Frustrated magnetism with neural network wave functions

Attila Szabó

Rudolf Peierls Centre, University of Oxford, U. K.
ISIS Neutron and Muon Source, U. K.
→ MPI-PKS, Dresden, Germany

Benasque, 16 August 2023

Science and Technology Facilities Council
ISIS Neutron and Muon Source

mpipks

Outline

- Neural quantum states (NQS)
  - "Sign problem"
- Symmetric ansätze for VMC
- Group-convolutional networks (GCNN)
- Lattice problems
  PRB **108,** 054410 (2023)
  - Square-lattice $J_1 - J_2$
  - Triangular-lattice $J_1 - J_2$
- Antiferromagnets on fullerene structures
  work in progress

In collaboration with

Chris Roth
UT Austin → Flatiron
PRB **108**, 054410 (2023)

Sylvain Capponi   Fabien Alet
CNRS Toulouse
work in progress

Variational Monte Carlo (VMC)

Computationally, wave functions are a many-variable complex function:

$$|\psi\rangle = \sum_{s_1,\dots,s_N} \psi(s_1,\dots,s_N)|s_1,\dots,s_N\rangle$$

- ED: $\psi(s_1,\dots,s_N)$ comes from a lookup table
- Hartree–Fock: $\psi(s_1,\dots,s_N) = N \times N$ determinant
- MPS: $\psi(s_1 \dots s_N) = \mathrm{tr}[A_1(s_1)\dots A_N(s_N)]$

Variational Monte Carlo (VMC)

A wave-function *ansatz* has additional parameters θ:

$$|\psi(\theta_1, \dots, \theta_P)\rangle = \sum_{s_1,\dots,s_N} \psi(s_1, \dots, s_N; \theta_1, \dots, \theta_P)|s_1, \dots, s_N\rangle$$

- Hartree–Fock: $\theta \equiv$ single-particle orbitals
- MPS: $\theta \equiv$ matrix entries

Variational Monte Carlo is a set of algorithms to optimise θ for *any* ansatz

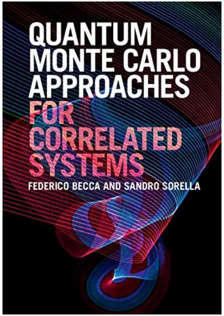
Variational Monte Carlo (VMC)

A wave-function *ansatz* has additional parameters θ:

$$|\psi(\theta_1, \dots, \theta_P)\rangle = \sum_{s_1,\dots,s_N} \psi(s_1, \dots, s_N; \theta_1, \dots, \theta_P)|s_1, \dots, s_N\rangle$$

- Hartree–Fock: $\theta \equiv$ single-particle orbitals
- MPS: $\theta \equiv$ matrix entries

Variational Monte Carlo is a set of algorithms to optimise θ for *any* ansatz

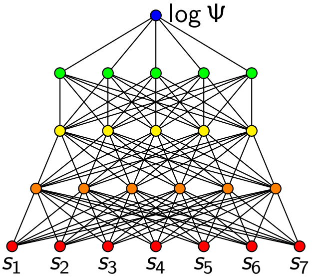


Neural quantum states (NQS)

Would like a very flexible and expressive variational ansatz

- Deep neural networks can represent images, language, etc.
- Try them as an ansatz!
  - Input of NN: basis state $s_1, \ldots, s_N$
  - Output: $\psi(s_1, \ldots, s_N)$
  - VMC optimises the layers of the network
- Many different kinds of neural networks

$\Rightarrow$ Pick the best one for the application!



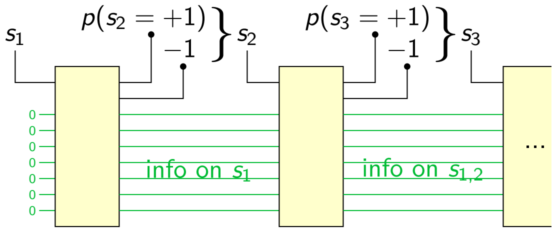
Example: recurrent neural networks

Used to generate sequential data (e.g. text)

- Probability of sequence $(s_1, s_2, \dots)$ as a string of conditional probabilities:

$$p(s_1, s_2, \dots) = p(s_1)p(s_2|s_1)p(s_3|s_1, s_2)\cdots$$

- Each $p(s_n)$ from the same NN
- Extra variables to compress $s_{i<n}$



Can also represent wave functions:

- $\psi(s_1, \dots) = \sqrt{p(s_1, \dots)}$
- Direct sampling without Markov chains

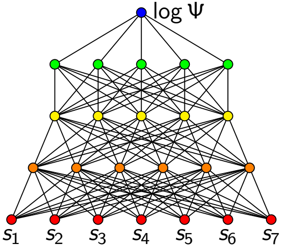
$\Rightarrow$ Promising for spin-glass problems

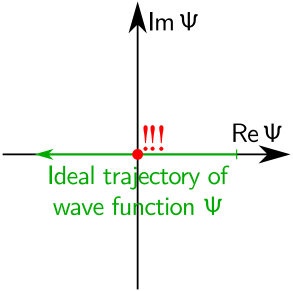
Hibat-Allah *et al.*, Nat. Mach. Intell. **3,** 952 (2021)

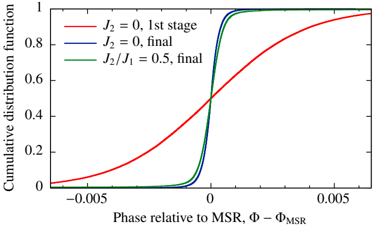
The sign problem in NQS

What if $\psi(s_1, \dots)$ aren't all positive?

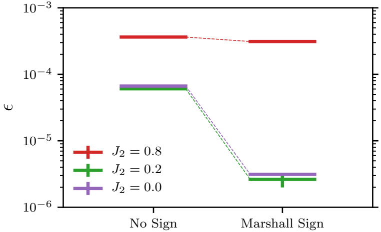
- **First idea:** $\psi(s_1, \dots, s_N) = A(s_1, \dots, s_N) e^{i\phi(s_1,\dots,s_N)}$
  - sign flip $\Longleftrightarrow$ $i\pi$-jump
  - destructive interference hard to follow
  - nearby "simple" sign structures easier to learn







ASz & Castelnovo, PRR (2020)



Hibat-Allah *et al.*, PRR (2020)

Symmetries in VMC

- Wigner's theorem: eigenstates have well-defined symmetry quantum numbers
  - restricting on symmetry sectors makes calculations more efficient
  - can look for low-lying excited states in different sectors
- Local symmetries (e.g., spin rotation)
  - can often encode directly in tensor networks
  - some variational ansätze respect it, but no general recipe
  - can enforce some symmetries with sampling (e.g., fixed $\sum \sigma^z$)
- Geometric symmetries (e.g., translation)
  - maps computational basis states on one another

  $\Rightarrow$ easy to represent in VMC!

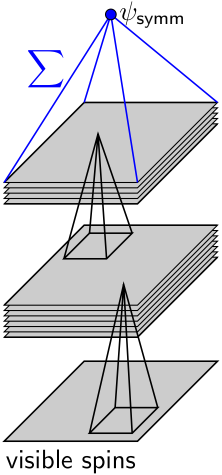
Convolutional neural networks (CNNs)

- Naïve enforcement of translation symmetry: take any ansatz and symmetrise it

$$\psi_{\text{symm}}(\boldsymbol{\sigma}) = \sum_{\mathbf{r}} \psi(\mathcal{T}_{\mathbf{r}}\boldsymbol{\sigma})$$

- This is computationally expensive: use *convolutional* networks
  - each layer is a replica of the input geometry
  - mapping is *equivariant:* translated input $\Longrightarrow$ translated output
  - summing all outputs in last layer is fully symmetric
- Can also represent nonzero wave vectors:

$$\psi_{\mathbf{k}}(\boldsymbol{\sigma}) = \sum_{\mathbf{r}} e^{-i\mathbf{k}\cdot\mathbf{r}} \psi(\mathcal{T}_{\mathbf{r}}\boldsymbol{\sigma})$$

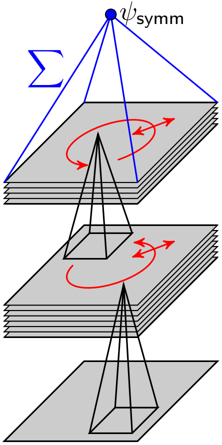


Group convolutional neural networks (GCNNs)

- Generalise the idea to nonabelian symmetry groups
  Cohen & Welling, ICML 2016
  - embedding layer: lattice → symmetry group
  - group convolutional layers: group → group
  - equivariance: $\sigma'(\mathbf{r}) = \sigma(g\mathbf{r}) \implies f'(h) = f(gh)$ in all layers
- Symmetric ψ from summing all entries:

$$\psi_0 = \sum_{i,g} e^{f_i(g)}; \qquad \psi_\chi = \sum_{i,g} \chi_g^* e^{f_i(g)}$$

- Implemented in NQS library NetKet



(G)CNNs vs. the sign problem

Wave function is a sum of terms

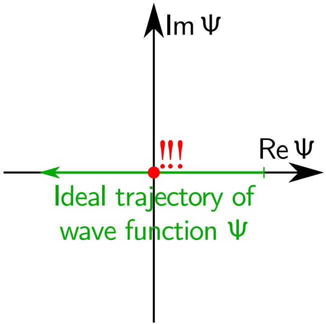
- $\psi = 0$ is no longer special
- Destructive interference: cancellation of (a few) terms
- Much better at learning complicated sign structures

Sum of terms structure seems critical to get nontrivial signs

Nomura, JPCM **33,** 174003 (2021)

Reh, Schmitt, Gärttner, PRB **107,** 195115 (2023)...

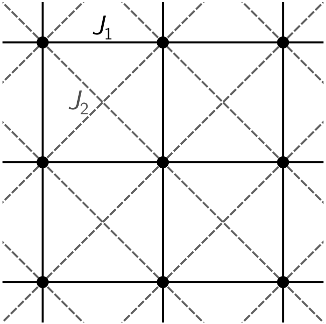
"Get it for free" with GCNNs!



Square-lattice $J_1 - J_2$ model

$$H = J_1 \sum_{\langle ij \rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j \qquad (J_1, J_2 \geq 0)$$

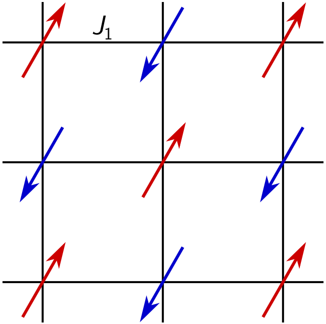
- J_1 only: unfrustrated Néel order
- J_2 only: unfrustrated stripy order
- $J_1 - J_2$ model: frustrated, many phases
  - Néel (stripy) for J_1 (J_2) dominant
  - in between: valence bond solid, spin liquid
- We focus on $J_2/J_1 = \begin{cases} 0.5 & \text{(spin liquid)} \\ 0.55 & \text{(VBS)} \end{cases}$



Square-lattice $J_1 - J_2$ model

$$H = J_1 \sum_{\langle ij \rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j \qquad (J_1, J_2 \geq 0)$$

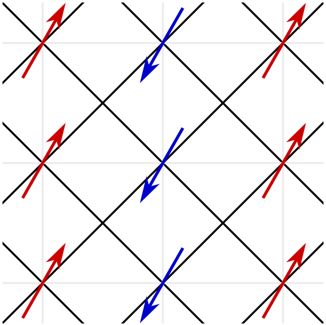
- J_1 only: unfrustrated Néel order
- J_2 only: unfrustrated stripy order
- $J_1 - J_2$ model: frustrated, many phases
  - Néel (stripy) for J_1 (J_2) dominant
  - in between: valence bond solid, spin liquid
- We focus on $J_2/J_1 = \begin{cases} 0.5 & \text{(spin liquid)} \\ 0.55 & \text{(VBS)} \end{cases}$



Square-lattice $J_1 - J_2$ model

$$H = J_1 \sum_{\langle ij \rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j \qquad (J_1, J_2 \geq 0)$$

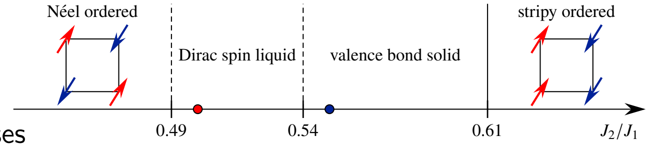
- J_1 only: unfrustrated Néel order
- J_2 only: unfrustrated stripy order
- $J_1 - J_2$ model: frustrated, many phases
  - Néel (stripy) for J_1 (J_2) dominant
  - in between: valence bond solid, spin liquid
- We focus on $J_2/J_1 = \begin{cases} 0.5 & \text{(spin liquid)} \\ 0.55 & \text{(VBS)} \end{cases}$



Square-lattice $J_1 - J_2$ model

$$H = J_1 \sum_{\langle ij \rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j \qquad (J_1, J_2 \geq 0)$$

- J_1 only: unfrustrated Néel order
- J_2 only: unfrustrated stripy order
- $J_1 - J_2$ model: frustrated, many phases
  - Néel (stripy) for J_1 (J_2) dominant
  - in between: valence bond solid, spin liquid

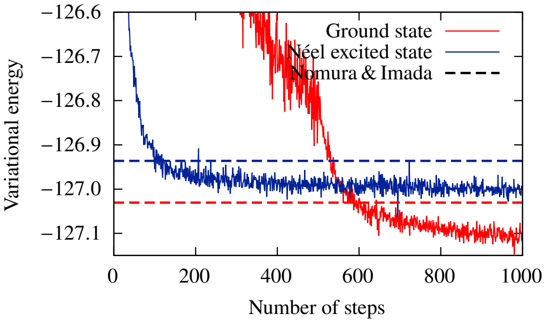


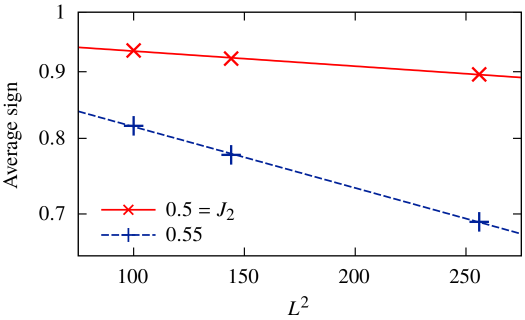
Nomura & Imada, PRX (2021)

- We focus on $J_2/J_1 = \begin{cases} 0.5 & \text{(spin liquid)} \\ 0.55 & \text{(VBS)} \end{cases}$

Ground states

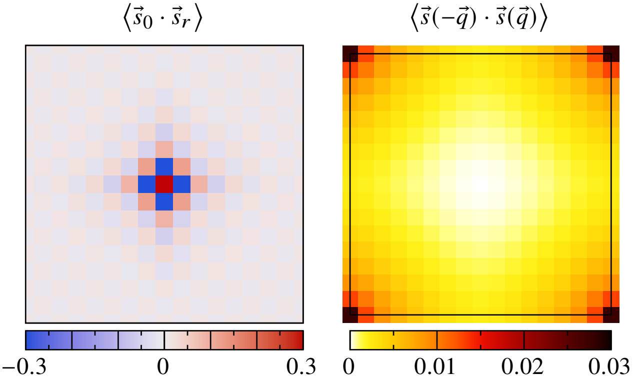
- Used GCNNs with L layers for $L \times L$ systems ($L = 10, 12, 16$)
- VMC converges well in ~ 100 GPU hours
  - best variational energies in the literature
  - exponentially decaying average sign ✓





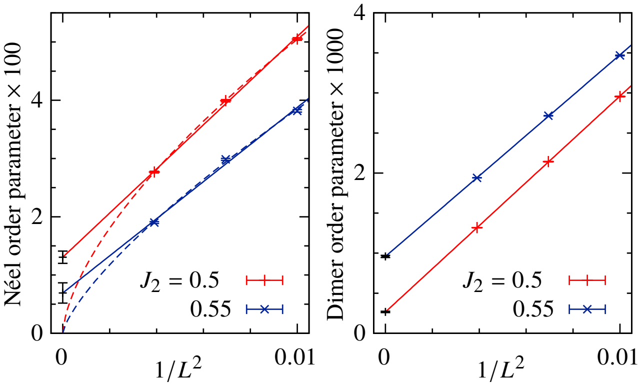
Correlations

- Spin correlators as expected
- Dimer correlators as covariances of $\vec{s}_i \cdot \vec{s}_j$ estimators
- Order parameters:
  - Néel $\sim L^{-z} \implies$ no order ✓
  - VBS $\sim D_\infty + \alpha L^{-2}$
  - D_∞ very small for $J_2 = 0.5$
  - but clearly finite for 0.55 ✓



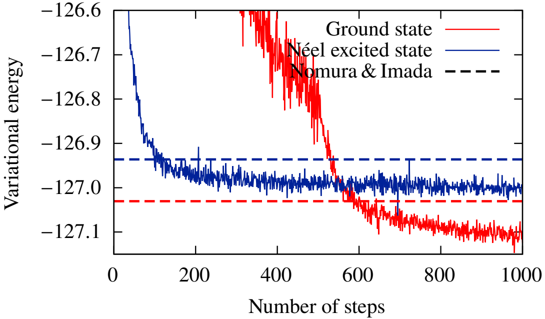
Correlations

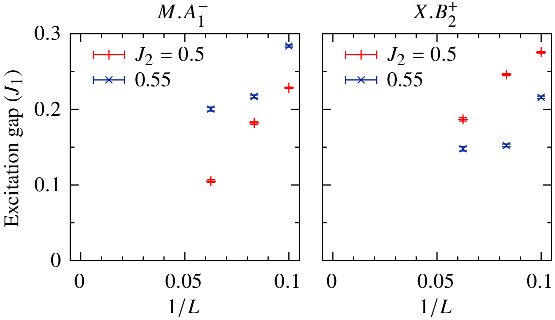
- Spin correlators as expected
- Dimer correlators as covariances of $\vec{s}_i \cdot \vec{s}_j$ estimators
- Order parameters:
  - Néel $\sim L^{-z} \implies$ no order ✓
  - VBS $\sim D_\infty + \alpha L^{-2}$
  - D_∞ very small for $J_2 = 0.5$
  - but clearly finite for 0.55 ✓



Excited states

- Restrict GCNN to nontrivial space-group irrep: $\psi_\chi = \sum_{i,g} \chi_g^* e^{f_i(g)}$
- VMC returns lowest energy in symmetry sector $\implies$ symmetry gaps
- Considered (π, π) triplet (Néel order), $(\pi, 0)$ singlet (VBS)
  - gaps similar to best estimates in literature
  - not accurate enough for extrapolation



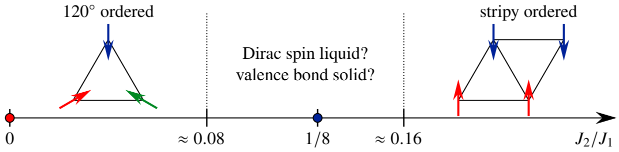


Triangular-lattice $J_1 - J_2$ model

$$H = J_1 \sum_{\langle ij \rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j \qquad (J_1, J_2 \geq 0)$$

Frustrated at every $J_1, J_2 \geq 0$

- We focus on $J_2/J_1 = \begin{cases} 0 & (120^\circ \text{ order}) \\ 1/8 & (\text{paramagnet}) \end{cases}$
- Requires deeper networks to converge
  - use residual networks for stability
- Converged energies less accurate
  - "Lanczos step:" $|\psi\rangle \mapsto |\psi\rangle + \alpha H|\psi\rangle$
    Sorella, PRB **64,** 024512 (2001)

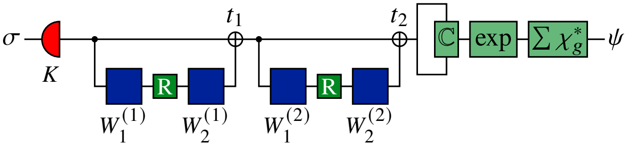


Triangular-lattice $J_1 - J_2$ model

$$H = J_1 \sum_{\langle ij \rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j \qquad (J_1, J_2 \geq 0)$$

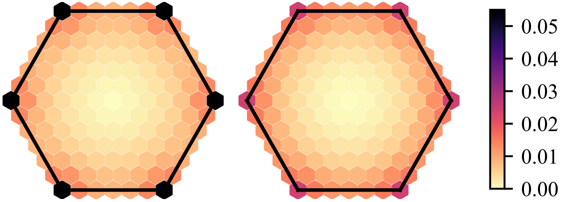
Frustrated at every $J_1, J_2 \geq 0$

- We focus on $J_2/J_1 = \begin{cases} 0 & (120^\circ \text{ order}) \\ 1/8 & (\text{paramagnet}) \end{cases}$
- Requires deeper networks to converge
  - use residual networks for stability
- Converged energies less accurate
  - "Lanczos step:" $|\psi\rangle \mapsto |\psi\rangle + \alpha H |\psi\rangle$
    Sorella, PRB **64,** 024512 (2001)



Triangular-lattice $J_1 - J_2$ model – results

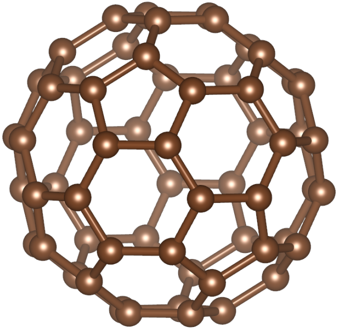
- Best known variational energies for $108, 144$ sites
- Recover Bragg peak at $J_2 = 0$
- Outlook:
    - larger systems
    - accurate order parameters
    - accurate gap estimates



C_{60} Heisenberg model

$$H = \sum_{\langle ij \rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j$$

- "Buckyball" geometry
  - 12 pentagon, 20 hexagon faces
  - I_h point group (120 symmetries)
- 8-layer GCNNs (3-layer performs well too)
  - Ground state: singlet, trivial irrep
  - First excited state: triplet, T_{2g}
  - Lowest $S = 2$: $\mathrm{A}_g \approx \mathrm{H}_g$
  - Energies match DMRG
    Rausch et al., SciPost Phys. 10, 087 (2021)

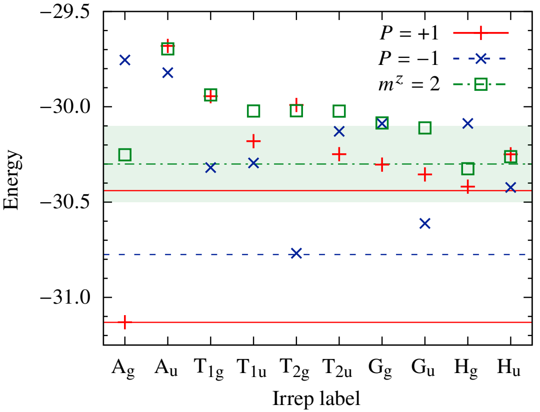


C_{60} Heisenberg model

$$H = \sum_{\langle ij \rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j$$

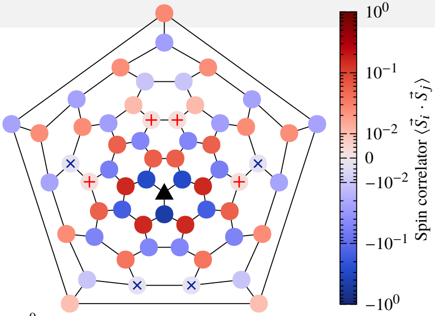
- "Buckyball" geometry
  - 12 pentagon, 20 hexagon faces
  - I_h point group (120 symmetries)
- 8-layer GCNNs (3-layer performs well too)
  - Ground state: singlet, trivial irrep
  - First excited state: triplet, T_{2g}
  - Lowest $S = 2$: $A_g \approx H_g$
  - Energies match DMRG

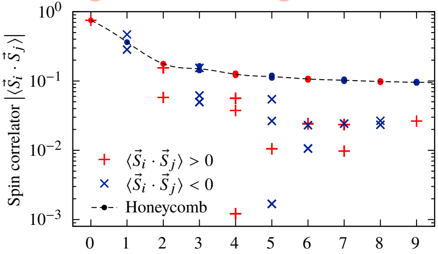
    Rausch et al., SciPost Phys. 10, 087 (2021)



Precursors to ordering

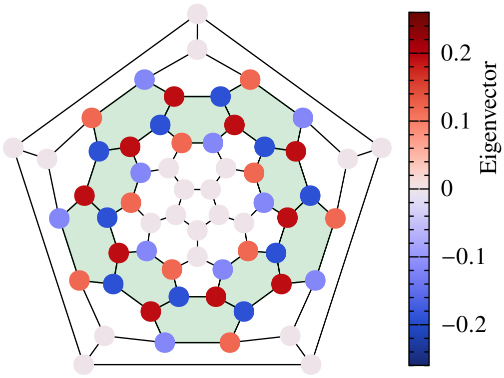
- Correlations level off at $\approx \pm 0.03$
- Analogies to honeycomb Néel order:
  - "Bragg peak:" leading eigenvector v of correlation matrix $\sim T_{2g}$
  - Néel pattern on unfrustrated component
  - "Magnon operator:" $\vec{S}_v = \sum_i v_i \vec{\sigma}_i$
    $\vec{S}_v|\text{GS}\rangle$ has 92% overlap with lowest triplet
  - "Tower of states:" $S = 2$ states consistent with applying $\vec{S}_v$ twice





Precursors to ordering

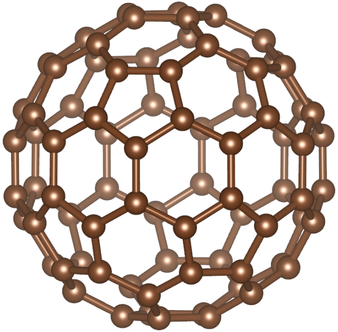
- Correlations level off at $\approx \pm 0.03$
- Analogies to honeycomb Néel order:
  - "Bragg peak:" leading eigenvector v of correlation matrix $\sim T_{2g}$
  - Néel pattern on unfrustrated component
  - "Magnon operator:" $\vec{S}_v = \sum_i v_i \vec{\sigma}_i$ $\vec{S}_v|\mathrm{GS}\rangle$ has 92% overlap with lowest triplet
  - "Tower of states:" $S = 2$ states consistent with applying $\vec{S}_v$ twice



C_{80} Heisenberg model

$$H = \sum_{\langle ij \rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j$$

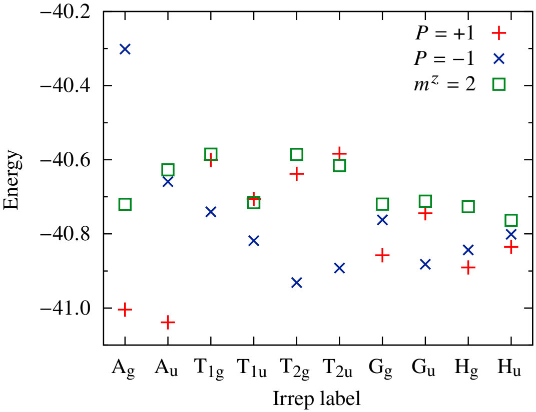
- Next smallest I_h-symmetric geometry
  - 12 pentagon, 30 hexagon faces
- Near-degenerate lowest-energy pairs
  - $S = 0$: $\mathrm{A_u} \lesssim \mathrm{A_g}$
  - $S = 1$: $\mathrm{T_{2g}} \lesssim \mathrm{T_{2u}}$
  - $S = 2$: $\mathrm{H_u} \lesssim \mathrm{H_g}$
- Inversion-symmetry breaking?



C_{80} Heisenberg model

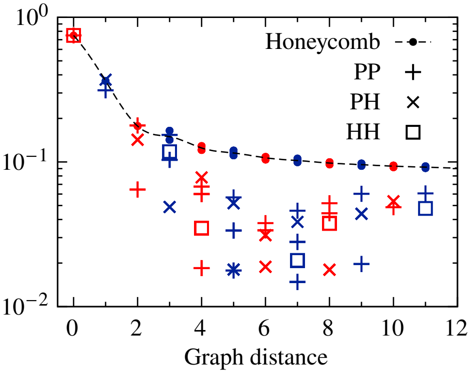
$$H = \sum_{\langle ij \rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j$$

- Next smallest I_h-symmetric geometry
  - 12 pentagon, 30 hexagon faces
- Near-degenerate lowest-energy pairs
  - $S = 0$: $\mathrm{A_u} \lesssim \mathrm{A_g}$
  - $S = 1$: $\mathrm{T_{2g}} \lesssim \mathrm{T_{2u}}$
  - $S = 2$: $\mathrm{H_u} \lesssim \mathrm{H_g}$
- Inversion-symmetry breaking?



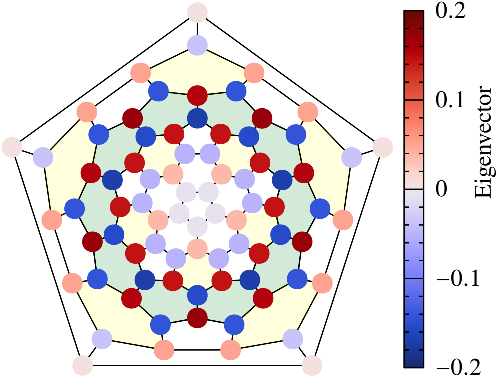
Precursors to chiral ordering

- Correlations level off at $\approx \pm 0.06$
  - higher than C_{60}: lower frustration
- "Bragg peak" eigenvector of similar structure
  - transforms under T_{2u} irrep
  - threefold degenerate $\implies$ three "magnon" operators $\vec{S}_{v_\alpha} = \sum_i v_i^{(\alpha)} \vec{\sigma}_i$
  - $\vec{S}_{v_1} \cdot (\vec{S}_{v_2} \times \vec{S}_{v_3})$ is an A_u singlet
  - analogous to tetrahedral order in triangular-lattice $J_1 - J_2 - J_\chi$ model
  - do the inversion-broken ground states break time reversal too?
- Future work: study larger molecules



Precursors to chiral ordering

- Correlations level off at $\approx \pm 0.06$
  - higher than C_{60}: lower frustration
- "Bragg peak" eigenvector of similar structure
  - transforms under T_{2u} irrep
  - threefold degenerate $\implies$ three "magnon" operators $\vec{S}_{v_\alpha} = \sum_i v_i^{(\alpha)} \vec{\sigma}_i$
  - $\vec{S}_{v_1} \cdot (\vec{S}_{v_2} \times \vec{S}_{v_3})$ is an A_u singlet
  - analogous to tetrahedral order in triangular-lattice $J_1 - J_2 - J_\chi$ model
  - do the inversion-broken ground states break time reversal too?
- Future work: study larger molecules



Summary

- Neural quantum states for frustrated magnets
  - Do not separate amplitudes and phases!
  - Group convolutional networks: enforce spatial symmetries efficiently
- $J_1 - J_2$ Heisenberg antiferromagnet
  - Square lattice: state-of-the-art energies, accurate order parameters
  - Triangular lattice: more expensive, need larger system sizes
- Heisenberg antiferromagnets on fullerene geometries
  - Low-lying spectrum explained by emerging Néel-like order
  - On some geometries, may get inversion and/or time-reversal breaking!

Outlook

- Need more accurate results for large systems and excited states
- Physically motivated ansätze
  - e.g., Gutzwiller-projected parton variational states
  - use neural networks for more flexible multi-particle orbitals
  - successfully used in quantum chemistry: FermiNet, PauliNet,...
    Pfau et al., Phys. Rev. Research **2,** 033429 (2020)
    Hermann et al., Nat. Chem. **12,** 891 (2020)
  - need to scale them to many more particles (on a lattice)