Augmenting soft computational machines with the addition of an intelligence prior

1st Christopher N. Singh
Department of Physics, Applied Physics, and Astronomy
Binghamton University
Binghamton, New York 13902, USA
csingh5@bighamton.edu

2nd Matthew Redell
Department of Physics, Applied Physics, and Astronomy
Binghamton University
Binghamton, New York 13902, USA
csingh5@bighamton.edu

3rd Mohannad Elhamod
Dept. of Computer Science
Virginia Tech
Blacksburg, VA 24060, USA
email address

4th Jie Bu
Dept. of Computer Science
Virginia Tech
Blacksburg, VA 24060, USA
email address

5th Anuj Karpatne
Dept. of Computer Science
Virginia Tech
Blacksburg, VA 24060, USA
email address

6th Wei-Cheng Lee
Department of Physics, Applied Physics, and Astronomy
Binghamton University
Binghamton, New York 13902, USA
csingh5@bighamton.edu

6th Wei-Cheng Lee
Department of Physics, Applied Physics, and Astronomy
Binghamton University
Binghamton, New York 13902, USA
csingh5@bighamton.edu

Abstract—**Improving the predictive power of machine learning models is one of the greatest challenges to the science of learning. Here we demonstrate with the simplest of neural networks that the addition of an intelligence prior can drastically improve the learning capabilities. We outline this simple mechanism to decrease the number of exposures, and enhance the predictive power with a number of examples relevant to the study of quantum phase transitions. We find that guided networks are uniquely capable to identify key features of quantum phases where unguided models fail, and that while the mean square error of topologically equivalent models may be commensurate, the structure of the predictions produced by the models is qualitatively very different. In many situations where knowledge of a physical system is available, but direct sampling of the entire phase space is computationally intractable, this approach offers a superior learning alternative.**

Index Terms—**component, formatting, style, styling, insert**

I. INTRODUCTION

The distillation of knowledge from data defines one of the most fundamental human endeavors – learning, and when viewed as the means by which actionable information is extracted from raw, unfiltered data, learning forms the basis of all intelligence. This fundamental principle has revitalized the science of learning in recent times, with great advancements in both hard and soft intelligence. And even though many of today's most challenging problems can be stated elegantly, consonant solutions are rarely, if ever, available. The consequence of this situation is that finding the best solution is almost always reduced to an optimization problem, where learning is maximized over computation.

Formulation as a minization problem, being unavoidable, affixes serious challenges to computation of expectations, namely, updating proceeds by estimation, with uncertainty inversely proportional to effort. In a world without infinite computational power, one must decide either to have fewer, accurate updates, or many, poorer updates. Computational intelligence, also known as machine learning techniques, have found success in tackling the optimization problem by mimicking biological systems, but still ultimately suffer from this fundamental complication. An exponentially or combinatorically exploding state space, in many ways, remains the final resolute barrier that still compels human mediations.

This work posits a mechanism to enhance the capability of soft intelligence machines' performance in this domain with the addition of an intelligence prior. In this approach, the validity of expectations are gauged with respect not only to the targets, but also required to be consistent with previously accumulated intelligence. We show that this extension pushes soft computational machines' extrapolative power beyond the feature space to which they were exposed, simply because each update is constrained to consistency not only with the targets, but also with a more general behavioural model.

As proof of concept, we apply this approach to identify quantum phase transitions in rings of qubits, as well as localization transitions in disordered, electron-lattice systems. The transverse field Ising chain is a quintessential physical model, thus is goverened by the Schrödinger equation; maintains an exponentially exploding state space, and has an analytical solution. As such, it is a perfect system from which to

extract the viability of this technique. Similarly, localization transitions manifest in the simplest of electron models, with the governing parameters being well studied analytically and numerically, and recent applications being suggested in hard computational intelligence, closing an elegant loop between the two fields.

Ultimately, we demonstrate that the ability of the most primitive of soft intelligence machines, the fully connected feedforward neural network, to predict quantum phase transitions can be drastically improved with the simple addition of a physics prior. The mean square error on the testing datasets drops much quicker in the physics guided models than in the black box counterparts, and the predictive power is greatly enhanced. In detecting a quantum phase transition, the physics guided machine can predict qualitative changes in the eigenspectrum that the black box machine entirely overlooks. While identifying a phase transition has been done before with black box models, these models were exposed to training instances in both phases. The guided approach codifies an approach that can predict a phase transition having never been explicitly exposed to it. Therefore, the evolution of physical observables like the magnetization and the fidelity susceptibility accross the transition, when sought by two networks of identical topology, can only be realized with learning guided by an intelligence prior. This result has deep implications for the field of computational intelligence, because it provides a route to augment learning in such a way that predictive power is maximized.

II. Introduction

The distillation of knowledge from data defines one of the most fundamental human endeavors – learning, and when viewed as the means by which actionable information is extracted from raw, unfiltered data, learning forms the basis of all intelligence. This fundamental principle has revitalized the science of learning in recent times, with great advancements in both hard and soft intelligence. And even though many of today's most challenging problems can be stated elegantly, consonant solutions are rarely, if ever, available. The consequence of this situation is that finding the best solution is almost always reduced to an optimization problem, where learning is maximized over computation.

Formulation as a minization problem, being unavoidable, affixes serious challenges to computation of expectations, namely, updating proceeds by estimation, with uncertainty inversely proportional to effort. In a world without infinite computational power, one must decide either to have fewer, accurate updates, or many, poorer updates. Computational intelligence, also known as machine learning techniques, have found success in tackling the optimization problem by mimicking biological systems, but still ultimately suffer from this fundamental complication. An exponentially or combinatorically exploding state space, in many ways, remains the final resolute barrier that still compels human mediations.

This work posits a mechanism to enhance the capability of soft intelligence machines' performance in this domain with the addition of an intelligence prior. In this approach, the validity of expectations are gauged with respect not only to the targets, but also required to be consistent with previously accumulated intelligence. We show that this extension pushes soft computational machines' extrapolative power beyond the feature space to which they were exposed, simply because each update is constrained to consistency not only with the targets, but also with a more general behavioural model.

As proof of concept, we apply this approach to identify quantum phase transitions in rings of qubits, as well as localization transitions in disordered, electron-lattice systems. The transverse field Ising chain is a quintessential physical model, thus is governed by the Schrödinger equation; maintains an exponentially exploding state space, and has an analytical solution. As such, it is a perfect system from which to extract the viability of this technique. Similarly, localization transitions manifest in the simplest of electron models, with the governing parameters being well studied analytically and numerically, and recent applications being suggested in hard computational intelligence, closing an elegant loop between the two fields.

Ultimately, we demonstrate that the ability of the most primitive of soft intelligence machines, the fully connected feedforward neural network, to predict quantum phase transitions can be drastically improved with the simple addition of a physics prior. The mean square error on the testing datasets drops much quicker in the physics guided models than in the black box counterparts, and the predictive power is greatly enhanced. In detecting a quantum phase transition, the physics guided machine can predict qualitative changes in the eigenspectrum that the black box machine entirely overlooks. While identifying a phase transition has been done before with black box models, these models were exposed to training instances in both phases. The guided approach codifies an approach that can predict a phase transition having never been explicitly exposed to it. Therefore, the evolution of physical observables like the magnetization and the fidelity susceptibility accross the transition, when sought by two networks of identical topology, can only be realized with learning guided by an intelligence prior. This result has deep implications for the field of computational intelligence, because it provides a route to augment learning in such a way that predictive power is maximized.

III. Methods

Because learning the energy whilst having the actual data, combined with the issues of outputs on various domains, does not work and is wasteful with information we already do have, we have decided to formulate the physics loss portion of the cost function in a way that is more transparent, and applicable to a wider range of problems. The condition we were using before is that the coefficients of the ground state wavefunction should satisfy the Schrödinger equation. However, the entire spectrum can be constrained in the exact same way. The

following of course will hold for any λ_i if every $\vec{\psi_i}$ is an eigenstate.

$$\sum_i \lambda_i(\hat{H}\vec{\psi_i} - E_i\vec{\psi_i}) = \vec{0} \tag{1}$$

This means we can formulate a matrix representation for the scalar physics cost with the help of two auxiliary definitions for an energy matrix and a Lagrange matrix. The matrix formulation of the equations is important for achieving vectorized operations at the implementation phase. If we define the energy matrix as

$$\hat{E} = \hat{I} \odot (\vec{1}^T \otimes \vec{E}), \tag{2}$$

where $\vec{E}$ is a column vector of the eigenvalues. Likewise we can define for the Lagrange matrix

$$\hat{\lambda} = \hat{I} \odot (\vec{1}^T \otimes \vec{\lambda}). \tag{3}$$

We always stick with Hessian notation, so that the primary vectorial objects are column vectors and matrices are collections of column vectors. In this sense $\vec{1}$ is a column vector of ones and $\vec{0}$ is a column vector of zeros.

Using these definitions, the scalar physics cost function can be defined as

$$C_2 = \frac{1}{2}||\hat{H}\hat{\Psi} - \hat{\Psi}\hat{E}||_F^2. \tag{4}$$

The backpropagation equations require the derivative of the cost function, so here we can take the matrix derivative

$$\frac{\partial C_2}{\partial \hat{\Psi}} = \hat{H}^2\hat{\Psi} - 2\hat{H}\hat{\Psi}\hat{E} + \hat{E}^2\hat{\Psi} \tag{5}$$

In this way, the elements of the result correspond to the gradient of the physics portion of the cost. This can be much more straight forwardly implemented with faster vectorized operations. In the matrix implementation of backpropagation, this expression is evaluated for each instance in the mini batch.

Finally, the method that has shown the most success has been a direct modulation of the error at the output layer.

$$\hat{\delta}^L + = \lambda(\hat{H}\hat{\Psi} - \hat{\Psi}\hat{E}) \tag{6}$$

IV. RESULTS

V. DISCUSSION

We have presented two different methods for guided learning of the neural networks, one in which the landscape of the loss function is altered via an additional term C_2, the other in which the error term is directly modulated by an additional guiding function. To illustrate the differences in these methods and determine which method will result in faster learning, we want to consider a simple example. Consider a simple feed forward network with two input layer neurons $\vec{a}^0 = (a_1^0, a^0, 2)^T$, which connect to a single output neuron a^1 via a weight vector $\vec{w} = (w_1, w_2)$ and an activation function $\sigma(z)$ such that

$$a_i^1 = \sigma(z_i) \tag{7}$$

with $z_i = w_1 a_{1,i}^0 + w_2 a_{2,i}^0$ for the ith data point. A schematic diagram of this network can be seen in Fig.**input a figure**.

Now, following the standard backpropagation algorithm, we can calculate the error in the output layer for each input data point i as

$$\delta_i^1 = \frac{\partial C}{\partial a_i^1}\sigma'(z_i) \tag{8}$$

where C is the loss function used and $\sigma'(z)$ is the derivative of the chosen activation function. Furthermore, after we iterate through the training set, we want to update the weight vector in accordance with this error. We do so as

$$\vec{w} = \vec{w} - \frac{\eta}{N}\sum_{i=1}^N \delta_i^1 \vec{a}_i^0 \tag{9}$$

where η is the prescribed learning rate and N is the number of training data points. To distinguish which of our methods will learn at a faster rate, let's assume that there exists a set number of training epochs M after which the network will be sufficiently trained. Thus, the initial weight vector $\vec{w}(0)$ will differ from the converged weight vector $\vec{w}(M)$ by a vector

$$\vec{s} = \vec{w}(M) - \vec{w}(0) \tag{10}$$

So, we can write the backpropagation of the error for the mth training epoch as

$$\vec{w}(m+1) = \vec{w}(m) - \frac{\eta}{N}\sum_{i=1}^N \delta_i^1(m)\vec{a}_i^0, \tag{11}$$

where $\delta_i^1(m)$ is calculated with respect to the output neuron value of during the mth training epoch $a_i^1(m)$. This means we can characterize the distance that the mth training epoch is from the converged result as

$$\vec{s}(m) = \vec{w}(M) - \vec{w}(m). \tag{12}$$

A simple measure of the speed of training for the network thus could be found from $\Delta_m \vec{s}(m)$, the change in the difference vector as a function of m

$$\begin{aligned} \Delta_m \vec{s}(m) &= \Delta_m(\vec{w}(M) - \vec{w}(m)) \\ &= -\Delta_m \vec{w}(m) \\ &= \frac{\eta}{N}\sum_{i=1}^N \delta_i^1(m)\vec{a}_i^0. \end{aligned} \tag{13}$$

This indicates that the most important indicator of convergence speed will be the quality of the error function $\delta_i^1(m)$.

Now, discuss the intracacies of the 3 different methods, BB, C_2, and δ

VI. CONCLUSION

We have demonstrated that the addition of an intelligence prior can greatly improve the learning capabilities and predictive power of soft computational intelligence engines. In the most rudimentary of circumstances, the improvements are vast. With the ever increasing importance of precipitating valuable information from effecitvely infinite amounts of chaotic data, succint training routines for neural networks are critical. Although we have demonstrated the efficacy of this approach as applied to quantum phase transitions in model systems, the

basic premise is widely applicable to any system that follows known laws of motion in the generalized sense. This strongly suggests the art of machine intelligence should be refined to include the aforementioned precepts, and we suspect that leveraging this technique will find application accross many areas of physics and engineering.

ACKNOWLEDGMENT

REFERENCES

REFERENCES

[1] P. W. Anderson. Absence of diffusion in certain random lattices. *Phys. Rev.*, 109:1492–1505, Mar 1958.

[2] Sergei Baranovski. *Charge transport in disordered solids with applications in electronics*, volume 17. John Wiley & Sons, 2006.

[3] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953–17979, Dec 1994.

[4] Tobias Brandes and Stefan Kettemann. *Anderson localization and its ramifications: Disorder, phase coherence, and electron correlations*, volume 630. Springer Science & Business Media, 2003.

[5] David E Carlson and Cristopher R Wronski. Amorphous silicon solar cell. *Applied Physics Letters*, 28(11):671–673, 1976.

[6] Lidia Ceriani and Paolo Verme. The origins of the gini index: extracts from variabilità e mutabilità (1912) by corrado gini. *The Journal of Economic Inequality*, 10(3):421–443, 2012.

[7] Jae-yoon Choi, Sebastian Hild, Johannes Zeiher, Peter Schauß, Antonio Rubio-Abadal, Tarik Yefsah, Vedika Khemani, David A. Huse, Immanuel Bloch, and Christian Gross. Exploring the many-body localization transition in two dimensions. *Science*, 352(6293):1547–1552, 2016.

[8] L. Chua. Memristor-the missing circuit element. *IEEE Transactions on Circuit Theory*, 18(5):507–519, Sep. 1971.

[9] Leon Chua. Resistance switching memories are memristors. *Applied Physics A*, 102(4):765–783, Mar 2011.

[10] Alexander L Despotuli and Alexandra V Andreeva. Spatial averaging of electrostatic potential differences in layered nanostructures with ionic hopping conductivity. *Journal of Electroanalytical Chemistry*, 829:1–6, 2018.

[11] Philip G Emma. Understanding some simple processor-performance limits. *IBM journal of Research and Development*, 41(3):215, 1997.

[12] Ferdinand Evers and Alexander D. Mirlin. Anderson transitions. *Rev. Mod. Phys.*, 80:1355–1417, Oct 2008.

[13] M. V. Fischetti. Master-equation approach to the study of electronic transport in small semiconductor devices. *Phys. Rev. B*, 59:4901–4917, Feb 1999.

[14] Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500–544, 1952.

[15] Arkadiusz Jdrzejewski, Anna Chmiel, and Katarzyna Sznajd-Weron. Kinetic ising models with various single-spin-flip dynamics on quenched and annealed random regular graphs. *Phys. Rev. E*, 96:012132, Jul 2017.

[16] Anton Kokalj. Xcrysden—a new program for displaying crystalline structures and electron densities. *Journal of Molecular Graphics and Modelling*, 17(3-4):176–179, 1999.

[17] G. Kresse and J. Hafner. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B*, 47:558–561, Jan 1993.

[18] Olga Krestinskaya, Aidana Irmanova, and Alex Pappachen James. Memristors: Properties, models, materials. In *Deep Learning Classifiers with Memristive Networks*, pages 13–40. Springer, 2020.

[19] Aliaksandr V. Krukau, Oleg A. Vydrov, Artur F. Izmaylov, and Gustavo E. Scuseria. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *The Journal of Chemical Physics*, 125(22):224106, 2006.

[20] Ryogo Kubo. Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12(6):570–586, 1957.

[21] Deok-Hwang Kwon, Kyung Min Kim, Jae Hyuck Jang, Jong Myeong Jeon, Min Hwan Lee, Gun Hwan Kim, Xiang-Shu Li, Gyeong-Su Park, Bora Lee, Seungwu Han, et al. Atomic structure of conducting nanofilaments in tio2 resistive switching memory. *Nature nanotechnology*, 5(2):148, 2010.

[22] Wei Lu. Going active. *Nature Materials*, 12:93–94, 2012.

[23] Abdul N. Malmi-Kakkada, Oriol T. Valls, and Chandan Dasgupta. Ising model on a random network with annealed or quenched disorder. *Phys. Rev. B*, 90:024202, Jul 2014.

[24] Koichi Momma and Fujio Izumi. Vesta 3 for three-dimensional visualization of crystal, volumetric and morphology data. *Journal of applied crystallography*, 44(6):1272–1276, 2011.

[25] Don Monroe. Hopping in exponential band tails. *Phys. Rev. Lett.*, 54:146–149, Jan 1985.

[26] AA Mostofi, JR Yates, G Pizzi, YS Lee, I Souza, D Vanderbilt, and N Marzari. An updated version of wannier90 a tool for obtaining maximally-localised wannier functions. *Computer Physics Communications*, 185(8):2309 – 2310, 2014.

[27] C. Nico, T. Monteiro, and M.P.F. Graça. Niobium oxides and niobates physical properties: Review and prospects. *Progress in Materials Science*, 80:1 – 37, 2016.

[28] Kentaro Nomura and A. H. MacDonald. Quantum transport of massless dirac fermions. *Phys. Rev. Lett.*, 98:076602, Feb 2007.

[29] Debashis Panda, Paritosh Piyush Sahu, and Tseung Yuen Tseng. A collective study on modeling and simulation of resistive random access memory. *Nanoscale Research Letters*, 13(1):8, Jan 2018.

[30] John P. Perdew, Adrienn Ruzsinszky, Gábor I. Csonka, Oleg A. Vydrov, Gustavo E. Scuseria, Lucian A. Constantin, Xiaolan Zhou, and Kieron Burke. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.*, 100:136406, Apr 2008.

[31] Matthew D Pickett, Gilberto Medeiros-Ribeiro, and R Stanley Williams. A scalable neuristor built with mott memristors. *Nature materials*, 12(2):114, 2013.

[32] James R Powell. The quantum limit to moore's law. *Proceedings of the IEEE*, 96(8):1247–1248, 2008.

[33] J. Rau and B. Müller. From reversible quantum microdynamics to irreversible quantum transport. *Physics Reports*, 272(1):1 – 59, 1996.

[34] Muhammad Sahimi, Barry D. Hughes, L. E. Scriven, and H. T. Davis. Stochastic transport in disordered systems. *The Journal of Chemical Physics*, 78(11):6849–6864, 1983.

[35] H. Scher and M. Lax. Stochastic transport in a disordered solid. i. theory. *Phys. Rev. B*, 7:4491–4502, May 1973.

[36] H. Scher and M. Lax. Stochastic transport in a disordered solid. ii. impurity conduction. *Phys. Rev. B*, 7:4502–4519, May 1973.

[37] Karlheinz Schwarz, Peter Blaha, and Georg KH Madsen. Electronic structure calculations of solids using the wien2k package for material sciences. *Computer Physics Communications*, 147(1-2):71–76, 2002.

[38] Joshua C Shank, M Brooks Tellekamp, Matthew J Wahila, Sebastian Howard, Alex S Weidenbach, Bill Zivasatienraj, Louis FJ Piper, and W Alan Doolittle. Scalable memdiodes exhibiting rectification and hysteresis for neuromorphic computing. *Scientific reports*, 8(1):12935, 2018.

[39] T Siegrist, P Jost, H Volker, M Woda, P Merkelbach, C Schlockermann, and M Wuttig. Disorder-induced localization in crystalline phase-change materials. *Nature materials*, 10(3):202, 2011.

[40] Dmitri B Strukov, Gregory S Snider, Duncan R Stewart, and R Stanley Williams. The missing memristor found. *nature*, 453(7191):80, 2008.

[41] David J Thouless. Electrons in disordered systems and the theory of localization. *Physics Reports*, 13(3):93–142, 1974.

[42] Léon van Hove. Quantum-mechanical perturbations giving rise to a statistical transport equation. *Physica*, 21(1):517 – 540, 1954.

[43] Zhiyong Wang, Laiyuan Wang, Masaru Nagai, Linghai Xie, Mingdong Yi, and Wei Huang. Nanoionics-enabled memristive devices: Strategies and materials for neuromorphic applications. *Advanced Electronic Materials*, 3(7):1600510, 2017.

[44] Rainer Waser, Regina Dittmann, Georgi Staikov, and Kristof Szot. Redox-based resistive switching memories–nanoionic mechanisms, prospects, and challenges. *Advanced materials*, 21(25-26):2632–2663, 2009.

[45] J Joshua Yang, Matthew D Pickett, Xuema Li, Douglas AA Ohlberg, Duncan R Stewart, and R Stanley Williams. Memristive switching mechanism for metal/oxide/metal nanodevices. *Nature nanotechnology*, 3(7):429, 2008.

[46] Wei Zhang, Riccardo Mazzarello, Matthias Wuttig, and Evan Ma. Designing crystallization in phase-change materials for universal memory and neuro-inspired computing. *Nature Reviews Materials*, page 1, 2019.

[47] Wei Zhang, Alex Thiess, Peter Zalden, R Zeller, PH Dederichs, Jean-Yves Raty, Matthias Wuttig, S Blügel, and Riccardo Mazzarello. Role of vacancies in metal–insulator transitions of crystalline phase-change materials. *Nature materials*, 11(11):952, 2012.

[48] Robert Zwanzig. Ensemble method in the theory of irreversibility. *The Journal of Chemical Physics*, 33(5):1338–1341, 1960.