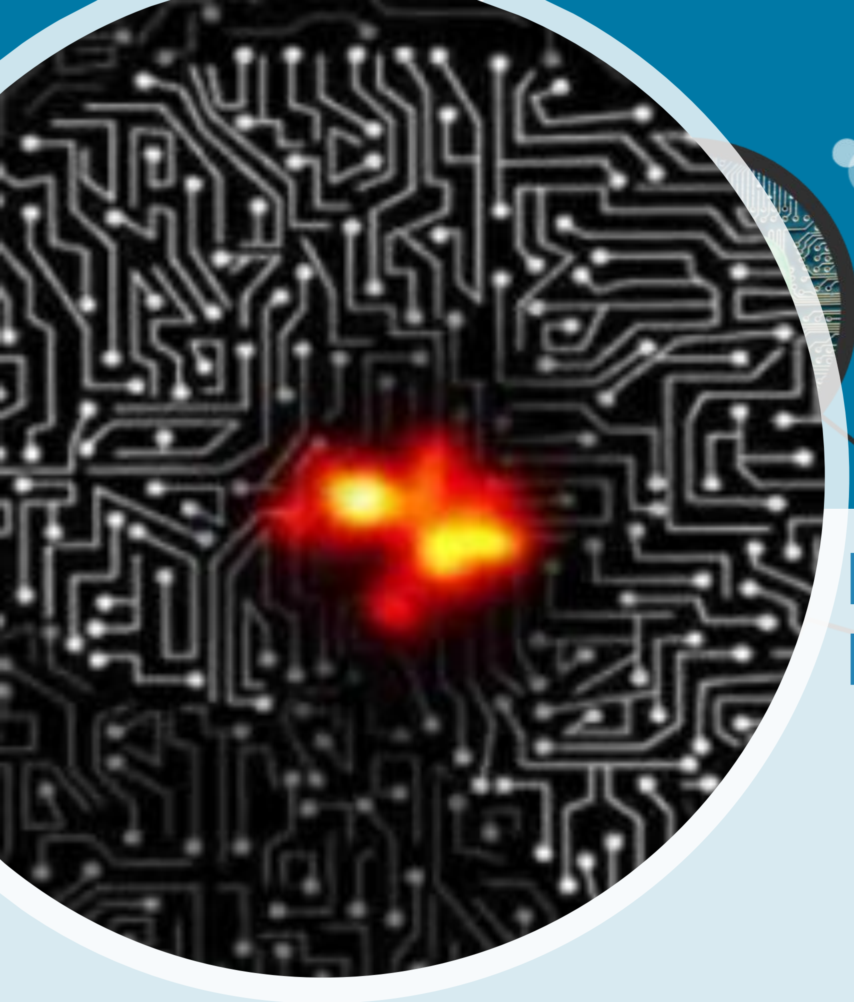


FIAS Frankfurt Institute for Advanced Studies

Deep Learning for Heavy Quark Potential

Kai Zhou (FIAS)

华大QCD讲习班： AI4Physics

From LQCD to in-medium HQ interactions via Deep Learning

- Introduction (potential model, lQCD measurements)
- Methodology (DNN+Shroedinger, uncertainty)
- Proof of concept
- Consistency check
- Results-Conclusions

With Shuzhe Shi, Jiaxing Zhao, Swagato Mukherjee, Pengfei Zhuang

arXiv: 2105.07862

Introduction

Large mass scale : $m_Q >> \Lambda_{QCD}, T, p$

- Produced via <u>Hard Processes</u> from early stage

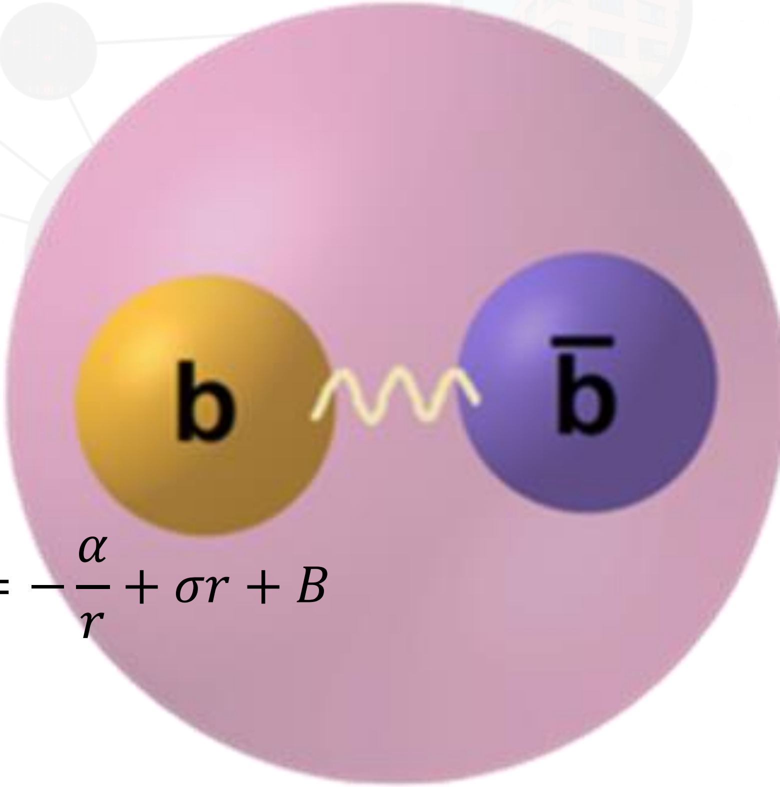
- 'Calibrated' <u>QCD Force</u> – HQ interaction

  In Vacuum : NR potential (NRQCD) , Cornell-like

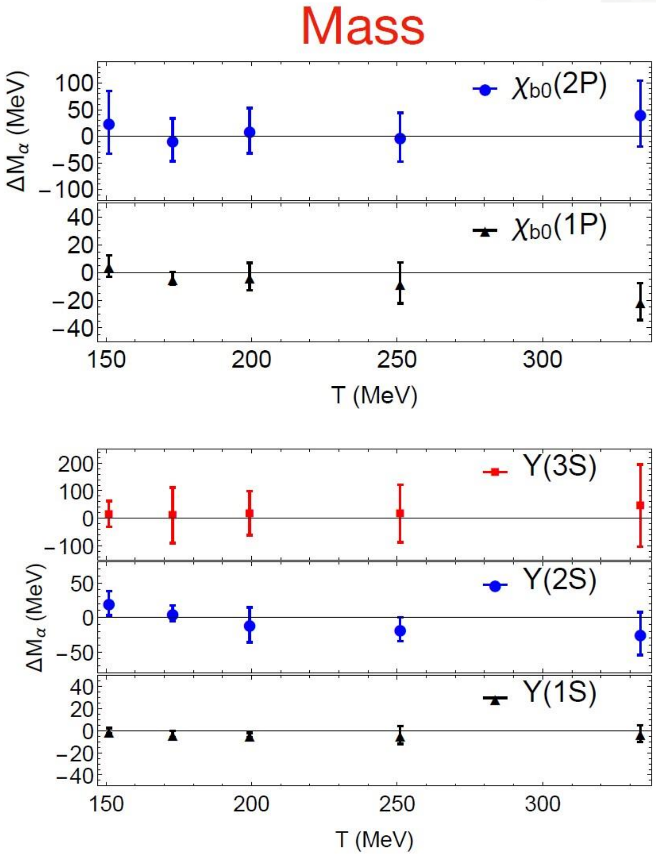
$$V(r) = -\frac{\alpha}{r} + \sigma r + B$$

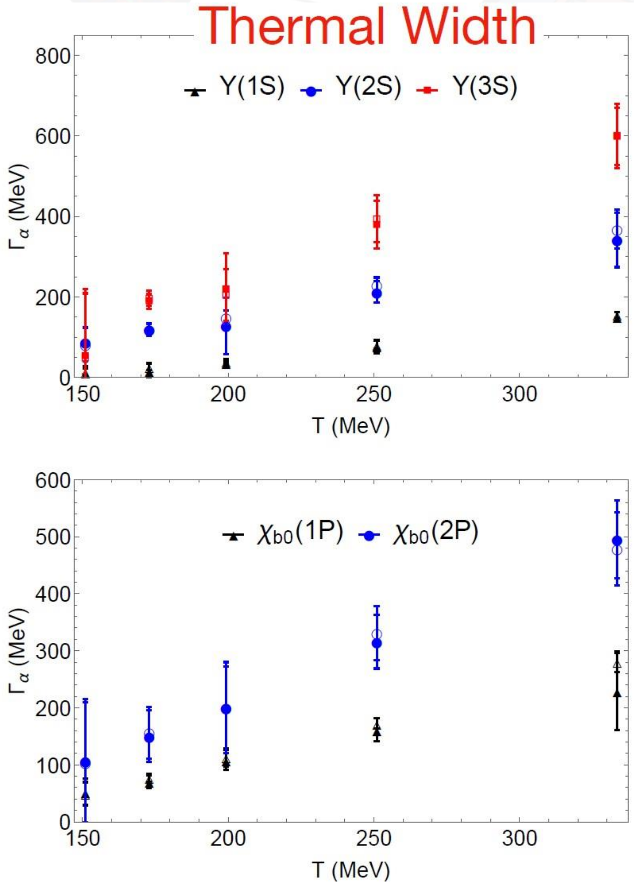
  In Medium : Color Screening , Thermal Width

Laine, et.al, JHEP(2007)



LQCD measured in medium Mass and Width for Bottomonium



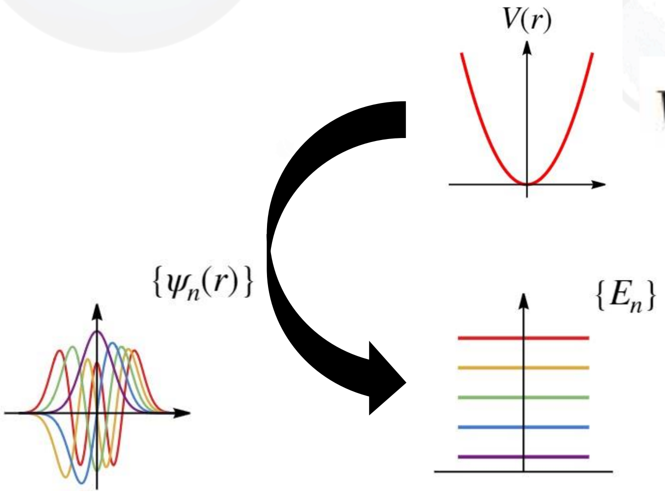


R. Larsen, et.al, PRD(2019), PLB(2020), PRD(2020)

Potential model : Shroedinger equation

$$\hat{H}\psi_n = -\frac{\nabla^2}{2m_\mu}\psi_n + V(r)\psi_n = E_n\psi_n$$

M. Strickland, et.at., PRC(2015) PRD(2018), PLB(2020)



$$V(T,r) = V_R(T,r) + i \cdot V_I(T,r)$$

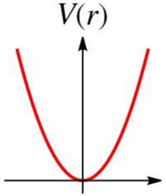
$$\begin{cases} \mathrm{Re}[E_n] = m - 2m_b \\ \mathrm{Im}[E_n] = -\Gamma \end{cases}$$

Inverse Power method

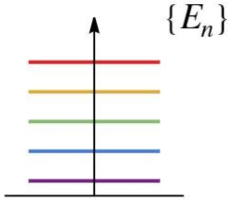
H.W.Crater, JCP(1994)

Potential model : Shroedinger equation

$$\hat{H}\psi_n = -\frac{\nabla^2}{2m_\mu}\psi_n + V(r)\psi_n = E_n\psi_n$$



$$V(T,r) = V_R(T,r) + i \cdot V_I(T,r)$$

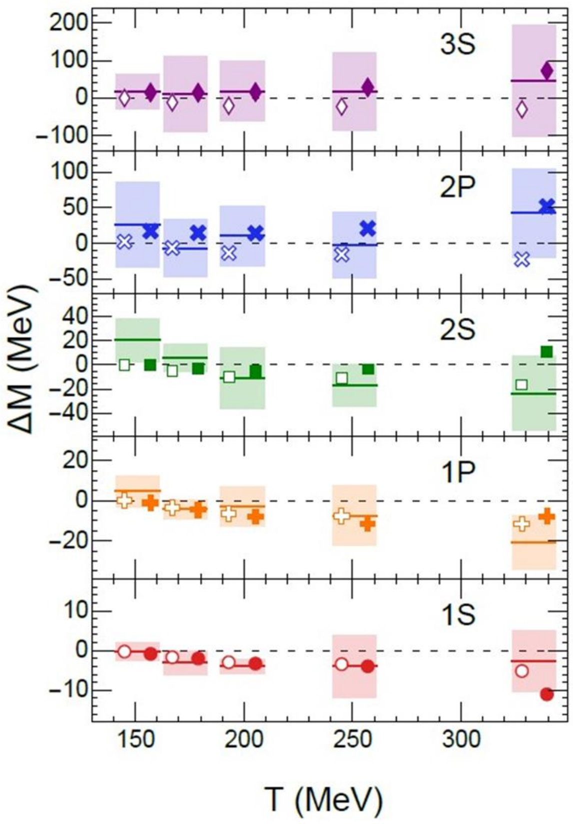


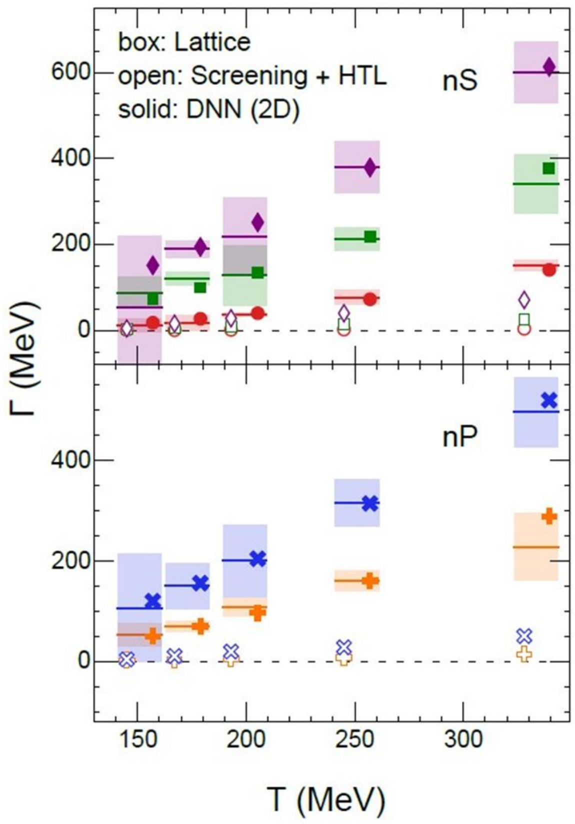
$$\begin{cases} \mathrm{Re}[E_n] = m - 2m_b \\ \mathrm{Im}[E_n] = -\Gamma \end{cases}$$

?

Inverse Problem !

LQCD data (color box) vs. best fit of HTL (open symbol) and of DNNs (solid symbol)



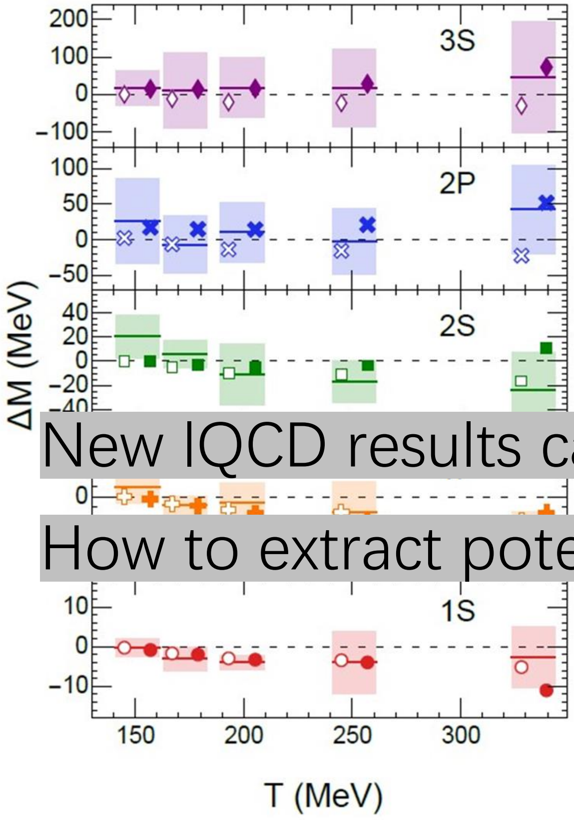


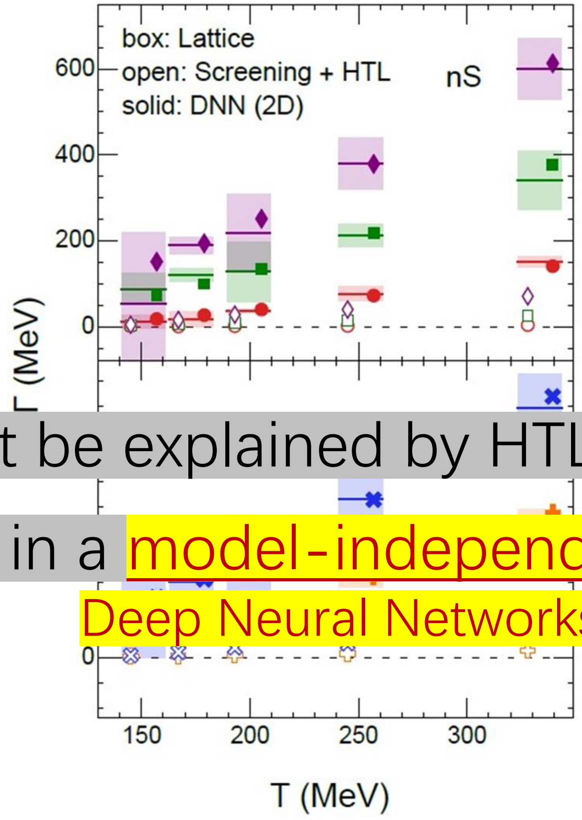
$$V_R(T,r) = \frac{\sigma}{\mu_D}\left(2 - (2 + \mu_D r)e^{-\mu_D r}\right) - \alpha\left(\mu_D + \frac{e^{-\mu_D r}}{r}\right) + B\,,$$

$$V_I(T,r) = -\frac{\sqrt{\pi}}{4}\mu_D T \sigma r^3 G^{2,2}_{2,4}\left(\begin{matrix}-\frac{1}{2},-\frac{1}{2} \\ \frac{1}{2},\frac{1}{2},-\frac{3}{2},-1\end{matrix}\middle| \frac{\mu_D^2 r^2}{4}\right) - \alpha T\,\phi(\mu_D r)\,,$$

HTL from "A. Rothkopf, et.al, PRD(2020)"

LQCD data (color box) vs. best fit of HTL (open symbol) and of DNNs (solid symbol)





$$V_R(T,r) = \frac{\sigma}{\mu_D}\left(2-(2+\mu_D r)e^{-\mu_D r}\right) - \alpha\left(\mu_D + \frac{e^{-\mu_D r}}{r}\right) + B\,,$$

$$V_I(T,r) = -\frac{\sqrt{\pi}}{4}\mu_D T \sigma r^3 G^{2,2}_{2,4}\left(\begin{matrix}-\frac{1}{2},-\frac{1}{2} \\ \frac{1}{2},\frac{1}{2},-\frac{3}{2},-1\end{matrix}\middle|\frac{\mu_D^2 r^2}{4}\right) - \alpha T\,\phi(\mu_D r)\,,$$

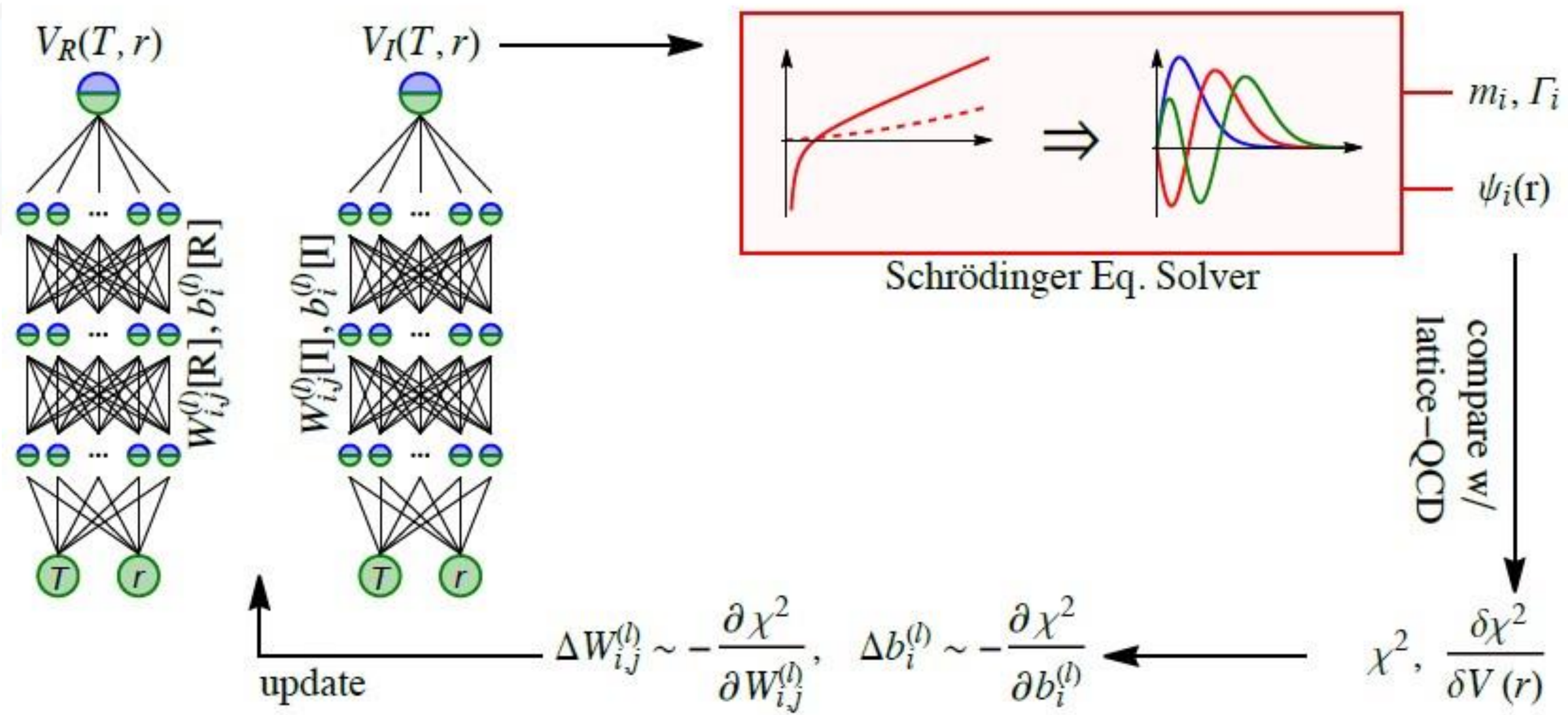
New lQCD results cannot be explained by HTL-inspired potential

How to extract potential in a model-independent way ?

Deep Neural Networks (DNN)

HTL from "A. Rothkopf, et.al, PRD(2020)"

Flow chart of HQ potential reconstruction with DNN



DNN basic : Universal Function Approximator

$$(f: \mathbb{R}^n \to \mathbb{R}^m) \quad \vec{x} \to \vec{y}$$

$$z_i^{(l)} = b_i^{(1)} + \sum_j W_{ij}^{(l)} a_j^{(l-1)}, \qquad a_i^{(l)} = \sigma^{(l)}\left(z_i^{(l)}\right)$$

ELU

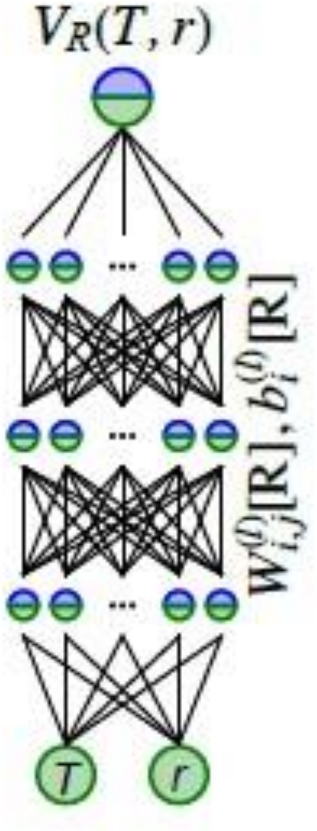
$$\longrightarrow \quad a^{(N)} = \tilde{y}(x;\theta) \qquad \theta \equiv \left\{W_{ij}^{(l)}, b_i^{(l)}\right\}$$

Gradient Descent for parameter tuning :

$$\Delta\theta \equiv \theta^{[k+1]} - \theta^{[k]} \sim -\nabla_\theta J(\theta)$$

Cost, e.g. : $J(\theta) = \frac{1}{2} \sum_{\mathrm{x} \in \text{data set}} \left|\tilde{\mathrm{y}}(\theta, \mathrm{x}) - \mathrm{y}(\mathrm{x})\right|^2 + \frac{\lambda}{2} \theta \cdot \theta$

DNN basic : Universal Function Approximator



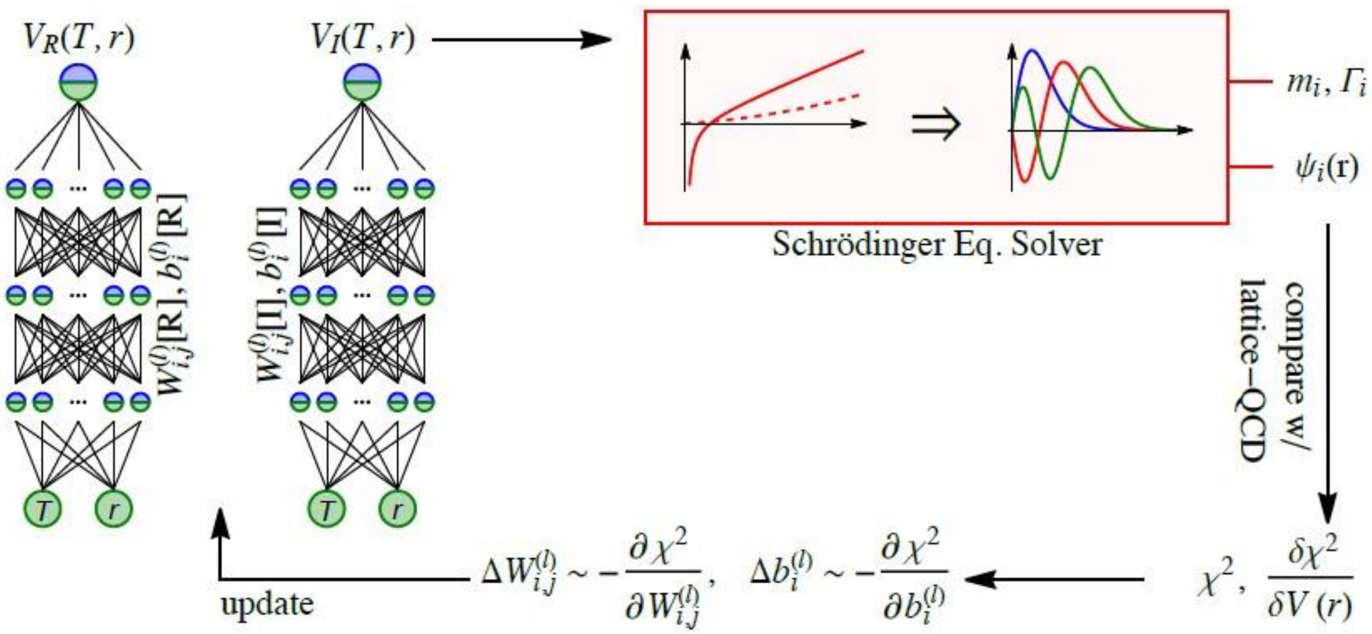
$$\frac{\partial J}{\partial \theta_i} = \sum_{x \in \text{data set}} \left(\tilde{y}(\theta, x) - y(x)\right) \cdot \frac{\partial \tilde{y}(\theta, x)}{\partial \theta_i} + \lambda \theta_i$$

$$z_i^{(l)} = b_i^{(1)} + \sum_j W_{ij}^{(l)} a_j^{(l-1)}, \quad a_i^{(l)} = \sigma^{(l)}\left(z_i^{(l)}\right)$$

$$\longrightarrow \quad \frac{\partial J}{\partial w_{ij}^{[l]}} = a_j^{[l-1]} \frac{\partial J}{\partial z_i^{[l]}} \qquad \frac{\partial J}{\partial b_i^{[l]}} = \frac{\partial J}{\partial z_i^{[l]}}$$

$$\frac{\partial J}{\partial z_i^{[l]}} = \sigma'\left(z_i^{[l]}\right) \sum_j W_{ji}^{[l+1]} \frac{\partial J}{\partial z_j^{[l+1]}}$$

Cost function for "DNN + Schroedinger Eq."



$$\chi^2 = \sum_{T,i} \frac{(m_{T,i} - m_{T,i}^{\text{lattice}})^2}{(\delta m_{T,i}^{\text{lattice}})^2} + \frac{(\Gamma_{T,i} - \Gamma_{T,i}^{\text{lattice}})^2}{(\delta\Gamma_{T,i}^{\text{lattice}})^2}$$

$$J(\boldsymbol{\theta}) = \frac{1}{2}\chi^2(\boldsymbol{\theta}) + \frac{\lambda}{2}\boldsymbol{\theta}\cdot\boldsymbol{\theta},$$

$$T \in \{0, 151, 173, 199, 251, 334\}\ \text{MeV}$$

$$i \in \{1S, 2S, 3S, 1P, 2P\}$$

Perturbation on Shroedinger Eq.

$$\left(\frac{\widehat{p}^2}{2m} + V(r)\right)|\psi_i\rangle = E_i|\psi_i\rangle,$$

$$\left(\frac{\widehat{p}^2}{2m} + V(r) + \delta V(r)\right)|\psi_i'\rangle = (E_i + \delta E_i)|\psi_i'\rangle.$$

Perturbation on Shroedinger Eq.

$$\left(\frac{\hat{p}^2}{2m} + V(r)\right)|\psi_i\rangle = E_i|\psi_i\rangle,$$

$$\left(\frac{\hat{p}^2}{2m} + V(r) + \delta V(r)\right)|\psi_i'\rangle = (E_i + \delta E_i)|\psi_i'\rangle.$$

$$\Rightarrow \quad \delta E_i = \langle\psi_i|\delta V(r)|\psi_i\rangle, \qquad |\psi_i'\rangle = |\psi_i\rangle + \sum_{j\neq i}\frac{\langle\psi_j|\delta V(r)|\psi_i\rangle}{E_i - E_j}|\psi_j\rangle.$$

Hellmann-Feynman theorem

Phys. Rev. (1939)

Perturbation on Shroedinger Eq.

$$\left(\frac{\hat{p}^2}{2m} + V(r)\right)|\psi_i\rangle = E_i|\psi_i\rangle,$$
$$\left(\frac{\hat{p}^2}{2m} + V(r) + \delta V(r)\right)|\psi_i'\rangle = (E_i + \delta E_i)|\psi_i'\rangle.$$

$$\Rightarrow \quad \delta m_i = \langle\psi_i|\delta V_R(r)|\psi_i\rangle, \quad \delta\Gamma_i = -\langle\psi_i|\delta V_I(r)|\psi_i\rangle.$$

$$|\psi_i'\rangle = |\psi_i\rangle + \sum_{j\neq i} \frac{\langle\psi_j|\delta V(r)|\psi_i\rangle}{E_i - E_j}|\psi_j\rangle.$$

Perturbation on Shroedinger Eq.

$$\left(\frac{\hat{p}^2}{2m} + V(r)\right)|\psi_i\rangle = E_i|\psi_i\rangle,$$
$$\left(\frac{\hat{p}^2}{2m} + V(r) + \delta V(r)\right)|\psi_i'\rangle = (E_i + \delta E_i)|\psi_i'\rangle.$$

$\Rightarrow$

$$\delta m_i = \langle\psi_i|\delta V_R(r)|\psi_i\rangle,$$
$$\delta\Gamma_i = -\langle\psi_i|\delta V_I(r)|\psi_i\rangle.$$

$$|\psi_i'\rangle = |\psi_i\rangle + \sum_{j\neq i} \frac{\langle\psi_j|\delta V(r)|\psi_i\rangle}{E_i - E_j}|\psi_j\rangle.$$

$$\delta V(r) = v\,\delta(r - r_k),$$

$\Rightarrow$

$$\frac{\delta m_i}{\delta V_R(r)} = -\frac{\delta\Gamma_i}{\delta V_I(r)} = |\psi_i(r)|^2,$$
$$\frac{\delta m_i}{\delta V_I(r)} = \frac{\delta\Gamma_i}{\delta V_R(r)} = 0.$$

Gradients for the Cost

$$\chi^2 = \sum_{T,i,j} \Big(R_{ij}^{(T)} \Delta m_{T,i} \Delta m_{T,j} + I_{ij}^{(T)} \Delta\Gamma_{T,i} \Delta\Gamma_{T,j} + 2 M_{ij}^{(T)} \Delta m_{T,i} \Delta\Gamma_{T,j} \Big) ,$$

⇒

$$\frac{\partial \chi^2}{\partial \theta_{R,n}} = \sum_{T,i,k} \frac{\partial \chi^2}{\partial m_{T,i}} \frac{\partial V_R(T, r_k)}{\partial \theta_{R,n}} |\psi_i(T, r_k)|^2 \mathrm{d}r ,$$

$$\frac{\partial \chi^2}{\partial \theta_{I,n}} = -\sum_{T,i,k} \frac{\partial \chi^2}{\partial \Gamma_{T,i}} \frac{\partial V_I(T, r_k)}{\partial \theta_{I,n}} |\psi_i(T, r_k)|^2 \mathrm{d}r ,$$

⇒

$$\frac{\partial J}{\partial \theta_{R,n}} = \sum_{T,i} \Big\{ \Big[\sum_k \frac{\partial V_R(T, r_k)}{\partial \theta_{R,n}} |\psi_i(T, r_k)|^2 \mathrm{d}r \Big] \times \sum_j \Big[R_{i,j}^{(T)} \Delta m_{T,j} + M_{ij}^{(T)} \Delta\Gamma_{T,j} \Big] \Big\} + \lambda \theta_{R,n} ,$$

$$\frac{\partial J}{\partial \theta_{I,n}} = -\sum_{T,i} \Big\{ \Big[\sum_k \frac{\partial V_I(T, r_k)}{\partial \theta_{I,n}} |\psi_i(T, r_k)|^2 \mathrm{d}r \Big] \times \sum_j \Big[I_{i,j}^{(T)} \Delta\Gamma_{T,j} + M_{ij}^{(T)} \Delta m_{T,j} \Big] \Big\} + \lambda \theta_{I,n} ,$$

Uncertainty Estimation – Bayesian Inference

$$\text{Posterior}(\boldsymbol{\theta}|\text{data}) \propto L(\boldsymbol{\theta}|\text{data}) \cdot \text{Prior}(\boldsymbol{\theta}).$$

$$L(\boldsymbol{\theta}|\text{data}) = P(\text{data}|\boldsymbol{\theta}) \propto \exp[-\chi^2(\boldsymbol{\theta})/2].$$

$$\text{Prior}(\boldsymbol{\theta}) \propto \exp[-\frac{\lambda}{2}\boldsymbol{\theta} \cdot \boldsymbol{\theta}].$$

$$\text{Posterior}(\boldsymbol{\theta}|\text{data}) = N_0 \exp\left[-\frac{\chi^2(\boldsymbol{\theta})}{2} - \frac{\lambda}{2}\boldsymbol{\theta} \cdot \boldsymbol{\theta}\right]$$

Sample potentials ~ $P(V_{\boldsymbol{\theta}}(T, r)) = \text{Posterior}(\boldsymbol{\theta}|\text{data})\,.$

Reference Sampler ~

$$\tilde{P}(\boldsymbol{\theta}) = (2\pi)^{-N_\theta/2}\sqrt{\det[\Sigma^{-1}]} \times \exp\left[-\frac{\Sigma^{-1}_{ab}}{2}(\theta_a - \theta_a^{\text{opt}})(\theta_b - \theta_b^{\text{opt}})\right]$$

$$\left(\Sigma^{-1}_{ab} \equiv \frac{\partial^2 J(\boldsymbol{\theta})}{\partial\theta_a \partial\theta_b}\right)$$

re-weighting with :

$\omega(\theta) = p\,(V_\theta(T, r))/\tilde{p}\,(\theta)$ to grantee posterior sampling

Vacuum potential & b-quark mass Calibration

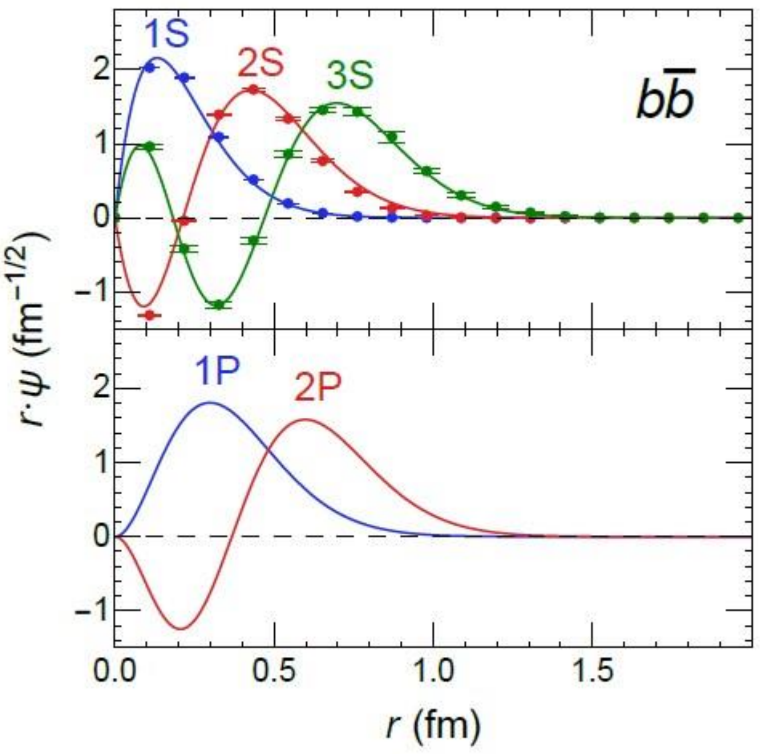
Cornell-Potential

$$V(r) = -\frac{\alpha}{r} + \sigma r + B$$

$m_b = 6.00$ GeV $\alpha = 0.406$

$\sigma = 0.221$ GeV^2 $B = -2.53$ GeV

	1S	2S	3S	1P	2P
experiment (MeV)	9445	10017	10352	9891	10254
model (MeV)	9449	10003	10356	9893	10258
difference (MeV)	+4	−14	+4	+2	+4

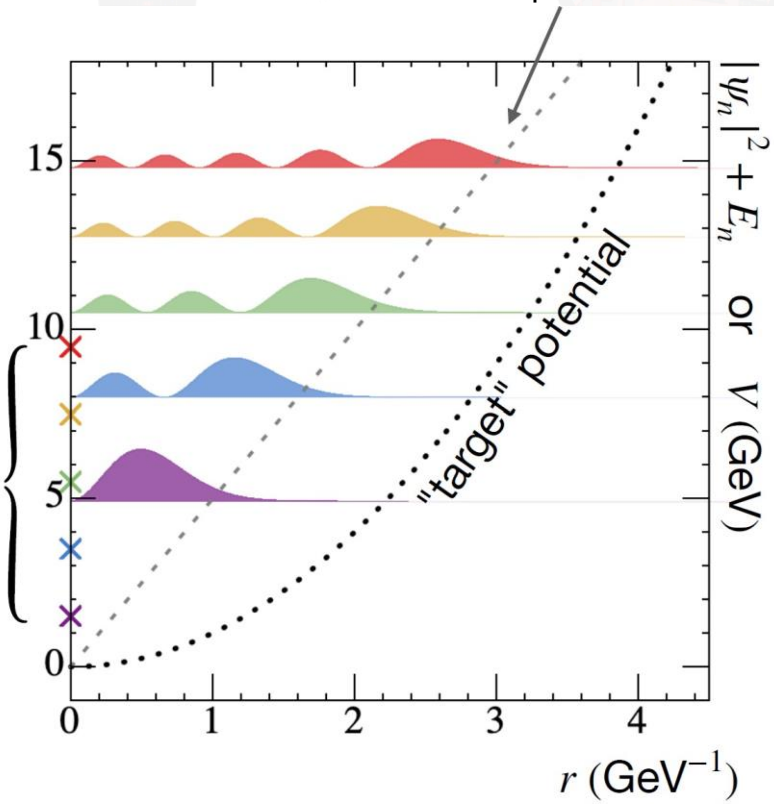


Proof of Concept

limited spectrum { En } to continuous interaction V(r) ?

Learn V(r) from 5 eigenvalues :

{ En } = {3/2, 7/2, 11/2, 15/12, 19/2} GeV

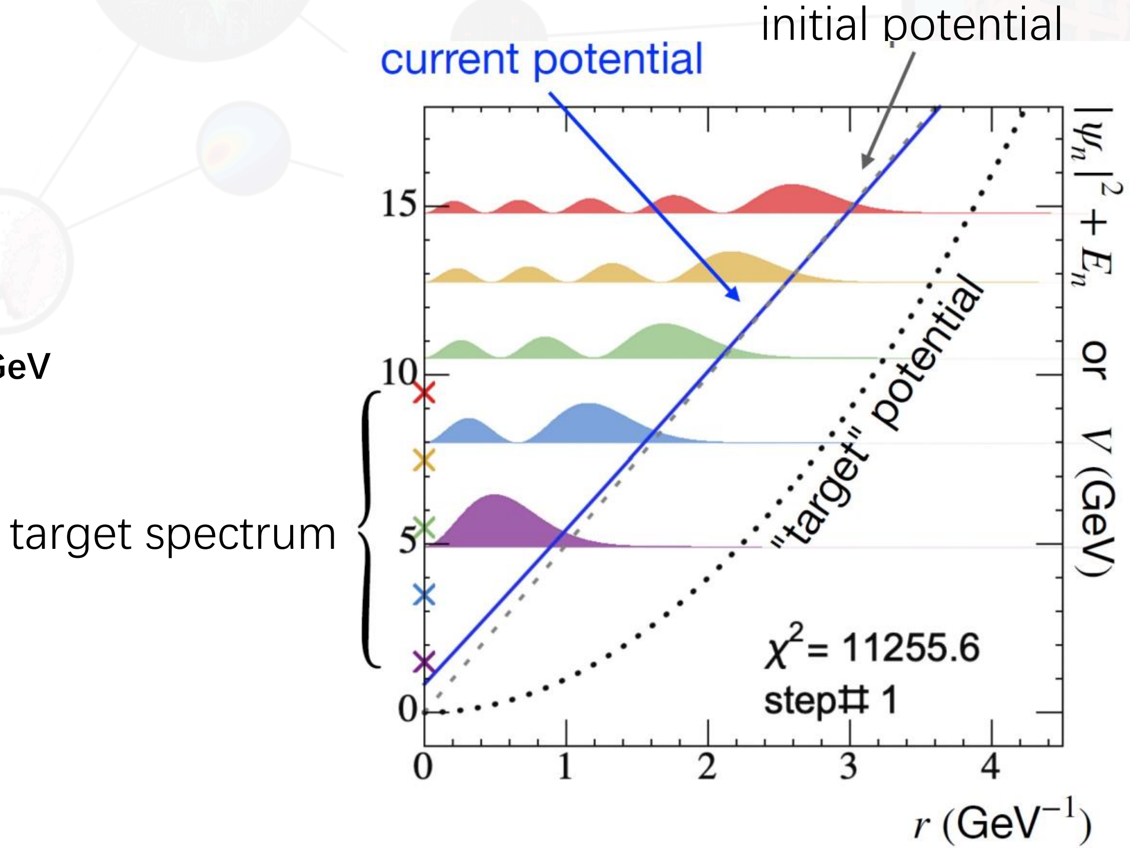


Proof of Concept

limited spectrum { En }   to   continuous interaction V(r) ?

Learn V(r) from 5 eigenvalues :

{ En } = {3/2, 7/2, 11/2, 15/12, 19/2} GeV

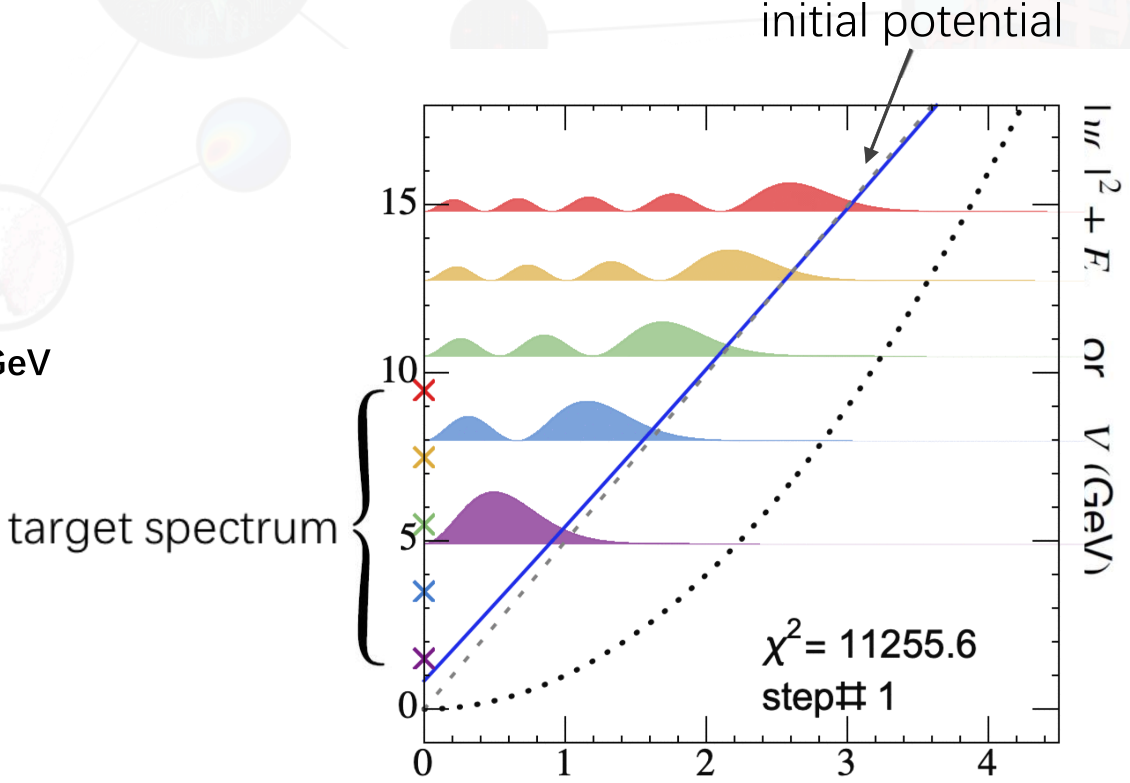


Proof of Concept

limited spectrum { En } to continuous interaction V(r) ?

Learn V(r) from 5 eigenvalues :

{ En } = {3/2, 7/2, 11/2, 15/12, 19/2} GeV



Proof of Concept

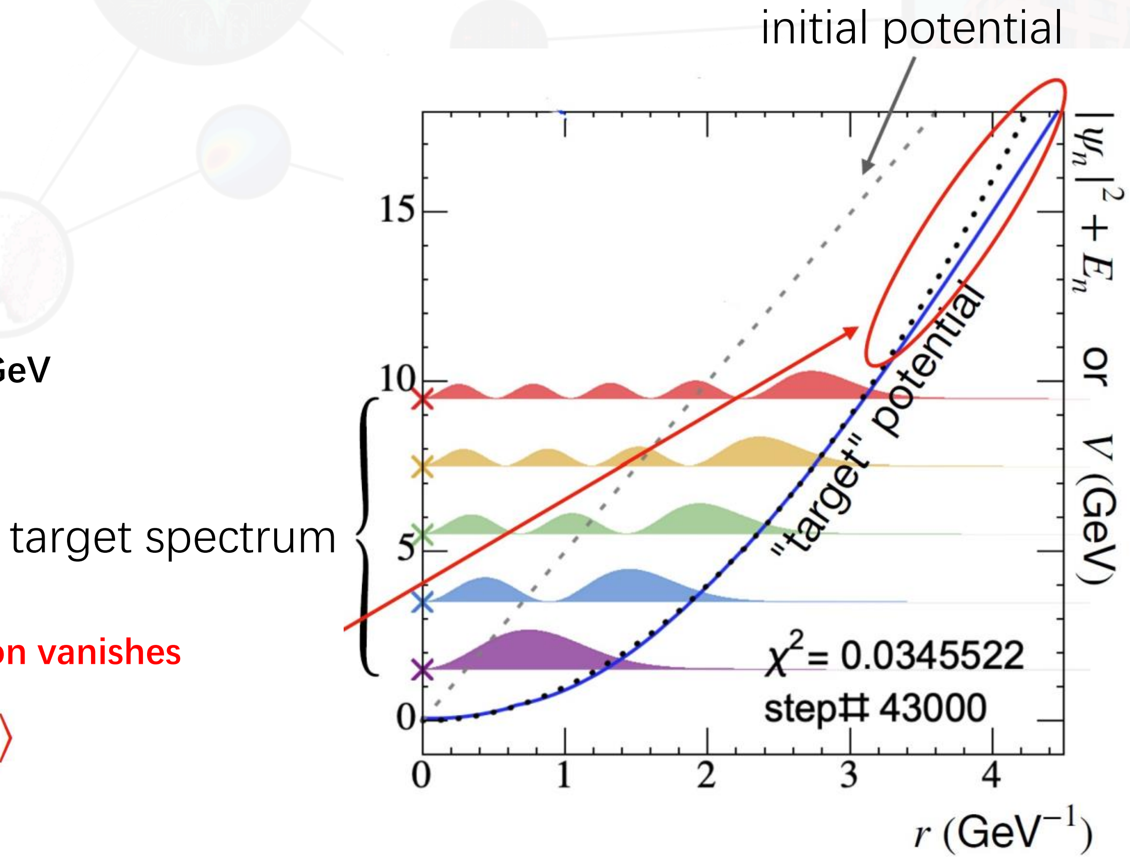
limited spectrum { En } to continuous interaction V(r) ?

Learn V(r) from 5 eigenvalues :

{ En } = {3/2, 7/2, 11/2, 15/12, 19/2} GeV

Deviation @ given states' wavefunction vanishes

$$\delta E_n = \langle \psi_n | \delta V(r) | \psi_n \rangle$$



Closure Test – suppose HTL is true

$$V_R(T,r) = \frac{\sigma}{\mu_D}\left(2-(2+\mu_D r)e^{-\mu_D r}\right) - \alpha\left(\mu_D + \frac{e^{-\mu_D r}}{r}\right) + B\,,$$

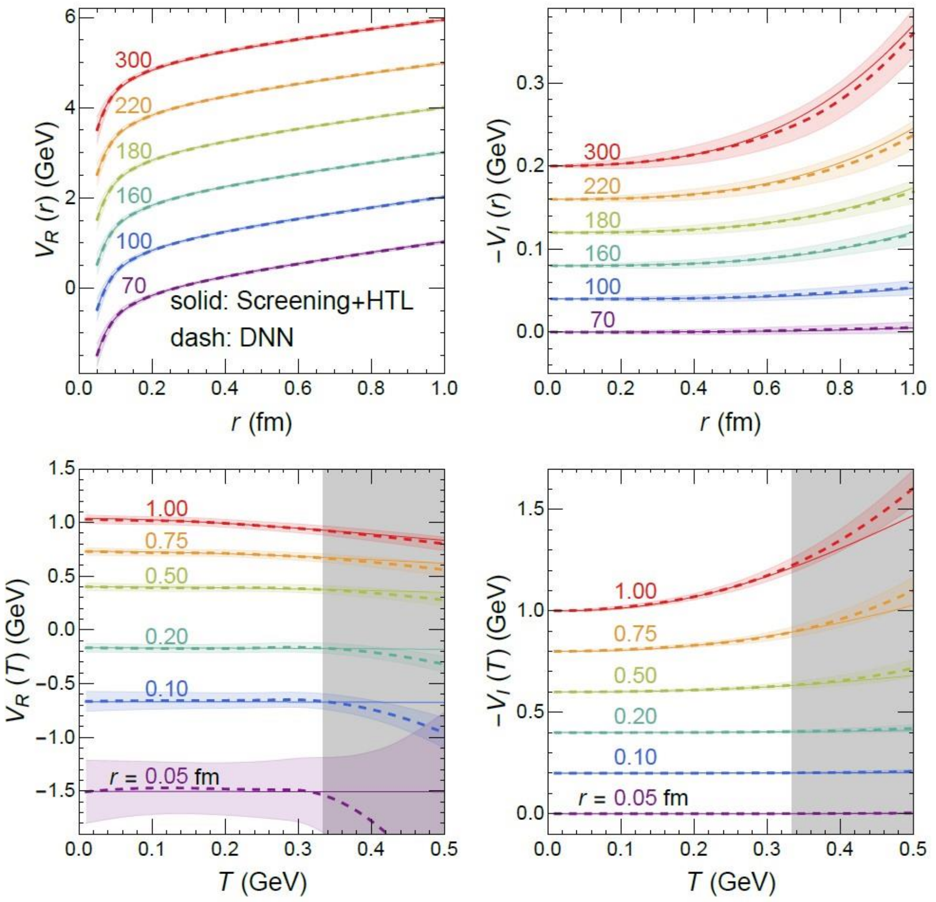
$$V_I(T,r) = -\frac{\sqrt{\pi}}{4}\mu_D T\sigma r^3 G^{2,2}_{2,4}\left(\begin{matrix}-\frac{1}{2},-\frac{1}{2}\\ \frac{1}{2},\frac{1}{2},-\frac{3}{2},-1\end{matrix}\middle|\frac{\mu_D^2 r^2}{4}\right) - \alpha T\,\phi(\mu_D r)\,,$$

$m_b = 4.676$ GeV, $\alpha = 0.39$,
$\sigma = 0.223$ GeV^2, $B = 0$ GeV,
assume that $\mu_D(T) = T/2$.

Provide mass and width of

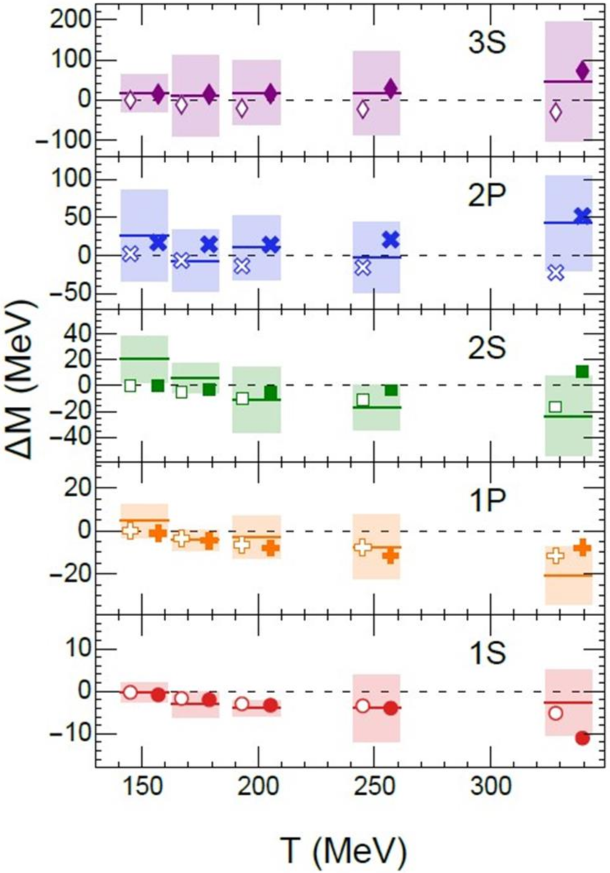
1S, 2S, 3S, 1P, and 2P states.

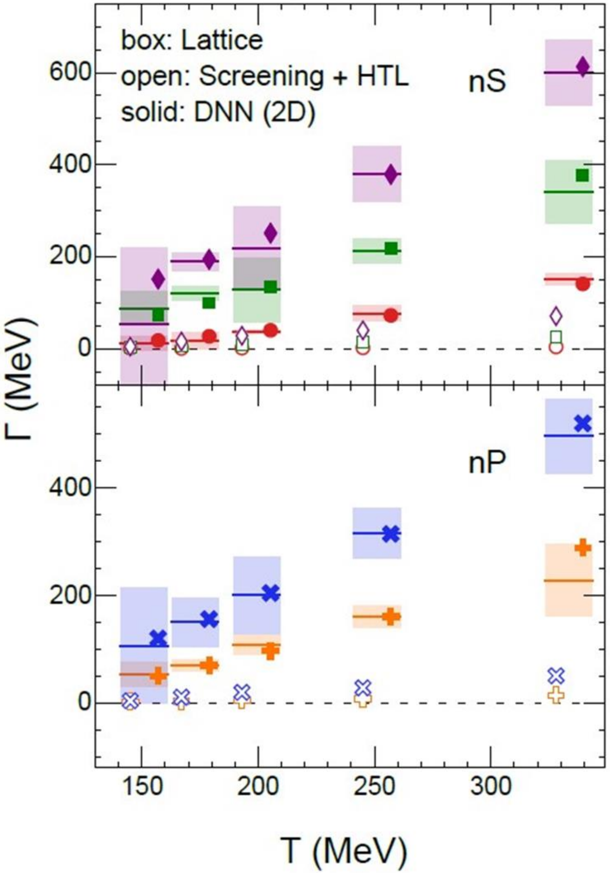
@$(0, 151, 173, 199, 251, 334)$ MeV



Best Fit of lQCD mass and width from HTL (open symbols) and DNNs (solid symbols)

Chi2-per-data=16.5/30





Consistency Check : with different parameterization

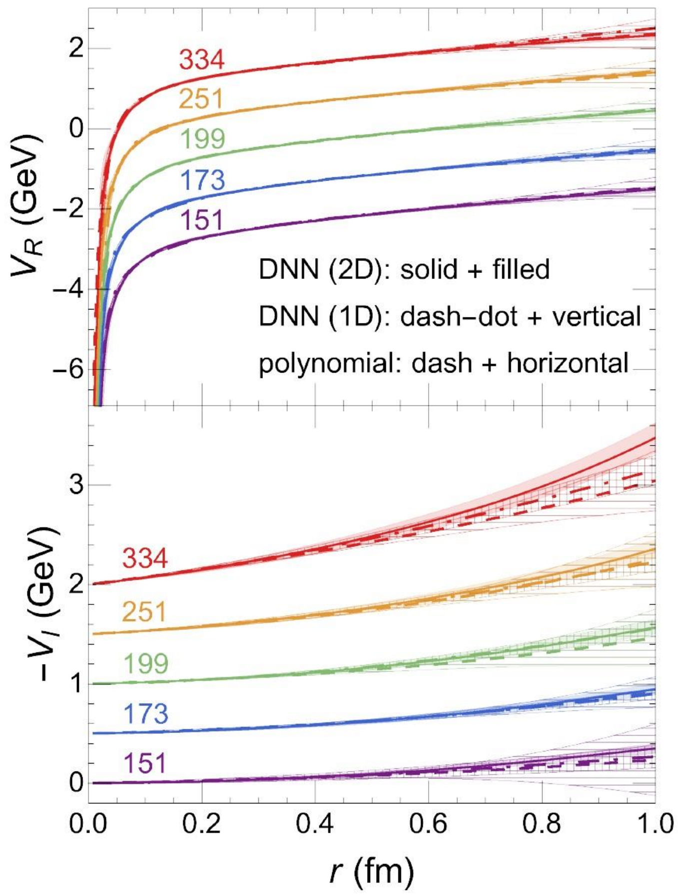
1, **DNN(2D)** :
T & r dependency

2, **DNN(1D)** :
only r dependency

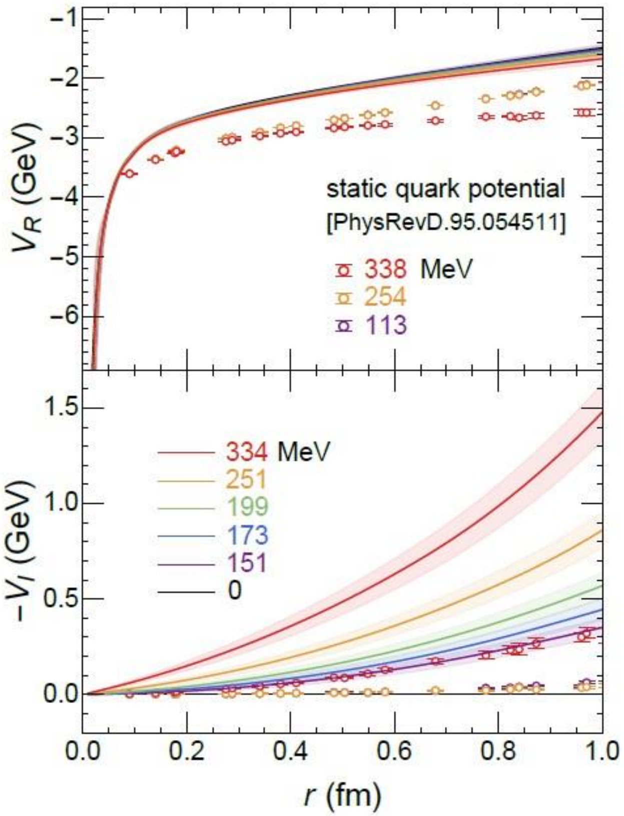
3, **Polynomial** :

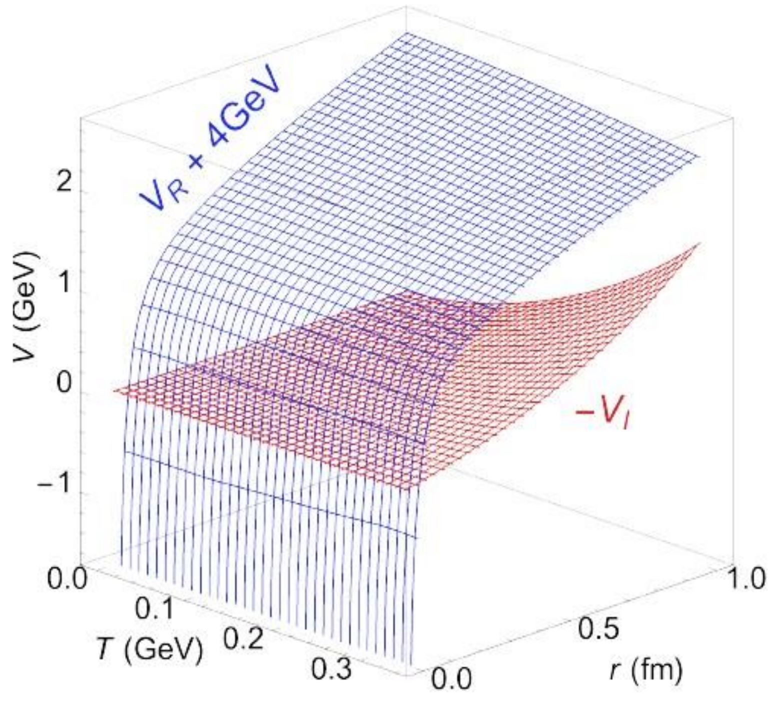
$$V_R(r) = \sum_{i=-1}^{3} c_{R,i}\, r^i,$$

$$V_I(r) = -\sum_{i=1}^{3} c_{I,i}\, r^i.$$



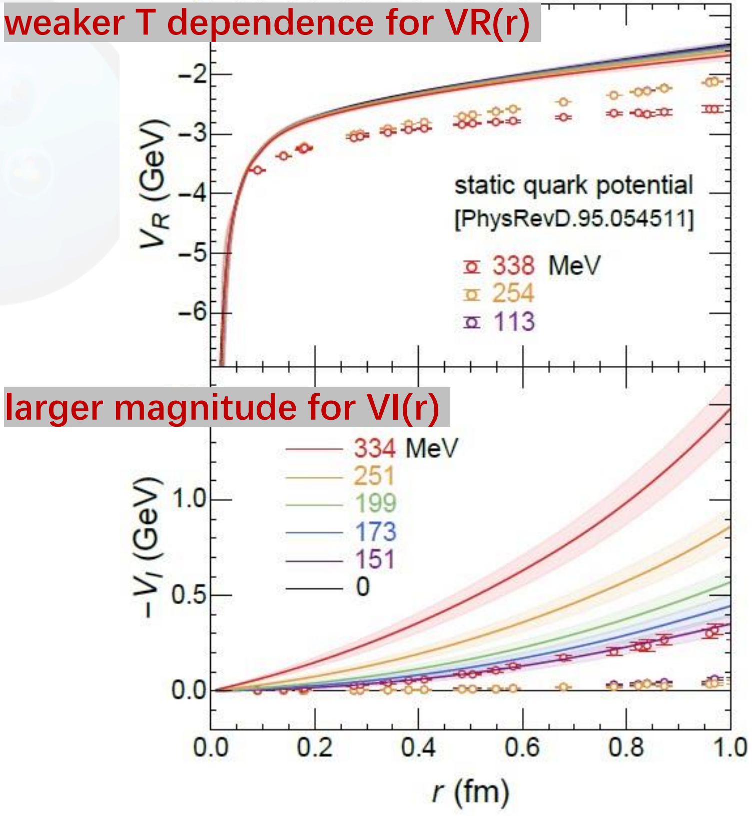
The (B) interaction potentials





A.Rothkopf, PRD(2017)

The (B) interaction potentials



- Traditional picture,

V & F show **platform** at large r, and decrease in height with increase T, So : binding energy decrease, average size increase, until a **melting Temperature**

- New picture,

Im[V] induced thermal width are so significant (**continuous dynamic diss ociation**), its enhancement compensates the vanishing of the melting effect (mild T-dependence of Re[V])

Summary

- Bias-free HQ complex interaction is reconstructed from our novel methodology **'NN+perterb.+Bayesian'**
- Both T and r dependence of the interaction potential are captured via **network representation**
- We found mild T-dependent screening effect for Re[V], while the strength of the Im[V] increases significantly with T
- Color Screening melting to **Continuous dynamic dissociation**

Opportunities as phycists:

"Computers will not completely replace human, at least for one kind, which is those who can set the objective function. If you are able to take a real-world problem and formulate it into a mathematical form for the objective function, you are going to be a master of the future AI system"

– Yang Qiang, HKUST

(1) physics and related (e.g. chemistry, engineering) problems are much <u>better defined</u> than conventional deep learning ones (e.g. image/natural language processing) – much more economic and efficient in tackling

(2) deep learning is a <u>black box</u> – simple physical systems as benchmark

*Renormalization Group *Statistical Physics *collective modes