NEURAL-NETWORK QUANTUM STATES

GIUSEPPE CARLEO

[EPFL - SWITZERLAND]

USEFUL REFERENCES

GENERAL ML	"DEEP LEARNING" BY GOODFELLOW ET AL.
APPLICATIONS TO PHYSICS	CARLEO ET AL. RMP 91, 045002 (2019)
VARIATIONAL MONTE CARLO	"QMC FOR CORRELATED SYSTEMS" BY BECCA, SORELLA
SOFTWARE	MORE IN THE TUTORIALS (JAX, FLAX, NETKET...)

MAIN GOALS

- KNOW ABOUT VARIATIONAL METHODS IN MANY-BODY QUANTUM PHYSICS

- QUANTUM GROUND STATE OF INTERACTING HAMILTONIANS, DYNAMICS | NO EXTERNAL DATA

- LEARNING FROM EXPERIMENTS | EXTERNAL DATA

THE MAIN PROBLEM

CASE OF A SINGLE SPIN 1/2

$$|\psi\rangle_1 = \underline{c_\uparrow}|\uparrow\rangle + \underline{c_\downarrow}|\downarrow\rangle$$

$$|c_\uparrow|^2 + |c_\downarrow|^2 = 1$$

CASE OF MANY SPINS

$$|\psi\rangle_N = \underline{c_{\uparrow\uparrow\ldots\uparrow}}|\uparrow\uparrow\ldots\uparrow\rangle + \underline{c_{\uparrow\uparrow\ldots\downarrow}}|\uparrow\uparrow\ldots\downarrow\rangle + \ldots \ldots \underline{c_{\downarrow\downarrow\downarrow\ldots\downarrow}}|\downarrow\downarrow\ldots\downarrow\rangle =$$

$$= \sum_k c_k |k\rangle$$

$$|k\rangle = |\sigma_1^z \sigma_2^z \ldots \sigma_N^z\rangle = |\sigma_1^z\rangle \otimes |\sigma_2^z\rangle \ldots \otimes |\sigma_N^z\rangle$$

$$k \in [1, 2^N]$$

THE QUANTUM MANY-BODY PROBLEM

$\hat{H}$, IS SOME HAMILTONIAN

$$\hat{H}|\Psi_0\rangle = E_0|\Psi_0\rangle \qquad E_0 \leq E_1 \dots \leq E_M$$

ON A COMPUTER

$$H_{ij} = \langle i|\hat{H}|j\rangle \rightsquigarrow 2^N \times 2^N \text{ MATRIX}$$

DIAGONALIZE IT $\rightsquigarrow$ COST IS

$|\Psi_K\rangle$, E_K $\qquad O(2^N)$ FOR

"LOCAL HAMILTONIANS"

LOCAL HAMILTONIANS

(QUANTUM INFORMATION COMMUNITY)

$\langle k | H | k' \rangle \neq 0$ | # k' is only poly(N)

↳ FIX k

$$\hat{H} = \begin{pmatrix} & & & & & \\ \times & — & \times & — & \times & — \\ & & & & & \end{pmatrix}$$

"SPARSENESS"

LANCZOS METHOD

EXAMPLE

$$\hat{H} = \sum_{i<j}^{N} J_{ij}\, \hat{\sigma}_i^z \hat{\sigma}_j^z + \Gamma \sum_i^{N} \hat{\sigma}_i^x$$

$\langle k | \hat{\sigma}_i^z \hat{\sigma}_j^z | k' \rangle \neq 0$ ONLY IF

$|k'\rangle = |k\rangle$

$\langle k | \hat{\sigma}_i^x | k' \rangle \neq 0$ ONLY IF

$|k'\rangle = | \sigma_1^z \ldots -\sigma_i^z \ldots \sigma_N^z \rangle$

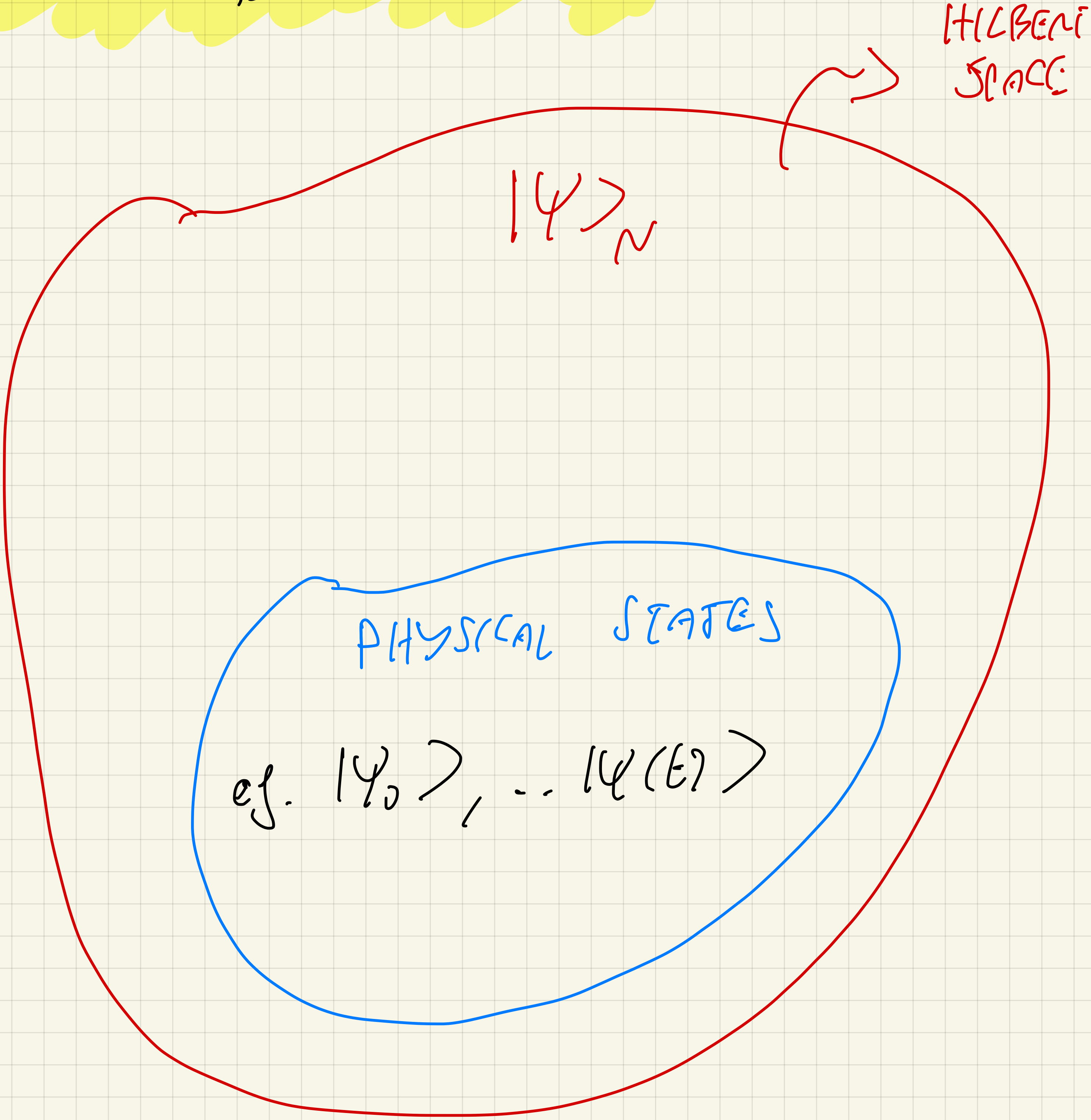
FOR FIXED k $\rightarrow$ # NONZERO MATRIX ELEMENTS $\boxed{(N+1)}$

BEYOND EXACT DIAGONALIZATION

ONLY UP TO ~ 40 SPINS

$2^N \longrightarrow 2^{120} \longrightarrow$ # ATOMS ON EARTH

VARIATIONAL METHODS



ACCESSING CORNERS

$$|\Psi(w)\rangle = \sum_{k}^{2^N} C_k(w)\, |k\rangle$$

PARAMETERS (VARIATIONAL PARS) $\sim$ POLY(N)

FIND "OPTIMAL" PARAMETERS BY OPTIMIZATION

$$E(w) = \frac{\langle\Psi(w)|\hat{H}|\Psi(w)\rangle}{\langle\Psi(w)|\Psi(w)\rangle} \geq E_0$$

EXACT GS ENERGY

$$E(w) = \frac{\sum_k |b_k|^2 E_k}{\sum_k |b_k|^2} \geq E_0$$

EXACT EIGEN-STATES

$$b_k = \langle\Psi_k|\Psi(w)\rangle$$

TWO MAIN CLASSES OF VARIATIONAL STATES

1. COMPUTE $E(w)$ "EXACTLY"
(APART FROM ROUNDING ERRORS)

2. COMPUTE $E(w)$ "APPROXIMATELY"
(IN A CONTROLLED WAY)

EXAMPLES OF FIRST KIND

(a) MEAN-FIELD STATES

$$|\Psi(w)\rangle = |\phi_1\rangle \otimes |\phi_2\rangle \dots |\phi_N\rangle$$

$$\langle \phi_i | \phi_j \rangle = \delta_{ij}$$

$$\langle \sigma_i^z | \phi_i \rangle = w^{(i)} \begin{cases} w_\uparrow^{(i)} \\ w_\downarrow^{(i)} \end{cases}$$

$$|w_\uparrow^{(i)}|^2 + |w_\downarrow^{(i)}|^2 = 1 \,, \quad \sim 2N \text{ PARAMETERS}$$

e.g. $\langle \Psi(w) | \hat{\sigma}_i^x | \Psi(w) \rangle = \langle \phi_i | \sigma_i^x | \phi_i \rangle$

(b) $|\psi(w)\rangle = \sum_k C_k(w)|k\rangle$

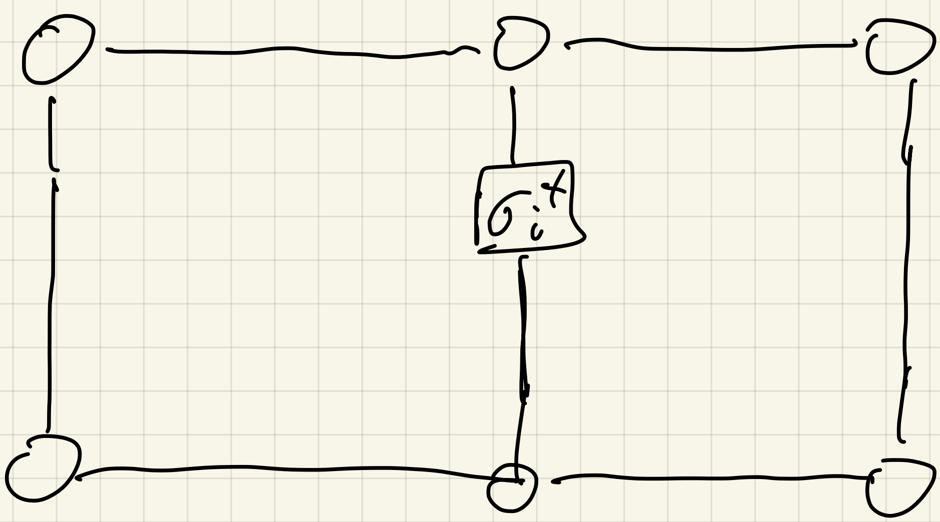
MATRIX PRODUCT STATE

$$C_k(w) = M_1(\sigma_1^z)\cdot \underline{M_2}(\sigma_2^z) \dots M_N(\sigma_N^z)$$

MATRIX OF SIZE $\chi \times \chi'$

BOND DIMENSION $\leftarrow$ $\chi \times \chi'$

$\langle\psi|\sigma_i^x|\psi\rangle$

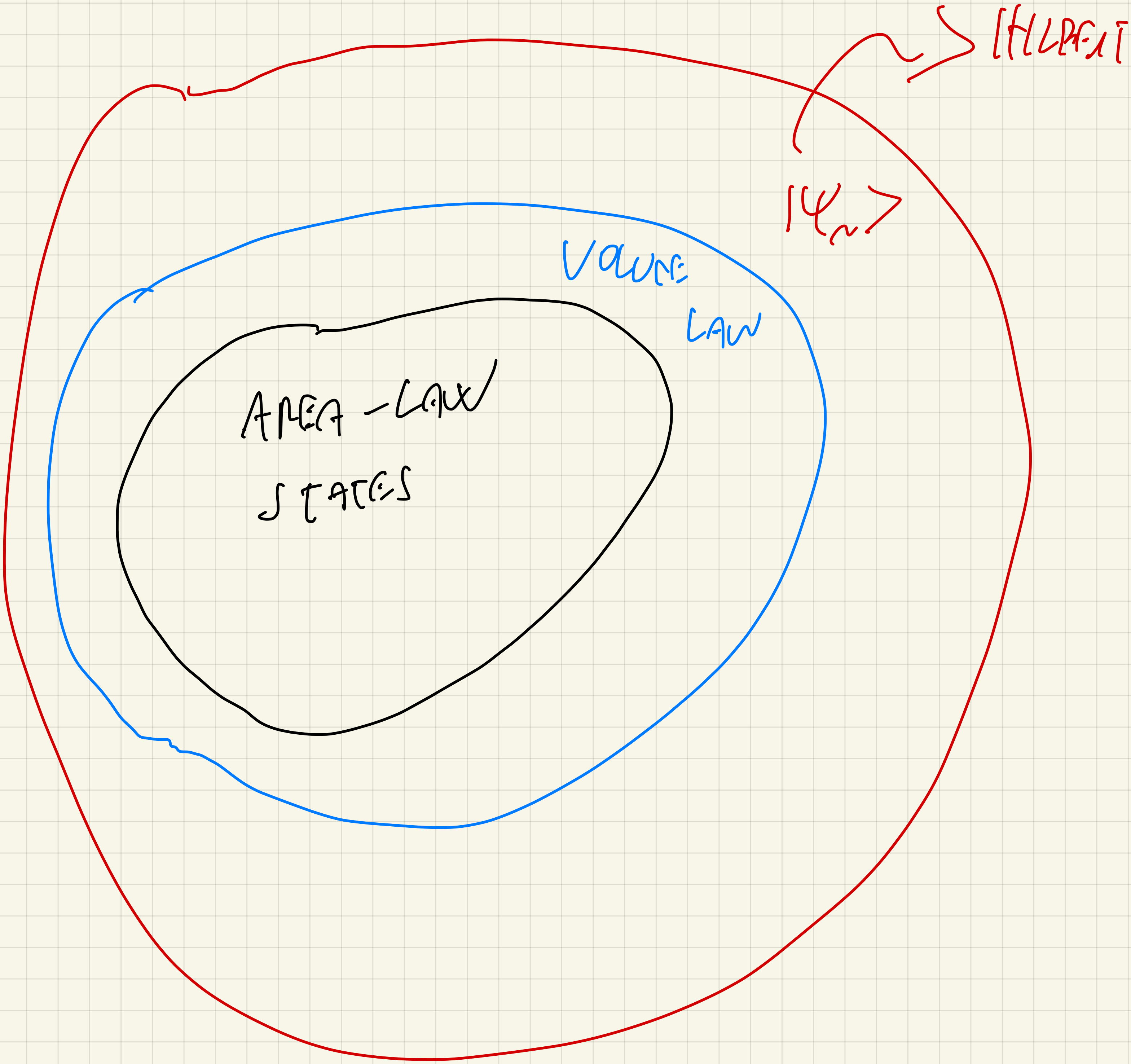


POLY PARAMETERS

$O(N \times \chi^2)$

COMPUTING IT $\sim \chi^3$

LIMITATIONS STEMMING FROM LOCALITY



SECOND FAMILY OF STATES

FORMALIZED IN VAN DEN NEST

ARXIV: 0911.1624 (2009)

USED SINCE McMILLAN ~1960

[VARIATIONAL MONTE CARLO]

DEFINITION

"COMPUTATIONALLY TRACTABLE STATES"

(a) COMPUTE $\langle K | \psi \rangle$ EFFICIENTLY

(b) SAMPLE FROM $P(K) = \frac{|\langle K|\psi\rangle|^2}{\langle\psi|\psi\rangle}$ EFFICIENTLY

THEOREM

(VAN DEN NEST)

$\frac{\langle\psi|\hat{O}|\psi\rangle}{\langle\psi|\psi\rangle}$ ~ CAN BE COMPUTED IN POLY(N) WITH FIXED ACCURACY

ESTIMATING PROPERTIES

$$\langle \hat{O} \rangle = \frac{\langle \psi | \hat{O} | \psi \rangle}{\langle \psi | \psi \rangle} =$$

$$= \frac{\sum_{k k'} \langle \psi | k \rangle \langle k | \hat{O} | k' \rangle \langle k' | \psi \rangle}{\sum_k |\langle \psi | k \rangle|^2} =$$

$$\rightsquigarrow O_{loc}(k)$$

$$= \frac{\sum_k |\langle \psi | k \rangle|^2 \sum_{k'} \langle k | \hat{O} | k' \rangle \frac{\langle k' | \psi \rangle}{\langle k | \psi \rangle}}{\sum_k |\langle \psi | k \rangle|^2} =$$

$$= \sum_k \underline{P}(k) \, O_{loc}(k) \, ,$$

STOCHASTIC ESTIMATE

1. GENERATE

$$k^{(1)}, k^{(2)}, \ldots k^{(M)} \sim P(k) = \frac{|\langle k|\psi\rangle|^2}{\langle\psi|\psi\rangle}$$

2. ESTIMATE

$$\langle \hat{O} \rangle_M \sim \frac{1}{M} \sum_{i=1}^{M} O_{loc}(k^{(i)})$$

$\boxed{\text{ACCURACY}}$

$$\text{ERROR}\left(\langle \hat{O} \rangle_M\right) \sim \sqrt{\frac{\text{VAR}(O_{loc})}{M}}$$

$$\text{ERROR}\left(\langle \hat{O} \rangle_M\right) \xrightarrow{M \to \infty} 0$$

SHOW THAT $\text{VAR}(O_{loc})$

MARKOV CHAIN MONTE CARLO

$$k^{(1)} \rightarrow k^{(2)} \cdots \rightarrow k^{(T)}$$

$$T\left(k^{(i)} \longrightarrow k^{(i+1)}\right)$$

GLOBAL TRANSITION KERNEL

$$P(k)\, T(k \rightarrow k') = P(k')\, T(k' \rightarrow k)$$

DETAILED BALANCE CONDITION

$$k^{(i)} \sim P(k)$$

METROPOLIS-HASTINGS ALGORITHM

$$\underbrace{\tilde{T}(k \to k')}_{\text{GLOBAL KERNEL}} = \underbrace{T(k \to k')}_{\text{LOCAL TRANSITION KERNEL}} \underbrace{A(k \to k')}_{\text{ACCEPTANCE PROBABILITY}}$$

$$A(k \to k') = \min\left(1, \frac{P(k')}{P(k)} \frac{T(k' \to k)}{T(k \to k')}\right)$$

$$\frac{P(k')}{P(k)} = \left|\frac{\langle k' | \psi \rangle}{\langle k | \psi \rangle}\right|^2 ,$$

EFFICIENT TO ESTIMATE THIS RATIO

ALGORITHM

0. $k^{(0)} \sim$ UNIFORM $\in [1, 2^N]$

1. $T(k^{(i)} \to k')$, GENERATE k'

2. $R = \frac{P(k')}{P(k)} \frac{T(k' \to k)}{T(k \to k')}$

3. DRAW $\xi \in [0, 1)$ UNIFORM DISTRIBUTED

4. if $\xi < R$ $\quad k^{(i+1)} = k'$

$\xi > R$ $\quad k^{(i+1)} = k^{(i)}$

$k^{(i)} \to k^{(i+1)} \ldots \quad k^{(M)} \sim P(k)$

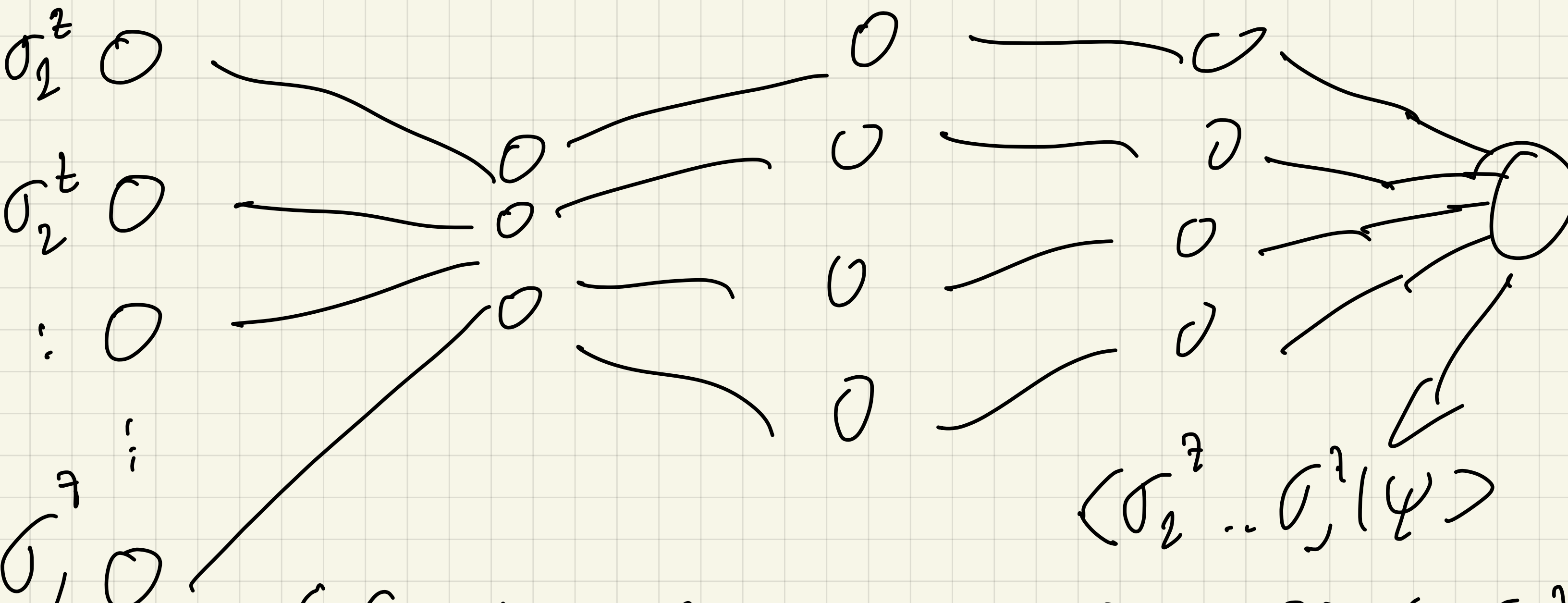
REPRESENTING HIGH-DIMENSIONAL WAVE FUNCTIONS

ARTIFICIAL NEURAL NETWORK

$$\langle k|\psi\rangle = \psi(\sigma_1^z \dots \sigma_N^z ; W)$$

$$\psi(\sigma_1^z, \sigma_2^z \dots \sigma_N^z ; W) = C_k(W) =$$

$$= g^{(D)} \circ W^{(D)} \dots g^{(2)} \circ W^{(2)} g^{(1)} \circ W^{(1)} \sigma$$



[CARLEO AND TROYER, SCIENCE 355, 602 (2017)]

GENERAL REPRESENTATION THEOREMS

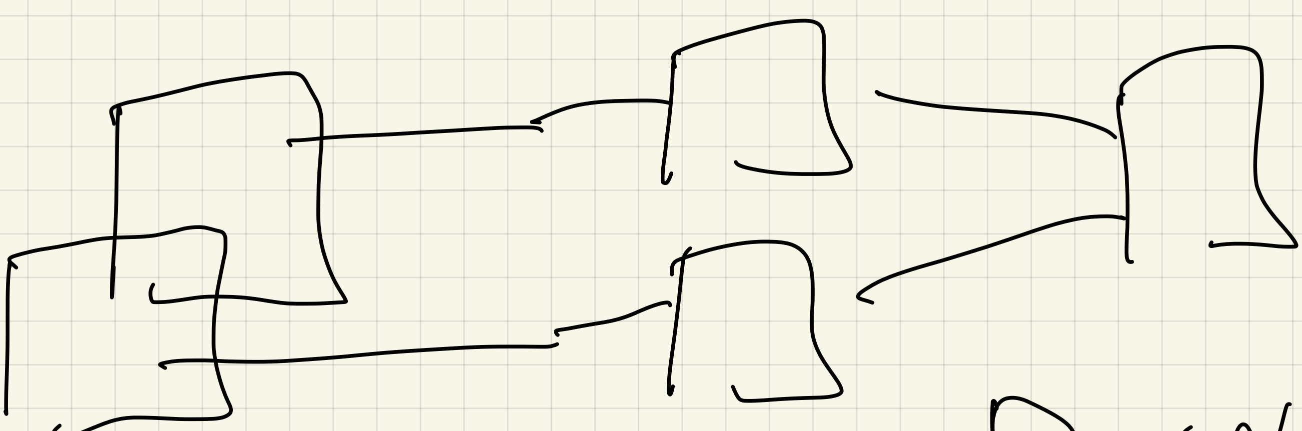
$$\langle k|\psi\rangle = \sum_{q=0}^{2N} \Phi_q\left(\sum_{p=1}^{N} \phi_{qp}(k_p)\right)$$

$$\vec{k} = (k_1 \dots k_N)$$

KOLMOGOROV AND ARNOLD (1956)	CYBENKO (1989)

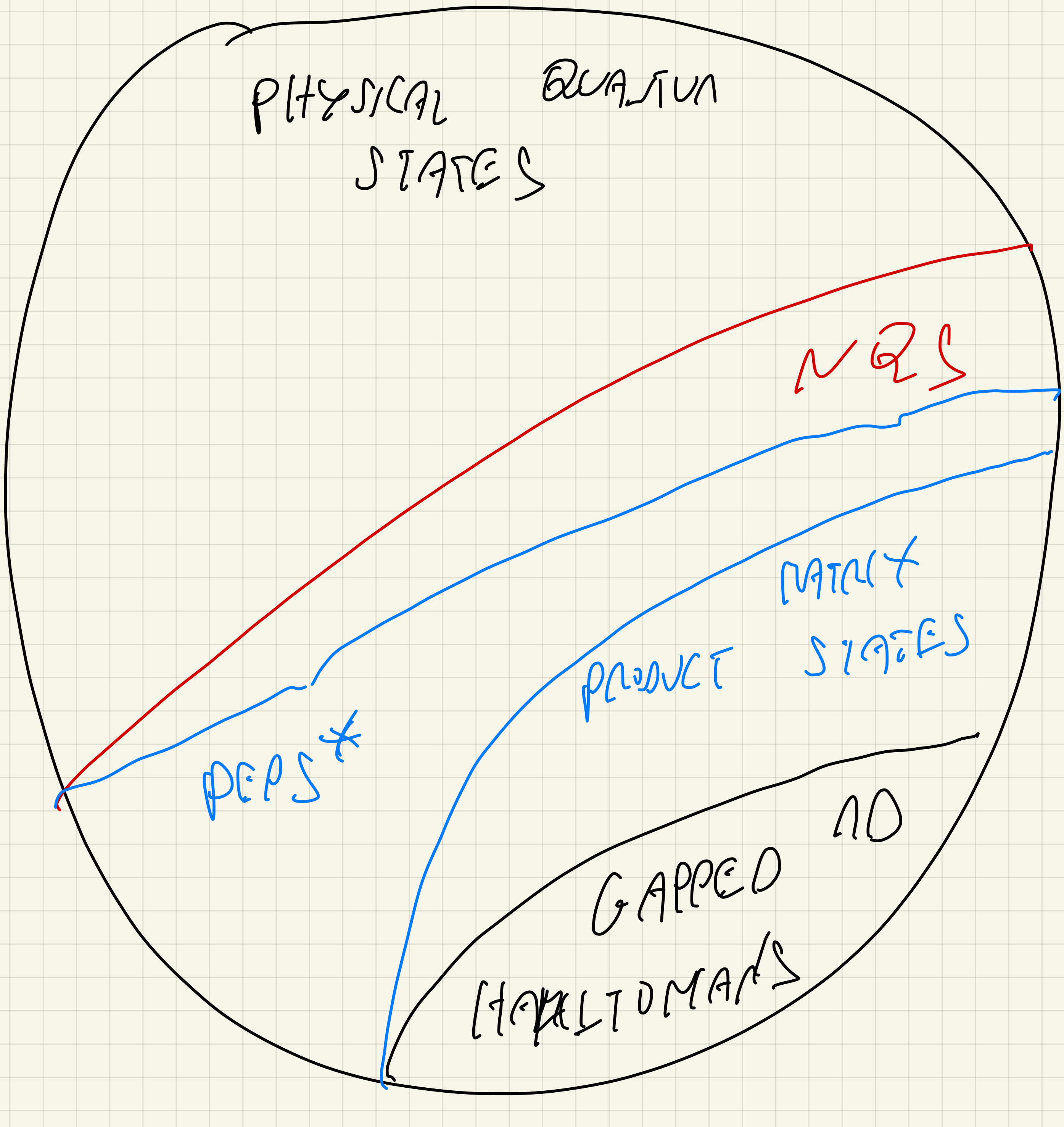
VOLUME LAW ENTANGLEMENT

LEVINE ET AL.
PRL 122, 065301 (2019)



$$D \sim N^d$$

GENERAL TENSOR NETWORKS



SHAMIR ET AL.

ARXIV: 2203.10239
(2022)

RECAP

VARIATIONAL STATES $|\psi(w)\rangle = \sum_k C_k(w)|k\rangle$

(I) $\langle \hat{O} \rangle$ COMPUTE "EXACTLY"

(II) COMPUTE STATISTICALLY $\langle \hat{O} \rangle$

(II) (a) $\langle k|\psi\rangle$ $\rightarrow \langle \hat{O} \rangle \sim \frac{1}{M}\sum_i^M O_{LOC}(k^{(i)})$

(b) $P(k) = \frac{|\langle k|\psi\rangle|^2}{\langle\psi|\psi\rangle}$

SAMPLING STRATEGIES BASED ON MCMC

NEURAL-NETWORK QUANTUM STATES

$\langle k|\psi\rangle = g^{(D)} \cdot W^{(D)} \dots W^{(1)} \vec{k}$

FEED-FORWARD

AUTOREGRESSIVE QUANTUM STATES

EXACT SAMPLING

$$\langle \sigma_1^z \ldots \sigma_N^z | \Psi \rangle = \prod_i^N \Psi_i(\sigma_i^z | \sigma_{i-1}^z \ldots \sigma_1^z)$$

SHARIR ET. AL

PRL 124, 020503

(2020)

$$\sum_{\{\sigma\}} |\Psi_i(\sigma | \sigma_{i-1}^z \ldots \ldots \sigma_1^z)|^2 = 1$$

$$\downarrow$$

$$\sum_{\{\sigma_1 \ldots \sigma_N\}} |\Psi(\sigma_1^z \ldots \sigma_N^z)|^2 = 1$$

$$P(\vec{\sigma}) = \prod_i^N |\Psi_i(\sigma_i^z | \sigma_{i-1}^z \ldots \sigma_1^z)|^2$$

LEARNING THE GROUND STATE

$$\mathcal{L}(w) = E(w) = \frac{\langle \Psi(w) | \hat{H} | \Psi(w) \rangle}{\langle \Psi(w) | \Psi(w) \rangle}$$

$$E(w) \simeq \frac{1}{M} \sum_{i}^{M} E_{loc}\left(k^{(i)}\right) = L(w)$$

$$E_{loc}(k) = \sum_{k'} \langle k | H | k' \rangle \frac{\langle k' | \Psi \rangle}{\langle k | \Psi \rangle}$$

GRADIENTS OF ENERGY

$$\frac{\partial E(W)}{\partial W_p} = 2\,\mathrm{Re}\Big[\langle E_{loc}(k)\, O_p^*(k)\rangle + - \langle E_{loc}(k)\rangle \langle O_p^*(k)\rangle\Big]$$

$$O_p(k) = \frac{\partial}{\partial W_p} \log \langle k|\psi\rangle = \langle k|\hat{O}_p|k\rangle$$

LEARNING ALGORITHM

0. INITIALIZE WEIGHTS $W^{(0)}$
1. SAMPLE $P(k; W^{(S)}) \sim k^{(1)} \ldots k^{(M)}$
2. MEASURE $E(W) \simeq \langle E_{loc}(k)\rangle = \langle \hat{H} \rangle$

   $$\frac{\partial E(W)}{\partial W_p} \simeq \langle \;\; \rangle - \langle\;\rangle\langle\;\rangle$$

3. $$W_p^{(S+1)} = W_p^{(S)} - \eta \frac{\partial E}{\partial W_p}$$

HOW CAN WE LEARN DYNAMICS?

"SOLVE" VARIATIONALLY

$$i \frac{\partial}{\partial t} |\phi\rangle = \hat{H}(t) |\phi\rangle \longrightarrow |\phi(t)\rangle$$

"STANDARD" VARIATIONAL PRINCIPLE

$$E(w) = \frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle} \geq E_0$$

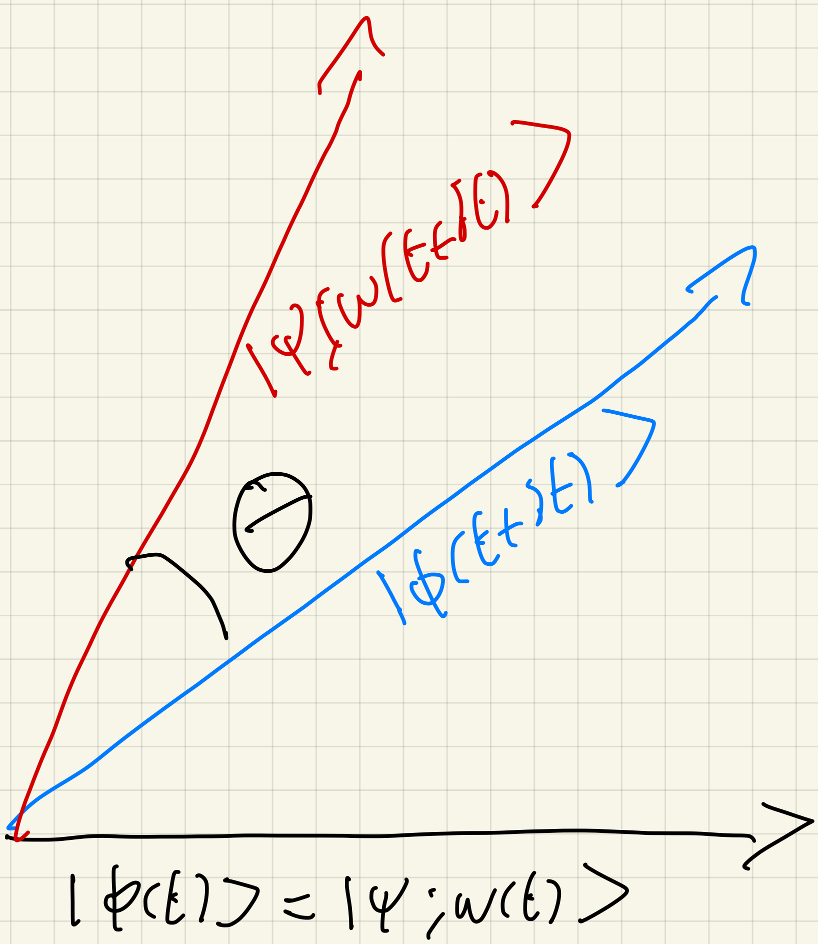
TIME-DEPENDENT VARIATIONAL STATES

$$|\phi(t)\rangle \simeq |\Psi ; W(t)\rangle$$

MOVE TIME DEPENDENCE INSIDE THE VAR. PARAMETERS

TIME-DEPENDENT VARIATIONAL PRINCIPLE

[DIRAC, FRENKEL]



(I) $|\phi(t)\rangle = |\psi; w(t)\rangle$ INITIAL CONDITION

(II) $$|\phi(t+\delta_t)\rangle = (1 - i\delta_t \hat{H})|\phi(t)\rangle + O(\delta_t^2)$$

$$|\psi; w(t+\delta_t)\rangle = (1 + \delta_t \hat{\Delta})|\phi(t)\rangle + O(\delta_t^2)$$

WHAT IS $\hat{\Delta}$?

$$\langle k|\psi; w(t+\delta_t)\rangle = \langle k|\psi; w(t)\rangle +$$

$$+ \delta_t \sum_p \frac{\partial}{\partial w_p} \langle k|\psi; w(t)\rangle \dot{w}_p + O(\delta_t^2)$$

$$\hat{\Delta} \Rightarrow \sum_p \hat{O}_p \dot{w}_p(t)$$

$$|A\rangle = \left(\hat{1} - i\delta_t \hat{H}\right)|\phi(t)\rangle$$

$$|B\rangle = \left(\hat{1} + \delta_t \sum_p \hat{O}_p \dot{w}_p\right)|\phi(t)\rangle$$

$$F_{AB}(\dot{w}) = \frac{|\langle A|B\rangle|^2}{\langle A|A\rangle\langle B|B\rangle} \quad , \quad \boxed{\frac{\partial F_{AB}}{\partial \dot{w}_p} = 0}$$

$$\max_{\dot{w}_p} F_{AB}(\dot{w}) \longrightarrow \text{EQUATION OF MOTION FOR } \dot{w}_p$$

$$i \sum_{p'} \underbrace{S_{pp'}}_{\text{QUANTUM GEOMETRIC TENSOR}} \dot{w}_{p'} = \underbrace{\mathrm{Re}\left[\langle O_p^* E_{loc}\rangle - \langle O_p^*\rangle\langle E_{loc}\rangle\right]}_{G_p}$$

$$S_{pp'} = \mathrm{Re}\left[\left\langle \frac{\partial \psi}{\partial w_p} \middle| \frac{\partial \psi}{\partial w_{p'}} \right\rangle - \left\langle \frac{\partial \psi}{\partial w_p} \middle| \psi \right\rangle \left\langle \psi \middle| \frac{\partial \psi}{\partial w_{p'}} \right\rangle\right]$$

$$S_{pp'} \simeq R_0 \left[\langle O_p^* O_{p'} \rangle - \langle O_p^* \rangle \langle O_{p'} \rangle \right]$$

$$\langle (\dots) \rangle \simeq \frac{1}{M} \sum_i^M P(t^{(i)}) (\dots\dots)$$

TIME-DEPENDENT VARIATIONAL MONTE CARLO

(t-VMC) [SCIENTIFIC REPORTS 2, 243 (2012)]

0. $W(t=0)$

I. SAMPLE $P(t;t) = \frac{|\langle t|\psi; W(t)\rangle|^2}{\langle\psi|\psi\rangle}$

II. "MEASURE" $S_{pp'}$, G_p

III. SOLVE LINEAR SYSTEM $\sum_{p'} S_{pp'} \dot{W}_p = G_p$

IV. $W_p(t+\delta t) = W_p(t) + \delta t\, \dot{W}_p(t)$

ALSO IN IMAGINARY TIME

$$|\phi(\tau)\rangle = e^{-\tau\hat{H}} |\phi(0)\rangle \longleftrightarrow |\phi(t)\rangle = e^{-it\hat{H}} |\phi(0)\rangle$$

EXACT GS OF $\hat{H}$

$$\lim_{\tau \to \infty} |\phi(\tau)\rangle = |\Psi_0\rangle \quad \text{if } \langle\Psi_0|\phi(0)\rangle \neq 0$$

$$\sum_{p'} S_{pp'} \dot{\psi}_p = \mathrm{Re}\left[\langle O_p^*\rangle\langle E_{loc}\rangle - \langle E_{loc} O_p^*\rangle\right]$$

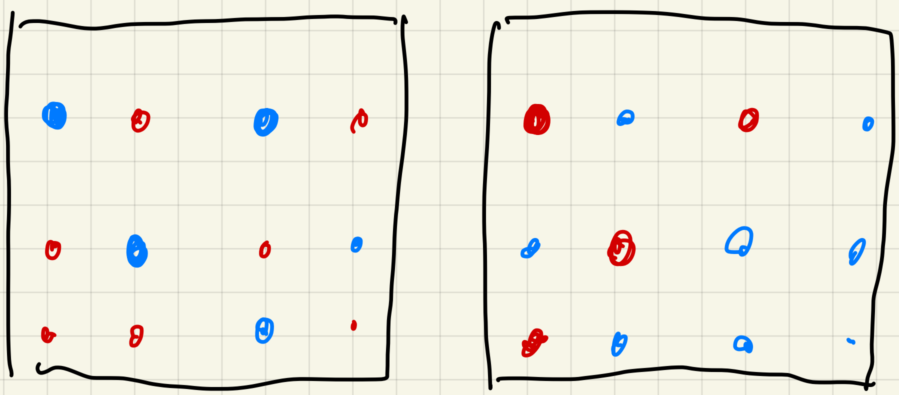
EQUIVALENT TO "STOCHASTIC RECONFIGURATION"
[SORELLA, SEE "QMC BOOK"]

AND VERY CLOSE TO "NATURAL GRADIENT" DESCENT [AMARI, 1998]

LEARNING FROM EXPERIMENTS

$$Q(k) = |\langle k|\phi\rangle|^2 \rightarrow$$ PROJECTIVE MEASUREMENT

$$\langle k|\phi\rangle = |\langle k|\phi\rangle|\, e^{i\theta(k)}$$



$k^{(1)}$ $\quad$ $k^{(2)}$ $\quad$ ... $\quad$ $k^{(M)}$

$\boxed{\text{GOAL}}$ RECONSTRUCT $\langle k|\phi\rangle$ FROM PROJECTIVE MEASUREMENTS

[QUANTUM STATE RECONSTRUCTION]

$$\langle k|\phi\rangle \simeq \langle k|\psi(w)\rangle$$

VARIATIONAL PRINCIPLE ?

$$P(k) = \frac{|\langle k|\psi;w\rangle|^2}{\langle\psi|\psi\rangle}, \quad Q(k) = |\langle k|\phi\rangle|^2$$

$$= \frac{F(k;w)}{N(w)}$$

UNSUPERVISED LEARNING

$$D_{KL}(P \,\|\, Q) = \sum_k Q(k)\left[\log Q(k) - \log P(k)\right]$$

if $P = Q$, $D_{KL} = 0$ in general

$$D_{KL} \geq 0$$

$$L(w) = D_{KL}(w)$$

GRADIENT OF KL

$$\frac{\partial}{\partial w_p} D_{KL}(w) = \left\langle \frac{\partial}{\partial w_p} \log F(k) \right\rangle_{\underline{P}} - \left\langle \frac{\partial}{\partial w_p} \log F(k) \right\rangle_Q$$

$$\frac{\partial}{\partial w_p} \log F(k) = 2\, O_p(k)$$

MEASURING IN MORE BASES

$$Q_B(k) = |\langle k | U_B^* | \phi \rangle|^2$$

$$D_{KL} = \sum_{kB} Q_B(k)\left[\log Q_B(k) - \log P_B(k)\right]$$

$$P_B(k) = \frac{|\langle k | U_B^* | \psi \rangle|^2}{N_B}$$

EFFICIENT IF U_B IS A k-LOCAL ROTATION

TORLAI ET AL., NATURE PHYSICS 14, 447 (2018)