Neural networks as "hidden" variable models for quantum systems

Steven Weinstein[*]

Perimeter Institute for Theoretical Physics
University of Waterloo

July 10, 2018

Abstract

We successfully model the behavior of two-spin systems using neural networks known as conditional Restricted Boltzmann Machines (cRBMs). The result gives local "hidden" variable models for product states and entangled states, including the singlet state used in the EPR-Bohm experiment. Bell's theorem is circumvented because the state of the system is dependent not only on the preparation but also on the measurement setup (the detector settings). Though at first glance counterintuitive, the apparent "retrocausality" in these models has a historical precedent in the absorber theory of Wheeler and Feynman [39] and an intuitive analog in the simple AC circuit of an electric guitar.

1 Introduction

In 1935, Einstein, Podolsky and Rosen (EPR) gave a powerful argument that quantum mechanics is incomplete [17]. Their central point is that if the predictions of quantum theory are correct, then the assumption of relativistic causality implies that there are facts about the world – which EPR called 'elements of reality' – that are not reflected in the quantum state.[1] Thirty years later, Bell [5] proved that it is impossible to find a complete theory that satisfies certain seemingly natural constraints having to do with locality. The take-away for many has been that any theory describing quantum phenomena must be nonlocal in a way that would have been anathema to EPR, violating the spirit, if not the letter, of the constraints causality enshrined in special relativistic theories.

In this paper, we offer a framework that provides a completion of quantum theory which shares much with classical statistical mechanics, and with a certain class of neural networks called Boltzmann machines [25, 1, 23]. Predictions in this framework are derived from the properties of a quasi-thermal ensemble, the boundary conditions of which are set by the classical configurations of the preparation and measurement apparatuses. Bell's argument is evaded because the assumption that the complete description of the system prior to measurement is *independent* of the final measurement settings – an assumption called Statistical Independence (SI) – is violated.[2] In fact, we will see in section 8 that the framework generalizes in a natural way to allow for maximal violations of the Bell-inequality, giving a model that becomes both deterministic and nonlocal in a way that nevertheless does not allow superluminal signaling.[3]

---

[*]ssweinstein@pitp.ca

[1]See [18] for a helpful reconstruction of the argument.

[2]SI is related to, but not the same as, measurement independence (MI). The latter postulates that the detector settings are independent of the properties of the system being under measured. It is also known as the "free choice" assumption, reflecting the idea that this is tantamount to the assumption that the settings can be freely chosen.

[3]The model differs significantly from hidden-variable models like Bohmian mechanics [7, 8] and Nelson's stochastic mechanics [29, 30], in that the stochasticity is not essential to avoiding the spectre of superluminal signaling.

2   Restricted Boltzmann Machines

A Restricted Boltzmann Machine (RBM) [33] is a type of generative neural network with the topology of a bipartite graph (see Figure 1). For the models we work with, the nodes or units
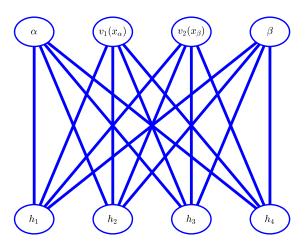

Figure 1: Simple RBM for EPR with four hidden units

(as the 'neurons' are called) are binary, as in the Ising model. Edges are weighted connections between the units. Positive weights (excitatory connections) and negative weights (inhibitory connections) respectively encode correlations and anti-correlations between units. Each unit has a bias, which corresponds to an adjustable firing or activation threshold. The *visible* units (one side of the graph) are intended to represent (binary) properties of the data, while the *hidden* units (the other side of the graph) encode correlations between these properties.

For example, suppose our machine is supposed to learn a set of black-and-white images containing 16 x 16 = 256 pixels. We will want to have 256 visible units $\mathbf{v} = (v_1, v_2, ..., v_{256})$, one per pixel. The weights and biases will make some visible configurations more probable than others according to the *ansatz* below. Training the machine involves systematically adjusting these weights and biases so that the machine assigns high probability to images like those in the data set. Thus a machine trained on 16 x 16 images of dogs and cats should generate images of dogs and cats, including but perhaps not limited to those found in the training data. The hidden units $\mathbf{h} = (h_1, h_2, ..., h_n)$ encode properties of dogs and cats, and the visible units manifest various combinations of these properties so that sampling them will yield images of dogs and cats.

An RBM is a type of energy-based model. To each configuration $(\mathbf{v}, \mathbf{h})$ there corresponds an energy

$$E(\mathbf{v}, \mathbf{h}) = -\left(\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij} v_i h_j + \sum_{i=1}^{m} c_i v_i + \sum_{j=1}^{n} d_j h_j \right). \tag{1}$$

and the probability of the configuration is given by the Boltzmann distribution for a system in thermal equilibrium at temperature T:

$$P(\mathbf{v}, \mathbf{h}) = \frac{e^{-E(\mathbf{v}, \mathbf{h})/T}}{Z}, \text{ where } Z = \sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})/T}. \tag{2}$$

2

This will look familiar to any condensed matter physicist, as it has the form of the Ising model Hamiltonian.[4] Just as in the Ising model, w_{ij} is a real-valued weight, a connection strength between units i and j, while c_i and d_j are real-valued biases of visible unit v_i and hidden unit h_j, respectively. (In the Ising model, these would correspond to the strength of the transverse magnetic field at different sites.) The important difference between an RBM and an Ising model is the topology; whereas an Ising model typically has the connectivity of a lattice, an RBM is a bipartite graph, where each visible node is connected to every hidden node, and vice-versa. Also, whereas the Ising model units take values ± 1, the Boltzmann machine units take values $\{0, 1\}$. Appropriate adjustment of the biases allows one to move from one representation to the other. Unless otherwise indicated, we will henceforth work at fixed temperature and set $T = 1$.

The probabilities of *observable* quantities are the probabilities of various configurations of the visible units. These are obtained by using the *free energy* $\mathcal{F}$ associated with a visible configuration $\mathbf{v}$:

$$\mathcal{F}(\mathbf{v}) \equiv -\log \sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})}. \tag{3}$$

This gives the effective energy needed to determine the marginal probability of all configurations that share the visible configuration $\mathbf{v}$. The probability itself is then given by

$$P(\mathbf{v}) = \frac{e^{-\mathcal{F}(\mathbf{v})}}{Z} \text{ with } Z = \sum_{\mathbf{v}'} e^{-\mathcal{F}(\mathbf{v}')}. \tag{4}$$

RBMs are local in the same sense that the Ising model is local. The probability that a given unit will be 'on' depends only on the the state of the units to which it's connected. More specifically,

$$P(v_i = 1) = \frac{1}{1 + e^{-\Delta E_i}} \tag{5}$$

where

$$\begin{aligned} \Delta E_i &= E_{v_i=0} - E_{v_i=1} \\ &= c_i + \sum_j w_j\, h_j, \end{aligned} \tag{6}$$

so that the probability that v_i is on depends only on the state of the hidden units h_j to which it is directly connected. The bias c_i and the weights w_j are fixed parameters in the model.

3   RBMs for EPR

An RBM is a useful way of learning and encoding a joint probability distribution. Consider the modern version of the EPR experiment [17] due to Bohm [6]. It involves spin measurements on a pair of particles prepared in a maximally entangled 'Bell' state, specifically two spin $\frac{1}{2}$ particles in the singlet state

$$\psi_{singlet} = \frac{1}{\sqrt{2}} \big(|{+}{-}\rangle - |{-}{+}\rangle \big). \tag{7}$$

One particle travels to station A and the other to B. At each station is a detector with two settings, each setting selecting a component of spin to be measured. The detector settings $\alpha \in \{a, a'\}$ and $\beta \in \{b, b'\}$ and measurement outcomes $x_\alpha \in \{+1, -1\}$ and $x_\beta \in \{+1, -1\}$ are two-valued, so we can treat them as binary random variables. A single experimental trial $(\alpha, \beta, x_\alpha, x_\beta)$ can thus be represented as a binary vector $\mathbf{v} = (v_1, v_2, v_3, v_4)$, where $v_i \in \{0, 1\}$. This vector represents the values taken by the four visible units that make up one layer of the RBM. (We map $+1$ to 0 and -1 to 1.)

---

[4]This is no coincidence: the precursor to the Boltzmann machine was an associative memory model invented by physicist John Hopfield, who found his inspiration in the theory of spin glasses [27].

We need to pick a sufficient number of hidden units $\mathbf{h} = (h_1, h_2, ... h_j)$ to encode the dependencies between the visible units. In [38], we used four hidden units, which was sufficient to allow the RBM to represent the quantum-mechanically predicted statistics of an EPR experiment to very high accuracy. The topology of the machine is that shown in Figure 1. Beginning with randomly chosen weights and biases, we trained the machine on a simulated data set from an EPR experiment using persistent contrastive divergence [36] to adjust the weights and biases, until the probabilities generated by the machine aligned with those in the data. For example, if $\mathbf{v} = (1, 1, 1, 0)$, appears in 5.2% of the trials, we would expect that the probability assigned to that vector by the machine would be appropriately close to 5.2%, so that $P(\mathbf{v}) = .052$, where the probability is given by eqn. 4.

In encoding the *joint* probability $P(\alpha, \beta, x_\alpha, x_\beta)$, we have arguably encoded more information than necessary, if what we are trying to do is reproduce the predictions of quantum mechanics. After all, quantum mechanics gives us only *conditional* probabilities $P(x_\alpha, x_\beta | \alpha, \beta)$. It tells us what we can expect to observe if we measure certain observables, but it does not tell us how often we can expect to measure those observables. Yet in specifying a joint probability distribution, we are effectively specifying both of these things. At best, this means we are squandering the representational capacity of our model on quantities we're not interested in. So we utilize a variant on an RBM called a conditional RBM (cRBM) to eliminate this redundancy.

The essence of a cRBM is that it is able to encode conditional probability distributions directly, i.e., without deriving them from a joint probability distribution. This is done by allowing the visible biases, the hidden biases, or the weights to vary according to the condition or conditions in question. In our cRBM, we will allow each detector setting to influence the bias of the hidden units. For the purposes of illustration, consider a cRBM with a single hidden unit (figure 2). We have two visible units to represent the two outcomes x_α and x_β, and two

(a) Conditional RBM

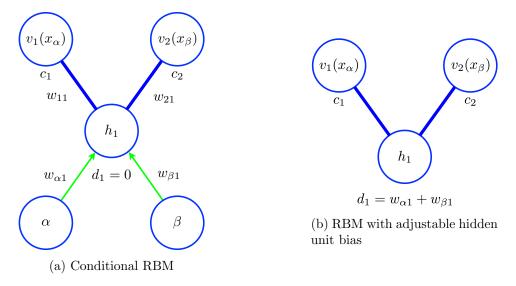(b) RBM with adjustable hidden unit bias

Figure 2: Equivalent representations

additional conditional units representing the detector settings. They are composite 'one-hot' units each consisting of a set of binary units, one for each setting. At any time, one of the (binary) units is on and the others off, so the unit as a whole is always on. As a result, the links between these conditional units and the hidden units are effectively one-way, directed links, since the hidden units have no effect on the conditional units. In fact, the conditional units act like adjustable biases on the hidden units. If the bias $d_1 = 0$, then one can view the machine shown in figure 2a as an ordinary RBM with an adjustable bias $d_1 = w_{\alpha 1} + w_{\beta 1}$, as in figure 2b. This is what we want: the detector settings are external inputs which have an effect on the parameters (here, the hidden unit biases) of the RBM model.

Training a cRBM is similar to training an RBM. As with an RBM, the machine learns

4

weights and biases so as to reproduce, in the activations of the visible units, the patterns found in the data. Let $\mathbf{u_1}$ correspond to the setting α of detector A, and $\mathbf{u_2}$ correspond to setting β of detector B. For a given pair of settings $\mathbf{u} = (\mathbf{u_1}, \mathbf{u_2})$, the energy of the RBM is

$$E(\mathbf{v}, \mathbf{u}, \mathbf{h}) = -\left(\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij} v_i h_j + \sum_{i=1}^{m} c_i v_i + \sum_{j=1}^{n} d_j h_j + \frac{1}{2} \sum_{j=1}^{n} \sum_{l=1}^{2} \sum_{k=1}^{2} w_{jlk} u_{lk} h_j \right). \quad (8)$$

where w_{jlk} refers to the weight connecting hidden unit h_j to the k'th setting of one-hot unit $\mathbf{u_l}$. Then the conditional probabilities are simply

$$P(\mathbf{v}|\mathbf{u}) = \frac{\sum_h e^{-E(\mathbf{v},\mathbf{u},\mathbf{h})}}{\sum_{v',u,h} e^{-E(\mathbf{v'},\mathbf{u},\mathbf{h})}}$$
$$= \frac{e^{-\mathcal{F}(\mathbf{v},\mathbf{u})}}{Z}. \quad (9)$$

Training involves tailoring *all* the weights, including the w_{jlk} weights connecting the conditional units to the hidden units. These weights encode the influence of the detector settings on the hidden units.

What differentiates the cRBM from a regular RBM is that the units corresponding to the conditioning variables – in this case, the detector settings – are *not* dynamical variables. Though the energy associated with a configuration is a function in part of the weights connecting the conditioning units to the hidden units, the sum over the configurations in the partition function involves only configurations in which the detectors and the preparation (the input state) have some specific value. Thus there are no probabilities assigned to detector settings or states. The only probabilities generated by the cRBM are conditional probabilities, conditional on the states and detector settings, just as in ordinary quantum mechanics.

4   Results

We constructed a variety of cRBM models, and learned the weights and biases using simple stochastic gradient descent, as described in, e.g., [24]. All of the models have *three* conditioning units, one for each of two detectors (as above) and an additional one corresponding to the choice of input state. They vary in the number of possible settings for each detector (2 or 8), the number of states (singlet only, or singlet plus two product states), and the number of hidden units. The intent was not only to model one particular experiment, the EPR experiment, but also to explore the ability of cRBMs to model generic two-particle systems.

A simple cRBM model with three hidden units easily reproduces the success of the (non-conditional) RBM model with four hidden units discussed previously. The topology of the model is shown in figure 3, and the results after training are shown (to two decimal places) in figure 4.

Aiming to generalize the result, we constructed a model with *eight* possible settings between 0 and $7\pi/8$ for each detector, still for a single entangled state. The topology is the same as before. The results are shown in figure 5. Even with eight settings, three hidden units were sufficient to give very good results. Taking a look at the directed links from the units modeling the detectors and the three hidden units in figure 6, we see a more or less linear relationship between detector setting and weight for the connections between both of the detectors and two of the three hidden units, while the other unit displays only subtle variation for each of the detectors.

The next step was to try to model the same measurements on systems prepared in different quantum states. In addition to the singlet (eqn 7), we chose the two states $|+-\rangle$ and $|-+\rangle$. Given that a general quantum state of two particles requires seven real numbers to specify (four complex numbers c_i such that $\sum_{i=1}^{4} |c_i|^2 = 1$) , it is not that surprising that we needed eight hidden units to get the level of accuracy seen in figures 7a, 7b and 7c.
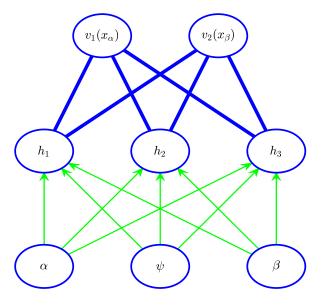
5

Figure 3: Conditional RBM for EPR with three hidden units

5   Bell's theorem and cRBM models

The cRBM models provide a template for a "completion" of quantum mechanics. Bell's theorem is widely understood to show that no *local* theory can reproduce all the predictions of quantum mechanics. So how does our model get around this?

Bell works in a framework where each choice of state (each choice of values for the dynamical variables or 'beables', hidden or not) leads to probabilities for the two possible outcomes of spin measurements on an entangled pair of spin-$\frac{1}{2}$ particles. If the model that gives rise to the probabilities satisfies certain properties, it must satisfy the Bell-CHSH inequality [11]: that is Bell's theorem. For appropriate choice of detector settings – appropriate choice of components of spin to be measured – quantum mechanics predicts, and experiment appears to corroborate [11, 19, 3], that this inequality is violated. This shows that a theory which satisfies the conjunction of Bell's 'locality' condition and in which the detector settings and dynamical variables are independent cannot reproduce the predictions of quantum mechanics. The cRBM models reproduce the quantum predictions but do not violate Bell's theorem because although they are 'local', they do not satisfy SI, the condition of statistical independence. We have previously dubbed this *nonlocality without nonlocality* [37].

Bell's analysis is intended to characterize a wide variety of candidate replacements or completions of quantum mechanics. The supposition is that in any given trial the system is in some state λ drawn from a space of states Λ defined by the model, and that to each such state there corresponds a prediction (in general, stochastic) about the results of measurements on each of the two particles. Naturally, the probability that a given state λ will be realized is going to depend in part on the conditions c that generate the pair of particles. Thus we speak of the the conditional probability $P(\lambda|c)$. This much conditional dependence of the λ on external boundary conditions is uncontroversial, and is assumed by Bell.

Now, each state λ implies probabilities $P(x_\alpha x_\beta|\alpha, \beta, \lambda)$ which yield an expectation value
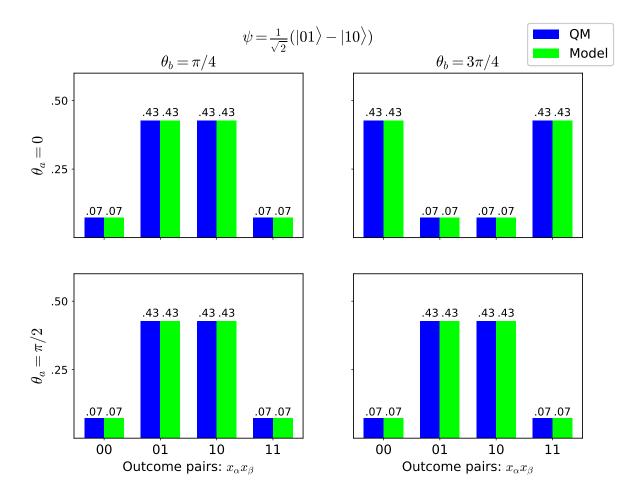
6

Figure 4: Predictions for cRBM model with three hidden units

$-1 \leq E(\alpha, \beta | \lambda) \leq +1$ for the product of the outcomes $x_\alpha, x_\beta \in \{+1, -1\}$:

$$E(\alpha, \beta | \lambda) \equiv \sum_\lambda \sum_{\alpha, \beta} x_\alpha x_\beta P(x_\alpha x_\beta | \alpha, \beta, \lambda) \tag{10}$$

The expectation value is then the weighted sum:

$$E_c(\alpha, \beta) = \sum_\lambda E(\alpha, \beta | \lambda) P(\lambda | c) \tag{11}$$

Expressions of the form $S = \sum_{\alpha, \beta} | \pm E(\alpha, \beta)|$ must take a value $S \leq 4$, since each expectation value has a maximum value of $+1$ and a minimum value of -1. In particular,

$$S = |(E(a, b) + E(a', b) + E(a, b') - E(a', b'))| \leq 4. \tag{12}$$

For spatially separated measurements, one might suppose that, given a particular state, the probabilities factorize as follows:

$$P_\lambda(x_\alpha, x_\beta | \alpha, \beta) = P_\lambda(x_\alpha, \alpha) * P_\lambda(x_\beta | \beta). \tag{13}$$

This is Bell's 'locality' condition, later distinguished by him from the 'local causality' condition, which is the conjunction of locality and the additional condition of statistical independence (SI),

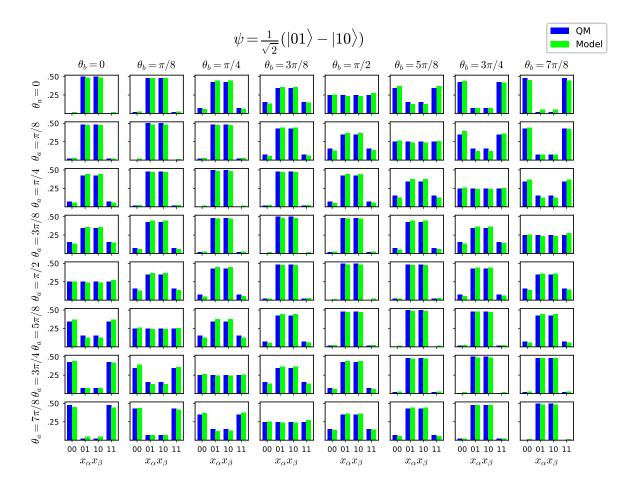$$P(\lambda | c) = P(\lambda | c, \alpha, \beta), \tag{14}$$

7

Figure 5: Predictions for cRBM model with eight hidden units: Singlet state

which implies that the detector settings do not constrain the state λ. Together, locality and SI imply the CHSH-Bell inequality: $S \leq 2$. But if the conditions c are such that the quantum-mechanical description is the singlet state, and the detector settings $(a, a', b, b') = (0, \pi/2, \pi/4, 3\pi/4)$, then quantum mechanics predicts that $S = 2\sqrt{2}$.

6 Physical models with future boundary conditions

In the CRBM model, the state of the hidden units is a function of both the input conditions (which we refer to as ψ), and the detector settings α and β. The machine is in a state of quasi-thermodynamic equilibrium at temperature $T = 1$ with future and past boundary conditions, so that it forms a kind of spacetime ensemble. Arrayed in spacetime, with time running from bottom to top, we can view the machine depicted in figure 3 in the manner depicted in figure 8.

Models in which "freely specifiable" future boundary conditions constrain the state of the system earlier in time are often called 'retrocausal' [12, 13, 13, 2, 32, 28], and have been the subject of much recent discussion. Yet the idea is arguably not that novel. The most prominent example of such a theory is indeed ordinary classical mechanics, formulated in the Lagrangian manner. Though the usual use of the Lagrangian formulation is to derive the equations of motion from arbitrary initial and final system configurations, one can also specify particular initial and final configurations and derive the trajectory from the principle of least action. What one
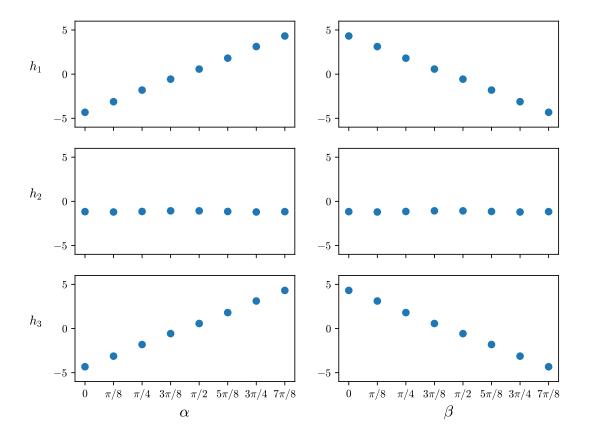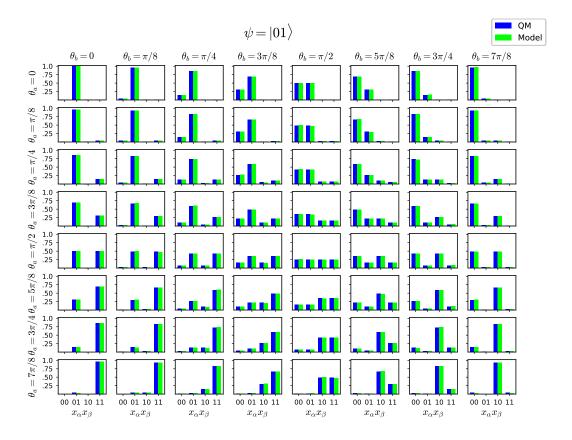
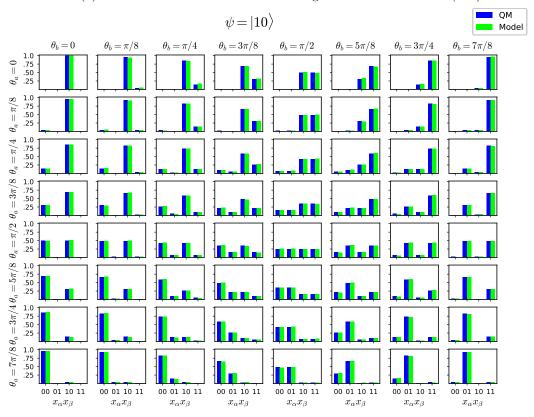Figure 6: Input weights for detectors A and B

cannot do, of course, is to specify initial and final configurations *and* velocities. In that case, the problem is overdetermined, meaning that in general there will be no solution at all. But specifying partial data at the initial and final times leads to a perfectly well-posed problem. And given that we typically regard configurations of physical systems as "freely specifiable", Lagrangian models would superficially appear to be candidates for the retrocausal appellation. Why they are not is a subtle and interesting matter we will not take up here.

The Lagrangian formulation was indeed the inspiration for the reformulation of Maxwell's electrodynamics known as the Wheeler-Feynman absorber theory [39, 15, 16]. This theory treated the electromagnetic field as a useful fiction, and derived the standard behavior of charged particles under the assumption that all radiation – if it actually existed – would be absorbed in the future. Thus in this theory, there is no source-free radiation, and the initial and final charge distributions determine the behavior of charges in the intermediate region. The primary drawback is that the universe does not appear to be configured such that all radiation is ultimately absorbed, since the expansion renders it largely transparent [26].
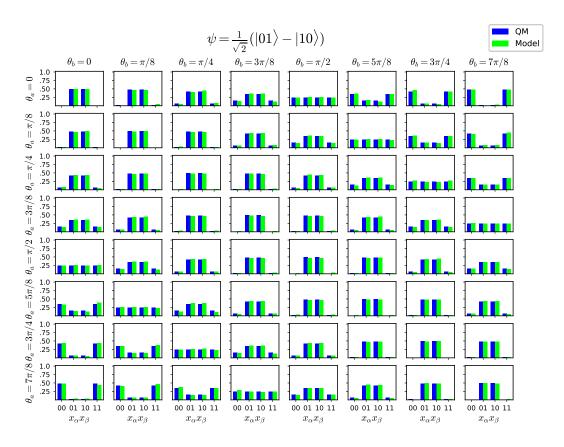
An everyday example of a dynamical system governed by past and future boundary conditions is the most important instrument of the 20th century, the electric guitar. At its heart is a simple, damped resonator called a 'pickup', which consists of a magnet within or underneath a coil of wire. The magnet magnetizes the strings, and the strings' oscillation drives the circuit, inducing an alternating current in the coil via electromagnetic induction in accord with Faraday's law. Thus the pickup serves as a transducer, converting the mechanical oscillation of the string into an electrical oscillation. But the pickup's frequency response is not flat, so it colors

(a) Predictions for cRBM model with eight hidden units: State $|+-\rangle$

(b) Predictions for cRBM model with eight hidden units: State $|-+\rangle$

(c) Predictions for cRBM model with eight hidden units: Singlet state

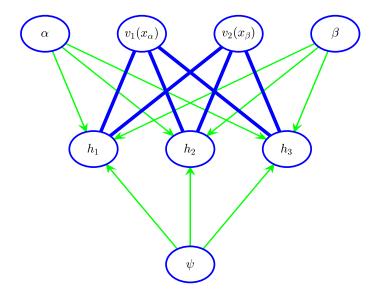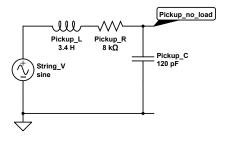Figure 7: Predictions for cRBM model with eight hidden units: Three states

Figure 8: Conditional RBM for EPR with three hidden units

the signal: it is a second-order low-pass filter. Just like a mechanical oscillator, the pickup has a resonant frequency at which it is most easily driven. (See figure 9.)
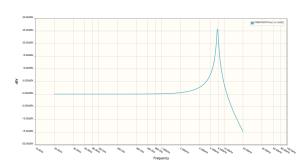
The frequency response of the pickup depends not only on its internal properties (most importantly the inductance of the coil), but on the properties of the objects *receiving* the signal, later in time. The resonant frequency of the pickup $f_r = \sqrt{1/LC}$ is a function not only of the inductance L_p and capacitance C_p of the coil, but also the capacitance of the cable connecting the guitar to the amplifier.

Under no external load, the pickup (figure 9a) reacts to the input signal as depicted in figure 9b. When the guitar is plugged into an amplifier, the frequency response changes, depending on the load. In figure 9d, we see the frequency response at an amplifier with input impedance of 1 MOhm using cables with capacitance ranging from 50 pf to 450 pf. What is of great interest for our purposes is that the behavior at the pickup is the same as the behavior at the amp. The cable affects the way the input signal is altered, but it does so in an effectively retrocausal way, as one can see by examining the frequency response at the pickup (figure 9f). The signal proceeds unchanged from the pickup through the cable to the amplifier, even though the properties of the cable affect the signal. The signal "knows" about the cable before it gets there.
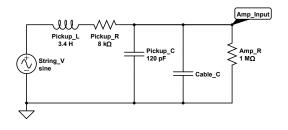
In this simple AC circuit, the signal is affected not only by the source (the preparation) but by the receiver (the measuring device). Viewed this way, it is very much analogous to the cRBM models discussed here. Of course, there is no true backward causation here, because changes made later in the chain take a finite, if miniscule, time to propagate back to the pickup. But this timescale is irrelevant for understanding the behavior of the system, since the signal of interest consists of oscillations whose period (at minimum $5 * 10^{-5}$ seconds for the highest audible frequency) is several orders of magnitude larger than what one might call the relaxation time of the system as a whole (around 10^{-8} seconds). In other words, the time it takes for a change at the amplifier input to propagate back to the pickup is 5000 times less than the time of one cycle of the highest audible frequency in the signal. Thus it is effectively instantaneous, so that changes made downstream at the cable or amplifier influence the signal upstream in the
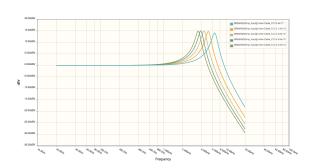
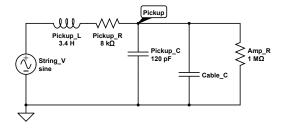(a) Electric guitar pickup driven by vibrating string. No load.

(b) Frequency response of the unloaded pickup

(c) Pickup connected to amplifier with cables of varying capacitance. Sampled at amplifier.

(d) Frequency response at amplifier

(e) Pickup connected to amplifier with cables of varying capacitance. Sampled at pickup.

(f) Frequency response at pickup

Figure 9: Frequency response of electric guitar pickup under varying capacitive loads

pickup itself.

7 Statistical independence, measurement independence, and "fine-tuning"

Models of EPR correlations that allow interdependence of some sort between the states λ and the detector settings α, β have been studied under various rubrics. Models that violate SI are sometimes called *retrocausal* as they postulate a dependence of the states at earlier times on the detector settings at the time of measurement. Models that violate what is sometimes called the *measurement independence* postulate that the conditional dependence goes the other way around, so that the seemingly-freely chosen detector settings are at least partially constrained by the prior states [9, 20, 21, 4, 34, 35, 22]. These are often called *superdeterministic*, and are said to violate the "free will" assumption. Of course it's possible to have the dependence running in both directions, in case the model has a well-defined joint probability over detector settings and states.

Wood and Spekkens [40] showed that all models that postulate some sort of mutual dependence of states and settings must be "fine-tuned" if they are to rule out superluminal signaling. The motivation for this is the idea that if the hidden state is influenced by the choice of setting at A, and if it in turn affects the outcome at B, then it will take a very special choice of model parameters (e.g., the weights and biases in the cRBM model) in order to preclude the possibility of faster-than-light signaling from A to B. They prove that any such model which does *not* permit signaling can be turned into a model which does allow signaling by altering the model parameters. Essentially, this means that the causal/topological structure of the model is by itself insufficient to preclude signaling.

The argument about fine-tuning is intended to suggest the unnatural or even conspiratorial nature of SI-violating models. Almada *et al.* [14] have countered that the various symmetry constraints that we commonly appeal to as constraints on our theories would appear to represent a kind of fine-tuning on the Wood and Spekkens characterization. Though further study is needed, it is indeed likely that it is the symmetries in the weights and biases of the cRBM model that preclude superluminal signaling. These are quite evident already in the simple model shown in figure 6.

8 Nonlocal boxes as the $T \to 0$ limit

The cRBM models exhibit interesting behavior in the $T \to 0$ limit. As noted above, probabilities are generated via a Boltzmann distribution at $T = 1$. Thus we have a quasi-thermal ensemble at fixed temperature. Given that this is a four-dimensional (spacetime) ensemble bounded by two times, rather than a spatial volume, it seems plausible that the physical counterpart of this temperature in our application is another constant, Planck's constant $\hbar$, this one having dimensions of action rather than energy. Indeed, the temperature in our cRBM model is playing a role similar to that played by $\hbar$ in the path integral formalism. One might regard the study of the $T \to 0$ limit as the analogous study of the $\hbar \to 0$ limit in quantum mechanics or any other theory in which $\hbar$ plays a fundamental role.

Whether or not one regards the temperature as analogous to Planck's constant, the $T \to 0$ limit is interesting because it reproduces "PR boxes" [31], hypothetical deterministic devices the behavior of which yields maximal violation of the Bell-inequality ($S = 4$) while not allowing superluminal signaling. They are maximally nonlocal boxes. In figure 10 we see the results of the model for $T = 0.2$, well above $T = 0$ but nevertheless sufficient to give PR boxes with high accuracy. Thus the cRBM model has a well-defined deterministic limit which is maximally nonlocal.
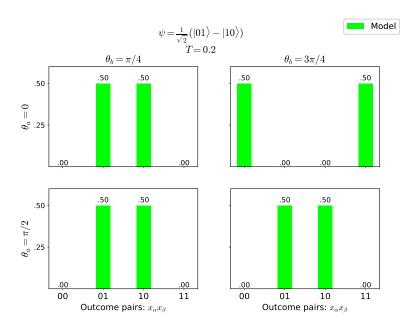
14

Figure 10: cRBM model with three hidden units at $T = 0.2$ approximates a PR box

9 Conclusion

By training conditional Restricted Boltzmann machines, we have constructed a family of stochastic hidden-variable models for particle pairs, both entangled and not, for a variety of measurement settings. The methods used are quite general, and the patterns exhibited in the weights and biases suggest that one might be able to generalize the model to arbitrary detector settings and state preparations, and then to arbitrary numbers of particles. The biggest barrier to this would seem to be the need for an exponential number of hidden units as a parallel to the growth of the Hilbert space dimension in the traditional quantum mechanical representation. Issues of practicality are indeed at the forefront of the attempt to use neural networks to learn particularly salient properties of large-scale quantum systems [10].

For our purposes, however, the proof of principle points in another, more interesting direction. That is, it points to the fact that a hidden-variable theory that depends on future boundary conditions associated with measuring devices may not tell us anything at all about the physical world for systems which do not have any sort of effective measurement in their future. In cosmological settings, this is the norm, since the future lightcone of an arbitrary point in an expanding universe is transparent; there is no interaction or absorption. This in turn suggests the possibility of a new sort of limit on the applicability of quantum mechanics.

Acknowledgements

References

[1] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive Sci.*, 9(1):147–169, 1985.

[2] Y. Aharonov, E. Cohen, and A. C. Elitzur. Can a future choice affect a past measurement's outcome? *Ann. Phys.*, 355:258–268, 2015.

[3] A. Aspect, J. Dalibard, and G. Roger. Experimental tests of Bell's inequalities using time-varying analysers. *Phys. Rev. Lett.*, 49:1804–1807, 1982.

[4] J. Barrett and N. Gisin. How much measurement independence is needed to demonstrate nonlocality? *Phys. Rev. Lett.*, 106(10):100406, 2011.

[5] J. S. Bell. On the Einstein–Podolsky–Rosen paradox. *Physics*, 1:195–200, 1964.

[6] D. Bohm. *Quantum Theory*. Prentice-Hall, Englewood Cliffs, 1951.

[7] D. Bohm. A suggested interpretation of the quantum theory in terms of "hidden" variables. I. *Phys. Rev.*, 85(2):166, 1952.

[8] D. Bohm. A suggested interpretation of the quantum theory in terms of "hidden" variables. II. *Phys. Rev.*, 85(2):180, 1952.

[9] Carl H. Brans. Bell's theorem does not eliminate fully causal hidden variables. *Int. J. Theor. Phys.*, 27(2):219–226, 1988.

[10] G. Carleo and M. Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

[11] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, Oct 1969.

[12] O. Costa De Beauregard. S Matrix, Feynman Zigzag and Einstein Correlation. *Phys. Lett.*, A67:171–174, 1978.

[13] J. G. Cramer. The transactional interpretation of quantum mechanics. *Rev. Mod. Phys.*, 58(3):647, 1986.

[14] D. D. Almada, K. Chng, S. Kintner, B. Morrison, and K.B. Wharton. Are retrocausal accounts of entanglement unnaturally fine-tuned? *Intl. Jnl. Quant. Found.*, 2:1–16, 2015.

[15] P.C.W. Davies. Extension of Wheeler-Feynman quantum theory to the relativistic domain. i. Scattering processes. *J. Phys. A: General Physics*, 4(6):836, 1971.

[16] P.C.W. Davies. Extension of Wheeler-Feynman quantum theory to the relativistic domain. ii. Emission processes. *J. Phys. A: General Physics*, 5(7):1025, 1972.

[17] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of reality be considered complete? *Phys. Rev.*, 47:777–780, 1935.

[18] A. Fine. *The Shaky Game. Einstein and the Quantum Theory*. U. Chicago Press, Chicago, 1986.

[19] S. J. Freedman and J. F. Clauser. Experimental test of local hidden-variable theories. *Phys. Rev. Lett.*, 28:938–941, Apr 1972.

[20] M.J.W. Hall. Local deterministic model of singlet state correlations based on relaxing measurement independence. *Phys.Rev. Lett.*, 105(25):250404, 2010.

[21] M.J.W. Hall. The significance of measurement independence for Bell inequalities and locality. In Torsten Asselmeyer-Maluga, editor, *At the Frontier of Spacetime: Scalar-Tensor Theory, Bell's Inequality, Mach's Principle, Exotic Smoothness*, pages 189–204. Springer, Switzerland, 2016.

[22] J. Handsteiner, A.S. Friedman, D. Rauch, J. Gallicchio, B. Liu, H. Hosp, J. Kofler, D. Bricher, M. Fink, C. Leung, A. Mark, H.T. Nguyen, I. Sanders, Fabian S., R. Ursin, S. Wengerowsky, A.H. Guth, D.I. Kaiser, T. Scheidl, and A. Zeilinger. Cosmic bell test: Measurement settings from milky way stars. *Phys. Rev. Lett.*, 118:060401, Feb 2017.

[23] G. E. Hinton. Boltzmann machine. *Scholarpedia*, 2(5):1668, 2007. revision #91075.

[24] G. E. Hinton. A practical guide to training Restricted Boltzmann Machines. In G. Montavon, G. B. Orr, and K. Müller, editors, *Neural Networks: Tricks of the Trade*, pages 599–619. Springer, Berlin, Heidelberg, second edition, 2012.

[25] G. E. Hinton and T. J. Sejnowski. Optimal perceptual inference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 448–453. IEEE New York, 1983.

[26] J.E. Hogarth. Cosmological considerations of the absorber theory of radiation. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 267(1330):365–383, 1962.

[27] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *P. Natl. Acad. Sci. U.S.A.*, 79(8):2554–2558, 1982.

[28] M. S. Leifer and M. F. Pusey. Is a time symmetric interpretation of quantum theory possible without retrocausality? *Proc. R. Soc. A*, 473(2202), 2017.

[29] E. Nelson. Derivation of the Schrödinger equation from Newtonian mechanics. *Phys. Rev.*, 150(4):1079, 1966.

[30] E. Nelson. *Quantum Fluctuations*. Princeton Univ. Press, Princeton, 1985.

[31] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. *Found. Phys.*, 24(3):379–385, 1994.

[32] H. Price and K. Wharton. Dispelling the quantum spooks: A clue that Einstein missed? In C. Bouton and P. Huneman, editors, *Time of Nature and the Nature of Time*, pages 123–137. Springer, 2017.

[33] P. Smolensky. Information processing in dynamical systems: Foundations of harmony theory. In J. L. McClelland, D. E. Rumelhart, PDP Research Group, et al., editors, *Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1*, pages 194–281. MIT press, Cambridge, MA, 1986.

[34] G. 't Hooft. Can quantum mechanics be reconciled with cellular automata? *Int. J. Theor. Phys.*, 42(2):349–354, 2003.

[35] G. 't Hooft. The cellular automaton interpretation of quantum mechanics. *arXiv:1405.1548*, 2014.

[36] T. Tieleman. Training Restricted Boltzmann Machines using approximations to the likelihood gradient. In *Proceedings of the 25th International Conference on Machine Learning*, pages 1064–1071. ACM, 2008.

[37] S. Weinstein. Nonlocality without nonlocality. *Found. Phys.*, 39(8):921–936, 2009.

[38] S. Weinstein. Learning the Einstein-Podolsky-Rosen correlations on a Restricted Boltzmann Machine. *arXiv:1707.03114*, 2017.

[39] J.A. Wheeler and R.P. Feynman. Interaction with the absorber as the mechanism of radiation. *Rev. Mod. Phys.*, 17:157–181, Apr 1945.

[40] C. J. Wood and R. W. Spekkens. The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning. *New J. Phys.*, 17(3):033002, 2015.