Deep feedforward functionality by equilibrium-point control in a shallow recurrent network

Anonymous Author(s)
Affiliation
Address
email

Abstract

Recurrent neural network based machine learning systems are typically employed for their sequential functionality in handling time-varying signals, such as for speech processing. However, neurobiologists find recurrent connections in the vision system and debate about equilibrium-point control in the motor system. Thus, we need a deeper understanding of how recurrent dynamics can be exploited to attain combinational stable-input stable-output functionality. Here, we study how a simplified Cohen-Grossberg neural network model can realize combinational multi-input Boolean functionality. We place our problem within the discipline of algebraic geometry, and solve a special case of it using piecewise-linear algebra. We demonstrate a connectance-efficient realization of the parity function as a proof-of-concept. Small-scale systems of this kind can be easily built, say for hobby robotics, as a network of two-terminal devices of resistors and tunnel diodes. Large-scale systems may be energy-efficiently built as an interconnected network of multi-electrode nanoclusters with non-monotonic transport mechanisms.

1 Introduction

Shallow recurrent neural networks are being investigated for more context-aware object recognition [25] and brain-like behaviour [23]. They can be more compact (by trading space for time) and are a naturally robust alternative to deep neural networks (which are easily fooled by input perturbations or transformations [18, 29, 1]) when the role of recurrent dynamics is not to produce time-varying output but instead to produce transient (hidden) state-dynamics that facilitate deep, robust and transformation-invariant fixed-input fixed-output functionality. To better engineer such dynamics, we shall study equilibrium-point control, which can be defined as the process of steering to a target in state-space by fixing the input signal, instead of driving it by a continuously varying input signal. Historically, equilibrium-point control [14, 5] was first formulated to provide a plausible solution to the degrees of freedom problem in motor control [3], that is, we mentally represent intermediate destination points rather than a continuum of velocity information required to execute a movement.

Here, we shall focus on using equilibrium-point control to realize multi-input Boolean functionality, in particular the parity function, which is a canonical proxy for nonlinear classification. Theoretical results in circuit complexity are known already for realizing Boolean functionality out of feedforward neural networks, with weighted-sum thresholded binary-output neurons [35]. It has been shown that arbitrary N-input Boolean functions can be realized in depth-3 feedforward networks with fewer neurons ($m = \mathcal{O}(2^{N/2})$ instead of the $\mathcal{O}(2^N)$ in total required for depth-2). However, with the advent of nanoelectronics, the size of an artificial neuron has been downscaled to such an extent that it is rather the interconnect wiring that now occupies a greater area in chip design. Thus for a fully-connected deep network, the area scales as the number of interconnects $m^2 = \mathcal{O}(2^N)$. Such a

$\mathcal{O}(2^N)$ scaling law was earlier obtained by Shannon [34] for realizing arbitrary N-input Boolean functions by an interconnection of input-controlled switches (or equivalently a feedforward network of 2-input Boolean gates). Thus, unless we employ higher-order neurons [16], we can say that a *Shannon bottleneck* limits the maximum N-input Boolean logic realizable in a given area by (nanoscale) feedforward networks. We aim to circumvent this Shannon bottleneck by employing recurrent physical networks. It is known that certain combinational logic functions can be realized by fewer logic gates in a cyclic network than in an acyclic network [31], and with analog signal processing the improvement factor could be even higher.

In the following section, we introduce a state-space model formalism to study equilibrium-point control, and commit to a physically realizable model, and discuss how a general solution for its equilibrium points is a difficult problem in algebraic geometry. Thus, we proceed to idealize the non-monotonic output of the physical system as a piecewise-linear function and solve for the equilibrium points. Finally, a piecewise-quadratic Lyapunov function is obtained for stability analysis and conditions for a unique equilibrium-point are provided.

After the theory, in the results section, we provide a connectance-efficient realization of the parity function. The discussion section puts our results into a broader context and offers avenues for further research. Our objective here is to work at the intersection of nonlinear dynamical systems, neural networks, unconventional neuromorphic hardware, cyclic Boolean circuits, piecewise-linear control systems, and algebraic geometry.

2 Theory

2.1 State-space model

For equilibrium-point control, in general we have an input vector $\boldsymbol{x}$, a state $s_i(t)$ for $i = 1 : N$, and an output y obtained from a system of equations

$$\dot{s}_i(t) = F_i(\boldsymbol{s}(t), \boldsymbol{x}),\ y = \lim_{t\to\infty} G(\boldsymbol{s}(t)). \tag{1}$$

In this paper, we commit to a physically realizable recurrent network with voltage nodes s_i from $i = 1 : N$, with a capacitive time-constant τ_i, using resistors (of a constant conductance f_{ij}) and tunnel diodes (of a voltage-dependent conductance $G_i(s_i)$) as shown in Fig. 1, yielding a state-space model of the form

$$\tau_i \dot{s}_i = x_i - \sum_{j\neq i} f_{ij}(s_i - s_j) - G_i(s_i),\ y = G_1(\hat{s}_1) \tag{2}$$

where $f_{ij} \geq 0$, G_i is a nonlinear passive function such that $G_i(s)s \geq 0$ and $\hat{s}_1 \equiv \lim_{t\to\infty} s_1(t)$ is the stable equilibrium-point if one exists (note: $y(\boldsymbol{x})$ can be multi-valued and depend on the basin of attraction that the initial state $\boldsymbol{s}(0)$ lies in). Brain-scale systems of this kind may be realized by an interconnected network of nanoclusters with non-monotonic transport mechanisms as proposed in [24, Chapter 5]. However, finding suitable network parameters that result in practical functionality remains a challenge. Note that, although not the focus of this work, Eq. (2) can also represent state-space models with noisy rectified-linear units, for which semi-analytical results are known from a computational neuroscience [12] and a machine learning [33] perspective.

2.2 Algebraic geometry of the equilibrium points

A study of the set of equilibrium points of a state-space model, $\mathcal{S}_0(\boldsymbol{x}) \equiv \{\boldsymbol{s} \ni F_{1:N}(\boldsymbol{s}, \boldsymbol{x}) = \boldsymbol{0}\}$, can not only help in characterising the stable equilibrium-points $\hat{\boldsymbol{s}} \in \mathcal{S}_* \subseteq \mathcal{S}_0$, but also provide necessary (but not sufficient) conditions in the parameters defining the functions $F_{1:N}$ and G, to realize desired equilibrium-point functionality $y(\boldsymbol{x})$. For example, to realize a Boolean function $y : \{0,1\}^N \to \{0,1\}$, the following property has to be satisfied:

$$\min_{\boldsymbol{s}\in\mathcal{S}_0(\boldsymbol{x})} G(\boldsymbol{s}) \leq 1 \wedge \max_{\boldsymbol{s}\in\mathcal{S}_0(\boldsymbol{x})} G(\boldsymbol{s}) \geq 0\ \forall \boldsymbol{x} \in \{0,1\}^N. \tag{3}$$

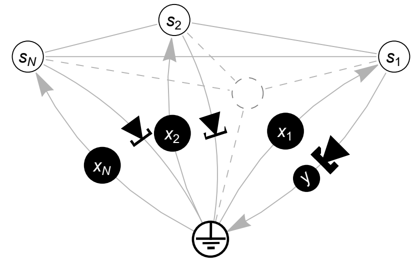

Figure 1: Recurrent physical network corresponding to the state-space model (2) where the inputs $x_{1:N}$ are currents, the states $s_{1:N}$ are voltages, the output y is a measured current, the linear interactions are due to resistors with a conductance f_{ij} between node i and j, and nonlinear interactions are due to tunnel diodes from node i to GND with conductance $G_i(s_i)$.

The set of equilibrium points of our recurrent physical network model (2) are the roots of the system of nonlinear equations

$$-f_{i,1:N} \cdot s_{1:N} + G_i(s_i) = x_i \tag{4}$$

where the linear-interaction matrix $f_{N\times N}$ has terms $f_{ii} \equiv -\sum_{j\neq i} f_{ij}$.

Solving the multivariate nonlinear equation (4) is a difficult problem in algebraic geometry, a discipline of mathematics which classically grew around efforts to understand the roots of multivariate polynomials and later metamorphosed by the study of integer-coefficient piecewise-linear functions, with an abstract language that has even recently been applied to explain circuit complexity results of deep feedforward networks [35, 30] through the lens of rational piecewise-linear functions [40].

Algebraic geometry originally dealt with a qualitative approach by geometrical arguments [15], in contrast to a quantitative approach by numerical methods. An example of that kind is Harnack's curve theorem [19] which states that for a 2-D polynomial curve of degree n, the maximum number of connected components is $(n^2 - 3n + 4)/2$. Now, with the advent of computer algebra, the roots of multivariate nonlinear equations are studied by the elimination of variables, using techniques such as resultants [13, 38] and Groebner bases [7, 8] for polynomial systems, and as an instance of the linear-complementarity problem [11] or equivalently as absolute-value equations [26] for piecewise-linear systems [37]. However, computer algebra is not scalable for higher dimensions. Thus there is a need to convey the richness in algebraic geometry using analytical expressions. While it is unlikely that analytical expressions may be obtained for any general form of nonlinearity, we may hope that the set of exactly solvable models can be extended well beyond linear equations, a hope banking on our successful experience from other areas of mathematics such as integral calculus [9, section IX] and iterated mappings [39, page 1098].

2.2.1 Piecewise-linear algebra

In this paper, we commit to a piecewise-linear analysis by considering

$$G_i(s) = \begin{cases} g_{i1}s & 0 \le s \le g_{i2} \\ (g_{i1}+g_{i3})g_{i2} - g_{i3}s & g_{i2} \le s \le g_{i2}(1+\frac{g_{i1}}{g_{i3}}) \end{cases}, \tag{5}$$

where $g_{i1,2,3} > 0$ so that G_i is a triangular peak function in a limited range of s, thus defining an idealized negative-differential behaviour. Shifting the state-space about its inflection points as $\boldsymbol{z} \equiv \boldsymbol{s} - \boldsymbol{g}_2$ and then combining (5) with (4) yields

$$x_i = \begin{cases} -f_{i,1:N}\cdot \boldsymbol{z} + g_{i1}z_i - f_{i,1:N}\cdot \boldsymbol{g}_2 + g_{i1}g_{i2} & -g_{i2} \le z_i \le 0 \\ -f_{i,1:N}\cdot \boldsymbol{z} - g_{i3}z_i - f_{i,1:N}\cdot \boldsymbol{g}_2 + g_{i1}g_{i2} & 0 \le z_i \le g_{i2}(\frac{g_{i1}}{g_{i3}}) \end{cases}, \tag{6}$$

which can be simplified to

$$x_i = -f_{i,1:N}\cdot \boldsymbol{z} + g_{i\ominus}z_i - g_{i\oplus}|z_i| - f_{i,1:N}\cdot \boldsymbol{g}_2 + g_{i1}g_{i2} \tag{7}$$

for $-g_{i2} \leq z_i \leq g_{i2}(\frac{g_{i1}}{g_{i3}})$ with $g_{i\ominus} \equiv (g_{i1} - g_{i3})/2$ and $g_{i\oplus} \equiv (g_{i1} + g_{i3})/2$.

The system in (7) can be expressed in the absolute-value equation normal form

$$\mathbf{A}\boldsymbol{z} - |\boldsymbol{z}| = \boldsymbol{b} \tag{8}$$

with $A_{ij} = (-f_{ij} + \mathrm{I}_{ij}\, g_{i\ominus})/g_{i\oplus}$, $b_i = (x_i + f_{i,1:N} \cdot \boldsymbol{g}_2 + g_{i1}g_{i2})/g_{i\oplus}$ and the bounds

$$-g_2 \leq z \leq g_2 g_1/g_3. \tag{9}$$

Similarly, (2) can be expressed as

$$\boldsymbol{\tau}\dot{\boldsymbol{z}} = \boldsymbol{g}_\oplus(\boldsymbol{b} - \mathbf{A}\boldsymbol{z} + |\boldsymbol{z}|). \tag{10}$$

Two sufficient conditions are known for the absolute value equation (8) to have a unique solution based on the largest singular value $\sigma_{\min}$ [26] and the spectral radius ρ [32]:

$$\sigma_{\min}(\mathbf{A}) > 1, \tag{11}$$

$$\rho(|\mathbf{A}^{-1}|) < 1. \tag{12}$$

However, those are not yet sufficient conditions for a unique equilibrium-point solution for (2) and (10) because the bounds in (9) were not enforced. Thus, we shall proceed to obtain a Lyapunov function to guarantee that a stable equilibrium-point is reached.

2.3 Lyapunov stability analysis

Equilibrium-point stability for large complex systems is not guaranteed in general [17, 27], and the effective dimensionality of stable-input stable-output responses is richly dependent on the parameter space [2]. However, the interaction matrix for our physical system (2) is symmetric, and hence the system is a special case of the Cohen-Grossberg model [10]

$$\dot{s_i} = a_i(s_i)[b_i(s_i) - \sum_{j=1}^{N} c_{ij} d_j(s_j)], \tag{13}$$

with $a_i(s_i) = 1/\tau_i$, $b_i(s_i) = x_i - G_i(s_i)$, $c_{ij} = -f_{ij}$ and $d_j(s_j) = s_j$. Thus, it is known to be globally absolute stable, with a Lyapunov function

$$V = -\sum_i \int_0^{s_i} b_i(u) d_i'(u)\, \mathrm{d}u + \sum_{i,j} \frac{c_{ij}}{2} d_i(s_i) d_j(s_j) \tag{14}$$

$$= \sum_i \Big(P_i(s_i) - x_i s_i - \sum_{j>i} f_{ij} s_i s_j - \frac{f_{ii}}{2} s_i^2\Big), \tag{15}$$

$$\text{where output power } P_i(s_i) \equiv \int_0^{s_i} G_i(u)\, \mathrm{d}u. \tag{16}$$

Alternatively, since our system (5) is piecewise-linear, a piecewise-quadratic Lyapunov function may be obtained by a piecewise-affine system [22] analysis. While this approach is more powerful and holds even for asymmetric interaction matrices, it also seems to be analytically complex. From another angle, global asymptotic stability [21, Theorem 3] is guaranteed if the Jacobian matrix $\mathbf{J}$ satisfies $J_{ii} + 1/2 \sum_{j\neq i} |J_{ij} + J_{ji}| < 0 \iff G_i'(s_i) > 0$ because in our system $J_{ii} = -\sum_{j\neq i} f_{ij} - G_i'(s_i)$ and $J_{ij} = f_{ij}$. Since our network employs non-monotonic functionality, $G_i'(s_i) > 0$ cannot be guaranteed for all reachable states s_i, and thus the above criteria is unfortunately inapplicable. Hence, we shall proceed with the Cohen-Grossberg approach.

The power function (21) simplifies to

$$P(s) = \begin{cases} \int_0^s g_1 u\, \mathrm{d}u = g_1 s^2/2 & 0 \leq s \leq g_2 \\ g_1 g_2^2/2 + \int_{g_2}^s (g_1 + g_3) g_2 - g_3 u\, \mathrm{d}u & g_2 \leq s \\ = g_1 s^2/2 - g_\oplus (s - g_2)^2, & \end{cases} \tag{17}$$

and using the rectifier function $[x) \equiv \max(x, 0)$ may be expressed conveniently as

$$P(s) = g_1 s^2/2 - g_\oplus [s - g_2)^2, \tag{18}$$

when the system is within its operational bounds.

3 Results

Given the Lyapunov stability result of our system, it is computationally efficient to simulate our state-space model and probe for combinational functionality. Here, we will simulate for the simplest proof-of-concept for deep functionality in a shallow recurrent - solving a parity problem.

Using a cascade of 2-input XOR gates, the N-bit parity function can be realized with $N/2 + N/4 + ... + 1 = N - 1$ gates and $2N - 1$ connections. Thus its total cost in area is at least $3N - 2$ units. A minimally-connected network has N input wires, 1 output wire, and $N - 1$ interconnect wires with a total area cost of $2N$ units, assuming that the area occupied by the remaining components is negligible. Thus for $N = 3$, while a conventional digital circuit costs 7 units, our recurrent physical network takes just 6 wiring units.

Our simple model has $N = 3$, $f_{12} = f_{13} = f$, $f_{23} = 0$, $g_{11} = g_1$, $g_{13} = g_3$, $g_{21} = g_{31} = \gamma_1$, $g_{23} = g_{33} = \gamma_3$, $g_{12} = g_2$ and $\gamma_{22} = \gamma_{32} = \gamma_2$. We find from a symbolic evaluation that $\sigma_{\min}(\mathbf{A}) \neq 1/\rho(|\mathbf{A}^{-1}|)$ in general, and conditions for unique stability were not obtainable (which is not surprising due to the $s_2 - s_3$ symmetry). Thus, parity functionality was found by trial-and-error yielding the parameters $\{f = 1.751, g_1 = 1.876, g_2 = g_3 = 0.126, \gamma_1 = 0.876, \gamma_2 = 1.6, \gamma_3 = 0.751\}$ and simulated using Wolfram Mathematica 13 (code in Appendix). When $x_1 = x_2 = x_3 = 1$, the states were forced to transition beyond the bounds in (5), so its range was extended by taking an absolute value. The results are plotted in Fig. 2.

4 Discussion

Our result should be seen as a theoretical proof-of-concept and as a motivation for continued research in this area. Future work must extend our simulations to much higher dimensions to serve as a practical demonstration of deep functionality by shallow recurrent networks. Moreover, the theoretical formalism introduced here is not yet fully exploited. We hope to find an analytical method to design functionality out of piecewise-linear Cohen-Grossberg networks.

Our style of reasoning to circumvent the Shannon bottleneck may also be applied to other systems such as networks of coupled oscillators [28]. Our non-modular mode of signal processing, offers an alternative to not just circuit designers, but also to systems biologists who typically understand chemical reaction networks [6] as a composition of modules [20]. While, we have discussed equilibium-point functionality in a state-space model driven by an additive input, it is also worth investigating autonomous systems where the input is set as an initial state. An example is realizing unboundedly-finite parity functions using just a radius-4 cellular automaton [4]. Finally, we hope that this paper can serve as a call to action for neuromorphic engineers to look at physical reservoir computing [36] from another angle, besides temporal input-output functionality.

References

[1] M. A. Alcorn, Q. Li, Z. Gong, C. Wang, L. Mai, W.-S. Ku, and A. Nguyen. Strike (with) a pose: Neural networks are easily fooled by strange poses of familiar objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4845–4854, 2019.

[2] R. D. Beer. Parameter space structure of continuous-time recurrent neural networks. *Neural computation*, 18(12):3009–3051, 2006.

[3] N. Bernstein. The co-ordination and regulation of movements. *The co-ordination and regulation of movements*, 1966.

[4] H. Betel, P. P. de Oliveira, and P. Flocchini. Solving the parity problem in one-dimensional cellular automata. *Natural Computing*, 12(3):323–337, 2013.

[5] E. Bizzi, N. Hogan, F. A. Mussa-Ivaldi, and S. Giszter. Does the nervous system use equilibrium-point control to guide single and multiple joint movements? *Behavioral and brain sciences*, 15(4):603–613, 1992.

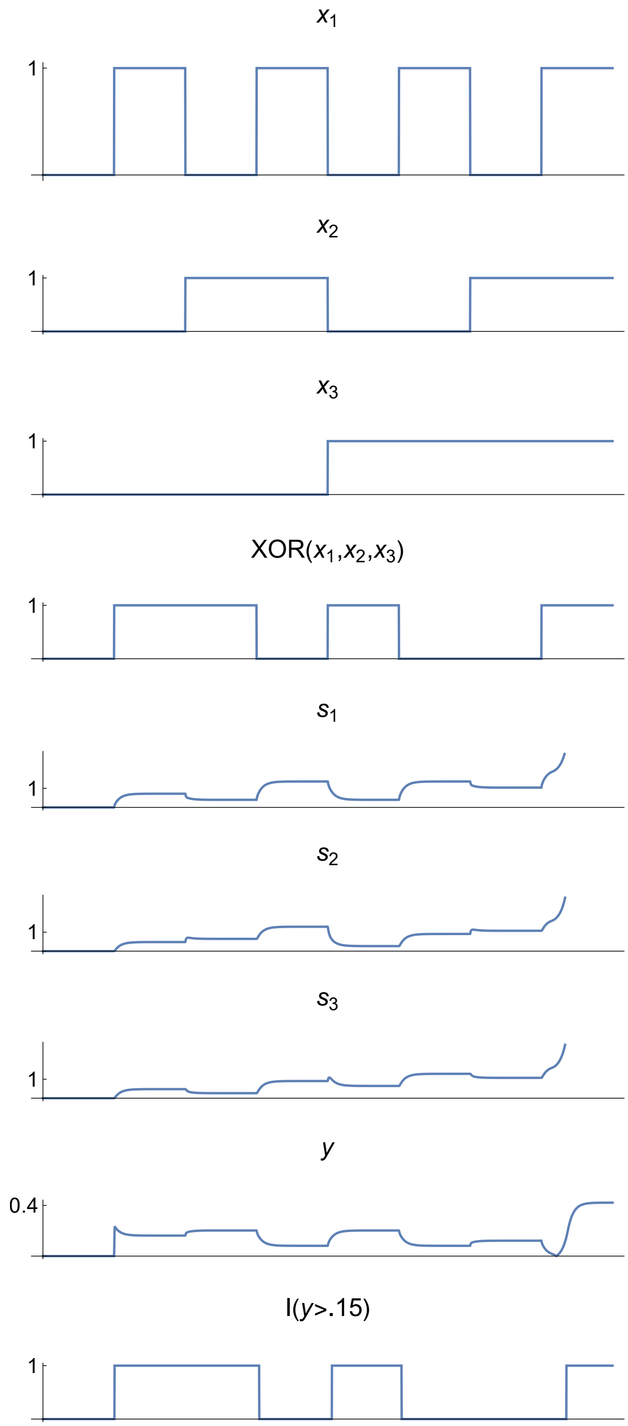

Figure 2: Numerical simulation of our 3-state network over 200 timesteps.

[6] D. Bray. Protein molecules as computational elements in living cells. *Nature*, 376(6538):307–312, 1995.

[7] B. Buchberger. Ein algorithmus zum auffinden der basiselemente des restklassenringes nach einem nulldimensionalen polynomideal. *PhD thesis, Universitat Insbruck*, 1965.

[8] B. Buchberger. Bruno buchberger's phd thesis 1965: An algorithm for finding the basis elements of the residue class ring of a zero dimensional polynomial ideal. *Journal of symbolic computation*, 41(3-4):475–511, 2006.

[9] G. S. Carr. Synopsis of elementary results in pure mathematics. 1886.

[10] M. A. Cohen and S. Grossberg. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE transactions on systems, man, and cybernetics*, (5):815–826, 1983.

[11] R. W. Cottle. *Linear complementarity problem*, pages 1873–1878. Springer US, Boston, MA, 2009.

[12] D. Durstewitz. A state space approach for piecewise-linear recurrent neural networks for identifying computational dynamics from neural measurements. *PLoS computational biology*, 13(6):e1005542, 2017.

[13] I. Z. Emiris. On the complexity of sparse elimination. *Journal of Complexity*, 12(2):134–166, 1996.

[14] A. G. Feldman. Functional tuning of the nervous system with control of movement or maintenance of a steady posture-ii. controllable parameters of the muscle. *Biofizika*, 11:565–578, 1966.

[15] W. Fulton. *Intersection theory*, volume 2. Springer Science & Business Media, 2013.

[16] S. Gao, M. Zhou, Y. Wang, J. Cheng, H. Yachi, and J. Wang. Dendritic neuron model with effective learning algorithms for classification, approximation, and prediction. *IEEE transactions on neural networks and learning systems*, 30(2):601–614, 2019.

[17] M. R. Gardner and W. R. Ashby. Connectance of large dynamic (cybernetic) systems: critical values for stability. *Nature*, 228(5273):784–784, 1970.

[18] I. J. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples, 2015.

[19] A. Harnack. Ueber die vieltheiligkeit der ebenen algebraischen curven. *Mathematische Annalen*, 10(2):189–198, 1876.

[20] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray. From molecular to modular cell biology. *Nature*, 402(6761):C47–C52, 1999.

[21] M. W. Hirsch. Convergent activation dynamics in continuous time networks. *Neural networks*, 2(5):331–349, 1989.

[22] M. Johansson and A. Rantzer. Computation of piecewise quadratic lyapunov functions for hybrid systems. In *1997 European Control Conference (ECC)*, pages 2005–2010. IEEE, 1997.

[23] J. Kubilius, M. Schrimpf, K. Kar, R. Rajalingham, H. Hong, N. Majaj, E. Issa, P. Bashivan, J. Prescott-Roy, K. Schmidt, et al. Brain-like object recognition with high-performing shallow recurrent anns. *Advances in neural information processing systems*, 32, 2019.

[24] C. P. Lawrence. *Evolving Networks To Have Intelligence Realized At Nanoscale*. PhD thesis, University of Twente, 2018.

[25] M. Liang and X. Hu. Recurrent convolutional neural network for object recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3367–3375, 2015.

[26] O. Mangasarian and R. Meyer. Absolute value equations. *Linear Algebra and Its Applications*, 419(2-3):359–367, 2006.

[27] R. M. May. Will a large complex system be stable? *Nature*, 238(5364):413–414, 1972.

[28] S. N. Menon and S. Sinha. "defective" logic: Using spatiotemporal patterns in coupled relaxation oscillator arrays for computation. In *2014 International Conference on Signal Processing and Communications (SPCOM)*, pages 1–6. IEEE, 2014.

[29] S.-M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard. Universal adversarial perturbations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1765–1773, 2017.

[30] M. Raghu, B. Poole, J. Kleinberg, S. Ganguli, and J. Sohl-Dickstein. On the expressive power of deep neural networks. In *international conference on machine learning*, pages 2847–2854. PMLR, 2017.

[31] M. D. Riedel and J. Bruck. Cyclic boolean circuits. *Discrete Applied Mathematics*, 160(13-14):1877–1900, 2012.

[32] J. Rohn, V. Hooshyarbakhsh, and R. Farhadsefat. An iterative method for solving absolute value equations and sufficient conditions for unique solvability. *Optimization Letters*, 8(1):35–44, 2014.

[33] D. Schmidt, G. Koppe, Z. Monfared, M. Beutelspacher, and D. Durstewitz. Identifying nonlinear dynamical systems with multiple time scales and long-range dependencies. In *International Conference on Learning Representations*, 2021.

[34] C. E. Shannon. The synthesis of two-terminal switching circuits. *The Bell System Technical Journal*, 28(1):59–98, 1949.

[35] K.-Y. Siu, V. P. Roychowdhury, and T. Kailath. Depth-size tradeoffs for neural computation. *IEEE Transactions on Computers*, 40(12):1402–1412, 1991.

[36] G. Tanaka, T. Yamane, J. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose. Recent advances in physical reservoir computing: A review. *Neural Networks*, 115:100–123, 2019.

[37] W. M. Van Bokhoven and D. M. Leenaerts. Explicit formulas for the solutions of piecewise linear networks. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 46(9):1110–1117, 1999.

[38] M. P. Williams. Solving polynomial equations using linear algebra. *Johns Hopkins APL Technical Digest*, 28(4):354–363, 2010.

[39] S. Wolfram. *A new kind of science*, volume 5. Wolfram media Champaign, IL, 2002.

[40] L. Zhang, G. Naitzat, and L.-H. Lim. Tropical geometry of deep neural networks. In *International Conference on Machine Learning*, pages 5824–5832. PMLR, 2018.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] The concrete result is the realization of a parity function by our recurrent physical network by using just 6 wiring units, while a conventional digital circuit costs 7 units. That being said, the paper is written to cover a much broader scope - this is a matter of taste (an earlier version of this manuscript recieved both positive and negative comments about the scope of this article).

(b) Did you describe the limitations of your work? [Yes] It is mentioned that future work must extend our simulations to much higher dimensions to serve as a practical demonstration of deep functionality by shallow recurrent networks. Also the simulation parameters were found by trial and error, instead of being derived analytically from the theoretical formalism - these limitations are mentioned in the discussion.

(c) Did you discuss any potential negative societal impacts of your work? [N/A]

(d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]

2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Check Appendix for the code to reproduce Figure 2.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A] It is evident that Figure 2 is not a large-scale deep learning experiment but a small-scale conceptual simulation which takes less than 2 seconds on a modern desktop CPU.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [No]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Appendix

Wolfram Mathematica code to reproduce Figure 2.

```
In[•]:= simulate[f_, g_, γ_] := (sys = NonlinearStateSpaceModel[{
          {x1 - (2 f) s1 + f (s2 + s3) - Abs[g⟦1⟧ * s1 - (g⟦1⟧ + g⟦3⟧) Ramp[s1 - g⟦2⟧]],
           x2 - (f) s2 + f (s1) - Abs[γ⟦1⟧ s2 - (γ⟦1⟧ + γ⟦3⟧) Ramp[s2 - γ⟦2⟧]],
           x3 - (f) s3 + f (s1) - Abs[γ⟦1⟧ s3 - (γ⟦1⟧ + γ⟦3⟧) Ramp[s3 - γ⟦2⟧]]},
          {x1, x2, x3, Xor[x1, x2, x3], s1, s2, s3, y = Abs[g⟦1⟧ s1 - (g⟦1⟧ + g⟦3⟧) Ramp[s1 - g⟦2⟧]], HeavisideTheta[y - .15]}
         }, {s1, s2, s3}, {x1, x2, x3}];
      inputs = {.5 - .5 * SquareWave[ t / 50], .5 - .5 * SquareWave[ t / 100], .5 - .5 * SquareWave[ t / 200]};
      out = OutputResponse[{sys, {0, 0}}, inputs, {t, 0, 200}];
    GraphicsColumn@Table[Plot[out⟦i⟧, {t, 0, 200}, PlotRange → All, Ticks → {Automatic, {0, 1 / 5, 1, 2}}], {i, 9}])

    simulate[1.751, {1.876, .126, .126}, {.876, 1.6, .751}]
```