1

Introduction

1.1 Quantum Many-Body Problems

According to legend, chess was invented by Grand Vizier Sissa Ben Dahir and given to King Shirham of India. The king was so pleased with the game that he offered to grant Sissa any request within reason. The Grand Vizier asked the king for one grain of wheat to be placed on the first square of the chessboard, two grains on the second square, four on the third, and so on, doubling the amount each time until all 64 squares were occupied. The king, arithmetically unaware, accepted the request. However, all the wheat of his kingdom was not enough to fulfill this offer because the total number of grains the Grand Vizier asked for was

$$1 + 2 + 4 + 8 + \cdots + 2^{63} = 2^{64} - 1 = 18,446,744,073,709,551,615.$$

This amount of wheat is approximately 80 times what would be produced in one harvest at current yields if all of Earth's arable land could be devoted to wheat. This number is significant because the number of grains grows exponentially as the number of chessboard squares increases. It demonstrates the rapid growth of exponential sequences.

This wheat-and-chessboard problem is just what is encountered in the study of quantum many-body theory, in which the Hilbert space grows exponentially with the system size. It limits the application range of many computational methods, such as exact diagonalization. Considering, for example, a system of N interacting electrons, if each electron has d-degrees of freedom, then the total dimension of the Hilbert space of the system is $M = d^N$. To determine the ground state of the system, one needs to minimize the energy within this M-dimensional space. The maximal M that is feasible with the best computer software and hardware currently available is about 10^9. It means that only $N \sim 30$ electrons can be handled, even if $d = 2$ and no extra information (for example, symmetry) is used.

https://doi.org/10.1017/9781009398671.002 Published online by Cambridge University Press

The exponential growth of the Hilbert space with the system size is one of the most severe problems we face in the theoretical study of quantum many-body systems. As vividly pointed out by the Nobel Laureate Walter Kohn [1], it raises an *exponential wall* that hinders the development of quantum many-body theory. In the early stage of quantum mechanics, soon after the successful interpretation of optical spectra of hydrogen atoms and other simple atoms by the Schrödinger equation, many physicists, including particularly Paul Dirac, declared that chemistry had come to an end because its content was entirely contained in the powerful Schrödinger equation. However, as it was realized later, this equation was far too complex to allow a solution for a system with many electrons.

In the search for fundamental interactions and elementary particles, the philosophy of *reductionism* has played an important role. It assumes that the nature of a complex world can be understood by reducing it to the interactions of its parts and that a complex system is nothing but the sum of its parts and can reduce to the account of individual constituents.

From the end of the nineteenth century, with the fast progress in the experimental exploration of the microscopic quantum world, a great many novel phenomena, including superconductivity, helium superfluid, and quantum Hall effect, were discovered. It turns out that the behavior of large and complex aggregates of particles is difficult to understand in terms of the extrapolation of a few particles. Instead, at each level of complexity, entirely new phenomena appear. For example, two hydrogen atoms can form a hydrogen molecule via the valence bond, and many hydrogen molecules can form liquid hydrogen at low temperature and even solid hydrogen under high pressure. On the other hand, if hydrogen molecules mix with oxygen molecules, they will undergo a chemical reaction to form water molecules. Hydrogen atoms behave differently under different circumstances.

Emergence is the way complex systems and phenomena arise out of relatively simple interactions and is central to the theories of integrative levels and complex systems [2]. The exponential wall is an unavoidable problem in the investigation of emergent phenomena. It is a *non-perturbative problem* and cannot be completely solved by conventional quantum field theory, which is established based on perturbation. A detailed description of quantum many-body systems poses formidable difficulties due to the exponentially large dimension of the associated Hilbert space. Certain approximations have to be taken to solve this problem.

In general, two kinds of approximations are used. One is to take a single particle approximation to convert a many-body problem into a single-particle one. The most commonly used methods in this direction include the self-consistent mean-field theory and the first-principles density functional theory. The other is to take a many-body approximation by selecting a finite many-body basis set to represent a target state or physical quantity defined in an intractably large Hilbert space.

It includes, for example, the configuration interaction, coupled-cluster expansion, quantum Monte Carlo methods, and tensor network renormalization group methods to be introduced in this book.

The *renormalization group* (RG) was initially devised in particle physics. It serves as the primary means for constructing the connections between theories at different length scales and becomes a powerful and efficient tool for exploring the systems where perturbation theory fails. Starting from some microscopic Hamiltonian, certain degrees of freedom are iteratively integrated and accounted for by modifying the original Hamiltonian. The new Hamiltonian contains modified couplings by a priori elimination of degrees of freedom. It is tempting to believe that this "renormalized" effective Hamiltonian captures the essential physics of the system on a narrower energy scale.

This method is rooted in modern physics and has played a crucial role in the fundamental theory of microscopic particles and the theory of condensed matter physics, and continuous phase transitions. In quantum field theory, renormalization was first introduced to cancel infinities by redefining parameters at different energy or momentum scales. In condensed matter and statistical physics, a more general RG framework based on scaling analyses was introduced to explain the universality properties of continuous phase transitions. It was even extended to compute wave functions for quantum lattice models directly.

The idea of RG was first anticipated by Stueckelberg and Peterman in 1951 [3] and reformulated in 1953 [4]. It emerges from the renormalization of field variables under a scale transformation. RG is not a group. It just forms a semigroup since the RG transformation is not reversible. It was dubbed as a group probably because in the 1950s, when the RG concept was first proposed, a doctrine of physics was that our world could be understood in terms of symmetries and their implementation through groups. However, the connection between RG transformation and group structure is highly formal, and its group interpretation is almost useless in practical applications.

The modern RG in terms of flow equations was introduced by Gell-Mann and Low [5] in 1954 and reformulated by Callan [6] and Symanzik [7] in 1970. Their essential assumption was that a renormalizable quantum field theory is scale-invariant, which depends on the parameters but not the scale. Thus under the scale transformation, the effective theory makes a self-similar replica of itself, with tiny changes in coupling constants determined by the flow equations.

The application of RG has achieved great success in studying quantum field theory. For example, a renormalization scheme developed by Feynman, Schwinger, and Tomonaga led them to solve the ubiquitous problem of infinities by expressing physical observables in terms of parameters. Their theory was spectacularly successful in quantum electrodynamics. They received the Nobel Prize in 1965.

In the early 1970s, 't Hooft and Veltman showed that Yang-Mills gauge theories are renormalizable, for which they shared the 1999 Nobel Prize. In 1973, Gross, Politzer, and Wilczek found that the beta function describing the renormalization flow equation of the strong interaction is negative and discovered the asymptotic freedom in quantum chromodynamics. They won the Nobel Prize in 2004.

In the second half of the 1960s, ideas to recursively generate flows of coupling constants arose independently in condensed matter physics, which led to a deeper understanding of the physical meaning of RG. In 1966, Kadanoff [8] proposed a *block-spin* idea to define the interactions at large distances as aggregates of components at shorter distances. A block spin transformation consists of the scale and the block spin transformations. This block spin approach, together with many vital contributions of Kenneth Wilson, laid the foundation of the scaling theory of second-order phase transitions and critical phenomena. The success of their approach rests on scale separation. At a critical point, the correlation length diverges and short-range fluctuations can affect long-wavelength behavior quantitatively but not qualitatively.

In the first half of the 1970s, Wilson began to apply the RG techniques to problems not pertaining to a critical point or the computation of ground-state wave functions for quantum systems. His work pioneered the idea of RG and allowed it to be implemented for any physical system, even away from critical points where the scaling invariance breaks. It extended the field of RG from the simple flow equations of coupling parameters to the whole Hilbert space and opened the field of the *numerical renormalization group* (NRG) [9]. He demonstrated the potential of this powerful method by constructing a successive RG solution of the famous Kondo impurity model, which describes a single magnetic impurity in a nonmagnetic metal. Wilson was awarded the Nobel prize partly for this contribution in 1982.

The success of Wilson's method in the single magnetic impurity problem relies on two peculiar features of the Kondo model. First, the width of the Kondo resonance sets an energy scale such that the contributions of energy levels far from the resonance can be integrated out. Second, the Kondo Hamiltonian can always be mapped onto an effective one-dimensional model since the Kondo interaction couples only with the *s*-wave part of the electron wave function. It dramatically lowers the barrier to solving this problem.

There are two ways in which NRG differs from the conventional (analytic) RG: First, in the conventional RG of quantum field theory, it is the charge, mass, interacting coupling constants, and a few other physical parameters renormalized during the reduction of energy scales. In contrast, in NRG, it is the wave function of a quantum state or the partition function of a Hamiltonian that is calculated using the RG transformation. A wave function or partition function contains the information

needed to evaluate all physical observables. Second, in the conventional RG calculation, the scaling invariance is assumed to keep the formula of interacting potentials unchanged. This assumption is no longer needed in the implementation of NRG.

1.2 From NRG to DMRG

Shortly after Wilson's dramatic success in solving the Kondo problem, there was considerable excitement about the possibility of applying the real-space NRG together with the Kadanoff blocking spin scheme [8] to a variety of quantum lattice problems with lattice sites replacing energy levels [10, 11, 12, 13]. The basic idea was to take Kadanoff's block spins as building cells and aggregate every two cells into a new block at each RG transformation. It was hoped that the ground-state properties of other many-body systems, for example, the Hubbard model [14], could be solved by this approach. However, the performance of Wilson's NRG was poor in treating these many-body models. For example, the error of the ground-state energy obtained with this method for the one-dimensional Hubbard model with 16 sites by keeping about 1,000 basis states is about 5–10% [15].

The essence of Wilson's NRG is to select a small set of basis states to represent the ground state or other targeted states of a Hamiltonian through successive local basis transformations. In the NRG calculations, the basis space is not fixed, unlike in the exact diagonalization. Instead, it changes dynamically at each step of RG iterations. The truncation error is determined by the criterion used in the basis space truncation. Of course, an optimal RG scheme should use a criterion that minimizes the error in the basis truncation. In addition, the number of basis states retained is a key parameter that can affect the truncation error – the more basis states retained, the smaller the truncation error.

A basic assumption of Wilson's NRG is that the ground state of a large system is determined by the low-lying excitations of the building blocks, whose weight is determined by the *thermal density matrix*

$$\rho = e^{-\beta H} \tag{1.1}$$

up to a normalization constant. Here, $\beta = 1/k_B T$, k_B is the Boltzmann constant, and T is temperature. Using the eigenvalues and eigenstates of the Hamiltonian

$$H|\Psi_n\rangle = E_n|\Psi_n\rangle, \tag{1.2}$$

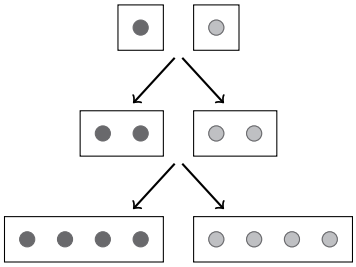
the density matrix can also be expressed as

$$\rho = \sum_n e^{-\beta E_n}|\Psi_n\rangle\langle\Psi_n|. \tag{1.3}$$

https://doi.org/10.1017/9781009398671.002 Published online by Cambridge University Press

The weight of the eigenstate $|\Psi_n\rangle$ in the partition function is proportional to $\exp(-\beta E_n)$. Thus, the lower the energy, the larger the weight. This criterion is not optimal because some highly excited states in a small system may lower their energies and become increasingly important with the increase of the system size. However, in Wilson's NRG, energy is the only parameter used for judging whether a state is retained or discarded.

The block spin NRG starts by considering a block B where the Hamiltonian H_B and the operators at the two ends of the blocks are defined. Then a new block of the same size is added to the system, which augments the Hamiltonian to H_{BB}. After diagonalizing this Hamiltonian, the enlarged block is replaced by a new effective block B_{new}, formed by D lowest eigenstates of H_{BB}, and the iteration continues, where D is the number of basis states retained. A schematic representation of these NRG iteration steps is



At each step, a block (left) is added with a duplicated block (right) to form an augmented block after the basis truncation.

This block spin NRG scheme was applied to the one-dimensional Hubbard model by Bray and Chui [10] in 1979, and later to the one-dimensional Heisenberg model by Pan and Chen [11], and by Kovarik [12]. Unlike the Kondo impurity model, their results were not that encouraging. The reason for this lies in the physical difference between the Kondo problem and the translationally invariant quantum lattice models. The most significant difference between the Kondo system and a one-dimensional lattice model is that the couplings between adjacent sites decrease exponentially in the Kondo system under Wilson's logarithmic discretization scheme. In contrast, it remains constant in the Hubbard or Heisenberg model. This exponential decrease is the key to the success of the method for the Kondo impurity systems. However, it is flawed in treating quantum lattice problems in a real-space blocking form.

The failure of Wilson's block spin NRG, as pointed out by White and Noack [16], results from the boundary or interface effect between blocks. This can be understood by considering a toy model – a single particle on a tight-binding chain. For this model, the ground state is a standing wave that vanishes at the two ends of each block. As an example, Fig. 1.1 shows how the ground-state wave function

https://doi.org/10.1017/9781009398671.002 Published online by Cambridge University Press

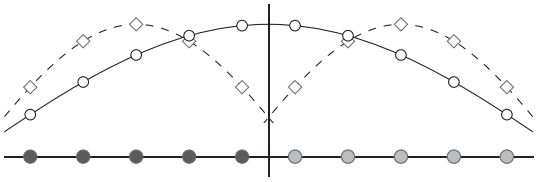


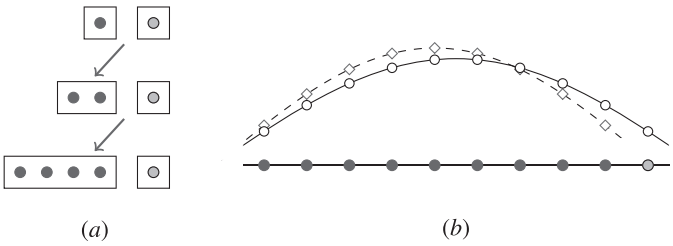
Figure 1.1 Ground-state wave functions of a particle hopping on a tight-binding chain. Two five-site blocks form a large block of ten sites. The ground state is a standing wave on each block. The wave function from the ground state of the two five-site blocks (diamonds) vanishes at the interface where the ground-state wave function of the ten-site block (circles) shows a maximum.

changes when a ten-site block is formed by combining two five-site blocks. The ground state is a standing wave on each block. The wave function takes a minimum at the interface of two small blocks. By combining two blocks into a larger one, the lowest-lying states of each block have nodes at the compound block center. On the contrary, the ground state of the compound block shows a maximum there. It suggests that the basis states of a larger block cannot be accurately approximated by a restricted number of states of smaller blocks without adequately considering the interface effect.

In 1992, Xiang and Gehring showed that Wilson's NRG could be improved by adding just one spin, instead of a block spin, at each iteration [13]. Their approach reduces both the truncation error and the interface effect and improves the accuracy of the ground-state energy of the one-dimensional Heisenberg model by a few orders of magnitude [13, 17]. By keeping about 200 states, for example, they obtained an error of about 0.5% in the ground-state energy. Their approach is similar to that used by Wilson for treating a single-impurity Kondo problem. The difference is that in the Kondo impurity problem, there are distinct energy scales to separate each lattice point in the effective Hamiltonian, and it is natural to add one site at each step of RG transformation.

This simple approach has two advantages compared to the block spin NRG. First, as shown in Fig. 1.2, this approach reduces the boundary error encountered in the double-blocking NRG scheme. Second, in the conventional NRG algorithm, the total number of basis states is squared by combining two blocks into a larger block. Namely, if the number of states at each block is D, then the total number of states is D^2. After truncation, only D of them are retained. However, in Xiang and Gehring's approach [13], the total number of states is dD with d the number of basis states at each lattice site. After truncation, the number of states retained is still D, but its percentage in the total number of states is much higher than in the former case since $d \ll D$. It significantly reduces the truncation error. Furthermore, because the total number of states in this improved NRG scheme is just dD, a larger D can be used in the RG calculation, which can also improve the accuracy.

https://doi.org/10.1017/9781009398671.002 Published online by Cambridge University Press



(*a*) (*b*)

Figure 1.2 (a) Illustration of the one-site growing scheme. After obtaining the new block from the previous step, a new site is added. The size of the block grows by one at each iteration step. (b) The lowest-energy states for a particle on a tight-binding chain. A ten-site block is obtained by adding the rightmost site to a nine-site block.

In 1992, White invented the *density matrix renormalization group* (DMRG) [18]. This iterative and variational method relies on an exact diagonalization of the Hamiltonian. The premise is to obtain a wave function in a reduced Hilbert space that approximates the actual ground state, minimizing the loss of information. It was originally developed for studying one-dimensional quantum lattice models in real space at zero temperature. But its core idea to construct the RG flow in terms of *reduced density matrices* works very generally. It yields an optimal scheme to truncate Hilbert space and a powerful approach for evaluating static, dynamical, and thermodynamic properties of low-dimensional quantum many-body systems without introducing any external bias. Its applicability has now been extended successfully to statistical mechanics, quantum chemistry, nuclear and high-energy physics, quantum information, machine learning, and other fields.

The idea of making a systematic approximation by truncating the basis space is not new. It is also the idea used in NRG and Monte Carlo simulations. Unfortunately, this truncation is often not very effective, especially in strongly correlated electronic systems. In DMRG, however, the basis states are first rotated to reduce the truncation error so that the ground state could be accurately represented just by a small set of basis states.

The DMRG technique splits a system into two blocks, called left (L) and right (R) blocks, and two sites that are often placed between these two blocks. These two blocks need not be of equal size. A set of representative states is retained for each block during the warmup. This system of left block plus two sites plus right block, abbreviated as ($L \bullet \bullet R$), is defined as a superblock. We call the left block plus the left-added site as the *system* block, denoted as $S = L\bullet$, and the right-added site plus the right block as the *environment* block, denoted as $E = \bullet R$. At each step of the iteration, the ground-state wave function of the superblock is calculated by diagonalizing the Hamiltonian. The reduced density matrix for the system block is then evaluated from the density matrix of the ground state by tracing out all the

https://doi.org/10.1017/9781009398671.002 Published online by Cambridge University Press

degrees of freedom in the environment. The eigenvalues of the reduced density matrix determine the probabilities of the corresponding eigenvectors in the ground state. These eigenvectors form a new set of basis states of the system block, which are now truncated according to their probabilities. In this scheme, the system block is augmented by just adding one site at each iteration, similar to the conventional NRG scheme of Xiang and Gehring [13].

Recently, DMRG has become one of the most reliable and versatile methods developed in modern computational physics. In particular, the real-space DMRG is one of the most accurate and efficient methods for studying quantum lattice models with short-range interactions in one dimension. It can treat large systems with controllable precision. In particular, in many one-dimensional systems, the error can be reduced to the level at which the DMRG results can be regarded as *quasi-exact* if the ground state has a finite excitation gap. A striking illustration of the precision of this method was first given by White and Huse [19]. They showed, by taking the $S = 1$ antiferromagnetic Heisenberg spin chain as an example, that precision of 10^{-12} for the ground-state energy, $E_0 = -1.401484038971(4)$ in the unit of the exchange constant, could be achieved just using modest computational resources.

1.3 From DMRG to Tensor Network Algorithms

Since the invention of DMRG, it has been successfully applied to study ground-state properties of one-dimensional and quasi-one-dimensional quantum lattice systems. Meanwhile, several methods have been developed to extend DMRG from zero temperature to finite temperatures, from local to nonlocal basis space, and from the calculation of static quantities to time-dependent or dynamical correlation functions. Such progress has extended the application scope of DMRG and stimulated the development of a new breed of algorithms based on the so-called *tensor network states*. These algorithms treat quantum states as products of interconnected tensors and offer new tools for probing quantum many-body systems.

Matrix Product State (MPS)

In 1995, Östlund and Rommer [20] pointed out that the wave function generated by DMRG is an MPS, and the success of DMRG is related to the fact that it is a variational method within the MPS space. They also pointed out that an MPS can be viewed as a variational wave function that can be optimized without invoking DMRG [20, 21]. Their work sheds light on the understanding of DMRG. It offers not just a systematical way to parameterize quantum many-body states but also a new route to extend the DMRG algorithm, especially to higher dimensions.

An MPS provides a highly versatile parametrization for the ground states of local Hamiltonians [22, 23]. It is a collection of so-called local tensors defined on each lattice site. A local tensor contains a dangling physical bond, whose dimension is just the total number of basis states at each lattice site, and two virtual bonds, whose dimensions equal the number of states retained at the DMRG calculation. The local tensors are connected by virtual bonds. The number of parameters required to specify these tensors grows linearly with the system size, significantly smaller than the exponentially large dimension of the Hilbert space.

The matrix product representation of quantum states is not new. It has been introduced under various names over the past five decades. The concept of MPS was first introduced in classical statistical physics by Baxter [24], and later on in the context of quantum Markov chains [25]. In 1987, Affleck, Kennedy, Lieb, and Tasaki (AKLT) proposed an extended Heisenberg model and showed that its ground state is a valence bond solid (VBS) state [26, 27]. Their work provided crucial insight into the physics of Haldane's conjecture that an integer spin antiferromagnetic Heisenberg chain has a gap in the excitation spectrum [28, 29, 30]. In 1991, Klümper *et al.* showed that the one-dimensional VBS state can be parameterized as a translationally invariant MPS [31, 32]. These works stimulated extensive studies of the translationally invariant subclass of MPS in the community of mathematical physics under the name of *finitely correlated states* [33].

The structure of MPS is simple and conceptually useful. However, in the first ten years of DMRG, MPS was not frequently used because the standard DMRG calculation does not benefit much from this kind of representation. The potential of MPS was released when this kind of wave function was explored using the language of quantum information. In particular, it was shown that the expression power of MPS (or DMRG) could be quantified by the *entanglement entropy*, a concept first introduced in quantum information theory [34, 35]. DMRG correctly characterizes the entanglement structure of the ground state governed by the *area law of entanglement entropy* in one dimension [36, 37]. This is the reason why it is so successful.

In 2004, Vidal introduced the *time-evolving block decimation* (TEBD) method to evaluate the time evolution of a quantum state represented by an MPS [38] without invoking direct renormalization of the Hamiltonian and other physical operators. Verstraete and Cirac also showed that one could variationally calculate the ground-state energy for a periodic system as accurately as for an open boundary system if the ground-state wave function is represented by a periodic MPS [39]. Their works revealed a deep connection between DMRG and quantum information, which has allowed the extension of DMRG to fields quite far from its origin.

From the perspective of MPS, DMRG can be regarded as an algorithm that locally updates the matrix elements of an MPS at two lattice sites at a time. By

exploiting the translational invariance, it is possible to develop algorithms for infinite-size systems with MPS. This means it is feasible to explore the physical properties of a given system directly in the thermodynamic limit without relying on finite-size extrapolations.

In the framework of MPS, the low-lying excitations can also be evaluated under the *single-mode approximation* that was first introduced by Bijl *et al.* [40] and by Feynman [41, 42] in the 1950s. In 1995, Östlund and Rommer proposed a variational ansatz for a single-particle excitation by representing it as a momentum superposition of a locally excited state from the ground state in the framework of MPS [20]. The local excitation is implemented through a bond matrix that is variationally optimized. In 2009, instead of using a bond matrix, Chung *et al.* suggested using some relevant local operators on the lattice sites to generate local excitations [43]. Pirvu *et al*. [44], on the other hand, generated the local excitation by replacing one of the local tensors in the ground-state MPS with a perturbed local tensor that is variationally optimized. Haegeman *et al.* [45] first explored the extension of the single-mode approximation with MPS in the thermodynamic limit.

As mentioned, MPS faithfully represents a quantum state that satisfies the area law of entanglement entropy in one dimension [36, 37]. However, in a critical system where the correlation length diverges, the entanglement entropy grows logarithmically with the system size [46, 47, 48]. This *logarithmic correction* to the entanglement entropy does not affect the calculation of short-range correlation functions much. However, as it diverges with the system size, it can strongly affect the calculation of long-range correlation functions. An MPS cannot describe this logarithmic correction with a finite bond dimension because the bond dimension has to grow polynomially with system size. Nevertheless, it could be described by an infinite-dimensional MPS [49, 50]

In 2007, Vidal proposed a new kind of tensor network state, called *multiscale entanglement renormalization ansatz* (MERA), to represent a critical state effectively [51]. A MERA yields a tensorial representation of a critical system with fixed bond dimensions independent of system size. It consists of a network of isometric tensors that are connected by another class of unitary tensors, called *disentanglers*. These tensors are introduced to mix up different length scales and to locally minimize the entanglement between neighboring sites on the same length scale. Physically, a MERA could be understood as a network on which quantum entanglement propagates. It opens up exciting perspectives for studying one-dimensional quantum or two-dimensional classical systems. The price to pay is an increase in the overall computational cost.

MPS was originally defined in one-dimensional lattice systems. In 2010, Verstraete and Cirac extended it into the continuous limit and introduced *continuous MPS* [52]. Their work allows, for the first time, variational tensor network

https://doi.org/10.1017/9781009398671.002 Published online by Cambridge University Press

algorithms to be applied to quantum field theories in one spatial dimension without lattice discretization. In 2013, Haegeman *et al.* proposed a continuous MERA theory for constructing RG flows of quantum field theories in real space [53]. In 2015, Jennings *et al.* [54, 55] introduced a higher-dimensional generalization of continuous MPS in the framework of path-integral representation.

Nonlocal Basis Space DMRG

In 1996, Xiang extended DMRG from real space to momentum space [56]. In momentum space, a "lattice" is simply a set of properly ordered momentum points allowed by periodic boundary conditions, with each momentum point representing a lattice site. An advantage of implementing DMRG in momentum space is that it enables simple manipulation of the kinetic energy term and treats the momentum as a good quantum number to block diagonalize the Hamiltonian. However, the interaction is highly "nonlocal" in the momentum space. It contains a sum of N^3 terms of four fermion or boson operators, which link two or more momentum points if considering the momentum conservation, where N is the lattice size. To compute and store the matrix elements of these N^3 operators is computationally costly. Xiang solved this problem by introducing a *regrouping method* to reduce the number of operators whose matrix elements need to be computed and stored [56]. This regrouping method works in arbitrary basis space. It paved the way for the application of DMRG in quantum chemistry [57], nuclear physics [58], and quantum Hall effects [59, 60, 61].

The DMRG algorithm was first applied to a molecular system of cyclic polyene chain modeled by an extended Hubbard [62, 63] or the Pariser–Parr–Pople Hamiltonian [64, 65]. It was followed by the first DMRG calculation with the full electronic Hamiltonian of 25 Hartree–Fock orbitals for a water molecule by White and Martin in 1999 [57]. A more accurate DMRG calculation for water molecules outperforming the best coupled-cluster result was first made by Chan and Head-Gordon using 41 orbitals in 2003 [66]. Since then, DMRG has been broadly used to benchmark other molecules. It has also been shown that DMRG is an efficient method for optimizing single-particle basis states from a large reservoir of Hartree–Fock orbitals [67].

Extension to Finite Temperatures

Extension of DMRG to finite temperature was first made by Nishino in 1995 [68] when he realized that the DMRG idea could also be used to diagonalize transfer matrices of two-dimensional classical statistical models. This led to the method

of *transfer matrix renormalization group* (TMRG). In 1996, Nishino and Okunishi combined the idea of *corner transfer matrix*, first introduced by Baxter [69], with DMRG and developed the *corner transfer matrix renormalization group* (CTMRG) method [70]. A variant of this CTMRG, called the *directional corner transfer matrix* method, was introduced by Orus and Vidal in 2009 [71]. This method treats the four directions of the lattice independently and is more flexible in application. It is commonly used, for example, in the contraction of two-dimensional tensor network states.

In 1996, Bursill, Xiang, and Gehring extended DMRG to finite temperatures for one-dimensional quantum systems [72] by introducing a quantum transfer matrix representation for the partition function using the *Trotter–Suzuki decomposition* [73, 74]. In this case, the lattice size is infinite, and the finiteness is at the level of the Trotter approximation along the imaginary time (i.e. inverse temperature) direction. As a quantum transfer matrix is generally non-symmetric and its left and right eigenvectors may not be a conjugate pair, a crucial step in establishing an efficient *quantum transfer matrix renormalization group* (QTMRG) was the introduction of *nonsymmetric density matrix*. This was done by Wang and Xiang [75] and independently by Shibata [76]. However, the nonsymmetric density matrix is not always positive definite, although it should be physically. The accumulated truncation error may ruin the positivity of this density matrix and lead to numerical instability, which dramatically reduces the accuracy of results at low temperatures. To avoid this instability, Huang introduced a *biorthonormalization scheme* to perform the QTMRG calculation [77]. In this scheme, a pair of basis sets obtained from the *singular value decomposition* (SVD) of the density matrix are biorthonormalized and used to construct the renormalized transfer matrix. It reduces the numerical instability and allows much lower temperatures to be reliably accessed by QTMRG.

Thermodynamic properties can also be calculated on a finite lattice system using TEBD based on the *purification* of the thermal density matrix [78]. The purification enlarges the Hilbert space with auxiliary sites, called ancillas. An ancilla contains the same number of basis states as a physical site. It pairs with a physical site to form a maximally entangled state at infinite high temperatures. The whole system resembles a ladder with the ancilla lattice appearing geometrically as another chain parallel to the original one. This ancilla approach is convenient to implement within the traditional framework of DMRG [79].

Time Evolution and Dynamical Correlation Functions

Dynamical correlation functions, such as optical conductivity or single-particle spectral function, can be measured experimentally. However, the calculation of

https://doi.org/10.1017/9781009398671.002 Published online by Cambridge University Press

these quantities is rather challenging. During the last three decades, several approaches have been proposed to calculate spectral functions using DMRG or related methods.

In 1987, Gagliano and Baliseiro [80] proposed a continued-fraction method (also called the Lanczos vector method) to evaluate dynamical correlation functions based on the Lanczos diagonalization. Their method was first adopted in the DMRG calculation by Hallberg [81] in 1995. While this method requires only modest numerical resources, reliable results that can be obtained are only limited to low frequencies.

The continued-fraction method uses sequential basis states generated by the Lanczos iteration to represent the whole Hilbert space of the system. It is not accurate nor efficient for describing the correlation at a particular frequency. To remedy this deficiency, the *correction vector method*, initially proposed by Soos and Ramasesha [82], was introduced in the DMRG calculation, first by Ramasesha *et al.* [83] in 1997, and later refined by Kühner and White [84] in 1999. The critical idea is to take Green's function at a particular frequency as a target state and calculate it by solving a set of large but sparse linear equations. This method can generate extremely accurate results at a given frequency, but the computational cost is very high. In 2002, Jeckelmann [85] improved the correction vector method and showed that it is much more efficient and accurate to determine the correction vector by minimizing a cost function. In 2011, Dargel *et al.* proposed an *adaptive Lanczos vector method* [86] to improve efficiency in the calculation of dynamical correlation functions. In 2016, Nocera and Alvarez [87] proposed a Krylov-space approach to replace the *conjugate gradient method* in the calculation of correction vectors.

It is more convenient to investigate dynamical spectral functions using either a Chebyshev expansion [88, 89] or a Lanczos expansion [90] in the framework of MPS. The key idea is to use MPS to represent the Chebyshev or Lanczos vectors generated in the corresponding expansion. However, as each vector obtained at each step of expansion is represented by an MPS, the recurrence or orthogonal relation between different vectors is satisfied only approximately. Thus, a reorthogonalization of these vectors is desired to improve the accuracy of dynamical correlation functions [89, 90]. This kind of method offers a balanced scheme between cost and accuracy. It yields results with accuracies comparable to those of the correction-vector DMRG but at dramatically reduced computational cost. Furthermore, one can also use the orthogonalized Chebyshev or Lanczos vectors to represent the Hamiltonian. Diagonalizing this Hamiltonian can also accurately determine the spectral weight at each energy eigenvalue, offering a simple but accurate approach to performing the finite-size scaling in the entire energy range [89].

Dynamic quantities can also be calculated from the Fourier transform of the time-dependent Green's function. However, to obtain good frequency resolution, one has to calculate the correlation functions over a long time interval, which is limited by either a loss of accuracy due to the approximation used or by finite-size effects such as reflections from the two open ends. In 2002, Cazalilla and Marston evaluated the time-dependent correlation functions for a one-dimensional system under an applied bias using DMRG [91]. In 2003, Luo *et al.* [92] proposed a *pace-keeping approach* to optimize the basis states that are retained with the time evolution. TEBD, on the other hand, provides a method to efficiently simulate the time evolution of one-dimensional quantum systems with short-range interactions using MPS [38]. Based on this approach, an *adaptive time-dependent DMRG* was proposed by Daley *et al.* [93] and by White and Feiguin [94].

TEBD is implemented by taking the Trotter–Suzuki decomposition. At each time of evolution, an MPS with a larger bond dimension than the original one is obtained. To proceed, one has to truncate the MPS by discarding less important parameters. In order to reduce both the truncation and the Trotter errors, Haegeman *et al.* introduced a *time-dependent variational principle* to optimize an MPS that preserves all symmetries of the system [95].

Investigation of long-time dynamics, however, remains challenging due to the linear growth of the entanglement entropy with the evolving time [96]. To catch up with the growing speed of the entanglement entropy, the bond dimension of MPS has to grow exponentially with time. In 2009, Bañuls *et al.* [97] proposed a *folding scheme* to reduce the entanglement of transfer matrices defined in the time direction. It yields an accurate approach for evaluating the long-time dynamics of a quantum state in the thermodynamic limit. A QTMRG extension of this method to finite temperatures was proposed by Huang *et al.* [98]. DMRG has also been extended to simulate time evolution at finite temperatures by purifying thermal density matrices with ancillas [99, 100].

Two- or Higher-Dimensional Quantum Systems

Investigation into physical properties of two-dimensional quantum systems with strong correlations, including high-temperature superconductivity, quantum Hall effects, frustrated antiferromagnets, and quantum spin liquids, has been at the heart of condensed matter physics in the past four decades. After the successful application of DMRG in one dimension, it is natural to extend this method to two or more dimensions. Liang and Pang made the first attempt to apply DMRG to two dimensions [101]. Unlike in one dimension, they found that DMRG is only moderately successful in two dimensions because computational resources need to increase exponentially with the width of the lattice, implying that DMRG can reliably access

only narrow two-dimensional lattices. From an analytical study for the spectra of a two-dimensional system of coupled oscillators, Chung and Peschel also found that the eigenvalues of density matrix decay much slower than in one dimension [102].

A better understanding of this exponential breakdown of DMRG in two dimensions is revealed by the entanglement area law, which provides a systematic way to quantify quantum correlations. The entanglement area law is believed to hold in both one and two dimensions if the ground state is gapped so that the correlation length is finite. According to the area law, the entanglement entropy of the ground state between two subsystems scales with the size of the interface. In one dimension, the interface contains just one site, independent of the lattice size. In two dimensions, however, the interface grows with the lattice size.

To capture the entanglement feature of the ground state correctly, the minimal number of many-body states should grow exponentially with the entanglement entropy. It puts an upper bound on the system sizes that can be accurately simulated. Nevertheless, DMRG is still one of the most promising methods for studying two-dimensional quantum lattice models because other methods, such as the exact diagonalization and the quantum Monte Carlo, all have their limitations. DMRG has been successfully applied to study, for example, the t-J model [103], Hubbard model [104], frustrated quantum spin models [105, 106, 107], and other quantum lattice models in two dimensions.

A simple extension of DMRG to two dimensions would be to replace every site added between the left and right blocks with a column of sites. However, the extra degrees of freedom added to the system would make the size of the Hilbert space prohibitively large. A practically feasible extension of DMRG to two dimensions is to add a single site to a block at a time. To do this, one needs to map a two-dimensional lattice onto a one-dimensional one or select a path to order all lattice points at the price of breaking the lattice symmetry and introducing long-range interactions.

A typical mapping is to fold a two-dimensional lattice, generally a multichain system, into a one-dimensional zipper by ordering the lattice sites in a snake path extending along the chain direction. Alternatively, one can zip a two-dimensional lattice along the diagonal direction [34]. This diagonal map would allow one to build up a $L \times L$ lattice using the blocks of an $(L-1) \times (L-1)$ lattice.

In a two-dimensional DMRG calculation for a multichain system, open or cylindrical boundary conditions are generally assumed instead of fully periodic boundary conditions. This is to avoid squaring the number of basis states required for a given accuracy. By cylindrical boundary conditions, we mean open boundary conditions along the chain direction and periodic boundary conditions along the direction perpendicular.

https://doi.org/10.1017/9781009398671.002 Published online by Cambridge University Press

As mentioned before, the wave function generated by DMRG is an MPS. In two dimensions, the entanglement entropy scales with the linear size of a lattice in a gapped system. To capture this effect, the bond dimension of MPS has to increase exponentially with the cylinder's circumference. To resolve this exponentially growing problem, one needs to find a scalable representation of the ground-state wave function that captures the entanglement area law.

Niggemann *et al.* made the first tensor network extension of MPS to two dimensions in the study of an extended spin-3/2 Heisenberg model on a honeycomb lattice [108]. They called this kind of tensor network state a *vertex-state model* to emphasize its resemblance to a standard vertex model. Similar to an MPS, they defined a local tensor at each lattice site that maps a set of bond variables onto a physical state. Later, Sierra and Martin-Delgado made a similar ansatz for the tensor network wave function [109]. The use of a tensor network state as a variational wave function for the three-dimensional classical lattice model was also suggested by Nishino [110].

In 2004, the idea of two- or higher-dimensional tensor network states was discussed by Verstraete and Cirac from a perspective of quantum information theory [111]. They reinvented the tensor network wave function first suggested by Niggemann *et al.* [108], and called it a *projected entangled pair state* (PEPS), since it can be understood in terms of pairs of maximally entangled virtual basis states defined on the bond linking two sites. Their work has attracted significant attention because it reveals more clearly the physical picture embedded in the tensor network representation of quantum states. Furthermore, they pointed out that PEPS satisfies the entanglement area law since the number of entangled bonds between any two blocks grows linearly with the size of their interface [112]. In other words, PEPS presents a faithful representation of a ground state that is governed by the entanglement area law.

PEPS accurately describes the entanglement between any two neighboring sites in the ground state. It is, for example, an exact representation of the ground state of the two-dimensional AKLT model [27], which is a VBS. In principle, it can be used to accurately describe an arbitrary quantum state that satisfies the entanglement area law if there is no limitation on the bond dimensions of local tensors. However, in practical calculations, the bond dimension of PEPS that can be reliably handled is generally not much more than ten without imposing any symmetry. In this case, PEPS is no longer an efficient representation of a quantum state in which the entanglements among three or more particles become important. An example of this kind of quantum state is the so-called *simplex solid state* (SSS), first introduced by Arovas [113].

A simplex is a building block of a lattice. For example, a four-site square is a simplex of a square lattice, and a three-site triangle is a simplex of a Kagome lattice.

https://doi.org/10.1017/9781009398671.002 Published online by Cambridge University Press

Unlike VBS, SSS emphasizes the group entanglement within a simplex. The wave function that characterizes the SSS-type state is given by the so-called *projected entangled simplex state* (PESS) first introduced by Xie *et al.* in 2014 [114]. PESS presents a natural extension of PEPS to a system in which many-particle or cluster entanglement becomes important. It also satisfies the entanglement area law. In the limit that a simplex contains two sites, it reduces to PEPS.

Either PEPS or PESS can accurately describe a quantum state with short-range entanglement and keep the locality of interactions. The minimal bond dimension that is needed for accurately describing a quantum state does not depend on the system size in a gapped system. This is an advantage of using this kind of ansatz. Furthermore, it can be directly used in a thermodynamic limit like an MPS.

Unlike an MPS with open boundary conditions, we cannot choose a gauge to simultaneously orthogonalize all the bond indices of PEPS so that all local tensors are canonicalized. It also happens for MPS with periodic boundary conditions or, more generally, if there is a loop in a tensor network state. A loop in a tensor network state means that one cannot split the network into two parts by just cutting one bond.

The biggest challenge in developing a two-dimensional tensor network algorithm is the optimization of local tensors, which give the best approximation to the ground state for a given Hamiltonian. Two kinds of approaches can be used to determine the local tensors of PEPS (similarly to PESS). One is to determine the local tensors by variationally minimizing the ground-state energy [111]. The other is the so-called update approach in which a PEPS is determined by performing an imaginary time evolution, that is, by applying the projection operator, or the density-matrix operator, $\exp(-\beta H)$ onto an arbitrary initialized PEPS which is not orthogonal to the true ground state. This projection is made by dividing β into many small pieces so that the projection operator can be readily evaluated using the Trotter–Suzuki decomposition formula [73, 115]. At each projection, an approximation has to be taken to truncate the bond indices of PEPS. The PEPS obtained with this approach would converge to the ground state in the zero-temperature limit, $\beta \to \infty$.

There are two kinds of update approaches: the simple update [116] and the full update [117]. The simple update is essentially an entanglement mean-field approach. It approximates the environment tensors by entanglement spectra defined on the bonds separating the system and environment sites. In other words, these bond entanglement spectra serve as an effective mean field acting on the system tensors whose elements are updated. This simple update approach becomes exact on the Bethe lattice or any other kind of lattice without loops [118].

The simple update is a local update approach. It underestimates the long-range correlations but works very efficiently and allows a PEPS or PESS with a bond

dimension as large as 100 or more to be approximately determined. With this approach, a system block with one or a few tensors is updated at one time. A PEPS determined by the simple update can be directly used to evaluate physical observables. It can also be used as an initial state in the full update calculation. One can also update a relatively large cluster of tensors at one time. In this case, the simple update is referred to as a cluster update approach [119]. The entanglement mean-field approximation used in the simple update can also be used in the variational minimization of PEPS or PESS [120]. It allows us to efficiently determine PEPS or PESS with a relatively large bond dimension.

The full update evaluates the environment tensor by directly contracting all the local tensors in the environment block [117] using, for example, a boundary MPS or CTMRG method. It is more accurate than the simple update but has a much higher computational cost. The bond dimension this approach can handle is generally less than ten if the tensors are not block diagonalized by imposing symmetries.

Both PEPS and PESS are variational wave functions and satisfy the variational principle. Hence, the energy calculated with either PEPS or PESS should be an upper bound of the true ground-state energy. However, as it is difficult to rigorously contract a PEPS or PESS, certain approximations have to be introduced in computing observables. These approximations may violate the variational principle and ruin the condition for the calculated energy being a rigorous upper bound.

The PEPS or PESS can also represent a low-lying excitation in a translation-invariant system under the single-mode approximation. This approach was first explored by Vanderstraeten *et al.* in 2015 [121]. The single-particle excitation is assumed to be a momentum superposition of a locally excited state by replacing one of the local tensors in a PEPS ground state with a perturbed local tensor that is variationally optimized. Dynamical spectral functions can be evaluated using this kind of tensor network states [122, 123].

Besides PEPS and PESS, a number of other tensor network states have also been introduced in two or more dimensions. This includes but is not limited to the tree tensor network state [124], string-bond state [125], entangled-plaquette state [126], and branching MERA [127].

Coarse-Graining Tensor Renormalization

Before the tensor network wave function was introduced in quantum systems, tensor network representations for the partition functions of classical lattice models were already used in statistical physics. In fact, as discussed in Chapter 3, all classical lattice models with local interactions can be represented as *tensor network models*. A tensor network model is a special kind of tensor network state that does

https://doi.org/10.1017/9781009398671.002 Published online by Cambridge University Press

not have any dangling physical bonds. Moreover, this representation is rigorous, unlike a quantum TNS whose local tensors have to be variationally determined. The partition function is determined by contracting all local tensors.

Similarly, in the determination of a PEPS, either by the variational minimization or by the full update, or in the evaluation of its observables, one needs to calculate a scalar product of two PEPS. This scalar product of two PEPS forms a double-layer tensor network model. The physical bonds connect the two layers of local tensors. By contracting out these physical bonds, the double-layer tensor network model becomes a single-layer one.

To exactly solve these tensor network models by contracting out all local tensors is, unfortunately, an exponentially hard problem. The computational cost scales exponentially with the number of sites no matter how the local tensors are contracted. To resolve this problem, we have to rely on some approximate schemes.

A two-dimensional tensor network model can be approximately solved, for example, using TMRG or CTMRG. It can also be solved by employing a modified TEBD method, which is referred to as a boundary MPS method in the literature. These methods are efficient and accurate in performing this kind of calculation, and their costs scale linearly with system size.

The scheme of coarse-graining tensor renormalization has also been developed to solve two-dimensional tensor network models. The idea is to perform a scaling transformation by coarse-graining local tensors until some fixed-point tensors are reached, similar to the Kadanoff block spin scheme [8] in classical statistical models. In 2007, Levin and Nave introduced the first coarse-graining renormalization group method of tensor network models [128]. They coined it *tensor renormalization group* (TRG). This method truncates the basis space according to the singular-value spectra of local tensors. It provides a local optimization of the truncation space. However, it does not consider the renormalization effect of the environment tensors on the singular-value spectra, which is key to the success of DMRG. To solve this problem, Xie *et al.* introduced a *second renormalization group* (SRG) method [129, 130] to account for the environment contribution. The SRG improves the accuracy of TRG significantly. For example, for the two-dimensional Ising model, the accuracy is improved by more than two orders of magnitude at the critical point and more than five orders of magnitude far away from the critical point by keeping 24 states.

The difference between TRG and SRG is similar to the difference between Wilson's NRG and DMRG. In Wilson's NRG, a block Hamiltonian is optimized without considering the interaction between its building cells. On the contrary, in DMRG, the truncated basis states of the system are fully optimized by considering the interplay between the system and the environment through the reduced density matrix, which measures the entanglement spectra between these two blocks.

The TRG algorithm can also be improved by just taking *higher-order singular value decomposition* (HOSVD) or more generally *Tucker decomposition* for local tensors [131]. This *HOSVD-based tensor renormalization group* (HOTRG) method determines the entanglement spectra on all the bonds of a local tensor at the same time. It presents a better scheme to truncate the basis states than TRG. HOTRG can also incorporate SRG to improve the accuracy of results further. This HOSVD-based SRG method is called HOSRG [131].

However, TRG and HOTRG do not remove all short-range entanglement embedded in each loop that is gauge invariant during the coarse-graining process. As a result, the effective tensor network at a given length scale still contains irrelevant correlations belonging to shorter length scales. The accumulation of these short-range correlations over successive coarse-graining transformations would ruin the scaling invariance at the critical point, leading to a large truncation error around the critical point. Nevertheless, universal information, such as critical exponents, can still be obtained from the fixed-point tensors.

To remove short-range entanglement, Gu and Wen proposed a tensor entanglement filtering renormalization approach and pointed out that it is crucial to optimize the tensor configurations that contain a loop [132]. However, this approach is computationally inefficient. A better approach for removing short-range entanglement or correlations at each coarse-graining step was introduced by Evenbly and Vidal based on the idea of MERA [133]. This approach, which is referred to as *tensor network renormalization* (TNR), is based on the insertion of optimized unitary disentanglers and isometric tensors into the tensor network. To a large extent, TNR recovers scale invariance at criticality.

In 2017, Yang *et al.* proposed a loop tensor network renormalization (Loop-TNR) approach to remove short-range entanglement [134]. In this approach, the short-range entanglement within a loop is removed by optimizing the rewiring tensors at each coarse-graining step. Like TNR, Loop-TNR improves the RG flow around a critical point. It produces accurately critical exponents, including central charge and scaling dimensions, with a lower computational cost than TNR.

The methods introduced here each have advantages and disadvantages. Some are easier to implement, and some are more stable in catching critical behaviors. In general, the cost of DMRG- or MPS-based methods is lower than that of coarse-graining methods. But the coarse-graining methods scale logarithmically with system size and allow the scaling exponents to be directly computed.

Among these methods, only HOTRG and HOSRG are applicable to three-dimensional tensor network models. These two methods can be used to directly contract three-dimensional classical lattice models or equivalent (2+1)-dimensional quantum lattice models. A HOTRG calculation for the

three-dimensional Ising model, for example, produces a very accurate estimation for the critical temperature [135] as well as the temperature dependence of the specific heat and magnetic susceptibility [131].

Other methods are either not applicable or difficult to extend to study of three-dimensional tensor network models. In order to use these methods, we have to make a dimension reduction to contract a three-dimensional tensor network model using a boundary PEPS, whose local tensors are determined using a variational optimization approach. The computational cost for determining this PEPS and its expectation values is generally higher than HOTRG.

1.4 Applications

Density-matrix and tensor network renormalization provides a state-of-the-art method for the classical simulation of quantum systems. It was initially introduced to investigate classical statistical and quantum lattice models in condensed matter physics, statistical mechanics, and quantum information, which significantly deepened our understanding of so-called correlated quantum phenomena. Later they were extended to new fields and generated significant impact, for example, in quantum chemistry, cold atoms, quantum computing, artificial intelligence, and, more generally, in the study of complex systems with a large number of degrees of freedom or variables. During the past thirty years, more than ten thousand scientific papers have been published in the development and applications of density-matrix and tensor network renormalization.[1] It is beyond my ability to comprehensively review the vast literature of papers published in this field. A survey of DMRG with its application up to late 1998 was given in a collection of lectures and articles [136]. To gain a more comprehensive picture of the DMRG or other tensor network-related methods and their applications, we refer the interested readers to the review articles [137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151] and references therein. A comparison between the early real-space renormalization group methods and the newly developed tensor-network-based renormalization group methods can be found in Ref. [152].

One Dimensional Quantum Systems

As an algorithm that is simple to implement, DMRG, together with its finite temperature and finite frequency extensions, has now become the most effective and accurate numerical method for studying not just the ground state but also thermodynamic and dynamic properties in one dimension. It has been successfully applied

[1] See the DMRG Home Page, http://quattro.phys.sci.kobe-u.ac.jp/dmrg.html, run by Tomotoshi Nishino.

to solve nearly all fundamentally interesting models with short-range interactions in one dimension (see Refs. [136, 137, 138, 139, 141]), including, but not limited to, the following quantum systems:

(i) Quantum spin models
This includes the antiferromagnetic Heisenberg models [19, 20, 153, 154, 155, 156, 157, 158] and their experimentally relevant generalizations by adding anisotropy [159, 160], dimerization [161], frustration [162, 163, 164], defects [165, 166, 167], randomness [168], or other interactions [163, 169, 170] to the Hamiltonians. A number of interacting spin models with higher symmetries, such as the SU(N) symmetry ($N > 2$) [171, 172], have also been studied.

The quantum spin models are physically interesting for three reasons. First, as the charge degree of freedom is frozen at each lattice site, this kind of model is relatively simple to study compared to the interacting fermion models, which serves as a playground for testing various numerical methods. Second, in 1983, Haldane predicted that the one-dimensional Heisenberg model with integer spin has an excitation gap and a finite correlation length [28, 29]. Since then, much experimental and theoretical effort has gone into understanding the difference between half-integer and integer spin chains. Third, there are a lot of quasi-one-dimensional compounds whose physical properties can be adequately described within the framework of interacting spin chains governed by quantum spin models.

The spin-1 Heisenberg model was the first quantum lattice model used by White to demonstrate the efficiency and accuracy of DMRG [18]. A benchmark calculation of DMRG was made by White and Huse [19] to evaluate the spin excitation gap and the correlation length for this model with an accuracy that was difficult to achieve with other numerical methods. The results confirmed the valence-bond-solid state picture proposed by AKLT [26, 27].

(ii) Interacting fermion models
This includes the Hubbard model [14, 173, 174, 175], extended Hubbard model [176, 177], multiband Hubbard model [178], SU(N) Hubbard model [179], t-J model [180, 181], periodic Anderson [182, 183], and Kondo lattice [184, 185] models, with randomness [186, 187], or other interactions [188]. The study of interacting fermion models is somewhat more complicated because of the larger number of degrees of freedom at each lattice site. However, DMRG is not bothered by the negative-sign problem that hampers quantum Monte Carlo simulations.

The Hubbard model [14], named after J. Hubbard, is a basic model for describing the Mott metal-insulator transition, charge-spin separation,

quantum magnetism, charge density waves, and other fundamental properties of interacting electrons. It reduces to the t-J model in the strong coupling limit. The Hubbard [189] and t-J models [190] are also the two simplest models used for understanding the mechanism of high-temperature superconductivity discovered in copper oxides. The Kondo lattice model is equivalent to the periodic Anderson lattice model in the strong coupling limit [191]. They are the two basic models for understanding the physical properties of heavy fermion systems in which itinerant electrons interact strongly with localized spins.

(iii) Systems with bosons

Bosonic systems are generally more challenging to treat than fermionic ones in DMRG because a single bosonic mode has an infinite state space. The basis space of a bosonic mode must be truncated before carrying out DMRG calculations [192, 193, 194, 195]. A commonly adopted approximation is to restrict the number of bosons in a finite interval around its average at each lattice site. This approximation is valid when the fluctuation in the number of bosons at a site is small.

There are three kinds of bosonic models that have been investigated with DMRG. The first includes the systems that contain just bosons [196, 197]. A typical example is the Bose–Hubbard model [196]. This model mimics the dynamic properties of ultracool atoms in optical lattices generated by laser beams [198]. Both the doping of particle numbers and the interaction strength can be readily controlled in an optical lattice of ultracool atoms. The phase diagram of this model, particularly the Mott insulator to superfluid condensate transition, has been thoroughly investigated with DMRG in conjunction with experimental measurements [199, 200]. The second model includes the systems of bosons coupled with fermions, such as the Su–Schrieffer–Heeger model or other electron–phonon coupled systems [193]. The third includes the interacting systems of bosons coupled with spins, such as the spin-Peierls system [195, 201] and the multiconnected Jaynes–Cummings lattice model [202]. A multiconnected Jaynes–Cummings lattice can be realized by coupling qubits, described by spins, with some cavity photon modes.

(iv) Quantum field theory and quantum gravity

The application of tensor network methods to the lattice gauge theory is a forefront field undergoing rapid development [148]. In particular, DMRG and other tensor network methods offer a powerful tool for defeating the notorious sign problem encountered in the simulation of lattice quantum chromodynamics [203, 204] and for exploring physical properties of the (1+1)-dimensional Abelian Schwinger model [205, 206, 207], $\lambda\phi^4$ scalar field theory [208, 209], Z_2 lattice gauge theory [210], SU(2) [211] and SU(3) [212] gauge theory, O(N) ($N = 2, 3, 4$) nonlinear sigma models [213, 214, 215, 216],

Gross–Neveu–Wilson model [217], lattice Abelian Higgs model [218], and massive Thirring model [219].

Continuous MPS provides a powerful variational ansatz for the ground state of strongly interacting quantum field theories in one dimension [52]. This ansatz, formulated directly in the continuum, enables us to parameterize the ground-state wave function without resorting to a lattice discretization scheme. It was first illustrated with the Lieb–Liniger model [52, 220], which is a basic model of quantum field theory describing one-dimensional bosons interacting through a repulsive contact potential. This approach was then extended to the study of excited states of the Lieb–Liniger model [221], free massive Dirac fermions and the Gross–Neveu model [222], coupled Lieb–Liniger models [223], and systems of two species of fermions [224]. It was also generalized for studying thermodynamic properties of one-dimensional quantum lattice models in the continuous imaginary time representation of quantum transfer matrices [225].

Tensor networks are relevant to quantum gravity. This connection was first pointed out by Swingle [226]. It was proposed that MERA [51] is linked to the geometry of space through a conjectured relationship or correspondence between the anti-de Sitter space in theories of quantum gravity and conformal field theories. Particularly, it was assumed that MERA could be understood as a lattice realization of an anti-de Sitter space with some geometry, in which the curvature is somehow linked to entanglement. The correspondence implies that space-time geometry may emerge from the underlying entanglement structure in a complex quantum state. This intriguing connection has attracted great interest in the community of superstrings. However, a report also claims that MERA is actually a lightcone geometry rather than an anti-de Sitter space [227].

Two- or Higher-Dimensional Quantum Lattice Models

A thorough investigation of two-dimensional quantum systems is essential to the microscopic understanding of high-temperature superconductivity, frustrated magnetism, quantum spin liquids, and many other novel quantum phenomena discovered in quasi-two-dimensional materials. However, the entanglement entropy grows much faster in two dimensions than in one dimension. As a result, it leads to a dramatic increase in the demand for computational resources compared to one-dimensional calculations. Yet geometrical constraints in two dimensions are much more relaxed than in one dimension, allowing exotic elementary excitations, such as anyons, and competing quantum fluctuations to exist.

Two-dimensional quantum models have been investigated with both DMRG and the tensor network methods based on the PEPS representation of wave functions

[138, 140, 141, 142, 143, 144, 149, 150]. DMRG calculations start by unzipping a two-dimensional lattice into a one-dimensional one [34, 101]. Due to the constraint imposed by the entanglement entropy, DMRG is limited to the study of two-dimensional lattices with a relatively small width. PEPS maintains the lattice structure and distributes the entanglement more evenly on the whole lattice. It reduces the dimension of the virtual basis state on each bond and allows even an infinite lattice system to be studied.

The first applications of both DMRG- and the PEPS-related RG methods were in the area of magnetism [228]. The results obtained from the calculation of the two-dimensional Heisenberg model were encouraging, demonstrating the potential of these methods in solving strongly correlated problems in two dimensions [34, 105, 106, 107, 116, 229, 230] as well as in three dimensions [231]. For example, both the ground-state energy and the magnetization of the square lattice Heisenberg model obtained with either DMRG [34, 229] or PEPS [230] are consistent with the quantum Monte Carlo results [232].

The tensor network RG study of the frustrated Heisenberg models on the triangular [229], kagome [105, 106, 107], honeycomb [233], and the Shastry–Sutherland lattice [234], as well as models frustrated by further-neighbor [235, 236, 237] or multispin interactions [238, 239], has yielded fruitful results, particularly in search of quantum spin liquids. It has settled several problems that have been long debated [105, 106, 107]. A quantum spin liquid is difficult to identify because it is topologically nontrivial and may possess long-range entanglement. Furthermore, it does not break any symmetry and cannot be characterized by a local order parameter.

Another set of systems extensively simulated by DMRG or PEPS is the interacting fermionic systems described by the t-J model [240, 241, 242], the Hubbard model [243, 244], and the Kondo–Heisenberg lattice model [245]. A common feature revealed is the formation of the so-called stripe phase in lightly hole-doped systems [240, 243, 244]. A stripe is a charge-density wave state separated by magnetically ordered states. It was first discovered in cuprate superconducting materials by neutron scattering spectroscopy [246]. Superconducting pairing states have also been studied for these models by DMRG [245]. It seems that the long-range pairing order is absent in the thermodynamic limit of the Hubbard model if there is only nearest-neighbor hopping.

The DMRG and PEPS have also been applied to interacting boson systems [231, 247, 248]. Particular attention has been devoted to the study of so-called supersolid [249] in the extended Bose–Hubbard model and Bose metal [250]. Supersolid is a phase with simultaneous charge-density wave and superfluid order. On the other hand, Bose metal is not a phase characterized by an order parameter but by a pattern of correlations associated with a surface of gapless modes.

Quantum models defined on a Bethe or, more generally, a tree lattice have also been extensively studied with tensor network methods. It is natural to represent a quantum state as a *tree tensor network* on a tree lattice. The lack of loops implies that a tree tensor network state can be rigorously canonicalized by minimizing the ground-state energy with DMRG or by taking an imaginary time evolution with a proper local tensor update scheme like simple update [116]. The simple update is an entanglement mean-field approach for canonicalizing a tensor network state, which works particularly well on a tree lattice [118]. It leads to a thorough understanding of magnetic orders and other physical properties of the spin [251, 252, 253, 254], and interacting fermion [255, 256] models on the Cayley or Bethe trees. However, the correlation lengths are always finite, even at a critical point [118]. This implies that the entanglement entropy is upper bounded, not like in a regular lattice, which can be used to evaluate accurately low-temperature thermodynamic quantities just by taking imaginary time evolution [257].

Quantum Systems with Nonlocal and Off-Diagonal Interactions

In an arbitrary basis space, one can order all the single-particle basis states to form an effective "one-dimensional lattice" and carry out the DMRG calculation similarly as in real space. Unlike in real space, the interaction of particles, represented using these single-particle basis states, is generally nonlocal and off-diagonal. For example, the Hubbard interaction is local and diagonal in real space. In momentum space, on the other hand, each momentum point in the first Brillouin zone serves as a basis site, and the Hubbard interaction becomes off-diagonal, which involves the coupling from two to four momentum points, and long-ranged.

For two reasons, implementation of DMRG in momentum or other non-local basis space is technically more challenging than in real space. First, the basis states are not naturally ordered. One should determine the order of these basis states that optimizes the final results [56, 258, 259]. Local optimization of orbital ordering could be achieved by taking an adaptive scheme to update the active basis states with a unitary transformation for the fermion operators on the two middle sites in the superblock of DMRG [259]. This unitary transformation optimizes the basis states by minimizing the entanglement between the augmented system and environment blocks. It defines a new set of basis states and orders them optimally. One could further improve this optimization scheme by sorting the orbitals according to their mutual entanglement structures [258]. The purpose is to reduce long-range correlations by placing strongly entangled orbitals close to one another. Second, as the interaction becomes nonlocal and off-diagonal, there are more operators whose matrix elements need to be evaluated and stored, leading to a dramatic increase in computational time and memory space. This problem could be significantly

https://doi.org/10.1017/9781009398671.002 Published online by Cambridge University Press

ameliorated by factorizing the Hamiltonian and introducing the so-called *complementary operators* that take partial sums over certain combinations of operators [56] within each subblock.

(i) Momentum space calculations
The application of momentum-space DMRG benefits from momentum conservation. It allows the Hamiltonian to be block diagonalized according to the values of the total momentum so that more basis states can be retained at each DMRG iteration. Furthermore, in a given interacting fermion model, different scattering processes, such as the forward, backward, or Umklapp scattering, could be readily identified in momentum space. This provides a unique scheme, not feasible in other basis spaces, to understand the effect of each scattering process by screening other scattering terms in the model.
The momentum-space DMRG has been mainly applied to explore the ground-state properties of the Hubbard model [56, 260, 261]. In both one and two dimensions, the bipartite entanglement entropy is found to satisfy the volume law and scale quadratically with the Hubbard interaction in the weak coupling limit. Nevertheless, the momentum-space DMRG can still provide accurate results for the two-dimensional Hubbard model with moderate system sizes [261]. A one-dimensional Hubbard model with long-range hopping is also investigated with DMRG [260]. This model is difficult to investigate with the real-space DMRG because the hopping is long-ranged.
One possibility to combat the volume-law increase of the entanglement entropy is to carry out DMRG calculations using a hybrid real- and momentum-space representation. This approach aims to study a two-dimensional cylindrical system by taking a real-space representation in the direction along the axis of the cylinder and a momentum space representation in the direction around the circumference. In this way, the translational invariance and good momentum quantum numbers are preserved in the transverse direction. This hybrid approach was used to study the interacting fermionic Hofstadter model [262] and the Hubbard model [263] in two dimensions. It leads to a considerable reduction in computation time and memory space compared with the pure real-space approach [262, 263].

(ii) Fractional quantum Hall effect
The fractional quantum Hall effect occurs in a strong magnetic field where the kinetic energy of electrons is quenched into highly degenerate Landau levels and the electron–electron interaction is the only relevant term in the Hamiltonian. This leads to the emergence of highly entangled and nonperturbative ground states with fractionalized particles.

Due to the nonperturbative nature of the fractional quantum Hall effect, numerical methods have played a crucial role in revealing its microscopic picture. In the Landau gauge, the noninteracting Landau orbitals are Gaussian localized [264], and the system could be mapped onto an effective "one-dimensional" chain with long-range interaction. It was shown that certain model wave functions of fractional quantum Hall effects, such as the Laughlin [265] and Moore–Read states [266], can be exactly represented by MPS [267]. This exact MPS representation has an infinite bond dimension (as the virtual space has support on conformal towers of states), but it can be truncated to a finite-dimensional MPS with high fidelity, which can be used for efficiently computing physical quantities [267, 268]. Furthermore, the topological order of fractional quantum Hall liquids can be identified from the entanglement spectrum obtained by DMRG through the conjecture made by Li and Haldane [269].

Like in the Hubbard model, the total momentum on a torus (or angular momentum on a sphere) of the fractional quantum Hall system is conserved. Therefore, the momentum-space DMRG can be extended to apply to the fractional quantum Hall system. It yields a powerful numerical tool for accessing a fractional quantum Hall system whose size is significantly larger than what could be handled by exact diagonalizations. Shibata and Yoshioka [59] made the first attempt along this line, emphasizing the investigation of the physical properties of electrons in higher Landau levels. More DMRG calculations were followed by a number of groups [60, 61, 270, 271]. As a result, the efficiency of DMRG has been significantly improved, which allows more than twenty thousand states to be retained [61]. In addition, DMRG has also been applied for studying fractional quantum Hall effects in a bosonic system [272], as well as in an extended Kagome Heisenberg model [273].

(iii) Quantum chemistry calculations

The DMRG implementation in quantum chemistry was first presented in the study of π-electrons of conjugated polyenes whose electronic properties are modeled by an extended Hubbard [62, 63] or the Parisier–Parr–Pople [64, 65] Hamiltonian. The fully *ab initio* DMRG determination of the electronic structure of molecules, on the other hand, started from the work of White and Martin in 1999 for the calculation of the ground-state energy of a water molecule with 25 active molecular orbitals [57]. Their work demonstrated the potential of DMRG for *ab initio* quantum chemistry calculations. The early applications focused on the multiconfigurational calculations for small molecules [66, 274, 275, 276, 277]. It was also applied to solve full configuration interaction problems for quasi-one-dimensional molecules, including linear hydrogen chains [278, 279], polyenes [278, 280] and other

https://doi.org/10.1017/9781009398671.002 Published online by Cambridge University Press

π-conjugated organic systems [281], transition-metal complexes or clusters [282, 283]. DMRG allows more than 100 active orbitals to be included in the calculation. This size of active basis space is not reachable by full configuration interaction algorithms.

In the quantum chemical DMRG, molecular orbitals play the role of lattice sites. As the Coulomb interaction is inherently long-ranged and off-diagonal in the orbital space, the lattice formed by the active orbitals are often far from one-dimensional-like and relatively large bond dimensions are required to use DMRG for diagonalizing the Hamiltonian. As the number of complementary operators whose matrix elements need to be evaluated and stored scales quadratically with the dimension of the active space, not many active orbitals can be used in the DMRG iteration. Luo *et al.* introduced a useful scheme to tame this problem, which allows DMRG to optimize the orbitals by exchanging one or two least-active orbitals in the active space with an equal number of inactive orbitals from a larger pool of orbitals after each DMRG sweep [67]. An orbital is chemically inactive if it is almost empty or fully occupied. The occupation number of an active orbital is determined by the eigenvalue of the single-particle density matrix obtained from the ground-state wave function [67, 284, 285].

The rapid development of DMRG has turned it into a reference approach for large-scale multiconfigurational calculations [147, 286], which dramatically broadens its range of applications in quantum chemistry. To more efficiently encode the entanglement structure of active orbitals, a higher-dimensional tensor network extension, based on the tree tensor network states [287, 288], is used to represent the ground-state wave function. This kind of tensor network state encodes entanglement in a tree-like structure, allowing for a more feasible description of molecules. Furthermore, the relativistic effects have also been explored [289, 290, 291].

Besides the static properties, DMRG has also been extended to study dynamical correlation functions. However, dynamical correlations involve the contribution of excited states whose wave functions are described by some orbitals not included in the active space. For the inclusion of those omitted orbitals, DMRG has to be combined with an approach that can capture these contributions, including perturbation [292, 293], coupled-cluster [294], or other methods [295, 296].

(iv) Nuclear structures

The DMRG scheme also provides a good practical truncation strategy for large-scale nuclear structure calculations. It works in the framework of nuclear shell models with an effective Hamiltonian in which a nucleus is modeled by filled core shells and partially filled valence orbitals of protons

and neutrons. The core is assumed to be inert. This significantly reduces the dimension of the active shell subspace.

The potential of DMRG in the nuclear structure calculation was first exemplified in a large angular momentum shell interacting through a pairing and a quadrupolar force in an oblate nucleus [58]. Several schemes were then proposed to implement DMRG in different symmetry channels, including the particle-hole [297], the z-component of the total angular momentum [298], and the total angular momentum symmetries [299, 300]. Their applications led to accurate treatment for quite a number of nuclei, including ^{28}Si [301], ^{56}Ni [298, 301], and ^{48}Cr [299, 301]. Further improvement to the accuracy could be achieved by optimally arranging the order of the proton and neutron orbitals according to the criterion that minimizes the sum of the distance between any two orbitals weighted by their quantum mutual information [302]. It reduces the DMRG error for the ground-state energy of ^{56}Ni by one order of magnitude [301, 302].

Classical Statistical Models

In the formalism of path integral, a d-dimensional quantum system is mapped onto a $(d + 1)$-dimensional classical statistical model. This correspondence between quantum and classical models implies that a method developed in a d-dimensional quantum system can be extended, with certain modifications, to a $(d + 1)$-dimensional classical system and vice versa.

As the partition function of a classical statistical model can be always represented as a network product of local tensors, various tensor network RG methods, including TMRG [68, 70, 303], coarse-graining TRG [128, 133, 134], and SRG [129, 131], have been developed in studying these systems in the past decades. Among them, TMRG [68] is a direct generalization of DMRG. In a classical system, the transfer matrix plays a similar role to the Hamiltonian in a quantum system. TMRG [68] is a method for diagonalizing transfer matrices based on the reduced thermal density matrix. As an exactly soluble system, the Ising model often serves as an ideal system for testing each method.

Tensor network methods have been applied to investigate nearly all two-dimensional classical statistical models, especially their critical behaviors around the phase transition points. This includes, for example, the Ising model [303, 304], Potts model [305, 306, 307, 308], clock model [309, 310, 311], vertex model [312, 313], self-avoiding-walk model [314], XY model [315, 316], Heisenberg model [317], and lattice gauge models with different symmetries [203, 204]. A number of three-dimensional classical statistical models have also been studied with tensor network methods [110, 131, 135, 318, 319, 320].

Quantum Information

The reduction of the Hilbert space carried out in DMRG or other tensor network renormalization group methods is a problem of quantum data compression. In the language of quantum information, it is to find an optimal scheme to minimize the quantum information loss or to carry out an optimal lossy quantum data compression. The von Neumann entropy, or the entanglement entropy, is the most fundamental measure in quantum information. It quantifies the nonlocal quantum correlations between a subsystem and its complement. In fact, the entanglement entropy was used in the DMRG calculations even before this terminology was formally introduced into the field of DMRG [34, 258].

By exploring the entanglement structures of quantum states, we understand why tensor network states approximate ground states of quantum lattice models with short-range interactions so well [47, 111]. At the heart of this insight is the area law of entanglement entropy [37], which confines the space of physical interest to a small corner of the whole Hilbert space. In other words, the area law places strong bounds on quantum entanglement that a many-body system can generate in its ground state. It translates directly to the amount of memory and time required to compute a quantum state in actual calculations.

In one dimension, if the entanglement of a bipartite system is bounded or grows logarithmically with its size, an efficient simulation with tensor network methods is possible. This explains why DMRG works so well in one dimension because the entanglement of one-dimensional ground states is bounded by the area law in a gapped system or grows logarithmically with the lattice size in a critical system [47]. However, the simulation of the time evolution may not be efficient even in one dimension since the bipartite entanglement grows linearly with time.

In two dimensions, the boundary grows with the system size. It is impossible to catch this fast-growing entanglement by using an MPS or other one-dimensional tensor network state to represent a ground-state wave function when the system size becomes sufficiently large. To resolve this problem, a two-dimensional representation of quantum states, like PEPS [111] or PESS [114], has to be introduced to spread the entanglement onto all bonds on the boundary.

Quantum information also plays a vital role in the DMRG calculations of quantum lattice models with nonlocal interactions, such as the Hubbard model in momentum space [56], and quantum chemical systems [57]. In treating nonlocal quantum models with DMRG, the ordering of the lattice sites or molecular orbitals seriously affects the accuracies of converged results. An optimal scheme is to order the single-particle states or orbitals by minimizing their overall distances determined by their mutual quantum information [258].

On the other hand, tensor networks provide a natural framework for a fundamental understanding of inherent entanglement structures of physical systems. Particularly, tensor networks, such as PEPS, could be used to explore the relationship between the edge and bulk states through the so-called entanglement Hamiltonians [321]. The entanglement Hamiltonian H_e is defined from the reduced density matrix. Its eigenvalues encode essential information about the boundary states. If the system is gapped and not topologically ordered, H_e is usually a boundary Hamiltonian with local interactions. In a critical system, however, H_e is generally a Hamiltonian with long-range interactions.

Tensor networks also provide a neat and unique representation for classifying topological phases in a gapped system [322]. As a short-range correlated state or a gapped state is well described by an MPS in one dimension, it was shown that all quantum states are equivalent to trivial product states in the absence of any symmetry [323]. This means no topological order without symmetry protection in one dimension. A topologically nontrivial phase may exist in a phase space restricted by symmetries. This symmetry-protected topological state is short-range entangled. However, by definition, an intrinsic topologically ordered state is long-ranged entangled.

Machine Learning

Tensor networks were initially introduced to provide an efficient tool for attacking quantum many-body problems. One emerging application direction of tensor networks in a seemingly unrelated field is machine learning. Deep neural networks can characterize complex learning tasks such as image classification or speech recognition. The reason why the neural networks work so well could be understood using the language of quantum entanglement [324]. On the other hand, quantum wave functions could be modeled by making use of fully connected neural networks and restricted Boltzmann machines [325]. It was shown that a Boltzmann machine is essentially equivalent to a tensor network [326].

Tensor networks, including MPS [327], PEPS [328] and the tree tensor network [329, 330], provide a natural way to parameterizing machine learning models. It was used for a variety of tasks, such as supervised learning for images with MPS [327, 331], MPO [332], or PEPS [328], mixed unsupervised and supervised learning with multiscale tensor networks [329]. Compared with other machine learning approaches, tensor networks offer clear theoretical insight and interpretation, more sophisticated training algorithms and strategies originally developed for solving quantum many-body problems, a dramatic reduction in memory needed, and a few other advantages.

https://doi.org/10.1017/9781009398671.002 Published online by Cambridge University Press

However, caution should be taken in applying tensor networks to machine learning since neural networks are characterized by nonlinear functions, whereas tensor networks are generally linear. It is still unclear how the entanglement is embedded in the data set. In a study on the supervised image classification using the MNIST data set of handwritten digits, it was found that the entanglement grows very fast with the number of images in the training data set, and the bond dimension of MPS used for training should be as high as the image number in order to capture the actual image structures [333]. Nevertheless, other kinds of tensor networks can be exploited to resolve this problem by block isolating or squeezing entanglement in these kinds of data sets [333].