An equivalent complex-domain representation in terms of voltages and currents in an LC network can be arrived at if we consider $\alpha_i = |V_i|^2 + |I_i|^2$ $\forall i$. In this case, we consider that the network is globally energy constrained, and all the individual units in the network have the same frequency ω. The redefined learning problem is as follows:

$$\begin{aligned}\min_{\{|V_i|,|I_i|\}} \quad & \mathcal{H} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(|V_i|^2+|I_i|^2)K(\mathbf{x_i},\mathbf{x_j}) \\ & \times(|V_j|^2+|I_j|^2) + h\sum_{i=1}^{N}\Psi(|V_i|,|I_i|,\nu) \\ \text{s.t.} \quad & \sum_{i=1}^{N}(|V_i|^2+|I_i|^2) = 1. \end{aligned} \tag{30}$$

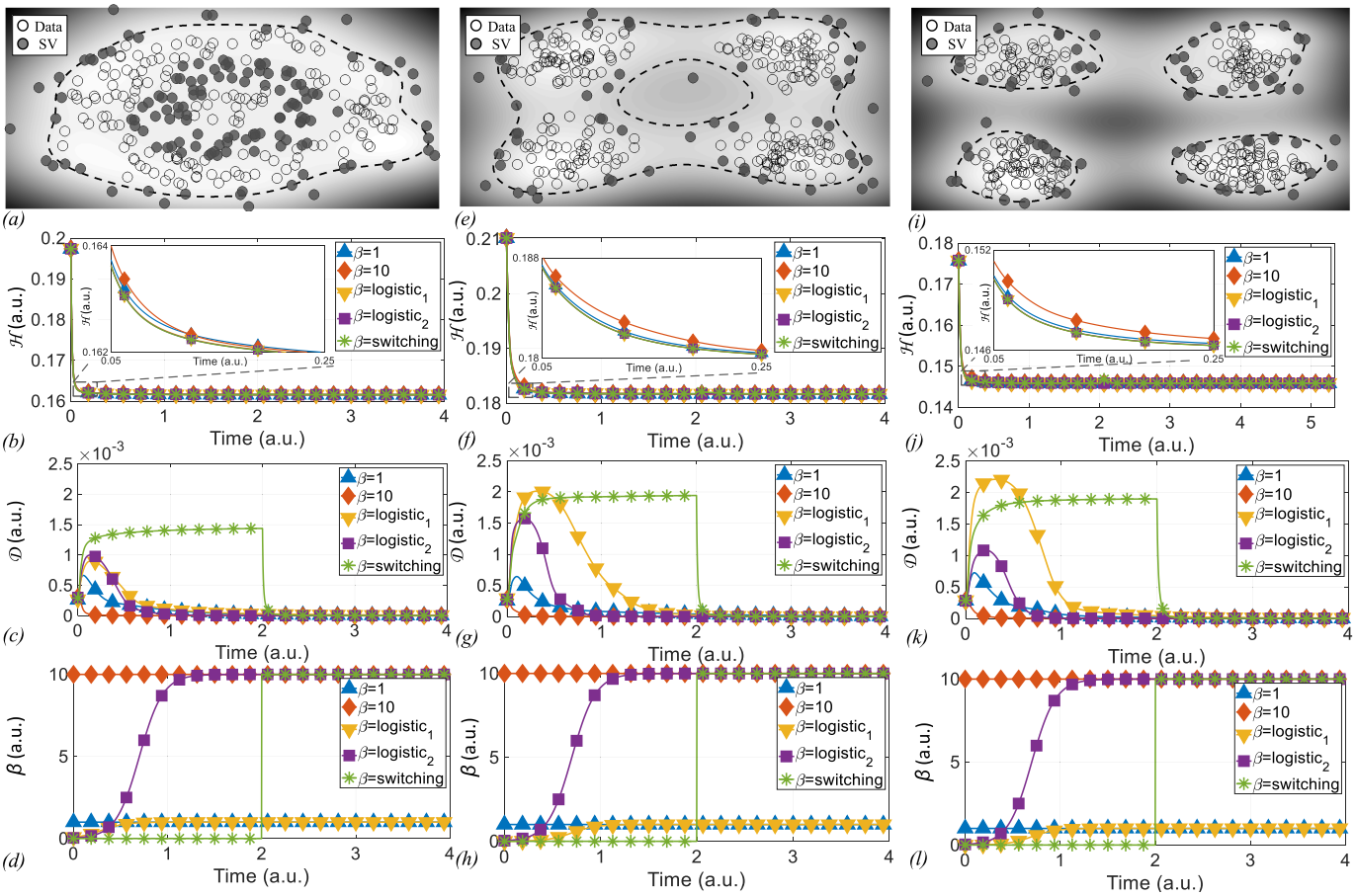

Fig. 8. Comparison of the performance of the resonant model $\mathcal{M}_r$ for different choices of annealing schedules for β for the one-class SVM problem, on three different two-dimensional synthetic datasets for a simulation duration of $t = 4\ a.u.$ In all cases, $N = 300, \nu = 0.1, \omega = \pi/20$, and $K(\cdot,\cdot)$ was chosen to be a Gaussian kernel with parameter values $\sigma = 1$, 10 and 20 for synthetic Datasets I-III, respectively. In addition, for each dataset, V_i, I_i, ϕ_i were randomly initialized $\forall i = 1, \ldots, N$. (a) Contour plot with the decision boundary around the data points and SVs. (b) Time evolution of $\mathcal{H}$ (inset shows a zoomed-in view of the cost evolution in the transient phase). (c) Time evolution of $\mathcal{D}$. (d) Time evolution of β for different annealing schedules. The curves corresponding to $\beta = 1$ and 10 denote the cases when β takes a constant value throughout the simulation duration; $\beta = \text{logistic}_1$ and $\beta = \text{logistic}_2$ correspond to the cases when β is slowly increased following a logistic curve and takes on maximum values of $\beta_{\max} = 1$ and $\beta_{\max} = 10$, respectively; $\beta = \text{switching}$ corresponds to the case when β switches from a minimum value ($\beta_{\min} = 0$) to a maximum value ($\beta_{\max} = 10$) at $t = 2\ a.u.$, after the system has converged to the optimal solution. (e)–(h) Similar plots on Dataset II. (i)–(l) Plots corresponding to Dataset III.

Introducing the active-power dissipation regularization, we arrive at the following problem:

$$\min_{\{|V_i|,|I_i|,\phi_i\}} \quad \mathcal{L} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(|V_i|^2 + |I_i|^2)K(\mathbf{x_i}, \mathbf{x_j}) \times (|V_j|^2 + |I_j|^2) + h\sum_{i=1}^{N}\Psi(|V_i|, |I_i|, \nu) + \beta\sum_{i=1}^{N}|V_i|^2|I_i|^2\cos^2\phi_i$$

$$\text{s.t.} \quad \sum_{i=1}^{N}(|V_i|^2 + |I_i|^2) = 1, \quad |\phi_i| \leq \pi \ \ \forall i. \tag{31}$$

The update equations in this case are of the form shown in (15)–(18). Fig. 7 shows a comparison of the active and reactive power metrics of each node of the non-resonant model $\mathcal{M}_{nr}$ and the resonant model $\mathcal{M}_r$ for a synthetic one-class SVM problem on a two-dimensional dataset (Dataset I). The dataset was generated by uniformly selecting 300 random points within a circle having a fixed radius. Here, $N = 300, \nu = 0.1$, and $\omega = \pi/4$, with random initializations for $V_i, I_i, \phi_i\ \forall i = 1, \ldots, N$. A constant value of the regularization hyperparameter $\beta = 1$ was considered throughout the duration of the optimization process for $\mathcal{M}_r$. Fig. 7(a) and (c) shows the values of the active power metric ($= |V_i||I_i|$) at each node in the initial and final stages, respectively, for $\mathcal{M}_{nr}$, whereas Fig. 7(b) and (d) shows the values of the reactive power metric ($= 0$) at each node in the initial and final stages, respectively, for $\mathcal{M}_{nr}$. Similarly, Fig. 7(e) and (g) shows the values of the active power metric ($= |V_i||I_i|\cos\phi_i$) at each node in the initial and final stages, respectively, for $\mathcal{M}_r$ whereas Fig. 7(f) and (h), finally, shows the values of the reactive power metric ($= |V_i||I_i|\sin\phi_i$) at each node in the initial and final stages, respectively, for $\mathcal{M}_r$.

Fig. 8 shows a comparison of the performance of the resonant model $\mathcal{M}_r$ for different choices of annealing schedules for β on three different two-dimensional synthetic datasets for a simulation duration of $t = 4\ a.u.$ In all cases, $N = 300, \nu = 0.1, \omega = \pi/20$, and $K(\cdot,\cdot)$ was chosen to be a Gaussian kernel with parameter values

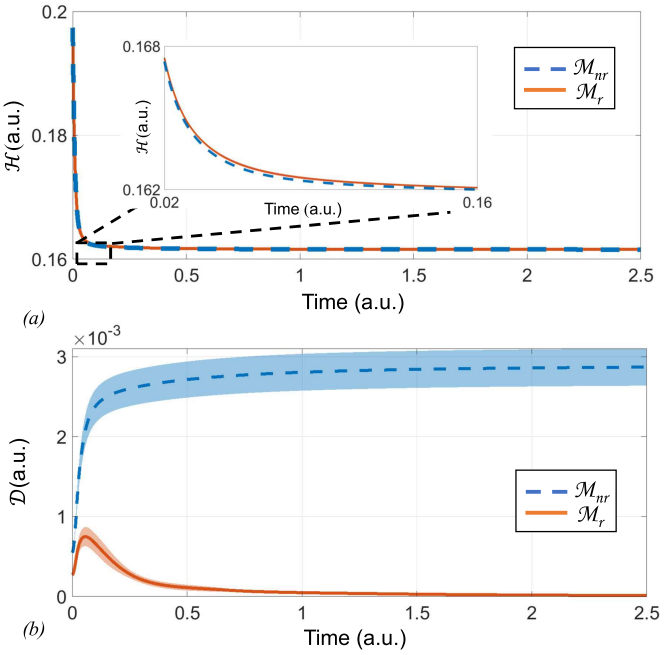

Fig. 9. Robustness to random initialization. Comparison of (a) time evolution of the cost $\mathcal{H}$ and (b) dissipation metric profile $\mathcal{D}$ ($\sum_{n=1}^{N}|V_i|^2|I_i|^2$ for the non-resonant model $\mathcal{M}_{nr}$ and $\sum_{n=1}^{N}|V_i|^2|I_i|^2\cos^2\phi_i$ for the resonant model $\mathcal{M}_r$, respectively) for the synthetic one-class SVM problem on Dataset I for $N = 300$, $\nu = 0.1$, and $\omega = \pi/8$ over ten random initializations for V_i, I_i, ϕ_i $\forall i = 1, \ldots, N$. The regularization parameter was chosen to be $\beta = 1$ for the entire simulation duration of $t = 4$ $a.u.$ of the optimization process for $\mathcal{M}_r$. For all the curves, the solid line indicates the mean response, whereas the shaded region shows the standard deviation about the mean across the trials.

$\sigma = 1, 10$, and 20, respectively, for synthetic Datasets I, II and III. In addition, for each dataset, V_i, I_i, and ϕ_i were randomly initialized $\forall i = 1, \ldots, N$. Fig. 8(a) shows the contour plot with the decision boundary around the data points, along with the support vectors (SVs) for Dataset I, whereas Fig. 8(b)–(d) shows the time evolutions of the cost $\mathcal{H}$, the dissipation metric $\mathcal{D}$, and the hyperparameter β for different annealing schedules. The curves corresponding to $\beta = 1$ and 10 denote the cases when β takes a constant value throughout the simulation duration. $\beta = \text{logistic}_1$ and $\beta = \text{logistic}_2$ correspond to the cases when β is slowly increased following a logistic curve of the form $\beta(t) = \beta_{\min} + (\beta_{\max} - \beta_{\min})/(1 + \exp(-k(t + t_0)))$ and takes on maximum values of $\beta_{\max} = 1$ and $\beta_{\max} = 10$, respectively, starting from $\beta_{\min} = 0$. Finally, $\beta =$ switching corresponds to the case when β switches from a minimum value ($\beta_{\min} = 0$) to a maximum value ($\beta_{\max} = 10$) at $t = 2$ $a.u.$, after the system has converged to the optimal solution. Fig. 8(e)–(h) shows the similar plots for Dataset II, whereas Fig. 8(i)–(l) shows the plots corresponding to Dataset III. Dataset I, as described before, was generated by selecting 300 points uniformly within a two-dimensional circle having a fixed radius. Datasets II and III also consist of 300 data points generated using a Gaussian mixture model consisting of four different clusters, with a fixed cluster mean and variance associated with each cluster. Datasets II and III differ only in terms of the cluster means associated with their respective

TABLE II
PERFORMANCE ON REAL BENCHMARK DATASETS FOR $\nu = 0.1$

Dataset	Majority Class Size	Correct	Outliers	SVs
Iris	(50,4)	48	2	8
Heart	(139,13)	1210	18	56
Diabetes	(500,4)	466	34	90
Ecoli	(327,7)	309	18	42
Adulta3a	(2412,123)	2111	301	356
Mammography	(10923,6)	9608	1315	5969

constituent clusters, while the cluster variances are the same for all the clusters for both the datasets. It can be seen that since the optimization problem is convex, the model always converges to the optimal solution for every dataset, irrespective of the annealing schedule for β or the dataset complexity. However, the dissipation profiles corresponding to a particular annealing schedule strongly depend on the complexity of the dataset. In general, however, higher values of β would lead to lower dissipation profiles during the learning process. Also, the model shows a much slower convergence in terms of the actual objective function for a constant nonzero value of β throughout the optimization, compared with the case when β is slowly annealed to a sufficiently high value. The choice of a proper annealing schedule for β would thus involve a tradeoff between the speed of convergence and the rate of power dissipation.

Finally, Fig. 9(a) and (b) shows the robustness of the proposed framework to different initial conditions by providing a comparison of the time evolutions of the cost $\mathcal{H}$ and the dissipation metric $\mathcal{D}$ ($\sum_{n=1}^{N}|V_i|^2|I_i|^2$ for the non-resonant model $\mathcal{M}_{nr}$ and $\sum_{n=1}^{N}|V_i|^2|I_i|^2\cos^2\phi_i$ for the resonant model $\mathcal{M}_r$), when applied to Dataset I. We show the results for ten random initializations of V_i, I_i and ϕ_i, $\forall i$. In all cases, $\nu = 0.1, \omega = \pi/8$, and $K(\cdot,\cdot)$ was chosen to be a Gaussian kernel with parameter value $\sigma = 1$. Note that even though the ten trails had different initializations of V_i and I_i, they were chosen to have the same initial value of α_i's in all cases, because of which there is no deviation between the cost evolution curves for both $\mathcal{M}_{nr}$ and $\mathcal{M}_r$. The dissipation evolution is, however, different across different trials for both the models. However, the dissipation attains a final value of zero for $\mathcal{M}_r$ for all the trials, while there is a finite dissipation for $\mathcal{M}_{nr}$ in all cases.

We also conducted experiments on real-life benchmark datasets of varying sizes and dimensionality, considering the majority class as inliers in all cases. In all the experiments, we used similar parameter settings and annealing schedules as used in the experiments shown in Fig. 8, and $\nu = 0.1$. Table II shows the dataset description along with the performance [number of inliers, number of outliers, and number of support vectors (SVs)] of the one-class SVM classifier for the following datasets: "Iris" (inlier: class "setosa"),

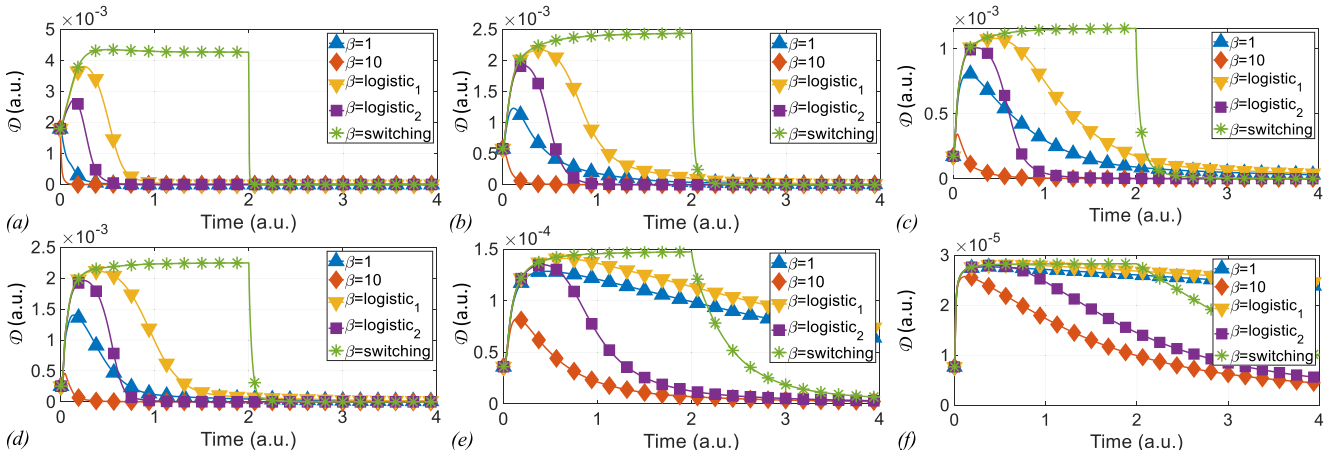

Fig. 10. Comparison of the dissipation profiles of the resonant model $\mathcal{M}_r$ for different choices of annealing schedules for β for different real-life benchmark datasets. (a) 'Iris'. (b) 'Heart'. (c) 'Diabetes'. (d) 'Ecoli'. (e) 'Adulta3a'. (f) 'Mammography'.

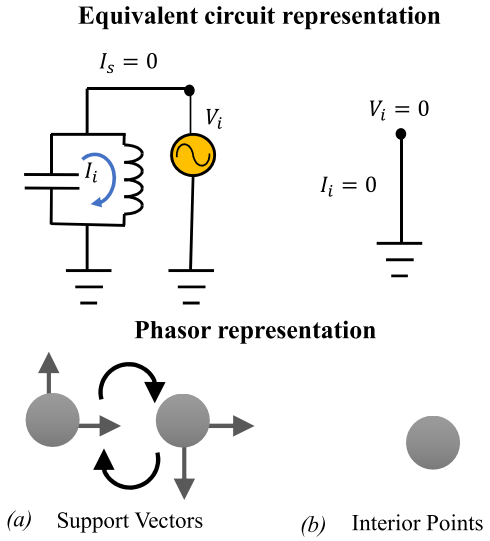

Fig. 11. Circuit- and phasor-based representations for a one-class SVM problem. (a) Support vectors, corresponding to resonant LC tanks. (b) Interior points, corresponding to sinks/ground ($V_i, I_i = 0$).

"Heart" (inlier: "healthy" heart), "Diabetes" (inlier: non-diabetic patients), "Ecoli" (inlier: classes "cp," "im," 'pp," "imU," and "om"), "Adulta3a" (inlier: individuals with income $\leq$ \$50K), and "Mammography" (inlier: non-calcifications) [34], [35]. Fig. 10(a)–(f) shows the dissipation profiles corresponding to different annealing schedules for the different datasets. It can be seen that an optimal choice of the annealing schedule depends on both the dataset size and complexity, even though the dissipation decreases over time for all the cases, irrespective of the schedule chosen.

Interestingly, the solution of the one-class SVM can be interpreted in terms of an electrical network as follows, as shown in Fig. 11: 1) the support vectors have voltage and current magnitudes with a $\pm\pi/2$ phase shift between them, and hence can be interpreted as resonant LC tanks; and 2) the interior points well inside the boundary have both zero voltage and current magnitudes and can be essentially treated as floating sinks.

VI. Conclusion and Discussion

In this article, we proposed a complex-domain formulation of a machine-learning network that ensures that the network's active power is dissipated only during the process of learning, whereas the network's reactive power is maintained to be zero at all times. We showed that the active power dissipation during learning can be controlled using a phase regularization parameter. Also, the framework is robust to variations in the initial conditions and to the choice of the input/driving frequency ω. The proposed approach thus provides a physical interpretation of machine-learning algorithms, where the energy required for storing learned parameters is sustained by electrical resonances due to nodal inductances and nodal capacitances. Using the one-class support vector machine problem as a case study, we have shown how the steady-state solution of a learning problem can be interpreted in terms of these nodal circuit elements.

Future directions involve exploring the implications of incorporating network dynamics in the free frequency variable ω, and utilizing the phase information associated with each node in the learning process. Also, the experimental results presented in this article were based on an unsupervised learning setting. Exploring an energy-efficient framework for supervised learning problems would also be a part of our future work.

In this paper, we also proposed a dynamical system model based on complex-domain growth transforms. The formulation is general enough to be applicable to other complex-domain learning models [19]–[23]. Our proposed framework also preserves both the magnitude and phase information and provides additional flexibility compared with other complex-domain learning models in terms of phase manipulation/cancellation [20].

In addition to implementing classic machine learning algorithms, the complex growth transform dynamical system can also be used for designing synchronized networks of coupled oscillators. Such networks can be potentially used for solving different computing tasks such as optimization, pattern matching etc., as is achievable using coupled oscillatory computing

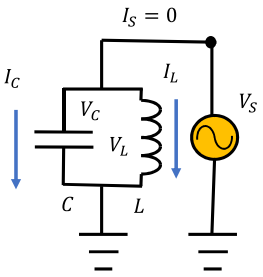

Fig. 12. LC tank resonator.

models [36], [37]. An oscillator network designed in this fashion is capable of demonstrating stable, self-sustaining oscillatory dynamics, whereby the network can return to its initial stable limit cycle configuration following small perturbations, while simultaneously minimizing some underlying system-level objective function. The framework could also be used to study connections between learning and synchronization, or the emergence of a rhythmic periodic pattern exhibited by a group of coupled oscillators, which provides the key to understanding periodic processes pervading complex networks of different biological, physical, social, and quantum ensembles [9], [38]. In this regard, the existing mathematical models for such collective behavior are mostly phenomenological or bottom–up, and in general, do not provide a network-level perspective of the underlying physical process. The proposed growth transform formulation, thus, could provide new network-level insights into the emergence of phase synchronization, phase cohesiveness, and frequency synchronization in coupled-oscillator networks.

Note here that since we are implicitly assuming an analog implementation, the learning network should converge to the steady-state solution of the optimization, where the time of convergence depends on the network's time constant [39]. Hence, the notion of time complexity is not well defined. However, in a digital implementation, the time complexity would depend on the learning algorithm under consideration, dataset size and dimensionality, angular frequency ω, and time constants τ_i for the phase updates.

APPENDIX

A. Resonance in an LC Tank

Consider the parallel LC tank circuit shown in Fig. 12, with V_C and V_L being the voltages across the capacitor C and inductor L, respectively. I_C and I_L denote the corresponding currents flowing through the elements. Thus, $V_S = V_L = V_C$ and $I_S = I_L + I_C$. Considering the LC tank to be driven by the voltage source V_S at frequency ω, we have the following condition in steady state:

$$I_S(\omega) = \frac{V_S(\omega)}{j\omega L}[1 - \omega^2 LC]. \tag{32}$$

Resonant condition of the circuit is achieved when

$$\omega = \frac{1}{\sqrt{LC}} \implies I_S(\omega) = 0. \tag{33}$$

This implies that the apparent power $S_N = P_N + jQ_N = V_S I_S^* + V_L I_L^* + V_C I_C^*$, where the active power $P_N = 0$. In addition, at resonance, the reactive power $Q_N = Q_C + Q_L = V_L I_L^* + V_C I_C^* = -(j/\omega L)|V(\omega)|^2 + (j/\omega L)|V(\omega)|^2 = 0$, where Q_C and Q_L are the reactive powers associated with the capacitance and inductance, respectively.

B. Mapping a Generic Optimization Problem to the Equivalent Network Model

Let us consider an optimization problem defined over a probabilistic domain, given by the following generic form:

$$\min_{\{x_i\}} \ \mathcal{H}(\{x_i\}) \quad \text{s.t.} \ \sum_{i=1}^{N} x_i = 1, \ x_i \geq 0. \tag{34}$$

We can map the above to the electrical network-based model introduced in Section III by replacing $x_i = |V_i|^2 + |I_i|^2$, which leads to the following problem in the $\{|V_i|^2, |I_i|^2\}$ domain:

$$\min_{\{|V_i|, |I_i|\}} \ \mathcal{H}(\{|V_i|, |I_i|\}) \quad \text{s.t.} \ \sum_{i=1}^{N} (|V_i|^2 + |I_i|^2) = 1. \tag{35}$$

Note that the method also works for optimization problems defined over non-probabilistic domains of the following form:

$$\min_{\{x_i\}} \ \mathcal{H}(\{x_i\}) \quad \text{s.t.} \ |x_i| \leq 1, \ x_i \in \mathbb{R} \ \ \forall i = 1, \ldots, N. \tag{36}$$

This can be done by considering $x_i = x_i^+ - x_i^- \ \forall i$, where both $x_i^+, x_i^- \geq 0$. Since by triangle inequality, $|x_i| \leq |x_i^+| + |x_i^-|$, enforcing $x_i^+ + x_i^- = 1 \ \forall i$ would automatically ensure $|x_i| \leq 1 \ \forall i$. Thus we have

$$\underset{\{x_i\}}{\text{argmin}} \ \mathcal{H}(\{x_i\}) \equiv \underset{\{x_i^+, x_i^-\}}{\text{argmin}} \ \mathcal{H}\left(\{x_i^+, x_i^-\}\right) \quad \text{s.t.} \ |x_i| \leq 1, \ x_i \in \mathbb{R} \qquad \text{s.t.} \ x_i^+ + x_i^- = 1, \ x_i^+, x_i^- \geq 0. \tag{37}$$

In this case, we replace $x_i^+ = |V_i|^2$, $x_i^- = |I_i|^2$, and the equivalent problem in the $\{|V_i|^2, |I_i|^2\}$ domain is thus as follows:

$$\min_{\{|V_i|, |I_i|\}} \ \mathcal{H}(\{|V_i|, |I_i|\}) \quad \text{s.t.} \ |V_i|^2 + |I_i|^2 = 1 \ \ \forall i = 1, \ldots, N. \tag{38}$$

For example, the variables $\{x_i\}$ can represent the Lagrangian multipliers in the primal space of a support vector machine network, or the weights and biases of a generic neural network.

C. Complex Growth Transform Dynamical System (Algorithm 1)

Let us consider the optimization problem in (34) again. We can use the Baum–Eagon inequality [40], [41] to converge

to the optimal point of $\mathcal{H}$ in steady state, by using updates of the form

$$x_i \leftarrow \frac{x_i\left(-\frac{\partial\mathcal{H}(\{x_i\})}{\partial x_i}+\lambda\right)}{\sum_{k=1}^{N} x_k\left(-\frac{\partial\mathcal{H}(\{x_k\})}{\partial x_k}+\lambda\right)}. \tag{39}$$

where $\mathcal{H}$ is assumed to be Lipschitz continuous [7] on the domain $D=\{x_1,\ldots,x_N : \sum_{i=1}^{N} x_i=1,\ x_i\geq 0\ \forall i\}\subset\mathbb{R}_+^N$. The constant $\lambda\in\mathbb{R}_+$ is chosen such that $|-\frac{\partial\mathcal{H}(\{x_i\})}{\partial x_i}+\lambda|>0\ \forall i$.

We can solve the optimization problem given by (10) by using the growth transforms discussed earlier. The outline of the proof is as follows.

1) We will start with a generic magnitude domain optimization problem without any phase regularizer and derive the form for the growth transform dynamical system which would converge to the optimal point asymptotically.
2) We derive a complex-domain counterpart of the above, again without phase constraints.
3) We derive the complex-domain dynamical system by incorporating a phase regularizer in the objective function.

Since the time evolutions of V_i and I_i are symmetric because of the conservation constraints, for the rest of the section, we will consider only the update equations for the voltages V_i, and similar results would also apply to the updates for I_i.

1) ***Condition 1****: Considering $\beta=0$ in (10) and $\mathcal{H}(\{|V_i|,|I_i|\})$ to be Lipschitz continuous over the domain $D=\{|V_i|,|I_i| : \sum_{i=1}^{N}(|V_i|^2+|I_i|^2)=1\}$, we can use the growth transforms to arrive at the following discrete-time update equations in terms of the voltage and current magnitudes:*

$$|V_{i,n}|^2 \leftarrow g_{V_i,n-1}(\{|V_{i,n-1}|^2,|I_{i,n-1}|^2\}) \tag{40}$$

where

$$g_{V_i,n-1}(\{|V_{i,n-1}|^2,|I_{i,n-1}|^2\}) = \frac{|V_{i,n-1}|^2}{\mu_{n-1}}\left(-\frac{\partial\mathcal{H}}{\partial|V_{i,n-1}|^2}+\lambda\right) \tag{41}$$

$$\mu_{n-1}=\sum_{k=1}^{N}\left(|V_{k,n-1}|^2\left[-\frac{\partial\mathcal{H}}{\partial|V_{k,n-1}|^2}+\lambda\right] + |I_{k,n-1}|^2\left[-\frac{\partial\mathcal{H}}{\partial|I_{k,n-1}|^2}+\lambda\right]\right) \tag{42}$$

and $\lambda\in\mathbb{R}_+$ is chosen to ensure that $\left(-\frac{\partial\mathcal{H}}{\partial|V_i|^2}+\lambda\right)>0$ and $\left(-\frac{\partial\mathcal{H}}{\partial|I_i|^2}+\lambda\right)>0\ \forall i$.

Writing $g_{V_i,n-1}=g_{V_i,n-1}(\{|V_{i,n-1}|^2,|I_{i,n-1}|^2\})$ for notational convenience and taking $g_{V_i,n-1}=|V_{i,n-1}|^2\sigma_{V_i,n-1}^2$, we get

$$|V_{i,n}|^2 \leftarrow |V_{i,n-1}|^2\sigma_{V_i,n-1}^2. \tag{43}$$

2) ***Condition 2****: Considering $\beta=0$ in (10) and $V_i, I_i\in D^C=\{V_i,I_i\in\mathbb{C} : \sum_{i=1}^{N}(|V_i|^2+|I_i|^2)=1\}$, a time evolution of the form given below converges to limit cycles corresponding to the optimal point of a Lipschitz continuous objective function $\mathcal{H}(\{|V_i|,|I_i|\})$:*

$$\begin{aligned} V_{i,n} &\leftarrow V_{i,n-1}\sigma_{V_i,n-1}e^{j\theta}\\ I_{i,n} &\leftarrow I_{i,n-1}\sigma_{I_i,n-1}e^{j(\theta+\phi_i)}\end{aligned} \tag{44}$$

where $\sigma_{V_i},\sigma_{I_i}\rightarrow 1\ \ \forall i=1,\ldots,N$, in steady state, θ is the instantaneous global phase difference of the system of phasors with respect to an absolute stationary reference frame, and ϕ_i is the instantaneous phase difference between V_i and I_i.

Proof: Since $|V_{i,n}|^2=V_{i,n}V_{i,n}^*$ and $|I_{i,n}|^2=I_{i,n}I_{i,n}^*$, the update equations can be rewritten as

$$\begin{aligned} V_{i,n}V_{i,n}^* &\leftarrow [V_{i,n-1}\sigma_{V_i,n-1}e^{j\theta}]\times[V_{i,n-1}\sigma_{V_i,n-1}e^{j\theta}]^*\\ I_{i,n}I_{i,n}^* &\leftarrow [I_{i,n-1}\sigma_{I_i,n-1}e^{j(\theta+\phi_i)}]\times[I_{i,n-1}\sigma_{I_i,n-1}e^{j(\theta+\phi_i)}]^*\end{aligned} \tag{45}$$

where $\sigma_{V_i,n-1}$ is used to represent $\sigma_{V_i,n-1}(\{V_{i,n}V_{i,n}^*,I_{i,n}I_{i,n}^*\})$ for ease of notation, and similarly for $\sigma_{I_i,n-1}$. Considering $\mathcal{H}(\{V_iV_i^*,I_iI_i^*\})$ to be analytic in D^C and applying Wirtinger's calculus [42], since

$$\begin{aligned}\frac{\partial\mathcal{H}}{\partial V_{i,n-1}} &= \frac{\partial\mathcal{H}}{\partial V_{i,n-1}V_{i,n-1}^*}.\left(\frac{\partial V_{i,n-1}V_{i,n-1}^*}{\partial V_{i,n-1}}\right)\\ &= \frac{\partial\mathcal{H}}{\partial V_{i,n-1}V_{i,n-1}^*}.V_{i,n-1}^*\end{aligned} \tag{46}$$

we have

$$\sigma_{V_i,n-1}=\sqrt{\frac{1}{V_{i,n-1}^*\eta}\left(-\frac{\partial\mathcal{H}}{\partial V_{i,n-1}}+\lambda V_{i,n-1}^*\right)} \tag{47}$$

where

$$\eta=\sum_{k=1}^{N}\left(V_{k,n-1}\left[-\frac{\partial\mathcal{H}}{\partial V_{k,n-1}}+\lambda V_{k,n-1}^*\right] + I_{k,n-1}\left[-\frac{\partial\mathcal{H}}{\partial I_{k,n-1}}+\lambda I_{k,n-1}^*\right]\right). \tag{48}$$

The discrete-time update equations for $V_{i,n}$ are thus given by

$$V_{i,n} \leftarrow V_{i,n-1}\sigma_{V_i,n-1}e^{j\theta}. \tag{49}$$

Similar expressions can also be derived for the current phasors. ■

Lemma: A continuous-time variant of the model given by (44) is given by

$$\frac{\partial V_i(t)}{\partial t}=j\omega\sigma_{V_i}(t)V_i(t)-\Delta\sigma_{V_i}(t)V_i(t). \tag{50}$$

Proof: The difference equations for the voltage phasors are given by

$$V_{i,n}-V_{i,n-1}\leftarrow V_{i,n-1}\sigma_{V_i,n-1}[e^{j\theta}-1]+V_{i,n-1}[\sigma_{V_i,n-1}-1] \tag{51}$$

which can be rewritten as

$$V_{i,n}-V_{i,n-1}\leftarrow V_{i,n-1}\sigma_{V_i,n-1}[e^{j\omega\Delta t}-1]-V_{i,n-1}\Delta\sigma_{V_i,n-1} \tag{52}$$

where $\Delta\sigma_{V_i,n-1} = 1 - \sigma_{V_i,n-1}$ and $\theta = \omega\Delta t$, ω being the common oscillation frequency of all the phasors, and Δt the time increment. In the limiting case, when $\Delta t \to 0$, this reduces to the following continuous-time differential equation for the complex variable $V_i(t)$:

$$\frac{\partial V_i(t)}{\partial t} = j\omega\sigma_{V_i}(t)V_i(t) - \Delta\sigma_{V_i}(t)V_i(t). \tag{53}$$

■

In the steady state, since H is Lipschitz continuous in D, $\sigma_{V_i}(t) \overset{t\to\infty}{\longrightarrow} 1$, hence the dynamical system above reduces to

$$\frac{\partial V_i(t)}{\partial t} = j\omega V_i(t) \tag{54}$$

which implies that the steady-state response of the complex variable $V_i(t)$ corresponds to limit cycle oscillations with a constant angular frequency of ω.

The difference equations in terms of the nodal currents can be similarly written as

$$I_{i,n} - I_{i,n-1} \leftarrow I_{i,n-1}\sigma_{I_i,n-1}[e^{j(\omega+\omega_{\phi_i})\Delta t} - 1] - I_{i,n-1}\Delta\sigma_{I_i,n-1} \tag{55}$$

where $\omega_{\phi_i} = \frac{d\phi_i}{dt}$. The equivalent continuous domain differential equation is then given by

$$\frac{\partial I_i(t)}{\partial t} = j(\omega + \omega_{\phi_i})\sigma_{I_i}(t)I_i(t) - \Delta\sigma_{I_i}(t)I_i(t). \tag{56}$$

■

3) ***Condition 3****: Considering $\beta \neq 0$ in (10), additional phase constraints can be incorporated in the dynamical system represented by Equations (15)–(18). In steady state, for $|V_i|^2|I_i|^2 \neq 0$, the system settles to $\phi_i = \pm\pi/2\ \forall i$. Additionally, for sufficiently small values of β (or if β is slowly increased during the optimization process), the system converges to the optimal point of the modified objective function $\mathcal{H}(\{|V_i|, |I_i|\})$.*

Proof: Since $\mathcal{L}(\{|V_i|, |I_i|, \phi_i\})$ is Lipschitz continuous in both $|V_i|^2$ and $|I_i|^2$, the same form of the update equations proved in Lemma 2 can be applied. For arriving at the updates for the phase angles ϕ_i, we will use a similar approach as shown in (37). We can split ϕ_i as $\phi_i = \phi_i^+ - \phi_i^-,\ \phi_i^+, \phi_i^- > 0$, which implies that $\phi_i^+ + \phi_i^- = \pi$. We can then apply the growth transform dynamical system [7] to get

$$\tau_i\omega_{\phi_i} + \phi_i(t) = g_{\phi_i}(t) \tag{57}$$

where $\omega_{\phi_i} = \frac{d\phi_i(t)}{dt}$ and

$$g_{\phi_i} = \pi\frac{\phi_i^+\left(-\frac{\partial\mathcal{L}}{\partial\phi_i^+} + \lambda\right) - \phi_i^-\left(-\frac{\partial\mathcal{L}}{\partial\phi_i^-} + \lambda\right)}{\phi_i^+\left(-\frac{\partial\mathcal{L}}{\partial\phi_i^+} + \lambda\right) + \phi_i^-\left(-\frac{\partial\mathcal{L}}{\partial\phi_i^-} + \lambda\right)}. \tag{58}$$

Since $\frac{\partial\mathcal{L}}{\partial\phi_i} = \frac{\partial\mathcal{L}}{\partial\phi_i^+} = -\frac{\partial\mathcal{L}}{\partial\phi_i^-}$, the above can be simplified to $g_{\phi_i} = \pi(\lambda\phi_i - \pi(\partial\mathcal{L}/\partial\phi_i))/(-\phi_i(\partial\mathcal{L}/\partial\phi_i) + \lambda\pi)$. This implies that the voltage and current phasors corresponding to the ith node in the network may be phase shifted by an additional amount ϕ_i with respect to the absolute reference.

Since for optimality, $\phi_i = \pm\pi/2$ for $|V_i|^2|I_i|^2 \neq 0$ in the steady state, the net energy dissipation in steady state is zero, i.e., $\sum_{n=1}^{N}|V_i|^2|I_i|^2\cos^2\phi_i = 0$. Also, in the steady state, for sufficiently small values of the hyperparameter β

$$\min_{\{|V_i|,|I_i|,\phi_i\}} \mathcal{L}(\{|V_i|, |I_i|, \phi_i\}) = \min_{\{|V_i|,|I_i|\}} \mathcal{H}(\{|V_i|, |I_i|\}) \tag{59}$$

which implies that the system reaches the optimal solution with zero active power in the post-learning stage. ■

REFERENCES

[1] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, and F. Huang, "A tutorial on energy-based learning," *Predict. Struct. Data*, vol. 1, pp. 1–59, Aug. 2006.

[2] J. J. Hopfield and D. W. Tank, "Neural computation of decisions in optimization problems," *Biol. Cybern.*, vol. 52, no. 3, pp. 141–152, 1985.

[3] J. Edminister, *Theory and Problems of Electric Circuits*. New York, NY, USA: McGraw-Hill, 1965.

[4] M. E. V. Valkenburg, *Network Analysis*. Upper Saddle River, NJ, USA: Prentice-Hall, 1964.

[5] J. Dixon, L. Moran, J. Rodriguez, and R. Domke, "Reactive power compensation technologies: State-of-the-art review," *Proc. IEEE*, vol. 93, no. 12, pp. 2144–2164, Dec. 2005.

[6] W. Hofmann, J. Schlabbach, and W. Just, *Reactive Power Compensation: A Practical Guide*. Hoboken, NJ, USA: Wiley, 2012.

[7] O. Chatterjee and S. Chakrabartty, "Decentralized global optimization based on a growth transform dynamical system model," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6052–6061, Dec. 2018.

[8] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, "Support vector method for novelty detection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2000, pp. 582–588.

[9] S. H. Strogatz, "From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators," *Phys. D, Nonlinear Phenomena*, vol. 143, nos. 1–4, pp. 1–20, Sep. 2000.

[10] R. Feynman, *The Character of Physical Law*. Cambridge, MA, USA: MIT Press, 2017.

[11] Y.-L. Boureau *et al.*, "A unified energy-based framework for unsupervised learning," in *Proc. Artif. Intell. Statist.*, 2007, pp. 371–379.

[12] G. E. Hinton and T. J. Sejnowski, "Optimal perceptual inference," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1983, pp. 448–453.

[13] R. Salakhutdinov, A. Mnih, and G. Hinton, "Restricted Boltzmann machines for collaborative filtering," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 791–798.

[14] B. Scellier and Y. Bengio, "Equilibrium propagation: Bridging the gap between energy-based models and backpropagation," *Frontiers Comput. Neurosci.*, vol. 11, p. 24, May 2017.

[15] J. Ngiam, Z. Chen, P. W. Koh, and A. Y. Ng, "Learning deep energy models," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 1–8.

[16] S. Zhai, Y. Cheng, W. Lu, and Z. Zhang, "Deep structured energy based models for anomaly detection," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1100–1109.

[17] T. Haarnoja, H. Tang, P. Abbeel, and S. Levine, "Reinforcement learning with deep energy-based policies," in *Proc. 34th Int. Conf. Mach. Learn. (JMLR)*, vol. 70, 2017, pp. 1352–1361.

[18] J. Lawson, G. Tucker, B. Dai, and R. Ranganath, "Energy-inspired models: Learning with sampler-induced distributions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8499–8511.

[19] A. Hirose, *Complex-Valued Neural Networks: Theories and Applications*, vol. 5. Singapore: World Scientific, 2003.

[20] C. Trabelsi *et al.*, "Deep complex networks," 2017, *arXiv:1705.09792*. [Online]. Available: http://arxiv.org/abs/1705.09792

[21] C. J. Gaudet and A. S. Maida, "Deep quaternion networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

[22] N. Guberman, "On complex valued convolutional neural networks," 2016, *arXiv:1602.09046*. [Online]. Available: http://arxiv.org/abs/1602.09046

[23] P. Bouboulis, S. Theodoridis, C. Mavroforakis, and L. Evaggelatou-Dalla, "Complex support vector machines for regression and quaternary classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1260–1274, Jun. 2015.

[24] J. Bruna, S. Chintala, Y. LeCun, S. Piantino, A. Szlam, and M. Tygert, "A mathematical motivation for complex-valued convolutional networks," 2015, *arXiv:1503.03438*. [Online]. Available: http://arxiv.org/abs/1503.03438

[25] T. Nitta, "Orthogonality of decision boundaries in complex-valued neural networks," *Neural Comput.*, vol. 16, no. 1, pp. 73–97, Jan. 2004.
[26] T. Nakashika, S. Takaki, and J. Yamagishi, "Complex-valued restricted Boltzmann machine for direct learning of frequency spectra," in *Proc. INTERSPEECH*, 2017, pp. 4021–4025.
[27] B. D. Tellegen, "A general network theorem, with applications," *Philips Res. Rep.*, vol. 7, pp. 256–269, Aug. 1952.
[28] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin, "The elements of statistical learning: Data mining, inference and prediction," *Math. Intell.*, vol. 27, no. 2, pp. 83–85, Mar. 2005.
[29] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.
[30] Z. Ghahramani, "Probabilistic machine learning and artificial intelligence," *Nature*, vol. 521, no. 7553, pp. 452–459, May 2015.
[31] S. Chakrabartty and G. Cauwenberghs, "Gini support vector machine: Quadratic entropy based robust multi-class probability regression," *J. Mach. Learn. Res.*, vol. 8, pp. 813–839, Apr. 2007.
[32] D. M. J. Tax and R. P. W. Duin, "Support vector domain description," *Pattern Recognit. Lett.*, vol. 20, nos. 11–13, pp. 1191–1199, Nov. 1999.
[33] A. Gangopadhyay, O. Chatterjee, and S. Chakrabartty, "Extended polynomial growth transforms for design and training of generalized support vector machines," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1961–1974, May 2018.
[34] D. Dua and C. Graff. (2017). *UCI Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml
[35] S. Rayana. (2016). *ODDS Library*. [Online]. Available: http://odds.cs.stonybrook.edu
[36] A. Raychowdhury *et al.*, "Computing with networks of oscillatory dynamical systems," *Proc. IEEE*, vol. 107, no. 1, pp. 73–89, Jan. 2019.
[37] J. Torrejon *et al.*, "Neuromorphic computing with nanoscale spintronic oscillators," *Nature*, vol. 547, no. 7664, pp. 428–431, Jul. 2017.
[38] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[39] A. Gangopadhyay, O. Chatterjee, and S. Chakrabartty, "Continuous-time optimization using sub-threshold current-mode growth transform circuits," in *Proc. IEEE 61st Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2018, pp. 246–249.
[40] L. E. Baum and G. Sell, "Growth transformations for functions on manifolds," *Pacific J. Math.*, vol. 27, no. 2, pp. 211–227, Nov. 1968.
[41] P. S. Gopalakrishnan, D. Kanevsky, A. Nadas, and D. Nahamoo, "A generalization of the Baum algorithm to rational objective functions," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 1989, pp. 631–634.
[42] M. F. Amin, M. I. Amin, A. Y. H. Al-Nuaimi, and K. Murase, "Wirtinger calculus based gradient descent and Levenberg–Marquardt learning algorithms in complex-valued neural networks," in *Proc. Int. Conf. Neural Inf. Process.* Berlin, Germany: Springer, 2011, pp. 550–559.

Oindrila Chatterjee (Student Member, IEEE) received the B.E. degree in electrical engineering from Jadavpur University, Kolkata, India, in 2013, and the M.S. degree in electrical engineering from Washington University in St. Louis, St. Louis, MO, USA, in 2017. She is presently a doctoral candidate in the Electrical and Systems Engineering Department at Washington University in St. Louis.

Her current research interests lie at the intersection of machine learning, optimization, nonlinear dynamical systems, and analog computing.

Shantanu Chakrabartty (Senior Member, IEEE) received the B.Tech. degree from IIT Delhi, New Delhi, India, in 1996, and the M.S. and Ph.D. degrees in electrical engineering from Johns Hopkins University, Baltimore, MD, USA, in 2002 and 2004, respectively.

From 1996 to 1999, he was with Qualcomm Inc., San Diego, CA, USA. In 2002, he was a Visiting Researcher with The University of Tokyo, Tokyo, Japan. From 2004 to 2015, he was an Associate Professor with the Department of Electrical and Computer Engineering, Michigan State University (MSU), East Lansing, MI, USA. He is currently a Clifford Murphy Professor with the Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO, USA. His work covers different aspects of analog computing. His current research interests include self-powered sensors and neuromorphic and hybrid circuits and systems.

Dr. Chakrabartty was a fellow of the Catalyst Foundation from 1999 to 2004. He was a recipient of the National Science Foundation's CAREER Award, the University Teacher-Scholar Award from MSU, and the 2012 Technology of the Year Award from MSU Technologies. He has previously served as an Associate Editor for the IEEE TRANSACTIONS OF BIOMEDICAL CIRCUITS AND SYSTEMS and a Review Editor for *Frontiers of Neuromorphic Engineering* journal.