- The Ising Model & Quantum Mechanical Ground State
  - Fixed spins rotate (magnet needle of a compass)
  - Interact with each other, but also influenced by external fields
  - Schrödinger equation
  - Calculate and find lowest energy: lower equation is derived how to calculate the energy observable (expectation value) in our case
  - Linear chain solution is obvious, for triangle already for three sites non-trivial
  - We have: triangular, hexagonal, linear, square lattices. With or without periodicity
- Solution through exact Diagonalization
  - Same Problem as before, but new Hamiltonian -> orthogonal external field
  - Canonical spin base system
  - Hamiltonian can be evaluated to matrix 2^n time, solving for eigenvalues even harder
- Solutions with neural Quantum States
  - Example for a denser Parametrization: positions of out-of-line spins, in reality more complicated
  - Iterative methods require not solving the whole equation, but approximating the solutions to a very high degree
  - Monte Carlo (random sampling and evaluating).
  - We do variational monte Carlo, diffusion monte Carlo is alternative method also described in the thesis. Based on imaginary time evolution
- New Idea: Why Graphs should help
  - Probably custom encoding not available for each problem
  - Universal representation, no need for custom encoding
  - Non needed connections are masked automatically (explain example)
- Transformers: short overview
  - Motivated by translation
  - Attention pictured, but not enough time to explain. Ask each site, what information about each other site they want, then incorporate as weighted sum
  - Expensive, but proven to work and effective
- From the Transformer to the Metaformer
  - First explain the modifications done to the transformer: Input embedding different (positional, not mentioned later but important), head not classifier token, but pooling head
  - Replace the elements attention, convolution, pooling
  - Explain, there are different pooling/convolution operations, as well as the graph versions. Convolution comes now
- Depthwise Convolutions and Separable Convolutions
  - Idea out of the mobile net paper
  - Expensive large kernels get replaced by independent, separated kernels.
  - Depthwise convolution sufficient for us, as recombination is done by the MLP
- Now where exactly come Graphs into play?
  - Already established operators / networks can be used on graph data
  - Example for pooling
  - Augmentation of the attention operation to only attend to relevant neighbors
- Image Classification
  - Stress, that we now talk about results of the image processing portion
  - Same hyperparameters were used for all models!
- Conventional Convolutions and Graph Convolutions
  - Show the theory: on tensor data, convolutions and graph convolutions are equivalent
  - Experiment confirms this
  - Difference is caused by the different use of the deterministic random number generator
- Comparison of different Token Mixers: Visualized
  - Accuracy & Convergence: Transformer -> Graph Transformer -> Conformer -> Graph Conformer -> Poolformer

- Comparison of different Token Mixers: Quantized
    - Explain metrics briefly: lower is always better except for accuracy.
    - Poolformer most efficient and fast. Then Conformer -> show trend is the other way around
    - Depending on use-case need to choose different framework
    - Graph element may even bring benefits in performance -> inductive bias
- Comparison to Established Architectures
    - Stress, that we now talk about Ground state search again!!
    - Here every model approximately same number of parameters
    - RBM (Restricted Boltzmann Machine) was not able to parametrize wavefunction, all others were
    - Show that there is kind of a trend, but no definite line
    - One SPECIFIC lattice and one SPECIFIC model was chosen
- Resiliency to the Choice of Lattice Encoding
    - Explain, how the structure may influence the memory encoding
    - To break symmetry and provide fair comparison, random shuffeling
    - Worse performance for CNN than for graph network
- Differences across the Phase Diagram
    - Explain what the parameters h and J were again and what lambda is.
    - Motivate why phase transitions would make the problem harder
    - Show that metaformer barely performs different, as state gets more complicated. CNN tanks precision
- Conclusion
    - Compatibility
    - Resiliency encoding
    - Customizability
    - Computational potential (transformer backbone)
    - Stability in phase transition region
- Future Work
    - Employ in related technical fields
    - Better implementation of graph algorithm (in thesis worse, but should be more efficient)
    - Make wavefuntions more complex and see if it utilizes the metaformers potential more