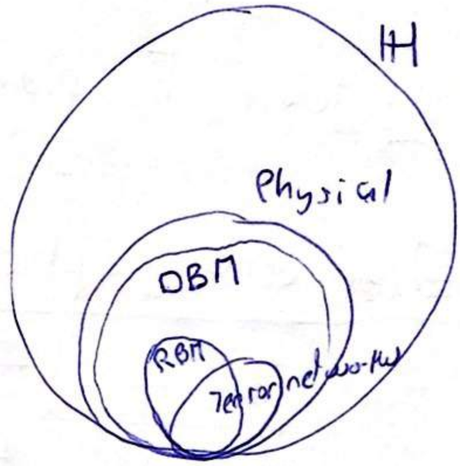
Lecture III: The physical corner of Hilbert's space (II)

"Neural-network quantum states"



- G.E. Hinton & R.R. Salakhutdinov, Science 313, 504 (2006)
- G. Carleo & M. Troyer, Science (17)
- D.-L. Deng, X. Li & S. Das Sarma PRX (17)
- X. Gao & L.-M. Duan, Nat. Comms (17)

Boltzmann Machines (classical machine learning)

RBMs:

- ML battles the same "curse of dimensionality" problems as many-body Q systems
- NNs unprecedented success in practice

$$\boxed{P(\vec{v}) = \frac{1}{Z} \sum_{\vec{h}} e^{-E(\vec{h},\vec{v})}}$$

with $\boxed{Z = \sum_{\vec{h},\vec{v}} e^{-E(\vec{h},\vec{v})}}$

Generative modelling with probabilistic graphical models

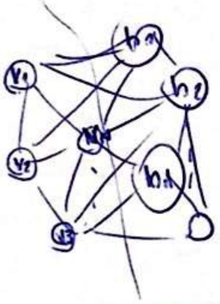
↓

Main applications: – unsupervised (and als supervised) learning

Ex: learn $\bar{P}$ from samples $(\bar{a}_1 \dots \bar{a}_{N_s})$ from $\bar{P}$

$(h_i, v_j = +1, -1)$

⇒ Energetic graphical models:



$$\boxed{P(\bar{v}) = \frac{\sum_{\bar{h}} e^{-E(\bar{v},\bar{h})}}{Z}}$$

Prob of visible string $\bar{v} = (v_1 \dots v_N)$

with $$\boxed{Z = \sum_{\bar{v},\bar{h}} e^{-E(\bar{v},\bar{h})}}$$ (the partition function)

⇒ For $$\boxed{E = -\Big[\underbrace{\bar{a}\cdot\bar{v} + \bar{b}\cdot\bar{h}}_{\text{local biases}} + (\bar{v},\bar{h})\,\underbrace{\bar{\bar{W}}}_{\substack{\text{weights}\\ \text{(interactions)}}}\begin{pmatrix}\bar{v}\\ h\end{pmatrix}\Big]}$$

$\bar{a} \in \mathbb{R}^N$
$\bar{b} \in \mathbb{R}^M$
$W \in \mathbb{R}^{N+M} \times \mathbb{R}^{N+M}$

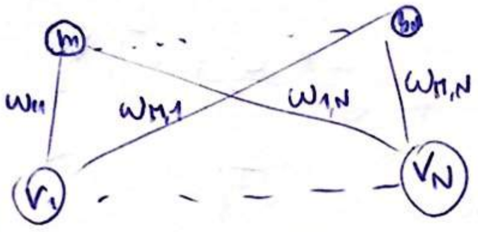
"classical Ising model"

⇒ The architecture is called a Boltzmann M

But fully-connected BMs are intractable (to sample and to train) as N+M grows ⇒

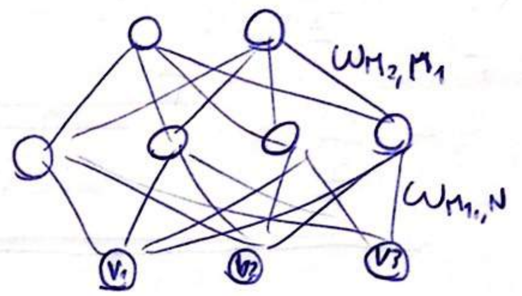
$\Rightarrow$ RBMs : (no intra-layer interactions)

RBM (shallow)



$$E = -\left[\bar{a}.\bar{v} + \bar{b}.\bar{h} + \bar{v}.\bar{W}\bar{h} \right]$$

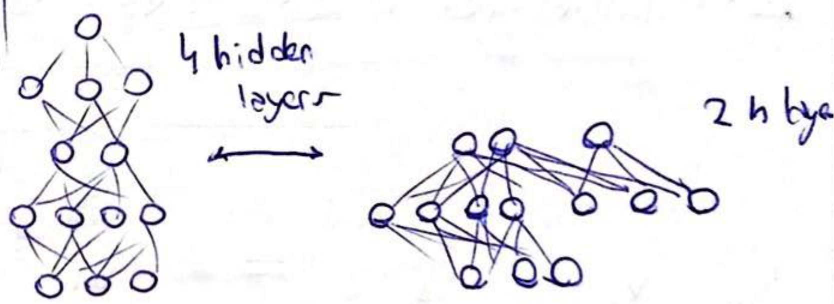
DRBM (deep)



$$E = -\left[\bar{a}.\bar{v} + \bar{b}_1\bar{h}_1 + \bar{v}\,\bar{W}_1\bar{h}_1 + \bar{h}_1\bar{W}_2\bar{h}_2 \right.$$

$\Downarrow$

2 hidden layers always suffice



Properties:

i) RBMs can represent arbitrary $P(\bar{v})$ if $M = O(2^N)$ (completeness - expressive power)

ii) RBMs are hard to sample:

[Given $\bar{\Omega} = (\bar{a}, \bar{b}, \bar{w})$ with $\Omega_{max} = O(N^{1+\epsilon})$

$\Rightarrow$ sample from $P_{\bar{\Omega}}$ with the classical circuit and input random bits $\Rightarrow$ collapse: RP = NP]

$\Rightarrow$ hard to train in the worst case scenario

However, in practice: extremely successfull !!

iii) Hidden-units Factorization : Because of no-intralayer interaction

$$\Rightarrow \boxed{P(\bar{v}) = \frac{1}{Z} \sum_{\bar{h}} e^{+(\bar{a}\cdot\bar{v} + \bar{b}\cdot\bar{h} + \bar{v}\,\bar{\omega}\,\bar{h})}} =$$

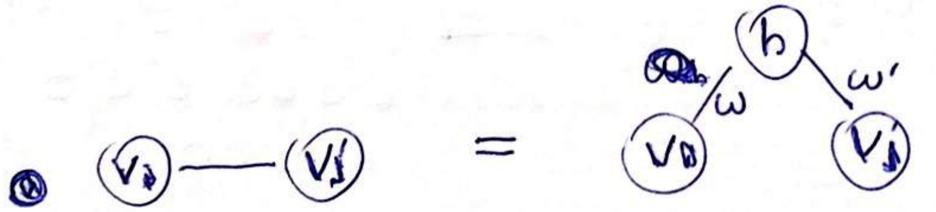
$$= \frac{1}{Z} e^{\bar{a}\cdot\bar{v}} \prod_{j\in[M]} \left[e^{b_j + \bar{\omega}_j \bar{v}} + e^{-(b_j + \bar{\omega}_j \bar{v})} \right]$$

(j^{th} column of $\bar{\omega}$)

(Exercise)

$$\boxed{= \frac{1}{Z} \prod_{j\in[M]} \left[2\, e^{a_i v_i} \cosh\left(b_j + \bar{\omega}_j \bar{v}\right) \right]}$$

(IV) Intralayer interaction simulatable by complex weights



with $\omega, \omega' \in \mathbb{C}$

(we'll see examples of these later)

However : $\omega \in \mathbb{R} \Rightarrow$ simpler to train

---

Training an RBM for unsupervised learning

Task: - given an N_s-sample $\{ v_1, \ldots \bar{v}_{N_s} \}$ (data set) generated by an unknown dis $\bar{P}_{Data}$

$\Rightarrow$ Find $\Omega = \{ \bar{a}, \bar{b}, \bar{\omega} \}$ (reals) s. that

$\bar{P}_{\Omega} \simeq \bar{P}_{Data}$ (as close as possible)

- Distinguishability of $\bar{P}$ from $\bar{Q}$:

$$\boxed{S(\bar{P} \| \bar{Q}) = -\sum_{\bar{v}} P_{\bar{v}} \ln\left(\frac{Q(\bar{v})}{P(\bar{v})}\right) = \sum_{\bar{v}} P_{\bar{v}} \ln\left(\frac{P(\bar{v})}{Q(\bar{v})}\right)}$$

"Kullback-Leibler divergence of $\bar{P}_\Omega$ from $\bar{P}_{Dat}$

or Relative (Shannon) Ent of $\bar{P}_\Omega$ given $\bar{P}_D$.
(info theory)

or Information gain if $\bar{P}_\Omega$ is used instead of $\bar{P}_D$"
(machine learning) → (amount of info lost Q is used to approximate P)

typically
P = real (posterior) data
Q = model (prior)
(we'll use it inverted)

- Physical interpretation (hypothesis testing)

→ Minimal-achievable probability of mistaking samples from $\bar{Q}$ for samples from $\bar{P}$ $= \boxed{O\left(e^{-S(\bar{P}\|\bar{Q})\, N_s}\right)}$

(Promise: either P or Q)

Now, consider:

$$S(\bar{P}_{Dat} \| \bar{P}_\Omega) = \sum_{\bar{v}} P_{Dat}(\bar{v}) \log\left[\frac{P_{Dat}(\bar{v})}{P_\Omega(\bar{v})}\right] = \mathbb{E}_{\bar{P}_{Dat}}\left(\log\left[\frac{P_{Dat}(\bar{v})}{P_\Omega(\bar{v})}\right]\right)$$

$$\simeq \frac{1}{N_s} \sum_{k=1}^{N_s} \log\left[\frac{P_{Dat}(\bar{v}_k)}{P_\Omega(\bar{v}_k)}\right]$$

$\bar{v}_k \sim \bar{P}_{Dat}$

$\Rightarrow$ $\boxed{\text{Training} = \text{minimize this over } \Omega}$

$$\Rightarrow \boxed{\min_{\Omega} \left\{ S\left(P_{Dat} \| P_{\Omega}\right) \right\}} \simeq \min_{\Omega} \left\{ \frac{1}{N_S} \sum_{\substack{i=1 \\ \bar{v}_i \sim \bar{P}_{Dat}}}^{N_S} \log\left[\frac{P_{Dat}(\bar{v}_i)}{P_{\Omega}(\bar{v}_i)}\right] \right\}$$

$$\log\left[P_{Dat}(\bar{v}_i)\right] - \log\left[P_{\Omega}(\bar{v}_i)\right]$$

$\log\left[P_{Dat}(\bar{v}_i)\right]$ $\downarrow$ independent of Ω

$$= \min_{\Omega} \left\{ \frac{-1}{N_S} \sum_{\substack{i=1 \\ \bar{v}_i \sim \bar{P}_{Dat}}}^{N_S} \log\left[P_{\Omega}(\bar{v}_i)\right] \right\}$$

prob of observing the given sample

$$\log\left[\prod_{i=1}^{N} P_{\Omega}(\bar{v}_i)\right]$$

$$\boxed{= -\max_{\Omega} \left\{ \frac{1}{N_S} \log\left[\prod_{\substack{i=1 \\ i \sim \bar{P}_{Dat}}}^{N} P_{\Omega}(\bar{v}_i)\right] \right\}}$$

Likelihood of Ω given the sample $\bar{v}_1, \ldots \bar{v}_{N_S}$

log-likelihood

factorizes further by property (xiii)

$\Rightarrow$ "Minimizing the KL div $\equiv$ maximizing the log-likelihood of the observed sample"

($\Rightarrow$ max likelihood estimation of Ω)

- Numerics: optimization solved by gradient descent

(Block-Gibbs MCMC and contrastive divergence)

Monte Carlo
Markov chain

Neural-Network quantum states: ~~2017~~ 2016-2017) ④

<u>Ansatz</u>: parametrize $C_{\bar{i}}$ in $|\Psi\rangle = \sum_{\bar{i}} c_{\bar{i}} |\bar{i}\rangle$ with a Boltzman machine.

<u>RBM states</u>: $\boxed{c_{\bar{v}} = \sum_{\bar{h}} e^{-E(\bar{v},\bar{h})}}$ → ($\frac{1}{Z}$ absorved as an additive constant in $E(\bar{v},\bar{h})$)

complex coefficients, $\bar{a} \in \mathbb{C}^N, \bar{b} \in \mathbb{C}^M \quad \omega \in \mathbb{C}^{M+N \times (M+N)}$

$\Rightarrow$ ~~$O((N+M)(N+M))$~~ $\boxed{O\left[(N+M)^2\right]}$ complex parameters in the description

---

<u>Properties</u> (in connection with tensor networks).

i) <u>Area law</u>: Any short-range RBM state obeys an entanglement area law in any lattice dimension and for any bipartition

constant range

$$\boxed{\omega_{i,j} = 0 \text{ if } |\bar{i}-\bar{j}| > R \quad \Rightarrow \quad S_A \leq 2\, S(A|B)\, R \log 2}$$

i: lattice position of visible unit; j: position of hidden unit; S_A: ent entropy of ρ_A; $S(A|B)$: area of border surface of bipartition $A:B$

(Exercise)

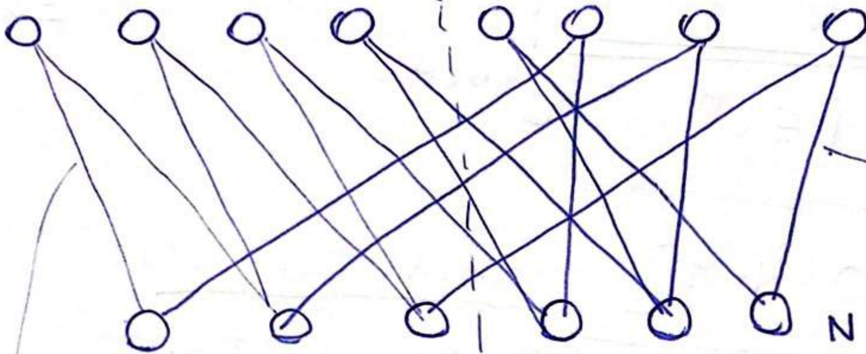
ii) ~~Area law~~

Volume law: for long-range RBM states

Example in 1D

A: $b_k = -i\frac{\pi}{4} \ \forall k \in [N-1]$ B: $b_k = i\frac{\pi}{2} \ \forall k \in [N, M]$

$M = \left\lfloor \frac{3N}{2} \right\rfloor - 1$



$$\omega_{k,k'} = \begin{cases} i\frac{\pi}{4} & k \in [N, M], \ k' = k+1-N \text{ or } k+1-\left\lceil \frac{N}{2} \right\rceil \\ 0 & \end{cases}$$

$a_k = 0 \ \forall k \in [N]$

$\boxed{O(N) \text{ parameters!!}}$

$$\omega_{k,k'} = \begin{cases} i\frac{\pi}{4} & k \in [N-1], \ k' = k \text{ or } k+1 \\ 0 & \text{otherwise} \end{cases}$$

1D volume law!

$$\Rightarrow \quad \boxed{\rho_A = \frac{\mathbb{1}}{2^L} \Rightarrow S_A = L \log(2)}$$

$\forall$ A consisting of L contiguous qubits

Exercise show it!

"If we wanted to reproduce this with an MPS/tensor-network $\Rightarrow$ bond dimension $R = O(2^M) \Rightarrow$ intractable!

iii) Expressive power of quantum RBM Ansatz:

a) All graph states admit an efficient RBM exact representation:

Change convention: $v_i, h_j = -1, +1 \Rightarrow v_i, h_j = 0, 1$

$$\Rightarrow \boxed{|G_E\rangle = \prod_{\underbrace{e \in E}_{\text{edges}}} CZ_e \prod_{\underbrace{i \in V}_{\text{vertices}}} |+\rangle_i}$$

Now

$$|+\rangle = \frac{|0\rangle + e^{a}|1\rangle}{\sqrt{1+e^{2a}}} \qquad a=0, \ b=0, \ w=0$$

and

$$CZ_{ii'} \sum_{\bar{v}} \psi(\bar{v})|\bar{v}\rangle = \sum_{h,\bar{v}} e^{-(E_{cz}(v_i,h) + E_{cz}(v_{i'},h))} \psi(\bar{v})|\bar{v}\rangle$$

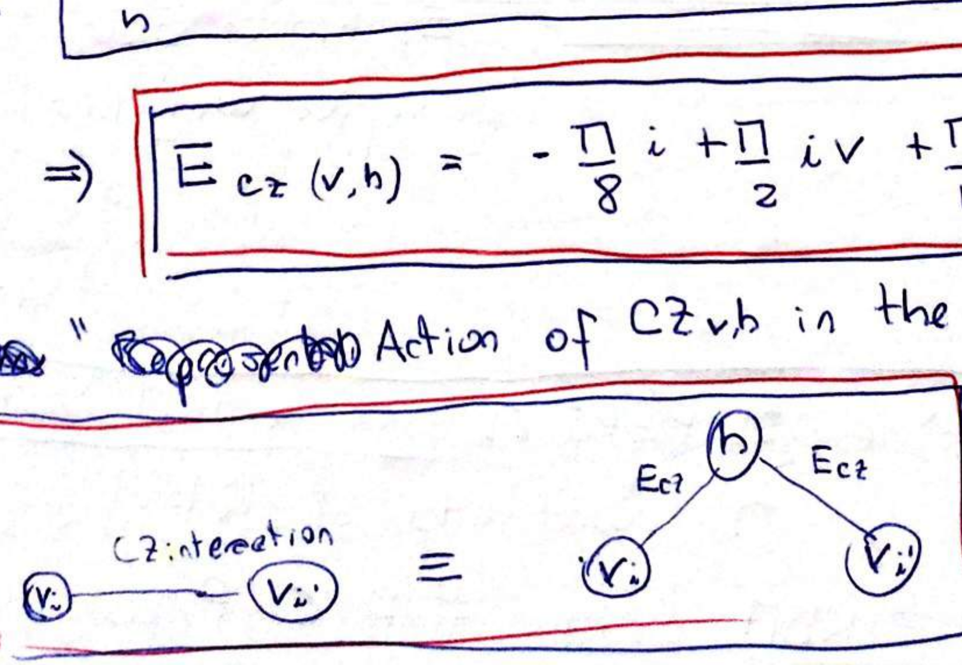
if

$$\boxed{\sum_h e^{-E_{cz}(v_i,h) + E_{cz}(v_{i'},h)} = (-1)^{v_i, v_{i'}} \quad \forall i, i'}$$

Exercise show!

$$\Rightarrow \boxed{E_{cz}(v,h) = -\frac{\pi}{8} i + \frac{\pi}{2} i v + \frac{\pi}{4} i h + i\pi v h}$$

"Action of CZ_{vh} in the RBM representation"

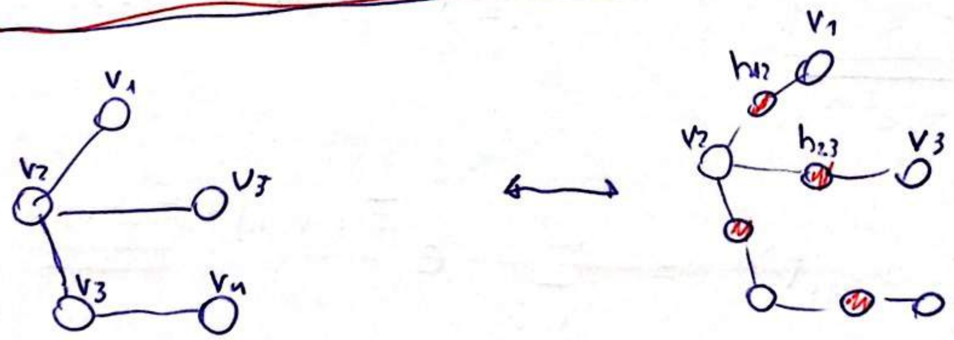


• In fact, any intralayer (Ising) interaction can be replaced ~~unitation~~ by a hidden neuron with complex weights

fully connected BM

Exercise: show!

Then:



Graph state of $V = 5$ and $|E| = 4$

~~RBM~~ short-ranged RBM with V visible units and |E| hidden ones with complex weights

b) Limitations

(poly sized circuit even if circuit not efficient to construct)

Theorem: $\exists\ |\psi\rangle \in$ 2D PEPS $\cap$ constant depth Q circuits states $\cap$ ground states of local gapped H

s.t. $|\psi\rangle \notin$ RBM unless $\#P \subset P/poly$

sharp P ~~poly sized circuit~~

$\Rightarrow$ Since $|\Psi\rangle \in$ constant-depth Q circuit $\Rightarrow$

same topological phase as 2D cluster state

$\Rightarrow$ RBM representability not closed under unitaries preserving the quantum phase

---

Quantum DBM states

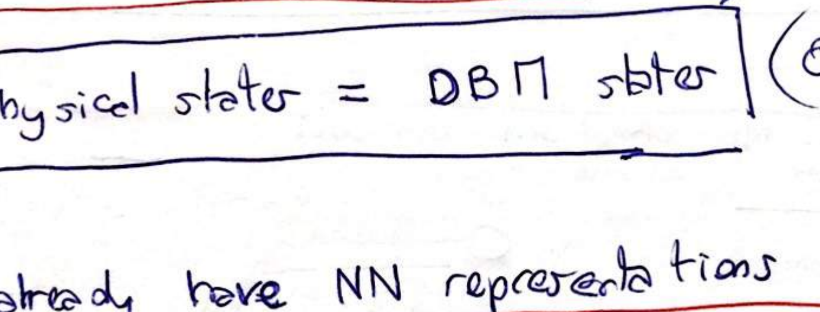
(exponential gap between RBM and DBM)

$\boxed{\text{Physical states = DBM states}}$ (QC in the NN representation)

- we already have NN representations of $|+\rangle$ and CZ

$\Rightarrow$ Add Hadamard H and Phase $Z(\theta)$ and we get universality!

- Hadamard gate:

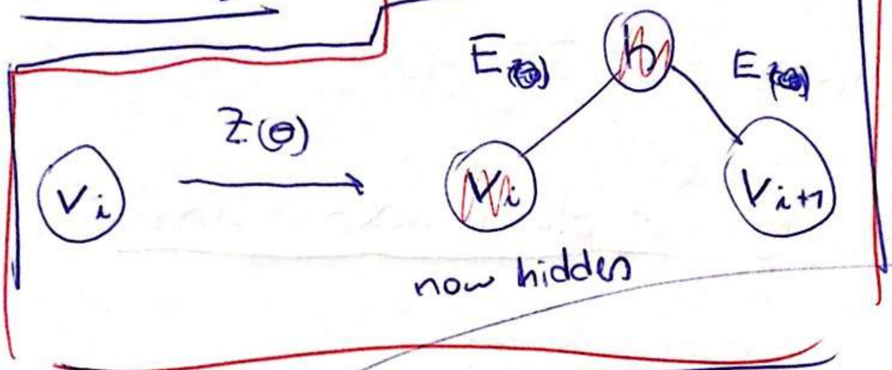


with $E_{H(v,h)} = E_{CZ(v,h)} + \frac{\ln(2)}{2}$

"Action of H"

(so that $\sum_h e^{-[E_H(v_i,h) + E_H(v_{i+1},h)]} = \frac{(-1)^{v_i v_{i+1}}}{\sqrt{2}}$)

- Phase gate:

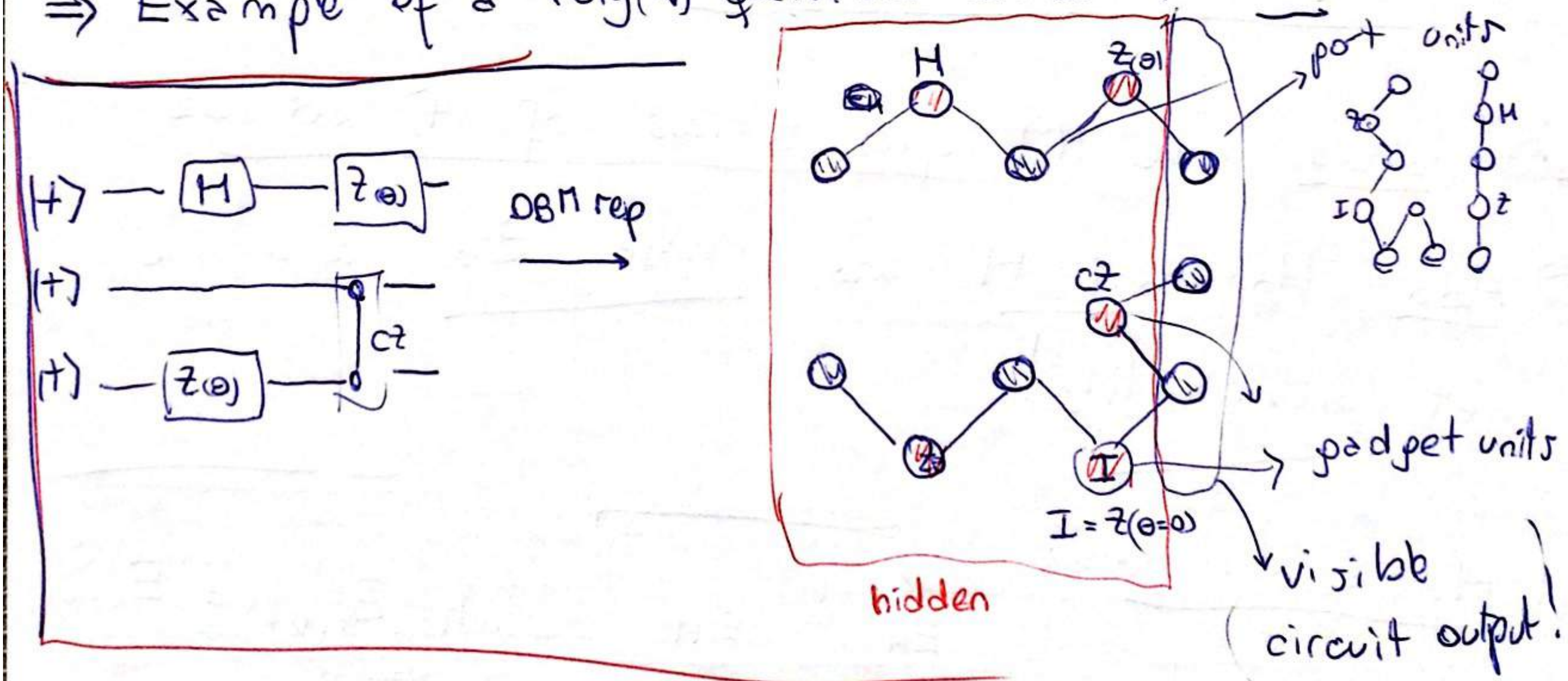


with
$$E_{\theta(v,h)} = \frac{\ln(2)}{\sqrt{2}} - \frac{\theta i v}{2} - i\pi v h$$

"Action of $Z(\theta)$ in the ~~RBM~~ NN rep"

so that
$$\sum_h e^{-\left(E_\theta(v,h) + E_\theta(v',h)\right)} = e^{i\theta v}\,\delta_{v,v'}$$

⇒ Example of a Poly(N) quantum circuit:



"All Poly(N) Q circuits states are short-ranged sparse DBMs with Poly(N) neurons!!"

More precisely:

"Any depth-T $\mathbb{Q}$ circuit has an exact sparse DBM representation with $O(NT)$ neurons"

↓

Corollaries:

1) "Any tensor network state with bond dimension R maximum coordination number c, and L local tensors can be represented ~~by~~ exactly by a sparse DBM with $\boxed{O(LR^{2c})}$ neurons"

2) The ground state of any Hamiltonian H can be ε-approximated by a sparse DBM with $\boxed{O\left(\frac{1}{\Delta}\left[N+\log\left(1/\varepsilon\right)\right]|X|^2\right)}$ neurons,

gap — number of particle — number of interaction terms

("Taylor truncation of $e^{-\beta H}$ and tensor-network representation of it. (imaginary time DMRG)")

⇒ All physical statos admit a sparce DBM efficient representation !

- Representation efficient ✓
- Manipulation Not !