

Lecture Notes in Physics

Editorial Board

R. Beig, Vienna, Austria
J. Ehlers, Potsdam, Germany
U. Frisch, Nice, France
K. Hepp, Zürich, Switzerland
R. L. Jaffe, Cambridge, MA, USA
R. Kippenhahn, Göttingen, Germany
I. Ojima, Kyoto, Japan
H. A. Weidenmüller, Heidelberg, Germany
J. Wess, München, Germany
J. Zittartz, Köln, Germany

Managing Editor

W. Beiglböck
Springer-Verlag, Physics Editorial Department
Tiergartenstrasse 17, D-69121 Heidelberg, Germany

Springer

*Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo*

The Editorial Policy for Proceedings

The series Lecture Notes in Physics reports new developments in physical research and teaching – quickly, informally, and at a high level. The proceedings to be considered for publication in this series should be limited to only a few areas of research, and these should be closely related to each other. The contributions should be of a high standard and should avoid lengthy redraftings of papers already published or about to be published elsewhere. As a whole, the proceedings should aim for a balanced presentation of the theme of the conference including a description of the techniques used and enough motivation for a broad readership. It should not be assumed that the published proceedings must reflect the conference in its entirety. (A listing or abstracts of papers presented at the meeting but not included in the proceedings could be added as an appendix.)

When applying for publication in the series Lecture Notes in Physics the volume's editor(s) should submit sufficient material to enable the series editors and their referees to make a fairly accurate evaluation (e.g. a complete list of speakers and titles of papers to be presented and abstracts). If, based on this information, the proceedings are (tentatively) accepted, the volume's editor(s), whose name(s) will appear on the title pages, should select the papers suitable for publication and have them refereed (as for a journal) when appropriate. As a rule discussions will not be accepted. The series editors and Springer-Verlag will normally not interfere with the detailed editing except in fairly obvious cases or on technical matters.

Final acceptance is expressed by the series editor in charge, in consultation with Springer-Verlag only after receiving the complete manuscript. It might help to send a copy of the authors' manuscripts in advance to the editor in charge to discuss possible revisions with him. As a general rule, the series editor will confirm his tentative acceptance if the final manuscript corresponds to the original concept discussed, if the quality of the contribution meets the requirements of the series, and if the final size of the manuscript does not greatly exceed the number of pages originally agreed upon. The manuscript should be forwarded to Springer-Verlag shortly after the meeting. In cases of extreme delay (more than six months after the conference) the series editors will check once more the timeliness of the papers. Therefore, the volume's editor(s) should establish strict deadlines, or collect the articles during the conference and have them revised on the spot. If a delay is unavoidable, one should encourage the authors to update their contributions if appropriate. The editors of proceedings are strongly advised to inform contributors about these points at an early stage.

The final manuscript should contain a table of contents and an informative introduction accessible also to readers not particularly familiar with the topic of the conference. The contributions should be in English. The volume's editor(s) should check the contributions for the correct use of language. At Springer-Verlag only the prefaces will be checked by a copy-editor for language and style. Grave linguistic or technical shortcomings may lead to the rejection of contributions by the series editors. A conference report should not exceed a total of 500 pages. Keeping the size within this bound should be achieved by a stricter selection of articles and not by imposing an upper limit to the length of the individual papers. Editors receive jointly 30 complimentary copies of their book. They are entitled to purchase further copies of their book at a reduced rate. As a rule no reprints of individual contributions can be supplied. No royalty is paid on Lecture Notes in Physics volumes. Commitment to publish is made by letter of interest rather than by signing a formal contract. Springer-Verlag secures the copyright for each volume.

The Production Process

The books are hardbound, and the publisher will select quality paper appropriate to the needs of the author(s). Publication time is about ten weeks. More than twenty years of experience guarantee authors the best possible service. To reach the goal of rapid publication at a low price the technique of photographic reproduction from a camera-ready manuscript was chosen. This process shifts the main responsibility for the technical quality considerably from the publisher to the authors. We therefore urge all authors and editors of proceedings to observe very carefully the essentials for the preparation of camera-ready manuscripts, which we will supply on request. This applies especially to the quality of figures and halftones submitted for publication. In addition, it might be useful to look at some of the volumes already published. As a special service, we offer free of charge L^AT_EX and T_EX macro packages to format the text according to Springer-Verlag's quality requirements. We strongly recommend that you make use of this offer, since the result will be a book of considerably improved technical quality. To avoid mistakes and time-consuming correspondence during the production period the conference editors should request special instructions from the publisher well before the beginning of the conference. Manuscripts not meeting the technical standard of the series will have to be returned for improvement.

For further information please contact Springer-Verlag, Physics Editorial Department 11 Tiergartenstrasse 17, D-69121 Heidelberg, Germany

Ingo Peschel Xiaoqun Wang
Matthias Kaulke Karen Hallberg (Eds.)

Density-Matrix Renormalization

A New Numerical Method in Physics

Lectures of a Seminar and Workshop Held at the
Max-Planck-Institut für Physik komplexer Systeme
Dresden, Germany, August 24th to September 18th, 1998



Springer

Editors

Prof. Dr. Ingo Peschel
Dr. Matthias Kaulke
FB Physik
Freie Universität Berlin
Arnimallee 14, D-14195 Berlin, Germany

Dr. Xiaoqun Wang
Institut Romand de Recherche Numérique en Physique des Matériaux
PPH 333-EPFL, CH-1015 Lausanne, Switzerland

Dr. Karen Hallberg
Centro Atomico Bariloche
8400 Bariloche, Argentina

Library of Congress Cataloging-in-Publication Data.

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Density matrix renormalization : a new numerical method in physics ; lectures of a seminar and workshop, held at the Max-Planck-Institut für Physik Komplexer Systeme, Dresden, Germany, August 24th to September 18th, 1998 / Ingo Peschel ... (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 1999
(Lecture notes in physics ; Vol. 528)
ISBN 3-540-66129-8

ISSN 0075-8450
ISBN 3-540-66129-8 Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 1999
Printed in Germany

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by the authors/editors
Cover design: *design & production*, Heidelberg

SPIN: 10731988 57/3144 - 5 4 3 2 1 0 – Printed on acid-free paper

Preface

This book deals with a method which is still not widely known, although it caused a revolution in the numerical treatment of low-dimensional quantum systems. How this started about ten years ago, is described in detail in a separate contribution. The reason for the impact it had, on the other hand, is easy to give. Density-matrix renormalization has two features which normally do not come together. With it, one can treat large systems containing, for example, several hundred spins and, at the same time, reach spectacular accuracies such as ten decimal places in ground-state energies. Thus the method combines the advantages of Monte Carlo calculations and of exact diagonalisations and it is not surprising that it was applied soon after its invention to a variety of problems, both old and new, and mostly connected with quantum chains. Parallel to that, it was extended in other directions, for example to finite temperatures, which increased the area of applications further. The number of groups using it grew also and in 1997 about sixty articles appeared in which it played an essential role.

This seemed a good time to bring the experts in the field together. On the one hand to summarize the status and to discuss open questions as well as future directions as, for example, the extension of the method to higher dimensions. On the other hand to share their experience with younger researchers interested in learning and applying the technique, including practical work on the computer. This was the concept of a four-week seminar and workshop at the Max-Planck-Institut für Physik komplexer Systeme in Dresden in the fall of 1998. It was the first meeting on density-matrix renormalization, and among its about 50 participants from all over the world were the leaders in the field. The workshop was viewed as very successful and the proposal to collect the contributions in the form of lecture notes was strongly supported from all sides, even though this had not been planned initially. This has led to the book which is presented here.

The volume is divided into two parts of roughly equal size which are closely related. The first one contains a series of seven longer lectures which present the method from various angles. This serves as a general introduction on the level of an advanced textbook. A central chapter is the second one which explains the basic principles and also contains, in line with the aims of the workshop, a computer program for a simple example. Other chapters deal with more advanced aspects, like the application to time-dependent problems or to finite temperatures. Two-dimensional systems, both classsical and quantum-mechanical, are also discussed. Finally, to put things into a broader perspective, there are contributions on Wilson's original renormalization method as well as on variational approaches.

The second part deals with the application of the technique to various physical problems, arranged in five different groups. The main themes of

the first part appear again, but also some additional aspects. However, the contributions are shorter and the focus is somewhat different. The physical questions and properties stand in the center and the method is in general only briefly discussed, except in those cases where new aspects come in. In this way, the second part is also a review of actual one-dimensional problems, interesting questions and their present status. Of course, it reflects the research topics of the participants, but it covers a large portion of the applications and thus gives a very good overview even though the additional posters are not included.

Throughout the book considerable effort has been made to obtain a clear, structured and self-contained presentation of the material, but also as much coherence as possible. Due to the structure, a certain amount of duplication in the references was inevitable, but we think that it can be accepted. We thank all authors for their, sometimes enthusiastic, cooperation in the project. We are particularly glad that Steven White agreed to write down his personal recollections on how it all began.

Of course, the whole undertaking would have been impossible without the Max-Planck-Institut in Dresden and its outstanding facilities. We thank its director, Peter Fulde, and its staff for their constant support, and also Springer-Verlag which agreed to publish the book in this well-known series. We hope that, besides being a source for everyone interested in the topic, it will also stimulate new research and thereby contribute to the further development of this field.

Berlin, April 1999

Ingo Peschel
Xiaoqun Wang
Matthias Kaulke
Karen Hallberg

Contents

Preface v

How It All Began: A Personal Account
S. White vii

Part I Introductory Lectures

1 Wilson's Numerical Renormalization Group
T. Costi 3

2 The Density Matrix Renormalization Group
R. Noack and S. White 27

3 Thermodynamic Limit and Matrix-Product States
S. Rommer and S. Östlund 67

4 A Recurrent Variational Approach
M. Martín-Delgado and G. Sierra 91

5 Transfer-Matrix Approach to Classical Systems
T. Nishino, K. Okunishi, Y. Hieida
T. Hikihara and H. Takasaki 127

6 Quantum Transfer-Matrix and Momentum-Space DMRG
T. Xiang and X. Wang 149

7 Calculation of Dynamical Properties
X. Wang, K. Hallberg and F. Naef 173

Part II Physical Applications

1 Chains and Ladders

1.1 Properties of the Hubbard Chain
R. Noack, S. Daul and S. Kneer 197

1.2 Soliton Bound-States in Dimerized Spin Chains
E. Sørensen, I. Affleck, D. Augier and D. Poilblanc 211

1.3 Haldane Phase, Impurity Effects and Spin Ladders
X. Wang 221

1.4 Spin-Chain Properties
K. Hallberg 231

2 Molecules and Polymers	
2.1 Electronic Structure using DMRG	
<i>Steven R. White</i>	237
2.2 Symmetrized DMRG Method for Conjugated Polymers	
<i>S. Ramasesha and K. Tandon</i>	247
2.3 Conjugated One-Dimensional Semiconductors	
<i>W. Barford and R. Bursill</i>	261
2.4 Strongly Correlated Complex Systems	
<i>S. Qin, J. Lou, Z. Su and L. Yu</i>	271
3 Classical Statistical Physics	
3.1 Non-Hermitian Problems and Some Other Aspects	
<i>I. Peschel and M. Kaulke</i>	279
3.2 Walls, Wetting and Surface Criticality	
<i>E. Carlon</i>	287
3.3 Critical Two-Dimensional Ising Films with Fields	
<i>A. Drzewiński</i>	295
4 Thermodynamic Properties	
4.1 One-Dimensional Kondo Lattices	
<i>N. Shibata</i>	303
4.2 Impurities in Spin Chains	
<i>S. Rommer and S. Eggert</i>	311
4.3 Thermodynamics of Ferrimagnets	
<i>U. Schollwöck</i>	321
4.4 Thermodynamics of Metallic Kondo Lattices	
<i>S. Moukouri and L. Caron</i>	329
5 Phonons and Disorder	
5.1 Methods for Electron-Phonon Systems	
<i>E. Jeckelmann, C. Zhang and S. White</i>	337
5.2 Disordered One-Dimensional Fermi Systems	
<i>P. Schmitteckert</i>	345

How It All Began: A Personal Account

Steven R. White

Department of Physics and Astronomy, University of California
Irvine, CA 92697-4575, USA

In this chapter I'd like to take the opportunity to relate some of the history leading up to the development of DMRG. I think it is always interesting to look back, from a perspective where we know what works and what doesn't, and see the sometimes tortured path that got us here. The history I present here is a personal one: the work I was involved in, and the development of my ideas, which eventually led me to develop DMRG. I am not attempting to give proper credit to others who may have had similar ideas earlier, which an ordinary review should do – for that I refer to other chapters in this book. I will include many incidental details which one would not find in a review. Some of the events that happened along the way I found quite amusing, and I hope you will too.

Much of the development of DMRG traces back to my years as a graduate student at Cornell, from 1982 to 1987. I had first wanted to work for Michael Fisher, who worked on statistical mechanics and phase transitions, but during the first year that I would be working on my thesis, he was going to be on sabbatical at Caltech and so he wasn't taking any new students.

During the fall of 1983, the start of my second year at Cornell, Ken Wilson announced a set of lectures on "Chemistry". In these lectures, Wilson described an approach to solving the electronic structure of atoms and molecules using renormalization group and other ideas borrowed from particle physics and statistical mechanics. None of Wilson's approach had been carried out beyond the planning stages, and it is a testament to Wilson's prestige at the time that he could announce a set of lectures on what he hoped to do on a very difficult problem, and have everyone come. Wilson was strongly influenced by his father, E. Bright Wilson, a noted chemist at Harvard, who, among other things, wrote one of the first quantum mechanics books with a chemistry orientation in collaboration with Linus Pauling. This influence led Wilson to have as a long term goal the development of new electronic structure techniques, and Wilson's only publication during his first several years as an assistant professor at Cornell was on "Model Atomic Hamiltonians" [1], in which a lattice model was introduced for first and second row atoms.

I had worked for a chemist for three years as an undergraduate at UC San Diego, doing molecular dynamics simulations. (Coincidentally, this chemist's name is Kent Wilson. Kent, like Ken, is an unusual and remarkable person.) I thought Wilson's plan sounded like a good thing to work on, and I approached him about being his student. I remember one of the first things he said: "You

know there will be some computer work involved?” I assured him that that was right up my alley. However, he told me that he was only going to have shared students, not students of his own, so I asked John Wilkins, who worked on a variety of topics in solid state theory, if he would be my other thesis advisor. He agreed, and so I was a shared student with Wilson and Wilkins, and the eventual title of my thesis was “New methods for electronic structure calculations.”

The main ideas of Wilson’s plan were: 1) One can use RG language to describe very weak bound states of model “atoms”. 2) With Grassmann algebra, one can integrate out fermions in an imaginary time path integral, and thus apply renormalization group ideas to electronic structure. This involved setting up a fine real-space lattice around a molecule, and then applying block-spin RG methods, using Grassmann integration, to “decimate” the lattice until the lattice spacing was comparable to an atomic radius. At that point, the problem would become nonperturbative, and decimation would stop. 3) The final part of the problem would be to apply the Langevin equation as a simulation tool to finish the problem. The Langevin approach had been developed by Klauder and others as a sort of early quantum Monte Carlo method. The effective action from the blocking procedure would be used as input to the Langevin equation.

When I started, I really didn’t understand the big picture of what we were doing, but I was willing to dive in right away and learn as I went along. I spent a year working on a toy model which tested the block-spin approach with Grassmann variables, and then I spent a year working on the Langevin approach. This second year was actually spent at UC Santa Barbara – Wilkins had taken six students with him to spend a year there. At the end of this time, however, we realized the entire project was a failure. We ran into two serious, seemingly insurmountable problems: first, we weren’t able to devise an effective method for integrating out core electrons in the Grassmann RG approach; and second, we found that the Langevin approach became numerically unstable and also gave the wrong answer, due to something similar to the fermion minus sign problem in quantum Monte Carlo.

I spent my last year at Cornell trying out some other approaches to electronic structure. Wilson didn’t seem especially interested in these, and so I mostly talked to Wilkins about them. One of these was a variation on the configuration interaction method used in quantum chemistry. This approach is interesting in retrospect because a little of the flavor of DMRG is there: the wavefunction was expanded in an optimal way using the largest eigenvalues and eigenvectors of a matrix. I also worked on a variation of quantum Monte Carlo in the canonical ensemble and in applying the finite element method to electronic structure. Wilson continued to be interested in electronic structure methods in general, and he attended a regular group meeting of mostly Wilkins people in which this was discussed.

In the fall of my last year at Cornell, Wilson gave an interesting talk on the application of his numerical RG approach, which he developed for the Kondo impurity problem, to 1D lattice systems. This approach used a Hamiltonian, as opposed to an Lagrangian, and there was no extra time dimension. Another student of Wilkins at that time was Barbara Jones, who was applying this approach to two Kondo impurities. I had talked with Barbara about this method and thought it was very interesting. After Wilson had applied it to impurity problems, others applied it to 1D lattice systems, where it was called “real space RG”. Wilson did not think very highly of real space RG, and in this talk Wilson used a toy model – a discretized particle in a 1D box – to analyze why it failed. Wilson showed how the truncation of the interactions at the edge of a block was unjustified, and led to nonsensical results. In pictorial terms, ignoring the interactions between blocks means that the retained states vanish at the edge of the blocks, and any subsequent states at later iterations always have kinks where the two blocks were joined. How would one fix this problem? Wilson suggested correcting for the neglected terms using a complicated form of perturbation theory. However, he concluded that this would be very difficult to carry out even in the 1D noninteracting case.

Perhaps Wilson was hoping someone would try out his ideas for fixing real space RG, but it was to be five years before this happened.

I had gotten to know Doug Scalapino, Bob Sugar, and their students quite well during our year at UCSB, and in September of 1987 I went back to UCSB for a post doc with Scalapino. There I focused on quantum Monte Carlo (QMC), Hubbard models, and in general the field of strongly correlated electrons and high temperature superconductivity, which I didn’t know very much about at the start. We made good progress in improving the Monte Carlo methods, and produced some influential papers on the Hubbard model. For example, we found some of the first indications that d-wave pairing, rather than s-wave pairing, was prevalent. After two years as a post doc, I went to UC Irvine as an Assistant Professor.

While still at UCSB and for my first year or two at UCI I worked on methods for extracting dynamics from Monte Carlo. Some of our early results using a method I developed prompted Richard Silver and Jim Gubernatis to apply the maximum entropy method, which turned out to be superior. I switched to using maximum entropy coupled with QMC, and studied some of the dynamical properties of the Hubbard model. However, I also wanted to become more familiar with exact diagonalization methods, and a student and I developed a Lanczos program to study dynamics on 4×4 Hubbard systems.

During this period, I would occasionally try out some new idea to solve the fermion minus sign problem, the main bottleneck in the QMC method for fermions. Each of these attempts failed. (I would also occasionally hear about attempts by others to fix the minus sign problem, with at best only partial success.) For example, my student Marco Vekić and I tried a variation on the

world-line QMC method in which the basic unit was a block of sites rather than a single site. We found that this actually made the minus sign problem much worse – Marco found that there was a new type of “internal” minus sign which isn’t there when you work with single sites! I began to conclude that the minus sign problem was really quite fundamental, and not likely to allow any easy solutions. I was also getting discouraged by the limitations of QMC on strongly-coupled systems, such as very slow evolution through phase space, large statistical errors, and poor scaling with system size, even in cases when there was no minus sign problem. This led me to once again think about numerical RG methods. (Since then, there has been more progress in dealing with the minus sign problem, but it is still a fundamental difficulty.)

I was not tempted to return to the Grassmann imaginary time formulation I had used as a graduate student. My work with quantum Monte Carlo had led me to conclude that it was very hard, and not very natural, to reach zero temperature by making the time direction of the lattice very long. Furthermore, I was very impressed with how the Lanczos method in exact diagonalization could pick out the ground state even with a very small gap to the first excited state. So, I began thinking back to Wilson’s Hamiltonian-based numerical RG. Besides working at zero temperature directly, this approach allowed one to deal directly with composite particles or states: for example, suppose all of the electrons in a particular system (say, one with an attractive on-site interaction) were tightly bound in pairs. In the Hamiltonian approach, most of your states describing the Hilbert space would consist of pairs, with little trace of the original electrons. In the Grassmann approach, you would have to work with the original electrons, which seemed to me to be severe drawback.

However, because of the problems with real space RG, I thought that it might be promising to try to solve a finite Hubbard lattice using the numerical RG method in momentum space. In other words, one sets up a “lattice” where each site represents a particular allowed momentum, and the states of the site indicate whether a particle with that momentum is present. Another reason for trying out the momentum space RG was previous experience: my Lanczos work with a student on Hubbard systems had been in momentum space, in order to implement momentum conservation in a simple way. I carried out this approach on a 4×4 Hubbard system, and wrote a paper on it. I concluded that the results were pretty reasonable for energy differences, although total energies were poor. I wanted to know what Wilson thought of this work, and I sent him some email and a copy of the paper [2].

Wilson did not seem overly impressed. He stated emphatically that the particle-in-a-box problem of real space RG needed to be solved first before trying out many particle systems. He also suggested the perturbation approach to fix it, which he had mentioned in his talk at Cornell. He said I should talk to Wilkins and one of his post docs, who had also asked him recently about the particle-in-a-box problem again.

I took Wilson's opinions very seriously. I contacted Wilkins, who referred me to his post doc Kevin Ingersent. I proposed to Kevin, who had some additional notes on Wilson's perturbation approach, that we try out Wilson's ideas together.

Kevin and I tried Wilson's perturbation theory and many other variations, as many as we could think of. This was during the early spring of 1991. All of these approaches failed – in fact, the results were almost always worse than not doing the perturbative corrections at all. We ran out of ideas, concluding it was hopeless, and gave up. In retrospect, I know that the starting point for the perturbation theory was just too poor.

When the summer came, I couldn't resist thinking again about the particle-in-a-box problem, and thought more about boundary conditions(BCs). I knew that simply changing the BCs didn't fix the problem, although it changed what the errors looked like: for example, with fixed BCs, you had problems with slopes at the edges of a block, while with free BCs you got discontinuities. Trying a single BC which was in between fixed and free was possible, but it didn't work any better. Finally I got the idea of trying two different BCs at the same time – getting some states from one BC, some from another. I was still thinking in terms of fixed and free BCs, and it seemed that based on what the wavefunctions for each boundary condition looked like, this might work much better—you could combine the different states to get rid of the various discontinuities. I immediately tried out the “fixed-free” approach numerically – I always try things out numerically very soon after I get an idea, since I enjoy programming and trying out the idea more than trying to work out the details on paper. The test consisted of comparing energies of low lying states after ten RG iterations with exact results.

When I saw the results from the first fixed-free test, I was astonished! Previously, the results from all my numerical tests were garbage, orders of magnitude off from the exact solution. Not only had the fixed-free method worked, the results now seemed exact! The results agreed to at least ten digits, keeping only eight states, and I didn't know whether the minuscule differences were simply the results of round off error. (Some time later, I carried out the calculations again in quadruple precision, and found that the results were good only to ten digits, no more.)

I played around with this for just a short while more, and then I returned to thinking about interacting systems. I felt that this had indeed given me the key to understanding the many particle problem as well as the single particle one. I first tried out the fixed free technique on Heisenberg spin chains. I ended up trying several variations of this approach, including the antiperiodic/periodic combination of boundary conditions, without success. However, this led me to the conclusion that I needed extra boundary sites, which were not part of the block, in order to induce the right boundary conditions. This led to the idea of solving a “superblock” system, and then projecting the superblock state onto the block. I devised a projection method, which was

conceptually similar to the optimal configuration interaction approach from my thesis. The status of this work, as of September 1991, was written up in a paper for a conference in Osaka [3]. Already I was becoming confident that this was going to be a very promising approach, and I said so at the conference.

In September, 1991 Reinhard Noack came to work for me as a post doc. Reinhard had been a student of Scalapino, where he worked on quantum Monte Carlo calculations of a two chain Hubbard model. Reinhard had had a lot of trouble in studying the two chain system because of the minus sign problem. I suggested to him that he take over the particle-in-a-box work and turn it into a paper. He got going very quickly on it and the paper was written by December [4]. Subsequently, Reinhard worked on applying these single particle RG methods to problems in Anderson localization. (Eventually, Reinhard returned to studying the two chain Hubbard model, this time using DMRG and with much more success.)

In the late fall or early winter of 1992 I returned to the interacting problem and realized that my projection idea of the superblock onto a block was actually equivalent to a singular value decomposition:

$$\psi_{ij} = U_{i\alpha} D_{\alpha\alpha} V_{\alpha j} \quad (1)$$

where i denotes states of the block, and j denotes states of the remainder of the superblock. I had taught graduate statistical mechanics a few times since coming to UCI, for which I had read parts of Feynman's Lectures on Statistical Mechanics [5]. One of my favorite parts of the book was his introduction to density matrices. I soon realized that performing the singular value decomposition was equivalent to diagonalizing a density matrix:

$$\rho_{ii'} = \psi_{ij} \psi_{i'j} = U_{i\alpha} D_{\alpha\alpha}^2 U_{i'\alpha}. \quad (2)$$

This made me very happy: now the technique was becoming quite elegant.

I then began numerical tests and more extensive program development. One early test was to diagonalize a finite Hubbard chain, split it in half, and look at the density matrix eigenvalues for the two sides. This behaved quite nicely. I then began writing a program to do the infinite-system DMRG method for the Heisenberg spin 1/2 chain, for which there is the exact Bethe ansatz solution. I was getting odd-even bond dimerization, so it didn't look like it was working well. However, I remember checking the average energy over two adjacent bonds and being very surprised at how close that was to the exact result – about 5 or 6 digits were correct! It was at that point that I realized things could work very well with this method even with interacting systems.

At that point I had a good understanding of all the pieces of the puzzle. I figured out how the finite system method should work. When I got that program working, and found it worked extremely well, I remember being most concerned about how I would explain this method to others – it was

completely unlike any other numerical method I had heard of. I felt the finite system method would look very strange and ad hoc, and I dreaded trying to write a paper to explain the algorithm and explain why, for example, there were two sites in the middle in the standard superblock configuration. It was not that I couldn't explain why, it was just that the explanation would be complicated, as would the description of the algorithm in general. I tried out periodic boundary conditions, found that they didn't work as as open BCs, and also developed an intuitive understanding of why that was.

I wanted a good problem to demonstrate the technique on and I settled on the spin-1 Heisenberg chain, where I got some nice results. I wanted to publicize the method a little. I had a family trip arranged to the East coast in May, and so I contacted friends at Bell Labs and IBM Yorktown Heights, telling them I had some results from an exciting new method and I wanted to give a talk. (I remember Andy Millis at Bell Labs introducing me, saying something like "This is Steve White who has some new results that *he* is very excited about...") The reception I got at Bell was very enthusiastic, and I started a collaboration with David Huse to study the spin-1 Heisenberg chain in more detail [6]. (At IBM, however, no one seemed particularly interested – they weren't working on those kinds of problems.)

I'd like to finish this chapter with some stories about the first DMRG-related papers and some interactions with the referees. For the particle-in-a-box paper [4], with Reinhard Noack, we felt that the editors would have a hard time finding someone to referee the paper, since this was not an active field. We suggested in our submittal letter to Physical Review Letters several people as referees, including Ken Wilson and Patrick Lee, who had worked in numerical RGs in the past. After some time we got a very sensible referee report back, pointing out that while this was very interesting, it was still just for a single particle, and might not generalize to many particle systems. Therefore, the referee recommended that it appear in Physical Review B, Rapid Communications. I thought this was completely reasonable and we told the editors to put it in Rapid Communications.

Very soon after I got a very surprising letter from the editors of PRL. It said that because of a late referee's report, they were now going to accept it for PRL. (We could also still send it to Rapid Communications if we wanted.) This didn't make any sense to me – if you get one good and one bad report, you always have to resubmit, and then maybe you get the paper in because the third referee likes it. So what happened is very unusual – I call it *Spontaneous Unrejection*. Then I read the new referee's report, and I realized why. The report was incredibly insightful. After saying that we had solved an important (toy) problem, and saying that the paper should be accepted, it went on to speculate on the prospects for applying the methods to many particle systems. It said that the fixed-free method probably wouldn't work for interacting systems, since the boundary conditions would apply to all particles identically, whereas it would need to apply in different ways to

different particles. It said that the superblock method might work, but that putting three or more blocks together would be very hard to diagonalize. Now, I had already come to these conclusions (or very similar ones) after months of further work. I realized that this report had to be from Wilson – only he had this level of insight about this problem – and that this explained why the editor would now accept the paper: Wilson was the final word in RG methods.

A few months after this exchange I submitted to PRL the original DMRG paper [7]. I got back one positive and one negative report. Both reports were critical of the presentation, which I was sympathetic with – I found it very hard to explain the method in a modest number of words. I rewrote it, taking out a description of the finite system method, which was too brief to make sense, and resubmitted. I got back a third report, which said, “I agree with Referee B. This paper is not appropriate for publication in Physical Review Letters. The reason is as follows. This paper is making a calculation on the Heisenberg linear chain problem. In my opinion, the method is not very new and the above problem has been widely worked in connection with Haldane problem. Hence this paper is inappropriate for publication in PRL...”

I have never appealed a PRL decision before or since, but in this case I did. I asked it be sent to a divisional editor, and I believe they sent it to Peter Young. The divisional editor accepted it at once.

It took me quite a while to write the longer paper describing the algorithms in detail [8]. I also had a hard time getting started because it seemed like too much work and I enjoyed applying DMRG to problems more than I liked writing. However, the long paper did finally get written, and shortly thereafter DMRG began to be used by a number of other groups.

I have been absolutely delighted by the marvelous work done by many other groups using DMRG since then. One of the biggest compliments to a scientist is to have his work used by others, and seeing all the uses, improvements, and extensions to DMRG, as discussed in the rest of this book, is a source of great pleasure to me.

References

1. K.R. Piech and K.G. Wilson, *J. Chem. Phys.* **48**, 3740 (1968)
2. S.R. White, *Phys. Rev. B* **45**, 5752 (1992).
3. S.R. White, in *Computational Approaches in Condensed-Matter Physics*, edited by S. Miyashita, M. Imada, and H. Takayama, Springer, p 97 (1992)
4. S.R. White and R.M. Noack, *Phys. Rev. Lett.* **68**, 3487 (1992)
5. R.P. Feynman, *Statistical Mechanics: A Set of Lectures*, Benjamin (1972)
6. S.R. White and D.A. Huse, *Phys. Rev. B* **48**, 3844 (1993)
7. S.R. White, *Phys. Rev. Lett.* **69**, 2863 (1992)
8. S.R. White, *Phys. Rev. B* **48**, 10345 (1993)

1 Wilson's Numerical Renormalization Group

Theo Costi

Theoretische Physik III, Universität Augsburg
D-86135 Augsburg, Germany

The idea of the numerical renormalization group (NRG) for a quantum-mechanical system with Hamiltonian H is to obtain the many-body eigenstates and eigenvalues on all energy scales $\omega_1 > \omega_2 > \dots$ in a sequence of steps, with each step corresponding to a distinct energy or length scale [1]. This is achieved by a formal procedure of tracing out high energy states to give effective Hamiltonians $H_N, N = 0, 1, \dots$ describing the physics on successively lower energy scales $\omega_N, N = 0, 1, \dots$. The renormalization group (RG) transformation R relates effective Hamiltonians describing the physics on successive energy scales : $H_{N+1} = R[H_N]$. Although the idea is straightforward, its practical implementation turns out to be a difficult task, especially if one wants the RG transformation to give the eigenstates and eigenvalues of H accurately down to the lowest energies.

A complete implementation of the method for a quantum mechanical problem was given by Wilson, who succeeded in constructing a non-perturbative RG transformation for the Kondo model valid on all energy scales [1]. This gave the first convincing calculation of the crossover from the high temperature weak coupling regime of the Kondo model to the low temperature strong coupling regime. The same transformation has since been applied to a range of similar quantum impurity models with equal success [2]. On the other hand, it has proven more difficult to develop similar RG transformations for quantum lattice models. For example, calculations using an approximate RG transformation for the one-dimensional Hubbard model [3] showed that the spectra became inaccurate after four RG iterations, corresponding to a maximum system size of 16 lattice sites, clearly not sufficient to access the lowest energies of interest. These difficulties with the real space approaches provided the motivation for the development of the density matrix renormalization group (DMRG) [4,5] described in this book.

In this chapter we describe a specific implementation of the NRG idea, Wilson's non-perturbative NRG method for the Kondo model [1]. This may serve as a useful background for the later chapters on the DMRG. A detailed description of the formalism of the NRG, the analysis of fixed points and full details on the calculation of thermodynamic quantities can be found in [1,2]. We also describe the calculation of dynamic and transport properties of quantum impurity models via the NRG. The outline is as follows: Sect. 1 introduces the Anderson and Kondo models [6], which are used in subsequent

sections to illustrate various aspects of the NRG. The basic ideas of the RG, such as fixed points, crossover scales, RG trajectories and associated universal scaling functions are also discussed there. Wilson's method is described in Sect. 2 and its application to dynamic and transport properties of quantum impurity models in Sect. 3. Finally, Sect. 4 concludes with some remarks on possible future directions using the NRG.

1 Quantum Impurity Models

Examples of quantum impurity models include single and multi-channel Kondo models, the Anderson impurity model and the dissipative two-state system. The simplest versions of these models will be introduced in detail below. They describe a large number of physical systems of current experimental and theoretical interest. Thus, transport through single-electron devices, such as quantum dots, and certain local properties of heavy fermion compounds are modeled by Kondo or Anderson impurity like models. Quantum tunneling between macroscopic fluxoid states in a superconducting quantum interference device [7], tunneling between edge states in the $\nu = 1/3$ fractional Quantum Hall effect [8], two-level atoms coupled to the electromagnetic field in optical fibers [9] and tunneling of defects in solids [10] are examples which can be modeled by the dissipative two-state system [11,12]. The latter is also equivalent to the anisotropic Kondo model [13,14].

The Anderson model

Let us start with the Hamiltonian of the Anderson impurity model, the prototype model of strongly correlated impurity systems. This was introduced in [16] as a microscopic model for local moment formation in non-magnetic metals. Its Hamiltonian is

$$H_{AM} = \varepsilon_d n_d + U n_{d\uparrow} n_{d\downarrow} + \sum_{\mathbf{k}, \mu=\uparrow, \downarrow} (V_{\mathbf{k}d} c_{\mathbf{k}\mu}^+ d_\mu + H.c) + \sum_{\mathbf{k}, \mu=\uparrow, \downarrow} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\mu}^+ c_{\mathbf{k}\mu}$$

The first two terms describe the impurity, which, for simplicity, is represented here by a non-degenerate s-level of energy ε_d . Electrons in the local level are subject to a Coulomb repulsion U which acts between spin-up and spin-down electrons. The local level hybridizes with the Bloch states of a non-interacting s-wave conduction band, the last term in H_{AM} , with amplitude $V_{\mathbf{k}d}$. The properties of the model are determined by the hybridization function $\Delta(\omega) = \pi \sum_{\mathbf{k}} |V_{\mathbf{k}d}|^2 \delta(\omega - \varepsilon_{\mathbf{k}})$, which, like the conduction density of states $\rho(\omega) = \sum_{\mathbf{k}} \delta(\omega - \varepsilon_{\mathbf{k}})$, will in general be a complicated function of energy. For most magnetic impurity systems, where the main interest is in the low energy behaviour, it is a good approximation to set $\Delta(\omega) \approx \Delta(\varepsilon_F) \equiv \Delta$. It can also be shown that a *weak* energy dependence of $\Delta(\omega)$ does not affect the low energy behaviour of the model [2].

For a numerical treatment, it is useful to reformulate the Anderson model in the form of a linear chain model. This allows the model to be iteratively diagonalized by a procedure to be described in Sect. 2. The first step in this procedure will be carried out explicitly below. This simple calculation also shows how the Kondo model arises as the low energy effective model of the Anderson model in the strong correlation limit. We consider only the case of orbital symmetry, $V_{kd} = V_{kd}$ and $\varepsilon_k = \varepsilon_k$. Then, only s-wave conduction states couple to the impurity (for details see Appendix A of [2]), and we can replace $c_{k,\mu}$ by $c_{k,l=m=0,\mu}$ and abbreviate the latter to $c_{k,\mu}$. We first notice that the impurity state in the Anderson model hybridizes with a local Wannier state $|0, \mu\rangle = f_{0,\mu}^+ |vac\rangle$, with $|vac\rangle$ the vacuum state, and $f_{0,\mu}$ given by

$$V f_{0,\mu} = \sum_k V_{kd} c_{k,\mu}. \quad (1)$$

The value of V follows from the normalization $\{f_{0,\mu}, f_{0,\mu}^+\} = 1$

$$V = (\sum_k |V_{kd}|^2)^{1/2}. \quad (2)$$

Using the above local state one can apply the Lanczos procedure for tridiagonalizing a hermitian operator, such as H_c , to obtain

$$H_c = \sum_{k,\mu} \varepsilon_k c_{k,\mu}^+ c_{k,\mu} \rightarrow \sum_{\mu,n=0}^{\infty} \epsilon_n f_{n,\mu}^+ f_{n,\mu} + \lambda_n (f_{n,\mu}^+ f_{n+1,\mu} + H.c.) \quad (3)$$

with site energies, ϵ_n , and hoppings, λ_n , depending only on the dispersion ε_k and hybridization matrix elements V_{kd} through the hybridization function $\Delta(\omega)$ [2]. The Anderson model then takes the linear chain form

$$\begin{aligned} H_{AM} = & \varepsilon_d n_d + U n_{d\uparrow} n_{d\downarrow} + V \sum_{\mu} (f_{0,\mu}^+ d_{\mu} + d_{\mu}^+ f_{0,\mu}) \\ & + \sum_{\mu,n=0}^{\infty} \epsilon_n f_{n,\mu}^+ f_{n,\mu} + \lambda_n (f_{n,\mu}^+ f_{n+1,\mu} + f_{n+1,\mu}^+ f_{n,\mu}) \end{aligned}$$

depicted in Fig. 1. Although, formally, this model looks like the one-dimensional

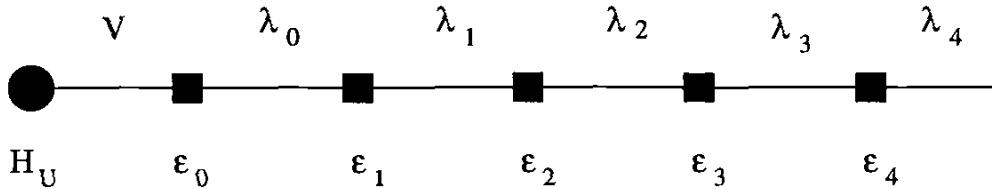


Fig. 1. The linear chain form of the Anderson model. $H_U = \varepsilon_d + U n_{d,\uparrow} n_{d,\downarrow}$. The “site energies” ε_n and “hoppings” λ_n follow from $\Delta(\omega)$.

real-space models treated by the DMRG method in this book, the interpretation here is not in terms of electrons hopping on a one-dimensional lattice

in real-space. Instead, as will become clearer in Sect. 2, each successive site added along the chain corresponds to adding lower energy degrees of freedom, measured relative to the Fermi level. By considering longer chains one can then access lower energies.

A zeroth-order (high-energy) approximation to the spectrum of the Anderson model can be obtained by considering just the coupling of the $n = 0$ Wannier state to the impurity and neglecting all others,

$$H_{AM} \approx H_0 = \varepsilon_d n_d + U n_{d\uparrow} n_{d\downarrow} + V \sum_{\mu} (f_{0,\mu}^+ d_{\mu} + d_{\mu}^+ f_{0,\mu}) \quad (4)$$

There are 16 many-electron states $|n_d, n_0\rangle$, which can be classified by the conserved quantum numbers of total electron number N_{el} , total z-component of spin S_z^{tot} and total spin \mathbf{S} . Using these symmetries we can diagonalize the block matrices H_{N_e, S, S_z}^0 to obtain the many-body eigenstates $|N_{el}, S, S_z, r\rangle$ and the corresponding eigenvalues. They are shown schematically in Fig. 2 for the symmetric case $\varepsilon_d = -U/2$ in the strong correlation limit $U \gg V^2$. The spectrum separates into two groups of states, one group of low energy states lying close to the groundstate with spacings $\mathcal{O}(V^2/U)$ and one group of high energy states lying at energies $\mathcal{O}(U/2)$ higher and also split by $\mathcal{O}(V^2/U)$.

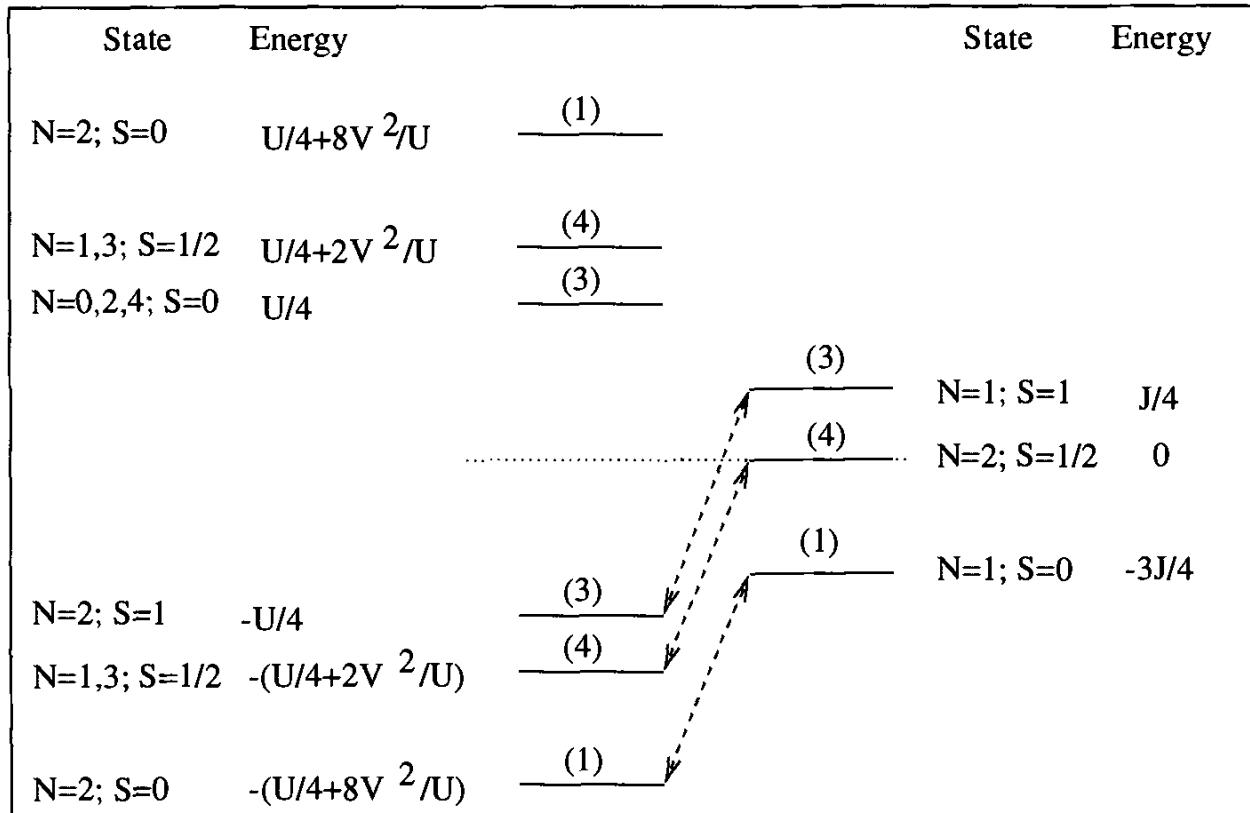


Fig. 2. Zeroth-order approximation to the spectrum of the symmetric Anderson model (left) and the isotropic Kondo model (right). The spectrum of the Anderson model has been shifted by $-U/4$ to show the particle-hole symmetry of the model. The states are labeled by electron number N and total spin S . The numbers in brackets are the degeneracies of the energy levels.

The Kondo model

Consider, now, the (orbitally) isotropic Kondo model,

$$H_{KM} = \sum_{k,\mu} \varepsilon_{k,\mu} c_{k,\mu}^+ c_{k,\mu} + JS \cdot s_0 \quad (5)$$

where \mathbf{S} is the impurity spin and

$$s_0 = f_{0,\mu}^+ \sigma_{\mu\nu} f_{0,\nu}$$

is the conduction electron spin-density at the impurity site with $f_{0,\mu} = \sum_k c_{k,\mu}$ the local Wannier state. This describes a $S = 1/2$ local moment interacting antiferromagnetically ($J > 0$) with the conduction electron spin-density at the impurity. As for the Anderson model, a zeroth-order (high-energy) approximation to the spectrum of the Kondo model is obtained by considering only the coupling of the impurity spin to the $n = 0$ Wannier state,

$$H_{KM} \approx H_0 = JS \cdot s_0 = \frac{J}{2} [(S + s_0)^2 - S^2 - s_0^2]. \quad (6)$$

Its spectrum of 8 eigenstates is also shown in Fig. 2. It consists of a single group of states close to the singlet groundstate with spacings $\mathcal{O}(J)$. This is the same group of states found above for the low energy part of the zeroth approximation to the Anderson model. So at energies, $\omega \ll U/2$, we have that $H_{AM} \approx H_{KM}$, i.e. the low energy behaviour of the Anderson model is governed by the Kondo model with a weak antiferromagnetic exchange interaction J . Identifying the splittings in the spectra for the two models (e.g. that between the lowest two states $E_2^{AM} - E_1^{AM} = 6V^2/U = 3J/4 = E_2^{KM} - E_1^{KM}$) gives the relation between the bare couplings $J = 8V^2/U$ (in agreement with the value found from the Schrieffer-Wolff transformation [17]). Within this zeroth order approximation the spin ($E_3^{KM} - E_1^{KM}$) and single-particle ($E_2^{KM} - E_1^{KM}$) excitations are not strongly renormalized (being of order the bare scale J). The key ingredient of Wilson's NRG, to be discussed in Sect. 2, is a controlled procedure for adding the remaining states $n = 1, 2, \dots$ neglected in the above approximation. As we shall see in the calculation of dynamical quantities below this leads to a drastic renormalization of the spin and single-particle excitations, such that the relevant excitations of the Kondo model (those at $\omega \ll J$) are on a scale of $k_B T_K \ll J$ (e.g., see Figs. 7-8 in Sect. 3). The bare scale J is renormalized down to the Kondo scale $k_B T_K = D(\rho J)^{1/2} \exp(-1/\rho J)$, where D is the band width and $\rho = 1/2D$ the density of conduction states (a physical interpretation of this large renormalization $J \rightarrow k_B T_K$ is given below in terms of tunneling and dissipation).

The Anisotropic Kondo model

The anisotropic Kondo model (AKM) played a key role in the development of the RG approach to quantum impurity models. It was introduced by Anderson as a device for studying the RG flow of the isotropic Kondo model [18].

The AKM also turns out to be of interest in its own right. It is closely related to one-dimensional models with long-range interactions, such as the inverse-square Ising model and the Coulomb gas model [18], which, in contrast to one-dimensional models with short-range interactions, exhibit *finite*-temperature phase transitions (of the Kosterlitz-Thouless type). The connection of the AKM to the dissipative two-state system with Ohmic dissipation (the Ohmic two-state system) [11,12], discussed below, makes the model relevant to the dynamics and thermodynamics of a large class of physical systems. We shall see below that it gives a new perspective on the Kondo effect in terms of the familiar notions of tunneling and dissipation.

The AKM results from the Kondo model by breaking the isotropy of the exchange interaction into a transverse (J_{\perp}) and a longitudinal (J_{\parallel}) part:

$$H = \sum_{k,\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \frac{J_{\perp}}{2} \sum_{\mathbf{k},\mathbf{k}'} (c_{k\uparrow}^\dagger c_{k'\downarrow} S^- + c_{k\downarrow}^\dagger c_{k'\uparrow} S^+) \\ + \frac{J_{\parallel}}{2} \sum_{\mathbf{k},\mathbf{k}'} (c_{k\uparrow}^\dagger c_{k'\uparrow} - c_{k\downarrow}^\dagger c_{k'\downarrow}) S^z + g\mu_B h S_z. \quad (7)$$

For later purposes we have introduced, in the last term, a local magnetic field, h , coupling only to the impurity spin. The significance we attach to this model is the following: it describes quantum mechanical tunneling between two impurity spin states denoted by \uparrow and \downarrow with a level asymmetry $\varepsilon = g\mu_B h$ and connected by a bare tunneling matrix element J_{\perp} . The longitudinal term J_{\parallel} acts as a dissipative coupling of the impurity spin to the electronic environment. Large J_{\parallel} favours spin-flips (tunneling) whereas small J_{\parallel} reduces spin-flips (tunneling). The former case therefore corresponds to weak dissipation (indeed the case $J_{\parallel} = \infty$ reduces to a two-state system decoupled from the electronic environment, i.e. zero dissipation) and the latter case to strong dissipation.

The Ohmic two-state system

It can be shown in several ways, including bosonization, that the AKM reduces to the Ohmic two-state system [11,13,14], whose Hamiltonian is given by [11]

$$H_{SB} = -\frac{1}{2}\hbar\Delta_0\sigma_x + \frac{1}{2}\varepsilon\sigma_z + \sum_i \omega_i (a_i^\dagger a_i + \frac{1}{2}) + \frac{1}{2}\sigma_z \sum_i \lambda_i (a_i + a_i^\dagger). \quad (8)$$

The first two terms describe a two-level system with a bare tunneling matrix element Δ_0 and a level asymmetry ε . The third term represents the environment. It consists of an infinite set of harmonic oscillators with frequency spectrum ω_i . In the mapping from the AKM these arise from the particle-hole excitations of the electron sea. The interaction between the two-level system

and the environment is described by the last term. The form of the spectral function $J(\omega) = \frac{\pi}{2} \sum_i \lambda_i^2 \delta(\omega - \omega_i)$ characterizes the model. The explicit derivation of the model from the AKM leads to a linear frequency dependence $J(\omega) = 2\pi\alpha\omega$, which is termed *Ohmic*, and which holds for $\omega \ll \omega_c$, where $\omega_c = 2D$ is a high energy cut-off. The dimensionless parameter, α , characterizes the strength of the dissipation. The precise correspondence between the models is given by $\epsilon = g\mu_B h$, $\frac{\Delta_0}{\omega_c} = \rho J_\perp$, and $\alpha = (1 + \frac{2\delta}{\pi})^2$ where $\tan \delta = -\frac{\pi\rho J_\parallel}{4}$ [11,14].

Scaling analysis

The scaling idea of Anderson for the AKM consists of eliminating high energy conduction electron states lying in a thin shell of width $|dD|$ close to the band edges $\pm D$ to obtain a new effective Hamiltonian with reduced band edges $\pm(D - |dD|)$ and renormalized couplings. The renormalizations of the dimensionless couplings $\rho J_\perp, \rho J_\parallel$ with decreasing D are described by differential equations, known as the “Poor Man’s” scaling equations when the derivation is perturbative in both ρJ_\perp and ρJ_\parallel [19]. The “Anderson-Yuval” scaling equations are perturbative only in ρJ_\perp . By using the correspondence between the AKM and the Ohmic two-state system given above, we can translate the well-known Anderson-Yuval scaling equations for the AKM [18] into the following scaling equations for the Ohmic two-state system [11,14]

$$\frac{d\alpha}{d \ln \omega_c} = \alpha \left(\frac{\Delta}{\omega_c} \right)^2 + \mathcal{O} \left(\frac{\Delta}{\omega_c} \right)^4, \quad \alpha(\omega_c = \omega_0) = \alpha_0 \quad (9)$$

$$\frac{d(\Delta/\omega_c)}{d \ln \omega_c} = -(1 - \alpha) \left(\frac{\Delta}{\omega_c} \right) + \mathcal{O} \left(\frac{\Delta}{\omega_c} \right)^3, \quad \Delta(\omega_c = \omega_0) = \Delta_0. \quad (10)$$

In these equations α and Δ are variables depending on the running cut-off, ω_c . For the purposes of this section, we denote by α_0 and ω_0 the parameters appearing in the bare Hamiltonian (where they formerly appeared as α and ω_c). The infinitesimal RG transformation (9–10) can be restated in terms of a RG transformation R relating effective Hamiltonians H, H' with different cutoffs $\omega_c, \omega_c + d\omega_c$, ($d\omega_c < 0$), $H = H(\omega_c, \alpha, \Delta/\omega_c) \rightarrow H' \equiv R[H]$. In addition to renormalizing the bare interactions, such transformations will also generate new interactions. These are difficult to treat analytically, so usually only the few most relevant interactions are included. In the numerical approach, described in Sect. 2, one works with general effective Hamiltonians, which allow a large number of new interactions to be included.

From (9) and (10) we see that there is a line of *fixed points* at $\Delta_0/\omega_0 = 0$ for $\alpha_0 \geq 0$ (at which the couplings remain invariant on reducing ω_c). Their stability to a finite Δ_0/ω_0 follows from (10), which states that Δ/ω_c is *relevant*, *marginal* or *irrelevant* depending on whether the dissipation strength α_0 is less than, equal to or larger than 1. Hence, the fixed points at $\Delta_0/\omega_0 = 0$ and $\alpha_0 > 1$ are stable low-energy fixed points (the ferromagnetic fixed points

of the AKM or the “self-trapped” fixed points of the two-state system), whereas the line of fixed points at $\Delta_0/\omega_0 = 0$ for $\alpha_0 \leq 1$ are unstable high-energy fixed points. The *scaling trajectories* (shown in Fig. 3) are calculated by dividing the two equations (9–10) and integrating the resulting equation from ω_0 down to ω_c :

$$\frac{1}{2} \left[\left(\frac{\Delta}{\omega_c} \right)^2 - \left(\frac{\Delta_0}{\omega_0} \right)^2 \right] = -((\ln \alpha - \alpha) - (\ln \alpha_0 - \alpha_0)) \quad (11)$$

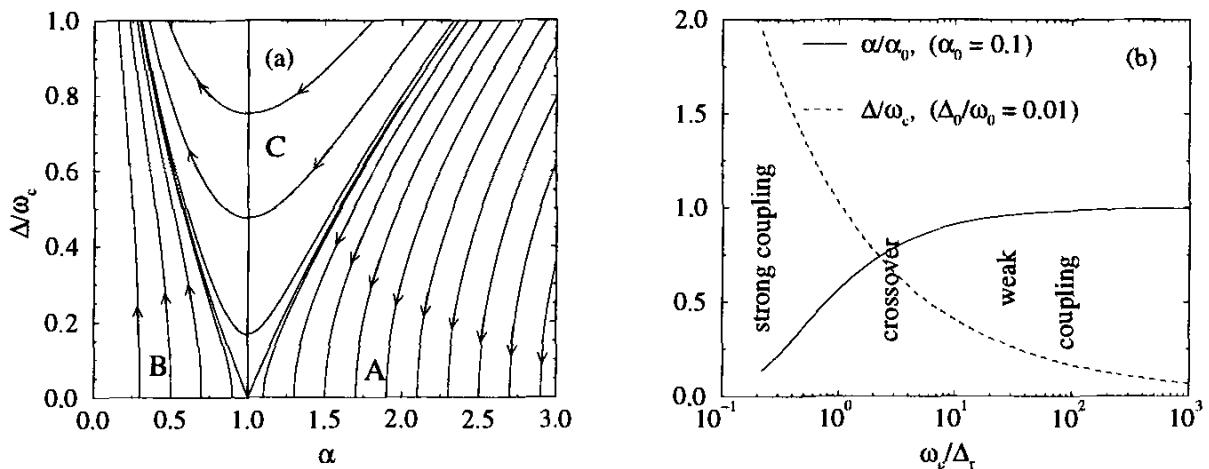


Fig. 3. (a) The scaling trajectories of the Ohmic two-state system [14]. The left and right separatrices at $\alpha = 1, \Delta/\omega_c = 0$ define the regions labeled *A*, *B* and *C* and the arrows indicate the direction of decreasing ω_c . In (b), the flow of the running couplings in the tunneling regime *B* is shown.

In regime A of Fig. 3 ($\alpha_0 > 1$), the dimensionless tunneling amplitude Δ/ω_c is irrelevant and the scaling equations remain valid as $\omega_c \rightarrow 0$. Integrating (10) down to ω_c gives a renormalized tunneling amplitude $\Delta_r \equiv \Delta(\omega_c) = \Delta_0(\omega_c/\omega_0)^{\alpha_0}$ which vanishes at $T = 0$ at low energies. Quantum mechanical tunneling is absent for $\alpha_0 > 1$ at $T = 0$. This region corresponds to the localized regime of the Ohmic two-state system (or the ferromagnetic sector of the AKM).

In regime B ($\alpha < 1$), which corresponds to the antiferromagnetic sector of the AKM ($J_{\parallel} > 0$), Δ/ω_c is relevant and the flow is towards the strong-coupling fixed point at $\alpha = 0$ and $\Delta/\omega_c = \infty$. This is suggested by the numerical solution of (9) and (10), which is shown in Fig. 3. The scaling analysis, which is perturbative in $\Delta(\omega_c)/\omega_c$, of course, breaks down when $\Delta/\omega_c = \mathcal{O}(1)$. The NRG confirms (Fig. 6 in Sect. 2) that the low energy fixed point is at $\Delta/\omega_c = \infty$ and $\alpha = 0$ (corresponding to a flow of the AKM towards the isotropic strong-coupling fixed point $\rho J_{\perp} = \rho J_{\parallel} = \infty$). In this regime, $\Delta(\omega_c)$ tends to a finite value, the renormalized tunneling amplitude Δ_r , as $\omega_c \rightarrow 0$. In the Kondo model this low-energy scale is the Kondo scale, generalized to the anisotropic case. Thus the renormalization of the

bare tunneling amplitude $\Delta \rightarrow \Delta_r$ by frictional effects of the environment corresponds to the renormalization of the bare transverse exchange interaction in the AKM from $J_\perp \rightarrow T_K$. This gives a physical significance to the Kondo scale as a *renormalized tunneling amplitude*. It can be estimated as the *crossover scale* separating weak ($\Delta/\omega_c \ll 1$) and strong-coupling ($\Delta/\omega_c \gg 1$) regimes of the model. Define $\Delta_r = \Delta(\tilde{\omega}_c)$ where $\tilde{\omega}_c$ is the crossover scale such that $\Delta(\tilde{\omega}_c)/\tilde{\omega}_c = 1$. Integrating (10) down to this crossover scale and approximating α by α_0 in the range $[\tilde{\omega}_c, \omega_0]$ gives $\Delta_r/\omega_0 = (\Delta_0/\omega_0)^{1/(1-\alpha_0)}$, which, up to prefactors depending on α_0 , is the correct low energy scale for the Ohmic two-state system [11,12]. The isotropic weak-coupling Kondo model corresponds to $\alpha_0 \rightarrow 1^-$. The frictional effects are largest in this case and give rise to the expected exponential dependence of $\Delta_r = T_K$ on $-1/\rho J$.

The RG flow described above consists of a one parameter family of scaling trajectories labeled by a parameter $\gamma = \gamma(\alpha, \Delta)$ which takes a constant value along each trajectory. This *scaling invariant* can be found from (11), to within an additive constant. We briefly mention the idea of *universality*, that Hamiltonians with different initial couplings $(\alpha_0, \Delta_0/\omega_0)$ and $(\alpha'_0, \Delta'_0/\omega'_0)$ but with the same scaling invariant $\gamma = \gamma(\alpha_0, \Delta_0/\omega_0) = \gamma(\alpha'_0, \Delta'_0/\omega'_0)$ flow along the same RG trajectory specified by γ and that therefore their physical properties are characterized by the same *universal scaling functions*. The scaling functions for the isotropic Kondo model are those on the RG trajectory labeled by initial $\alpha = \alpha_0 = 1^-$. They have also been calculated for general α (arbitrary anisotropy) in [14] by using the Bethe-Ansatz solution of the AKM [15], and in [42] via the NRG.

2 Wilson's Numerical Approach

Wilson's formulation of the RG for the Kondo model is similar in spirit to Anderson's scaling method. The main difference lies in the non-perturbative construction of the RG transformation. The scaling approach uses perturbation theory in the initially small dimensionless coupling (ρJ_\perp) to construct such a transformation, but since ρJ_\perp increases with decreasing energy scale this approach eventually becomes inaccurate. In the Wilson approach the RG transformation is perturbative only via a small parameter $\Lambda^{-1/2} < 1$ which is related to the momentum rescaling factor $\Lambda > 1$. The accuracy of the transformation is the same at each step and is independent of the size of the running couplings. For this reason it gave the first correct description of the crossover from the weak coupling to the strong coupling regime of the Kondo model.

Separation of scales

In the Kondo problem, as in other quantum impurity problems, the behaviour of the system changes qualitatively over many energy scales as it

passes through a crossover between fixed points (e.g. from behaviour characteristic of a well defined magnetic moment at high temperature to behaviour characteristic of a Fermi liquid at temperatures below the crossover scale). In order to describe this crossover the idea is to separate out the many energy scales in the problem and to set up a procedure for treating each scale in turn. A separation of energy scales can be achieved by discretizing the conduction band $[-1 < \varepsilon_k/D < 1]$ into positive and negative energy intervals, $D_n^+ = [\Lambda^{-(n+1)}, \Lambda^{-n}]$ and $D_n^- = [-\Lambda^{-n}, -\Lambda^{-(n+1)}]$, $n = 0, 1, \dots$, about the Fermi level $\varepsilon_F = 0$ as shown in Fig. 4. Assuming orbital isotropy, only s-

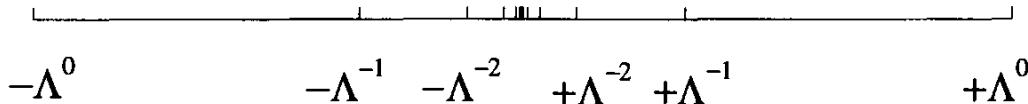


Fig. 4. Logarithmic discretization of the conduction band

wave conduction electrons interact with the impurity, and we can rewrite the continuum version of the Kondo Hamiltonian in the following dimensionless form (henceforth, summation over repeated spin indices is implied)

$$H = \int_{-1}^{+1} d\varepsilon \varepsilon c_{\varepsilon,\mu}^+ c_{\varepsilon,\mu}^- + J\rho \underbrace{\int_{-1}^{+1} d\varepsilon \int_{-1}^{+1} d\varepsilon' c_{\varepsilon,\mu}^+ \sigma_{\mu,\nu} c_{\varepsilon',\nu}^-}_{2J\rho f_{0,\mu}^+ \sigma_{\mu\nu} f_{0,\nu}^-} \cdot S, \quad (12)$$

A constant density of states, ρ , has been assumed and, as in Sect. 1,

$$f_{0\mu} = \frac{1}{\sqrt{2}} \int_{-1}^{+1} d\varepsilon c_{\varepsilon,\mu}^- \quad (13)$$

is the Wannier state at the impurity. The logarithmic discretization approximation now consists of choosing from each interval D_n^\pm just one state, the average electron state

$$c_{-n,\mu} \sim \int_{-\Lambda^{-n}}^{-\Lambda^{-(n+1)}} d\varepsilon c_{\varepsilon,\mu}^-$$

and the average hole state

$$c_{+n,\mu} \sim \int_{+\Lambda^{-(n+1)}}^{+\Lambda^{-n}} d\varepsilon c_{\varepsilon,\mu}^-$$

These states have energies

$$\varepsilon_{\pm n} = \pm \frac{1}{2}(\Lambda^{-n} + \Lambda^{-(n+1)}) = \pm \frac{1}{2}\Lambda^{-n}(1 + \Lambda^{-1}) \quad (14)$$

Of all the states one can construct in each interval D_{-n}^\pm , these are the states which are most localized near the impurity [2]. The infinite number of states $p = 1, 2, \dots$ neglected in each interval D_n^\pm are required to be orthogonal to the states defined above. This suggests that the states neglected $p = 1, 2, \dots$ will be centred at sites away from the impurity. A more precise argument shows that they are centred at distances $r \sim \Lambda^p$ from the impurity and that they only couple indirectly to the impurity [2]. Consequently they can be neglected for the calculation of impurity properties. We therefore arrive at the discretized Kondo Hamiltonian

$$H \approx \sum_{n=0}^{\infty} (\epsilon_{-n} c_{-n,\mu}^+ c_{-n,\mu} + \epsilon_{+n} c_{+n,\mu}^+ c_{+n,\mu}) + 2J\rho f_{0,\mu}^+ \boldsymbol{\sigma}_{\mu\nu} f_{0,\nu} \cdot \mathbf{S}$$

which as in (3) can be put into the linear chain form

$$H = \frac{1}{2}(1 + \Lambda^{-1}) \sum_{n=0}^{\infty} \Lambda^{-n/2} (f_{n,\mu}^+ f_{n,\mu} + f_{n+1,\mu}^+ f_{n,\mu}) + 2J\rho f_{0,\mu}^+ \boldsymbol{\sigma}_{\mu\nu} f_{0,\nu} \cdot \mathbf{S}. \quad (15)$$

Here, we have used the explicit form of the Lanczos coefficients ϵ_n, λ_n appearing in (3) which were calculated analytically in [1] for a logarithmically discretized conduction band: $\epsilon_n = 0$ and $\lambda_n \approx \frac{1}{2}(1 + \Lambda^{-1})\Lambda^{-n/2}, n \gg 1$. This form of the Hamiltonian provides a clear separation of the energy scales $\frac{1}{2}(1 + \Lambda^{-1})\Lambda^{-n/2}, n = 1, 2, \dots$ in H and allows the diagonalization of the Hamiltonian in a sequence of controlled steps, each step corresponding to adding an orbital $f_{n,\mu}$ which is a relative perturbation of strength $\Lambda^{-1/2} < 1$.

RG transformation

A RG transformation relating effective Hamiltonians on successive energy scales $\Lambda^{-n/2}$ and $\Lambda^{-(n+1)/2}$ can be set up as follows. First, H in (15) is truncated to N orbitals to give H_N , whose lowest scale is $D_N = \frac{1}{2}(1 + \Lambda^{-1})\Lambda^{-(N-1)/2}$. In order to look for fixed points we define rescaled Hamiltonians $\bar{H}_N \equiv H_N/D_N$:

$$\begin{aligned} \bar{H}_N &= \Lambda^{(N-1)/2} \left[\sum_{n=0}^{N-1} \Lambda^{-n/2} (f_{n,\mu}^+ f_{n,\mu} + f_{n+1,\mu}^+ f_{n,\mu}) + \tilde{J} f_{0,\mu}^+ \boldsymbol{\sigma}_{\mu\nu} f_{0,\nu} \cdot \mathbf{S} \right], \\ \tilde{J} &= \frac{2J\rho}{\frac{1}{2}(1 + \Lambda^{-1})}, \end{aligned}$$

from which we can recover H as

$$H = \lim_{N \rightarrow \infty} \frac{1}{2}(1 + \Lambda^{-1})\Lambda^{-(N-1)/2} \bar{H}_N. \quad (16)$$

The sequence of rescaled Hamiltonians \bar{H}_N satisfies the recursion relation

$$\bar{H}_{N+1} = \Lambda^{1/2} \bar{H}_N + (f_{N,\mu}^+ f_{N+1,\mu} + f_{N+1,\mu}^+ f_{N,\mu}), \quad (17)$$

and allows a RG transformation R to be defined:

$$\bar{H}_{N+1} = R[\bar{H}_N] \equiv \Lambda^{1/2} \bar{H}_N + (f_{N,\mu}^+ f_{N+1,\mu} + f_{N+1,\mu}^+ f_{N,\mu}) - \bar{E}_{G,N+1} \quad (18)$$

with $\bar{E}_{G,N+1}$ the ground-state energy of \bar{H}_{N+1} .

Iterative scheme

The transformation R relates effective Hamiltonians $H_N = D_N \bar{H}_N$ and $H_{N+1} = D_{N+1} \bar{H}_{N+1}$ on decreasing scales $D_N > D_{N+1}$. It can be used to iteratively diagonalize the Kondo Hamiltonian by the following sequence of steps:

1. the local part

$$\bar{H}_0 = \Lambda^{-1/2} \tilde{J} f_{0,\mu}^+ \sigma_{\mu\nu} f_{0,\nu} \cdot S, \quad (19)$$

which contains the many-body interactions, is diagonalized (the “zeroth”-order step described in Sect. 1),

2. assuming that \bar{H}_N has been diagonalized,

$$\bar{H}_N = \sum_{\lambda} \bar{E}_{\lambda}^N |\lambda\rangle\langle\lambda| \quad (20)$$

we add a “site” and use (18) to set up the matrix for \bar{H}_{N+1} within a product basis $|\lambda, i\rangle = |\lambda\rangle_N |i\rangle_{N+1}$ consisting of the eigenstates $|\lambda\rangle_N$ of \bar{H}_N and the 4 states $|i\rangle_{N+1}$ of the next orbital along the chain (i.e. $|i\rangle_{N+1} = |0\rangle, |\uparrow\rangle, |\downarrow\rangle, |\uparrow\downarrow\rangle$). The resulting matrix

$$\begin{aligned} \langle \lambda, i | \bar{H}_{N+1} | \lambda', i' \rangle &= \Lambda^{1/2} \delta_{i,i'} \delta_{\lambda,\lambda'} \bar{E}_{\lambda}^N \\ &+ (-1)^{N_{e,\lambda'}} \langle \lambda | f_{N,\mu}^+ | \lambda' \rangle \langle i | f_{N+1,\mu} | i' \rangle \\ &+ (-1)^{N_{e,\lambda}} \langle i | f_{N+1,\mu}^+ | i' \rangle \langle \lambda | f_{N,\mu} | \lambda' \rangle, \end{aligned}$$

with $N_{e,\lambda}, N_{e,\lambda'}$ the number of electrons in $|\lambda\rangle, |\lambda'\rangle$ respectively, is diagonalized and the procedure is repeated for the next energy shell as depicted in Fig. 5. Since \bar{H}_N is already diagonalized, the off-diagonal matrix elements, involving ${}_N\langle \lambda | f_{N,\mu} | \lambda' \rangle_N$, can be expressed in terms of the known eigenstates of \bar{H}_N (see [2] for explicit expressions).

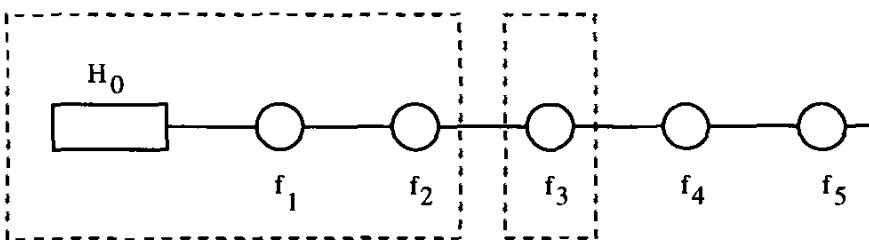


Fig. 5. Iterative diagonalization scheme for H , starting with H_0 and then adding successive orbitals f_1, f_2, \dots

In practice since the number of many-body states in \bar{H}_N grows as 4^N it is not possible to retain all states after about $N = 5$. For $N > 5$ only the lowest 1000 or so states of \bar{H}_N are retained. The truncation of the spectrum of \bar{H}_N restricts the range of eigenvalues in $H_N = D_N \bar{H}_N$ to be such that $0 \leq E_\lambda^N \leq K D_N$ where $K = K(\Lambda)$ depends on Λ and the number of states retained. For 1000 states and $\Lambda = 3$, $K(\Lambda) \approx 10$. However, eigenvalues below D_N are only approximate eigenvalues of the infinite system H , since states with energies below D_N are calculated more accurately in subsequent iterations $N + 1, N + 2, \dots$. Therefore the part of the spectrum of H_N which is close to the spectrum of H is restricted to $D_N \leq E_\lambda^N \leq K(\Lambda)D_N$. This allows the whole spectrum of H to be recovered by considering the spectra of the sequence of Hamiltonians H_N , $N = 0, 1, \dots$. In this way the many-body eigenvalues and eigenstates are obtained on all energy scales.

There is no guarantee that the neglect of the higher energy states in the procedure described above will not make inaccurate the low-energy part of the spectrum calculated in subsequent iterations. However, in the Kondo model, as well as in other quantum impurity models, the addition of an energy shell is a perturbation of relative strength $\Lambda^{-1/2} < 1$ and the neglect of the high-energy states turns out not to spoil the low-energy spectrum. Herein lies the difference to the naive real-space NRG approaches to one-dimensional lattice problems. For the latter, the addition of a site is never a small perturbation since the hopping matrix elements on a one-dimensional lattice are constant. Consequently real-space RG methods which neglect states on the basis of energy are uncontrolled approximations. Another factor which makes the Kondo calculation accurate is the relatively large fraction of states which can be retained at each iteration. This is $1/4$ for the Kondo model and can be made slightly larger by using symmetries. For models where this fraction is smaller (e.g. $1/6$ for the local Cu – O model [51], $1/16$ for the two-channel Kondo model) explicit calculations show that the neglect of high-energy states remains a valid approximation.

The smallness of the perturbation when adding successive energy shells appears to be the main reason for maintaining the accuracy of the low-energy spectrum for each N and therefore for each energy scale. How accurate is the spectrum obtained from the iterative procedure? For the Kondo model, the fixed point spectrum (that of \bar{H}_N for $N \rightarrow \infty$) corresponds to setting $\tilde{J} = \infty$ in H and can be calculated exactly. Typically, the relative error in the lowest eigenvalues calculated from the iterative procedure is found to be of $\mathcal{O}(10^{-7})$ [1]. This holds even for $N \sim 100$, which for $\Lambda = 2$ corresponds to eigenvalues with absolute value $\Lambda^{-(N-1)/2} \sim 10^{-15}$. That the eigenstates are also obtained accurately follows indirectly from the fulfilment of exact Fermi-liquid relations for dynamical properties (e.g. the Friedel sum rule of Sect. 4).

Comparison with DMRG

The DMRG method, described in the next chapter, differs from the usual NRG approach (and that used in the Kondo calculation) in several ways. The most important, and the reason for its success as applied to one-dimensional lattice models, is the criterion for choosing the basis states of the subsystems (blocks) used to extend the size of the system (superblock). These are chosen according to their weight in a reduced density matrix built from a few eigenstates of the larger system (superblock). That is, the states retained in the subsystems (similar to the lowest states retained in \bar{H}_N in the Kondo calculation) are in this case not necessarily the lowest energy states, but they are the states which couple most strongly, in the sense of having large eigenvalues in the reduced density matrix describing the subsystem, to the ones of interest, the target states of the larger system (in the Kondo calculation these might be taken to be the lowest few eigenstates of \bar{H}_{N+1}). The procedure gives highly accurate results for these target states, something which is not possible within the conventional real-space NRG. A second difference of the DMRG approach to the usual NRG approach, is precisely the focus of the DMRG on obtaining just a few low-lying states accurately, instead of a group of several hundred states on each energy scale which one is in principle trying to obtain within the usual NRG approach. However, as explained above, the latter aim is currently only realized for quantum impurity models. Conventional real-space approaches fail in this respect, and the DMRG is currently the most accurate method for obtaining low-lying eigenstates of one-dimensional lattice models.

3 Applications

Applications of the NRG to quantum impurity models fall into three areas: analysis of fixed points, calculation of thermodynamics and calculation of dynamic and transport properties.

Following Wilson's formulation of the NRG for the Kondo problem, the fixed points and thermodynamics of a number of other quantum impurity models were investigated, including the Anderson impurity model [2], the resonant level model [20], the two-impurity [21,22] and two-channel Kondo models [23,24], Anderson models with further screening [25,26] or conduction channels [24,27,28,32] and Kondo impurities in superconductors [29] and pseudogap systems [30,31,33].

A further development was the calculation of dynamical quantities, starting with the $T = 0$ absorption and photoemission spectra of the X-ray Hamiltonian [34], followed by the calculation of $T = 0$ spectral densities and response functions of the non-degenerate and degenerate Anderson impurity models [35–40] and related models [41–44]. Finite-temperature dynamic and transport properties of the Anderson impurity model have also been calculated [40,45–47]. The linear conductance of a quantum dot, modeled by an

Anderson impurity model, has been calculated as in [40] and used to interpret recent experiments [48] (see [49,50] for further applications to mesoscopic systems). The calculation of non-equilibrium transport through correlated impurity systems has been discussed in [51].

Another recent direction for applications of the NRG has been to the study of Anderson and Kondo impurity models with energy-dependent hybridization functions, $\Delta(\omega)$, or momentum-dependent exchange interactions. These are of interest in the context of pseudogap systems, Kondo insulators and in the dynamical mean field theory of strongly correlated lattice systems [53]. The possibility of using the NRG for the latter was noted in [45,52]; applications to the metal-insulator transition in the Hubbard model include [54,55].

Fixed points and thermodynamics

A fixed point H^* of R is defined by

$$R[H^*] = H^*. \quad (21)$$

In practice, proximity to a fixed point is identified by ranges of N , $N_1 \leq N \leq N_2$, where the energy levels \bar{E}_p^N of \bar{H}_N are approximately independent of N : $\bar{E}_p^N \approx \bar{E}_p$ for $N_1 \leq N \leq N_2$. A typical energy level flow diagram showing regions of N where the energy levels are approximately constant is shown in Fig. 6 for the AKM [42]. There is an unstable high-energy fixed point (small N) and a stable low-energy fixed point (large N). The low-energy spectrum is identical to that of the isotropic Kondo model at the strong-coupling fixed

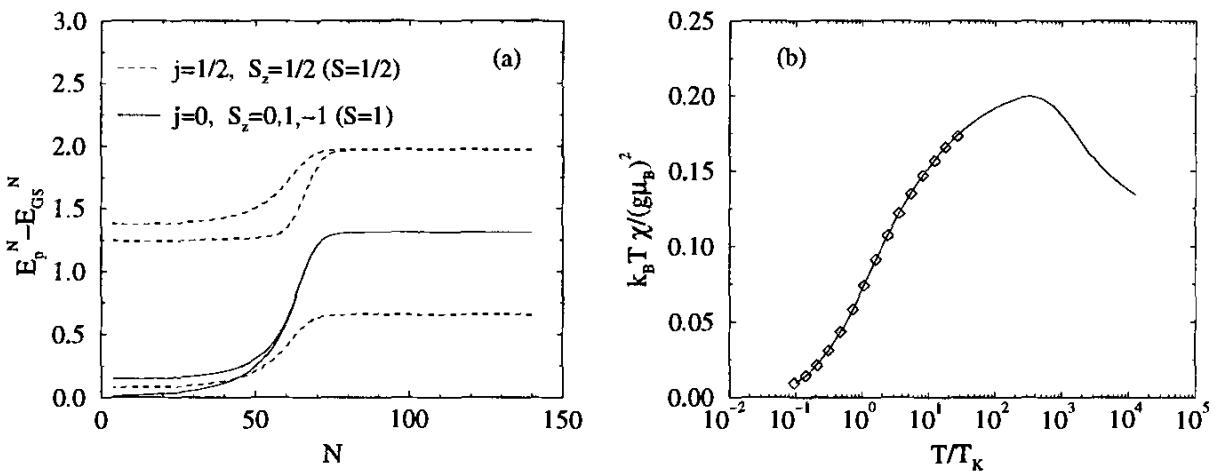


Fig. 6. (a) The lowest rescaled energy levels of the AKM for $J_{||} = 0.443$ and $J_{\perp} = 0.01$. The states are labeled by conserved pseudospin j and total S_z [42]. In (b) the static susceptibility of the Anderson impurity model for $U/\pi\Delta = 6$, $\varepsilon_d/\Delta = -5$ (full curve) is shown. The symbols are from the universal susceptibility curve for the isotropic Kondo model (taken from Table V of [2]), which agrees with the low temperature susceptibility of the Anderson model.

point $J = \infty$ in [1] (e.g. the lowest single-particle excitations in Fig. 6, $\eta_1 = 0.6555$, $\eta_2 = 1.976$ agree with the $\Lambda = 2$ results of the isotropic model in [1]). The crossover from the high-energy to the low-energy fixed point is associated with a scale $T_K(J_\perp, J_\parallel)$ (or Δ_r in the equivalent Ohmic two-state system of Sect. 1). Spin-rotational invariance, broken at high energies, is restored below this scale (e.g. the $j = 0$ states with $S_z = 0$ and $S_z = \pm 1$ become degenerate below T_K and can be classified by the same total spin S as indicated in Fig. 6). Analytic calculations can be carried out in the vicinity of fixed points by setting up effective Hamiltonians $H_{\text{eff}} = H^* + \sum_\lambda \omega_\lambda O_\lambda$, where the leading deviations O_λ about H^* can be obtained from general symmetry arguments. This allows, for example, thermodynamic properties to be calculated in a restricted range of temperatures, corresponding to the restricted range of N where \bar{H}_N can be described by a simple effective Hamiltonian H_{eff} . We refer the reader to the detailed description of such calculations in [1,2], and we turn now to the numerical procedure for calculating thermodynamics, which can give results at all temperatures, including the crossover regions.

Suppose we have diagonalized exactly the Hamiltonian for a quantum impurity model such as the Kondo model and that we have all the many-body eigenvalues E_λ and eigenstates $|\lambda\rangle$:

$$H = \sum_\lambda E_\lambda |\lambda\rangle\langle\lambda| \equiv \sum_\lambda E_\lambda X_{\lambda\lambda}. \quad (22)$$

We can then calculate the partition function

$$Z(T) \equiv \text{Tr } e^{-H/k_B T} = \sum_\lambda e^{-E_\lambda/k_B T}, \quad (23)$$

and hence the thermodynamics via the impurity contribution to the free energy $F_{\text{imp}}(T) = -k_B T \ln Z/Z_c$, where $Z_c = \text{Tr } e^{-H_c/k_B T}$ is the the partition function for the non-interacting conduction electrons. In the NRG procedure we can only calculate the "partition functions" Z_N for the sequence of truncated Hamiltonians H_N :

$$Z_N(T) \equiv \text{Tr } e^{-H_N/k_B T} = \sum_\lambda e^{-E_\lambda^N/k_B T} = \sum_\lambda e^{-D_N \bar{E}_\lambda^N/k_B T} \quad (24)$$

We will have $Z_N(T) \approx Z(T)$ provided

1. we choose $k_B T = k_B T_N \ll E_{\max}^N = D_N K(\Lambda)$ so that the contribution to the partition function from excited states $E_\lambda^N > D_N K(\Lambda)$, not contained in Z_N , is negligible, and
2. the truncation error made in replacing H by H_N in equating (23) and (24) is small. This error has been estimated in [2] to be approximately $\Lambda^{-1} D_N / k_B T_N$.

Combining these two conditions requires that

$$\frac{1}{\Lambda} \ll \frac{k_B T_N}{D_N} \ll K(\Lambda). \quad (25)$$

The choice $k_B T = k_B T_N \approx D_N$ is reasonable and allows the thermodynamics to be calculated at a sequence of decreasing temperatures $k_B T_N \sim D_N, N = 0, 1, \dots$ from the truncated partition functions Z_N . The procedure is illustrated in Fig. 6 for the impurity static susceptibility of the Anderson impurity model

$$\chi_{imp}(T) = \frac{(g\mu_B)^2}{k_B T} \left[\frac{1}{Z} \text{Tr} (S_z^{tot})^2 e^{-H/k_B T} - \frac{1}{Z_c} \text{Tr} (S_{z,c}^{tot})^2 e^{-H_c/k_B T} \right].$$

Dynamic properties

We consider now the application of the NRG method to the calculation of dynamic properties of quantum impurity models. For definiteness we consider the Anderson impurity model and illustrate the procedure for the impurity spectral density $\rho_{d,\mu}(\omega, T) = -\frac{1}{\pi} \text{Im} G_{d,\mu}(\omega, T)$, with $G_{d,\mu}(\omega, T)$ the impurity Green's function. Response functions can be calculated in a similar way.

Suppose we have all the many-body eigenstates $|\lambda\rangle$ and eigenvalues E_λ of the Anderson impurity Hamiltonian H . Then, the impurity Green's function can be written in the Lehmann representation as

$$G_{d,\mu}(\omega, T) = \frac{1}{Z(T)} \sum_{\lambda, \lambda'} |\langle \lambda | d_\mu | \lambda' \rangle|^2 \frac{e^{-E_\lambda/k_B T} + e^{-E_{\lambda'}/k_B T}}{\omega - (E_{\lambda'} - E_\lambda)} \quad (26)$$

and the corresponding impurity spectral density $\rho_{d,\mu}$ as

$$\rho_{d,\mu}(\omega, T) = \frac{1}{Z(T)} \sum_{\lambda, \lambda'} |M_{\lambda, \lambda'}|^2 (e^{-E_\lambda/k_B T} + e^{-E_{\lambda'}/k_B T}) \delta(\omega - (E_{\lambda'} - E_\lambda))$$

with $M_{\lambda, \lambda'} = \langle \lambda | d_\mu | \lambda' \rangle$.

Consider first the $T = 0$ case ($T > 0$ is described in the next section), then

$$\rho_{d,\mu}(\omega, T = 0) = \frac{1}{Z(0)} \sum_{\lambda} |M_{\lambda,0}|^2 \delta(\omega + E_\lambda) + \frac{1}{Z(0)} \sum_{\lambda'} |M_{0,\lambda'}|^2 \delta(\omega - E_{\lambda'}).$$

In order to evaluate this from the information which we actually obtain from an iterative diagonalization of H , we consider the impurity spectral densities corresponding to the sequence of Hamiltonians $H_N, N = 0, 1, \dots$,

$$\rho_{d,\mu}^N(\omega, T = 0) = \frac{1}{Z_N(0)} \sum_{\lambda} |M_{\lambda,0}^N|^2 \delta(\omega + E_\lambda^N) + \frac{1}{Z_N(0)} \sum_{\lambda'} |M_{0,\lambda'}^N|^2 \delta(\omega - E_{\lambda'}^N).$$

From the discussion on the spectrum of H_N in the previous section, it follows that the groundstate excitations of H_N which are representative of the infinite-system H are those in the range $D_N \leq \omega \leq K(\Lambda)D_N$. Lower energy

excitations and eigenstates are calculated more accurately at subsequent iterations, and higher energy excitations are not contained in H_N due to the elimination of the higher energy states at each N . Hence, for fixed N ,

$$\rho_{d,\mu}^N(\omega, T = 0) \approx \rho_{d,\mu}(\omega, T = 0) \quad (27)$$

provided that we choose $\omega \approx \omega_N \equiv k_B T_N$. A typical choice is $\omega = 2\omega_N$ for $\Lambda = 2$. This allows $\rho_{d,\mu}(\omega, T = 0)$ to be calculated at a sequence of decreasing frequencies $\omega = 2\omega_N, N = 0, 1, \dots$ from the quantities $\rho_{d,\mu}^N$. In practice

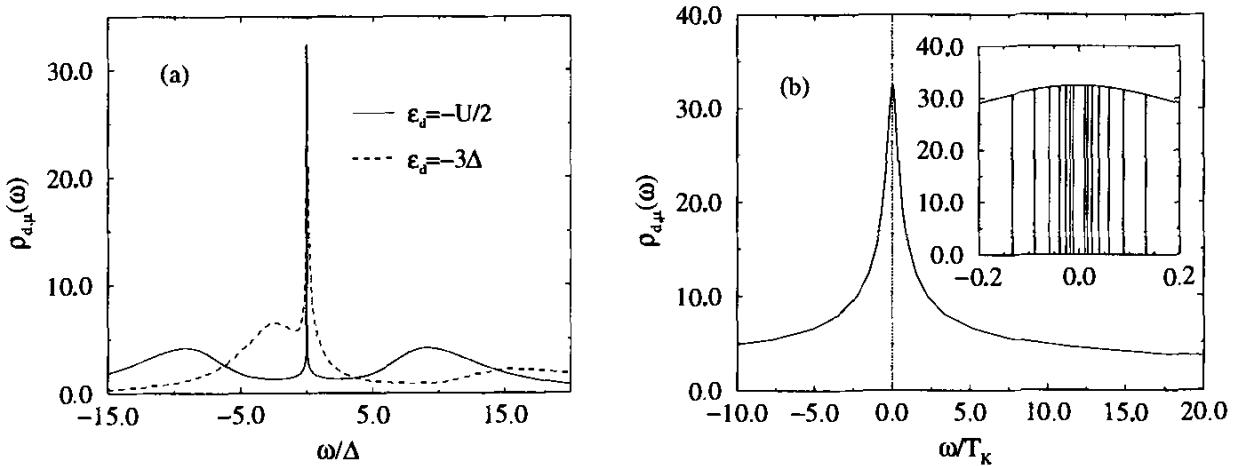


Fig. 7. (a) The impurity spectral density for the Anderson impurity model for $U/\pi\Delta = 6$ and different local level positions. The Kondo resonance for the case $\epsilon_d = -U/2$ is shown in more detail in (b). The vertical lines in the inset show the sequence of energies $\omega = 2\omega_N$ at which the spectral density is calculated and demonstrates the ability of the method to resolve low-energy scales.

we are not interested in the discrete spectra $\rho_{d,\mu}^N(\omega) = \sum_\lambda w_\lambda^N \delta(\omega - E_\lambda^N)$ of the Hamiltonians H_N but in continuous spectra which can be compared with experiment. Smooth spectra can be obtained from the discrete spectra by replacing the delta functions $\delta(\omega - E_\lambda^N)$ by smooth distributions $P_N(\omega - E_\lambda^N)$. A natural choice for the width η_N of P_N is D_N , the characteristic scale for the energy-level structure of H_N . Two commonly used choices for P are the Gaussian and the Logarithmic Gaussian distributions [36,40]. A peak of intrinsic width Γ at frequency Ω_0 will be well resolved by the above procedure provided that $\Omega_0 \ll \Gamma$, which is the case for the Kondo resonance and other low-energy resonances. In the opposite case, the low (logarithmic) resolution at higher frequencies may be insufficient to resolve the intrinsic widths and heights of such peaks. Usually such higher frequency peaks are due to single-particle processes and can be adequately described by other methods (exceptions include interaction dominated features in the Ohmic two-state system, see below, and in strongly correlated lattice models in high dimensions [53]). In both cases, $\Omega_0 \ll \Gamma$ and $\Omega_0 \gg \Gamma$, the positions and intensities of such peaks is given correctly. An alternative procedure for obtaining smooth spectra, which in principle resolves finite-frequency peaks with the same resolution

as the low-energy peaks, has been proposed in [56]. This involves a modified discretization of the conduction band with energies $\pm 1, \pm \Lambda^{-z}, \pm \Lambda^{-z-1}, \dots$ instead of the usual discretization $\pm 1, \pm \Lambda^{-1}, \pm \Lambda^{-2}, \dots$. By considering all z between 0 and 1 one recovers a continuous spectrum without the need to use a broadening function. The procedure requires diagonalizing H for many values of z . It has also proved useful for carrying out thermodynamic calculations at large Λ [57].

How accurate is the NRG for dynamic properties? In Fig. 7 we show results for $T = 0$ spectral densities of the Anderson impurity model [40]. A good measure of the accuracy of the procedure is given by the Friedel sum rule, a Fermi-liquid relation which states that [58]

$$\rho_{d,\mu}(0) = \frac{1}{\pi\Delta} \sin^2(\pi n_d/2), \quad n_d = \int_{-\infty}^0 d\omega \rho_{d,\mu}(\omega) \quad (28)$$

The integrated value of n_d , for the spectral density shown in Fig. 7, is 0.991. Including the renormalization in Δ due to the discretization, as discussed in [2], gives $\rho_{d,\mu}(0) = 32.779$. The value extracted directly from Fig. 7 is $\rho_{d,\mu}(0) = 32.31$ resulting in a 1.4% error, most of which is due to using the integrated value of n_d over all energy scales. Calculating n_d solely from the low energy part of the spectrum (e.g. as the limit $n_d(T \rightarrow 0)$ in a thermodynamic calculation) further reduces this error. More important, however, is that the error remains small independent of the interaction strength $0 \leq U \leq \infty$.

Two-particle Green's functions and response functions can also be calculated. Figure 8 shows the longitudinal spin-relaxation function

$$S(\omega) = -\frac{1}{\pi} \frac{\text{Im}\chi_{zz}(\omega)}{\omega}, \quad \chi_{zz} = \langle\langle S_z; S_z \rangle\rangle$$

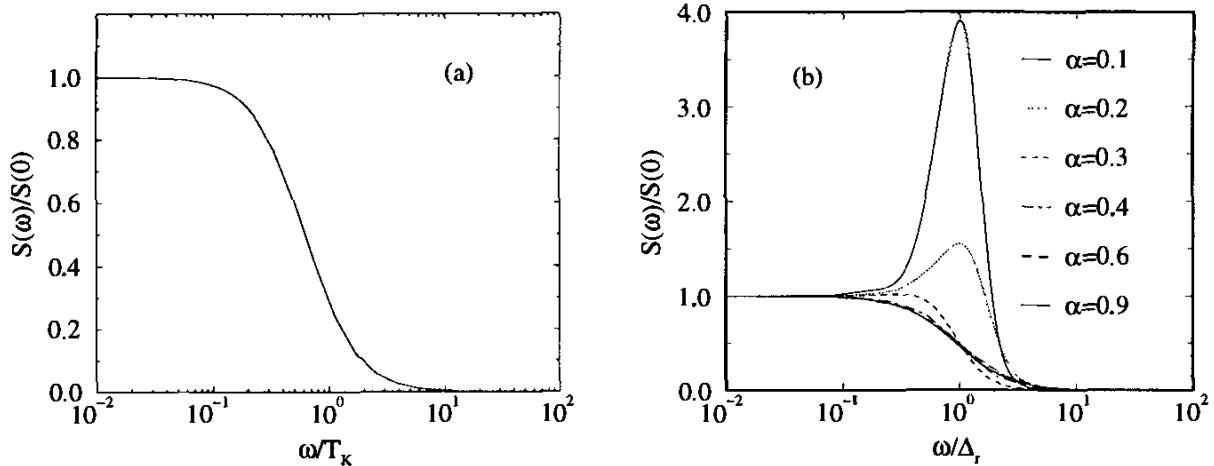


Fig. 8. The $T = 0$ longitudinal spin-relaxation function, $S(\omega)$, for (a) the Anderson impurity model for $U/\pi\Delta = 6$ and $\varepsilon_d = -5\Delta$, and (b) the AKM for increasing values of the coupling ρJ_{\parallel} corresponding to decreasing values of the dissipation strength α in the equivalent Ohmic two-state system [41] ($\Delta_r = T_K$).

of the Anderson impurity model and of the AKM [41,42]. The former always exhibits incoherent spin dynamics. It is interesting that the latter can exhibit coherent spin dynamics for sufficiently large ρJ_{\parallel} . This is more easily understood in terms of the equivalent Ohmic two-state system (Sect. 1), where large ρJ_{\parallel} corresponds to a sufficiently small dissipative coupling of the spin to its electronic environment leading to coherent dynamics of the spin or coherent tunneling between the two quantum-mechanical states. A subtlety here is that a quasielastic peak in $S(\omega)$ does not necessarily mean that the time-dependent dynamics of the spin is incoherent [59]. In fact the quasielastic peak in $S(\omega)$ for the AKM disappears at $\alpha = 1/3$ [41] whereas the incoherent dynamics (non-oscillatory behaviour in the *real-time* dynamics) only sets in at the Toulouse point $\alpha = 1/2$ ($\rho J_{\parallel} \approx 1$).

Transport properties

The transport properties of quantum impurity models, require knowledge of both the frequency and temperature dependence of the impurity spectral density.

The resistivity $\rho(T)$ of conduction electrons scattering from a single Anderson impurity, for example, is given by the expression

$$\rho(T)^{-1} = -e^2 \int_{-\infty}^{+\infty} \tau_{tr}(\omega, T) \frac{\partial f}{\partial \omega} d\omega \quad (29)$$

where the transport time $\tau_{tr}(\omega, T)$ is related to the impurity spectral density by $\tau_{tr}^{-1}(\omega, T) = \Delta \rho_{d,\mu}(\omega, T)$ and Δ is the hybridization strength. Similar expressions hold for the other transport coefficients.

The procedure for calculating finite-temperature dynamical quantities, like $\rho_{d,\mu}(\omega, T)$, required as input for calculating transport properties, is similar to that for $T = 0$ dynamics described above. The spectral density $\rho_{d,\mu}(\omega, T)$ at fixed temperature T is evaluated as above at frequencies $\omega \approx 2\omega_N$, $N = 0, 1, \dots, M$ until $2\omega_M$ becomes of order $k_B T$. To calculate the spectral density at frequencies $\omega < k_B T$ a smaller “cluster” is used. This is done because when $k_B T$ is larger than the frequency at which the spectral density is being evaluated, it is the excited states of order $k_B T$ contained in previous clusters that are important and not the excitations very much below $k_B T$.

Results for the resistivity and thermopower of the Anderson impurity model are shown in Fig. 9. The method gives uniformly accurate results at high and low temperatures, as well as correctly describing the crossover region $T \approx T_K$ (detailed comparisons of the resistivity with known results at high and low temperature can be found in [40,45]).

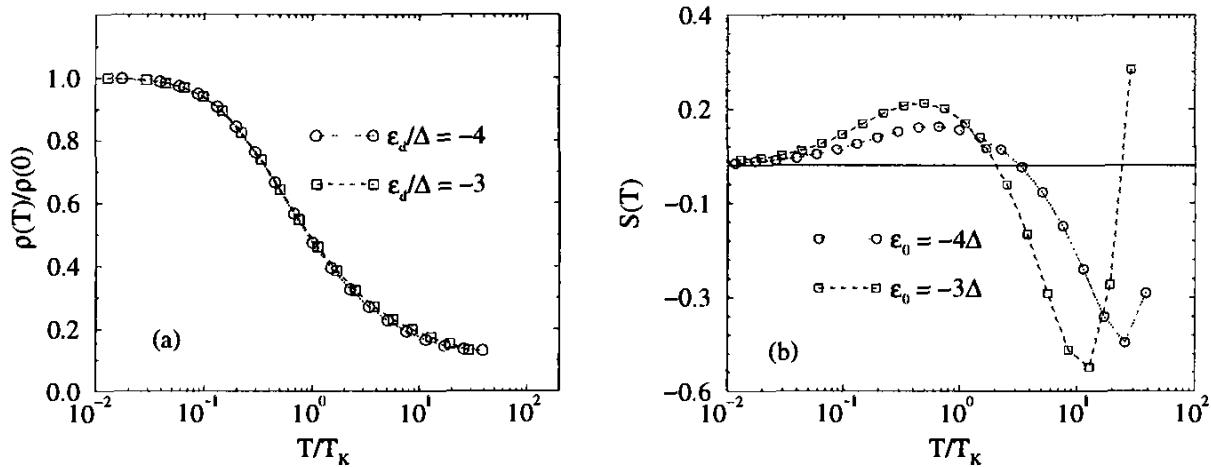


Fig. 9. (a) The scaled resistivity $\rho(T)/\rho(0)$ and (b) the thermopower $S(T)$ of the Anderson impurity model for $U/\pi\Delta = 4$ and two values of the local level position in the Kondo regime [40].

4 Conclusions

The NRG transformation for the Kondo model has developed into a powerful tool for the study of quantum impurity models. It gives information on the many-body eigenvalues and eigenstates of such models on all energy scales and thereby allows the direct calculation of their thermodynamic, dynamic and transport properties.

There is room for further improvement and extensions of the method. The use of a logarithmic discretization of the conduction band, for example, gives rise to insufficient resolution at higher energies. This limits the accuracy with which certain models can be investigated, for examples those arising in the dynamical mean field theory of correlated lattice systems [53]. New approaches [56] for overcoming these difficulties are therefore of interest.

The NRG also has potential to give information on the non-equilibrium transport through correlated impurity systems such as quantum dots [51]. However, away from equilibrium, the absence of a groundstate requires new criteria other than energy for eliminating unimportant states. Ideas based on the DMRG method described in this book are being explored and may prove useful in this context.

References

1. K.G. Wilson Rev. Mod. Phys. **47**, 773 (1975)
2. H.B. Krishnamurthy, J.W. Wilkins and K.G. Wilson, Phys. Rev. B **21**, 1044 (1980)
3. S.-T. Chui and J.W. Bray, Phys. Rev. **18**, 2426 (1978); J.W. Bray and S.-T. Chui, Phys. Rev. **19**, 4876 (1979)

4. S.R. White, Phys. Rev. Lett. **69**, 2863 (1992); Phys. Rev. B **48**, 10345 (1993); Phys. Rep. **301**, 187 (1998)
5. S.R. White and R.M. Noack, Phys. Rev. Lett. **68**, 3487 (1992)
6. A.C. Hewson, The Kondo Problem to Heavy Fermions, Cambridge University Press (1993) gives an extensive discussion of the Kondo model and related quantum impurity models.
7. A.J. Leggett, in Percolation, Localization and Superconductivity, NATO ASI Series: Physics Vol. **109**, edited by M. Goldman and S. A. Wolf
8. F. Lesage, H. Saleur and S. Skorik, Phys. Rev. Lett. **76**, 3388 (1996)
9. R. Konik and A. LeClair, Phys. Rev. B **58**, 1872 (1998)
10. K.S. Ralls and R.A. Buhrman, Phys. Rev. Lett. **60**, 998 (1988); B. Golding, N. Zimmermann and S. Coppersmith, Phys. Rev. Lett. **68**, 998 (1992)
11. A.J. Leggett, S. Chakravarty, A.T. Dorsey, M.P.A. Fisher, A. Garg and W. Zwerger, Rev. Mod. Phys. **59**, 1 (1987); **67**, 725 (1995)
12. U. Weiss, Series in Modern Condensed Matter Physics, Vol. 2, 2nd edition, World Scientific, Singapore (1999)
13. F. Guinea, V. Hakim and A. Muramatsu, Phys. Rev. B **32** 4410, (1985)
14. T.A. Costi and G. Zaránd, to appear in Phys. Rev. B **59** (1999)
15. A.M. Tsvelick and P.B. Wiegmann, Adv. Phys. **32**, 453 (1983)
16. P.W. Anderson, Phys. Rev. **124**, 41 (1961)
17. J.R. Schrieffer and P.A. Wolff, Phys. Rev. **149**, 491 (1967)
18. P.W. Anderson and G. Yuval, Phys. Rev. Lett. **23**, 89 (1969); Phys. Rev. B **1**, 4464 (1970)
19. P.W. Anderson, J. Phys. C **3**, 2439 (1970)
20. L.N. Oliveira and J.W. Wilkins, Phys. Rev. **23**, 1553 (1981)
21. B.A. Jones, in Field Theories in Condensed Matter, Ed. Z. Tesanovic, Addison-Wesley, p 87 (1990)
22. O. Sakai and Y. Shimizu, J. Phys. Soc. Japan, **61**, 2333, (1991); ibid **61**, 2348 (1991)
23. D.M. Cragg, P. Lloyd and P. Nozieres, J. Phys. C **13**, 803 (1980)
24. H.B. Pang and D.L. Cox, Phys. Rev. B **44**, 9454 (1991); H.B. Pang, Phys. Rev. Lett. **73**, 2736 (1994)
25. T.A. Costi and A.C. Hewson, Physica C. **185–189**, 2649 (1991); T.A. Costi and A.C. Hewson, J. Magn. Magn. Mat. **108**, 129 (1992)
26. I.E. Perakis, C.M. Varma and A.E Ruckenstein, Phys. Rev. Lett. **70**, 3467 (1993); I.E. Perakis and C.M. Varma, Phys. Rev. **49**, 9041 (1994)
27. K. Ingersent, B.A. Jones and J.W. Wilkins, Phys. Rev. Lett. **69**, 2594 (1992)
28. H. Kusunose, K. Miyake, Y. Shimizu and O. Sakai, Phys. Rev. Lett. **76**, 271 (1996)
29. K. Satori, H. Shiba, O. Sakai and Y. Shimizu, J. Phys. Soc. Japan, **61**, 3239 (1992); ibid, **62**, 3181 (1993)
30. K. Chen and C. Jayaprakash, J. Phys. Cond. Matt. **7**, 491 (1995)
31. K. Ingersent, Phys. Rev. B **54**, 11936 (1996); C. Gonzalez-Buxton and K. Ingersent, preprint cond-mat/9803256, (1998)
32. R. Bulla, A.C. Hewson and G.M. Zhang, Phys. Rev. B **56**, 1171 (1997)
33. R. Bulla, Th. Pruschke and A.C. Hewson, J. Phys. Cond. Matt. **9**, 10463 (1997)
34. L.N. Oliveira and J.W. Wilkins, Phys. Rev. B **24**, 4863 (1981); ibid **32**, 696 (1985); D.L. Cox, H.O. Frota, L.N. Oliveira and J.W. Wilkins, Phys. Rev. B **32**, 555 (1985)

35. H.O. Frota and L.N. Oliveira, Phys. Rev. B **43**, 7871 (1986)
36. O. Sakai, Y. Shimizu and T. Kasuya, J. Phys. Soc. Jap. **58** 3666 (1989)
37. O. Sakai, Y. Shimizu and T. Kasuya, Physica B **163**, (1990)
38. T.A. Costi and A.C. Hewson, Physica B **163**, 179-181 (1990)
39. J.J.S. Brito and H.O. Frota, Phys. Rev. B **42**, 6378 (1990)
40. T.A. Costi and A.C. Hewson, Phil. Mag. B **65**, 1165 (1992)
41. T.A. Costi and C. Kieffer, Phys. Rev. Lett. **76**, 1683 (1996)
42. T.A. Costi, Phys. Rev. Lett. **80**, 1038 (1998)
43. T.A. Costi, P. Schmitteckert, J. Kroha and P. Wölfle, Phys. Rev. Lett. **73**, 1275 (1994); T.A. Costi, P. Schmitteckert, J. Kroha and P. Wölfle, Physica C **235-240**, 2287 (1994); T.A. Costi, J. Kroha and P. Wölfle, Phys. Rev. B **53**, 1850 (1996)
44. O. Sakai, Y. Shimizu and N. Kaneko, Physica B **186-188**, 323, (1993); Y. Shimizu and O. Sakai, Physica B **186-188**, 891 (1993)
45. T.A. Costi and A.C. Hewson, J. Phys. Cond. Matt. **30**, 361 (1993)
46. T.A. Costi, A.C. Hewson and V. Zlatić J. Phys. Cond. Matt. **6**, 2251 (1994)
47. S. Suzuki, O. Sakai and Y. Shimizu, J. Phys. Soc. Japan **65**, 4034 (1996)
48. D. Goldhaber-Gordon, J. Göres, M.A. Kastner, H. Shtrikman, D. Mahalu and U. Meirav, Phys. Rev. Lett. **81**, 5225 (1998)
49. H.O. Frota and K. Flensberg, Phys. Rev. B **46**, 15207 (1992)
50. W. Izumida, O. Sakai and Y. Shimizu, J. Phys. Soc. Japan **66**, 717 (1997); ibid **67**, (1998)
51. T.A. Costi, Phys. Rev. B **55**, 3003 (1997)
52. R. Bulla, Th. Pruschke, J. Low Temp. Phys. **99**, 619 (1995)
53. W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989); A. Georges, G. Kotliar, W. Krauth and M.J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)
54. O. Sakai and Y. Shimizu, Sol. St. Commun. **89**, 307 (1994)
55. R. Bulla and Th. Pruschke and A.C. Hewson, J. Phys. Cond. Matt. **10**, 8365 (1998)
56. M. Yoshida, M. A. Whitaker and L.N. Oliveira, Phys. Rev. B **41**, 9403 (1990)
57. W.C. Oliveira and L.N. Oliveira, Phys. Rev. B **49**, 11992 (1994); S.C. Costa, C.A. Paula, V.L. Libero and L.N. Oliveira, Phys. Rev. B **55**, 30 (1997)
58. D.C. Langreth, Phys. Rev. **150**, 516 (1966)
59. K. Völker, Phys. Rev. B **58**, 1862 (1998)

2 The Density Matrix Renormalization Group

Reinhard M. Noack¹ and Steven R. White²

¹ Institut de Physique Théorique, Université de Fribourg
CH-1700 Fribourg, Switzerland

² Department of Physics and Astronomy, University of California
Irvine, CA 92697–4575, USA

The Density Matrix Renormalization Group [1] (DMRG) is a numerical technique for finding accurate approximations to the ground state and the low-lying excited states of strongly interacting quantum lattice systems such as the Heisenberg, t - J , and Hubbard models. DMRG traces its roots to Wilson's numerical renormalization group (RG) treatment of impurity problems [2] [see also Chap. 1(I)] and is related to real space renormalization groups. DMRG is remarkable in the accuracy that can be achieved for one-dimensional systems. For example, the ground state energy of the spin-one Heisenberg chain on lattices of hundreds of sites can be calculated to an accuracy of order 10^{-10} with a modest amount of computational effort. In addition, up to a dozen or so excited states with specified conserved quantum numbers as well as virtually any equal-time observable can be calculated. The principal limitation of the method at the present time is dimensionality, or, relatedly, range of the interaction, so that the majority of systems treated up to now have been one-dimensional or quasi-one-dimensional, i.e. strips of finite width. The computational effort necessary to achieve a given accuracy grows rapidly with the width of the system.

In this chapter we will attempt to introduce DMRG in a pedagogical manner. We will first briefly discuss some simpler methods to which DMRG is closely related, including exact diagonalization methods and Wilson's original Hamiltonian-based numerical method. Recently, there have been extensions of the method to the calculation of higher energy properties such as dynamical correlation functions [3], to two-dimensional classical systems [4], to one-dimensional quantum systems at finite temperature [5–7], and to quantum chemical calculation of the states of molecules [8]. Many of these extensions are discussed at length in other chapters of this book. In this chapter, we will concentrate on the original formulation for the ground state properties of quantum lattice systems.

The outline of this chapter is as follows: we first briefly introduce exact diagonalization in Sect. 1, and then discuss Wilson's original numerical renormalization group procedure as applied to quantum lattice systems in Sect. 2. We illustrate the reason for the failure of the Wilson RG for most quantum lattice models using the noninteracting tight-binding particle in Sect. 3. Two

algorithms that overcome these shortcomings for the noninteracting system are also introduced in this section. In Sect. 4, we discuss how to generalize one of these, the superblock procedure, to interacting systems via a projection using the reduced density matrix. The two classes of DMRG algorithms, the infinite system algorithm and the finite system algorithm, are discussed in Sect. 5, along with additional details of the algorithms for interacting systems.

This chapter contains two appendices. Appendix A describes the form of the density matrix for single particle systems, such as the tight-binding chain. Appendix B describes the DMRG algorithm for the tight-binding chain and gives a C++ program to carry out the calculation. We hope that using this program to carry out a simple calculation will provide an excellent introduction to the ideas in DMRG.

1 Exact Diagonalization

Let us assume that we want to find the low-energy properties of strongly interacting quantum lattice models such as the Heisenberg model,

$$H = J \sum_{\langle ij \rangle} \mathbf{S}_i \cdot \mathbf{S}_j , \quad (1)$$

or the Hubbard model,

$$H = -t \sum_{\langle ij \rangle, \sigma} \left(c_{i\sigma}^\dagger c_{j\sigma} + c_{j\sigma}^\dagger c_{i\sigma} \right) + U \sum_i n_{i\uparrow} n_{i\downarrow} , \quad (2)$$

where $c_{i\sigma}^\dagger$ creates an electron of spin $\sigma = \uparrow$ or \downarrow on site i , $n_{i\sigma} = c_{i\sigma}^\dagger c_{i\sigma}$, t is the hopping matrix element between neighboring sites, $\langle ij \rangle$ denotes a sum over pairs of nearest neighbor sites, and U is the energy cost due to the Coulomb repulsion of two electrons on the same site. The Heisenberg model can be obtained from the strong coupling (large U) limit of the half-filled Hubbard model with the quantum mechanical spin operator $\mathbf{S}_i = c_{i\sigma}^\dagger \boldsymbol{\tau}_{\sigma\sigma'} c_{i\sigma'}$, where $\boldsymbol{\tau}_{\sigma\sigma'}$ is a vector of Pauli matrices, and $J = 4t^2/U$. Here we will consider primarily the one-dimensional version of these models in order to illustrate the DMRG. The sum over nearest neighbors $\langle ij \rangle$ then reduces to a sum over lattice sites i with $j = i + 1$. Although these models can be written down in compact form, they involve many degrees of freedom. The spin-1/2 Heisenberg model on an L -site lattice has 2^L degrees of freedom, and the Hubbard model 4^L .

Unlike classical models whose ground states are usually trivial, but whose thermodynamics are interesting (e.g. the Ising model), the properties of the ground states of these quantum lattice models are difficult to calculate, and are, in general, poorly understood. For example, the question of whether or not the ground state of the doped two-dimensional Hubbard model on a

square lattice is superconducting with $d_{x^2-y^2}$ pairing symmetry and therefore relevant to the high- T_c superconductors has not yet been definitively answered.

Both the Hubbard and the Heisenberg models do have exact solutions in one dimension via the Bethe Ansatz. However, the solution is somewhat unwieldy and only a limited number of properties of the system can be easily extracted. Comparison with the exact solution will nevertheless provide an opportunity to test the performance of the DMRG. In more than one dimension, there are no well-controlled analytical methods to treat these models. Therefore, numerical techniques have become an important means for studying them. The most important numerical methods are quantum Monte Carlo, exact diagonalization, and the numerical renormalization group.

Perhaps the simplest approach to numerically treating a quantum lattice system is to diagonalize the Hamiltonian for a finite-size lattice. A convenient basis for the Hilbert space for such a system has states of the form

$$|1\rangle \otimes |2\rangle \otimes \dots \otimes |n\rangle , \quad (3)$$

where $|i\rangle$ labels a state of a single site of the system, and \otimes denotes a direct product. For example, for a spin-1/2 system, $|i\rangle = \uparrow$ or \downarrow . The size of the Hilbert space grows exponentially in the number of sites n . If A is a single-site operator belonging to site i , its matrix elements have the simple form

$$\langle n| \dots \langle 1| A | 1' \rangle \dots | n' \rangle = \delta_{11'} \dots \delta_{i-1,i-1'} \delta_{i+1,i+1'} \dots \delta_{n,n'} A_{i,i'} , \quad (4)$$

where $A_{i,i'}$ is the matrix elements of the operator for a single site. Similar expressions hold for products of single-site operators.

The δ -functions in these expressions make the matrix for the Hamiltonian operator in the many-site Hilbert space extremely sparse. Consequently, the usual dense matrix diagonalization methods, such as the Jacobi method or Householder transformation with diagonalization of the resulting tridiagonal matrix, are not the most efficient algorithms for these systems. We are primarily interested in the ground state and possibly a few low-lying excited states. Two widely used and very efficient methods for finding a few eigenstates of a large, sparse matrix are the *Lanczos method* and the *Davidson method*. These methods build up a small set of basis vectors, and minimize the energy within this basis. The reduced set of basis vectors is systematically expanded until convergence is reached. Typically, the original sparse Hamiltonian H only comes into play in the multiplication of a vector v by H , in which case only the nonzero elements of H are relevant. The Davidson method [9], also utilizes the diagonal elements of H in an attempt to generate an improved reduced basis. Consequently, if the Hamiltonian is at all dominated by its diagonal elements, the Davidson method will probably converge more quickly than the Lanczos method.

In efficient exact diagonalization programs, the Hamiltonian matrix is usually not explicitly generated. Instead, a procedure to multiply a vector

by H is used which generates the Hamiltonian matrix elements as they are needed.

Although these diagonalization algorithms work remarkably well for finding the lowest eigenvalues and eigenvectors of the large, sparse matrices found in quantum lattice problems, the maximum system size that can be treated is still severely limited by the exponential growth of the Hilbert space. The principal limitations are then that the computer memory required to store a Hilbert space vector becomes too large to handle, and operations on these vectors become too expensive. In practice, the spin-1/2 Heisenberg model can be diagonalized on up to about 36 sites and the Hubbard model on up to about 20 sites. It would therefore be advantageous to formulate a variational diagonalization scheme that also truncates the Hilbert space used to represent H in a controlled way. This can be done using the *numerical renormalization group*.

2 Wilson's Numerical Renormalization Group

In this section, we will sketch out the basic ideas of Wilson's numerical renormalization group [2], using a notation that will be useful later for discussions of the DMRG. See Chap. 1(I) for a more complete discussion of Wilson's numerical RG as applied to impurity problems. We will postpone discussion of the mechanics of keeping and transforming the operators to a later section, since they are the same as for the DMRG.

The basic idea of the renormalization group is to integrate out unimportant degrees of freedom progressively using a succession of renormalization group transformations. In order to investigate the behavior of the single-impurity Kondo problem, Wilson implemented the RG in a purely *numerical* way [2]. The procedure starts with a numerical representation of the Hamiltonian in a particular basis, then adds degrees of freedom by carrying out a renormalization group transformation, typically by increasing the size of the finite system, and finally numerically transforms the representation of the Hamiltonian to a reduced basis.

To be more concrete, let us consider a one-dimensional quantum lattice model such as the Heisenberg or Hubbard model. In fact, Wilson was able to map the spherically symmetric Kondo problem onto such a one-dimensional quantum lattice model. In the Kondo case, the impurity is represented by the first site and the spherically symmetric momentum shells (logarithmically discretized) of the conduction electrons are represented by a semi-infinite lattice with only near-neighbor interactions.

Wilson's numerical RG procedure then proceeds as follows:

1. Isolate a portion of the system containing L sites. Here L is chosen to be small enough so that the Hamiltonian H_L can be diagonalized exactly.
2. Diagonalize H_L numerically, obtaining the m lowest eigenvalues and eigenvectors.

3. Transform H_L and other operators in the “block” of length L to a new basis consisting of the m lowest eigenvectors of H_L , i.e. form $\bar{H}_L = O_L^\dagger H_L O_L$, $\bar{A}_L = O_L^\dagger A_L O_L$, etc., where the columns of O_L contain the m lowest eigenvectors of H_L , and A_L is an arbitrary operator in the block. Note that \bar{H}_L is a diagonal matrix with m elements.
4. Add a site to \bar{H}_L to form H_{L+1} . In order to do this, the interaction between the block of length L and the additional site added must be reconstructed. We will discuss in more detail how this is done in Sect. 5.
5. Repeat starting with step 2, substituting H_{L+1} for H_L .

This scheme is depicted pictorially in Fig. 1. For a Hamiltonian represented in a real-space basis, the RG step is a real-space blocking scheme. Typically, the number of states m kept at each step is held constant, so the time and memory required for each diagonalization stays the same, and the computer time needed is linear in L .

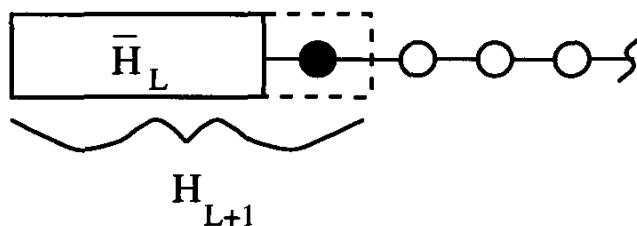


Fig. 1. A pictorial depiction of the Wilson numerical RG procedure.

The basic idea of this scheme is that only the low-energy eigenstates obtained for a system of size L will be important in making up the low-energy states of a system of size $L + 1$. Note that in isolating the block of length L , one has to decide how to treat the boundaries of the block. The simplest thing to do is to neglect connections to surrounding sites, which corresponds to applying open boundary conditions to the system being diagonalized.

This procedure worked well in Wilson’s original work on the single impurity Kondo problem and, with minor variations, is still used today for a variety of single and two Kondo and Anderson impurity problems [see Chap. 1(I)]. In his original paper [2], Wilson very carefully justified the truncation by perturbatively calculating the error, and compared the numerical results with analytical analysis of the behavior near the fixed points. Due to the transformation into an energy basis and the logarithmic discretization of the conduction electron “sites”, each site corresponds to a successively lower energy scale, and the coupling between successive sites decreases exponentially.

However, when this procedure is applied to other systems for which the lattice model does not include an intrinsic separation of energy scales, such as the one-dimensional Heisenberg or Hubbard models, the accuracy becomes quite poor after just a few iterations [10].

3 Numerical RG for the Particle on a Chain

In order to understand why the Wilson numerical renormalization group procedure breaks down for interacting quantum lattice systems, it is useful to consider first its application to a simple noninteracting quantum lattice problem, a single particle on a tight-binding chain. The Hamiltonian we will consider is

$$H = - \sum_{i=1}^{L-1} (|i\rangle\langle i+1| + |i+1\rangle\langle i|) + 2 \sum_{i=1}^L |i\rangle\langle i|, \quad (5)$$

where the state $|i\rangle$ corresponds to a localized tight-binding orbital on site i . The matrix elements $H_{ij} = \langle i|H|j\rangle$ are then

$$H_{ij} = \begin{pmatrix} 2 & -1 & 0 & 0 & \dots \\ -1 & 2 & -1 & 0 & \dots \\ 0 & -1 & 2 & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (6)$$

The value 2 on the diagonal is chosen so that this operator is just the discretization of the second derivative operator, $-\partial^2/\partial x^2$.

The Wilson procedure described in the previous section can be carried out on this system with just a few minor differences. First, since this is a one-particle problem, the dimension of the Hilbert space for a lattice of length L is L , rather an exponential of L as in an interacting system. Since the Hilbert space grows less rapidly with system size for the noninteracting systems, we will add two equal-sized blocks in the real-space blocking step, rather than adding a site at a time. Secondly, the mechanics of putting the blocks together is a little simpler.

In step 1 of the Wilson procedure, we isolate a block of length L . We can understand how this is done for the noninteracting system by considering a semi-infinite system broken up into blocks of length L . The Hamiltonian can then be written

$$H = \begin{pmatrix} H_L & T_L & 0 & 0 & \dots \\ T_L^\dagger & H_L & T_L & 0 & \dots \\ 0 & T_L^\dagger & H_L & T_L & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (7)$$

For $L = 1$, the H_1 is a 1×1 matrix with value 2 and T_1 is a 1×1 matrix with value -1. For larger L , H_L has the form of (6), and T_L connects only sites on the block boundaries, i.e. has all zero elements except for a -1 in the lower left corner. Isolating a block then consists of neglecting the T_L , and therefore applies fixed boundary conditions to H_L . Here we use the term “fixed” boundary conditions rather than the equivalent “open” boundary conditions to emphasize that the single-particle wavefunction vanishes at the

boundary. (On the lattice, the wavefunction is actually zero only at the sites *beyond* the boundary, i.e. site 0 and $L + 1$, see Fig. 2.)

In step 2, we diagonalize H_L and form

$$O_L = \begin{pmatrix} | & | \\ v_1 & \cdots v_m \\ | & | \end{pmatrix}, \quad (8)$$

where $v_1 \dots v_m$ are the eigenvectors corresponding to the $m \leq L$ lowest eigenvalues of H_L . In step 3, we form the diagonal $m \times m$ matrix $\bar{H}_L = O_L^\dagger H_L O_L$ and transform the connection between H_L and the rest of the system, T_L to the new basis by forming $\bar{T}_L = O_L^\dagger T_L O_L$.

We now increase the size of the system, step 4, by putting two blocks of size L together to form a system of size $2L$ with

$$H_{2L} = \begin{pmatrix} \bar{H}_L & \bar{T}_L \\ \bar{T}_L^\dagger & \bar{H}_L \end{pmatrix} \quad (9)$$

and

$$T_{2L} = \begin{pmatrix} 0 & 0 \\ \bar{T}_L & 0 \end{pmatrix}. \quad (10)$$

The procedure can then be repeated starting with step 2 by substituting H_{2L} and T_{2L} for H_L and T_L . The size of the system therefore doubles at each step, but the size of the matrix to be diagonalized is at most $2m \times 2m$. Notice that since we transform to a truncated basis at each step, the matrix elements of \bar{H}_L and \bar{T}_L can no longer be easily related to the original real-space basis. However, if m were equal to L at each step, the procedure would be exact; it would just be a complicated reshuffling of the original Hamiltonian.

As illustrated in the column labeled ‘‘Wilson’’ in Table 1, this procedure performs quite badly as soon as $m < L$. There are large errors in the energies of the lowest few states after only the first few truncating steps. This failure was pointed out by Wilson at an informal seminar at Cornell University in 1986 as an example of a numerical RG procedure which does not work. He also pointed out that in this simple system it is easy to understand why the procedure fails.

Table 1. Lowest energies after 10 blocking transformations for the noninteracting single particle on a 1-D chain with fixed boundary conditions.

	Exact	Wilson	Fixed-Free
E_0	2.3508×10^{-6}	1.9207×10^{-2}	2.3508×10^{-6}
E_1	9.4032×10^{-6}	1.9209×10^{-2}	9.4032×10^{-6}
E_2	2.1157×10^{-5}	1.9214×10^{-2}	2.1157×10^{-5}
E_3	3.7613×10^{-5}	1.9217×10^{-2}	3.7613×10^{-5}

In the continuum limit, the tight-binding model with fixed boundary conditions describes a particle in a box of length L with an infinitely high po-

tential at the walls. The eigenfunctions are therefore particle-in-a-box eigenfunctions, $\psi_n(x) \sim \sin n\pi x/L$ with n a positive integer, and vanish at the boundaries of the box. In the RG procedure, the lowest few eigenstates of a system of length L are combined to form the low-lying eigenstates of a system of length $2L$. The lowest eigenstates of a system of length L and of length $2L$ are plotted in Fig. 2. Clearly, a combination of the ground states of two systems of length L is a bad approximation to the ground state for a system of length $2L$. Since the wave vectors of the discrete system take on small but finite values on the first and last sites of the system, the “kink” at the boundary between the blocks can be removed, but only by using almost all of the eigenstates of the smaller block.

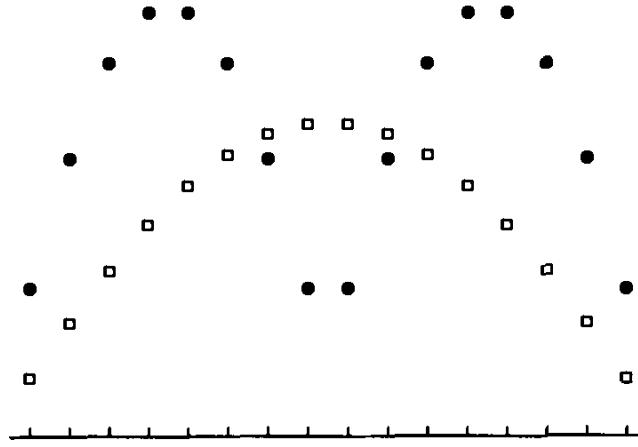


Fig. 2. The lowest eigenstates of two 8-site blocks (solid circles) and a 16-site block (open squares) for the one-dimensional tight-binding model with fixed boundary conditions.

The lesson that is learned is that the treatment of the boundaries of the blocks is crucial in formulating an accurate RG procedure. One could try to fix this problem by applying some other boundary conditions such as periodic boundary conditions to the block rather than fixed boundary conditions. While it is possible to formulate such an RG procedure, one finds that this procedure is also quite inaccurate after the first few truncations. The reason for this is that a periodic wave vector must have the same value at both boundaries of a block, so that the wave vector of the larger block is broken up into “steps”, leading to an inaccurate representation of the low-lying states.

White and Noack [11] formulated two types of RG procedures which solve these problems and work quite well for the single-particle problem. Both are based on choosing a new basis for \bar{H}_L which is *not* the eigenbasis of H_L . In the first procedure, called the *combination of boundary conditions* (CBC) method, the new basis is formed from the low-lying eigenstates of several different block Hamiltonians. The Hamiltonians are formed by applying a number of *different* boundary conditions to the edge of the block. For exam-

ple, fixed and free (for free boundary conditions the derivative of the wave function vanishes at the boundary) boundary conditions can be applied. One can form $H_L^{bb'}$ with $b, b' = \text{fixed or free}$. For example,

$$H_{L=2}^{\text{free,fixed}} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}. \quad (11)$$

We diagonalize $H_L^{bb'}$ for all 4 combinations of boundary conditions, and then form O_L from the $m/4$ eigenvectors associated with the lowest eigenvalues from each combination of boundary conditions. Since the columns of O_L are not orthogonal, we must explicitly orthogonalize them using the Gram-Schmidt procedure. We then form $\bar{H}_L^{bb'} = O_L^\dagger H_L^{bb'} O_L$ and $\bar{T}_L = O_L^\dagger H_L^{bb'} O_L$. Note that $\bar{H}_L^{bb'}$ is not diagonal. The Hamiltonian of the system of size $2L$ is then

$$H_{2L}^{bb'} = \begin{pmatrix} \bar{H}_L^{b,\text{fixed}} & \bar{T}_L \\ \bar{T}_L^\dagger & \bar{H}_L^{\text{fixed},b'} \end{pmatrix}. \quad (12)$$

That fixed boundary conditions must be used where the blocks are joined together can be understood by considering a procedure in which all the states are kept at each step and requiring it to be exact. The matrix T_{2L} is formed as in (10).

The fixed-free CBC procedure works amazingly well. As can be seen by comparing the “Exact” and “Fixed-Free” columns of Table 1, the ground-state energy is obtained almost exactly (the error is in the tenth digit) even after 10 iterations ($L = 2048$) when $m = 8$ states are kept. The CBC procedure can also be formulated with other combinations of boundary conditions such as periodic and antiperiodic. While the periodic-antiperiodic CBC procedure is not as accurate as the fixed-free procedure, it performs much better than the Wilson procedure with periodic or antiperiodic boundary conditions.

The second type of procedure developed in Ref. [11], called the *superblock* method, chooses a new basis for \bar{H}_{2L} and \bar{T}_{2L} based on the idea that they will eventually be used to make up part of a larger system. In order to do this, a “superblock” (with periodic boundary conditions) made up of $p > 2$ blocks is formed and diagonalized. For example,

$$H_{2L}^{p=4} = \begin{pmatrix} \bar{H}_L & \bar{T}_L & 0 & \bar{T}_L^\dagger \\ \bar{T}_L^\dagger & \bar{H}_L & \bar{T}_L & 0 \\ 0 & \bar{T}_L^\dagger & \bar{H}_L & \bar{T}_L \\ \bar{T}_L & 0 & \bar{T}_L^\dagger & \bar{H}_L \end{pmatrix}. \quad (13)$$

The transformation O_{2L} is then made up by projecting the m lowest-lying eigenstates of H_{2L}^p onto the coordinates of the first two blocks, and then orthonormalizing its columns. In other words, if u_j^α (with $j = 1, \dots, 4m$) is an eigenvector of H_{2L}^4 , then a nonorthonormalized column vector of O_{2L} is composed of the first $2m$ elements, $j = 1, \dots, 2m$, of u_j^α , assuming \bar{H}_L is an

$m \times m$ matrix. This new basis is used to transform $\bar{H}_{2L} = O_{2L}^\dagger H_{2L} O_{2L}$ and $\bar{T}_{2L} = O_{2L}^\dagger T_{2L} O_{2L}$, as defined in (9) and (10).

The idea is that the fluctuations in the additional blocks surrounding the portion of the system to be transformed effectively apply general boundary conditions, or equivalently, provide the conditions at the boundaries that the transformed blocks would see as part of a larger system. As p becomes large, this procedure becomes exact because it reduces to an exact diagonalization of the complete final system. Another interesting feature of this procedure is that the diagonalization step is decoupled from the real-space blocking step: a different size system is diagonalized than is blocked together. This procedure yields accurate results for the tight-binding particle eigenstates, although not quite as accurate as the fixed-free CBC procedure.

Of course, we are interested in developing an RG procedure for interacting quantum lattice systems, not the single-particle problem, so the crucial question is whether these procedures can be generalized to work on interacting systems. The CBC method is difficult to generalize because it is difficult to find a general enough set of boundary conditions for interacting systems. To see this, consider the many-body wavefunction for a system of noninteracting fermions, i.e. for the Hubbard model at $U = 0$. An arbitrary many-body state is composed of the Slater determinant of the single-particle wavefunctions of the individual particles, some of which may have nodes and some of which may have antinodes at the boundaries of a block. It is easy to find boundary conditions for which every particle on the block has a node or every particle has an antinode at the boundaries, but it is difficult to find boundary conditions which produce different behavior for different particles. However, a general representation for a block that is part of a larger system must provide a complete range of boundary behavior for each particle individually.

The superblock method seems more promising for application to interacting systems, since the general behavior at the boundaries is provided automatically by embedding the block of interest in a larger superblock. However, the projection of the wavefunction of the superblock onto the system block, which is a simple, single-valued coordinate projection in the noninteracting system, becomes multivalued for the interacting system: one state of the superblock can, in general, project onto many states of the system block. The Density Matrix Renormalization Group is based on choosing an optimal way to do this projection and on combining it with a version of the superblock procedure.

4 The Density Matrix Projection

In this section, we will discuss how to generalize the projection of the superblock described in the previous section for the noninteracting system to interacting systems. The procedure involves forming the reduced density matrix for the system block as part of the superblock. We will show that the

basis obtained using this density matrix projection is the *optimal* basis in a particular sense.

First, let us briefly review the properties of density matrices. An excellent treatment is given in Feynman's book on statistical mechanics [12]. The term "density matrix" is used to refer to a number of different, but related mathematical objects, both in quantum mechanics and quantum statistical mechanics. Here we consider a quantum mechanical system in a definite pure state, and consider the properties of a part of that system. Since we will later use this procedure as part of a superblock algorithm, we will label the entire system the *superblock*, the part that we are interested in constructing a basis for the *system block*, and the remainder of the system the *environment block*, as depicted in Fig. 3. Let $|i\rangle$ label the states of the system block, and $|j\rangle$ label the states of the environment block, i.e. the rest of the superblock. If ψ is a state of the superblock,

$$|\psi\rangle = \sum_{ij} \psi_{ij} |i\rangle |j\rangle . \quad (14)$$

The reduced density matrix for the system block is defined as

$$\rho_{ii'} = \sum_j \psi_{ij}^* \psi_{i'j} . \quad (15)$$

By normalization, $\text{Tr}\rho = 1$. The density matrix contains all the information needed from the wavefunction ψ to calculate any property restricted to the system block. If operator A acts only on the system block, then

$$\langle A \rangle = \sum_{ii'} A_{ii'} \rho_{i'i} = \text{Tr } \rho A . \quad (16)$$

Now let us diagonalize the density matrix. Let ρ have eigenstates $|u^\alpha\rangle$ and eigenvalues $w_\alpha \geq 0$. Since $\text{Tr}\rho = 1$, $\sum_\alpha w_\alpha = 1$. Then for any system block operator A ,

$$\langle A \rangle = \sum_\alpha w_\alpha \langle u^\alpha | A | u^\alpha \rangle . \quad (17)$$

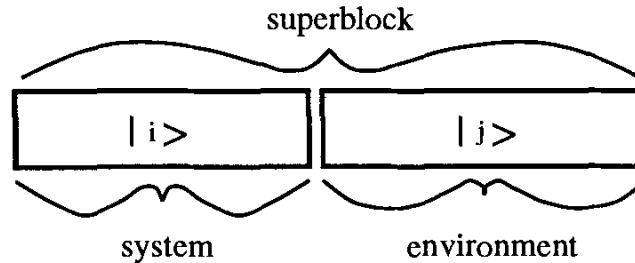


Fig. 3. A superblock divided into a system block and an environment block.

Equation (17) will apply immediately to our numerical renormalization group procedure. Suppose we wish to throw away some states from the system block. If for a particular α , $w_\alpha \approx 0$, we make no error in $\langle A \rangle$, for any A , if we discard $|u^\alpha\rangle$. We have found a way to find which states to keep (those with significant w_α) and which to discard.

This argument can be made much more precise. In particular, we can show that keeping the most probable eigenstates of the density matrix gives the most accurate representation of the state of the superblock, i.e., the system block plus the environment block. Let us assume we have diagonalized the superblock and obtained one particular state $|\psi\rangle$, typically the ground state. We wish to define a procedure for producing a set of states of the system block $|u^\alpha\rangle$, $\alpha = 1, \dots, m$, with $|u^\alpha\rangle = \sum_i u_i^\alpha |i\rangle$, which are optimal for representing ψ in some sense. Because we allow only m states, we cannot represent $|\psi\rangle$ exactly if $\ell > m$, where ℓ is the number of system block states $|i\rangle$. We wish to construct an accurate expansion for $|\psi\rangle$ of the form

$$|\psi\rangle \approx |\bar{\psi}\rangle = \sum_{\alpha,j} a_{\alpha,j} |u^\alpha\rangle |j\rangle . \quad (18)$$

In other words, we wish to minimize

$$\mathcal{S} = \left| |\psi\rangle - |\bar{\psi}\rangle \right|^2 \quad (19)$$

by varying over all $a_{\alpha,j}$ and u^α , subject to $\langle u^\alpha | u^{\alpha'} \rangle = \delta_{\alpha\alpha'}$. Without loss of generality, we can write

$$|\bar{\psi}\rangle = \sum_\alpha a_\alpha |u^\alpha\rangle |v^\alpha\rangle \quad (20)$$

where $v_j^\alpha = \langle j | v^\alpha \rangle = N_\alpha a_{\alpha,j}$, with N_α chosen to set $\sum_j |v_j^\alpha|^2 = 1$. Switching to matrix notation, we have

$$\mathcal{S} = \sum_{ij} (\psi_{ij} - \sum_{\alpha=1}^m a_\alpha u_i^\alpha v_j^\alpha)^2 , \quad (21)$$

and we minimize \mathcal{S} over all u^α , v^α , and a_α , given the specified value of m . The solution to this minimization problem is known from linear algebra. We now think of ψ_{ij} as a rectangular matrix. The solution is produced by the singular value decomposition [13] of ψ ,

$$\psi = UDV^T , \quad (22)$$

where U and D are $\ell \times \ell$ matrices, V is an $\ell \times J$ matrix (where $j = 1, \dots, J$, and we assume $J \geq \ell$), U is orthogonal, V is column-orthogonal, and the diagonal matrix D contains the singular values of ψ . Linear algebra tells us that the u^α , v^α , and a_α which minimize \mathcal{S} are given as follows: the m largest-magnitude diagonal elements of D are the a_α and the corresponding columns of U and V are the u^α and v^α . (We emphasize that the singular

value decomposition is not being used here as a numerical method, only as a convenient factorization which allows us to use a theoretical result from linear algebra.)

These optimal states u^α are also eigenvectors of the reduced density matrix of the block as part of the system. This reduced density matrix for the block depends on the state of the system, which in this case is a pure state $|\psi\rangle$. (The system could also be in a mixed state [see below] or at finite temperature.) The density matrix for the block in this case, where ψ_{ij} is assumed real, is given by

$$\rho_{ii'} = \sum_j \psi_{ij} \psi_{i'j} . \quad (23)$$

We see that

$$\rho = U D^2 U^T , \quad (24)$$

i.e. U diagonalizes ρ . The eigenvalues of ρ are $w_\alpha = a_\alpha^2$ and the optimal states u^α are the eigenstates of ρ with the largest eigenvalues. Each w_α represents the probability of the block being in the state u^α , with $\sum_\alpha w_\alpha = 1$. The deviation of $P_m \equiv \sum_{\alpha=1}^m w_\alpha$ from unity, i.e. the “discarded weight” of the density matrix eigenvalues, measures the accuracy of the truncation to m states.

To summarize, in the previous two paragraphs we have shown that when the superblock is assumed to be in a pure state, the optimal states to keep are the m most significant eigenstates of the reduced density matrix of the system block, obtained from the wavefunction of the superblock via (23).

We can also consider the superblock to be in a mixed state. This is the natural assumption for a system at finite temperature, and it is also useful to assume a mixed state when one wishes to obtain several of the lowest lying states: if we put the superblock with equal probability into each of several states, then the system block states obtained from the density matrix will equally well represent each of these superblock states. We represent the mixed case by saying that the superblock has probability W_k to be in state $|\psi^k\rangle$. If the superblock is at a finite temperature, then the W_k are normalized Boltzmann weights. In this case the appropriate definition for the error in the representation is

$$\mathcal{S} = \sum_k W_k \sum_{ij} \left(\psi_{ij}^k - \sum_{\alpha=1}^m a_\alpha^k u_i^\alpha v_j^{k,\alpha} \right)^2 . \quad (25)$$

Note that we are interested in determining a single set of optimal u^α , whereas we allow the rest of the system additional freedom to choose a different v^α for each state k . Minimizing over the u^α , $v^{k,\alpha}$, and a_α^k , we find

$$\rho u^\alpha = w_\alpha u^\alpha \quad (26)$$

with

$$\rho_{ii'} = \sum_k W_k \sum_j \psi_{ij}^k \psi_{i'j}^k \quad (27)$$

and

$$w_\alpha = \sum_k W_k (a_\alpha^k)^2. \quad (28)$$

This equation for ρ is the definition of the reduced density matrix when the superblock is in a mixed state, and the u^α are the eigenstates of ρ .

Thus the conclusion when the superblock is in a mixed state is identical to the result for a pure state: the optimal states to keep are the eigenvectors of the reduced density matrix with the largest eigenvalues.

The effectiveness of the truncation of the Hilbert space of the system block via the density matrix depends crucially on the distribution of the density matrix eigenvalues w_α . For certain exactly solvable systems, it is possible to determine the density matrix eigenvalues exactly. For these integrable systems, the distribution of the density matrix eigenvalues can be shown to decay exponentially, see Chap. 3.1(II).

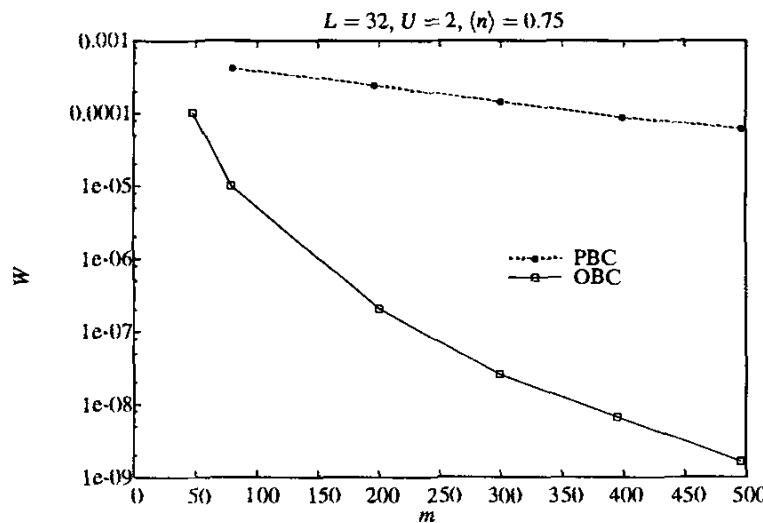


Fig. 4. Sum of the discarded weight of the density matrix eigenvalues for the finite system DMRG algorithm, as a function of the number of states kept m .

As a realistic example of the distribution of the density matrix eigenvalues, we give numerical results for the one-dimensional Hubbard model obtained from the finite size DMRG algorithm, which will be discussed in the next section. In Fig. 4, we plot the sum of the discarded density matrix eigenvalues, $\sum_{\alpha=m+1}^{m_{\max}} w_\alpha$, where m_{\max} is the size of the density matrix. This discarded weight is plotted in Fig. 4 as a function of m on a logarithmic scale for the 32-site Hubbard chain with open and periodic boundary conditions at $n = 0.75$. As can be seen, the discarded weight falls off rapidly with the number of states kept. For open boundary conditions, the decay is clearly slower than exponential for small m , in contrast to the original results found

for the spin-one Heisenberg chain [1]. This could be because the system is in the Luttinger liquid regime and therefore has no gap in either spin or charge excitations. It is possible that the decay becomes exponential at large enough m , when the error in the energy becomes less than the finite energy level spacing always found on a finite system. This discarded weight is strongly correlated with the error in the ground state energy and is often used as a measure of the error. It has been used to make an $m \rightarrow \infty$ extrapolation of the energy [14]. Also depicted in Fig. 4 is the discarded weight for the same system with periodic boundary conditions. The convergence of the discarded weight (and thus the ground state energy) with m is much slower than for open boundary conditions. It is generally found that the accuracy of the energy for a given m is many orders of magnitude worse for periodic than for open boundary conditions. It is therefore usually better to treat systems with open boundary conditions on much larger lattices rather than systems with periodic boundary conditions.

5 DMRG Algorithms

In this section, we will describe how to combine the superblock procedure with the density matrix projection in order to define efficient DMRG algorithms. There are three main ingredients needed to form a DMRG algorithm: first, we have to decide how to add degrees of freedom to the system, i.e. how to build up the system block; second, we have to determine the configuration of the superblock; and finally, we must choose which superblock eigenstate or eigenstates to use to construct the density matrix.

For interacting systems, it is clear that one wants to add the minimum number of degrees of freedom at once to the system block in order to keep as large a fraction of the system block states as possible, and to keep the size of the Hilbert space of the superblock as small as possible. Therefore, one usually wants to build up the system block one site at a time in a procedure similar to that described for the Wilson numerical RG in Sect. 2.

The algorithms then fall into two classes, depending on how the environment block is chosen to form the superblock: the infinite system algorithm and the finite system algorithm. We will discuss these algorithms in detail below.

We will call the superblock state or states used to form the reduced density matrix for the system block *target states*. If only ground state properties are desired, it is most accurate to target just the ground state of the superblock. (The Hamiltonian is usually block diagonal in particular quantum numbers such as S_z ; by ground state we will mean ground state for a particular quantum number.) If excited states or matrix elements between different states are required, more than one target state can be used. However, for fixed number of states kept m , the accuracy with which the properties of each individual state can be determined goes down as more states are tar-

geted. For simplicity, we will assume that only the ground state is targeted in the following.

The infinite system algorithm

The infinite system algorithm is the most straightforward extension of the Wilson procedure described in Sect. 2 that incorporates the superblock concept. We build up the system block one site at a time, just as in the Wilson procedure, but must choose some sort of environment block. The simplest way of forming the environment block is to use a reflection of the system block. The superblock configuration is shown in Fig. 5. Here \bar{H}_ℓ is the Hamiltonian for the system block in the reduced basis, as before, and the solid dots represent single sites. The right block, \bar{H}_ℓ^R , is formed by relabeling the sites in the system block so that they are reflected onto the right part of the lattice.

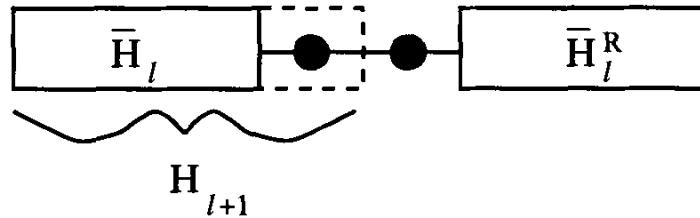


Fig. 5. The superblock configuration for the infinite-system algorithm.

The infinite system algorithm then proceeds as follows:

1. Form a superblock containing L sites which is small enough to be exactly diagonalized.
2. Diagonalize the superblock Hamiltonian H_L^{super} numerically, obtaining *only* the ground state eigenvalue and eigenvector $|\psi\rangle$ using the Lanczos or Davidson algorithm.
3. Form the reduced density matrix $\rho_{ii'}$ for the new system block from $|\psi\rangle$ using (15). Note that $\ell' = \ell = L/2 - 1$.
4. Diagonalize $\rho_{ii'}$ with a dense matrix diagonalization routine to obtain the m eigenvectors with the largest eigenvalues.
5. Construct $H_{\ell+1}$ and other operators in the new system block and transform them to the reduced density matrix eigenbasis using $\bar{H}_{\ell+1} = O_L^\dagger H_{\ell+1} O_L$, $\bar{A}_{\ell+1} = O_L^\dagger A_{\ell+1} O_L$, etc., where the columns of O_L contain the m highest eigenvectors of $\rho_{ii'}$, and $A_{\ell+1}$ is an operator in the system block.
6. Form a superblock of size $L + 2$ using $\bar{H}_{\ell+1}$, two single sites and $\bar{H}_{\ell+1}^R$.
7. Repeat starting with step 2, substituting H_{L+2}^{super} for H_L^{super} .

It is clear that this algorithm is very much in the spirit of the original Wilson procedure in that the system being diagonalized grows at each step. There

are, however, a few important differences. First, as in the noninteracting superblock procedure, the diagonalization step and the real-space blocking step take place on different size systems. Therefore, the energy and various expectation values are calculated during the superblock diagonalization, while the density matrix diagonalization rather than a Hamiltonian diagonalization is used to determine the new basis for the system block. Second, the size of the superblock grows by two sites rather than one site at every step. Third, we have assumed that the system is reflection symmetric. It is possible to formulate algorithms that do not assume reflection symmetry, but this is done most easily in the context of the finite system algorithm described below.

The finite system algorithm

In the finite size algorithm, the environment block is chosen in a different way: it is chosen so that the size of the superblock is kept fixed at each step. Suppose that we have run the infinite system algorithm until the superblock reaches size L , but have stored all the $\bar{H}_{\ell'}^R$ for $\ell = 1, \dots, L/2 - 2$ as well as all the additional operators needed to connect the blocks at each step. We can then continue to build up the system block, but keep $L = \ell + \ell' + 2$ fixed by using the appropriate previously stored $\bar{H}_{\ell'}^R$. The finite size algorithm then proceeds as follows:

0. Carry out the infinite system algorithm until the superblock reaches size L , storing \bar{H}_ℓ and the operators needed to connect the blocks at each step.
1. Carry out steps 3-5 of the infinite system algorithm to obtain $\bar{H}_{\ell+1}$. Store it. (Now $\ell \neq \ell'$.)
2. Form a superblock of size L using $\bar{H}_{\ell+1}$, two single sites and $\bar{H}_{\ell'-1}^R$. The superblock configuration is given by Fig. 6, where $\ell' = L - \ell - 2$.
3. Repeat steps 1-2 until $\ell = L - 3$ (i.e. $\ell' = 1$). This is the *left to right* phase of the algorithm.
4. Carry out steps 3-5 of the infinite system algorithm, reversing the roles of \bar{H}_ℓ and $\bar{H}_{\ell'}^R$, i.e. switch directions to build up the right block and obtain $\bar{H}_{\ell'+1}^R$. Store it.
5. Form a superblock of size L using $\bar{H}_{\ell-1}$, two single sites and $\bar{H}_{\ell'+1}^R$.
6. Repeat steps 4-5 until $\ell = 1$. This is the *right to left* phase of the algorithm.
7. Repeat starting with step 1.

A useful analogy is to think of this procedure as being like running a zipper repeatedly from left to right and then right to left through a superblock that is always the same size. Each time the zipper changes direction, a new set of stored blocks is used as the environment block. In this way, the representations of the stored blocks are iteratively improved and the zipping can be repeated until convergence is reached.

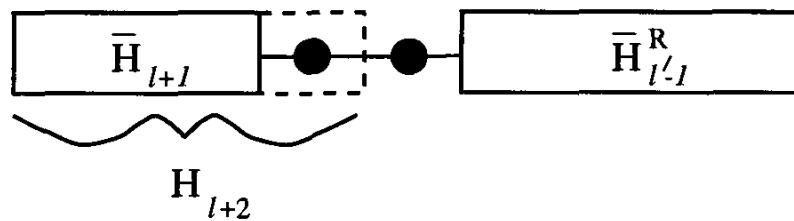


Fig. 6. A typical superblock configuration for the left-to-right phase of the finite-system algorithm.

Note that since the left block and the right blocks are stored independently, we do not have to assume that the lattice is reflection symmetric (at least after step 0). Since the same size superblock is always diagonalized, the algorithm is less dependent than the infinite system algorithm on translational invariance, i.e. on the optimum representation of different size superblocks being similar.

If reflection symmetry is present, it can be used at the point at which $\ell = \ell'$ to shorten the length of the zips. One way of formulating the algorithm in this case is to build up the left blocks from $\ell = 1$ to $\ell = L/2 - 1$, and build up the right blocks from $\ell' = L/2$ to $\ell' = L - 3$, i.e. to only zip from the left side of the superblock to the middle and then back to the left side. The fact that we have used reflection symmetry in the infinite system phase, step 0, is usually not important. However, it is also possible to formulate infinite system algorithms that do not use reflection symmetry. This issue will be discussed in more detail below in the context of algorithms for two-dimensional and fermion systems.

For a given system size L , the finite system algorithm almost always gives substantially more accurate results than the infinite system algorithm, and is therefore usually preferred unless there is a specific reason to go to the thermodynamic limit.

It is instructive to examine how the infinite and finite system algorithms work in detail on a particular system. Here we will utilize the one-dimensional Hubbard model, (2), as an example. This model is fairly complicated, containing both spin and fermionic degrees of freedom, but has a Bethe Ansatz exact solution [15], which has been extended to the case of open boundary conditions [16]. Here we will investigate how the energy obtained from the finite-system DMRG algorithm converges to the exact Bethe Ansatz energy on a particular system. This is done for a fractionally filled 32-site system with open boundary conditions in Fig. 7. The ordinate “Iteration” refers to the iteration of reflection symmetric finite system algorithm, for which each iteration consists of a right-to-left “zip” from center (the reflection symmetric configuration) to the left edge, followed by a left-to-right “zip” back to the center.

As can be seen from the curves which were made for different runs with different numbers of states m kept in the system block, there is virtually no

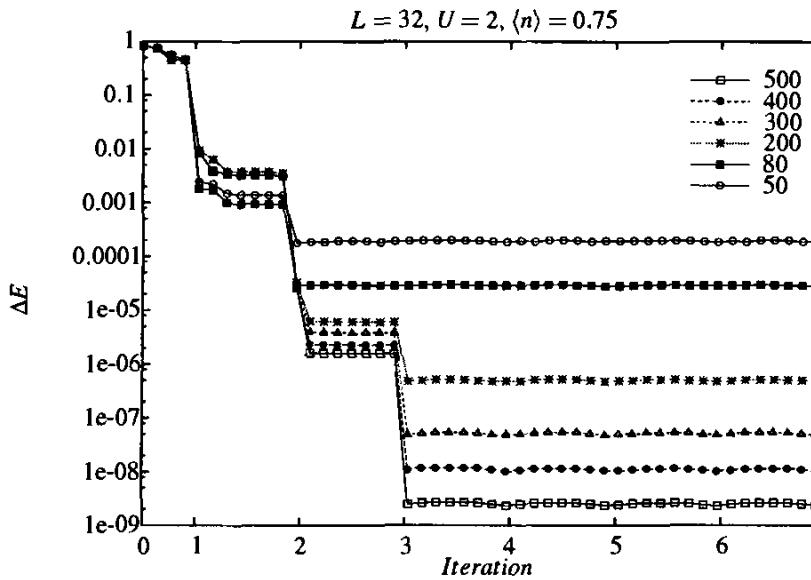


Fig. 7. The difference between the ground-state energy obtained using the finite system DMRG and the exact Bethe Ansatz energy for a 32-site system at $U = 2$ and filling $\langle n \rangle = 0.75$, plotted on a logarithmic scale as a function of iteration of the finite system algorithm. Each point represents a Davidson diagonalization of the superblock (every fourth diagonalization is plotted), and the different symbols represent different runs made with a fixed number $m = 50, \dots, 500$ of states kept.

dependence of the accuracy of the energy on m in the zeroth iteration, the infinite system phase. This behavior is typical for fractionally filled fermion systems and for finite-width systems. It is due to the poor job the infinite system algorithm does at representing systems with an intrinsic length, such as the wavelength of local density oscillations or the width of a two-dimensional system. After two to three iterations, the error in the energy saturates for a given m , with the saturation occurring after a larger number of iterations for larger m . Therefore, it is only helpful to increase m once sufficient convergence in the number of iterations has been achieved. The optimum strategy is then to increase m fairly rapidly at each iteration. If this is done, the majority of time is spent doing the last iteration, and the finite system algorithm only takes a factor of two to three longer than the infinite system algorithm for the same number of states, but can be orders of magnitude more accurate. Note that the rate of convergence with the number of iterations can vary strongly from model to model, so the optimum strategy will also vary.

Once the optimum convergence is achieved, the crucial question is how the accuracy of the energy depends on the maximum number of states kept in a run, i.e. on how the height of the plateaus in Fig. 7 depends on m . This dependence is illustrated in Fig. 8 for different U values. The form of the decay in the error in the ground state energy with m is very similar to the form of decay of the discarded weight of the density matrix eigenvalues, Fig. 4. The error in the energy is usually proportional to the discarded weight, once the algorithm has converged sufficiently. Another interesting point is that the rate

of convergence is almost completely independent of the interaction strength U . In fact, we have found that even adding a nearest-neighbor Coulomb interaction does not substantially change the convergence with m [17], as long as the system is the Luttinger liquid regime, i.e. has no spin or charge gap. The nature and density of the low-lying excitations in the superblock, which does not change much in a Luttinger liquid, seems to be more important than the nature and range of the interactions.

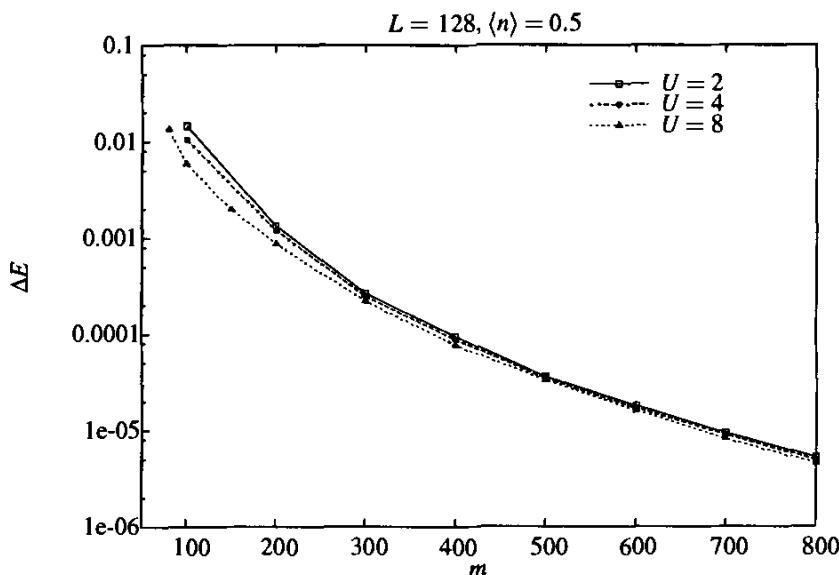


Fig. 8. Convergence of the finite system algorithm as a function of the number of states kept m for different U .

Details for interacting systems

Up to now, we have not considered in detail how to store and transform the operators necessary to carry out the renormalization group transformation for an interacting system. In this section, we will discuss how to do this efficiently.

In order to construct the Hamiltonian of the system, a block must have various operators stored as matrices connecting these states. For example, for the Heisenberg model with exchange terms

$$\mathbf{S}_i \cdot \mathbf{S}_{i+1} = S_i^z S_{i+1}^z + \frac{1}{2}(S_i^+ S_{i+1}^- + S_i^- S_{i+1}^+) \quad (29)$$

one needs to store $m \times m$ matrix representations of S_i^z , S_i^+ , and S_i^- for i equal to the left or right end sites of the block. (In practice, one need not store S_i^- , since it can be obtained by taking the Hermitian conjugate of S_i^+). For a Hubbard model one would have to store matrices for $c_{i\sigma}$, with $\sigma = \uparrow$ and \downarrow , in order to reconstruct the hopping term $\sum_{\sigma} (c_{i+1\sigma}^\dagger c_{i\sigma} + c_{i\sigma}^\dagger c_{i+1\sigma})$.

Consider joining two blocks B_1 and B_2 together in a Heisenberg system. In the Wilson procedure, B_2 will typically consist of a single site, and for the DMRG algorithms, it is sufficient to consider how to compose two blocks. If B_1 has m_1 states, and B_2 has m_2 states, the combined block B_1B_2 has m_1m_2 states. We label the combined states by two indices, ij . The matrix representing the Hamiltonian of B_1B_2 is then given by

$$\begin{aligned} [H_{B_1B_2}]_{ij;i'j'} &= [H_{B_1}]_{ii'} \delta_{jj'} + [H_{B_2}]_{jj'} \delta_{ii'} + [S_\ell^z]_{ii'} [S_{\ell+1}^z]_{jj'} \\ &\quad + \frac{1}{2} [S_\ell^+]_{ii'} [S_{\ell+1}^-]_{jj'} + \frac{1}{2} [S_\ell^-]_{ii'} [S_{\ell+1}^+]_{jj'} \end{aligned} \quad (30)$$

where H_{B_1} is the Hamiltonian matrix of block B_1 , and ℓ is its rightmost site. In order for the connections in the Hamiltonian to be restored when the two blocks are combined, each block must contain each of the matrices appearing in (30).

The most time-consuming part of a DMRG calculation is the diagonalization of the system Hamiltonian, which occurs once for every step. Only the ground state or a few low-lying states are needed, and so the Lanczos or Davidson iterative algorithms should be used. These algorithms both require repeated multiplications of superblock vectors ψ by the superblock Hamiltonian H^{super} . However, the actual Hamiltonian matrix should not be constructed and stored. Instead, the following procedure to directly multiply $H^{\text{super}}\psi$ uses much less memory and is also much faster. Consider again, for simplicity, a system formed from two blocks. The superblock Hamiltonian can be written in the general form

$$[H^{\text{super}}]_{ij;i'j'} = \sum_{\alpha} A_{ii'}^{\alpha} B_{jj'}^{\alpha}. \quad (31)$$

Then the product $H^{\text{super}}\psi$ can be written as

$$\sum_{i'j'} [H^{\text{super}}]_{ij;i'j'} \psi_{i'j'}^{\alpha} = \sum_{\alpha} \sum_{i'} A_{ii'}^{\alpha} \sum_{j'} B_{jj'}^{\alpha} \psi_{i'j'}^{\alpha}. \quad (32)$$

For each α , the last sum is performed first, as a matrix-matrix multiplication of B^{α} and ψ^T , to form a temporary matrix $C_{ji'}^{\alpha}$. Then a matrix-matrix multiplication of A^{α} and $[C^{\alpha}]^T$ forms a partial result, which is added into the result vector, giving a sum on α .

Whenever a site is added onto a block, or more generally two blocks are added, the operator matrices must be updated. The eigenstates of the density matrix can be written in the form u_{ij}^{α} , which we write as $O_{ij;\alpha}$, $\alpha = 1, \dots, m$. Here i and j represent state indices of the two blocks that are being added together. Then for each operator A that is needed, $A_{ij;i'j'}$ is replaced by $A_{\alpha\alpha'}$, where

$$A_{\alpha\alpha'} = \sum_{iji'j'} O_{ij;\alpha} A_{ij;i'j'} O_{i'j';\alpha'}. \quad (33)$$

The terms appearing in (30) show the various ways operators $A_{ij,i'j'}$ can be formed from single-block operators $A_{ii'}$.

Any efficient DMRG program must make use of quantum numbers to speed up the calculation and reduce storage. For example, in order to construct the system Hamiltonian it may be necessary to store for a block the matrix form of the operator \hat{S}_ℓ^+ , where ℓ is the right-most site of the block. If there are m states in the block, this is an $m \times m$ matrix. However, if states are labeled and grouped by block quantum number S^z , then this matrix is mostly zeroes, with the nonzero parts in rectangular blocks. These blocks connect states with specific quantum numbers, e.g. the states corresponding to the left index of the matrix may have $S_z = 0$, and for the right index $S_z = -1$. It is essential to store only the nonzero elements of this matrix. Although this can be done using sparse matrices, the best way to do it is as a set of dense matrices, one for each nonzero rectangular block. The multiplication of $H^{\text{super}}\psi$ described above takes place as described in the previous paragraph, except that now there is an additional sum or loop over quantum numbers, and the dense matrices which are multiplied are much smaller. Keeping track of all the matrices, each of a different size, can be very well organized in C++ by defining classes to represent operator matrices, submatrices, etc. The fact that these matrices are all different sizes and have different dimensions at each DMRG step makes it somewhat more difficult to use storage efficiently in Fortran 77, which does not have dynamic allocation of memory.

It is useful at this point to mention typical maximum numbers of states kept, m , for various systems on current computers. For the one-dimensional Hubbard model, $m = 800$ for a system of up to a few hundred sites can be treated on typical workstation, using a few hundred megabytes of main memory [17,18]. For the Heisenberg model, up to $m = 1100$ states have been kept for the spin-one chain with an impurity [19] and $m = 1700$ for the spin-two chain [20]. These Heisenberg-chain calculations have been carried out using a highly optimized Fortran code.

Measurements

Measurements are made using the superblock wavefunction $|\psi\rangle$ to evaluate expectation values of the form $\langle\psi|A|\psi\rangle$. Rather complicated operators can be evaluated fairly easily, but dynamical information is more difficult to obtain. In order to measure A , one must have kept operator matrices for the components of A . For example, to measure the on-site spin-density S_j^z for all sites ℓ , one must keep track of matrices $[S_\ell^z]_{ii'}$, for all each site ℓ in each of the blocks. These operators must be updated using (33) at every step of each iteration. We will once again divide superblock into two parts with states labeled by $|i\rangle$ and $|j\rangle$. One then obtains the expectation value using

$$\langle\psi|S_\ell^z|\psi\rangle = \sum_{i,i',j} \psi_{ij}^* [S_\ell^z]_{ii'} \psi_{i'j}, \quad (34)$$

etc. This procedure gives *exact* evaluations of $\langle \psi | A | \psi \rangle$ for the *approximate* eigenstate $|\psi\rangle$.

For a correlation function such as $\langle \psi | S_\ell^z S_m^z | \psi \rangle$, the evaluation depends on whether ℓ and m are on the same block or not. If they are on different blocks, then one need only have kept track of $[S_\ell^z]_{ii'}$ and $[S_m^z]_{jj'}$, and one has

$$\langle \psi | S_\ell^z S_m^z | \psi \rangle = \sum_{i,i',j,j'} \psi_{ij}^* [S_\ell^z]_{ii'} [S_m^z]_{jj'} \psi_{i'j'} . \quad (35)$$

If ℓ and m are on the same block, one *should not* use

$$\langle \psi | S_\ell^z S_m^z | \psi \rangle \approx \sum_{i,i',i'',j} \psi_{ij}^* [S_\ell^z]_{ii'} [S_m^z]_{i'i''} \psi_{i''j} . \quad (36)$$

This expression does not evaluate the correlation function exactly within the approximate state $|\psi\rangle$. The sum over i' should run over a complete set of states, but does not, whereas the sums over the other variables need run only over those states needed to represent $|\psi\rangle$, since they appear as a subscript in either the $|\psi\rangle$ on the left or on the right.

To evaluate this type of correlation function, one needs to have kept track of $[S_j^z S_k^z]_{ii'}$ throughout the calculation. One then evaluates

$$\langle \psi | S_\ell^z S_m^z | \psi \rangle = \sum_{i,i',j} \psi_{ij}^* [S_\ell^z S_m^z]_{ii'} \psi_{i'j} \quad (37)$$

to obtain the correlation function.

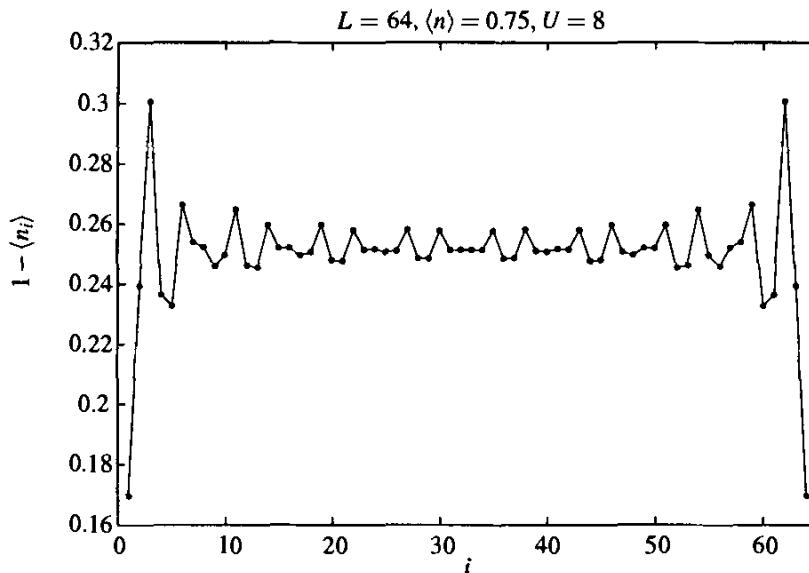


Fig. 9. Local hole density $1 - \langle n_i \rangle$ plotted as a function of i .

As an example of measurements of the first type, (34), we display the local hole density $1 - \langle n_i \rangle$, where $n_i \equiv n_{i\uparrow} + n_{i\downarrow}$ for the one-dimensional

Hubbard model in Fig. 9. The oscillations are Friedel oscillations due to the open boundary conditions.

In Fig. 10, we show the nearest-neighbor spin-spin correlations for the half-filled system as an example of a correlation function between different sites, calculated using (37). This correlation function also has oscillations due to the open boundaries. Since the correlation functions cannot be calculated directly from the Bethe Ansatz, it is not possible to directly calculate the error, as for the ground state energy. The issue of how to form a good approximation to correlation functions in the thermodynamic limit using these finite-lattice correlation functions, as well as a comparison with the asymptotic results obtain from a combination of bosonization and the Bethe Ansatz [21] is discussed in more detail in Chap. 1.1(II).

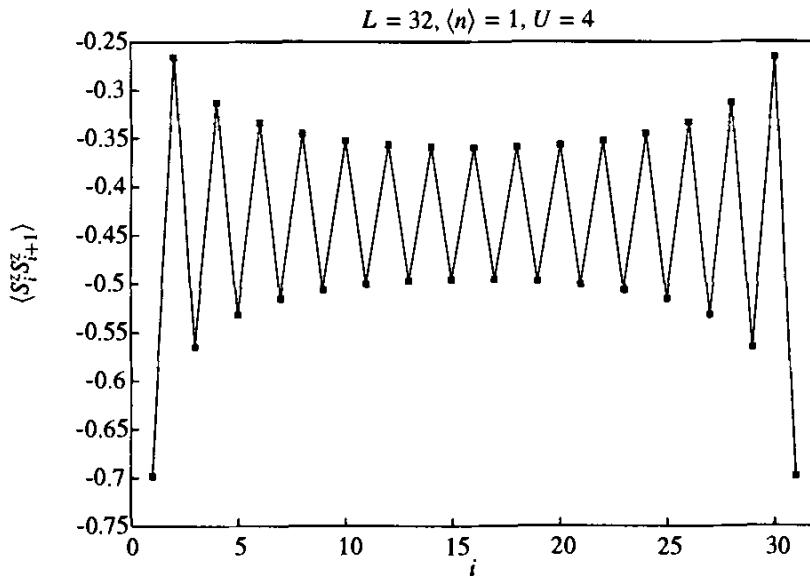


Fig. 10. Nearest neighbor spin-spin correlation function $\langle S_i^z S_{i+1}^z \rangle$ plotted as a function of i for the half-filled system.

Wave function transformations

An important improvement in the efficiency of DMRG comes from keeping track of the wavefunction from step to step [22]. The step referred to here is the process of adding a site to a block and requires the diagonalization of a system configuration of two blocks and two sites [1]. In each DMRG step, an iterative sparse matrix algorithm, such as the Davidson method, is used to find the ground state of the system. Thus far, we have not specified a starting point for the Davidson procedure. To ensure that the DMRG procedure is always stable and convergent, the system ground state usually has to be determined to rather high accuracy. (One diagonalization which converges to a low-lying eigenstate other than the ground state can ruin the accuracy of

the entire DMRG sweep.) Consequently, a substantial number of Davidson steps are necessary to converge to sufficient accuracy, typically 40-100. The total calculation time is proportional to the average number of Davidson steps.

If a very good initial guess is available for the Davidson procedure, the number of Davidson steps can be reduced substantially. An ideal initial guess, for the case of the finite system DMRG algorithm, is the final wavefunction from the previous DMRG step. This wavefunction, however, is in a different basis, corresponding to a different system configuration, but it can be transformed into the basis corresponding to the current configuration, as described below. Use of this transformation to obtain the initial state in a Davidson diagonalization can reduce the number of Davidson steps by one half, typically, assuming that one iterates Davidson until it converges to high accuracy. Use of this initial guess has an even more important advantage: it is not necessary to converge to high accuracy, since there is no danger of converging to an incorrect low-lying eigenstate. The initial guess not only has low energy, it approximately describes the correct eigenstate, as obtained in the previous step. Hence a lower accuracy convergence is possible without damaging the stability of DMRG. In fact, the algorithm can be made completely stable even if the number of Davidson steps is restricted to two or three! Thus one saves a factor of 20-50 in the time required by the Davidson procedure. The overall speedup is somewhat reduced from this factor because the calculation time to perform other parts of the DMRG procedure, such as diagonalizing the density matrix, becomes significant.

A DMRG step adds a site onto a block, constructing an appropriate basis for the new block. The new basis is defined by the eigenvectors of the density matrix, which is determined from the current system wavefunction. The eigenvectors defining the new basis (which we rewrite as matrices L and R below) can then be used to transform the current system wavefunction into the appropriate basis for the *next* step. Here we consider a finite-system DMRG step, in which we move from left to right, adding a site onto the left block, in the standard configuration with two sites between the left and right blocks. Let $|\alpha_\ell\rangle$ be the states of left block ℓ , where ℓ is the rightmost site of the block. The two sites in the middle are $\ell + 1$ and $\ell + 2$. Let $|s_\ell\rangle$ be the states of site ℓ , $|s_{\ell+1}\rangle$ for site $\ell + 1$, etc. Then the basis states for the new left block are given by

$$|\alpha_{\ell+1}\rangle = \sum_{s_{\ell+1}, \alpha_\ell} L^{\ell+1}[s_{\ell+1}]_{\alpha_{\ell+1}, \alpha_\ell} |\alpha_\ell\rangle \otimes |s_{\ell+1}\rangle. \quad (38)$$

This notation is similar to that of Östlund and Rommer [23]. For more details, see Chap. 3(I). The transformation matrix $L^{\ell+1}[s_{\ell+1}]_{\alpha_{\ell+1}, \alpha_\ell}$ is a slightly rewritten form of the truncated matrix of density matrix eigenvectors $u^{\alpha_{\ell+1}}$: specifically, $L^{\ell+1}[s_{\ell+1}]_{\alpha_{\ell+1}, \alpha_\ell} = u_{s_{\ell+1} \alpha_\ell}^{\alpha_{\ell+1}}$. L includes only the eigenvectors which are retained, i.e. whose corresponding eigenvalues are above a cut-off. Let the states of the right block be $|\beta_{\ell+3}\rangle$, where we note that $\ell + 3$ is the

leftmost site of the block. These states were formed at an earlier right-to-left DMRG step, using a different set of density matrix eigenvectors, which we write in terms of a transformation matrix $R^{\ell+3}$:

$$|\beta_{\ell+3}\rangle = \sum_{s_{\ell+3}, \beta_{\ell+4}} R^{\ell+3}[s_{\ell+3}]_{\beta_{\ell+3}, \beta_{\ell+4}} |s_{\ell+3}\rangle \otimes |\beta_{\ell+4}\rangle . \quad (39)$$

Note that reflection symmetry is not assumed here: the L and R matrices are independent.

The wavefunction is written in a basis for the two block plus two site superblock. This superblock basis has basis states of the form

$$|\alpha_\ell s_{\ell+1} s_{\ell+2} \beta_{\ell+3}\rangle = |\alpha_\ell\rangle \otimes |s_{\ell+1}\rangle \otimes |s_{\ell+2}\rangle \otimes |\beta_{\ell+3}\rangle . \quad (40)$$

A system wavefunction $|\psi\rangle$ is written in this basis as

$$|\psi\rangle = \sum_{\alpha_\ell s_{\ell+1} s_{\ell+2} \beta_{\ell+3}} \psi(\alpha_\ell s_{\ell+1} s_{\ell+2} \beta_{\ell+3}) |\alpha_\ell s_{\ell+1} s_{\ell+2} \beta_{\ell+3}\rangle . \quad (41)$$

One needs to transform this wavefunction into the basis appropriate for the next DMRG step, in which the superblock is shifted by one site, with basis states of the form $|\alpha_{\ell+1} s_{\ell+2} s_{\ell+3} \beta_{\ell+4}\rangle$. However, the transformation between the two bases cannot be exact, since there is a truncation in going from $|\alpha_\ell s_{\ell+1}\rangle$ to $|\alpha_{\ell+1}\rangle$. However, the states $|\alpha_{\ell+1}\rangle$ are formed using the density matrix to be ideally adapted for representing $|\psi\rangle$. This means that the wavefunction can be transformed in an approximate but controlled fashion, with the error in the transformation depending on the truncation error in the DMRG step. Since the error in the density matrix is given by the truncation error, and since the density matrix is, roughly speaking, the square of the wavefunction, the error in the wavefunction transformation should be roughly the square root of the truncation error.

The simplest way to derive the transformation is to assume, based on the above argument, that for the transformation of the wavefunction only, one can approximate

$$\sum_{\alpha_{\ell+1}} |\alpha_{\ell+1}\rangle \langle \alpha_{\ell+1}| \approx 1 . \quad (42)$$

With this approximation one readily obtains

$$\begin{aligned} \psi(\alpha_{\ell+1} s_{\ell+2} s_{\ell+3} \beta_{\ell+4}) &\approx \\ \sum_{\alpha_\ell s_{\ell+1} \beta_{\ell+3}} L^{\ell+1}[s_{\ell+1}]_{\alpha_{\ell+1}, \alpha_\ell} \psi(\alpha_\ell s_{\ell+1} s_{\ell+2} \beta_{\ell+3}) R^{\ell+3}[s_{\ell+3}]_{\beta_{\ell+3}, \beta_{\ell+4}} . \end{aligned} \quad (43)$$

This is the desired transformation.

The most efficient way to implement this transformation numerically is to first form the intermediate wavefunction

$$\psi(\alpha_{\ell+1} s_{\ell+2} \beta_{\ell+3}) = \sum_{\alpha_\ell s_{\ell+1}} L^{\ell+1}[s_{\ell+1}]_{\alpha_{\ell+1}, \alpha_\ell} \psi(\alpha_\ell s_{\ell+1} s_{\ell+2} \beta_{\ell+3}) , \quad (44)$$

and then form the final result

$$\psi(\alpha_{\ell+1}s_{\ell+2}s_{\ell+3}\beta_{\ell+4}) = \sum_{\beta_{\ell+3}} \psi(\alpha_{\ell+1}s_{\ell+2}\beta_{\ell+3}) R^{\ell+3}[s_{\ell+3}]_{\beta_{\ell+3}, \beta_{\ell+4}}. \quad (45)$$

In this form, the transformation requires very little computer time compared to other parts of the calculation.

This transformation is used for one half of the DMRG steps, when a site is being added to the left block. An analogous transformation is used for adding a site to the right block.

Implementing this transformation requires saving all the matrices L and R , which was not necessary in the original formulation of DMRG. The storage for these matrices is typically 20-30% of the storage required for the blocks themselves, so the extra storage is not a major concern. In an efficient DMRG implementation for a typical single processor workstation, both the blocks and the transformation matrices should be stored on disk.

Extension to two dimensions

The issue of how to optimally extend the DMRG to two or more dimensional quantum systems is difficult and is still a subject of active development. It is instructive to consider first how to extend the one-dimensional algorithm to quasi-one-dimensional systems with a finite width, e.g. to the Heisenberg or Hubbard model on a “ladder”. One straightforward way to do this would be to replace the single sites added between the blocks with a row of sites. However, for wide systems the extra degrees of freedom in the two center “sites” would make the size of the system’s Hilbert space prohibitively large. It is usually better to add single sites by mapping the 2D system onto a 1D system, simply by tracing a path through the lattice. A typical superblock configuration for a ladder system is shown in Fig. 11. The site added to the system block is enclosed by a dashed line and the dotted line shows the order in which sites are added in a sweep. An up-down-up-down path is shown; an up-up-up path would be just as efficient. One can see that it is not possible to reflect the left block onto a right block of the proper geometry at every step, so the finite system algorithm must be used. A similar difficulty appears in fermion systems at fractional filling because it is not possible to set the number of particles so that the average density stays constant at all system sizes. The denominator of the fraction of the filling is analogous to the row size for two-dimensional lattices. With the mapping onto a 1D system, the two-dimensional procedure differs from the one-dimensional finite system procedure only in that there are additional connections between the system and environment blocks along the boundary.

Several infinite system approaches can be used in the warmup sweep [24]. The simplest is just to use several sites for the environment block, without truncation. Usually the accuracy of the warmup sweep is not critical; a few

sweeps of the finite-system procedure can make up for a poor quality warmup sweep.

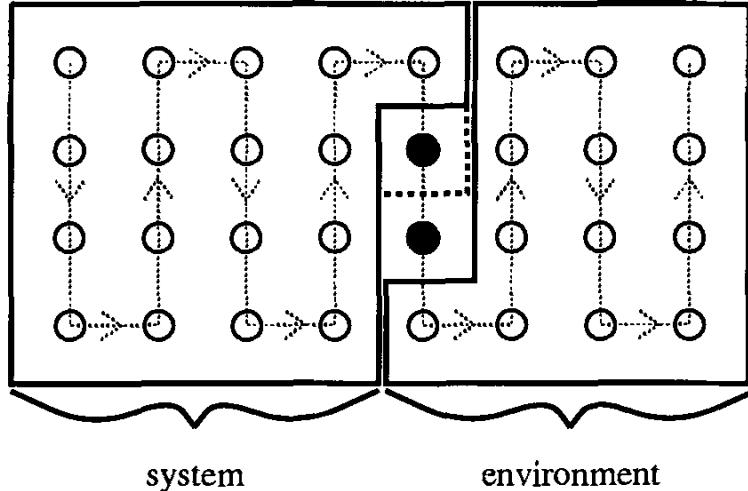


Fig. 11. The superblock configuration for the two-dimensional algorithm. The order in which sites are added to the system block on a series of iterations is given by the dotted line, and the site added to the approximate system block Hamiltonian is outlined by the dashed line.

The number of states needed to maintain a certain truncation error in the density matrix projection procedure depends strongly on the number of operators connecting the two parts of the system. Best accuracy is obtained when the number of connections between the system and environment blocks is minimized. Therefore, we usually study systems with open rather than periodic or antiperiodic boundary conditions. Also, we find that the number of states m needed to maintain a given accuracy depends strongly on the width and weakly on the length of the system.

Just how rapidly the truncation error increases with the width of the system is not clear in general. Liang and Pang [25] studied the error in the energy as a function of width for a gas of noninteracting spinless fermions and found that the number of states needed to maintain a given accuracy grew exponentially with the width of the system. In an interacting system such as the Hubbard model, the detailed structure of the energy spectrum seems to be important. For example, in the two chain Hubbard model at half-filling, where there is a spin and pairing gap, the truncation error for a given m is much smaller than away from half-filling, where the spin gap is reduced and the gap to pairing excitations is no longer present. For Heisenberg ladders, the presence or absence of a gap in the spin spectrum depends on whether the number of chains is even or odd, so the truncation error for a given m depends on the number of chains in a complicated way. In any case, for systems of more than one dimension, it is important to be able to keep as many states m per block as possible. Efficient algorithms are therefore crucial in this case.

6 Remarks

In this chapter, we have described the Density Matrix Renormalization Group as originally developed to determine the properties of low-lying states of interacting quantum systems. We have followed the historical development of the method in order to motivate the ingredients that go into the algorithms. We have also tried to strike a balance between a pedagogical explanation of the basic ideas and the inclusion of sufficient detail to allow the reader to produce efficient implementations. In order to provide a good starting point for understanding the basic ideas and limitations of various numerical RG schemes, the simplest problem, the tight-binding chain, was presented in considerable detail. As the best way to learn the technique is to implement a scheme on the computer and to play with it, we have included, in Appendix B, a program which treats this problem as an example.

The DMRG is still under quite active development. Since the basic ideas seem to be quite general and the algorithms well-suited for experimenting, there is hope that significant improvements will be made to both performance and applicability in the future. As we have described, the most serious limitation of the original ground-state algorithm is the strong dependence of its convergence on the degree of two-dimensionality and on the complexity and range of the interaction. Up to now, progress has been made through improved efficiency of the implementations and better performance of the computers. A variation of the algorithm that scales more favorably for higher dimensional systems would, however, be quite useful. The momentum-space formulation discussed in Chap. 6(I) could be a step in this direction.

There has also been much recent activity on extending the DMRG and in applying it to new problems. In particular, there have been interesting developments in extending the DMRG to calculate dynamic and finite-temperature properties of interacting quantum systems, in applying it to two-dimensional classical systems, and in studying its behavior analytically by examining its behavior near the infinite-system fixed point and by making contact with variational states. These developments are the subject of the remainder of Part I of this book, and a wide variety of applications are discussed in Part II.

A Density Matrix for Single Particle Systems

Here we show, in the case of a single particle, that the density matrix is equivalent to a simple projection of the wavefunction, used in the superblock method for the tight-binding chain described in Sect. 4.

Consider a wavefunction $\psi(k)$, where k runs over the sites of the system, $k = 1, \dots, L$. We will call sites $1, \dots, \ell$ the left block, labeled by i , and sites

$\ell + 1, \dots, L$ the right block, labeled by j . In order to write a single-particle wavefunction in a product form

$$|\psi\rangle = \sum_{ij} \psi_{ij} |i\rangle |j\rangle , \quad (\text{A.1})$$

it is necessary to construct an enlarged basis which includes zero and two-particle states. Specifically, we use the basis

$$\begin{aligned} &|0\rangle_L \\ &|1\rangle_L = c_1^\dagger |0\rangle_L \\ &\vdots \\ &|\ell\rangle_L = c_\ell^\dagger |\ell\rangle_L \end{aligned} \quad (\text{A.2})$$

for the left block, and similarly for the right block. Here c_i^\dagger creates a particle at site i . Then the wavefunction $\psi(k)$ can be written in matrix form as

$$\psi = \begin{pmatrix} 0 & \psi(\ell+1) \dots \psi(L) \\ \psi(1) & \\ \vdots & 0 \\ \psi(\ell) & \end{pmatrix} . \quad (\text{A.3})$$

In this matrix, the upper left zero represents the amplitude in the state $|0\rangle_L \otimes |0\rangle_R$. This state is included in the basis but since there is one particle, its coefficient is always zero. The rest of the first column represents the states $|i\rangle_L \otimes |0\rangle_R$, and similarly the rest of the first row represents $|0\rangle_L \otimes |j\rangle_R$. The large lower right block of zeros represents two particle states.

Then the density matrix $\rho_{ii'}$ in matrix form is

$$\rho = \psi\psi^\dagger = \begin{pmatrix} w_R & 0 & 0 & \dots & 0 \\ 0 & \psi_1\psi_1 & \psi_1\psi_2 \dots & \psi_1\psi_\ell \\ 0 & \psi_2\psi_1 & \psi_2\psi_2 \dots & \psi_1\psi_\ell \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \psi_\ell\psi_1 & \psi_\ell\psi_2 \dots & \psi_\ell\psi_\ell \end{pmatrix} , \quad (\text{A.4})$$

where $w_R = 1 - w_L = 1 - \sum_{i=1}^\ell |\psi_i|^2$ is the probability that the particle is in the right block. This density matrix has two nonzero eigenvalues, with corresponding eigenvectors $(1 \ 0 \ \dots \ 0)^T$ and $w_L^{-1/2}(\psi_1 \dots \psi_\ell)^T$. The first eigenvector does not need to be explicitly treated; the second is equivalent to the projection of the wavefunction considered in Sect. 4.

B A Program in C++ for the Particle on a Chain

General

In order to illustrate the DMRG method more explicitly, we describe here the DMRG algorithm and a complete DMRG C++ program to obtain the ground state energy and wave function for the particle on-a-chain problem numerically. We hope that after writing or playing with a program based on this appendix, the main part of this chapter should be much easier to understand. The algorithm which we will describe here is a simplified version of the finite system algorithm described in Sect. 5.

Our goal is to solve the particle on a chain problem with Hamiltonian (5) numerically. To make the program more like an efficient DMRG program for an interacting system, in which the dimensions of the matrices are much larger, we assume that our computer has only enough memory to diagonalize a 4×4 matrix. We will assume that it has enough “disk” storage to hold about two vectors of length L . By this we mean that the innermost loops involve manipulations of only a small amount of data, and access to the main data of size $\sim L$ occurs in outer loops, a little at a time. We will assume we do not know the analytic solution to the problem, and the algorithm we present will work equally well if the nonzero elements of H are perturbed slightly, either randomly or not. The solution we present is the DMRG method, in simplified form only because it is applied to a single particle problem.

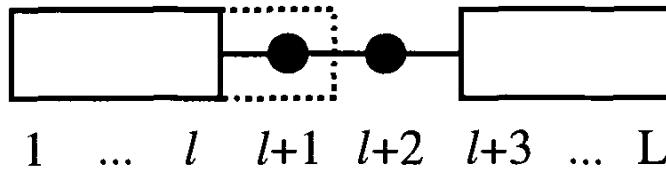


Fig. B.1. Breakup of a system, of length L , into two blocks and two sites.

Our procedure will divide the system up into four pieces, called blocks, as shown in Fig. B.1. (This is the “superblock” of Sect. 5.) Block 1 (the “left block”) will consist of sites $1 \dots \ell$; block 2 will consist of a single site, $\ell + 1$; block 3 will also consist of a single site, $\ell + 2$; and block 4 (the “right block”) will consist of sites $\ell + 3 \dots L$. During the course of the algorithm, the dividing point ℓ will be moved back and forth through the system, so that every site except the first and last is represented by one of the two middle sites during a “sweep” through the system.

Each block will be represented by a single basis state. For the middle blocks, this is not a limitation; for the left and right blocks, it is a severe one. A wavefunction ψ_j , $j = 1, \dots, L$, is written as

$$\psi_j = \begin{cases} a_1 L_j & j \leq \ell \\ a_2 & j = \ell + 1 \\ a_3 & j = \ell + 2 \\ a_4 R_j & j \geq \ell + 3 \end{cases} . \quad (\text{B.1})$$

The basis states L and R are defined only within the corresponding blocks, and they are normalized, $\langle L|L \rangle = \langle R|R \rangle = 1$. Note that any state can be represented in this way if L and R are chosen appropriately for that state. Using (6) and (B.1), we find that the Hamiltonian matrix element between two states ψ and ψ' can be written as

$$\langle \psi | H | \psi' \rangle = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}^T \begin{pmatrix} H_{11} & T_{12} & 0 & 0 \\ T_{12} & 2 & -1 & 0 \\ 0 & -1 & 2 & T_{34} \\ 0 & 0 & T_{34} & H_{44} \end{pmatrix} \begin{pmatrix} a'_1 \\ a'_2 \\ a'_3 \\ a'_4 \end{pmatrix} \quad (\text{B.2})$$

where

$$H_{11} = \langle L | H | L \rangle , \quad (\text{B.3})$$

$$H_{44} = \langle R | H | R \rangle , \quad (\text{B.4})$$

$$T_{12} = \langle L | H | l + 1 \rangle = -L_\ell , \quad (\text{B.5})$$

and

$$T_{34} = \langle l + 2 | H | R \rangle = -R_{\ell+3} . \quad (\text{B.6})$$

The 4×4 matrix H in (B.2) represents the Hamiltonian in our restricted basis. Its minimum eigenvalue is an upper bound for the exact ground state energy. When L and R represent the left and right parts of the exact ground state, then the minimum eigenvalue is the exact ground state energy.

In order to turn this into an algorithm for finding the ground state of the system, we will specify an iterative procedure for improving L and R . Let $L(\ell)$ and $R(\ell+3)$ represent the bases for step ℓ . Then we want to improve the set of all basis states, $\{L(\ell)\}$ and $\{R(\ell)\}$. Our basic step in this procedure will be to get an improved basis state $L(\ell+1)$, given $L(\ell)$ and $R(\ell+3)$. Given this improved basis state, we proceed through the lattice from left to right, improving the set $\{L(\ell)\}$. When we reach the right end, we will reverse the procedure, generating an improved $R(\ell+2)$, given $L(\ell)$ and $R(\ell+3)$, as we proceed through the lattice in the reverse direction, eventually improving all elements of the set $\{R(\ell)\}$.

The basic left-to-right step is this: First, we diagonalize our 4×4 Hamiltonian matrix, (B.2), obtaining a ground state wavefunction in this basis, a vector of length 4, which we write as (a_1, a_2, a_3, a_4) . We normalize a_1 and a_2

as $a'_1 = a_1/N$ and $a'_2 = a_2/N$, where $N = (a_1^2 + a_2^2)^{1/2}$. Then the new basis state, properly normalized, is

$$L(\ell+1)' = \begin{pmatrix} a'_1 L(\ell)_1 \\ \vdots \\ a'_1 L(\ell)_\ell \\ a'_2 \end{pmatrix}. \quad (\text{B.7})$$

The new Hamiltonian matrix element, needed to construct H for the next step, is

$$\langle L(\ell+1)' | H | L(\ell+1)' \rangle = {a'_1}^2 \langle L(\ell) | H | L(\ell) \rangle + 2{a'_2}^2 - 2a'_1 a'_2 L(\ell)_\ell. \quad (\text{B.8})$$

If $L(\ell)$ and $R(\ell+3)$ are the exact left and right pieces of the true ground state of the system, then $L(\ell+1)'$ obtained in this way is the exact left piece of the true ground state, when the system is divided at $\ell+1$ rather than ℓ . Thus if our procedure ever finds the correct ground state at some step ℓ , then after one additional sweep through the lattice, all the $\{L(\ell)\}$ and $\{R(\ell)\}$ will represent exact parts of the true ground state and the procedure will have converged. Furthermore, assume $|\psi(\ell)\rangle$ is the state at step ℓ . Then, with $L(\ell+1)'$ as specified in (B.7), in the new basis at step $\ell+1$, $|\psi(\ell)\rangle$ can be represented exactly. Therefore the energy at step $\ell+1$ cannot be higher than at step ℓ . However, the basis in step $\ell+1$ generally has additional degrees of freedom not in the basis at step ℓ , and consequently, the energy at step $\ell+1$, obtained from diagonalizing the 4×4 Hamiltonian matrix, will generally be lower than at step ℓ . Consequently, we expect a monotonic convergence to the true ground state with this procedure.

In order to initialize this iterative procedure, we need to generate a set of approximate initial blocks $\{L(\ell)\}$ and $\{R(\ell)\}$. A good initial approximation comes from applying our procedure to a set of systems of sizes $4, 6, \dots, L$. Starting with the system of size 4, the left and right blocks are just single sites, with trivial L and R . We use this system to get an approximate $L(2)$. Proceeding to the system of size 6, this $L(2)$ can be used, and for the block R we use the reflection of $L(2)$, switching the left and right sides of the basis state. At each additional step, a reflection of the left block is used for the right block. In this way, we gradually grow the system up to the full size L , generating an initial set of $\{L(\ell)\}$ and $\{R(\ell)\}$.

Program

We now present a complete C++ program to solve the particle on a chain problem using the steps outlined above. The object-oriented nature of C++ means that the main routine, for a particle on a chain, can be reused with rather little change for a non-trivial interacting problem. The program utilizes the “MatrixRef” matrix library, written by the authors. Both the MatrixRef matrix library and a complete version of the program described here are

currently available on the world-wide web at <http://hedrock.ps.uci.edu>. In what follows, we will list and discuss the important parts of the program.

Blocks: We first define a class for a **Block**.

```
class Block
{
public:
    Real H11, L_inner;
    Block()          // Default: construct a one-site block
                    : H11(2.0), L_inner(1.0) { }
    Block Reflect() const { return *this; }
};
```

Here type (**Real** is defined in the matrix library as a **double**, equivalent to a Fortran **real*8**). The two variables **H11** and **L_inner** store the values of either $\langle L|H|L \rangle$ or $\langle R|H|R \rangle$, and either L_ℓ or $R_{\ell+3}$, depending on whether the block is a left or right block. The constructor function **Block()** initializes the block in the way appropriate for a single site block, with $H_{11} = 2$ and $L_\ell = 1$. The **Reflect()** function, which reflects a block, does nothing in this simple case because of the dual usage of **L_inner**, but is included to illustrate the more general case.

System: Next we define a class **System**, which contains the four **Blocks** making up the lattice, for a particular step ℓ .

```
class WaveFunction;

class System
{
public:
    const Block &b1, &b2, &b3, &b4;
    System(const Block& bb1,const Block&bb2,
           const Block&bb3, const Block&bb4)
        : b1(bb1), b2(bb2), b3(bb3), b4(bb4) { }

    Real GetGroundState(WaveFunction& p);
};
```

A **System** has four data members, references to four blocks, **b1**, **b2**, **b3**, and **b4**. The function **GetGroundState**, returns a **Real** number, the ground state energy, and finds a ground state wavefunction. The definition of **WaveFunction** is

```
class WaveFunction
{
public:
```

```
Vector v;
WaveFunction() : v(4) {}
};
```

Wave function: A `WaveFunction` here is simply a `Vector` of length 4. Vectors are defined in the matrix library. The function `GetGroundState` is

```
Real System::GetGroundState(WaveFunction& p)
{
    Matrix H(4,4), evecs(4,4);
    Vector evals(4);
    H = 0.0;
    H(1,1) = b1.H11;    H(2,2) = b2.H11;
    H(3,3) = b3.H11;    H(4,4) = b4.H11;
    H(1,2) = H(2,1) = -b1.L_inner;
    H(2,3) = H(3,2) = -1.0;
    H(3,4) = H(4,3) = -b4.L_inner;
    EigenValues(H,evals,evecs);
    p.v = evecs.Column(1);
    if(p.v.sumels() < 0.0) p.v *= -1.0;
    Real energy = evals(1);
    return energy;
}
```

This function defines a 4×4 matrix `H`, a matrix of eigenvectors `evecs`, and a vector of eigenvalues `evals`. `H` is first initialized to contain all 0's, and then the non-zero elements are defined according to (B.2), (B.5), and (B.6). Then, the Matrix library routine `EigenValues` is called, which diagonalizes `H`. The `WaveFunction` `p` is set to the ground state vector, and the ground state energy is returned.

Density matrix: The last class we will define is a `DensityMatrix`. Thus far, in our description of the single particle DMRG algorithm, density matrices have not appeared. However, as shown in Appendix A, the elements a_1 and a_2 of the wavefunction ψ are equivalent to the reduced density matrix for the first two blocks as part of the system as a whole.

```
enum LR {Left, Right};

class DensityMatrix
{
public:
    Real a,b;
    DensityMatrix(const WaveFunction& psi,LR lr)
    {
        if(lr == Left)
            { a = psi.v(1); b = psi.v(2); }
```

```

    else
        { a = psi.v(4); b = psi.v(3); }
    }

Vector NewBasis()
{
    Vector res(2);
    Real norm = sqrt(a*a+b*b);
    res(1) = a / norm;
    res(2) = b / norm;
    return res;
}

};

```

A `DensityMatrix` has two data members, `a` and `b`, representing a_1 and a_2 or a_4 and a_3 , depending on the whether one is forming a density matrix for the left or right half of the system. The first line defines a new type, `LR`, which takes on only two symbolic values, `Left` and `Right`, indicating which half of the system we want. One makes a `DensityMatrix` out of a `WaveFunction` `psi` simply by assigning the appropriate values to `a` and `b`. The function `NewBasis` forms a new basis for $L(\ell + 1)$, which simply involves normalizing the two-element vector (a,b) .

In order to utilize the new basis to form a new block, there are functions `NewLeft` and `NewRight`, which combine two blocks (either the left two blocks [`NewLeft`], or the right two blocks [`NewRight`]) using a new basis, and using (B.8). Here we show `NewLeft`; `NewRight` is quite similar.

```

Block NewLeft(const Block& b1, const Block& b2, const Vector& bas)
{
    Block res;
    res.H11 = bas(1) * bas(1) * b1.H11 + 2 * bas(2) * bas(2)
              - 2 * bas(1) * bas(2) * b1.L_inner;
    res.L_inner = bas(2);
    return res;
}

```

Main program: The main program is relatively straightforward, using these classes and routines. First, we read in the length of the system and the number of sweeps to be performed, and initialize an array of blocks which will hold both $\{L(\ell)\}$ and $\{R(\ell)\}$.

```

int main()
{
    Block siteblock;
    cerr << "Input length, number of iterations: ";
    int i, length, nsweeps;
    cin >> length >> nsweeps;
    cout << "length, nsweeps = " << length SP nsweeps << endl;
}

```

```

cout << "Exact energy = " << exacten(length) << endl;
exlen = length;
Array1<Block> allblocks(length);
WaveFunction psi;
Real energy;

```

Next, we perform the warmup sweep, which uses the infinite-size algorithm to build up the system from four sites initially to the full system size.

```

// Warmup sweep, using the infinite-size algorithm
allblocks[1] = siteblock;
for(i = 1; i < length/2; i++)
{
    Block rightblock = allblocks(i).Reflect();
    System S(allblocks(i),siteblock,siteblock,rightblock);
    energy = S.GetGroundState(psi);
    cout << i+1 SP psi.v(2) SP energy SP 0 << endl;
    DensityMatrix rho(psi,Left);
    Vector basis = rho.NewBasis();
    allblocks[i+1] = NewLeft(allblocks(i),siteblock,basis);
}

```

The macro `SP` in the `cout` statement makes a space, and is defined in the matrix library. The final 0 to be printed out on each line indicates this is the warmup sweep. Once the system has reached its full size, `nsweeps` sweeps are performed to converge to the ground state. Each sweep consists of a right-to-left part, and a left-to-right part. In this program, because the system is symmetric under reflections, the sweeps start or end at the middle, symmetric configuration, rather than the far right site. The right-hand blocks in the middle configuration are obtained as reflections of the left hand blocks.

```

// Finite System sweeps
for(int swp = 1; swp <= nsweeps; swp++)
{
// We assume reflection symmetry:
    allblocks[length/2 + 2] = allblocks(length/2 - 1).Reflect();
    cout << endl;

// Right to left
    for(i = length/2+2; i > 3; i--)
    {
        System S(allblocks(i-3),siteblock,siteblock,allblocks(i));
        energy = S.GetGroundState(psi);
        cout << i-1 SP psi.v(3) SP energy SP swp - 0.5 << endl;
        DensityMatrix rho(psi,Right);
        Vector basis = rho.NewBasis();
        allblocks[i-1] = NewRight(siteblock,allblocks(i),basis);
    }
}

```

```

// Left to right
cout << endl << 1 SP psi.v(1) SP energy SP swp << endl;
for(i = 1; i < length/2-1; i++)
{
    System S(allblocks(i),siteblock,siteblock,allblocks(i+3));
    energy = S.GetGroundState(psi);
    cout << i+1 SP psi.v(2) SP energy SP swp << endl;
    DensityMatrix rho(psi,Left);
    Vector basis = rho.NewBasis();
    allblocks[i+1] = NewLeft(allblocks(i),siteblock,basis);
}
}
return 0;
}

```

At each step i , the energy and value of the wavefunction at site $i + 1$ is printed out. (We use i , rather than l , as the step variable because ℓ looks too much like 1 in program listings.)

Results

Results from this program are shown in Fig. B.2. The ground state energy is plotted as a function of the index of the site ℓ just added, for the warmup and first three sweeps. Note that during the warmup, the system length is

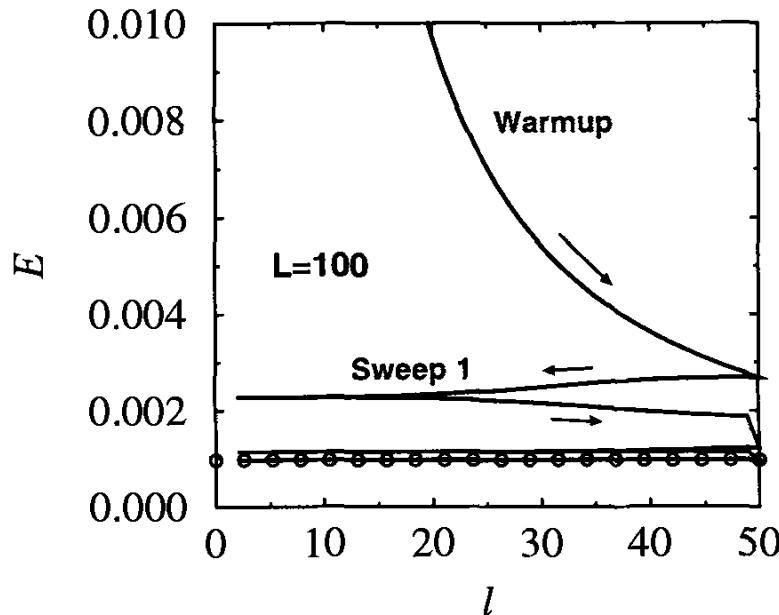


Fig. B.2. Results for the ground state energy for a particle on a chain of length $L = 100$.

less than the eventual system size, which accounts for most of difference in energy from the exact result. (Here $\ell = i + 1$ for left to right sweeps and $\ell = i - 1$ for right to left sweeps.) During the warmup sweep, the system size

is still growing, and at each step the length is 2ℓ . Starting with the beginning of the first sweep, the length is fixed at L . Each sweep starts in the middle, progresses to the left side, and returns. There is a discontinuous jump in the energy between sweeps because at those points a reflection of the left block is used to replace the first right block of the new sweep. During the warmup, this is done every step, but it occurs only once per sweep during the other sweeps. The open circles represent the exact energy, $E = 0.000967435$. (The general formula is $E = 2[1 - \cos(\frac{\pi}{L+1})]$.) Good convergence is obtained after three sweeps.

Results for the wavefunction are shown in Fig. B.3. The wavefunction at the site just added is plotted versus the site index for the first three sweeps. Each half-sweep is labeled by an integer, starting at 1. The open circles represent the exact wavefunction, which is given (in unnormalized form) by

$$\psi_j = \sin(qj) \quad (\text{B.9})$$

with $q = \pi/(L + 1)$, and $j = 1, \dots, L$. Convergence of the wavefunction is reasonable after three sweeps; another sweep or two would be needed for precise agreement with the exact results.

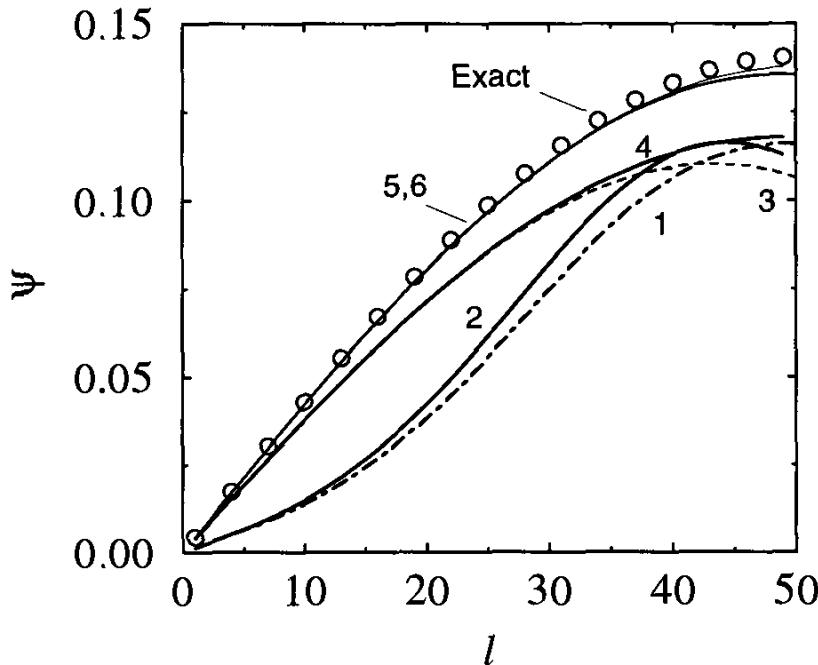


Fig. B.3. Results for the wavefunction for a particle on a chain of length $L = 100$.

The above program can be easily modified to find more than just one state, to deal with modifications to the Hamiltonian, including randomness, etc. A useful exercise would be to modify this program to include randomness or another local potential, to include longer range hopping, etc. Another useful exercise would be to translate this program into another programming language.

References

1. S.R. White, Phys. Rev. Lett. **69**, 2863 (1992), Phys. Rev. B **48**, 10345 (1993)
2. K.G. Wilson, Rev. Mod. Phys. **47**, 773 (1975)
3. K. Hallberg, Phys. Rev. B **52**, 9827 (1995); H. Pang, H. Akhlaghpour and M. Jarrell, Phys. Rev. B **53**, 5086 (1996)
4. T. Nishino, J. Phys. Soc. Jpn. **64**, 3598 (1995)
5. S. Moukouri and L.G. Caron, Phys. Rev. Lett. **77**, 4640 (1996)
6. R.J. Bursill, T. Xiang and G.A. Gehring, J. Phys. Cond. Matt. **8**, 583 (1996)
7. X.Q. Wang and T. Xiang, Phys. Rev. B **56**, 5061 (1997)
8. S.R. White and R.L. Martin, unpublished, cond-mat/9808118
9. E.R. Davidson, J. Comp. Phys. **17**, 87 (1975); E.R. Davidson, Computers in Physics **7**, No. 5, 519 (1993)
10. J.W. Bray and S.T. Chui, Phys. Rev. B **19**, 4876 (1979); T. Xiang and G.A. Gehring, Phys. Rev. B **48**, 303 (1993)
11. S.R. White and R.M. Noack, Phys. Rev. Lett. **68** 3487, (1992)
12. R.P. Feynman, Statistical Mechanics: A Set of Lectures, Benjamin (1972)
13. W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, Numerical Recipes: The Art of Scientific Computing, Cambridge (1986)
14. E. Jeckelmann and S.R. White, Phys. Rev. B **57**, 6376 (1998)
15. E.H. Lieb and F.Y. Wu, Phys. Rev. Lett. **20**, 1445 (1968)
16. H. Schulz, J. Phys. C: Solid State Phys. **18**, 581 (1985)
17. S. Kneer, Diploma Thesis, Universität Würzburg (1997)
18. G. Bedürftig, B. Brendel, H. Frahm and R.M. Noack, Phys. Rev. B **58**, 10225 (1998)
19. X.Q. Wang and S. Mallwitz, Phys. Rev. B **48**, R492 (1996)
20. S.J. Qin, X.Q. Wang, Lu Yu, Phys. Rev. B **56**, R14251 (1997)
21. H.J. Schulz, Int. J. Mod. Phys. B **5**, 57 (1991)
22. S.R. White, Phys. Rev. Lett. **77**, 3633 (1996)
23. S. Östlund and S. Rommer, Phys. Rev. Lett. **75**, 3537 (1995)
24. R.M. Noack, S.R. White and D.J. Scalapino, in Computer Simulations in Condensed Matter Physics VII, Eds. D.P. Landau, K.K. Mon and H.B. Schüttler Spinger, p 85 (1994)
25. S. Liang, Phys. Rev. B **49**, 9214 (1994)

3 Thermodynamic Limit and Matrix-Product States

Stefan Rommer^{**} and Stellan Östlund

Institute for Theoretical Physics, Chalmers University of Technology,
SE - 412 96 Göteborg, Sweden

In this lecture we study the nature and underlying principles of the density matrix renormalization group (DMRG) method. Our primary goal is not to reach best possible accuracy by doing full scale DMRG calculations. Instead we will keep a very modest number of states in the DMRG truncation and analytically analyze the renormalization process. Many of the results that will be reviewed in this lecture were originally presented in [1,2]. For background information on the DMRG method see Chap. 2 (I) and the original articles by White [3,4].

In Sect. 1 of this lecture we show that if the DMRG algorithm converges to a fixed point in the thermodynamic limit, the DMRG ground state leads to a special form for the wave function. Guided by this form we define a set of ansatz wave functions in matrix product (MP) form. To make things more concrete we apply our ideas to the antiferromagnetic Heisenberg spin-1 chain with bilinear and biquadratic interactions, defined by

$$H = \sum_{i=1}^n \mathbf{S}_i \cdot \mathbf{S}_{i+1} - \beta (\mathbf{S}_i \cdot \mathbf{S}_{i+1})^2 . \quad (1)$$

Section 2 contains a brief discussion of some aspects of this model that are of relevance for later sections.

In Sects. 3 and 4 we analyze the matrix product ansatz states and find that the DMRG ground state can actually be rederived by doing a variational calculation in the matrix product ansatz states. This demonstrates a relation of the DMRG to a variational calculation. In Sect. 4 we calculate several properties of the matrix product states. In particular we determine correlation functions in the MP states and find that the states are intrinsically either long range ordered or finitely correlated.

In Sect. 5 we extend the ansatz to include a set of Bloch states that describe elementary excitations in both finite and infinite systems. We show numerical results for the low-lying excitation spectrum of the spin-1 chain. The last section, Sect. 6, contains a summary. This lecture also contains two

^{**} Present address: Department of Physics and Astronomy, University of California, Irvine, CA 92697-4575

short appendices. Appendix A deals with parity of the MP wavefunction and App. B contains details on how to calculate an orthonormal basis in the MP states. These appendices give details on how calculations in the MP states can be done.

matrix-product states have been extensively used in both classical and quantum mechanics and it is out of the scope of this lecture to give a detailed account on this issue. Before we discuss their connection to the DMRG we will however briefly discuss some other uses of MP states for quantum systems that are closely related to this lecture.

When it comes to increasing numerical accuracy, the variational calculations that will be discussed in this lecture are probably not a serious alternative to big DMRG calculations. However, the matrix product states give us some insight into how the DMRG works. We can also better understand the limitations of the DMRG method. In this lecture we will see an example of this when we look at correlation functions in the MP states.

The matrix-product states are also interesting in their own right. For example they can be used as a convenient representation of valence bond ground states for spin-1 chains [5,6], like the ground state of the Affleck-Kennedy-Lieb-Tasaki (AKLT) model. The AKLT model, i.e. the Hamiltonian in (1) with $\beta = -1/3$, is one simple, but non-trivial, example of a model with MP ground state. This ground state will be introduced in Sect. 2.

MP states have also been extensively used to construct new classes of spin Hamiltonians with exact MP ground states [6–9]. The task is here in a sense opposite to the usual when solving a Hamiltonian; construct MP states and then try to find the Hamiltonians for which these states are the exact ground states. MP states have also been used as approximate trial states for spin chains [10,11] and spin ladder systems [12,13]. Even though the numerical results of, for example, the ground-state energy of the MP ansatz for the pure Heisenberg chain cannot compete in accuracy with a full scale DMRG calculation, the MP ansätze give us a picture of how the ground states of non-solvable spin chains might look like. As an example, the AKLT state is believed to be a good caricature of the true ground state of the ordinary spin-1 Heisenberg model ((1) with $\beta = 0$).

Klümper *et al.* [7] as well as Kolezhuk *et al.* [8,11] use a somewhat different MP formalism from the one that will be presented in this lecture. It is however not difficult to see that their MP states can be written in the form that we will use.

1 Thermodynamic Limit and the Matrix-Product State

In a renormalization scheme like the DMRG one typically starts with a very short 1D chain whose properties can be calculated exactly. The scheme then proceeds by iteratively adding a single site. When the chain gets longer we do

not use the full set of basis states for describing the system. Instead a reduced number of basis states is used to represent a system in each renormalization step. In this way we keep the Hilbert space at a manageable size as the system size grows.

In the literature, the DMRG algorithm is usually described by how the Hamiltonian and other operators of a system are constructed as the chain grows by one site and how these operators are calculated as the basis truncation is done. The basis vectors themselves, expressed in the original complete basis, are however not explicitly constructed. This is also not necessary and would not be numerically feasible. After all, it is the large dimension of the full Hilbert space we want to avoid. In this section we will however analyze a renormalization step and look at how the basis states change.

Assume we have a block that represents a chain with $n - 1$ sites. Let m_s be the number of possible states of a single lattice site. If we would treat this system exactly there would be m_s^{n-1} states in the Hilbert space basis for this system. In the case of a spin-1 chain, we could label the site with the z -component of the single spin-1, so that $m_s = 3$. The number of states in this complete basis rapidly becomes too large to handle when n is increased. Assume therefore that an approximation is made and our chain is represented by a smaller set of states labeled by $\{|\beta\rangle_{n-1}\}$. This set of states has been chosen by the previous iterations of the renormalization with the aim to describe the low energy physics. Assume there are m states in this basis, where $m \leq m_s^{n-1}$. If this is the first iteration, $\{|\beta\rangle_{n-1}\}$ is the complete basis.

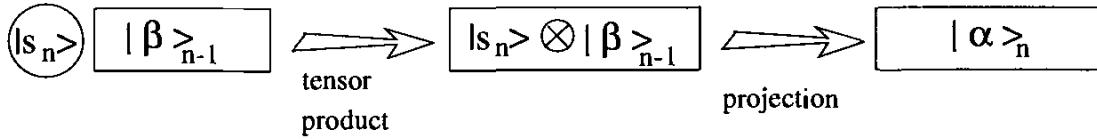


Fig. 1. A renormalization step. First the old block representing $n - 1$ sites is joined to a single site labeled by s_n . The tensor product then “glues” the different parts together. Finally there is a projection to a new block representing n sites.

We now add a single site, labeled by s_n , the z -component of spin, to the left hand side of our block resulting in a new block with n sites and $m_s * m$ states in its basis. The basis states are now generated by the product representation $\{|s_n\rangle \otimes |\beta\rangle_{n-1}\}$. See Fig. 1 for an illustration of this step. We now use a projection operator A_n to generate a new truncated basis with typically m states that represent the “important” states of the longer block. This whole process is written

$$|\alpha\rangle_n = \sum_{\beta, s_n} A_n^{\alpha, (\beta, s_n)} |s_n\rangle \otimes |\beta\rangle_{n-1}, \quad (2)$$

where we have indexed A by the chain length n and its matrix indices α and (β, s_n) . Note that (β, s_n) is thought of as a single index labeling a tensor product of the states $|s_n\rangle$ and $|\beta\rangle_{n-1}$. In the DMRG, a specific algorithm is used to calculate A , but this is not important in the present discussion.

We now make two crucial observations.

1. First we perform a simple change in notation: $A_n^{\alpha,\beta}[s_n] \equiv A_n^{\alpha,(\beta,s_n)}$, thus writing the $m \times (m_s * m)$ matrix as a set of m_s matrices with dimension $m \times m$.
2. Second, we assume that the recursion leads to a fixed point for the projection operator so that we can write $A_n[s] \rightarrow A[s]$, as $n \rightarrow \infty$.

By repeatedly applying the renormalization step in (2) we now find that

$$|\alpha\rangle_n = \sum_{s_n \dots s_1, \beta} (A[s_n]A[s_{n-1}] \dots A[s_1])^{\alpha, \beta} |s_n s_{n-1} \dots s_1\rangle \otimes |\beta\rangle_0 , \quad (3)$$

where $|\beta\rangle_0$ represents some initial state. We thus see that the renormalization procedure results in a wave function that can be written in a matrix product form.

Equation (3) now suggests a natural form for trial wavefunctions. We can in a compact way write linear combinations of the m^2 states

$$|\alpha, \beta\rangle_n = \sum_{s_n \dots s_1} (A[s_n]A[s_{n-1}] \dots A[s_1])^{\alpha, \beta} |s_n s_{n-1} \dots s_1\rangle$$

by introducing a matrix Q . For every $m \times m$ matrix Q we define the (unnormalized) state $|Q\rangle_n$

$$|Q\rangle_n \equiv \sum_{\{s\}} \text{tr}(Q A[s_n] \dots A[s_1]) |s_n \dots s_1\rangle . \quad (4)$$

The state $|Q\rangle_n$ can be viewed as a state that is uniform in the bulk, but with a linear combination of boundary conditions defined by $|\alpha\rangle_n$ on the left and $|\beta\rangle_0$ on the right. The special case of $Q = \mathbb{1}$, the identity matrix, leads to a translationally invariant state with periodic boundary conditions. This $Q = \mathbb{1}$ state we will later on use as our trial ground state.

If we now demand that the projection of (2) preserves orthonormal bases, $\langle \alpha | \alpha' \rangle = \delta_{\alpha, \alpha'}$, we can use the recursion formula (2) and the orthogonality of the local spin states and previous block states to find

$$\begin{aligned} \delta_{\alpha, \alpha'} &= \sum_{\beta, \beta', s, s'} (A^{\alpha', \beta'}[s'])^* A^{\alpha, \beta}[s] \langle s' | s \rangle \langle \beta' | \beta \rangle \\ &= \sum_s (A[s] A^\dagger[s])^{\alpha, \alpha'} . \end{aligned} \quad (5)$$

Hence in matrix form we have $\sum_s A[s] A^\dagger[s] = \mathbb{1}$. This constraint will be used later to reduce the number of free parameters in A .

2 The Model

We will now take a closer look at the spin model defined in (1) and review some of its properties that will be of interest in later sections.

The behavior of the model in (1) as a function of β has been extensively studied [14–22]. Depending on the ratio between the bilinear and the biquadratic terms in the Hamiltonian, this model is believed to exhibit several different phases. See e.g. [20] for a review of the phase diagram.

In this lecture we apply our methods to (1) with $\beta = -1/3$ and $\beta = 0$. It is believed that the model exhibits a single phase in the range $-1 < \beta < 1$, usually called the Haldane phase. At $\beta = 0$ we find the ordinary Heisenberg model, which Haldane conjectured [23] to have exponential decaying spin-spin correlations and a finite energy gap to the first excited state. This conjecture has been very well supported by numerical calculations [14, 24, 25]. At the point $\beta = 1$ the model is exactly solvable using the Bethe ansatz. The spectrum was calculated by Takhtajan [26] and Babujian [27] and is gapless with power-law decay of correlations. From field theory arguments [28] it is suggested that the gap opens up on both sides of the point $\beta = 1$. The numerical evidence for the opening of the gap is, however, inconclusive [19, 20]. At $\beta = -1$ we find the Lai-Sutherland model which is again solvable by Bethe ansatz techniques [29–31]. The model is gapless with algebraically decaying correlations.

The model with $\beta = -1/3$ is the AKLT Hamiltonian with a unique ground state that can be constructed as an exact nearest-neighbor valence bond state (VBS), as was shown by Affleck *et al.* [32]. The AKLT Hamiltonian has exponential decay of correlations and a finite energy gap.

The VBS state can easily be constructed in a pictorial way [32]. Imagine that instead of a spin-1 object on each site of a chain there are two spin-1/2 objects. These two spin-1/2 can be represented in the s_z -basis by the z -component of each of the spins. The two spins can also be represented in the total spin basis by their total spin and the z -component of total spin. The transformation between these two bases is given by the Clebsch-Gordan coefficients [33]. In the total spin basis the two spin-1/2 form a singlet (spin-0) or a triplet (spin-1). Now let the two spin-1/2 be combined into a triplet and let this triplet represent a single spin-1 site in a spin-1 chain. To construct the VBS state, all we now have to do is to combine neighboring spin-1/2 on *different* sites into singlets (see Fig. 2). As we will see later, this state can

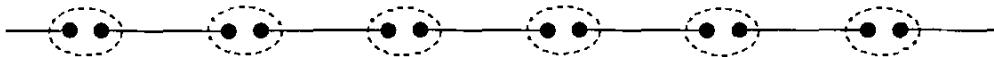


Fig. 2. The exact ground state of the AKLT Hamiltonian. The dots represent spin-1/2 variables. The dashed circles represent triplet bonds between the two spins on the same site and will thus be spin-1 objects. The solid lines represent singlet bonds between spin-1/2 objects on different sites.

also simply be written as a matrix-product state.

The valence-bond state is the exact ground state only for $\beta = -1/3$ but it is considered to give a reasonable picture of the ground state also for other values of β in this phase. For $-1 < \beta < 1$ our model is thus believed to have a gap, a unique ground state and exponentially decaying correlations.

From the valence-bond caricature of the ground state in the Haldane phase, there should be spin-1/2 degrees of freedom at each end of an open spin-1 chain. That these spin-1/2 edge states exist in the $\beta = -1/3$ model is clear from the construction of the exact ground state (see Fig. 2). For the ordinary Heisenberg model with $\beta = 0$, numerical calculations support the spin-1/2 edge states and thus the valence-bond picture [34].

Let us now look at the spin-1 chain from a total spin basis point of view. We will, in a DMRG fashion, start with a single site and then add a single spin-1 site at a time to the chain. Motivated by the spin-1/2 edge excitations of the open spin-1 chain, we will take a single spin-1/2 site as our first site. After adding the first spin-1 to our system, it thus consists of one spin-1/2 and one spin-1 site. In the total spin basis, the basis states of this system is a sum of two irreducible representations of total spin, one spin-1/2 and one spin-3/2,

$$\frac{1}{2} \otimes 1 = \frac{1}{2} \oplus \frac{3}{2}. \quad (6)$$

Note that since we have a spin-1/2 on the edge, the irreducible representations of total spin of our chain will be half-odd integers.

After adding another spin-1 site to the chain we find that the irreducible representations are

$$\left(\frac{1}{2} \oplus \frac{3}{2} \right) \otimes 1 = \frac{1}{2} \oplus \frac{3}{2} \oplus \frac{1}{2} \oplus \frac{3}{2} \oplus \frac{5}{2}. \quad (7)$$

In this way we can continue to add sites to our chain. We note that, as we add more spins to the chain, we do not only get higher values of total spin among our basis states, we also get multiple copies of irreducible representations of the same total spin. See Fig. 3 for an illustration of this process.

If we now, in a DMRG fashion, want to represent a finite chain by a reduced number of basis states, which basis states should we choose? The Hamiltonian of (1) is spin rotationally invariant since it commutes with all three components of the total spin $\mathbf{S}_{tot} = \sum_i \mathbf{S}_i$. In order to preserve this symmetry, we should pick complete representations of total spin to build up our basis. By this we mean that we want the basis to consist of complete multiplets of total spin, i.e. if one of the basis states for example has total spin $j = 3/2$ and z -component of total spin $m_z = 1/2$, then we also must have the states with $m_z = -3/2$, $m_z = -1/2$ and $m_z = 3/2$ of the same multiplet as basis states. Since the ground state of the spin-1 chain (without the spin-1/2 edge spin) is a singlet of total spin [35], we could guess that to describe this state the best representations to keep in each step would

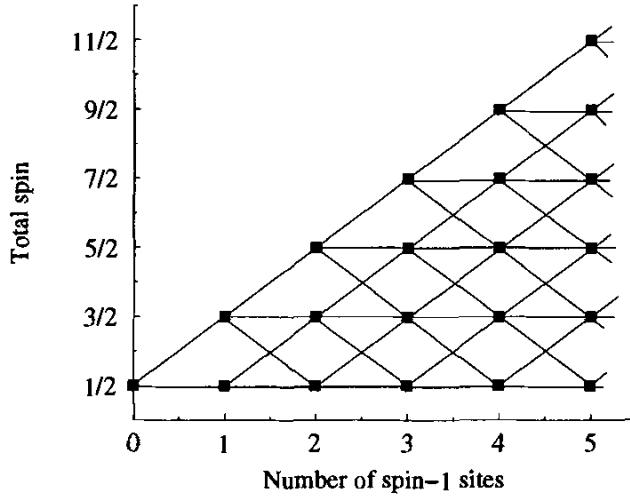


Fig. 3. Illustrates how the total spin representations of the spin-1 chain with a spin-1/2 edge spin evolve as we add sites to the system. Each square denotes one or more copies of an irreducible representation of total spin.

be those with small values of total spin. In the next section we will see an example of this.

3 Construction of the Matrices A

We now analyze the projection matrix A . We want our effective Hamiltonian, after the truncation of the basis, to have as many symmetries preserved as possible from the original Hamiltonian. As in the previous section, in order that the projection in each step preserves the spin rotational symmetry, our basis states of a block must form a representation of total spin.

Since we keep basis states with many different values of total spin as well as many states with the same total spin in each iteration, all the basis states together must form a sum of irreducible representations of total spin. Adding a spin one does not mix even or half-odd spin representations, thus the basis states must form a sum of either all half-odd or all integer spin representations. Most naturally for the spin-1 chain one would work with integer spin representations, but as we saw in the previous section, by placing a single spin-1/2 on the side of the entire chain one could use half integer spin representations to represent the blocks instead. This is consistent with the existence of a spin-1/2 edge state [34,32] and we have found that working with half-odd integer representations give better numerical results.

We now discuss how the projection operator A can be constructed for the case were we have kept 6 basis states in each iteration. A more complicated example with 12 basis states can be found in [2]. We have used the half-odd spin representations.

By doing a DMRG calculation on the spin-1 chain we have found that when approximately 6 states are kept the blocks are represented by one spin-

$1/2$ and one spin- $3/2$ irreducible representation. Since we in this case do not have multiple copies of representations with the same total spin, we can uniquely label each basis state by its total spin j and the z -component of total spin, m . The six “old” basis states that are the result of the previous iteration are thus denoted by $|j, m\rangle$, with $j = 1/2, m = -1/2, +1/2$ and $j = 3/2, m = -3/2, -1/2, +1/2, +3/2$

After adding a site and then truncating the Hilbert space we get the “new” basis states. Since we assume that we have reached the thermodynamic limit fixed point, the new basis states should fall into the same irreducible representations as the old basis states. We label the six basis states that represent the new block with one more site by $|j', m'\rangle$, with $j' = 1/2, m' = -1/2, +1/2$ and $j' = 3/2, m' = -3/2, -1/2, +1/2, +3/2$.

Let us now examine what happens in our example when going from the old $|j, m\rangle$ to the new $|j', m'\rangle$. When adding a single spin-1 to the old block of six states we get $6 \times 3 = 18$ “intermediate” states in the product representation of the old block states with a spin 1. This procedure is the same as is described by (7) and these 18 states thus fall into 5 irreducible representations. Since we in the intermediate states have more than one representation with a given total spin we cannot just label them with the total spin. If we however use the “old” total spin j as an additional label we can uniquely label the intermediate states by $|j, j'', m''\rangle$, where j'' is the total spin of the intermediate state and m'' is the corresponding z -component. These five intermediate reps are shown in the j'' -column in Fig. 4. We then project from these 5 reps back down to the two reps that we have chosen to keep. This projection must preserve the total spin symmetry, i.e. it cannot mix different j'' and it cannot depend on total m'' . We thus get only a few nonzero projection terms. Since there is exactly one intermediate spin- $1/2$ and one intermediate spin- $3/2$ for each of the two “old” representations j there is one projection term from each of the “old” j to each of the “new” j' . There are thus four nonzero projection terms which we denote by $P^{j', j}$. The $P^{j', j}$ are indicated by lines in Fig. 4. These projection terms are in fact not independent, but are related by the requirement that the new states are normalized.

Expressing all this mathematically, we let, as above, j uniquely label a rep of total spin of the old block. Each state is thus labeled by $|j, m\rangle$ where m is the z -component of total spin. The single spin to be added is labeled by $|s\rangle$, where s is the z -component of the spin-1.

The 18 intermediate states can be explicitly constructed in the total spin basis using the Clebsch-Gordan coefficients of the form $\langle(j_1, m_1)(j_2, m_2)|j, m\rangle$ as

$$|j, j'', m''\rangle = \sum_{m,s} \langle(j, m)(1, s)|j'', m''\rangle (|s\rangle \otimes |j, m\rangle). \quad (8)$$

From these 18 states $|j, j'', m''\rangle$ we now do a projection to 6 “new” basis states $|j', m'\rangle$, i.e. we define 6 new states by taking linear combinations of the 18 intermediate states. As above we demand that the projection preserves total

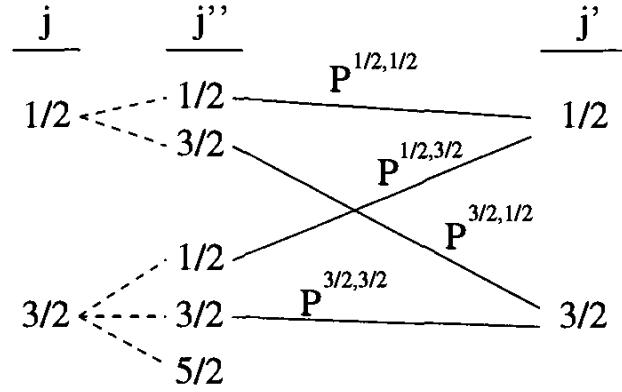


Fig. 4. Construction of the block states is shown when six states are kept in the basis. Old representations are on the left and new representations on the right. Intermediate representations are denoted by j'' . The z -component of total spin is not explicitly written out. Each solid line represents a nonzero projection $P^{j'',j}$ of basis representations.

spin and the z -component of total spin. The projection (the linear combinations) that define our new states will thus be very restrictive and we will have $j'' = j'$ and $m'' = m'$ in the projection. The projection is thus determined by a few projection parameters (coefficients) $P^{j'',j}$ and is written

$$|j', m'\rangle = \sum_j P^{j',j} |j, j', m'\rangle . \quad (9)$$

Note that $j'' = j'$ and $m'' = m'$ so that $|j, j', m'\rangle$ in (9) denotes the 18 intermediate states. Inserting (8) into (9) we find that (C.f. (2))

$$|j', m'\rangle = \sum_{s,(j,m)} A^{(j',m'),(j,m)} [s] (|s\rangle \otimes |j, m\rangle) ,$$

where

$$A^{(j',m'),(j,m)} [s] = P^{j',j} \langle (j, m) (1, s) | j', m' \rangle . \quad (10)$$

Thus, although the projection matrices A contain a total of $3 \times 6 \times 6 = 108$ numbers, they are in fact generated by the relatively few degrees of freedom available in $P^{j',j}$.

For this case with six basis states there are naively 4 parameters $P^{j',j}$ and we could write them in a 2×2 matrix \mathbf{P}

$$\mathbf{P} \equiv \begin{pmatrix} P^{\frac{1}{2},\frac{1}{2}} & P^{\frac{1}{2},\frac{3}{2}} \\ P^{\frac{3}{2},\frac{1}{2}} & P^{\frac{3}{2},\frac{3}{2}} \end{pmatrix} . \quad (11)$$

If we demand normalization of all basis states, $\langle j_1, m_1 | j_2, m_2 \rangle = \delta_{j_1, j_2} \delta_{m_1, m_2}$, a simple calculation using (5) and (10) yields the condition that the diagonal elements of $\mathbf{P}^T \mathbf{P}$ are all 1, where the superscript T denotes transpose. This gives two constraints. Since the spin-1/2 basis states are automatically

orthogonal to the spin-3/2 states, orthogonality of all basis states gives us no new constraints. We thus end up with only two free parameters. Under the given constraints we can choose to parameterize \mathbf{P} as

$$\mathbf{P} = \begin{pmatrix} \cos(\phi) \cos(\theta) \\ \sin(\phi) \sin(\theta) \end{pmatrix}, \quad (12)$$

where ϕ and θ represent the only free parameters in the projection matrices $A[s]$ in the case $m = 6$.

Similar but slightly more involved arguments are used in [2] for 12 basis states. In that case it is found that out of the $3 \times 12 \times 12 = 432$ numbers in A , there are only eight free parameters.

Recently, Dukelsky *et al.* have done a related study but with a slightly different choice of independent parameters [36]. They considered matrix-product states with up to six irreducible representations in the truncated basis.

With only a few free parameters we can use a variational principle for the energy to determine these. At this point it is clear that the DMRG plays no essential role in the construction aside from providing a guide to which representations to keep. Even this choice could be done variationally. In the next section we will make an ansatz for a ground state wavefunction and perform the variational calculation to find the ground-state energy and the projection matrices $A[s]$. For $m = 6$ this amounts to finding the parameters ϕ and θ of (12) that minimize the energy of our ground-state ansatz.

The Hamiltonian in (1) with $\beta = -1/3$ is a very special point in this context. As can be seen from the construction of the ground state of the AKLT-model (Fig. 2) the total spin of the ground state does not increase as we add sites to the system. After we have added a site, only the spin-1/2 representation on the right hand side of (6) will contribute to the ground state. Thus only a single spin-1/2 representation at each step is needed to represent the ground state. From the arguments of this section we now see that if there is only a single irreducible representation in our truncated basis we have no freedom at all in the projection matrix $P^{j',j}$. Without having to do a variational calculation, the matrix product state with a single spin-1/2 irreducible representation gives us the exact ground state of the AKLT-model.

The fact that the DMRG in the thermodynamic limit produces a ground state in matrix product form also suggests an explanation of the fact that the DMRG is exact for models that have exact matrix product ground states. This is also studied in Chap. 3.1 (II).

4 Expectation values

To do the variational calculation that was discussed in Sect. 3 we take as an ansatz for the ground-state wave function the translationally invariant state $Q = \mathbb{1}$ from (4) which we denote by $|1\rangle$. Thus

$$|1\rangle \equiv \sum_{\{s_j\}} \text{tr} (A[s_n] \dots A[s_1]) |s_n \dots s_1\rangle. \quad (13)$$

Note that although it is not explicitly written out, $|1\rangle$ has a definite number of lattice sites n . For the AKLT model [32] ($\beta = -1/3$) our ground-state ansatz is exact as are the matrix-product states of [5–7].

We will now show how expectation values in the matrix-product states can be calculated. We will do all calculations for the ground-state ansatz $|1\rangle$ but generalization to the general state $|Q\rangle$ in (4) is straightforward. We will calculate several properties of the states. In particular we will determine the form of correlation functions in the matrix-product states. The calculations done in this section will also be of help in the next section where we will look at excited states.

The expectation value of an operator \mathcal{O} , e.g. energy or correlation function, in this state is given by

$$\begin{aligned} \langle 1|\mathcal{O}|1\rangle &= \sum_{\{s_j\}, \{s'_j\}} \text{tr}(A^*[s'_n] \dots A^*[s'_1]) \text{tr}(A[s_n] \dots A[s_1]) \\ &\quad \times \langle s'_n \dots s'_1 | \mathcal{O} | s_n \dots s_1 \rangle . \end{aligned} \quad (14)$$

To write this expression in a simpler form we define the tensor product matrix $(B \otimes C)$ by $(B \otimes C)^{(\alpha, \beta), (\tau, \nu)} = B^{\alpha, \tau} C^{\beta, \nu}$. We will in the rest of the lecture interchangeably use ordinary matrix indices α, β and composite indices (α, β) , where composite indices are written with a parenthesis around them. This means that we can write a $m \times m$ “matrix” A as either a matrix $A^{\alpha, \beta}$ or as a m^2 vector $A^{(\alpha, \beta)}$. When the indices are not explicitly written out, the matrix or vector character of the symbol is assumed to be clear from the context. We now use the trace and matrix product identities $\text{tr}(B) \text{tr}(C) = \text{tr}(B \otimes C)$ and $(BCD) \otimes (EFG) = (B \otimes E)(C \otimes F)(D \otimes G)$ to find

$$\begin{aligned} \langle 1|\mathcal{O}|1\rangle &= \sum_{\{s_j\}, \{s'_j\}} \text{tr}\left((A^*[s'_n] \otimes A[s_n]) \dots (A^*[s'_1] \otimes A[s_1])\right) \\ &\quad \times \langle s'_n \dots s'_1 | \mathcal{O} | s_n \dots s_1 \rangle . \end{aligned} \quad (15)$$

To write this in a more compact form we define a mapping \widehat{M} from 3×3 spin matrices M to $m^2 \times m^2$ matrices \widehat{M} by

$$\widehat{M} \equiv \sum_{s', s} M_{s', s} (A^*[s'] \otimes A[s]) . \quad (16)$$

We denote by $S \equiv (S^x, S^y, S^z)$ the spin-1 representation of total spin and thus by $\hat{S} \equiv (\hat{S}^x, \hat{S}^y, \hat{S}^z)$ the “hat” mapping of the 3×3 spin matrices S . By $\hat{1}$ we denote the “hat” mapping of the 3×3 identity matrix. We now see from (15) that the norm and the expectation value of the spin at the site j is given by

$$\begin{aligned} \langle 1|1\rangle &= \text{tr}(\hat{1}^n), \\ \langle 1|S_j|1\rangle &= \text{tr}(\hat{1}^{n-1} \hat{S}) , \end{aligned} \quad (17)$$

where we in the last equation have used the cyclicity of the trace. Other expectation values are also easily obtained. Since we can factorize matrix elements like

$$\begin{aligned}\langle s'_j, s'_i | \mathbf{S}_i \cdot \mathbf{S}_j | s_j, s_i \rangle &\equiv (\mathbf{S}_i \cdot \mathbf{S}_j)_{s'_j, s'_i, s_j, s_i} \\ &= (\mathbf{S})_{s'_i, s_i} \cdot (\mathbf{S})_{s'_j, s_j},\end{aligned}$$

we find that expectation values of energy and spin-spin correlation function are given by

$$\begin{aligned}\langle 1 | \mathbf{S}_j \cdot \mathbf{S}_{j+1} | 1 \rangle &= \text{tr} \left(\hat{1}^{n-2} \hat{S} \hat{S} \right), \\ \langle 1 | \mathbf{S}_j \cdot \mathbf{S}_{j+l} | 1 \rangle &= \text{tr} \left(\hat{1}^{n-l-1} \hat{S} \hat{1}^{l-1} \hat{S} \right).\end{aligned}\quad (18)$$

A similar but a more complicated formula can be derived also for the operator $(\mathbf{S}_j \cdot \mathbf{S}_{j+1})^2$.

Formulas similar to (17) and (18) for expectation values in MP states have also been derived by Fannes *et al.* [6].

A possible problem with the construction of the projection operator is that parity is not built into the construction of the ground state since the projectors operate from the left to the right. There is therefore the possibility that parity is violated in the ground state $|1\rangle$. In App. A it is shown that the state $|1\rangle$ has parity $(-1)^n$, where n is the number of sites. This appendix also serves as another example for how calculations in the matrix product states can be done.

The ground-state energy

For $m = 6$ the freedom in the projection operators $A[s]$, and thus also in the ground-state ansatz $|1\rangle$, can be parameterized by the parameters ϕ and θ in (12). The energy of the state $|1\rangle$ is then a function of these two parameters,

$$E_{\text{MP}}(\phi, \theta) = \langle 1 | H | 1 \rangle. \quad (19)$$

Note that $\langle 1 | 1 \rangle = 1$ due to (5). Note also that although it is not explicitly written out, E_{MP} depends on the number of lattice sites n . The convergence of E_{MP} to the limit $n \rightarrow \infty$ is however rapid and the value in the thermodynamic limit is easily calculated.

We can now use the variational principle for the energy and determine the parameters ϕ and θ , and thus also the projection matrices $A[s]$ by minimizing E_{MP} over all values of ϕ and θ . For $m = 12$ a similar minimization can be done but with eight parameters instead of two.

For both cases the computational effort is modest and we implemented the minimization procedure in Mathematica on an ordinary desktop workstation. We used a conjugate gradient method [37] to do the minimization in the two- and eight-dimensional spaces.

The projection matrices $A[s]$ obtained by this variational technique were found to agree up to high numerical accuracy with the projection operator obtained from similar DMRG calculations. Note that to compare the projection matrices from the variational calculation with the projection matrices from the DMRG calculation, the DMRG has to be done so as to preserve the total spin symmetry. We used the infinite system DMRG algorithm with a spin-1/2 edge spin. The ground-state energy was estimated from the difference in total energy of two systems from consecutive iteration steps.

Table 1. ground-state energy per site

β	E_{MP} ($m = 6$)	E_{DMRG}	E_{MP} ($m = 12$)	E_{DMRG}	exact	best numerical
-1/3	-0.666667	-0.666667	-0.666667	-0.666667	-2/3	
0	-1.39966	-1.39962	-1.40138	-1.40138	-	-1.401484038971(4)

The result for the lowest energy state for $\beta = -1/3$ and $\beta = 0$ is found in Table 1. The best result known to us for $\beta = 0$ comes from DMRG calculations in [14]. The exact result at the AKLT point $\beta = -1/3$ can be found in [32].

The small difference for the ground-state energy in Table 1 between the variational calculation and the DMRG calculation for $\beta = 0$ can be explained by the fact that the MP ground state ansatz is by construction uniform while the DMRG method will not produce a uniform ground state, since the projection operators A_n of (2) will depend on the site index n .

Correlation functions

The matrix $\hat{1}$ is central in this formulation of matrix product states. In (18) it appears as a kind of transfer matrix for the spin-1 chain. We will now look at the ground-state correlation functions in more detail. It turns out that the matrix product states are either long range ordered or have exponentially decaying correlations and that the correlation lengths for any operator is given by the eigenvalues of $\hat{1}$. Generally the correlation function of two operators is given by (C.f. (18))

$$\begin{aligned} \langle 1 | \mathcal{O}_{i+l} \mathcal{O}_i | 1 \rangle &= \text{tr} \left(\hat{1}^{n-l-i} \hat{\mathcal{O}}_{i+l} \hat{1}^{l-1} \hat{\mathcal{O}}_i \hat{1}^{i-1} \right) \\ &= \text{tr} \left(\hat{1}^{n-l-1} \hat{\mathcal{O}}_{i+l} \hat{1}^{l-1} \hat{\mathcal{O}}_i \right). \end{aligned} \quad (20)$$

Now diagonalize $\hat{1}$,

$$\hat{1} = U X U^{-1}, \quad (21)$$

where U is the matrix containing the eigenvectors of $\hat{1}$ and X is the diagonal matrix with the eigenvalues on the diagonal. It is easy to show that

$\hat{1}$ must have an eigenvalue equal to one [1,2,36]. To do this, we take the m^2 -dimensional vector $v^{(\beta',\beta)} \equiv \delta_{\beta,\beta'}$. By using (5), we now find

$$\begin{aligned} \sum_{\beta',\beta} (\hat{1})^{(\alpha',\alpha),(\beta',\beta)} v^{(\beta',\beta)} &= \sum_{\beta',\beta} \sum_s (A^*[s] \otimes A[s])^{(\alpha',\alpha),(\beta',\beta)} v^{(\beta',\beta)} \\ &= \sum_{\beta',\beta} \sum_s (A^*[s])^{\alpha',\beta'} (A[s])^{\alpha,\beta} \delta_{\beta,\beta'} \\ &= \sum_{\beta,s} (A^*[s])^{\alpha',\beta} (A[s])^{\alpha,\beta} \\ &= \delta_{\alpha,\alpha'} \\ &\equiv v^{(\alpha',\alpha)}. \end{aligned} \quad (22)$$

The vector v is thus an eigenvector to $\hat{1}$ with eigenvalue 1. Further, it is found numerically that all other eigenvalues have absolute values less than one. Now insert (21) into (20) and let us assume that n is large. We then see that only the eigenvalue one will survive in the term $\hat{1}^{n-l-1}$ and we get

$$\langle 1 | \mathcal{O}_{i+l} \mathcal{O}_i | 1 \rangle = \langle \lambda_1 | \hat{\mathcal{O}}_{i+l} U X^{l-1} U^{-1} \hat{\mathcal{O}}_i | \rho_1 \rangle, \quad (23)$$

where $\langle \lambda_1 |$ and $| \rho_1 \rangle$ are the left and right eigenvectors of $\hat{1}$ corresponding to the eigenvalue one. The correlation function will thus be

$$\langle 1 | \mathcal{O}_{i+l} \mathcal{O}_i | 1 \rangle = \sum_{k=1}^{m^2} \langle \lambda_1 | \hat{\mathcal{O}}_{i+l} | \rho_k \rangle \langle \lambda_k | \hat{\mathcal{O}}_i | \rho_1 \rangle x_k^{l-1}, \quad (24)$$

where x_k is the k th eigenvalue of $\hat{1}$, with $x_1 = 1$ and $|x_k| < 1$ for $k = 2 \dots m^2$. If $l-1$ is large, we can extract the leading behavior of (24). However, it turns out that for some operators, the expectation value between $\langle \lambda_1 |$ and $| \rho_1 \rangle$ is zero. The leading behavior will thus be governed by the largest eigenvalue for which the expectation values on the right hand side of (24) does not vanish. We thus find that, if both $\langle \lambda_1 | \hat{\mathcal{O}}_{i+l} | \rho_1 \rangle$ and $\langle \lambda_1 | \hat{\mathcal{O}}_i | \rho_1 \rangle$ are nonzero, the state has long range order since the eigenvalue $x_1 = 1$ will prevent decay of the correlations. If however the rows of $\hat{\mathcal{O}}_{i+l}$ or the columns of $\hat{\mathcal{O}}_i$ are orthogonal to $\langle \lambda_1 |$ and $| \rho_1 \rangle$ respectively, we will have exponential decay of correlations. We get for the correlation length

$$\xi = -\frac{1}{\ln|x_k|}, \quad (25)$$

where x_k is the largest eigenvalue of $\hat{1}$ for which the expectation values in (24) does not vanish. It turns out that the spin-spin correlation function, $\langle S_{i+l}^z S_i^z \rangle$ is governed by the second largest eigenvalue and thus decays exponentially. The so called string correlation function,

$$g(l) = \langle S_{i+l}^z \left(\prod_{j=i+1}^{i+l-1} e^{i\pi S_j^z} \right) S_i^z \rangle, \quad (26)$$

however turns out to be long ranged.

Numerically we find for $m = 12$ the next leading eigenvalue of $\hat{1}$ to be -0.777 , giving an asymptotic spin-spin correlation length from (25) of $\xi = 3.963$, compared to best estimates [14] of $\xi = 6.03(1)$. For $m = 6$ we find the next leading eigenvalue to be -0.681 giving $\xi = 2.60$. We see that the severe truncation of our basis to only six or twelve states has resulted in the asymptotic correlations being quite poor, although we have verified that intermediate length spin-spin correlations are consistent with more precise calculations [38]. For the string correlation function in (26) we find with $m = 12$ that $g(\infty) = -0.3759$, whereas best estimates are [14] $g(\infty) = -0.374325096(2)$.

It is well established that the DMRG works best for gapful systems with short ranged correlations and that the convergence of calculated eigenvalues or correlation functions with m is slower when the gap is small or the correlation length is large. Legeza and Fáth report a decrease in accuracy by several orders of magnitude at a critical point as compared to off-critical points for the Ising model in a transverse field [39]. The number of states m that has to be kept for a certain accuracy thus grows very large as a critical point is approached [40]. Efforts to study phase transitions by investigating the divergence of the correlation function or the vanishing of the gap have not been successful because the difficulties to accurately determine these properties close to a critical point [22,20,41].

How do systems with algebraically decaying correlations (infinite correlation length) fit into the matrix product picture presented in this lecture? Does the DMRG converge to a fixed point in the thermodynamic limit also for these systems? In a recent study by Andersson *et al.* [42] it has been found that for a system of spinless fermions on a lattice the DMRG converges to a fixed point in the thermodynamic limit. This system is gapless and has algebraically decaying correlations. At intermediate range, where the correlations are given by the sum of exponentials in (24), it is found that this sum mimics the behavior of a powerlaw function. The asymptotic correlations on the other hand always turn out to be exponential. It is however found that the asymptotic correlation length defined by (25) increases with m as $m^{1.3}$ which indicates an infinite correlation length for the exact system [42].

5 Bloch States

In this section we use the states $|Q\rangle_n$ as defined in (4), where Q is a general $m \times m$ matrix, to make an ansatz for the low-lying excited states. For a translationally invariant system we can define our states to be Bloch states. A reasonable ansatz for a Bloch state $|Q, k\rangle_n$ defined by a matrix Q and a momentum k is given by

$$|Q, k\rangle_n \equiv \sum_{s_n \dots s_1} \sum_{j=1}^n e^{ijk} \text{tr}\left(A[s_n] \dots A[s_{j+1}] Q A[s_j] \dots A[s_1]\right) \times |s_n \dots s_1\rangle.$$

This wavefunction can be viewed as the ground state $|1\rangle$ with a disturbance Q introduced at some site, and then letting the disturbance run over all sites to form a state with a definite momentum. In this way we get a single “particle” excitation.

We will now calculate properties of these states. We will determine the Hamiltonian and normalization matrices

$$\langle Q', k | H | Q, k \rangle_n \quad (27)$$

$$\langle Q', k | Q, k \rangle_n \quad (28)$$

and find orthonormal eigenstates $|Q_\gamma, k\rangle$ of the matrix (27),

$$\langle Q_{\gamma'}, k | H | Q_\gamma, k \rangle = E_\gamma \delta_{\gamma', \gamma} \quad (29)$$

$$\langle Q_{\gamma'}, k | Q_\gamma, k \rangle = \delta_{\gamma', \gamma}, \quad (30)$$

where γ labels the different eigenstates. Note that states with different momentum k are automatically orthogonal. It should also be emphasized that there will not be any minimization done here. The projection matrices $A[s]$ are the same as for the ground state. The only freedom we have in the states $|Q, k\rangle_n$ reside in the $m \times m$ matrix Q and we will determine the set of eigenmatrices Q_γ from an eigenvalue equation.

We will now derive expressions for matrix elements (27) and (28). The method to calculate them are the same as was presented for the state $|1\rangle$ in the previous section. The Bloch states are however more complicated and the calculations are a bit tedious. We will therefore only summarize them here. The details can be found in [2]. See also App. B for some details regarding the state $|Q\rangle_n$ of (4).

One finds that the general structure of all matrix elements between the Bloch states is that they consist of traces with a convolution sum over matrix products inside each trace. For the overlap we find

$$\langle Q', k | Q, k \rangle_n = n \operatorname{tr} \left((Q' \otimes \mathbb{1}) \sum_{j=0}^{n-1} e^{ijk} \hat{1}^{n-j} (\mathbb{1} \otimes Q) \hat{1}^j \right), \quad (31)$$

with similar, but more complicated, expressions for the Hamiltonian and for the z -component of total spin. For finite-length chains, the sums in expressions of this type can be expediently calculated by a recursive scheme for the case when n is a power of two [2].

By doing a transformation which is described in some detail in App. B, we can calculate a norm matrix $G(k, n)$ defined by

$$\langle Q', k | Q, k \rangle_n \equiv n Q' G(k, n) Q, \quad (32)$$

where the right hand side of (32) is an ordinary vector-matrix-vector product and we regard Q' and Q as m^2 -dimensional vectors and where $G(k, n)$ is an $m^2 \times m^2$ matrix. The nice thing about (32) is that we have effectively turned

the computation of the trace in (31) for all Q and Q' into a matrix inner product between Q , Q' and a single $m^2 \times m^2$ matrix $G(k, n)$, independent of Q and Q' .

Similarly we can define a $m^2 \times m^2$ Hamiltonian matrix by

$$(Q', k | H | Q, k)_n \equiv n Q' H(k, n) Q . \quad (33)$$

A matrix $S_T^z(k, n)$ representing the z -component of total spin can be defined analogously.

Rewriting the traces over matrix products as ordinary vector-matrix-vector products is very useful since now we can diagonalize matrices like $H(k, n)$ in (33) and thereby find the eigenstates in the subspace spanned by all ansatz vectors $|Q, k\rangle_n$. The principle for calculating these matrices is described in App. B. Due to the number of terms in the expressions for the expectation values as in (31) it is however numerically cumbersome for finite length chains.

There is however an elegant way to extract the leading behavior of $H(k, n)$ and $G(k, n)$ as $n \rightarrow \infty$ [2], which we will briefly describe here.

Let us first define the z -transform (sometimes called a discrete Laplace transform) of a series $\{a_n\}_{n=0}^\infty$ by $F(\lambda) = \sum_{n=0}^\infty a_n e^{-n\lambda}$. Let us now denote the sum inside the trace in (31) by S_n , so that

$$S_n = \sum_{j=0}^{n-1} e^{ijk} \hat{1}^{n-j} (\mathbb{1} \otimes Q) \hat{1}^j .$$

We now define a series $\{S_n\}_{n=0}^\infty$, and take the z -transform of this sequence. By examining the analytical structure of the transformed series it is possible to extract the leading behavior of the sum S_n , as $n \rightarrow \infty$. In this way we can get the asymptotic form of $G(k, n)$ in the limit of large n . This procedure is then also applied to all sums in the expressions for the Hamiltonian matrix elements. This whole procedure finally results in the asymptotic forms

$$H(k, n) = n^2 H_2(k) + n H_1(k) + H_0(k) + \mathcal{O}(z)^n \quad (34)$$

$$G(k, n) = n G_1(k) + G_0(k) + \mathcal{O}(z)^n , \quad (35)$$

with $H(k, n)$ and $G(k, n)$ as defined in (33) and (32) and were $H_i(k)$, $i = 0, 1, 2$ and $G_i(k)$, $i = 0, 1$ are $m^2 \times m^2$ matrices. Here z represents the next leading eigenvalue of $\hat{1}$ which numerically is found to be $|z| \approx 0.8$ for $m = 12$. There are thus very small corrections to the asymptotic form. We also find that H_2 and G_1 are non-vanishing only when the momentum k is zero. The energy of a state $|Q, k\rangle_n$ is

$$E_Q(k, n) = \frac{Q H(k, n) Q}{Q G(k, n) Q} . \quad (36)$$

Since the normalization matrix $G(k, n)$ is not the identity matrix, the naive basis states $(Q_{i,j})^{\alpha,\beta} = \delta_{i,\alpha} \delta_{j,\beta}$, with $i = 1, \dots, m$ and $j = 1, \dots, m$, are

not orthonormal. We thus have to consider also the matrix $G(k, n)$ in the diagonalization. The eigenvalue equation which must be solved is thus

$$H(k, n)Q(k, n) = E_Q(k, n)G(k, n)Q(k, n), \quad (37)$$

where $Q(k, n)$ is an m^2 -dimensional vector. Combining (34), (35) and (37) for $k \neq 0$ we have

$$(nH_1(k) + H_0(k))Q(k, n) = (nE_0 + \Delta_k(n))G_0(k)Q(k, n), \quad (38)$$

where E_0 is the ground-state energy per site and $\Delta_k(n)$ is the excitation energy. E_0 denotes the ground-state energy per site in the limit $n \rightarrow \infty$, and is therefore independent of n . Since we are interested in the solutions to (38) when $n \rightarrow \infty$ we assume also Q and Δ_k to be independent of n . From the solution of (38) we get the ground-state energy, which is the same energy as we found from the minimization in Sect. 4. We also get the excitation spectrum Δ_k and the corresponding orthonormal eigenstates $|Q_\gamma, k\rangle$, where γ labels the different eigenstates. These states represent our “single magnon” states and will have the properties of (29) and (30). An asymptotic form for the z -component of total spin, $S_T^z(k, n)$, similar to the form for $H(k, n)$, containing terms up to order n^2 can be determined in a similar way.

We have computed the asymptotic forms when $n \rightarrow \infty$ for the Hamiltonian and norm matrices defined in (33) and (32) as well as for the total spin matrix for $\beta = 0$ and for different momenta k . We simultaneously diagonalize $H(k, n)$ and $S_T^z(k, n)$ which means that the eigenstates will be states with a definite z -component of total spin. We have numerically determined these states along with their energy and spin expectation values. For $m = 12$ we find the single-particle spectrum shown in Fig. 5.

The low-lying triplet branch defines the gap $\Delta_\pi = 0.4094$, which is very good compared to the most accurately known result [14, 24, 25] of 0.410502(1). Furthermore, we compute the spin wave velocity $v = 2.452$ to be compared to the calculations in [24], where $v = 2.49(1)$ was obtained. Clearly we reproduce the single-particle triplet excitations with high accuracy considering the few number of states in our basis. Our calculation also yields a detailed spectrum of lowest lying “single-magnon” excitations shown by dotted lines in Fig. 5. Our second lowest energy excitation at $k = \pi$ is a singlet shown by a dotted line in the same figure with $\Delta_\pi(\text{singlet}) = 2.348$.

As a function of k , the second lowest single-particle excitation is either a singlet or a spin-2 object, as has also been observed in exact finite size calculations [19]. Parity of each of the elementary excitations is verified by checking the relation (39) with Q as well as with the matrices A . The boundary to two particle excitations at a given value of k is computed explicitly by minimizing the sum of energies of excitations whose pseudo-momentum sums to k , and similarly for the three-particle excitations. These results are shown by the light and dark shaded regions in Fig. 5. The picture fits well with previously obtained results.

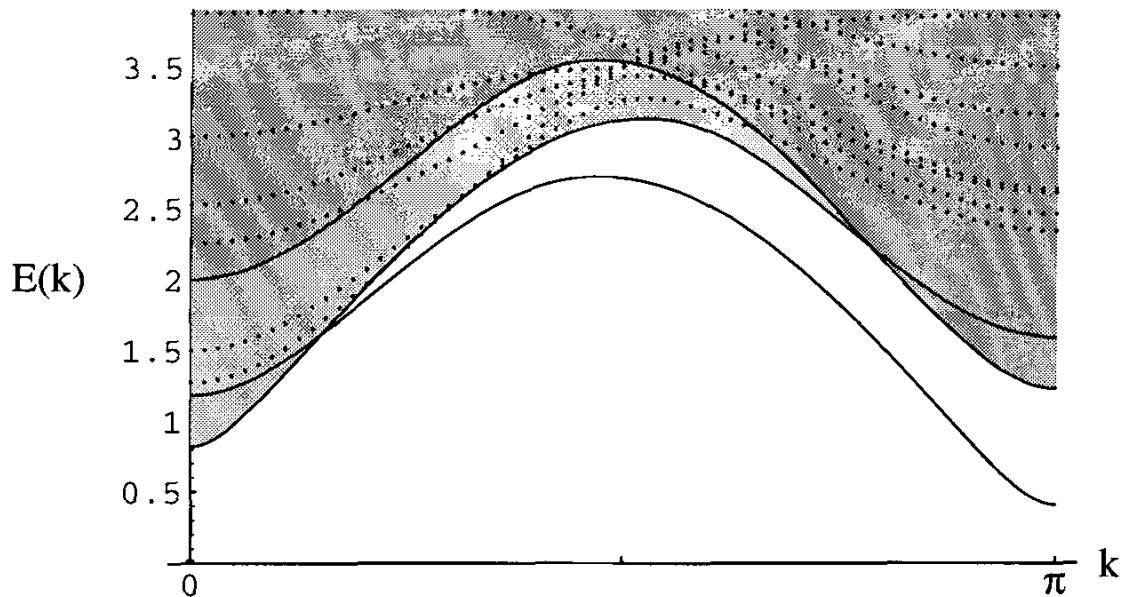


Fig. 5. The spectrum for $\beta = 0$ and $m = 12$ is shown. The lowest single particle triplet is shown as a solid line, with the lightly shaded region representing two-particle excitations and the dark region three-particle excitations. Solid lines define the boundaries to the two- and three-particle continuum. Dotted lines indicate the spectrum of higher energy single-magnon excitations. The spin of these dotted excitations are, in order of increasing energy at $k = \pi$: 0,1,2,2,3,1,1,0.

6 Summary and Conclusions

In this lecture we have studied the DMRG method as it reaches the thermodynamic limit. It was found that if the renormalization converges to a fixed point, the DMRG ground state can be written in a matrix product form. This form was used to define a set of matrix product ansatz wave functions that can be investigated on their own, with no direct connection to the DMRG. The numerical calculations were tested on the spin-1 chain. It was found that by using symmetry considerations, the number of free parameters in these ansatz wave functions can be greatly reduced. In particular, the spin rotational symmetry of the spin-1 chain was explored and the basis states in the truncated basis were chosen to be complete representations of total spin of the chain. With just a few free parameters, a variational principle for the energy was now used to find the lowest energy state in the few-parameter family of ansatz states. By performing the variational calculation it was found that the variational matrix product ground state agrees to a very high accuracy with the ground state found from similar DMRG calculations.

This suggests that the rapid convergence of the DMRG is explained by the fact that the states selected are optimally chosen eigenstates of total block spin. Properly chosen, these states are highly efficient for building wave functions with a small basis that have low total spin for all subblocks.

By further investigating the matrix-product states it was found that the asymptotic correlations in the states are either long ranged or decay exponentially. Intermediate range correlations were found to be either long ranged or to be a sum of exponentials. This suggests an explanation of the fact that the DMRG has been found to work best for gapful system with exponentially decaying correlations.

In this lecture we also defined a set of states describing elementary excitations in the spin-1 chain. These states could be explored in the thermodynamic limit and the excitation spectrum of the Heisenberg spin-1 chain was numerically estimated and found to agree well with previous results.

A Parity

In this appendix we will show in detail that the state $|1\rangle$ has a definite parity. Let \mathcal{P} be the parity operator. We thus have

$$\begin{aligned}\mathcal{P}|1\rangle &= \sum_{\{s_j\}} \text{tr} (A[s_n] \dots A[s_1]) \mathcal{P}|s_n \dots s_1\rangle \\ &= \sum_{\{s_j\}} \text{tr} (A[s_n] \dots A[s_1]) |s_1 \dots s_n\rangle.\end{aligned}$$

Suppose now that there exists an invertible $m \times m$ matrix $Q_{\mathcal{P}}$ such that

$$Q_{\mathcal{P}} A[s] = \text{sign}[\mathcal{P}] (A[s])^T Q_{\mathcal{P}}, \quad (39)$$

where A^T denotes transpose and $\text{sign}[\mathcal{P}]$ is a proportionality constant that will be seen to be the eigenvalue of the parity operator. Then it follows that

$$\begin{aligned}\mathcal{P}|1\rangle &= \sum_{\{s_j\}} \text{tr} (Q_{\mathcal{P}}^{-1} Q_{\mathcal{P}} A[s_n] \dots A[s_1]) |s_1 \dots s_n\rangle \\ &= \text{sign}[\mathcal{P}]^n \sum_{\{s_j\}} \text{tr} (A^T[s_n] \dots A^T[s_1]) |s_1 \dots s_n\rangle \\ &= \text{sign}[\mathcal{P}]^n \sum_{\{s_j\}} \text{tr} (A[s_1] \dots A[s_n]) |s_1 \dots s_n\rangle \\ &= \text{sign}[\mathcal{P}]^n |1\rangle.\end{aligned}$$

Thus, for the ground state to have definite parity, it is sufficient that such a $Q_{\mathcal{P}}$ exists. How do we find this matrix, if it exists? We multiply both sides of the defining relation (39) by $A^\dagger[s]$ and sum over s . Using (5) we find that

$$\begin{aligned}Q_{\mathcal{P}}^{\alpha,\beta} &= \text{sign}[\mathcal{P}] \sum_s (A^T[s])^{\alpha,\tau} Q_{\mathcal{P}}^{\tau,\nu} (A^T[s])^{\nu,\beta} \\ &= \text{sign}[\mathcal{P}] \left(\sum_s (A^T[s])^{\alpha,\tau} A^{\beta,\nu}[s] \right) Q_{\mathcal{P}}^{\tau,\nu} \\ &= \text{sign}[\mathcal{P}] \sum_s (A^T[s] \otimes A[s])^{(\alpha,\beta),(\tau,\nu)} Q_{\mathcal{P}}^{\tau,\nu}.\end{aligned}$$

Thus, Q_P , if it exists, is the eigenvector of the matrix $\sum_s (A^T[s] \otimes A[s])$ with eigenvalue ± 1 . The parity operator of (39) can numerically be computed and it is found that the ground state has parity $(-1)^n$, where n is the number of sites.

B The General State $|Q\rangle$

In this appendix we show in some detail how to rewrite matrix elements between the MP states $|Q\rangle_n$ defined in (4) into ordinary matrix-vector products. The Bloch states $|Q, k\rangle_n$ can be treated in an analogous way but the calculations will be more tedious.

To calculate the norm we use the same trace and tensor product identities as when deriving (15). We find that

$$\begin{aligned} (Q'|Q)_n &= \sum_{\{s_j\}} \text{tr} ((Q')^* A^*[s_n] \dots A^*[s_1]) \text{tr}(Q A[s_n] \dots A[s_1]) \\ &= \text{tr}((Q'^* \otimes Q) \hat{1}^n). \end{aligned} \quad (40)$$

We can rewrite this trace as ordinary matrix products. To do this we first define the generalized transpose $M^{T_{p_1, p_2, p_3, p_4}}$ of a matrix M by

$$(M^{T_{p_1, p_2, p_3, p_4}})^{(\alpha_1, \alpha_2), (\alpha_3, \alpha_4)} = M^{(\alpha_{p_1}, \alpha_{p_2}), (\alpha_{p_3}, \alpha_{p_4})}, \quad (41)$$

where $\{p_1, p_2, p_3, p_4\}$ is a permutation of $\{1, 2, 3, 4\}$. We also define a tilde operator \tilde{M} by the formula

$$\text{tr}((Q' \otimes Q)M) = \sum_{\alpha', \beta', \alpha, \beta} (Q')^{\alpha', \beta'} (\tilde{M})^{(\alpha', \beta'), (\alpha, \beta)} Q^{\alpha, \beta}, \quad (42)$$

so that the tilde operator effectively generates the matrix corresponding to the inner product of Q' and Q with M . One finds by writing out (42) in components that

$$\tilde{M} = M^{T_{3142}}.$$

Hence

$$((Q')^{\alpha', \beta'} | Q^{\alpha, \beta})_n = Q'^{(\alpha', \beta')} G(n)^{(\alpha', \beta'), (\alpha, \beta)} Q^{(\alpha, \beta)}, \quad (43)$$

with

$$G(n) = \widetilde{(\hat{1}^n)}. \quad (44)$$

We have now effectively turned the computation of the trace in (40) for all Q and Q' into a matrix inner product between Q , Q' and a single $m^2 \times m^2$ matrix G , independent of Q and Q' . Note that on the right side in (43) we write Q and Q' as vectors of length m^2 .

Similarly we can compute the expectation value of the Heisenberg Hamiltonian defined in (1) with $\beta = 0$ as

$$(Q'|H|Q)_n = \sum_{i=0}^{n-2} \text{tr} \left((Q' \otimes Q) \hat{1}^i \hat{S} \hat{S} \hat{1}^{n-2-i} \right), \quad (45)$$

where \hat{S} denotes the hat mapping in (16) of the spin-1 matrices. Similar to (43) we can define an $m^2 \times m^2$ matrix $H(n)$ by

$$(Q'|H|Q)_n = Q' H(n) Q, \quad (46)$$

where $H(n)$ is independent of Q' and Q . Note that the left hand side of (46) is again an ordinary vector-matrix-vector product where Q' and Q are regarded as m^2 -dimensional vectors. Other operators like the z -component of total spin can be determined in a similar way.

By rewriting the traces over matrix products as ordinary vector-matrix-vector products we can diagonalize matrices like $H(n)$ in (46) and thereby find the eigenstates in the subspace spanned by all ansatz vectors $|Q\rangle_n$. The energy of a state $|Q\rangle_n$ is given by

$$E_Q = \frac{(Q|H|Q)_n}{(Q|Q)_n} = \frac{Q H(n) Q}{Q G(n) Q}. \quad (47)$$

The eigenvalue equation to be solved is thus (C.f. (37))

$$H(n)Q = E_Q G(n)Q. \quad (48)$$

References

1. S. Östlund and S. Rommer, Phys. Rev. Lett. **75**, 3537 (1995)
2. S. Rommer and S. Östlund, Phys. Rev. B **55**, 2164 (1997)
3. S.R. White, Phys. Rev. Lett. **69**, 2863 (1992)
4. S.R. White, Phys. Rev. B **48**, 10345 (1993)
5. L. Accardi, Phys. Rep. **77**, 169 (1981)
6. M. Fannes, B. Nachtergael and R.F. Werner, Europhys. Lett. **10**, 633 (1989); J. Phys. A **24**, L185 (1991); Commun. Math. Phys. **144**, 443 (1992)
7. A. Klümper, A. Schadschneider and J. Zittartz, Europhys. Lett. **24**, 293 (1993); Z. Phys. B **87**, 281 (1992); C. Lange, A. Klümper and J. Zittartz, Z. Phys. B **96**, 267 (1994)
8. A.K. Kolezhuk and H.-J. Mikeska, Phys. Rev. B **56**, R11380 (1997); A.K. Kolezhuk, J. Phys.: Cond. Matt. **10** 6795 (1998)
9. G. Su, Phys. Lett. A **213**, 93 (1996)
10. A. Schadschneider and J. Zittartz, Ann. der Phys. **4**, 157 (1995)
11. A.K. Kolezhuk, H.J. Mikeska and S. Yamamoto, Phys. Rev. B **55**, R3336 (1997)
12. S. Brehmer, H.J. Mikeska and U. Neugebauer, J. Phys. C: Solid State Phys. **8**, 7161 (1996)
13. J.M. Román, G. Sierra, J. Dukelsky and M.A. Martín-Delgado, J. Phys. A: Math. Gen. **31**, 9729 (1998)

14. S.R. White and D.A. Huse, Phys. Rev. B **48**, 3844 (1993)
15. J. Solyom, Phys. Rev. B **36**, 8642 (1987)
16. G. Fáth and J. Sólyom, Phys. Rev. B **47**, 872 (1993)
17. Y. Xian, J. Phys.: Cond. Matt. **5**, 7489 (1993)
18. R.F. Bishop, J.B. Parkinson and Y. Xian, J. Phys.: Cond. Matt. **5**, 9169 (1993)
19. J. Oitmaa, J.B. Parkinson and J.C. Bonner, J. Phys. C: Solid State Phys. **19**, L595 (1986)
20. U. Schollwöck, T. Jolicœur and T. Garel, Phys. Rev. B **53**, 3304 (1996)
21. T. Xiang and G.A. Gehring, Phys. Rev. B **48**, 303 (1993)
22. R.J. Bursill, T. Xiang and G.A. Gehring, J. Phys. A **28**, 2109 (1995)
23. F.D.M. Haldane, Phys. Lett. A **93**, 464 (1983); Phys. Rev. Lett. **50**, 1153 (1983); J. Appl. Phys. **57**, 3359 (1985)
24. E.S. Sørensen and I. Affleck, Phys. Rev. Lett. **71**, 1633 (1993)
25. T. Sakai and M. Takahashi, Phys. Rev. B **42**, 1090 (1990)
26. L.A. Takhtajan, Phys. Lett. A **87**, 479 (1982)
27. H. Babujian, Phys. Lett. A **90**, 479 (1982)
28. I. Affleck, Phys. Rev. Lett. **55**, 1355 (1985); I. Affleck and F.D.M. Haldane, Phys. Rev. B **36**, 5291 (1987)
29. G.V. Uimin, JEPT Lett. **12**, 225 (1970)
30. C.K. Lai, J. Math. Phys. **15**, 1675 (1974)
31. B. Sutherland, Phys. Rev. B **12**, 3795 (1975)
32. I. Affleck, T. Kennedy, E.H. Lieb and T. Tasaki, Phys. Rev. Lett. **59**, 799 (1987); Commun. Math. Phys. **115**, 477 (1988)
33. See any textbook on Quantum Mechanics, e.g. J.J. Sakurai, Modern Quantum Mechanics, Addison-Wesley (1985)
34. S. Qin, T.K. Ng and Z.B. Su, Phys. Rev. B **52**, 12844 (1995)
35. W. Marshall, Proc. Roy. Soc (London) Ser. A **232**, 48 (1955); E. Lieb and D. Mattis, J. Math. Phys. **3**, 749 (1962)
36. J. Dukelsky, M.A. Martín-Delgado, T. Nishino and G. Sierra, Europhys. Lett. **43**, 457 (1998)
37. W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, Numerical Recipes, Cambridge University Press (1992)
38. E. S. Sørensen and I. Affleck, Phys. Rev. B **49**, 15771 (1994)
39. Ö. Legeza and G. Fáth, Phys. Rev. B **53**, 14349 (1996)
40. S.R. White and I. Affleck, Phys. Rev. B **54**, 9862 (1996)
41. A. Drzewiński and J.M.J. van Leeuwen, Phys. Rev. B **49**, 403 (1994)
42. M. Andersson, M. Boman and S. Östlund, preprint cond-mat/9810093

4 A Recurrent Variational Approach

Miguel A. Martín-Delgado¹ and Germán Sierra²

¹ Departamento de Física Teórica I, Universidad Complutense
E-28040 Madrid, Spain.

² Instituto de Matemáticas y Física Fundamental, C.S.I.C.
E-28006 Madrid, Spain.

The introduction of the Density Matrix Renormalization Group (DMRG) by S.R. White [1,2] has changed in many aspects the field of strongly correlated electrons. Originally intended as a new and powerful numerical tool [1–6], the DMRG has also spurred the interest of theorists [7–16] in the search for new analytic albeit non-perturbative methods to deal with the difficulties posed by strongly correlated systems. Introductions can be found in [17–20], and Chap. 2(1).

This chapter is a review of our work on the Recurrent Variational Approach (RVA) [21–25]. This method is a variation of the original DMRG algorithm, but it is based on a different physical picture. The main idea is to build the system step by step, starting from a small initial system and adding one site at a time. This allows us to understand the physics of the system analytically. Thus, we will start by giving a brief review of analytic approaches to the DMRG. The aim of this analytical approach is to try to understand the relevant physical degrees of freedom so that we can figure out what the underlying physics is in a strongly correlated system. This initial analytical goal has been also developed in order to later acquire more numerical precision. To do this, the method becomes more numerical and somehow stands in between an analytical formulation of the DMRG and a numerical one.

The chapter is based on a series of papers [21–25] and several seminars and courses. We focus our attention on the practical purpose of the RVA rather than on the purely theoretical aspects and thus we illustrate the application of the method with examples. Due to the great interest acquired by the subject of ladder systems (see [26] for a review), we have taken these systems as the testgrounds to set up the fundamentals of the method.

We present in Table 1 a sketch of the relationship and differences between the RVA and the DMRG methods. Both in the RVA and the DMRG there is a truncation procedure. In the RVA there is a single state ($m = 1$) which depends on a collection of variational parameters which serve to explore the ground-state manifold. In the DMRG one keeps $m > 1$ states to build the ground state of the system. Likewise, in the RVA the system (chain, ladder, etc) is constructed step by step adding a certain unit (site, rung, etc) and this leads to recursion relations satisfied by the RVA-state (usually the

ground state). This is the analogue of the Wilsonian onion scheme used by the DMRG. Finally, in both methods there is an optimization process to select the states. In the RVA it is done by minimizing the ground-state energy while in DMRG it is done by using the density matrix to project onto the most probable states.

Table 1. Schematic comparison between the RVA and the DMRG (RR stands for Recurrence Relation)

Comparison	RVA	DMRG
# States kept	$m = 1$	$m > 1$
Order RR's	$1^{\text{st}}, 2^{\text{nd}}, \dots$	1^{st}
Optimization	Lowest energy state	Most probable states

Originally, the variational RVA was introduced to solve the RVB (Resonating Valence Bond) for even-legged ladders [21] (see next section). Then it was realized that this recursive method was reminiscent of the Wilsonian RG method and contact was made with the matrix-product approach and the thermodynamic limit of DMRG (Östlund and Rommer [15], Chap. 3(I)). There are however some subtle differences between the RVA and the matrix-product (MP) method as presented in the latter references. The ansätze of [15] satisfy periodic boundary conditions, are uniform and translationally invariant. The RVA states satisfy open boundary conditions and can in fact be non uniform. In this respect the RVA is more faithful to the DMRG which can be considered as a highly performant non uniform variational matrix-product method.

On the other hand the RVA allows a simple treatment of systems with holes with and without the Gutzwiller constraint of non double occupancy. In these cases the recurrence relations (RR) are both for the size and the number of holes. We can easily generate doubly projected BCS-like states. The projections being on the number of electrons and the no double occupancy.

What seems to be lacking in the RVA is the concept of a density matrix. As we shall show in Sect. 4 the RVA and more generally the MP methods give rise, in the thermodynamic limit, to a density matrix which emerges as the left eigenvector of a non-symmetric transfer matrix. The latter transfer matrix is constructed out of the MP variational parameters.

The RVA, MP and DMRG methods deal very well with systems with a finite correlation length [37]. However, for gapless systems the DMRG requires a number of states kept m increasing with the size of the system [29].

The organization of this chapter is as follows. In Sects. 2 and 3 we introduce the RVA to solve analytically the resonating valence bond problem in a two-leg Heisenberg ladder. In Sect. 4 we present another extension of the RVA based on first-order recurrence relations, which leads to a matrix-product representation of the ansatz. We apply this to the spin-one Heisenberg chain. It is shown how a density matrix emerges in this formalism and allows us to make contact with the DMRG method. In Sect. 5 we extend the RVA to deal

with doped square ladders, taking into account both spin and charge degrees of freedom. In Sect. 6 the RVA is applied to study Heisenberg and t-J models on diagonal ladders. These types of ladders serve as new settings to make extensions of the RVA for different basis of states, such as S_z -states, and for non-uniform ansatz which are those having variational parameters depending locally on the position along the diagonal ladder. This corresponds to RVA applications not in the thermodynamic limit. We close this Chapter in Sect. 7 by discussing new developments and future problems.

1 Two-Legged Spin Ladders

The Recurrent Variational Approach was originally introduced in [21] to solve the computation of RVB states [30] in Heisenberg ladders with an even number of legs. The reason for this restriction comes from two facts. On one hand, theoretical studies have found that antiferromagnetic spin ladders are gapped (gapless) depending on whether the number of legs n_l is an even (odd) number. This prediction has been confirmed experimentally in several compounds [26]. On the other hand, we know that RVB states are gapped. Intuitively, the gap comes from the breaking of a valence bond (singlet) of the covered lattice and thereby forming a triplet state at that bond. This singlet-triplet excitation costs an energy of J which signals the presence of a spin gap in the spectrum of the RVB states. Thus, RVB states are a good basis for the description of even-legged ladders. Moreover, to be more concrete, we choose the simplest even ladder possible, namely, the two-leg ladder with an AF-Heisenberg model described by two coupling constants: J for the links in the legs, and J' for the vertical rungs, see Fig. 1.

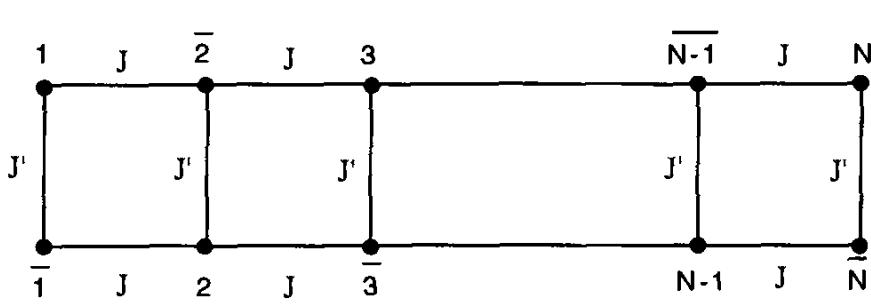


Fig. 1. The two-leg ladder with an AF-Heisenberg model described by two coupling constants: J for the links in the legs, and J' for the vertical rungs

The Hamiltonian reads

$$H = J \sum_{k=1}^{N-1} (\mathbf{S}_k \cdot \mathbf{S}_{\bar{k+1}} + \mathbf{S}_{k+1} \cdot \mathbf{S}_{\bar{k}}) + J' \sum_{k=1}^N \mathbf{S}_k \cdot \mathbf{S}_{\bar{k}} \quad (1)$$

where N denotes the number of vertical rungs of the ladder and open boundary conditions have been assumed. As the two-leg ladder is a bipartite lattice, we label the sites with and without bars to distinguish the two sublattices in which the ladder can be decomposed.

What if the number of legs is odd? White, Noack and Scalapino [31] came up with a nice picture to qualitatively distinguish the different behaviour of even-legged vs. odd-legged spin ladders. Their basic idea is to see when the resonance mechanism of valence bonds can be better achieved depending on the number of legs in a spin ladder. For a simple 2×2 plaquette lattice this resonating mechanism is shown in Fig. 2. In this simple lattice, the ground state (g.s.) of the AF-Heisenberg model is precisely the superposition of the two configurations of pair of bonds (vertical and horizontal) depicted in Fig. 2. Recall that a bond between two sites is a singlet state. Thus, the RVB state is exact and it has lower energy than either VB configurations (vertical or horizontal) that form it. We thus say that the *resonance of the VB states lowers the energy of the variational state*.

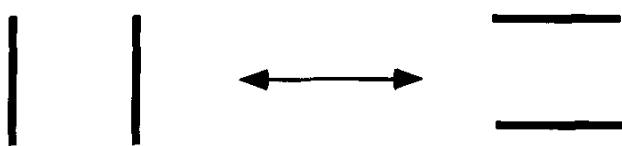


Fig. 2. The basic bond resonance mechanism between horizontal and vertical bonds in an elementary plaquette.

With this resonating mechanism in mind, let us see what happens on spin ladders starting first with the simplest two-leg ladder and then adding more legs. For a two-leg ladder, there are four types of VB configurations, namely,

Resonating: (Fig. 3 (a)) This consists of vertical and horizontal VB states arranged such that the basic resonance mechanism for a single plaquette (Fig. 2) can take place. In particular, this means that whenever there is a horizontal VB state in one leg, there must be another one located opposite in the other leg.

Staggered: (Fig. 3 (b)) This consists of only horizontal VB states arranged such that the basic resonance mechanism for a single plaquette (Fig. 2) cannot take place. The horizontal VB states are shifted by one link from one another.

As the staggered configuration is incapable of resonating, it is energetically unfavorable. However, it is possible to produce a local region of staggered VB states by introducing two spin defects (uncoupled spins in non-singlet states) at the borders of that region (Fig. 3 (c)). Again, within the short range RVB scenario, we expect that the two spin defects will get confined into a pair on the same rung (Fig. 3 (d)). The reason for this is that in this way we are reducing the staggering regions which are costly in energy. This qualitative picture for the confinement of spin defects is readily extrapolable to any even-legged ladder.

Now let us turn to a three-leg ladder and we immediately see that this setting is qualitatively different. First, in a three-leg ladder we only have one type of order (Fig. 3 (d)), unlike the two possibilities (resonating vs. staggered) in a two-leg ladder. Thus, when a spin defect is introduced there

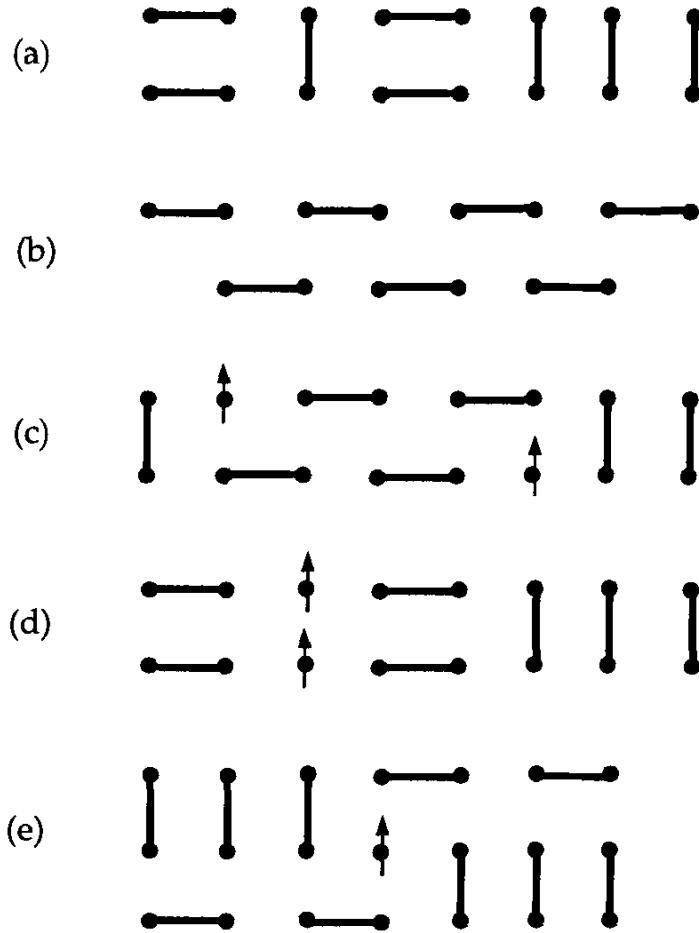


Fig. 3. Different Valence Bond configurations in two-leg and three-leg ladders. (a) Resonating. (b) Staggered. (c) Spin defects bordering a staggered local configuration. (d) Confinement of spin defects. (e) Non-confinement of spin defects in a three-leg ladder.

is no change in the type of order (Fig. 3 (e)) for it only amounts to a mere shift of configurations. This shift does not cost energy. The final outcome is that spin defects do not come in pairs in a three-leg ladder.

Thus, we arrive at the following qualitative picture distinguishing even-legged and odd-legged ladders:

Even-legged Ladders: They exhibit confinement of spin defects.

Odd-legged Ladders: They do not exhibit confinement of spin defects.

These qualitative results were confirmed with the DMRG method [31].

After this brief discussion of RVB states in general ladders we come to the point of the RVA for two-leg spin ladders.

Resonon states: In order to construct a variational RVB state, we need to classify all the possible short-range VB configurations (dimers) in a two-leg ladder. The guiding idea to achieve this is again the resonating mechanism of valence bonds in the elementary 2×2 plaquettes forming the ladder as explained above. We shall not consider staggered configurations for they raise the energy. To perform this classification, we shall choose a reference state formed by all bonds in the vertical positions for every rung out of N forming the ladder (Fig. 4). This state we denote as $|0\rangle$. There is a physical reason

underlying this state. It is the ground state for the J -term in the Hamiltonian (1). The strong-coupling limit of a two-leg ladder is defined as the coupling region where the vertical rungs are stronger than the horizontal links, namely, $J' \gg J$. In this limit, the most important state is precisely $|0\rangle$. It is known that the strong-coupling limit provides a reasonable picture of the physics for a two-leg ladder even down to the isotropic limit where $J' = J$. Thus, it is a convenient starting point for the ladder and it is a simple limit also because the diagonalization of the rungs states is trivial: they are singlets and triplets. Once the horizontal coupling J is turned on, it begins to be possible to have a pair of vertical bonds resonating to horizontal bonds in a certain plaquette of the ladder. Let us call a *resonon* to a resonating 2×2 plaquette (Fig. 5). We can place one resonon on any of the $N - 1$ plaquettes of the ladder. In this situation we say that the effect of the coupling J is to make the plaquettes start resonating. If we increase the value of J for J' fixed, it will also probably have two resonons present in the ladder (Fig. 6), and eventually, there will be $[N/2]$ plaquettes ([...] meaning integer part) resonating at the same time, see Fig. 7. In summary, the problem of classifying VB states reduces to first decomposing the Hilbert space of dimer states by sectors with a fixed number of resonons, as follows:

Zero-resonon sector: $|0\rangle$. (Fig. 4)



Fig. 4. The zero-resonon state $|0\rangle$ in a two-leg ladder. It is made of all vertical rung singlets. It is the reference state in the strong coupling limit $J' \gg J$.

One-resonon sector: $|x_1\rangle$. (Fig. 5)

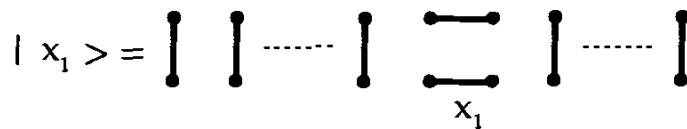


Fig. 5. A generic one-resonon state $|x_1\rangle$ in a two-leg ladder.

Two-resonon sector: $|x_1, x_2\rangle$. (Fig. 6)

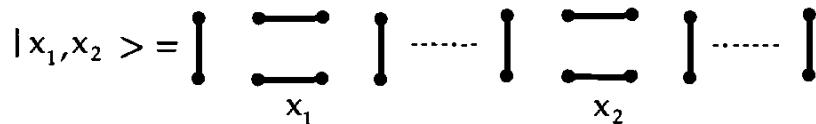


Fig. 6. A generic two-resonon state $|x_1, x_2\rangle$ in a two-leg ladder.

$\frac{N}{2}$ -resonon sector: $|x_1, x_2, \dots, x_M\rangle$. (Fig. 7)

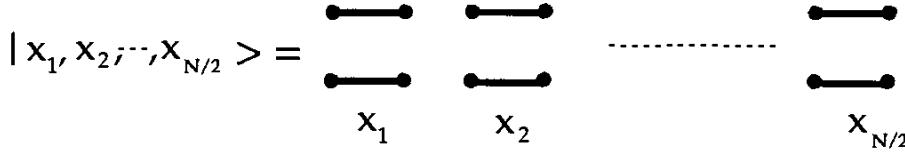


Fig. 7. The all-resonon state $|x_1, x_2, \dots, x_{\frac{N}{2}}\rangle$ in a two-leg ladder.

With this classification of states we may write the following Generalized Dimer-RVB state for a two-leg spin ladder,

$$|\psi(u)\rangle_{RVB} = \sum_{M=0}^{[N/2]} u^M \sum_{\{x_1, x_2, \dots, x_M\}} |x_1, x_2, \dots, x_M\rangle \quad (2)$$

where the sum is taken over all sectors with M resonons and u is the amplitude for a resonon. Here u is a variational parameter to be determined upon minimization of the ground-state energy. As a result, it is a function of the ratio of couplings, namely, $u = u(J/J')$.

Second, it is an interesting combinatorial problem to compute the number $D(N, M)$ of dimer states with N rungs and M resonons which is given by,

$$D(N, M) = \binom{N - M}{M} \quad (3)$$

Thus, the dimension, denoted by F_N , of the whole Hilbert space of dimer RVB states in a two-leg ladder of length N is,

$$F_N = \sum_{M=0}^{[N/2]} \binom{N - M}{M} \quad (4)$$

We have arrived at a well-known formula in the theory of partitions [32] and it corresponds to a combinatorial representation of the Fibonacci number F_N .

Recurrence relations for wave functions: The key point about the result (4) for the dimension of the Hilbert space of dimer RVB states is the fact that Fibonacci numbers are well-known to satisfy recursion relations which allow us to construct successive members of the series from previous members, namely,

$$F_{N+2} = F_{N+1} + F_N \text{ with initial values, } F_0 = F_1 = 1 \quad (5)$$

In fact, using this recursive idea we could have equally arrived to the generation of all possible dimer-RVB configurations present in the the classification by resonon sectors outlined above. To see this, let us use the Fibonacci

series (5) to generate all possible states: for $N = 1$ there is only one state $F_1 = 1$; for $N = 2$ there are $F_2 = 2$; for $N = 3$ there are $F_3 = 3$; for $N = 4$ there are $F_4 = 5$ and so on and so forth.

Denoting by $|N\rangle \equiv |\psi(u)\rangle_{RVB}$ a generic RVB state (2) for a two-leg ladder of N rungs, there are only two possibilities or *movements* to create generic RVB states of higher length, namely,

- Addition of one vertical rung to create the state $|N + 1\rangle$.
- Addition of one pair of horizontal bonds (resonon) to create $|N + 2\rangle$.

From these arguments we can establish that the dimer RVB states (2) satisfy a *recursion relation* given by,

$$|N + 2\rangle = |N + 1\rangle \otimes |\phi_1\rangle_{N+2} + u |N\rangle \otimes |\phi_2\rangle_{N+1,N+2} \quad (6)$$

where the state denoted by $|\phi_1\rangle_{N+2}$ is a vertical rung at position $N + 2$ and $|\phi_2\rangle_{N+1,N+2}$ is made up of a pair of horizontal bonds located between the rungs at $(N + 1, N + 2)$, i.e.,

$$|\phi_2\rangle_{N+1,N+2} = (N + 1, \overline{N + 2})(N + 2, \overline{N + 1}) \quad (7)$$

where (N, \overline{M}) denotes a singlet between the sites N and \overline{M} . The recursion relation (RR) (6) is symbolically represented in Fig. 8.

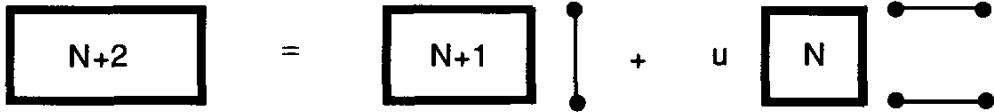


Fig. 8. Diagrammatic representation of the second order recurrence relation (RR).

Using (6) one can generate recursively the dimer RVB state $|N\rangle$ from previous states with lower length and estimate variationally the ground-state energy as,

$$e_N^{(g.s.)} \equiv \frac{\langle N | H_{AFM}^{two-leg} | N \rangle}{(2N) \langle N | N \rangle} = e(N, J'/J; u) \quad (8)$$

The value of u is fixed by minimization of (8). The RVA allows us to find analytically quantities such as, $\langle \psi | \psi \rangle(u)$, $\langle \psi | H_{AFM} | \psi \rangle(u)$, $\langle \psi | \mathbf{S}_i \cdot \mathbf{S}_j | \psi \rangle(u)$, etc ... without going through the tedious task of computing overlappings with the Sutherland rules [33].

It is worthwhile to mention that this recursive method for constructing the states of the ladder bears a great deal of similarity with the iterative method employed by the DMRG for constructing matrix elements of observables in chains, ladders etc..., see Chap. 2(I). In this sense we may consider the RVA as an analytical formulation of the DMRG.

Norms of states: To compute the norms of the states (1) let us define

$$Z_N = \langle N|N \rangle \quad (9)$$

It is also convenient to define an auxiliary function Y_N as follows,

$$Y_N = \langle_N (\phi_1| \otimes \langle N-1|)|N\rangle \quad (10)$$

The second order RR's for the states leads (6) to a closed set of RR's for the overlaps Z_N, Y_N , namely,

$$\begin{aligned} Z_{N+2} &= Z_{N+1} + uY_{N+1} + u^2Z_N \\ Y_{N+2} &= Z_{N+1} + \frac{u}{2}Y_{N+1} \end{aligned} \quad (11)$$

where we have made use of the result,

$$_{N+1}\langle\phi_1|\phi_2\rangle_{N,N+1} = \frac{1}{2}|\phi_1\rangle_N \quad (12)$$

The RR's (11) together with the initial conditions,

$$Z_0 = Z_1 = 1, Y_0 = 0, Y_1 = 1 \quad (13)$$

define values Z_N and Y_N for arbitrary values of N by iteration of (11).

Expectation values: The expectation values of the Hamiltonian can be determined using the same scheme outlined above. In fact, we can write the definition,

$$E_N = \sqrt{1/(Z_N)} \quad , \quad D_N = \langle_N (\phi_1| \otimes \langle N-1|) E_N |N\rangle \quad (14)$$

To obtain the RR's satisfied by E_N, D_N , one splits the Hamiltonian H_N of a two-leg ladder of length N (1) into two pieces:

$$H_N = H_{N-1} + H_{N-1,N} \quad (15)$$

where H_{N-1} is the Hamiltonian of length $N-1$ and $H_{N-1,N}$ is the rest of the whole H_N Hamiltonian, $H_{N-1,N} \equiv H_N - H_{N-1}$, which is made of one vertical rung and two horizontal links. With this splitting and using (6) and (9), we obtain,

$$\begin{aligned} E_{N+2} &= E_{N+1} + J'\epsilon_0 Z_{N+1} + u(D_{N+1} + (2J + J')\epsilon_0 Y_{N+1}) \\ &\quad + u^2(E_N + 2J\epsilon_0 Z_N) \\ D_{N+2} &= E_{N+1} + J'\epsilon_0 Z_{N+1} + \frac{u}{2}(D_{N+1} + (2J + J')\epsilon_0 Y_{N+1}) \end{aligned} \quad (16)$$

where $\epsilon_0 = -3/4$ is the lowest eigenvalue of the operator $\mathbf{S}_1 \cdot \mathbf{S}_2$.

The initial conditions for E_N and D_N are

$$E_0 = 0, E_1 = J'\epsilon_0, D_0 = 0, D_1 = J'\epsilon_0. \quad (17)$$

The RR's for the energies involve the norms of the states and depend both on u and the coupling constants J', J .

2 Results for the Variational Ground-State Energy

The ground-state energy for a ladder of length N is estimated using equation (9). In the thermodynamic limit $N \rightarrow \infty$ one can find a closed expression for the density energy per site [21],

$$e_\infty = \lim_{N \rightarrow \infty} \frac{1}{2N} \frac{E_N}{Z_N} = \frac{R(\alpha)}{2\alpha Q'(\alpha) P(\alpha)} \quad (18)$$

in terms of three polynomials P, Q, R evaluated at the biggest root α of the cubic polynomial $Q(y)$. Finally one looks for the absolute minimum of e_∞ by varying the parameter u . In Table 2 we show the ground-state energies per site for different values of the coupling constant ratio J/J' , varying through strong, intermediate and weak-coupling regimes.

Table 2. The values $-e_\infty^{\text{MF}}/J'$ are mean field (MF) values taken from [34], while $-e_\infty^{\text{Lan}}/J'$ are Lanczos values taken from [35].

J/J'	u	$-e_\infty^{\text{RVA}}/J'$	$-e_\infty^{\text{MF}}/J'$	$-e_\infty^{\text{Lan}}/J'$
0	0	0.375	0.375000	
0.2	0.128521	0.383114	0.382548	
0.4	0.323211	0.40835	0.405430	
0.6	0.578928	0.44853	0.442424	
0.8	0.87441	0.499295	0.489552	
1	1.18798	0.556958	0.542848	0.578
1.25	1.58519	0.63518	0.614473	0.6687
1.66	2.21853	0.772172	0.738360	0.8333
2.5	3.39153	1.06915	1.002856	1.18
5	5.9777	1.99285		2.265

In the strong coupling regime $J/J' < 1$ the RVA states give a slightly better ground-state energy than the mean field result. This latter state produces rather unphysical results for $J/J' > 1$, which does not occur in our case. The RVA results compare also well with the exact results for $J' \sim J$.

The RR's can also be used to compute the spin correlator $\langle \mathbf{S}_i \cdot \mathbf{S}_j \rangle$ which has an exponential decay behaviour $\exp(-|i-j|/\xi)$, with ξ the spin correlation length which satisfies the exact equation

$$u^3 \mathcal{L}^3 - (2+u)u^2 \mathcal{L}^2 - (2+4u)u^2 \mathcal{L} + 4u^3 = 0 \quad (19)$$

where $\mathcal{L} = e^{1/\xi}$. For $J = J'$ one finds $\xi = 0.737$ [21], which can be compared with its exact value given by 3.2 [31]. The latter results imply that at the isotropic point the NNN bonds are quite important. The simplest improvement of the dimer-RVB state is to add a bond of length $\sqrt{5}$ (analogue to a knight move [36]), which leads to a third order RR. The g.s. energy per site of this state is given by -0.5713 while the spin correlation length is 0.959 [21].

Hence there is an improvement in both quantities but still one needs longer bonds.

The RVA as well as the matrix-product method work well when the correlation length of the system is small. In the DMRG context this means that the spectrum of the density matrix is dominated by a few eigenvalues [27,28]. However for systems with a large correlation length one has to consider matrix products or DMRG's with high values of m . As shown recently in [29] the correlation length ξ scales as m^β with some exponent β .

3 Matrix-Product States for Integer-Spin Chains

In this section we consider another extension of the RVA which consists in setting up a recursion relation of first order instead of second order and higher RR's that we used before.

Simple examples of RR's of first order were used as a guide to construct the RVA recursion relations in Heisenberg and t-J ladders [21–23]. In those first order RR's the variational ansatz was too simple so as to capture the main features of the systems. Thus we had to resort to higher order RR's. Another alternative that we briefly examine in the following is to keep the RR's to first order at the expense of retaining all possible states available in the problem at that stage.

We shall consider a spin chain with spin S at each site, where S is an integer with a AF-Heisenberg interaction. We shall set up the recursion relations as an ansatz for the finite-size construction of the ground state of the spin chain. Then, we shall show how to achieve the thermodynamic limit $N \rightarrow \infty$ analytically and then compare the structure of this solution with the DMRG approach revealing a close relationship [24].

Let us denote the basis states as $|a, JM\rangle_N$, where N is the length of the chain, $a = 1, \dots, d_J$ denotes the multiplicity of the total spin J of the state and M is the third component of the spin. Counting the number of multiplets we have $m = \sum_J d_J$, which corresponds to a number $m_W = \sum_J (2J+1) d_J$ of states in the standard DMRG. Thus, let us write down the following variational ansatz for a finite size N spin chain:

$$|a_1, J_1 M_1\rangle_N = \sum_{a_2, J_2, M_2, M} A_{a_1 J_1, a_2 J_2} \times |SM\rangle_N \otimes |a_2, J_2 M_2\rangle_{N-1} \langle SM, J_2 M_2 | J_1 M_1 \rangle \quad (20)$$

where $\langle SM, J_2 M_2 | J_1, M_1 \rangle$ are Clebsch-Gordan (CG) coefficients which show up in this a rotationally invariant treatment of the spin chain. $A_{a_1 J_1, a_2 J_2}$ are the variational parameters of the ansatz. In (20) we keep not only singlets (valence bonds) but also higher spin states (triplet, etc...) which will improve the numerical results of this type of ansätze.

The ansatz (20) is assumed to be normalized unlike previous RVA treatments. This implies that the variational parameters $A_{a_1 J_1, a_2 J_2}$ are subject to the following orthonormalization conditions,

$$A_{a_1 J_1, a_2 J_2} = 0 \text{ unless } |J_2 - S| \leq J_1 \leq J_2 + S \quad (21)$$

$$\sum_{a_2, J_2} A_{a_1 J_1, a_2 J_2}^* A_{a'_1 J_1, a_2 J_2} = \delta_{a_1 a'_1} \quad (22)$$

Equation (21) follows from the CG decomposition $S \otimes J_2 \rightarrow J_1$ in (20).

The initial data of the recurrence relation (20) is given by choosing a spin $S/2$ irrep at the end of the chain. This choice eliminates the multiplicity associated to the effective spins $S/2$ at the ends of the chain. The sum in J' 's in (20) is of course restricted to a finite set of spins.

After many iterations the DMRG reaches a fixed point, and the ground state exhibits a similar structure as in the ansatz (20) which is called a matrix-product structure (MP) [15]. The parameters $A_{a_1 J_1, a_2 J_2}$ are determined as usual by minimizing the energy of the states $|a, JM\rangle_N$ in the limit where $N \rightarrow \infty$. For this purpose let us define the following quantity,

$$E_{aa'J}^{(N)} =_N \langle a, JM | H_N | a', JM \rangle_N \quad (23)$$

where H_N is the Hamiltonian acting on the chain with N sites. These matrix elements are analogous to the energy matrix elements in (14), etc. From this we can derive a recurrence relation for $E_{aa'J}^{(N)}$, similar to RR's (16),

$$E_{aa'J}^{(N+1)} = E_{aa'J}^{(N)} + \sum_{a''a'''} \frac{\langle a, JM | H_{N-1,N} | a'', JM' \rangle_N}{\langle a''a' | H_{N-1,N} | a'''a' \rangle_N} E_{a''a'''J'}^{(N)}, \quad (24)$$

where $H_{N-1,N}$ is the coupling between sites N and $N-1$. It is now clear that $A_{a_1 J_1, a_2 J_2}$ is given by

$$V_{aa'J} =_N \langle a, JM | H_{N-1,N} | a', JM \rangle_N \quad (25)$$

and $V_{aa'J}$ is the matrix element of the part of the Hamiltonian which couples the sites N and $N-1$, which does not dependent on N ,

$$V_{aa'J} =_N \langle a, JM | H_{N-1,N} | a', JM \rangle_N \quad (26)$$

The differences between RR's in (24) and (16) have a very definite origin:

1. The absence of auxiliary functions such as D_N (14) in (24) is because the latter are first order RR's instead of second order as in (16).
2. The absence of norm terms such as Z_N (9) in RR's (24) is because the states in the ansatz (20) are normalized.

For the Heisenberg model, $H_{N-1,N} = \mathbf{S}_{N-1} \cdot \mathbf{S}_N$, and applying the Wigner-Eckart theorem, we find the following expression for V in terms of 6-j symbols,

$$V_{a_1 a_2 J_1} = \sum_{a_3 J_2, a_4 J_3, a_5 J_4} \mathcal{H}_{J_1 J_2 J_3 J_4} \times \quad (27)$$

$$A_{a_1 J_1, a_3 J_2} A_{a_2 J_1, a_4 J_3} A_{a_3 J_2, a_5 J_4} A_{a_4 J_3, a_5 J_4},$$

$$\begin{aligned} \mathcal{H}_{J_1, J_2, J_3, J_4} = & (-)^{2S+J_1+J_2+J_3+J_4+1} S(S+1)(2S+1) \\ & \times \sqrt{(2J_2+1)(2J_3+1)} \left\{ \begin{array}{ccc} 1 & S & S \\ J_1 & J_2 & J_3 \end{array} \right\} \left\{ \begin{array}{ccc} 1 & S & S \\ J_4 & J_2 & J_3 \end{array} \right\} \end{aligned} \quad (28)$$

The solution of (24) can be expressed in matrix notation as,

$$|E^{(N)}\rangle = (1 + \mathbf{T} + \mathbf{T}^2 + \dots + \mathbf{T}^{N-3}) |V\rangle + \mathbf{T}^{N-2} |E^{(2)}\rangle \quad (29)$$

where $|E^{(N)}\rangle$ is regarded in (29) as a vector whose components are labeled by $(aa'J)$. The entries of \mathbf{T} are given by (25).

In the limit $N \rightarrow \infty$ the contribution from $|E^{(2)}\rangle$ drops off and, as we shall show below, $E_{aa'J}^{(N)}$ behaves as,

$$\lim_{N \rightarrow \infty} \frac{1}{N} E_{aa'J}^{(N)} = \delta_{aa'} e_\infty \quad (30)$$

where e_∞ can be identified with the ground-state energy density and it reads,

$$e_\infty = \sum_{aa'J} \rho_{aa'J} V_{aa'J} \quad (31)$$

In (31) $\rho_{aa'J}$ is the left eigenvector of the matrix \mathbf{T} with maximal eigenvalue 1, and plays the role of a density matrix in this formalism. The proof of (30) and (31) follows from the existence of an eigenvalue of the matrix \mathbf{T} equal to 1 [37,15]. This property can be deduced from the normalization condition (22). Let us call $|v\rangle$ and $\langle\rho|$ the right and left eigenvectors associated to the eigenvalue 1 of \mathbf{T} , which we shall assume to be unique,

$$\mathbf{T} |v\rangle = |v\rangle, \quad \langle\rho| \mathbf{T} = \langle\rho| \quad (32)$$

Table 3. Here m is total the number of multiplets, N_A is the number of independent variational parameters, d_J is the number of multiplets with spin J , $e_\infty^{\text{MP,DMRG}}$ is the g.s. energy density of the matrix-product state (DMRG), $1 - P_m$ is the probability of the states truncated out in the DMRG and ξ^{MP} is the spin correlation length of the MP state. The exact results are given by $e_\infty = 1.4014845$ and $\xi = 6.03$ [4].

m	N_A	$d_{1/2}$	$d_{3/2}$	$d_{5/2}$	$-e_\infty^{\text{MP}}$	$-e_\infty^{\text{DMRG}}$	$1 - P_m$	ξ^{MP}
1	0	1	0	0	1.333333	1.333333	1.6×10^{-2}	0.910
2	2	1	1	0	1.399659	1.369077	1.4×10^{-3}	2.600
3	4	2	1	0	1.401093	1.392515	1.3×10^{-5}	3.338
4	7	2	2	0	1.401380	1.401380	1.6×10^{-5}	3.937
5	10	2	2	1	1.401443	1.401436	7.6×10^{-6}	4.085
6	13	2	3	1	1.401474	1.401468	1.3×10^{-6}	4.453

Then (22) implies that $|v\rangle$ is given in components by $v_{aa'J} = \delta_{aa'}$. On the other hand the quantities $\rho_{aa'J}$ that appear in (31) are nothing but the

components of $\langle \rho |$, and they are found by solving the eigenvalue problem (32). In (31) we have normalized $\langle \rho |$ according to,

$$\langle \rho | v \rangle = 1 \rightarrow \sum_{aJ} \rho_{aaJ} = 1 \quad (33)$$

If one views $\rho_{aa'J}$ as a matrix which consists of blocks labelled by J , it has the properties of a density matrix, and it corresponds precisely to the reduced density matrix of the blocks in the DMRG formalism (see below) [24].

The results of the minimization (with constraints) procedure are summarized in Table 3 and compare with DMRG results [24].

To analyze in more detail the relation between the variational approach and the DMRG we give in Table 2 the eigenvalues of the matrix $\rho_{aa'J}$, and those of the DMRG reduced density matrix in the case where $m = 6$. The latter matrix has dimension $3m$ and the truncation DMRG method consists in choosing m states with highest eigenvalues w_{DMRG}^2 , which add up to $1 - P_m$ (see Table 3). For this reason we have to scale the DMRG weights of the states kept in order that they sum up to 1.

$$\bar{w}_{\text{DMRG}}^2 = w_{\text{DMRG}}^2 / P_m \quad (34)$$

Table 4. a and J are the labels of the irrep, w_{MP}^2 , are the eigenvalues of the MP density matrix, and \bar{w}_{DMRG}^2 are the corresponding DMRG eigenvalues kept in the RG process and normalized to 1. The data corresponds to $m = 6$

a	J	w_{MP}^2	\bar{w}_{DMRG}^2
1	1/2	.9695581	.9696232
2	1/2	.0007662	.0007599
1	3/2	.0295443	.0294877
2	3/2	.0001119	.0001089
3	3/2	.0000078	.0000085
1	5/2	.0000118	.0000118

Table 3 and Table 4 suggest that the predictions made by the matrix-product ansatz and the DMRG should become identical for large values of m .

On the other hand, the variational approach introduced above leads to a superblock configuration $[B] \bullet [B^r]$, [24,38] while in the standard DMRG it is $[B] \bullet \bullet [B^r]$. This is not a major difference though, and as a matter of fact, the superblock scheme $[B] \bullet [B^r]$ is employed in the momentum-space approach to DMRG by T. Xiang, Chap. 5(I). The extensions of the results presented in this section to the two-leg $S=1/2$ ladder can be found in [25].

4 Two-Legged t-J Ladders

So far we have been dealing only with models involving spin degrees of freedom. Now, we shall present the extension of the RVA when there are also

charge degrees of freedom. We shall do this by studying the t-J model in two types of ladders: rectangular and diagonal ladders. It is also possible to extend the method to the Hubbard model.

Together with the Hubbard model, the t-J model is the most studied model in connection to High- T_c Superconductivity. The t-J model is the strong coupling version of the Hubbard model and reduces to the Antiferromagnetic Heisenberg model (AFH) at half-filling. As a matter of fact, we have the following sequence of models,

$$\text{Hubbard Model} \xrightarrow{U/t \gg 1} \text{t-J Model} \xrightarrow{\text{half-filling}} \text{AF Heisenberg Model}$$

Heuristically, we can also motivate the study of the t-J model by simply starting from a AF-Heisenberg model to simulate undoped spin systems and then introducing movable holes with a t -kinetic term added to the Heisenberg model. By doping we always mean the deviation from half-filling.

It is convenient to restrict the study of a complicated model such as the t-J model to ladder lattices and especially we shall be dealing with the simplest case corresponding to the two-leg ladder. Then, the Hamiltonian we are interested in reads as follows,

$$\begin{aligned} H &= H_S + H_K & (35) \\ H_S &= J' \sum_{i=1}^N (\mathbf{S}_{i,1} \cdot \mathbf{S}_{i,2} - \frac{1}{4} n_{i,1} n_{i,2}) \\ &\quad + J \sum_{a=1}^2 \sum_{i=1}^{N-1} (\mathbf{S}_{i,a} \cdot \mathbf{S}_{i+1,a} - \frac{1}{4} n_{i,a} n_{i+1,a}) \\ H_K &= - P_G \left[\sum_{a=1}^2 \sum_{i=1;s}^{N-1} t c_{i,a;s}^\dagger c_{i+1,a;s} + \sum_{i=1;s}^N t' c_{i,1;s}^\dagger c_{i,2;s} \right] P_G + h.c. \end{aligned}$$

where the $c_{i,s}$ ($c_{i,s}^\dagger$) is the electron destruction (creation) operator for site i and spin s , n_i is the occupation-number operator, and P_G is the Gutzwiller projection operator which forbids doubly occupied sites. Here J and t are Heisenberg and hopping couplings, respectively, along the horizontal links of the legs while J' and t' are similar couplings along the vertical rungs of the ladder. We have made this distinction among couplings anticipating the study of the strong-coupling limit we shall carry out, as we did in previous sections with the AF-Heisenberg ladders. The isotropic point is defined as $J' = J$, $t' = t$. The label $a = 1, 2$ stands for the two legs of the ladder.

The big difficulty working with the t-J model is precisely this constraint of no double occupancy. Analytic approaches usually fail to implement the whole physics underlying this constraint or are subject to unproven assumptions. In the RVA treatment of the t-J model, the double occupancy constraint is *built-in* from the very beginning.

The problem we face is to generalize the RVA of the AF-Heisenberg ladder to the case of doping. Being the RVA a variational method, there is no mechanical way to proceed, unlike the perturbative methods. As usual with variational methods, we need a hint from the physics of the problem in order to set up the methodology. In the case of the two-leg t-J ladder there are some basic features coming from the strong-coupling analysis and other means which are helpful.

We know by now that there is a spin gap in the spectrum of the two-leg ladder when it is undoped. The question is what happens when we dope the system. For low doping, it is expectable that the spin gap remains, but there appear charge gapless excitations associated to the collective motion of pairs of holes which form a sort of superfluid. These facts are reflected in the low-energy physics of this model which falls into the universality class of the Luther-Emery (LE) liquid [43], unlike the one-dimensional version of the t-J model which is known to be a Luttinger Liquid (LL) [44]. In a LE liquid, the spin modes are gapped while the charge degrees of freedom are gapless.

Recurrent relations for doped states

In order to set up recursion relations for doped states in the t-J model it is convenient to recall the elementary recipies we have used so far:

- Set up the strong-coupling limit of the model.
- Select the basic local configurations in this regime.
- Solve the model in a small cluster, e.g. a 2×2 plaquette.

To understand this basic physics, let us take the undoped case $x = 0$ as a reference. In this case the second order RR is given in (6). This corresponds to having no pair of holes $P = 0$ and the state is simply $|N, P = 0\rangle$. It is worthwhile to point out that if we set $N = 2$ in (6) we precisely recover the *exact* ground state for the simple 2×2 plaquette, namely,

$$|N = 2, P = 0\rangle = |\phi_{1,0}\rangle_1 |\phi_{1,0}\rangle_2 + |\phi_{2,0}\rangle_{1,2} \quad (36)$$

where $|\phi_{2,0}\rangle_{1,2}$ is the state defined in (7) times u (see (41)).

We can apply this same reasoning for the construction of the RR for states $|N, P\rangle$ for the case $N = 2$ and $P = 1$. Let us label the sites of the 2×2 lattice as 1, 2, 3, 4 in a clockwise fashion. To solve for the state $|2, 1\rangle$ we use symmetry arguments to write down this state as (Fig. 9),

$$|2, 1\rangle = a |\text{edge}\rangle + b |\text{diagonal}\rangle \quad (37)$$

$$|\text{edge}\rangle = (12) + (34) + (23) + (14), \quad |\text{diagonal}\rangle = (24) + (13) \quad (38)$$

$$|2,1\rangle = a \left(\begin{array}{c} \bullet \\ | \\ \circ \end{array} + \begin{array}{c} \bullet \\ | \\ \circ \end{array} + \begin{array}{c} - \\ | \\ \circ \end{array} + \begin{array}{c} \circ \\ | \\ \bullet \end{array} \right) + b \left(\begin{array}{c} \bullet \\ \diagdown \\ \circ \end{array} + \begin{array}{c} \circ \\ \diagup \\ \bullet \end{array} \right)$$

Fig. 9. The exact ground state for a single plaquette with two holes (case N=2 and P=1).

$$\begin{aligned} \boxed{N+2,P+1} &= \boxed{N+1,P+1} \begin{array}{c} \bullet \\ | \\ \bullet \end{array} + \boxed{N+1,P} \begin{array}{c} \circ \\ | \\ \circ \end{array} \\ &\quad - u \quad \boxed{N,P+1} \begin{array}{c} - \\ | \\ - \end{array} \\ &\quad + b \quad \boxed{N,P} \left(\begin{array}{c} \circ \\ \diagdown \\ \bullet \end{array} + \begin{array}{c} \bullet \\ \diagup \\ \circ \end{array} \right) \\ &\quad + c \quad \boxed{N,P} \left(\begin{array}{c} - \\ | \\ \circ \end{array} + \begin{array}{c} \circ \\ | \\ - \end{array} \right) \end{aligned}$$

Fig. 10. A pictorial representation of the second-order recurrence relation (40) for two-leg t-J ladder.

where (i, j) symbolically denotes a bond between sites i and j . The ratio of the probability amplitudes a and b and the ground-state energy are given by,

$$\frac{a}{b} = \frac{1}{[2 + (J/4t)^2]^{1/2} - J/4t}, E(2, 1) = -\frac{1}{2}[J + \sqrt{J^2 + 32t^2}] \quad (39)$$

With the help of these configurations we can set up the RR that generates the dimer-hole RVB states as follows, (Fig. 10)

$$\begin{aligned} |N + 2, P + 1\rangle &= |N + 1, P + 1\rangle |\phi_{1,0}\rangle_{N+2} + |N + 1, P\rangle |\phi_{1,1}\rangle_{N+2} \\ &\quad + |N, P + 1\rangle |\phi_{2,0}\rangle_{N+1,N+2} + |N, P\rangle |\phi_{2,1}\rangle_{N+1,N+2} \end{aligned} \quad (40)$$

where we have gathered the basic local configurations in terms of pair field operators $\Delta_{i,j}^\dagger = \frac{1}{\sqrt{2}}(c_{i,\uparrow}^\dagger c_{j,\downarrow}^\dagger - c_{i,\downarrow}^\dagger c_{j,\uparrow}^\dagger)$,

$$\begin{aligned} |\phi_{1,1}\rangle_x &= |0\rangle_x \\ |\phi_{1,0}\rangle_x &= \Delta_{(x,1)(x,2)}^\dagger |0\rangle_x \\ |\phi_{2,0}\rangle_{x,x+1} &= -u \Delta_{(x,1)(x+1,1)}^\dagger \Delta_{(x,2)(x+1,2)}^\dagger |0\rangle_{x,x+1} \\ |\phi_{2,1}\rangle_{x,x+1} &= [b (\Delta_{(x,1)(x+1,2)}^\dagger + \Delta_{(x,2)(x+1,1)}^\dagger) \\ &\quad + c (\Delta_{(x,1)(x+1,1)}^\dagger + \Delta_{(x,2)(x+1,2)}^\dagger)] |0\rangle_{x,x+1} \end{aligned} \quad (41)$$

where $|0\rangle_x$ is the vacuum associated with the rung labelled by the coordinate x . The states $|\phi_{n,p}\rangle$, involve $n = 1, 2$ rungs and $p = 0, 1$ pairs of holes. The variational parameter u gives the amplitude of the resonance of a pair of bonds between vertical and horizontal positions, while b and c are the variational parameters associated with the diagonal and horizontal configurations of two holes, respectively [22].

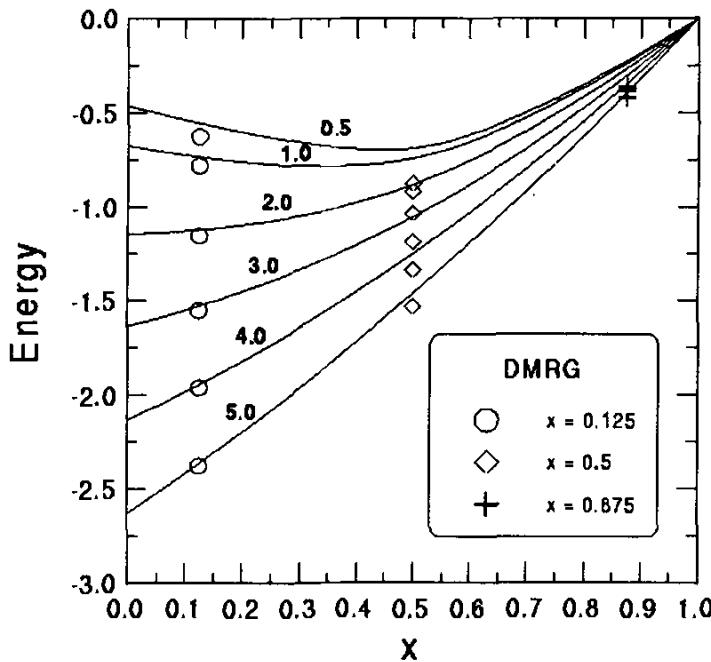


Fig. 11. Ground-state energy per site of the 2×32 ladder with $J = 0.5, t = t' = 1$ and $J' = 0.5, 1, 2, 3, 4, 5$. The continuum curves are obtained with the RVA, while the special symbols are the DMRG data corresponding to $x = 1/8, 1/2$ and $7/8$ respectively.

Norms and energies: The parameters u, b, c are found by the standard minimization of the mean value of the energy $\langle N, P | H_N | N, P \rangle / \langle N, P | N, P \rangle$. As usual, (40) implies recursion relations for the energy overlaps. Let us define the following norms $Z_{N,P}$, mean values $E_{N,P}$ and auxiliary functions $Y_{N,P}$, $D_{N,P}$ and $W_{N,P}$

$$\begin{aligned} Z_{N,P} &= \langle N, P | N, P \rangle & Y_{N,P} &= {}_N \langle \phi_{1,0} | (N-1, P | N, P \rangle) \\ E_{N,P} &= \langle N, P | H_N | N, P \rangle & D_{N,P} &= {}_N \langle \phi_{1,0} | (N-1, P | H_N | N, P \rangle) \\ W_{N,P} &= \langle N, P | n_N | N, P \rangle \end{aligned} \quad (42)$$

where n_N is the number operator acting on the rung N . The off-diagonal overlaps arise from the cross terms when applying (40) to the ket and the bras in $\langle N+2, P+1 | N+2, P+1 \rangle$ and $\langle N+2, P+1 | H_{N+2} | N+2, P+1 \rangle$.

The recursion relations satisfied by (42) are given by,

$$\begin{aligned}
Z_{N+2,P+1} &= Z_{N+1,P+1} + u^2 Z_{N,P+1} + u Y_{N+1,P+1} \\
&\quad + Z_{N+1,P} + 2(b^2 + c^2) Z_{N,P} \\
Y_{N+2,P+1} &= Z_{N+1,P+1} + u/2 Y_{N+1,P+1} \\
E_{N+2,P+1} &= E_{N+1,P+1} - J' Z_{N+1,P+1} + u^2 E_{N,P+1} \\
&\quad - (2J + J'/2)u^2 Z_{N,P+1} + E_{N+1,P} \\
&\quad + 2(b^2 + c^2)E_{N,P} - (2Jc^2 + 4bt + 8bct') Z_{N,P} \\
&\quad + uD_{N+1,P+1} - 2u(J + J'/2)Y_{N+1,P+1} - 4tbY_{N+1,P} \\
&\quad - \frac{1}{4}JW_{N+1,P+1} - \frac{1}{4}Ju^2W_{N,P+1} - \frac{1}{4}J(b^2 + c^2)W_{N,P} \\
D_{N+2,P+1} &= E_{N+1,P+1} - J'Z_{N+1,P+1} + u/2D_{N+1,P+1} \\
&\quad - u(J + J'/2)Y_{N+1,P+1} - 2tbZ_{N,P} - \frac{1}{4}JW_{N+1,P+1} \\
W_{N+2,P+1} &= 2Z_{N+1,P+1} + 2u^2 Z_{N,P+1} + 2(b^2 + c^2)Z_{N,P} \\
&\quad + 2uY_{N+1,P+1}
\end{aligned} \tag{43}$$

The initial conditions for these equations read as follows,

$$\begin{aligned}
Z_{0,0} &= 1, \quad Y_{0,0} = E_{0,0} = D_{0,0} = W_{0,0} = 0 \\
X_{N,P} &= 0, \text{ for } N < P \text{ and } X = Z, Y, E, D, W
\end{aligned} \tag{44}$$

Ground-state properties: It is possible to retrieve now many ground-state properties from the RR's in (43) performing the minimization process with respect to the variational parameters u, b, c for a fixed value of the coupling constants J, J', t, t' . In this case we must also take the finite density of holes limit to get meaningful variable-doping quantities. It is clear that the RR's in (43) are more complicated to carry out an analytic solution in the limit when $N \rightarrow \infty, P \rightarrow \infty, x = P/N$ fixed, and we shall not undertake this method of solution this time. Alternatively, we can keep the number of rungs N at a fixed value, though large enough to eliminate finite-size effects, and then vary the number of pairs of holes P in order to cover the whole range of the doping parameter $0 \leq x \leq 1$. This finite- (N, P) minimization process is done numerically by implementing the recursion equations in (43) by means of an iterative algorithm in a computer.

The computer needs are very modest both in memory space and CPU time and with a standard minimization library like those in [46,47] it is possible to minimize very large two-leg ladders, over $N = 200$ rungs, although here we shall present results for sizes 32×2 in order to compare with available DMRG results [22]. In Fig. 11 we show some results comparing the RVA and the DMRG. Observe the huge agreement for the strong-coupling region.

The parameter u starts from a positive value corresponding to the undoped ladder [21], and it decreases upon doping until a critical value $x_c(J/J')$, where it vanishes. For higher dopings u becomes negative. This behaviour is

very interesting and it suggests a continuous crossover between the RVB picture and the BCS picture of pairing in the strong coupling regime. We shall call the region below x_c underdoped and above it, overdoped.

For the undoped ladder the parameter u can be interpreted as the square of the RVB amplitude h_{RVB} for having a bond along the legs [21]. The analogous amplitude for a bond along the rungs has been implicitly normalized to 1. For low doping, i.e. $x < x_c$, since $u(x) > 0$, we can similarly define a doping-dependent amplitude for a leg-bond as follows,

$$u(x) = h_{\text{RVB}}^2(x) > 0, \quad (x < x_c) \quad (45)$$

In order to fulfill the Marshall theorem [40] for the undoped ladder one requires the RVB amplitude $h_{\text{RVB}}(0)$ to be positive [41], which explains why $u(0)$ is also positive. At $x = 0$, $h_{\text{RVB}}(0)$ increases with J/J' due to the resonance between rung and leg singlets, according to the RVB scenario. Upon doping, however, the holes act as a destructive interference which degrades progressively the aforementioned resonance mechanism. This explains why $u(x)$ and $h_{\text{RVB}}(x)$ decrease with x . For $x < x_c$ the ground state is dominated by the resonating valence bonds and the RVB picture remains qualitatively correct.

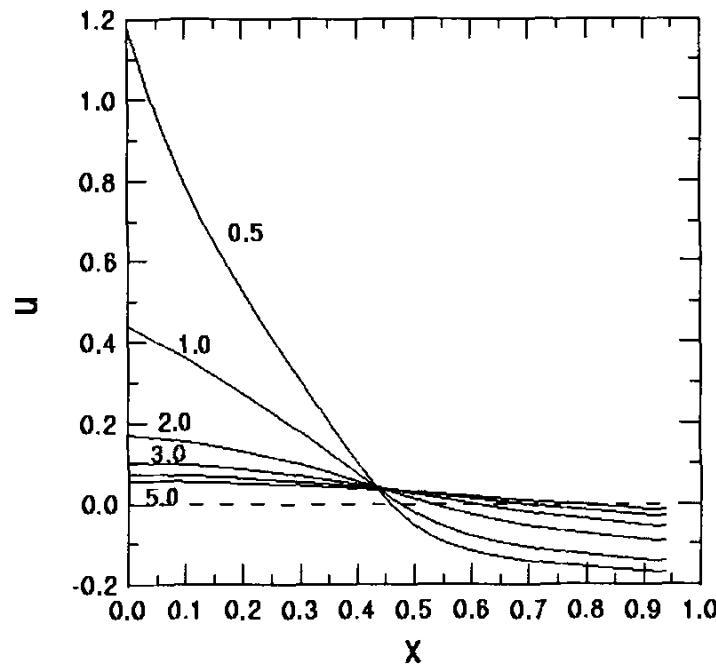


Fig. 12. The variational parameter u as function of the doping for a 2×32 ladder with $J = 0.5, t = t' = 1$ and $J' = 0.5, 1, 2, 3, 4, 5$

For $x > x_c$ the parameter u becomes (Fig. 12) negative and the definition (4) makes no longer sense since h_{RVB} is supposed to be positive. The physical interpretation of the overdoped region comes from the solution of the Cooper problem in the t-J two-legged ladder, and its BCS extension. It can be shown analytically that two electrons in the latter system form a bound state only under certain conditions (details will be given elsewhere). For $J = 0.5, t =$

$t' = 1$ one must have $J' > 3.3048$, (the binding of two electrons in the t-J chain requires $J/2t > 1$ [44]). The exact solution for four or more electrons is difficult to construct, but we expect it to be given essentially by a Gutzwiller projected BCS-like wave function with no double occupied sites. A short range version of the latter type of wave function can be generated from the recursion relation (40), with u a negative parameter, which can be written as,

$$u(x) = -h_{\text{BCS}}^2(x) < 0, \quad (x > x_c) \quad (46)$$

where h_{BCS} is the BCS amplitude for two electrons at distance 1 along the legs. Of course the interpretation of u as the square of a BCS amplitude requires it to be negative. As we put more electrons into the ladder the value of h_{BCS} decreases and for electron densities bigger than $1 - x_c$, the Cooper pairs start to resonate and we switch into the RVB regime.

The difference between the underdoped and overdoped regimes can be attributed to two different internal structures of the pairs [22]. In the low doping regime $x < x_c$, holes doped into the spin-liquid RVB state form pairs with an internal $d_{x^2-y^2}$ -like structure relative to the undoped system. However for $x > x_c$ one moves into the low-density limit characterized by electrons doped into an internal s -wave like symmetry. The difference between underdoped and overdoped regimes can be attributed to two different symmetries of the order parameter, which may be d -wave or s -wave type.

5 Diagonal t-J Ladders

In this section we shall show how the RVA can also be applied to lattices with shapes different from the standard square-lattice ladders studied so far. More concretely, we shall apply the RVA to a recently introduced class of ladders [23] in the context of strongly coupled electron systems, which are called *diagonal ladders* due to their particular diagonal structure as shown in Fig. 13.

There is a theoretical motivation to study diagonal ladders. They provide another route for going from 1D to 2D, as was one main goal to study ladders in Sec.2. Diagonal ladders are labelled by an integer $n_p = 1, 2, \dots$ which gives the number of elementary plaquettes needed to generate the entire structure (Fig. 13). The first member of this family, i.e. $n_p = 1$, is also known as the *necklace ladder* and it consists of a collection of N plaquettes joined along a common diagonal. In this chapter we shall focus on the necklace ladder for simplicity, much as we did with the two-leg ladder.

The original motivation for introducing diagonal ladders in the context of the t-J model was to understand the fully doped stripes in the (1,1) direction that have been observed experimentally in materials like $\text{La}_{1-x}\text{Sr}_x\text{NiO}_4$ [49], in Hartree-Fock studies of t-J and Hubbard models, and numerically in DMRG studies of the t-J model [50]. The simplest possible toy model of

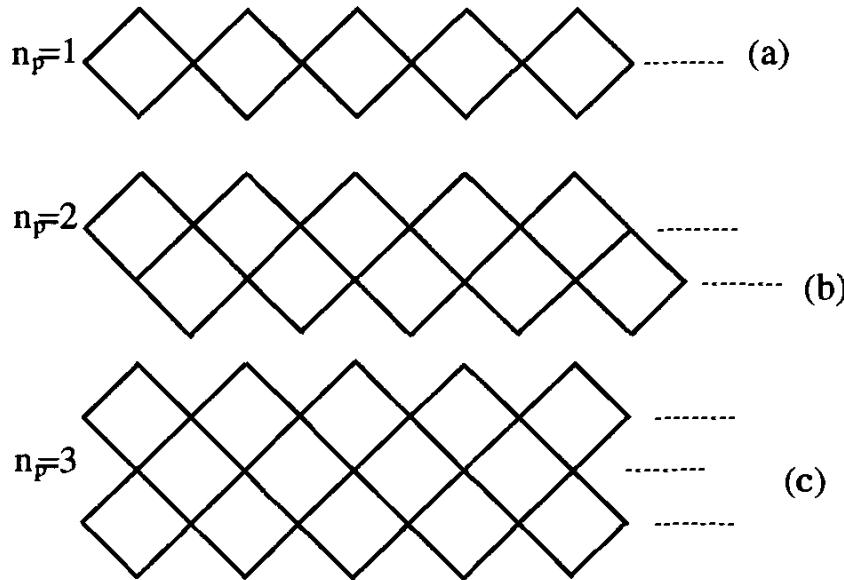


Fig. 13. Examples of diagonal ladders with a number of plaquettes $n_p = 1, 2, 3$ in the unit cell

this type of stripes is provided by a necklace ladder with a hole doping equal to $1/3$. As we shall see this doping plays an important role for the necklace ladder but we will be able to deal with doped diagonal ladders much as we did with the two-leg ladder.

Let us also point out that diagonal ladders have recently appeared as constituent parts of some interesting materials like $\text{Sr}_{0.4}\text{Ca}_{13.6}\text{Cu}_{24}\text{O}_{41.84}$ known for its superconducting properties at high pressure [39]. In addition there are a variety of materials which contain weakly coupled arrays of ladders [51,52].

From the point of view of the RVA, diagonal ladders are interesting lattices in several regards:

- In the undoped case, they provide a suitable setting to apply RVA ansätze in basis states which need not to be explicitly rotationally invariant. So far we have used basis of states in terms of VB states (singlets). In the diagonal ladder we shall use states which only have a definite quantum number, namely, the S_z -component. These states will be also used in generic doped diagonal ladders.
- Diagonal ladders also provide us with an example in which the RVA ansatz has to be set up in terms of non-uniform configurations, as we shall see later on. This means that we have to work with RR's for a fixed doping x of the ladder, i.e., the RR's do not mix diagonal ladders with different dopings, unlike the square two-leg ladder. The reason for this fact is the existence of a dominant configuration for each doping value which contribute the most to the variational method.

This RVA can be implemented analytically if the variational ansatz is sufficiently simple. Below we shall propose various RVA states for the necklace ladder with dopings $0 \leq x \leq 1/3$.

For a more detailed account on theoretical considerations for diagonal ladders and DMRG results, see [23].

The t-J model on the necklace ladder: It is useful to write the t-J model on the necklace ladder in a similar fashion as the AF-Heisenberg chain, namely, using the diagonal coordinates as shown in Fig. 14,

$$H = J \sum_{n=1}^N (\mathbf{S}_{1,n} + \mathbf{S}_{2,n}) \cdot (\mathbf{S}_{3,n} + \mathbf{S}_{3,n-1}) - \mathcal{P}_G \left(t \sum_{n=1,s}^N ((c_{1,n;s}^\dagger + c_{2,n;s}^\dagger) c_{3,n;s} + h.c.) \right) \mathcal{P}_G \quad (47)$$

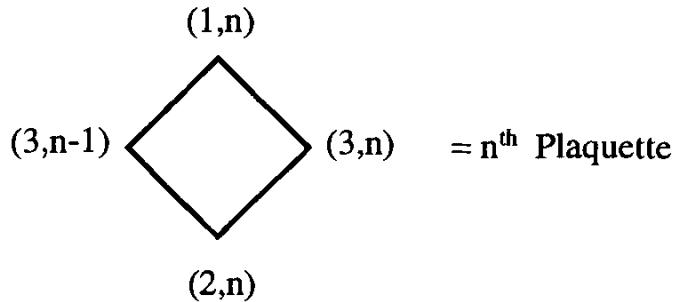


Fig. 14. Diagonal coordinates for a single plaquette

For the necklace t-J ladder, there is a parity plaquette conservation theorem [56]: the t-J Hamiltonian on a necklace ladder commutes with every permutation operator P_n associated with the minor diagonal of the n^{th} plaquette defined by the sites $(1, n)$ and $(2, n)$ in Fig. 14. Here the permutation operator P_n ($n = 1, \dots, N$) is defined by its action on the fermionic operators, which is trivial on all the sites except for those on the minor diagonal of the n^{th} plaquette where it acts as

$$P_n c_{(1,n),s} P_n^\dagger = c_{(2,n),s} , P_n c_{(2,n),s} P_n^\dagger = c_{(1,n),s} \quad (48)$$

Of course, the spin and the density-number operators at the sites $(1, n)$ and $(2, n)$ are also interchanged under the action of P_n . The above theorem is the statement that P_n commutes with H , (47), for all n ,

$$[H, P_n] = 0 \text{ for } n = 1, \dots, N \quad (49)$$

and can be easily proved. Equation (49) is not special to the t-J Hamiltonian, since any other lattice Hamiltonian having the permutation symmetry

between the two sites on the minor diagonal of every plaquette would share this same property.

The immediate consequence of (49) is that we can simultaneously diagonalize the Hamiltonian $H_{t,J}$ and the whole collection of permutations operators P_n , whose possible eigenvalues are given by $\epsilon_n = \pm 1$. The latter fact is a consequence of the following equation,

$$P_n^2 = 1 \quad (50)$$

Letting ϵ_n denote the parity of the n^{th} plaquette, the 9 possible states associated with the minor diagonal of a plaquette can be classified according to their parity, i.e. $\epsilon_n = 1$ for even parity states and $\epsilon_n = -1$ for odd-parity states (see Table 3).

Table 5. Parity of states in a minor diagonal

	State	ϵ
2 holes	$ 0\rangle$	1
bonding	$(c_{1,s}^\dagger + c_{2,s}^\dagger) 0\rangle$	1
singlet	$\Delta_{1,2}^\dagger 0\rangle$	1
antibonding	$(c_{1,s}^\dagger - c_{2,s}^\dagger) 0\rangle$	-1
triplet	$(c_{1,s}^\dagger c_{2,s'}^\dagger + s \leftrightarrow s') 0\rangle$	-1

The Hilbert space $\mathcal{H}_{\text{necklace}}$ of the t-J model can be split into a direct sum of subspaces \mathcal{H}_ϵ classified by the parity of their plaquettes, $\epsilon = \{\epsilon_n\}_{n=1}^N$, namely

$$\mathcal{H}_{\text{necklace}} = \bigoplus_\epsilon \mathcal{H}_\epsilon \quad (51)$$

Thus we can look for the ground state in the sector with N plaquettes, h holes, spin S , and parity ϵ . This seems to be a formidable task. It is more reasonable to look for the absolute minimum of the g.s. energy, keeping fixed the values of the number of plaquettes N , the number of holes h and the ratio J/t , i.e.

$$E_0(N, h, S_{\min}, \epsilon_{\min}, J/t) \leq E_0(N, h, S, \epsilon, J/t), \quad \forall S, \epsilon, \quad (52)$$

Even this question is not easy to answer with full generality. However we shall present the most representative cases.

Undoped case ($x = 0$)

A DMRG calculation shows that the g.s. of this spin ladder is ferrimagnetic as shown in the DMRG snapshot of Fig. 15 and that the parity of all plaquettes is $\epsilon = -1$. One can show that the spin necklace ladder is equivalent to an

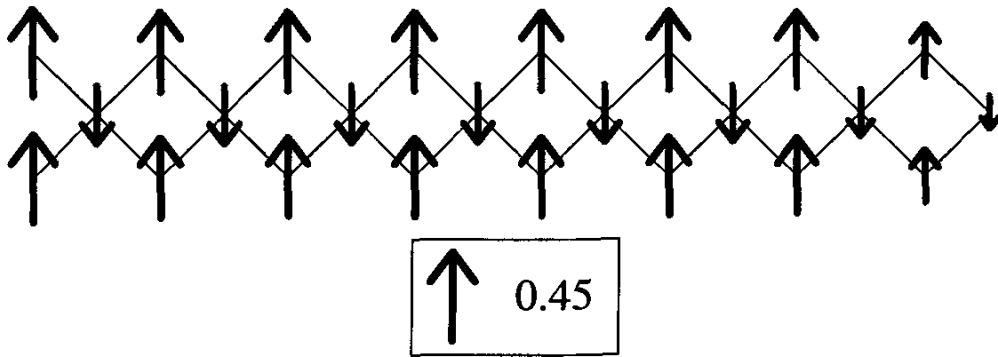


Fig. 15. DMRG results for the spin configuration for the g.s. of an undoped 8×3 necklace ladder. The state has total spin $s_G = 4$. The length of the arrow is proportional to $\langle S_z \rangle$, according the scale in the box

alternating spin-1/spin- $\frac{1}{2}$ chain, which is also known to be ferrimagnetic [53–55]. Having $\epsilon = -1$ for every plaquette amounts to having a spin 1 on every minor diagonal (see Table 5).

The most important configuration of the g.s. at $x = 0$ is the Néel state which is slightly perturbed by quantum fluctuations. This can be taken care of by the RVA state generated by the RR's given in Fig. 16.

$$\begin{aligned}
 \diamond_{2N+1} &= \diamond_{2N} \downarrow + u \diamond_{2N-1} (\downarrow \uparrow + \uparrow \downarrow) 1/\sqrt{2} \\
 &\quad + v \diamond_{2N-2} \uparrow \downarrow \uparrow \\
 \diamond_{2N} &= \diamond_{2N-1} \uparrow \uparrow + u \diamond_{2N-2} (\uparrow \downarrow + \uparrow \uparrow) 1/\sqrt{2}
 \end{aligned}$$

Fig. 16. Diagrammatic representation of the recurrent relations generating the ground state of an undoped necklace ladder using the variational RVA

Minimizing the g.s. energy of this RVA state in the limit $N \gg 1$ we find that its value per plaquette is given by $-0.4822 J$ [23], which corresponds to an energy per site of the associated alternating spin chain equal to $-0.7233 J$. The values of the variational parameters are given by $u = -0.3288$ and $v = 0.1691$. This RVA energy is to be compared with the extrapolated DMRG results $-0.72704J$ or the spin-wave value $-0.718J$ of Pati et. al. [53].

We also find with the RVA ansatz that the expectation values of the spins near the center of the system are given by $\langle S_{1,n}^z \rangle = \langle S_{2,n}^z \rangle = 0.396$ and $\langle S_{3,n}^z \rangle = -0.292$ in agreement with the results of [53] namely $\langle S_{1,n}^z \rangle = 0.39624$ and $\langle S_{3,n}^z \rangle = -0.29248$.

The case $x = 1/3$

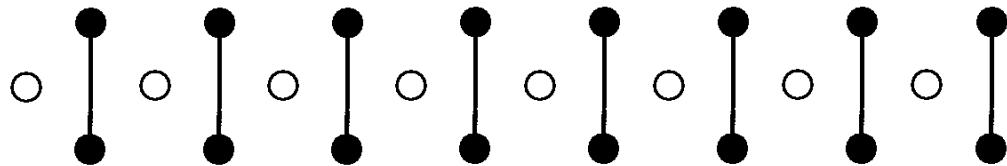


Fig. 17. Pictorial representation of the most probable configuration for doped $x = 1/3$ necklace ladder. Blank circles denote holes and vertical solid lines represent valence bond states

The DMRG shows that for a doping $x = 1/3$ the most important configuration is the one where the holes occupy the main diagonal of the necklace ladder while the minor diagonals are occupied by singlets (Fig. 17). The parity of all the plaquettes is $\epsilon = +1$. This state is a sort of diagonal stripe [23]. Of course, there are also quantum fluctuations which can be dealt with in a RVA fashion. The corresponding RR's that generate a $x = 1/3$ state are given in Fig. 18.

$$\begin{aligned}
 \diamondsuit_{2N+1} &= \diamondsuit_{2N} \circ + a \diamondsuit_{2N-1} \left(\begin{array}{c} \circ \\ \diagup + \diagdown \end{array} \right) 1/\sqrt{2} \\
 &+ b \diamondsuit_{2N-2} \begin{array}{c} \circ \\ \hline \circ \end{array} \\
 \diamondsuit_{2N} &= \diamondsuit_{2N-1} | + a \diamondsuit_{2N-2} \left(\begin{array}{c} \circ \\ \diagup + \diagdown \end{array} \right) 1/\sqrt{2}
 \end{aligned}$$

Fig. 18. A pictorial representation of the recurrence relations employed with the RVA to construct variational g.s. states for the doped $x = 1/3$ necklace ladder. The diagonal squares represent bulk states of a given length. Small circles represent holes and solid lines valence bonds

For a necklace with 7 plaquettes ($J = 0.35, t = 1$) the RVA g.s. energy is -16.33996 while the DMRG is -16.554153. We are using a RVA where there 3 variational parameters per plaquette and hence the ansatz is non uniform [23].

We have plotted in Fig. 20 the results for $x = 1/3$ in the necklace ladder and compared it with the DMRG results showing a good agreement.

Intermediate dopings ($0 \leq x \leq 1/3$)

From Figs. 16 and 18 we observe a similarity in the RR's that generate the RVA states for dopings $x = 0$ and $x = 1/3$. We have seen that for intermediate dopings the g.s. produced by the DMRG contains a dominant state $|\psi_0\rangle$ together with its local fluctuations [23]. The “classical state” $|\psi_0\rangle$ can be generated by a first order RR while its “quantum fluctuations” yield up to a third order RR. The skeleton of these RR's is depicted in Fig. 19 . Observe that there are 3 independent variational parameters a_N, b_N, c_N per plaquette.

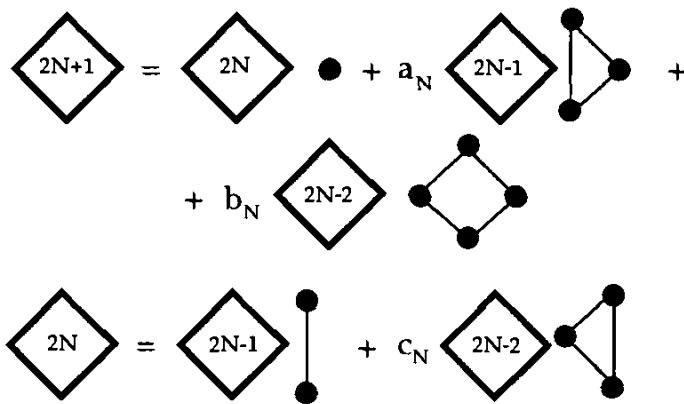


Fig. 19. A pictorial representation of the recursion relations employed to generate RVA states for the under-doped ($0 \leq x \leq 1/3$) necklace ladder. Here a, b and c are local variational parameters. A square denotes a bulk state on a ladder of a length given by its number inside. The black circles and solid lines represent generic fluctuation states

The procedure to find the ground state of the necklace ladder is summarized as follows,

- Fix the length N of the ladder, the number of holes h and the total S^z .
- Generate all the $|\psi_0\rangle$ configurations with those quantum numbers N, h, S^z . Generically the number of configurations grows exponentially. For example the seven-plaquette case has $N=15$ and a total of 3×10^4 configurations.
- Compute the energy of the state associated with the zero-order state $|\psi_0\rangle$ using the recursion relations and find the variational parameters which lead to a minimum energy. For the seven-plaquette ladder there are 21 independent variational parameters.
- Extract the state $|\psi_0\rangle$ which has the absolute minimum energy for a given N, h and S^z .

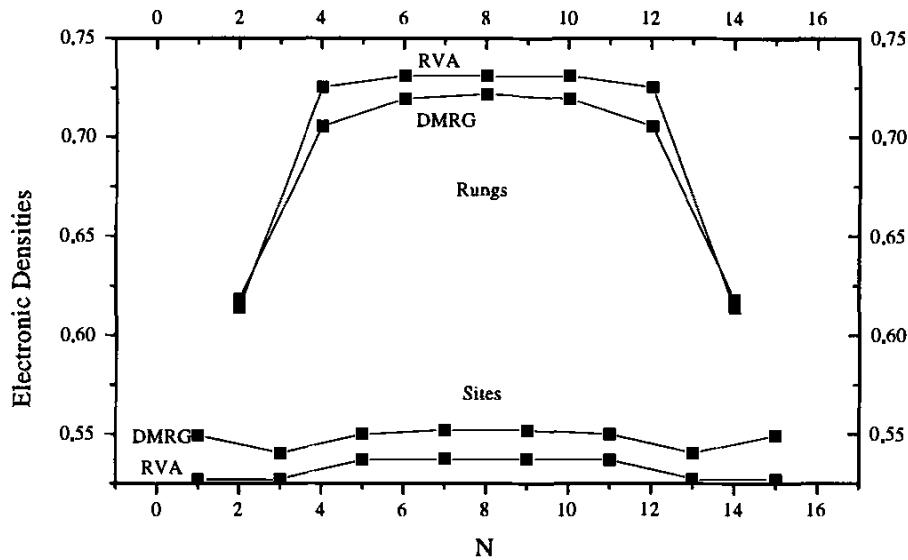


Fig. 20. Electronic densities for a necklace ladder. It corresponds to 7 plaquettes and 8 holes (doping $x = 1/3$). Below are shown the DMRG and RVA values for the sites in the main diagonal, while above are shown the values on the rungs corresponding to subdiagonals

In Table 6 we present some of the results obtained by this RVA and compare with DMRG results [23] finding a very good agreement.

Table 6. DMRG and RVA total energies for a 7-plaquette necklace ladder with h holes and total spin S_z . The string of epsilon is the pattern of parities in the subdiagonals (rungs).

h	S	$(\epsilon_1, \dots, \epsilon_7)$	E_0^{DMRG}	E_0^{RVA}
8	0	(+ + + + + + +)	-16.554153	-16.33996
8	1	(+ + + - + + +)	-16.284855	-16.00358
7	1/2	(+ + + + + + -)	-15.489511	-15.18141
6	0	(- + + + + + -)	-14.424805	-14.02286
6	1	(- + + + + + -)	-14.424798	-14.02286

Let us comment on the DMRG results shown in the following figures. In what follows, (h, S) denotes a system with h holes and total spin S as in Table 6.

- *Case (8,1):* The spin excitation of the $x = 1/3$ state is given by a spin-1 magnon strongly localized on an odd-parity plaquette located at the center of the ladder. The other plaquettes remain even and spinless. The value of the spin gap is given by $\Delta_s = 0.27$ (DMRG) and 0.32 (RVA).

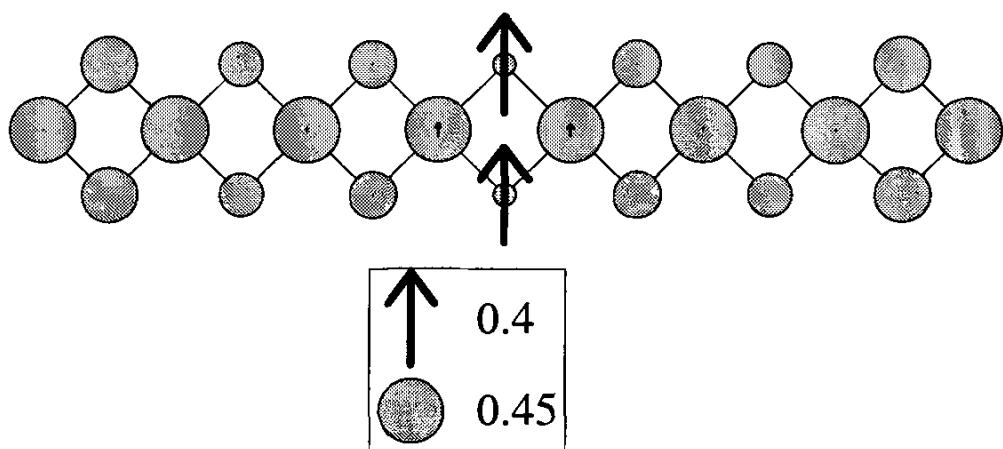


Fig. 21. Results using DMRG showing the necklace state with 8 holes and spin $S_z = 1$. The diameter of the circles are proportional to the hole density, and the length of the arrows are porportional to $\langle S_z \rangle$, according to the scale in the box

- *Case (7,1/2)*: This case is obtained by doping the $x = 1/3$ state with an electron. The additional electron goes into anyone of the boundary plaquettes. The corresponding plaquette changes its parity to -1 .

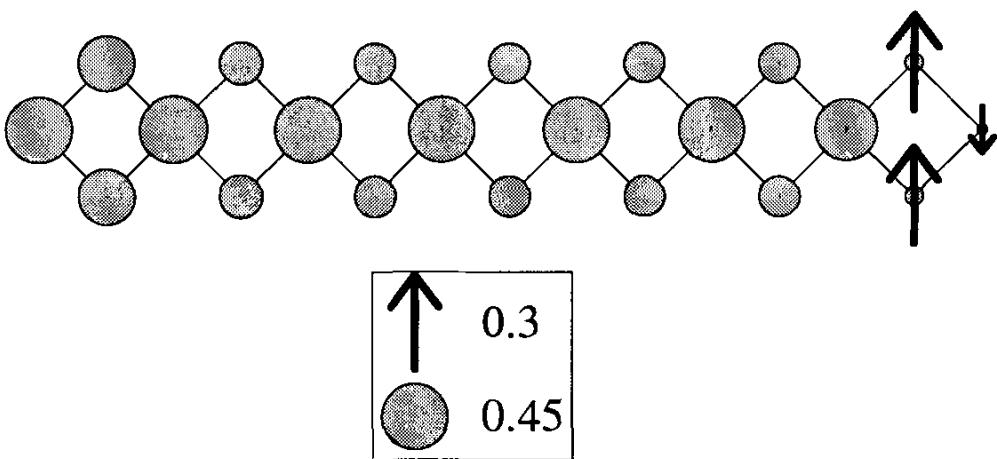


Fig. 22. DMRG results for the hole and spin densities of the necklace state with 7 holes and spin $S_z = 1/2$

- *Cases (6,0) and (6,1)*: The state $x = 1/3$ is now doped with two electrons, which go to the boundary plaquettes, changing their parity. There seems to be a small effective coupling between the two spin $1/2$ at the ends of the ladder, which leads to a breaking of the degeneracy between the triplet and the singlet. This is reminiscent of the effective spin $1/2$ at the ends of the Haldane and AKLT open spin chains [63]. There also exists a weak effective coupling that breaks the four-fold degeneracy of the open chains.

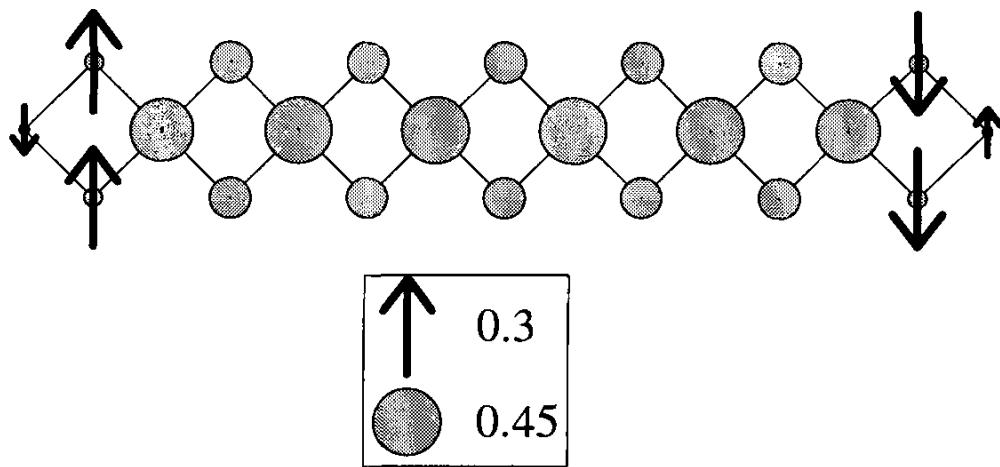


Fig. 23. DMRG results for the hole and spin densities of the necklace state with 6 holes and spin $S_z = 0$

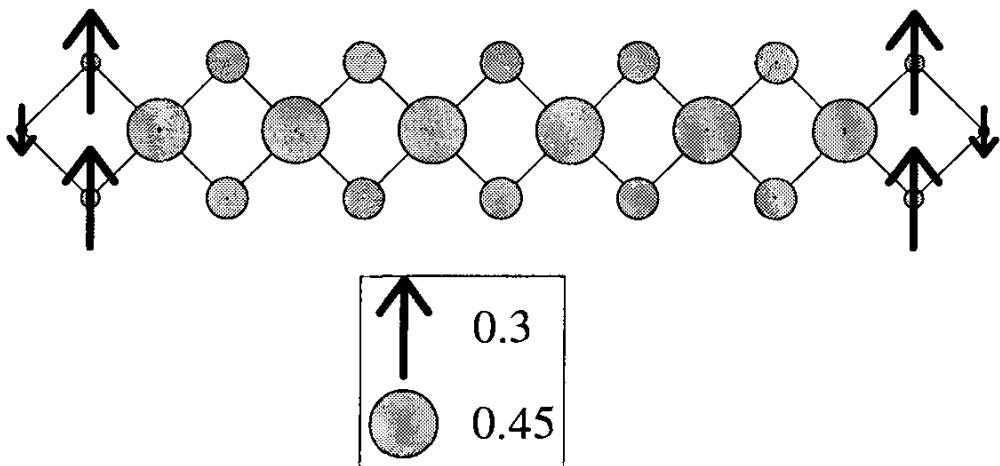


Fig. 24. Results from DMRG showing the necklace state with 6 holes and spin $S_z = 1$

6 Problems and Prospects

In this section we would like to mention some problems that may in principle be studied with the RVA.

Hubbard and polymer models

The application of the RVA goes beyond the specific models such as the Heisenberg and t-J considered in this chapter. There are other models where we can apply RVA techniques. We have already mentioned the study of a two-leg Hubbard ladder in Sect. 5 [57]. Similarly, RVA can be used to study polymers [58] which can be modeled with the Pariser-Parr-Pople (PPP) Hamiltonian [59,60] and Chaps. 2.2(II) and 2.3(II). The two hopping parameters in the PPP Hamiltonian play the role of the J and J' coupling constants in the two-leg Heisenberg ladder.

Other ladders

As explained in Sect. 2 the even-legged ladders have a finite correlation length and a spin gap and hence are good systems to apply the RVA or the MP method. The techniques developed for the 2 leg ladder can be easily extended to the 4 leg ladder. The dimer-RVB state can be generated again by a second-order RR but one needs more than one singlet state associated to the different shapes of the boundary of the four-leg ladder. The number of dimer states approaches from below the exponential behaviour found by Kasteleyn for the 2D dimers.

Stripes

Tranquada et al. have shown experimentally that in certain cuprates the charge degrees of freedom form quasi-1D structures called stripes along the (1,0) direction of the planes [49]. This happens when the doping takes a value $x = 1/8$. Motivated by these results, White and Scalapino have studied large clusters of spins and holes (t-J model) using the DMRG and shown that the (1,0) stripes are essentially two-leg ladders with doping $x = 1/4$ separated by two-leg undoped spin ladders [50].

This suggests that a simple ansatz for a stripe with doping x could be given as

$$|\text{stripe}\rangle = |A\rangle_1 |B\rangle_2 |A\rangle_3 \cdots |B\rangle_{N-1} |A\rangle_N \quad (53)$$

where $|A\rangle_i$ and $|B\rangle_i$ are two-leg ladder states located at the positions $2i - 1$ and $2i$, and dopings x_A and x_B satisfying $x = (x_A + x_B)/2$. In the example of Tranquada $x = 1/8$, $x_A = 1/4$, $x_B = 0$. It is clear that the state (53) can be generated by a first-order RR. To go beyond (53) one should consider the quantum fluctuations of the neighbour ladders which can be described by a RVA state generated by the following second order RR

$$\begin{aligned} |2N+1\rangle &= |2N\rangle |A\rangle + |2N-1\rangle |(BA)\rangle \\ |2N\rangle &= |2N-1\rangle |B\rangle + |2N-2\rangle |(AB)\rangle \end{aligned} \quad (54)$$

where $|N\rangle$ denotes a state made of N two-leg ladders, $|(AB)\rangle$ is a four-leg ladder state which results from the quantum fluctuations of two parallel two-leg ladders $|A\rangle|B\rangle$ (same for $|(BA)\rangle$).

Equation (54) is similar to the RR that generates the dimer-RVB state of the two-leg ladder with the replacement: rung \rightarrow stripe. The states $|A\rangle$, $|B\rangle$, $|(AB)\rangle$ and $|(BA)\rangle$ can be constructed using again RR's. Hence one can generate the g.s. of the plane in two steps: first building the two-leg and four-leg ladder states and next joining them to get a 2D state. The minimization of the g.s. of the whole system will couple the variational parameters involved in the construction of $|A\rangle$, $|B\rangle$, $|(AB)\rangle$ and $|(BA)\rangle$.

2D Matrix-product ansätze

The construction outlined above yields a possible way to build 2D states combining 1D algorithms in the x and y directions of the plane and seems to be appropriate to study stripe phases. However it is not clear that such an algorithm would be able to deal with genuine 2D systems. The problem is related to finding a truly 2D version of the DMRG ([61] and T. Nishino in Chap. 5(I)). The actual DMRG algorithm is essentially one-dimensional and it loses its accuracy for large clusters [62].

The connection between the DMRG and the RVA and MP ansätze may be here of some help. Rather than asking what is the 2D version of the DMRG we may ask whether there are 2D versions of the RVA or the MP. The answer to the latter question is positive since from the work of Affleck et al. we know that there are AKLT states in 2D [63]. The AKLT states can be seen as a simple type of MP states. In fact, Niggemann et al. have constructed a vertex-state model for a spin-3/2 hexagonal lattice which is an example of these 2D MP states [64]. Below we generalize this construction to square lattices.

In 2D there are two types of MP states which can be called, using Statistical-Mechanics terminology, vertex-MP and face-MP ansätze [65]. A vertex-MP ansatz is defined in terms of a set of amplitudes

$$A_{\alpha,\beta}^{\gamma,\delta}[s], \quad (\alpha, \beta, \dots = 1, \dots, m; \quad s = 1, \dots, m_s) \quad (55)$$

where the labels $\alpha, \beta, \gamma, \delta$ are associated with the links of the square lattice while s labels the quantum state, e.g. spin, associated to the vertex where the 4-links $\alpha, \beta, \gamma, \delta$ meet. $A_{\alpha,\beta}^{\gamma,\delta}[s]$ is a sort of Boltzmann weight of a vertex model. The vertex-MP wave function $\psi(s_1, s_2, \dots, s_N)$ can be obtained by multiplying all the Boltzmann weights $A_{\alpha_i,\beta_i}^{\gamma_i,\delta_i}[s_i]$ and contracting and summing over the link variables according to the same pattern of a vertex model in Statistical Mechanics [66]. Hence the value of the wave function $\psi(s_1, s_2, \dots, s_N)$ is given by the partition function of a vertex model where the Boltzmann weights depend on the value of the local states s_i . Moreover the norm $\langle \psi | \psi \rangle$ of the vertex-MP state is given by the partition function of another vertex model whose Boltzmann weights are defined as,

$$R_{\alpha\alpha',\beta\beta'}^{\gamma\gamma',\delta\delta'} = \sum_s A_{\alpha,\beta}^{\gamma,\delta}[s] A_{\alpha',\beta'}^{\gamma',\delta'}[s] \quad (56)$$

This R matrix is the 2D version of the T matrix defined in (25).

The face-MP models can be defined in a similar manner by a set of variational parameters as in (55) where now the variables α, β, \dots are now associated to the vertices of the squares while the quantum variable s is associated to the face whose vertices are $\alpha, \beta, \gamma, \delta$. This is similar to the face or Interaction Round a Face models (IRF) in Statistical Mechanics [66].

Hence in 2D there are two generic ways to produce MP ansätze which are in fact the straightforward generalization of the 1D MP ansätze. These two

generalizations suggest to use some well-know models such as the 6-vertex model to test some of the ideas presented above.

We would like first of all to thank J. Dukelsky, T. Nishino, S. R. White, D.J. Scalapino and J.M. Román for their collaboration and many discussions on the several aspects related to the common work presented in these lectures.

We would like to thank the organizers of the DMRG98 at the Max-Planck-Institut in Dresden for their kind invitation to lecture in this meeting, for their warm hospitality and for making these lecture notes come out, as well as Prof. P.Fulde for supporting the DMRG98 at the MPI Dresden.

We also acknowledge many interesting discussions during this meeting with S. White, R. Noack, K. Hallberg, T. Nishino, S. Pleutin, X. Wang, I. Peschel, S. Ramasesha, S. Rommer, T. Xiang.

We acknowledge support from the DGES under contract PB96-0906.

References

1. S.R. White, Phys. Rev. Lett. **69**, 2863 (1992)
2. S.R. White, Phys. Rev. B **48**, 10345 (1993)
3. S.R. White and R.M. Noack, Phys. Rev. Lett. **68**, 3487 (1992)
4. S.R. White and D.A. Huse Phys. Rev. B **48**, 3844 (1993)
5. R.M. Noack, S.R. White and D.J. Scalapino Phys. Rev. Lett. **73**, 882 (1994)
6. As a proof of the relevance acquired by this method in the recent years, it is impossible to fully account for all the existing bibliography on the DMRG, and the reader is referred to search the cond-mat archive for a complete documentation on the subject.
7. M.A. Martín-Delgado and G. Sierra, Int. J. Mod. Phys. A **11**, 3145 (1996)
8. M.A. Martín-Delgado and G. Sierra, Phys. Lett. B **364**, 41 (1995)
9. M.A. Martín-Delgado, J. Rodríguez-Laguna and G. Sierra, Nucl. Phys. B **473**, 685 (1996)
10. M.A. Martín-Delgado and G. Sierra, Phys. Rev. Lett. **76**, 1146 (1996)
11. M.A. Martín-Delgado and G. Sierra, in From Field Theory to Quantum Groups, World Scientific Publishers (1996)
12. G. Sierra and T. Nishino, cond-mat/9610221; Nucl. Phys. B **495**, 505 (1997)
13. T. Nishino, J. Phys. Soc. Jpn. **64**, 3958 (1995)
14. T. Nishino and K. Okunishi, J. Phys. Soc. Jpn. **65**, 891-894 (1996); cond-mat/9507087
15. S. Östlund and S. Rommer, Phys. Rev. Lett. **75**, 3537 (1996); Phys. Rev. B **55**, 2164 (1997)
16. T. Nishino and K. Okunishi. J. Phys. Soc. Jpn. **64**, 4084 (1995); cond-mat/9510004
17. S.R. White, Phys. Rep. **301**, 187 (1998)
18. J. González, M.A. Martín-Delgado, G. Sierra and A.H. Vozmediano, Quantum Electron Liquids and High-T_c Superconductivity, Chapter 11, Lecture Notes in Physics, Monographs, Springer **38** (1995)
19. T. Nishino and K. Okunishi, in Strongly Correlated Magnetic and Superconducting Systems, Lecture Notes in Physics, Springer, **478** (1997)

20. M.A. Martín-Delgado, in Strongly Correlated Magnetic and Superconducting Systems, Lecture Notes in Physics, Springer, **478** (1997)
21. G. Sierra and M.A. Martín-Delgado, Phys. Rev. B **56**, 8774 (1997)
22. G. Sierra, M.A. Martín-Delgado, J. Dukelsky, S. R. White and D. J. Scalapino, Phys. Rev. B **57**, 11666 (1998)
23. G. Sierra, M.A. Martín-Delgado, S.R. White, D.J. Scalapino, J. Dukelsky, cond-mat/9806251, to be published in Phys. Rev. B
24. J. Dukelsky, M.A. Martín-Delgado, T. Nishino, G. Sierra. Europhys. Lett. **43**, 457 1998
25. J.M. Román, G. Sierra, J. Dukelsky and M.A. Martín-Delgado: cond-mat/9802150; J. Phys. A: Math. Gen. **31**, 9729 (1998)
26. E. Dagotto and T M. Rice, Science **271**, 619 (1996)
27. I. Peschel, M. Kaulke and Ö. Legeza, Ann. Physik (Leipzig) **8**, 153 (1999); cond-mat/9810174
28. K. Okunishi, Y. Hieida and Y. Akutsu, cond-mat/9810239
29. M. Andersson, M. Boman and S. Östlund, cond-mat/9810093
30. P.W. Anderson, Science **235**, 1196 (1987)
31. S. R. White, R. M. Noack and D. J. Scalapino, Phys. Rev. Lett. **73**, 886 (1994)
32. G. Andrews, The Theory of Partitions, Encyclopedia of Mathematics and its Applications, Addison-Wesley **2** (1976)
33. B. Sutherland, Phys. Rev. B **37**, 3786 (1988)
34. S. Gopalan, T.M. Rice and M. Sigrist, Phys. Rev. B **49**, 8901 (1994)
35. T. Barnes, E. Dagotto, J. Riera and E. S. Swanson, Phys. Rev. B **47**, 3196 (1993)
36. G. Fano, F. Ortolani and L. Ziosi, Phys. Rev. B **54**, 17557 (1996)
37. M. Fannes, B. Nachtergaele and R.F. Werner, Commun. Math. Phys. **144**, 443 (1992)
38. H. Takasaki, T. Hikihara and T. Nishino, cond-mat/9810241
39. M. Uehara , T. Nagata, J. Akimitsu, H. Takahashi, N. Mori and K. Kinoshita, J. Phys. Jpn, **65**, 2764 (1997)
40. W. Marshall, Proc. Roy. Soc. A **232**, 48 (1955)
41. S. Liang, B. Doucot and P.W. Anderson, Phys. Rev. Lett. **50**, 1153 (1988)
42. W. Marshall, Proc. Roy. Soc. A **232**, 48 (1955)
43. M. Troyer, H. Tsunetsugu and T. M. Rice, Phys. Rev. B **53**, 251 (1996)
44. M. Ogata, M.U. Luchini, S. Sorella and F.F. Assaad, Phys. Rev. Lett. **66**, 2388 (1991)
45. S.R. White and D.J. Scalapino, Phys. Rev. B **55**, 6504 (1997)
46. The IMSL Fortran-77 Library. Visual Numerics Inc.
47. The Numerical Recipes in C. W. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Cambridge University Press (1993)
48. A. Klümper, A. Schadschneider and J. Zittartz, Europhys. Lett. **24**, 293 (1993)
49. J.M. Tranquada et al. Nature **375**, 561 (1995); Phys. Rev. B **54**, 7489 (1996)
50. S.R. White and D.J. Scalapino, Phys. Rev. Lett. **80**, 1272 (1998); cond-mat/9801274
51. S.A. Carter, B. Battlogg, R.J. Cava, J.J. Krajewski, W.F. Peck, Jr. and T.M. Rice, Phys. Rev. Lett. **77**, 1378 (1996)
52. M. Matsuda, K.M. Kojima, Y.J. Uemura, J.L. Zarestky, K. Nakajima, K. Kakurai, T. Yokoo, S.M. Shapiro and G. Shirane, Phys. Rev. B **57**, 11467 (1998)

53. S.K. Pati, S. Ramasesha and D. Sen, Phys. Rev. B **55**, 8894 (1997); cond-mat/9704057
54. A.K. Kolezhuk, H.J. Mikeska and S. Yamamoto, Phys. Rev. B **55**, 3336 (1997)
55. S. Yamamoto, S. Brehmer and H.J. Mikeska, Phys. Rev. B **57**, 13610 (1998); S. Brehmer, H.J. Mikeska and S. Yamamoto, J. Phys. Cond. Matt. **9**, 3921 (1997)
56. A. Giesekus, Phys. Rev. B **52**, 2476 (1995)
57. E. Kim, private communication.
58. S. Pleutin, private communication.
59. S. Pleutin and J.L. Fave, J. of Phys. : Condens. Matter **10**, 3941 (1998)
60. M.A. Garcia-Bach, R. Valenti, D.J. Klein, cond-mat/9705121 and references therein; Phys. Rev. B in press
61. H. Otsuka Phys. Rev. B **53**, 14004 (1996)
62. S. Liang and H. Pang, Phys. Rev. B **49**, 9214 (1994)
63. I. Affleck, T. Kennedy, E.H. Lieb and H. Tasaki, Commun. Math. Phys. **115**, 477 (1988)
64. H. Niggemann, J. Zittarz, Z. Phys. B **101**, 289 (1996)
H. Niggemann, A. Klümper, J. Zittarz, Z. Phys. B **104**, 103 (1997)
65. G. Sierra and M.A. Martín-Delgado, Proceedings of the Workshop on the Recent Developments in Exact RG Methods, Faro, Portugal, (1998)
66. R. J. Baxter, Exactly Solved Models in Statistical Mechanics, Academic Press (1982)

5 Transfer-Matrix Approach to Classical Systems

T. Nishino¹, K. Okunishi², Y. Hieida², T. Hikihara¹, and H. Takasaki¹

¹ Department of Physics, Graduate School of Science and Technology
Kobe University, Inoshishi-dai, Nada, Kobe 657-8501, Japan

² Department of Physics, Graduate School of Science, Osaka University
Toyonaka, Osaka 560-0043, Japan

The development of the DMRG method by White [1] is one of the major achievements in computational condensed-matter physics (see Chap. 2(I)). DMRG enables one to calculate ground states of relatively large-scale one-dimensional (1D) quantum systems. Since 1D quantum systems are deeply related to 2D classical systems [2–4], it is natural to import the DMRG method to 2D classical systems. The infinite-system algorithm was first applied to the Ising model by one of the authors [5]. Carlon *et al.* [6–8] applied the finite-system algorithm to the Potts model (Chap. 3.2(II)), and they calculated very accurate density profiles and critical indices. After that, the DMRG formulation for classical systems was applied to 1D quantum system at finite temperature [9,10,12] (Chaps. 6(I) and 4(II)).

In this review we explain the DMRG applied to 2D classical systems ('classical DMRG') by looking at the renormalization group (RG) transformation for the transfer matrix T . For the distinction, we call the DMRG for quantum systems 'quantum DMRG' in the following. We first introduce the infinite-system algorithm (Sects. 1–4). It then follows naturally that the classical DMRG is a kind of variational method that maximizes the partition function using a limited number of degrees of freedom, where the variational state is written as a product of local matrices [15] (Sect. 5). Actually, such a variational formulation has been used for a long time; we give a brief review in Sect. 6. The variational state obtained by the infinite-system algorithm can be improved systematically by use of the finite-system algorithm (Sect. 7). As a variation of the infinite-system algorithm, we introduce the corner transfer matrix (CTM) formulation of the classical DMRG, where the formulation can be generalized to higher-dimensional systems (Sect. 8). We finally discuss some remaining problems in the last section.

1 Transfer Matrix

As an example of a 2D classical system, let us consider the isotropic square-lattice Ising model [17] on a cylinder, which is defined by imposing periodic

boundary conditions in the vertical direction and open boundary conditions for the horizontal direction of an $(N \times \ell)$ spin lattice (Fig. 1) [16]. The system

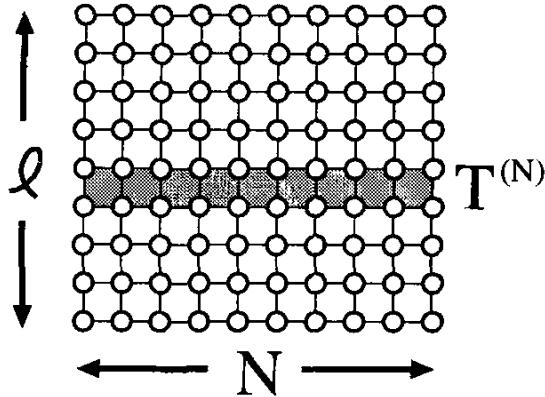


Fig. 1. Square-lattice Ising model on a N by ℓ transfer matrices $T^{(N)}$ defined in (1). Open boundary conditions are imposed in the horizontal direction and periodic ones in the vertical direction.

consists of ℓ rows of width N . We assume $\ell \gg N$, and refer to N as the system size. We label the spins in a row from left to right as $s_1, s_2, \dots, s_{N-1}, s_N$, and occasionally use the vector notation $\mathbf{s} = s_1 \dots s_N$ for simplicity. When the Ising interaction is restricted to the nearest neighbor pairs, the transfer matrix is expressed as a 2^N -dimensional matrix (Fig. 2)

$$\begin{aligned} T^{(N)}(\mathbf{s}' | \mathbf{s}) &= \exp \left\{ \frac{K}{2} \sum_{i=1}^{N-1} (s'_i s'_{i+1} + s_i s_{i+1}) + K \sum_{i=1}^N s'_i s_i \right\} \\ &= \exp \left(\frac{K}{2} s_1 s'_1 \right) \left\{ \prod_{i=1}^{N-1} W(s'_i s'_{i+1} | s_i s_{i+1}) \right\} \exp \left(\frac{K}{2} s_N s'_N \right), \end{aligned} \quad (1)$$

where $K = J/k_B T$ is the coupling and

$$W(s'_i s'_{i+1} | s_i s_{i+1}) = \exp \left\{ \frac{K}{2} (s_i s_{i+1} + s_{i+1} s'_{i+1} + s'_{i+1} s'_i + s'_i s_i) \right\} \quad (2)$$

is the local Boltzmann weight for a square between i and $i+1$. We have symmetrized the transfer matrix in order to simplify the following formulation. It may be helpful for the readers to recall the quantum-classical correspondence $T^{(N)} = \exp(-\Delta H^{(N)})$, where $H^{(N)}$ is the corresponding quantum Hamiltonian. (However, $H^{(N)}$ is normally non-local and the correspondence is more of a formal character.)

For a tutorial purpose, we only consider the symmetric transfer matrix in the following. Extensions to the asymmetric case are straightforward [5,18]; to treat an asymmetric transfer matrix is not rare in classical statistical mechanics. More details can be found in Chaps. 6(I) and 7(I) [10,11], as

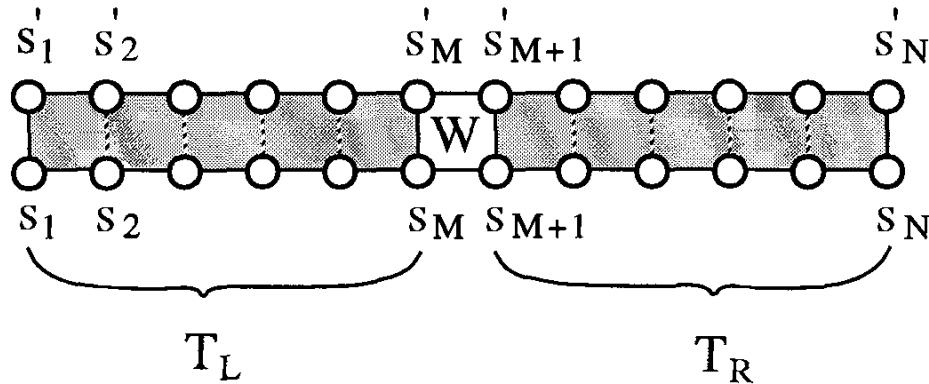


Fig. 2. Division of the transfer matrix $T^{(N)}$ into the half-row transfer matrices T_L and T_R , and a Boltzmann weight W , compare (3).

well as in Chap. 4.1(II) [12,13]. More generally, it is possible to treat various models such as

- The q -state Potts model [19,20] or the n -vector model, that have *discrete spin symmetry and short range interaction*.
- The interaction round a face (IRF) model [18] whose Boltzmann weight $W(s'_i s'_{i+1} | s_i s_{i+1})$ is represented by arbitrary matrix; *the weight can even be negative*. [14] (There is no sign problem.)
- The vertex model [18] that has spin variables in the middle of each bond.

in the framework of DMRG.

Following the convention of the quantum DMRG, let us divide the spin row into the left and the right as $s = s_L s_R$, where $s_L = s_1 \dots s_M$ and $s_R = s_{M+1} \dots s_N$. For the moment we assume that N is an even integer, and that $M = N/2$. According to the division, we can decompose $T^{(N)}$ into three factors

$$T^{(N)}(s' | s) = T_L(s'_L | s_L) W(s'_M s'_{M+1} | s_M s_{M+1}) T_R(s'_R | s_R), \quad (3)$$

where $T_L(s'_L | s_L)$ and $T_R(s'_R | s_R)$ are the half-row transfer matrices

$$\begin{aligned} T_L(s'_L | s_L) &= \exp\left(\frac{K}{2} s'_1 s_1\right) \prod_{i=1}^{M-1} W(s'_i s'_{i+1} | s_i s_{i+1}) \\ T_R(s'_R | s_R) &= \exp\left(\frac{K}{2} s'_N s_N\right) \prod_{i=M+1}^{N-1} W(s'_i s'_{i+1} | s_i s_{i+1}). \end{aligned} \quad (4)$$

2 Density Submatrix

The density matrix of the N by ℓ system (Fig. 1) is simply the ℓ -th power of the transfer matrix $\rho = (T^{(N)})^\ell$. The partition function is its trace

$$Z = \text{Tr } \rho = \text{Tr} (T^{(N)})^\ell = \sum_{\kappa} \lambda_{\kappa}^{\ell}, \quad (5)$$

where λ_{κ} is the eigenvalue of $T^{(N)}$ in the decreasing order $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$. (We have assumed that the transfer matrix is diagonalizable, and that all the eigenvalues are positive.) Since we have assumed that ℓ is sufficiently larger than N , the partition function can be approximated as $Z \simeq \lambda_1^\ell$, where the symbol ‘ \simeq ’ denotes that the ratio λ_1^ℓ/Z converges to unity in the limit $\ell/N \rightarrow \infty$. In this situation, the density matrix can be well approximated as

$$\rho \simeq |v^{(N)}\rangle \lambda_1^\ell \langle v^{(N)}| \quad (6)$$

where $|v^{(N)}\rangle$ is the eigenvector of $T^{(N)}$ that corresponds to the largest eigenvalue λ_1 . We assume the normalization $\langle v^{(N)}|v^{(N)}\rangle = 1$.

What is called ‘density matrix’ in the context of DMRG is not the density matrix itself, but the density submatrix (DSM) — also called ‘reduced density matrix’ — which is obtained by partially tracing out spin indices from ρ . The DSM for the left and the right halves of the system are defined as [21]

$$\begin{aligned} \rho_L(s'_L|s_L) &= \sum_{s_R} \rho(s'_L s_R|s_L s_R) \\ \rho_R(s'_R|s_R) &= \sum_{s_L} \rho(s_L s'_R|s_L s_R). \end{aligned} \quad (7)$$

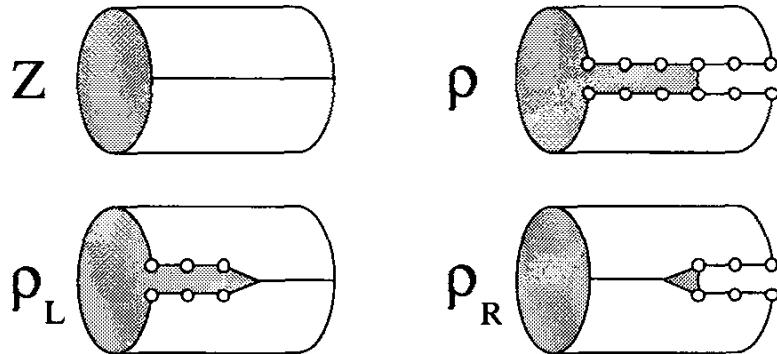


Fig. 3. ρ_L and ρ_R (7). The DSM can be interpreted as a cut in the 2D system shown in Fig. 1.

Figure 3 shows the graphical representation of the DSM, where ρ_L and ρ_R correspond to the cuts on the cylindrical system shown in Fig. 1 [22]. To take the trace of the DSM one has to join the system along the cut and to

reconstruct the cylinder: $Z = \text{Tr } \rho_L = \text{Tr } \rho_R$. Using the eigenvector $v^{(N)}$ in (6), we can express the DSM in the more familiar form

$$\begin{aligned} \rho_L(s'_L|s_L) &\simeq \sum_{s_R} v^{(N)}(s'_L s_R) \lambda_1^\ell v^{(N)}(s_L s_R) \\ \rho_R(s'_R|s_R) &\simeq \sum_{s_L} v^{(N)}(s_L s'_R) \lambda_1^\ell v^{(N)}(s_L s_R) \end{aligned} \quad (8)$$

that has been used for the quantum DMRG.

As for the quantum DMRG, the diagonalization of the DSM is one of the important steps for the classical DMRG. The DSM ρ_L and ρ_R can be diagonalized as

$$\rho_L = \sum_\xi \mathbf{a}_\xi \omega_\xi^2 \mathbf{a}_\xi^T, \quad \rho_R = \sum_\zeta \mathbf{b}_\zeta \mu_\zeta^2 \mathbf{b}_\zeta^T, \quad (9)$$

where \mathbf{a}_ξ and \mathbf{b}_ζ are the eigenvectors of ρ_L and ρ_R , respectively. The eigenvectors satisfy the orthogonality relations $\mathbf{a}_\xi^T \mathbf{a}_{\xi'} = \delta_{\xi\xi'}$ and $\mathbf{b}_\zeta^T \mathbf{b}_{\zeta'} = \delta_{\zeta\zeta'}$, and they are complete

$$I_L = \sum_\xi \mathbf{a}_\xi \mathbf{a}_\xi^T, \quad I_R = \sum_\zeta \mathbf{b}_\zeta \mathbf{b}_\zeta^T, \quad (10)$$

where I_L and I_R are 2^M - and 2^{N-M} -dimensional unit matrices, respectively; $I_L = I_R$ when $N = 2M$. It is convenient to use the matrix notation $A = (\mathbf{a}_1, \mathbf{a}_2, \dots)$ and $B = (\mathbf{b}_1, \mathbf{b}_2, \dots)$ when we explicitly show the spin indices. For example, (9) can be written as

$$\begin{aligned} \rho_L(s'_L|s_L) &= \sum_\xi A(s'_L|\xi) \omega_\xi^2 A(s_L|\xi) \\ \rho_R(s'_R|s_R) &= \sum_\zeta B(s'_R|\zeta) \mu_\zeta^2 B(s_R|\zeta). \end{aligned} \quad (11)$$

In (9) and (11) we have expressed the eigenvalues of ρ_L and ρ_R as squares of real numbers, because normally they are non-negative [23]. As to the eigenvalues of ρ , we assume decreasing order for both ω_ξ^2 and μ_ζ^2

$$\omega_1^2 \geq \omega_2^2 \geq \dots \geq 0, \quad \mu_1^2 \geq \mu_2^2 \geq \dots \geq 0, \quad (12)$$

in the following. In general, $\omega_i = \mu_i$ holds for $1 \leq i \leq \min(2^M, 2^{N-M})$ in the limit $\ell/N \rightarrow \infty$.

3 Renormalization-Group Transformation

Unlike the density matrix ρ , not only the largest eigenvalue is dominant in (12). This is because s'_L and s_L in $\rho_L(s'_L|s_L)$ are correlated through the

junction s_R (7)-(8). It has been known that the eigenvalues ω_i^2 (or μ_i^2) in (12) decay nearly exponentially when the correlation length of the system is finite [18,24,25] (Chap. 3.1(II)), and that the decay is rapid when the correlation length is short. Even at the critical temperature, we expect a certain decay in ω_i^2 , because the system size N is finite and the correlation length is of the order of N . As a result we can say that the partition function can be approximated by the partial sum of the DSM eigenvalues

$$\tilde{Z} = \sum_{i=1}^m \omega_i^2, \quad (13)$$

where m is the number of the eigenvalues kept. It is obvious that $Z \geq \tilde{Z}$, and the difference $Z - \tilde{Z}$ is fairly small when m is sufficiently large. We can also express \tilde{Z} as

$$\tilde{Z} = \text{Tr}(P_L \rho_L) = \text{Tr} \tilde{\rho}_L \quad (14)$$

where P_L is a projection operator

$$P_L(s'_L|s_L) = \sum_{\xi=1}^m A(s'_L|\xi)A(s_L|\xi), \quad (15)$$

and therefore $\tilde{\rho}_L$ is an m -dimensional diagonal matrix

$$\tilde{\rho}_L(\xi'|\xi) = \sum_{s'_L s_L} A(s'_L|\xi') \rho_L(s'_L|s_L) A(s_L|\xi) = \delta_{\xi' \xi} \omega_\xi^2 \quad (16)$$

where $\xi', \xi \leq m$. (If $m = 2^M$, P_L coincides with I_L in (10).) From (14)-(16), the transformation $\rho_L \rightarrow \tilde{\rho}_L$ by transformation from the row spin s_L to an m -state block spin ξ . In the same manner $B(s_R|\zeta)$ is related to the RG transformation $s_R \rightarrow \zeta$, that performs the mapping $\rho_R \rightarrow \tilde{\rho}_R$. In the following, we put a tilde over the renormalized matrices.

It should be noted that the RG transformation $\rho_L \rightarrow \tilde{\rho}_L$ is performed so that it maximizes the approximate partition function \tilde{Z} within the restricted number of degrees of freedom m . On the other hand, the RG transformation in quantum DMRG minimizes the ground-state energy; this is consistent with the thermodynamic relation $F = -k_B T \log Z = U - TS$ in the limit $T \rightarrow 0$.

4 Infinite-System Algorithm

In quantum DMRG, the infinite-system algorithm consists of the iterative use of the system expansion and the RG transformation for the Hamiltonian. It is easy to introduce the RG process to 2D classical system, using the quantum-classical correspondence $T^{(N)} = \exp(-\Delta H^{(N)})$; as far as the formulation is concerned, the classical DMRG can be obtained just by replacing the Hamiltonian $H^{(N)}$ in quantum DMRG by the transfer matrix $T^{(N)}$. As we show in

this section, the infinite-system algorithm constructs renormalized transfer matrices $\tilde{T}^{(6)}, \tilde{T}^{(8)}, \dots, \tilde{T}^{(N)}$ successively up to arbitrary N .

The infinite-system algorithm starts from the 4-site system, whose transfer matrix is given by

$$T^{(4)}(s'_1 s'_2 s'_3 s'_4 | s_1 s_2 s_3 s_4) = T_L(s'_1 s'_2 | s_1 s_2) W(s'_2 s'_3 | s_2 s_3) T_R(s'_3 s'_4 | s_3 s_4). \quad (17)$$

Diagonalizing $T^{(4)}$, we obtain the largest eigenvalue and the corresponding eigenvector $v^{(4)}(s_1 s_2 s_3 s_4)$, that satisfies $T^{(4)} v^{(4)} = \lambda v^{(4)}$. We then obtain $\rho_L(s'_1 s'_2 | s_1 s_2)$ and $\rho_R(s'_3 s'_4 | s_3 s_4)$ using (8) and get the RG transformation matrices $A(s_1 s_2 | \xi)$ and $B(s_3 s_4 | \zeta)$ from the diagonalization of ρ_L and ρ_R (11).

The next step is the enlargement of the system from $N = 4$ to $N = 6$ and the RG transformation via the substitution

$$\begin{aligned} \sum_{s'_1 s'_2 s_1 s_2} A(s'_1 s'_2 | \xi') T_L(s'_1 s'_2 | s_1 s_2) W(s'_2 s'_3 | s_2 s_3) A(s_1 s_2 | \xi) &\rightarrow \tilde{T}_L(\xi' s'_3 | \xi s_3) \\ \sum_{s'_5 s'_6 s_5 s_6} B(s'_5 s'_6 | \zeta') W(s'_4 s'_5 | s_4 s_5) T_R(s'_5 s'_6 | s_5 s_6) B(s_5 s_6 | \zeta) &\rightarrow \tilde{T}_R(s'_4 \zeta' | s_4 \zeta), \end{aligned} \quad (18)$$

where \tilde{T}_L and \tilde{T}_R are the renormalized half-row transfer matrices. (We have changed the spin indices of T_R according to the enlargement $N = 4$ to $N = 6$.) The greek indices ξ and ζ represent the block variables, that take (at most) m values. This is shown graphically in Fig. 4. Joining \tilde{T}_L and \tilde{T}_R , we can construct the renormalized transfer matrix for $N = 6$

$$\tilde{T}^{(6)}(\xi' s'_3 s'_4 \zeta' | \xi s_3 s_4 \zeta) = \tilde{T}_L(\xi' s'_3 | \xi s_3) W(s'_3 s'_4 | s_3 s_4) \tilde{T}_R(s'_4 \zeta' | s_4 \zeta). \quad (19)$$

At this stage, ξ and ζ represent the 2-spin blocks.

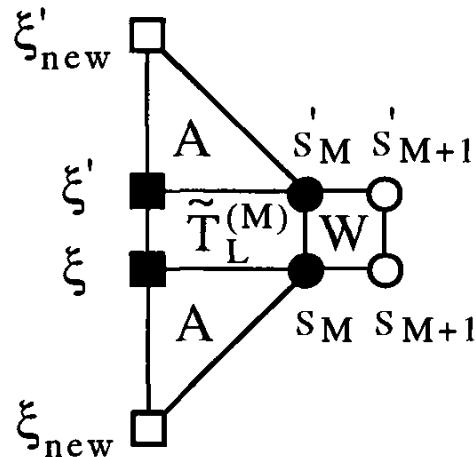


Fig. 4. Enlargement by W and RG transformation through A of the half-row transfer matrix \tilde{T}_L in (18)-(21). The filled circles and squares denote variables which are summed.

It is obvious that we can obtain $\tilde{T}^{(8)}$ in the same manner, i.e., diagonalizing $\tilde{T}^{(6)}$ in (19) to obtain the largest eigenvalue and the corresponding eigenvector $\tilde{v}^{(6)}(\xi s_3 s_4 \zeta)$, and create $\tilde{\rho}_L(s'_3 s'_4 | \xi s_3)$ and $\tilde{\rho}_R(s'_4 \zeta' | s_4 \zeta)$, using an extension of (8). We obtain the RG transformations $A(\xi s_3 | \xi_{\text{new}})$ and $B(s_4 \zeta | \zeta_{\text{new}})$ by diagonalizing $\tilde{\rho}_L$ and $\tilde{\rho}_R$, (an extension of (11)) respectively, where ξ_{new} and ζ_{new} are m -state variables that represent the 3-spin block. The $\tilde{T}^{(8)}$ is then constructed as

$$\begin{aligned} \tilde{T}^{(8)}(\xi'_{\text{new}} s'_4 s'_5 \zeta'_{\text{new}} | \xi_{\text{new}} s_4 s_5 \zeta_{\text{new}}) \\ = \tilde{T}_L(\xi'_{\text{new}} s'_4 | \xi_{\text{new}} s_4) W(s'_4 s'_5 | s_4 s_5) \tilde{T}_R(s'_5 \zeta'_{\text{new}} | s_5 \zeta_{\text{new}}), \end{aligned} \quad (20)$$

where $\tilde{T}_L(\xi'_{\text{new}} s'_4 | \xi_{\text{new}} s_4)$ and $\tilde{T}_R(s'_5 \zeta'_{\text{new}} | s_5 \zeta_{\text{new}})$ are created as

$$\begin{aligned} \sum_{\xi' s'_3 \xi s_3} A(\xi' s'_3 | \xi'_{\text{new}}) \tilde{T}_L(\xi' s'_3 | \xi s_3) W(s'_3 s'_4 | s_3 s_4) A(\xi s_3 | \xi_{\text{new}}) \rightarrow \tilde{T}_L(\xi'_{\text{new}} s'_4 | \xi_{\text{new}} s_4) \\ \sum_{s'_6 \zeta' s_6 \zeta} B(s'_6 \zeta' | \zeta'_{\text{new}}) W(s'_5 s'_6 | s_5 s_6) \tilde{T}_R(s'_6 \zeta' | s_6 \zeta) B(s_6 \zeta | \zeta_{\text{new}}) \rightarrow \tilde{T}_R(s'_5 \zeta'_{\text{new}} | s_5 \zeta_{\text{new}}). \end{aligned} \quad (21)$$

In this way, we further obtain $\tilde{T}^{(10)}$, $\tilde{T}^{(12)}$, $\tilde{T}^{(14)}$, ..., etc., up to arbitrary system size. This is the outline of the infinite-system algorithm.

Numerical diagonalization for $\tilde{T}^{(N)}$ is normally performed via the Lanczos method [26], that requires multiplication of $\tilde{T}^{(N)}$ with a $4m^2$ -dimensional vector x . Since $\tilde{T}^{(N)}$ is represented as a product of \tilde{T}_L , W , and \tilde{T}_R , the numerical multiplication $x''' = \tilde{T}^{(N)}x$ can be done very rapidly via the following three steps (Chap. 2(I))

$$\begin{aligned} x'(\xi' s'_M s_M s_{M+1} \zeta) &= \sum_{\xi} \tilde{T}_L(\xi' s'_M | \xi s_M) x(\xi s_M s_{M+1} \zeta), \\ x''(\xi' s'_M s'_{M+1} s_{M+1} \zeta) &= \sum_{s_M} W(s'_M s'_{M+1} | s_M s_{M+1}) x'(\xi' s'_M s_M s_{M+1} \zeta), \\ x'''(\xi' s'_M s'_{M+1} \zeta') &= \sum_{s_{M+1} \zeta} \tilde{T}_R(s'_{M+1} \zeta' | s_{M+1} \zeta) x''(\xi' s'_M s'_{M+1} s_{M+1} \zeta). \end{aligned} \quad (22)$$

One does not have to prepare the $4m^2$ -dimensional transfer matrix $\tilde{T}^{(N)}$, containing $16m^4$ matrix elements, explicitly. The Lanczos diagonalization can be accelerated further by an appropriate choice of the initial vector [44–46].

5 Variational State in DMRG

the classical (and also quantum) DMRG. Suppose that we have $\tilde{T}^{(N)}$ and the corresponding eigenvector $\tilde{v}^{(N)}(\xi s_M s_{M+1} \zeta)$ from the infinite-system algorithm, where $\langle \tilde{v}^{(N)} | \tilde{v}^{(N)} \rangle = 1$ is satisfied. If we form the expectation value

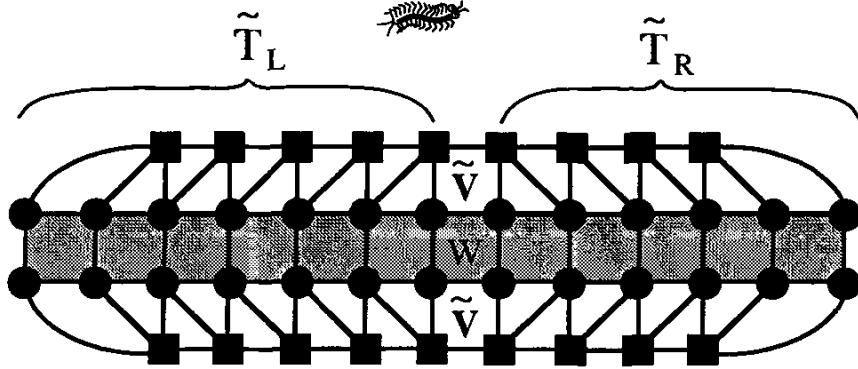


Fig. 5. Graphical representation of the variational eigenvalue of the transfer matrix $T^{(N)}$, see (23). The shaded region represents $T^{(N)}$. The upper and lower parts correspond to the variational state $v^{(N)}$ in (24).

$\tilde{\lambda} = \langle \tilde{v}^{(N)} | \tilde{T}^{(N)} | \tilde{v}^{(N)} \rangle$, this can be represented as shown in Fig. 5; the squares denote ξ, ξ', ζ , and ζ' variables which are defined in the renormalization steps. The left and the right halves of the diagram correspond to $\tilde{T}_L(\xi' s'_M | \xi s_M)$ and $\tilde{T}_R(s'_{M+1} \zeta' | s_{M+1} \zeta)$, respectively. As indicated by the filling, all the spin variables are all summed. If we recall that \tilde{T}_L and \tilde{T}_R are created through the iterative use of (21), it is possible to write

$$\langle \tilde{v}^{(N)} | \tilde{T}^{(N)} | \tilde{v}^{(N)} \rangle = \langle v^{(N)} | T^{(N)} | v^{(N)} \rangle, \quad (23)$$

where $v^{(N)}(s_1 \dots s_N)$ is the variational state for $T^{(N)}$ defined as

$$\sum_{\{\xi\}\{\zeta\}} \prod_{i=2}^{M-1} A(\xi_{i-1} s_i | \xi_i) \tilde{v}^{(N)}(\xi_{M-1} s_M s_{M+1} \zeta_{M+2}) \prod_{j=M+2}^{N-1} B(s_j \zeta_{j+1} | \zeta_j), \quad (24)$$

where $\xi_1 \equiv s_1$ and $\zeta_N \equiv s_N$. This corresponds to the portions in Fig. 5 above and below the original (= unrenormalized) transfer matrix, respectively. We have put the indices to ξ and ζ in order to distinguish the block spin variables. The matrices $A(\xi_{i-1} s_i | \xi_i)$ and $B(s_j \zeta_{j+1} | \zeta_j)$ depend on their position i and j , respectively. Note that $v^{(N)}(s_1 \dots s_N)$ is normalized because of the assumed normalization $\langle \tilde{v}^{(N)} | \tilde{v}^{(N)} \rangle = 1$ and the orthogonal relations

$$\begin{aligned} \sum_{\xi_{i-1} s_i} A(\xi_{i-1} s_i | \xi_i) A(\xi_{i-1} s_i | \xi'_i) &= \delta_{\xi_i \xi'_i} \\ \sum_{s_j \zeta_{j+1}}^m B(s_j \zeta_{j+1} | \zeta_j) B(s_j \zeta_{j+1} | \zeta'_j) &= \delta_{\zeta_j \zeta'_j}. \end{aligned} \quad (25)$$

Since $T^{(N)}$ is symmetric and positive definite, the largest eigenvalue $\tilde{\lambda}$ of $\tilde{T}^{(N)}$ is a variational lower bound for the largest eigenvalue λ of $T^{(N)}$

$$\tilde{\lambda} = \langle v^{(N)} | T^{(N)} | v^{(N)} \rangle \leq \lambda, \quad (26)$$

and the difference $\epsilon = \lambda - \tilde{\lambda}$ is a decreasing function of m .

It is possible to decompose $\tilde{v}^{(N)}(\xi_{M-1}s_M s_{M+1}\zeta_{M+2})$ further into a matrix product [27], and to write down the variational state in the form

$$\begin{aligned} v^{(N)}(s_1 \dots s_N) \\ = \sum_{\{\xi\}\{\zeta\}} \prod_{i=2}^M A(\xi_{i-1}s_i|\xi_i) \Omega(\xi_M|\zeta_{M+1}) \prod_{j=M+1}^{N-1} B(s_j\zeta_{j+1}|\zeta_j), \end{aligned} \quad (27)$$

where $\Omega(\xi_M|\zeta_{M+1})$ is a $2m$ -dimensional diagonal matrix

$$\Omega(\xi_M|\zeta_{M+1}) = \delta_{\xi_M \zeta_{M+1}} \frac{\omega_{\xi_M}}{\sqrt{\tilde{Z}}}. \quad (28)$$

We have imposed the normalization $\Omega^2 = 1$. (Strictly speaking, Ω need not to be diagonal, but $\Omega\Omega^T$ should be.) In the thermodynamic limit $N \rightarrow \infty$, the matrices $A(\xi_{i-1}s_i|\xi_i)$ and $B(s_j\zeta_{j+1}|\zeta_j)$ lose the position dependence, and the variational state in (27) coincides with Östlund's matrix-product state (See Chap. 3(I)).

As a result of the infinite-system algorithm, we obtain the variational free energy per site

$$\tilde{f} = -\frac{1}{N} k_B T \log \tilde{\lambda} \quad (29)$$

and other thermodynamic quantities via the numerical derivative of \tilde{f} . Since we have the variational state $v^{(N)}(s_1 \dots s_N)$ explicitly ((24) and (27)), we can calculate arbitrary spin correlation functions, such as $\langle s_i \rangle = \langle v^{(N)} | \hat{s}_i | v^{(N)} \rangle$ and $\langle s_i s_j \rangle = \langle v^{(N)} | \hat{s}_i \hat{s}_j | v^{(N)} \rangle$, etc. Generally speaking, the numerical precision of $\langle s_i s_j \rangle$ decreases as $|i - j|$ increases.

When we calculate spin correlation functions, we have to check that the energy scale ϵ_m introduced by the restriction of the degrees of freedom is sufficiently small compared with the excitation gap ϵ_g (of $H^{(N)} \equiv -\log \tilde{T}^{(N)}/\Delta$). If not, we have to increase m to keep the numerical precision. If the system is just at the critical temperature $T = T_c$, the gap ϵ_g is of the order of $1/N$. For this reason, it is not reliable to apply the infinite-system algorithm to a classical system just at T_c , and directly analyze the thermodynamic limit $N \rightarrow \infty$. It would be better to use the finite-system algorithm, that improves the variational state obtained by the infinite-system algorithm further (see Sect. 7).

Let us see how accurate the infinite-system algorithm is. Figure 6 shows the specific heat $C_v(T)$ of the square-lattice Ising model, which is obtained by taking the temperature derivative of the nearest-neighbor correlation function $E(T) = \langle s_M s_{M+1} \rangle$. We calculate $E(T)$ for the system up to $N = 2M = 2048$, which is sufficiently larger than the correlation length for each plotted temperature, except at $T = T_c$. The m -dependence for $E(T)$ is negligible when we keep $m = 60$ states. Since we know the exact solution for this

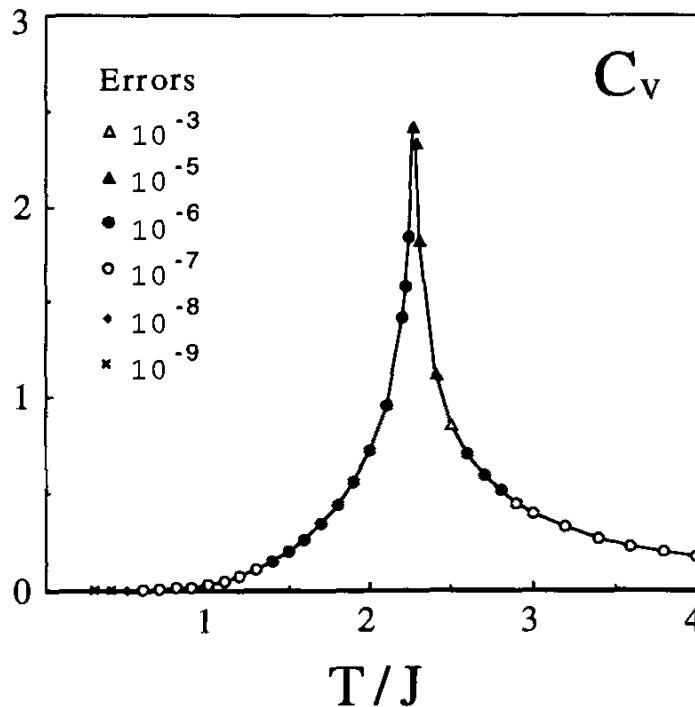


Fig. 6. Numerical results for the specific heat of the 2D Ising model [5].

model [28], we can directly observe the numerical error in $C_v(T)$; in Fig. 6, it is indicated by several marks. The numerical error is not negligible near the critical temperature T_c , partly because $E(T)$ is singular at T_c , and because the numerical derivative

$$\frac{E(T + \Delta T/2) - E(T - \Delta T/2)}{\Delta T} \quad (30)$$

is sensitive to ΔT ; typically, we have chosen $\Delta T = 10^{-4}$. The other source of numerical errors is the increase of the cut-off energy scale ϵ_m near T_c , that spoils the numerical precision of the block-spin transformation.

6 History of the Matrix-Product State

Variational states written in the form of matrix products, such as $v^{(N)}$ in (24), have been used for a long time. In this section we review briefly the history of the matrix-product state. In 1941 Kramers and Wannier (K-W) [29,30] investigated the square-lattice Ising model, assuming that the variational state can be written as

$$v(\dots, s_{i-1}, s_i, s_{i+1}, s_{i+2}, \dots) = \dots F(s_{i-1}|s_i)F(s_i|s_{i+1})F(s_{i+1}|s_{i+2})\dots, \quad (31)$$

where $F(s'|s)$ is a 2-dimensional symmetric matrix. The transition temperature T_c and the specific heat calculated from this variational state are more

accurate than those obtained by the molecular-field and the Bethe approximations [31]. It should be noted that the Gutzwiller approximation [32] for the Hubbard model [33,34] is quite similar to the K-W approximation.

Around 1960-70 Baxter improved the K-W approximation by introducing additional freedom [18,35]. His variational state is written as

$$\begin{aligned} v(\dots, s_{i-1}, s_i, s_{i+1}, s_{i+2}, \dots) \\ = \sum_{\dots, a, b, c, d, \dots} \dots F_{ab}(s_{i-1}|s_i) F_{bc}(s_i|s_{i+1}) F_{cd}(s_{i+1}|s_{i+2}) \dots, \end{aligned} \quad (32)$$

where \dots, a, b, c, d, \dots denote the additional m -state variables. Since $F_{ab}(s'|s)$ contains $4m^2$ adjustable parameters, the way of finding out the best $F_{ab}(s'|s)$ is non trivial. He performed the optimization using a self-consistent equation for the corner transfer matrix (CTM) [18].

Applications of the matrix-product formulation to quantum systems began with the investigations of Haldane's conjecture. In 1985 Nightingale and Blöte [36] used the K-W matrix product (31) as the initial vector of their projector Monte Carlo simulation. It is interesting that they commented on Baxter's method as follows “... *This method was formulated by Baxter for classical models in statistical mechanics. The generalization to quantum mechanical system is straightforward.*” (The RVA formulation by Martín-Delgado and Sierra can be seen as a realization of this concept; see Chap. 4(I).) In 1987 Affleck, Lieb, Kennedy, and Tasaki [37] showed that the ground state of a special $S = 1$ spin chain can be exactly expressed as

$$\sum_{\dots, a, b, c, d, e, \dots} \dots M_{ab}(s_{i-1}) M_{bc}(s_i) M_{cd}(s_{i+1}) M_{de}(s_{i+2}) \dots, \quad (33)$$

where $\dots, a, b, c, d, e, \dots$ are 2-state variables. They also showed that ground states of this kind exist also for a two-dimensional $S = 3/2$ quantum spin system. Fannes *et al.* generalized the above wave function (33) by assigning m degrees of freedom to $\dots, a, b, c, d, e, \dots$. Their variational state is known as ‘finitely correlated state,’ since the correlation length is always finite [38,39]. Although (33) does not look like (32), they are essentially the same; they are related via a duality transformation. Such a product state has been considered independently in the field of classical diffusion models [40–43].

The variational states in (31)-(33) are uniform. The advantage of the variational state in DMRG ((24) and (27)) is that it allows for a position dependence of the matrices. Because of this, it is possible to treat finite-size systems in the framework of DMRG.

7 Finite-System Algorithm

The variational state $v^{(N)}(s_1 \dots s_N)$ in (24) obtained by the infinite-system algorithm is a good approximation for the eigenvector of $T^{(N)}$, however, it

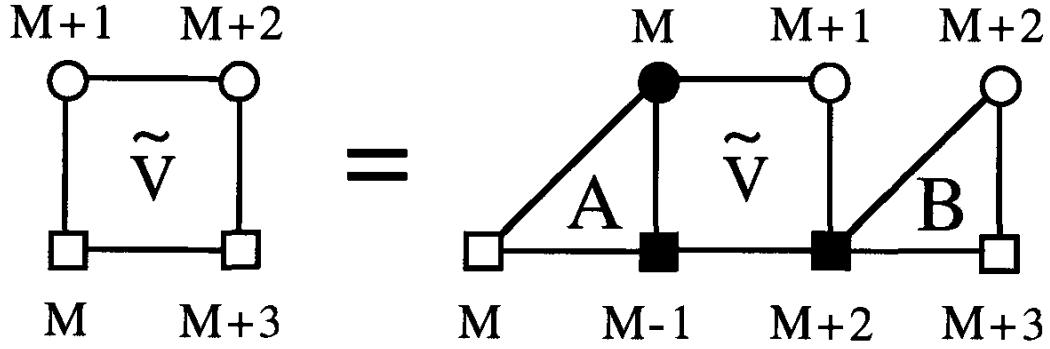


Fig. 7. Position shift of the renormalized state vector $\tilde{v}^{(N)}$ according to (35)-(36).

is not the best one. The finite-system algorithm improves the state, so that $\langle v^{(N)} | T^{(N)} | v^{(N)} \rangle$ is maximized under the constraint $\langle v^{(N)} | v^{(N)} \rangle = 1$, keeping at most m states for each block variable.

The algorithm gradually improves the whole variational state

$$\sum_{\{\xi\}\{\zeta\}} \prod_{i=2}^{M-1} A(\xi_{i-1} s_i | \xi_i) \tilde{v}^{(N)}(\xi_{M-1} s_M s_{M+1} \zeta_{M+2}) \prod_{j=M+2}^{N-1} B(s_j \zeta_{j+1} | \zeta_j). \quad (34)$$

by directly improving the vector $\tilde{v}^{(N)}$, and indirectly improving other parts by shifting the position of $\tilde{v}^{(N)}$. Assuming that we have already obtained all the matrices shown in (34), we explain the following numerical procedures of the finite-system algorithm of the classical DMRG; the main point is the position shift for $\tilde{v}^{(N)}$ [44,45]. Let us create the DSM

$$\begin{aligned} \tilde{\rho}_L(\xi'_{M-1} s'_M | \xi_{M-1} s_M) \\ = \sum_{s_{M+1} \zeta_{M+2}} \tilde{v}^{(N)}(\xi'_{M-1} s'_M s_{M+1} \zeta_{M+2}) \tilde{v}^{(N)}(\xi_{M-1} s_M s_{M+1} \zeta_{M+2}), \end{aligned} \quad (35)$$

and obtain $A(\xi_{M-1} s_M | \xi_M)$ by diagonalizing it. Then, by shifting the position of $\tilde{v}^{(N)}$ as shown in Fig. 7

$$\begin{aligned} \tilde{v}^{(N)}(\xi_M s_{M+1} s_{M+2} \zeta_{M+3}) = \\ \sum_{s'_M \xi'_{M-1} \zeta'_{M+2}} \tilde{A}(\xi'_{M-1} s'_M | \xi_M) \tilde{v}^{(N)}(\xi'_{M-1} s'_M s_{M+1} \zeta'_{M+2}) \tilde{B}(s_{M+2} \zeta_{M+3} | \zeta'_{M+2}), \end{aligned} \quad (36)$$

we can construct the new variational state

$$\begin{aligned} v_{\text{new}}^{(N)}(s_1 \dots s_N) = \\ \sum_{\{\xi\}\{\zeta\}} \prod_{i=2}^M A(\xi_{i-1} s_i | \xi_i) \tilde{v}^{(N)}(\xi_M s_{M+1} s_{M+2} \zeta_{M+3}) \prod_{j=M+3}^{N-1} B(s_j \zeta_{j+1} | \zeta_j). \end{aligned} \quad (37)$$

Compare (37) with (34), the position of $\tilde{v}^{(N)}$ is shifted by one.

The finite-system algorithm can be viewed as maximizing $\langle v_{\text{new}}^{(N)} | T^{(N)} | v_{\text{new}}^{(N)} \rangle$ via the tuning of $\tilde{v}^{(N)}(\xi_M s_{M+1} s_{M+2} \zeta_{M+3})$, under the constraint $\langle v_{\text{new}}^{(N)} | v_{\text{new}}^{(N)} \rangle = 1$. This maximization is equivalent to the diagonalization of the (shifted) renormalized transfer matrix

$$\begin{aligned} & \tilde{T}^{(N)}(\xi'_M s'_{M+1} s'_{M+2} \zeta'_{M+3} | \xi_M s_{M+1} s_{M+2} \zeta_{M+3}) \\ &= \tilde{T}_L(\xi'_M s'_{M+1} | \xi_M s_{M+1}) W(s'_{M+1} s'_{M+2} | s_{M+1} s_{M+2}) \\ & \quad \tilde{T}_R(s'_{M+2} \zeta'_{M+3} | s_{M+2} \zeta_{M+3}) \end{aligned} \quad (38)$$

to obtain its eigenvector $\tilde{v}^{(N)}(\xi_M s_{M+1} s_{M+2} \zeta_{M+3})$, where we already have $\tilde{T}_R(s'_{M+2} \zeta'_{M+3} | s_{M+2} \zeta_{M+3})$. The half-row transfer matrix $\tilde{T}_L(\xi'_M s'_{M+1} | \xi_M s_{M+1})$ can be easily obtained as

$$\sum_{\xi'_{M-1} s'_M \xi_{M-1} s_M} A(\xi'_{M-1} s'_M | \xi'_M) \tilde{T}_L(\xi'_{M-1} s'_M | \xi_{M-1} s_M) \\ W(s'_M s'_{M+1} | s_M s_{M+1}) A(\xi_{M-1} s_M | \xi_M) \quad (39)$$

Thus, by choosing $\tilde{v}^{(N)}$ in (36) as the initial vector of the Lanczos diagonalization for $\tilde{T}^{(N)}$ in (38) [44], one can rapidly improve $\tilde{v}^{(N)}(\xi_M s_{M+1} s_{M+2} \zeta_{M+3})$.

In such a way the finite-system algorithm shifts the position of $\tilde{v}^{(N)}$ to an arbitrary place and improves the variational state site by site [47]. The numerical procedures are basically the same as those in the quantum DMRG explained in Chap. 2(I). It is interesting that such a local improvement (or update) is also used in the zero-temperature QMC simulation for fermionic systems [48].

The advantage of the finite-system algorithm, compared to the infinite one, is its high numerical precision in the calculated thermodynamic quantities and spin correlation functions. The precision is high enough to determine the critical exponents and minor corrections to the scaling hypothesis of a classical system with the help of finite-size scaling [49,50]. Examples are shown in Chaps. 3.2(II) and 3.3(II) by Carlon and Drzewinski [6–8].

8 Corner Transfer Matrix Formulation

We have treated the 2D Ising model on a cylinder, and applied the RG transformation to the row transfer matrix $T^{(N)}$. Remember that the block spin transformation is obtained from the diagonalization of the DSM, which corresponds to a cut in the cylinder (Fig. 3). The pictorial image of the DSM suggests that we can define a DSM for any system with arbitrary geometry, just by creating a cut in it. Baxter considered such a construction of the DSM more than 30 years ago [18,35]. He took a square cluster, and expressed the DSM of the system as the fourth power of the so called corner transfer matrix (CTM).

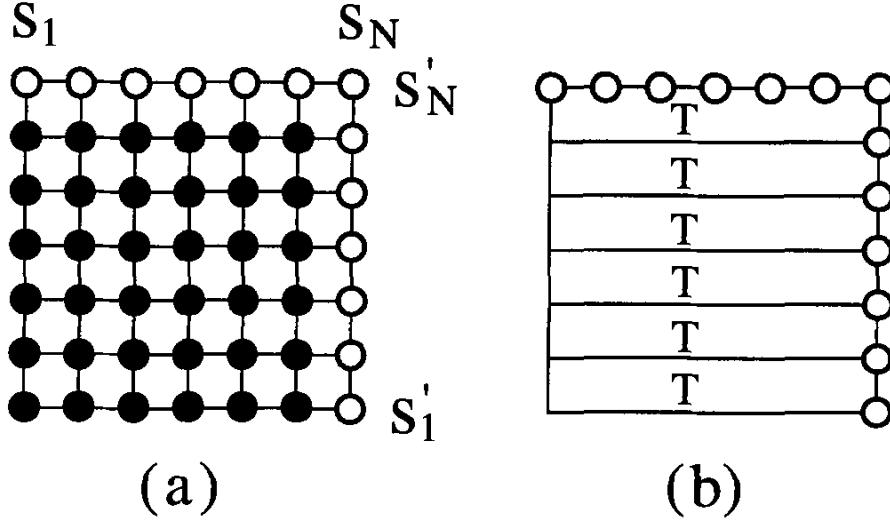


Fig. 8. Construction of the corner transfer matrix $C^{(N)}$, see (40). (a) The configuration sum is taken over the spins shown in black. (b) It is possible to regard $C^{(N)}$ as a stack of N copies of $T^{(N)}$ with appropriate boundary conditions.

The CTM represents the Boltzmann weight of a square system. Figure 8 (a) shows the corner transfer matrix of the Ising model

$$C^{(N)}(s'|s) = \sum_{\{s\}} \prod_{\langle i j k l \rangle} W(s_i s_j | s_k s_l), \quad (40)$$

where $\langle i j k l \rangle$ represents the neighboring spins around a plaquette W , and the sum is taken over all the spins shown in black. By definition, $C^{(N)}(s'|s)$ is block diagonal, because s'_N in $s' = s'_1 \dots s'_N$ and s_N in $s = s_1 \dots s_N$ are the same; in other words, $C^{(N)}(s'|s)$ for $s'_N \neq s_N$ is always zero. It is also possible to construct $C^{(N)}$ by stacking N transfer matrices: $N \times T^{(N)}$. (See Fig. 8 (b).) The system is then enlarged by joining another row and column to the square. This is done by multiplying horizontal and vertical transfer matrices T and a plaquette W as shown in Fig. 9 (a). Formally this means $C^{(N)}(s'|s) \rightarrow C^{(N+1)}(s'_{\text{new}}|s_{\text{new}}) = C^{(N+1)}(s'_1 \dots s'_{N+1}|s_1 \dots s_{N+1})$ with

$$C^{(N+1)}(s'_{\text{new}}|s_{\text{new}}) = \sum_{s''' s''} \delta_{s'_{N+1} s_{N+1}} W(s_N s_{N+1} | s''_N s'_N) T^{(N)}(s'|s''') T^{(N)}(s|s'') C^{(N)}(s'''|s''). \quad (41)$$

At the same time we extend the length of $T^{(N)}$ by joining a plaquette W as shown in Fig. 9 (b) in order to prepare for the next extension $C^{(N+1)} \rightarrow C^{(N+2)}$. The extension $T^{(N)} \rightarrow T^{(N+1)}$ is essentially the same as (21).

Baxter constructed the DSM as the fourth power of the CTM

$$\rho_c(s''''|s) = \sum_{s''' s'' s'} C^{(N)}(s''''|s''') C^{(N)}(s'''|s'') C^{(N)}(s''|s') C^{(N)}(s'|s). \quad (42)$$

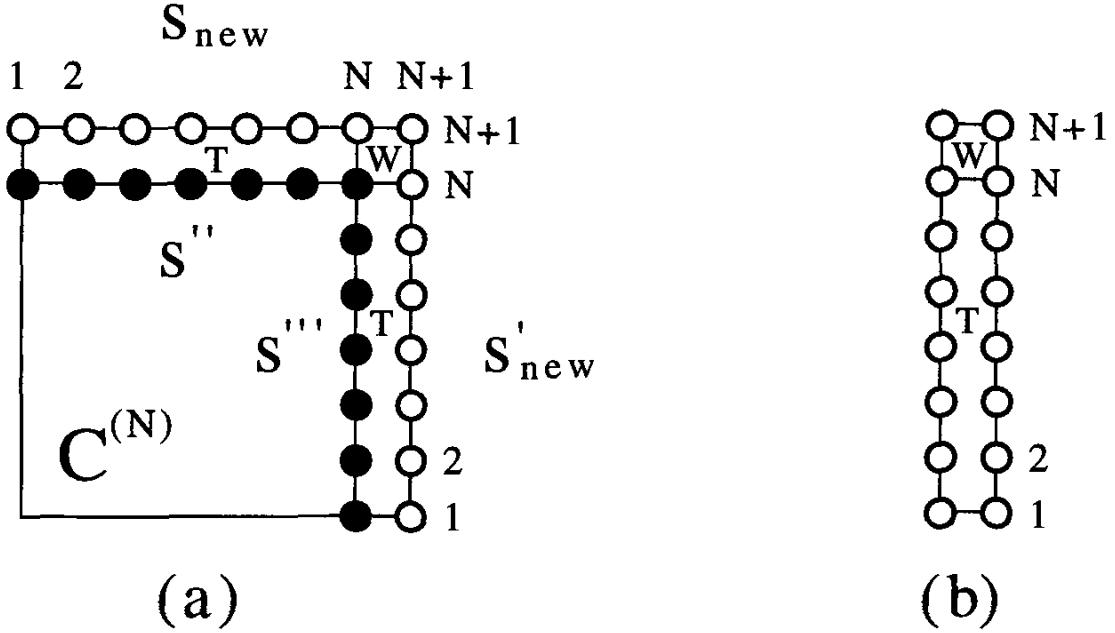


Fig. 9. Enlargement of the CTM: (a) $C^{(N+1)}$ is obtained by joining two $T^{(N)}$ and a W to $C^{(N)}$, see (41). (b) The extension $T^{(N)} \rightarrow T^{(N+1)}$, see (21).

This is not far from the conventional definition of DSM in DMRG (8), because ρ_c can be written as

$$\rho_c(s'_L|s_L) = \sum_{s_R} v(s'_L|s_R)v(s_L|s_R), \quad (43)$$

where the vector $v(s_L|s_R)$ is given by

$$v(s_L|s_R) = \sum_s C^{(N)}(s_L|s)C^{(N)}(s|s_R). \quad (44)$$

Note that $v(s_L|s_R)$ is block diagonal, as $C^{(N)}(s'|s)$ is. Figure 10 shows the graphical representation of ρ_c , that corresponds to a $2N - 1$ by $2N - 1$ square with a cut, where the cut extends from an edge to the center. Thus $Z^{(2N-1)} = \text{Tr } \rho_c$ is the partition function of the $2N - 1$ by $2N - 1$ square system.

Using the CTM, Baxter calculated the variational free energy (per site) of 2D lattice models in the limit $N \rightarrow \infty$ by solving a self-consistent equation. Okunishi pointed out that Baxter's method is basically the same as the infinite-system algorithm of the classical DMRG [51]. Compared to the classical DMRG, Baxter's method has the advantage that the numerical calculation is very fast, since it does not require diagonalizations of large-scale matrices. Nishino and Okunishi introduced this advantage into the DMRG and formulated a numerical RG algorithm, which is called 'corner transfer matrix renormalization group' (CTMRG) [52,53].

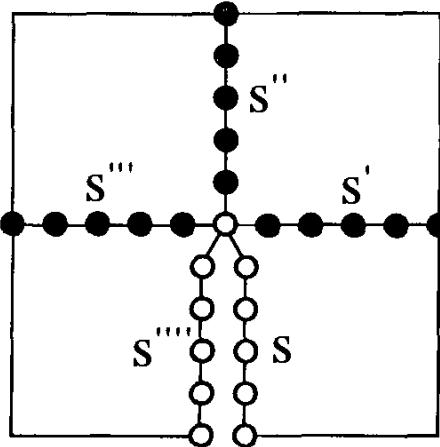


Fig. 10. Baxter's construction of the DSM as the fourth power of a CTM (42).

The outline of the CTMRG is as follows. The diagonalization of the DSM

$$\rho_c(s'|s) = \sum_{\xi} A(s'|\xi) \alpha_{\xi}^4 A(s|\xi) \quad (45)$$

defines the RG transformation $s \rightarrow \xi$ via the orthogonal matrix $A(s|\xi)$. For example, $C^{(N)}$ can be renormalized as

$$\tilde{C}^{(N)}(\xi'|\xi) = \sum_{s' s} A(s'|\xi') C^{(N)}(s'|s) A(s|\xi) = \delta_{\xi'|\xi} \alpha_{\xi}, \quad (46)$$

where ξ' and ξ are m -state block variables. One thing one has to keep in mind is that ρ_c is block diagonal, as $C^{(N)}$ is, and therefore the block variable ξ implicitly includes the spin variable s_N , the central spin in Fig. 10. The variational partition function of the $2N-1$ by $2N-1$ square is then expressed as

$$\tilde{Z}^{(2N-1)} = \text{Tr } \tilde{\rho}_c = \text{Tr } \left(\tilde{C}_c^{(N)} \right)^4 = \sum_{\xi=1}^m \alpha_{\xi}^4. \quad (47)$$

As the infinite-system algorithm, the CTMRG consists of the successive mappings $\tilde{C}^{(N)} \rightarrow \tilde{C}^{(N+1)} \rightarrow \tilde{C}^{(N+2)}$ using the enlargement of the system (41) and the RG transformation (46). As a result, we obtain $\tilde{Z}^{(2N-1)} \rightarrow \tilde{Z}^{(2N+1)} \rightarrow \tilde{Z}^{(2N+3)} \rightarrow \dots$ up to arbitrary system sizes, starting from a small N ($N = 2$ or 3).

Numerical data calculated by the CTMRG can be used for a finite-size scaling analysis of classical systems. Figure 11 shows the N -dependence of the local order parameter $\langle s \rangle$ of the Ising model at the center of a $(2N-1)$ by $(2N-1)$ square, when $T = T_c$. Fixed boundary conditions (all spins up at the boundary) are chosen. We plot representative data for both $m = 4$ and $m = 200$ up to $2N-1 = 19999$. At criticality, the local order parameter obeys the scaling formula

$$\langle s \rangle \propto (2N-1)^{-(d-2+\eta)/2}, \quad (48)$$

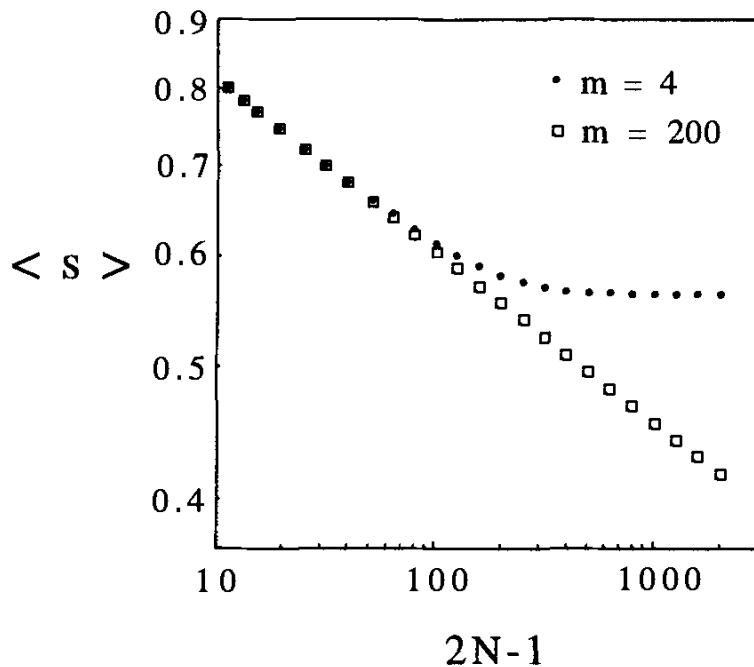


Fig. 11. Order parameter of the Ising model at the center of a $(2N-1)$ by $(2N-1)$ square when $T = T_c$.

where d is the spatial dimension ($d = 2$). Indeed, the calculated parameter $\langle s \rangle$ is almost proportional to $(2N-1)^{-1/8}$. The least-square fitting to the data in the range $199 \leq 2N-1 \leq 19999$ gives $\eta = 0.2501$, which is quite close to the exact one $\eta = 1/4$. In the same manner, we can determine another exponent ν from the nearest-neighbor spin correlation $E_{2N-1} = \langle s' s \rangle$ at the center of the $(2N-1)$ by $(2N-1)$ system, that obeys the scaling form

$$E_{2N-1} - E_\infty \propto (2N-1)^{1/\nu-d}. \quad (49)$$

From the calculated data for $199 \leq 2N-1 \leq 19999$, we obtain $\nu = 1.0006$. Again, the numerical result agrees with the exact exponent $\nu = 1$.

The CTMRG is also useful to detect a latent heat L . As an example, let us calculate L of the $q = 5$ Potts model [54], which shows a weak first-order transition. (It was impossible to determine L by Monte Carlo simulations because L is quite small.) Figure 12 shows the calculated local energy at the center of a $(2N-1)$ by $(2N-1)$ square system for both fixed (= ordered) and free (= disordered) boundary conditions, up to $2N-1 \leq 3999$ for $m = 40, 67$ and 200 . Though there is a non-negligible m -dependence, it is clear that the model does not show a second-order transition. A double extrapolation with respect to N and m gives the latent heat $L = 0.027$ which agrees with the exact result $L \sim 0.0265$ [18,20]. This is the first quantitative numerical estimate of L for the $q = 5$ Potts model.

Finally we mention that Baxter's construction of the DSM can be generalized to any dimension [55–57]. For example, consider a cubic cluster of a 3D Ising model, and divide it into 8 subcubes. As a direct extension of the CTM in 2D, we can imagine ‘corner tensors’ $C^{(N)}(a|b|c)$ that correspond to the

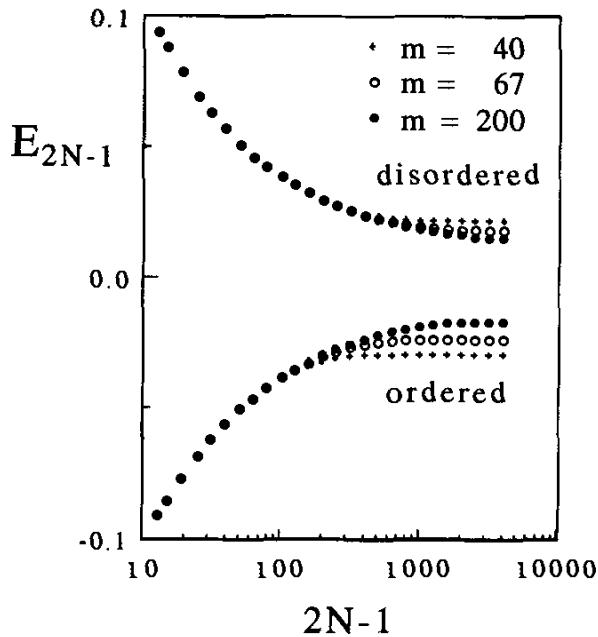


Fig. 12. Local energy E_{2N-1} of the $q = 5$ Potts model at the center of a $(2N - 1)$ by $(2N - 1)$ square, calculated for the ordered and the disordered phase. The zero of the energy has been shifted to the average of the two quantities [18,20].

Boltzmann weight for a subcube; the tensor indices $a = (a_{ij})$, $b = (b_{ij})$, and $c = (c_{ij})$ represent 2D spin arrays on the surfaces of the subcubes. Figure 13 shows the pictorial image of the DSM $\rho_{\text{cube}}(a|b)$, which is constructed as a contraction between 8 corner tensors. In 3D, the DSM corresponds to a cut in the cube. We obtain the RG transformation from the surfaces a , b , and c to m -state block spins by diagonalizing $\rho_{\text{cube}}(a|b)$. The enlargement of the corner cube can be performed via a 3D generalization of Fig. 9. Thus, as far as the formulation is concerned, we can extend CTMRG to 3D systems.

More generally, by breaking up a n -dimensional hypercube into 2^n -numbers of *hyper-corner-cubes*, we can define a DSM for a $(n - 1)$ -dimensional surface of the hyper-corner cube. Once we have obtained the DSM, we can define the block-spin transformation, and can apply RG transformations to the hypercubic system. Since the renormalized *hyper-corner tensor* has m^n numbers of elements, numerical calculations in higher dimension are much more difficult than in 2D. [58]

9 Discussion

We finally list several unsolved problems in classical DMRG.

- The matrix-product variational state for a finite-size system with periodic boundary conditions (= finite-size ring) should be translational invariant, however, the finite-system algorithm does not give such a translational invariant state [59].

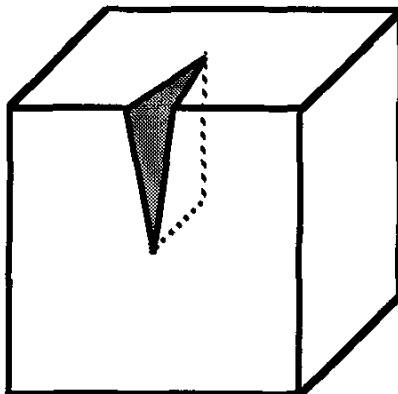


Fig. 13. The density matrix for a 3D cube, constructed from eight corner tensors. This is a generalization of Fig. 9.

- The way to apply classical DMRG to models with long-range interaction is not known; it is even difficult to define a transfer matrix for such models.
- It is not straightforward to treat random classical systems, because the transfer-matrix eigenvalue can be negative, and because the RG transformation depends on the position of block spins [60].
- To treat systems with continuous site variables or systems defined in continuous space (-time), we have to first discretize them [61]; we do not know the general principle for such a discretization that is suitable for DMRG.

We hope that these problems will be solved in the near future. [62]

The authors thank the organizers of the DMRG '98 workshop at the Max-Planck-Institut für Physik komplexer Systeme in Dresden. T. N. thanks I. Peschel, E. Carlon, G. Sierra, M. A. Martín-Delgado, X. Wang and T. Xiang for the discussions on classical DMRG. Numerical calculations were done using NEC SX-4 in the computer center of Osaka University. Y. H. is partly supported by the Sasakawa Scientific Research Grant from The Japan Science Society. K. O. and T. H. are partially supported by a Grant-in-Aid from the Ministry of Education, Science and Culture of Japan.

References

1. S.R. White, Phys. Rev. Lett. **69**, 2863 (1992); Phys. Rev. B **48**, 10345 (1993)
2. H.F. Trotter, Proc. Am. Math. Soc. **10**, 545 (1959)
3. M. Suzuki, Prog. Theor. Phys. **56**, 1454 (1976)
4. R.P. Feynmann and A. R. Hibbs, Quantum Mechanics and Path Integrals, McGraw-Hill (1965)

5. T. Nishino, J. Phys. Soc. Jpn. **64**, 3598 (1995)
6. E. Carlon and A. Drzewiński, Phys. Rev. Lett. **79**, (1997) 1591; Phys. Rev. E **57**, 2626 (1998)
7. E. Carlon and F. Igloi, Phys. Rev. B **57**, 7877 (1998); F. Igloi and E. Carlon, cond-mat/9805083
8. E. Carlon, A. Drzewiński and J. Rogiers, Phys. Rev. B **58**, 5070 (1998)
9. R.J. Bursill, T. Xiang and G.A. Gehring, J. Phys. Condensed Matter L583-L590 (1996)
10. X. Wang and T. Xiang, Phys. Rev. B **56**, 5061 (1997)
11. F. Naef, X. Wang, X. Zotos and W. van der Linden, cond-mat/9812117, to appear in Phys. Rev. B (1999)
12. N. Shibata, J. Phys. Soc. Jpn **66**, 2221 (1997)
13. B. Ammon, M. Troyer, T.M. Rice and N. Shibata, cond-mat/9812144
14. Y. Honda and T. Horiguchi, Phys. Rev. E **56**, 3920 (1997)
15. S. Östlund and S. Rommer, Phys. Rev. Lett **75**, 3537 (1995); S. Rommer and S. Östlund, Phys. Rev. B **55**, 2164 (1997); M. Andersson, M. Boman and S. Östlund, cond-mat/9810093
16. It is possible to fix the boundary spins to consider more general boundary conditions.
17. There is a bibliography of Ising in cond-mat/9605174.
18. R.J. Baxter, Exactly Solved Models in Statistical Mechanics, Academic Press, London, (1982)
19. R.B. Potts, Proc. Camb. Phil. Soc. **48**, 106
20. F.Y. Wu, Rev. Mod. Phys. **54**, 235 (1982) and references therein.
21. Baxter used another definition of the density submatrix in his variational method [18], where his density submatrix is block diagonal, see (42)-(44).
22. T. Nishino and K. Okunishi, in Strongly Correlated Magnetic and Superconducting Systems, Eds. G. Sierra and M.A. Martín-Delgado, Springer (1997)
23. Eigenvalues of the DSM can be negative when the system contains randomness.
24. I. Peschel, M. Kaulke and Ö. Legeza, Ann. Physik (Leipzig) **8**, 153 (1999); cond-mat/9810174
25. Strictly speaking, the density matrix eigenvalues do not decay exponentially; See K. Okunishi, Y. Hieida and Y. Akutsu, cond-mat/9810239
26. C. Lanczos: J. Res. Nat. Bur. Std. **45**, 255 (1950)
27. The Numerical recipes home page (<http://cfata2.harvard.edu/numerical-recipes/>) is useful to know about numerical linear algebra. Also it is worth reading J. Wilkinson, The Algebraic Eigenvalue Problem, Oxford (1965)
28. L. Onsager, Phys. Rev. **65**, 117 (1944)
29. H.A. Kramers and G.H. Wannier, Phys. Rev. **60**, 263 (1941)
30. R. Kikuchi, Phys. Rev. **81**, 988 (1951)
31. H.A. Bethe, Proc. Roy. Soc. A **150**, 552 (1935)
32. M.C. Gutzwiller, Phys. Rev. **137**, A1726 (1965)
33. J. Kanamori, J. Phys. Soc. Jpn. **30**, 275 (1963)
34. J. Hubbard, Proc. Roy. Soc. A **276**, 238 (1963); A **281**, 401 (1964)
35. R.J. Baxter, J. Math. Phys. **9**, 650 (1968); R.J. Baxter, J. Stat. Phys. **19**, 461 (1978)
36. N.P. Nightingale and H.W. Blöte: Phys. Rev. B **33**, 659 (1986)
37. I. Affleck, T. Kennedy, E.H. Lieb and H. Tasaki, Phys. Rev. Lett. **59**, 799 (1987)

38. M. Fannes, B. Nachtergale and R.F. Werner, *Europhys. Lett.* **10**, 633 (1989); M. Fannes, B. Nachtergale and R.F. Werner, *Commun. Math. Phys.* **144**, 443 (1992); M. Fannes, B. Nachtergale and R.F. Werner, *Commun. Math. Phys.* **174**, 477 (1995)
39. A. Klümper, A. Schadschneider and J. Zittartz, *Z. Phys. B* **87**, 281 (1992); H. Niggemann, A. Klümper and J. Zittartz, *Z. Phys. B* **104**, 103 (1997)
40. B. Derrida and M.R. Evans, *J. Phys. A: Math. Gen.* **26**, 1493 (1993)
41. N. Rajewsky, L. Santen, A. Schadschneider and M. Schreckenberg, *cond-mat/9710316*
42. A. Honecker and I. Peschel, *J. Stat. Phys.* **88**, 319 (1997)
43. Y. Hieida, *J. Phys. Soc. Jpn.* **67**, 369 (1998)
44. S.R. White, *Phys. Rev. Lett.* **77**, 3633 (1996)
45. T. Nishino and K. Okunishi, *J. Phys. Soc. Jpn.* **64**, 4084 (1995)
46. U. Schollwöck, *Phys. Rev. B* **58**, 8194 (1998)
47. A modification of the whole variational state is more time consuming. The situation is similar to the ‘Order N ’ problem in the density functional formalism.
48. S.R. White, D.J. Scalapino, R.L. Sugar, E.Y. Loh, J.E. Gubernatis and R.T. Scalettar, *Phys. Rev. B* **40**, 506 (1989)
49. M.E. Fisher, in *Proc. Int. School of Physics ‘Enrico Fermi’*, Ed. M.S. Green, Academic Press, **51**, p 1 (1971)
50. M.N. Barber, in *Phase Transitions and Critical Phenomena*, Ed. C. Domb and J.L. Lebowitz, Academic Press, **8**, p. 146 (1983) and references therein.
51. K. Okunishi, Thesis, Osaka University 1996 (in Japanese); Thesis, Osaka University 1999 (in English). (Contact to okunishi@godzilla.phys.sci.osaka-u.ac.jp)
52. T. Nishino and K. Okunishi, *J. Phys. Soc. Jpn.* **65**, 891 (1996); T. Nishino and K. Okunishi, *J. Phys. Soc. Jpn.* **66**, 3040 (1997)
53. T. Nishino, K. Okunishi, and M. Kikuchi, *Physics Letters A* **213**, 69 (1996)
54. T. Nishino and K. Okunishi, *J. Phys. Soc. Jpn.* **67**, 1492 (1998)
55. T. Nishino and K. Okunishi, *J. Phys. Soc. Jpn.* **67**, 3066 (1998)
56. G. Sierra and M.A. Martín-Delgado, *cond-mat/9811170*.
57. Y. Hieida, K. Okunishi and Y. Akutsu, *cond-mat/9901155*
58. A way to decrease the computational effort in higher dimension is to perform RG transformations *before* creating the new RG transformations; it is possible, because the (infinite-system) DMRG is a self consistent method. [1,45]
59. The numerical precision in DMRG for a system with periodic boundary conditions is lower than that for a system with open boundary conditions. The reason can be understood by looking at the variational state written as a matrix product.
60. Position dependence *in the quantum Hamiltonian* can be treated by quantum DMRG; K. Hida, *J. Phys. Soc. Jpn.* **65**, 895 (1996)
61. S.G. Chung, *J. Phys.: Cond. Matt.* **9**, L619 (1997); *Current Topics in Physics*, Ed. Y.M. Cho, J.B. Hong, and C.N. Yang, World Scientific, **1**, p 295 (1998)
62. The authors will list up the latest information about DMRG in this URL: <http://quattro.phys.sci.kobe-u.ac.jp/dmrg.html>

6 Quantum Transfer-Matrix and Momentum-Space DMRG

Tao Xiang¹ and Xiaoqun Wang²

¹ Institute of Theoretical Physics, Academia Sinica
Beijing 100080, P. R. China

² Institut Romand de Recherche Numerique en Physique des Materiaux
PPH-EPFL, CH-1015 Lausanne, Switzerland

The quantum transfer-matrix and momentum-space density-matrix renormalization group (DMRG) are two important tools for studying physical properties of quantum lattice models. Their development has greatly extended the field of application of the DMRG. This chapter is devoted to the technical aspects of both methods, emphasizing those features which are not present in the standard DMRG discussed in Chap. 2(I). From a technical point of view, there is no direct connection between these two methods. Readers who are interested in just one of them need to read the relevant section only.

The DMRG is a systematical numerical method for studying highly correlated systems [1]. It uses a small number of optimized basis states, say m states, to accurately expand the ground state (or other targeted states). In conventional numerical renormalization-group methods, one keeps the m lowest energy eigenstates of a block Hamiltonian [2–4]. In the DMRG, however, one keeps the m most probable basis states of a subblock in the ground state of a superblock. The standard DMRG has been successfully applied to many quasi one-dimensional systems at zero temperature. A detailed discussion has been given by Noack and White in Chap. 2(I).

The quantum transfer-matrix DMRG (TMRG) is by far the most accurate method for studying thermodynamic properties of quasi one-dimensional (1D) quantum lattice models with short range interactions. This method was first explored by Bursill, Xiang and Gehring [5] based on the classical TMRG method discussed in Chap. 3(I) [6]. Wang and Xiang [7] then introduced a non-symmetric density matrix in the context of a non-symmetric transfer-matrix representation of the partition function. In Sect. 1, we will discuss this method in detail. It has now been successfully applied to several interesting quasi one-dimensional systems [8–18]. Some applications can also be found in Chaps. 4.1-4.4 (II). Recently, the TMRG was extended by Mutou *et al.* and Naef *et al.* independently for the investigation of dynamic properties at finite temperatures [19,20]. The corresponding discussions can be found in the next chapter and in Chap. 4.2 (II).

The momentum-space DMRG is an extension of the standard DMRG method [1]. The application of the method is still in its early stage. The key technical problem in the momentum-space DMRG was solved by Xiang in 1996 [21]. He introduced the concept of composite operators to reduce the number of independent operators whose matrix elements need to be stored and updated, from the order of N^3 to the order of N . This has made DMRG iteration in momentum space practically possible. The development of this method provides a solution to the application of the DMRG in two or higher dimensions. Section 2 will focus on the detailed technical aspects of this method. The concept of composite operators can be used in other DMRG applications, such as the quantum-chemistry calculations discussed in Chaps. 2.1(II) and 2.2(II), respectively.

1 Transfer-Matrix DMRG

In this section, we illustrate how the TMRG works using a periodic isotropic spin- $\frac{1}{2}$ Heisenberg chain described by

$$H = J \sum_i \mathbf{S}_i \cdot \mathbf{S}_{i+1} - BS_i^z, \quad (1)$$

where J is the exchange constant and B is an external magnetic field. We will first discuss the quantum transfer-matrix theory [23] and show that all thermodynamic quantities can be obtained from the largest and the second-largest eigenvalue and corresponding eigenvectors of the transfer matrix. We will then discuss the problems encountered in applying the DMRG to the quantum transfer matrix. We define a non-symmetric reduced density matrix and show that the optimized basis vectors retained in the truncation of the Hilbert space are the eigenvectors of this non-symmetric density matrix.

Quantum transfer-matrix theory

The quantum transfer-matrix method is based on a Trotter-Suzuki decomposition of the partition function, which maps a d -dimensional quantum system to a $(d + 1)$ -dimensional classical one [24,25]. This decomposition has been widely used in quantum Monte Carlo simulations. It was first introduced by Trotter in 1959 [24] and justified by Suzuki in 1976 [25].

The first step of the quantum transfer-matrix method is to decompose the Hamiltonian into two parts, $H = H_1 + H_2$, with each part a sum of commuting terms:

$$\begin{aligned} H_1 &= \sum_{i=odd} h_{i,i+1}, \\ H_2 &= \sum_{i=even} h_{i,i+1}, \end{aligned} \quad (2)$$

where

$$h_{i,i+1} = J \mathbf{S}_i \cdot \mathbf{S}_{i+1} - \frac{B}{2} (S_i^z + S_{i+1}^z), \quad (3)$$

for the Heisenberg model (1).

The TMRG uses the second-order approximation of the Trotter formula [26]

$$Z = \text{Tr} e^{-\beta H} = \text{Tr} (V_1 V_2)^M + o(\epsilon^2), \quad (4)$$

where $\epsilon = \beta/M$ and M is the Trotter number.

$$V_1 = e^{-\epsilon H_1} = \prod_{i=odd} e^{-\epsilon h_{i,i+1}} = \prod_{i=odd} v^{i,i+1}, \quad (5)$$

$$V_2 = e^{-\epsilon H_2} = \prod_{i=even} e^{-\epsilon h_{i,i+1}} = \prod_{i=even} v^{i,i+1}, \quad (6)$$

where $v^{i,i+1}$ are local transfer matrices defined by $v^{i,i+1} = e^{-\epsilon h_{i,i+1}}$. By inserting $2M$ times the identity

$$\sum |s^1 \cdots s^N\rangle \langle s^1 \cdots s^N| = 1$$

between the V_1 and V_2 operators in (4) and labelling successively the complete bases with $k \in [1, 2M]$ (so-called times slices), the partition function can then be expressed as

$$\begin{aligned} Z &= \lim_{\epsilon \rightarrow 0} \sum_{\{s_k^i\}} \langle s_1^1 \cdots s_1^N | V_1 | s_2^1 \cdots s_2^N \rangle \langle s_2^1 \cdots s_2^N | V_2 | s_3^1 \cdots s_3^N \rangle \cdots \cdots \\ &\quad \langle s_{2M-1}^1 \cdots s_{2M-1}^N | V_1 | s_{2M}^1 \cdots s_{2M}^N \rangle \langle s_{2M}^1 \cdots s_{2M}^N | V_2 | s_1^1 \cdots s_1^N \rangle \\ &= \lim_{\epsilon \rightarrow 0} \sum_{\{s_k^i\}} \left(v_{1,2}^{1,2} v_{1,2}^{3,4} \cdots v_{1,2}^{N-1,N} \right) \left(v_{2,3}^{2,3} v_{2,3}^{4,5} \cdots v_{2,3}^{N,1} \right) \cdots \cdots \\ &\quad \left(v_{2M-1,2M}^{1,2} \cdots v_{2M-1,2M}^{N-1,N} \right) \left(v_{2M,1}^{2,3} \cdots v_{2M,1}^{N,1} \right), \end{aligned} \quad (7)$$

where $v_{k,k+1}^{i,i+1} = \langle s_k^i s_k^{i+1} | v^{i,i+1} | s_{k+1}^i s_{k+1}^{i+1} \rangle$ represents the matrix element of $v^{i,i+1}$. The superscripts i and subscripts k for s and v (subsequently σ and τ below) represent the spin coordinates in the space and Trotter directions, respectively. Equation (7) can be interpreted as an *evolution* of the product of V_1 and V_2 in the Trotter direction, as shown graphically in Fig. 1(a).

In the quantum transfer-matrix theory, the partition function is reformulated in terms of transfer matrices, *evolving along the spatial direction*, as illustrated in Fig. 1(b). If we collect all $v_{k,k+1}^{i,i+1}$ with the same $(i, i+1)$, then the partition function becomes

$$Z = \sum_{\{s_i^k\}} \left(v_{1,2}^{1,2} v_{3,4}^{1,2} \cdots v_{2M-1,2M}^{1,2} \right) \left(v_{2,3}^{2,3} v_{4,5}^{2,3} \cdots v_{2M,1}^{2,3} \right) \cdots \cdots$$

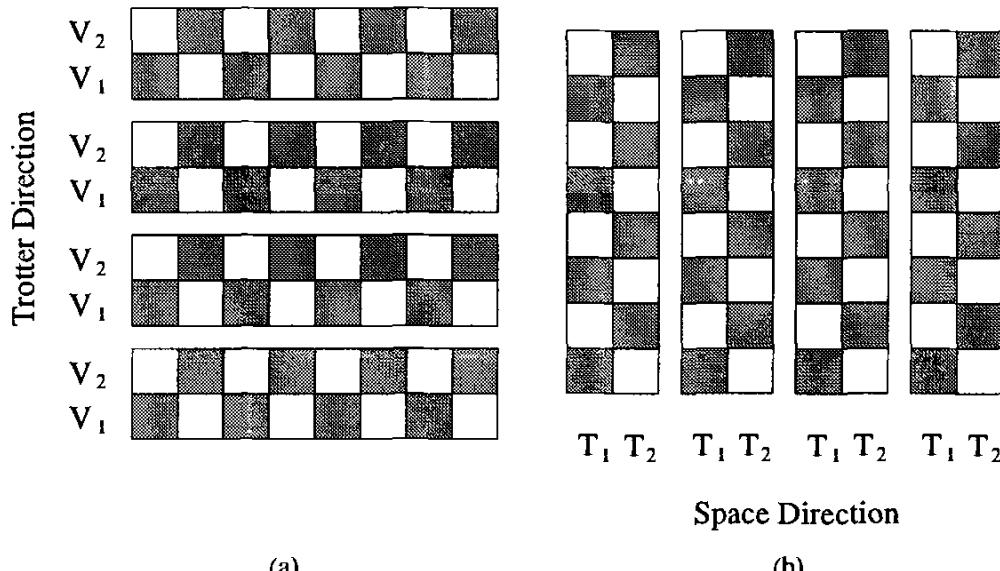


Fig. 1. Schematic representation of the partition function. Shaded squares denote local matrices $v^{i,i+1}$ in (a) and $\tau_{k,k+1}$ in (b).

$$\begin{aligned}
 & \left(v_{1,2}^{N-1,N} v_{3,4}^{N-1,N} \dots v_{2M-1,2M}^{N-1,N} \right) \left(v_{2,3}^{N,1} v_{4,5}^{N,1} \dots v_{2M,1}^{N,1} \right) \\
 &= \text{Tr} [(\tau_{1,2}\tau_{3,4}\dots\tau_{2M-1,2M})(\tau_{2,3}\tau_{4,5}\dots\tau_{2M,1})]^{N/2} \\
 &= \text{Tr} \mathcal{T}_M^{N/2},
 \end{aligned} \tag{8}$$

where $\mathcal{T}_M = T_1 T_2$ is called a quantum transfer matrix [23], T_1 and T_2 are products of local transfer matrices:

$$T_1 = \tau_{1,2}\tau_{3,4}\dots\tau_{2M-1,2M}, \tag{9}$$

$$T_2 = \tau_{2,3}\tau_{4,5}\dots\tau_{2M,1}. \tag{10}$$

The local transfer matrix $\tau_{k,k+1}$ is defined through the matrix $v^{i,i+1}$:

$$\tau(\sigma_k^i \sigma_{k+1}^i | \sigma_k^{i+1} \sigma_{k+1}^{i+1}) \equiv \langle s_k^i, s_k^{i+1} | v^{i,i+1} | s_{k+1}^i, s_{k+1}^{i+1} \rangle, \tag{11}$$

where $\sigma_k^i \equiv (-1)^{i+k} s_k^i$. σ_k^i differs from s_k^i only by a phase factor and is introduced for later convenience.

The transfer matrix \mathcal{T}_M defined above is not symmetric. In the case where both T_1 and T_2 are symmetric and semi-positive (i.e. all eigenvalues of T_1 and T_2 are larger than or equal to zero), the transfer matrix \mathcal{T}_M can be symmetrized. To see this more clearly, we rewrite (8) as

$$Z = \lim_{M \rightarrow \infty} \text{Tr} \left(T_1^{1/2} T_2 T_1^{1/2} \right)^{N/2}. \tag{12}$$

Since $T_1^{1/2}$ is space symmetric when T_1 is symmetric and semi-positive, $\mathcal{T}_M^S = T_1^{1/2} T_2 T_1^{1/2}$ is a real symmetric transfer matrix. This symmetrized transfer matrix, however, may not be of any practical use in the DMRG computations since it is a product of three matrices and the computer memory space and CPU time needed for storing and updating the matrix elements of \mathcal{T}_M^S is much

larger than for a non-symmetric transfer matrix. For a spin- S Heisenberg model with either ferromagnetic or antiferromagnetic coupling, the transfer matrix can be symmetrized in the absence of an external field. But in general T_1 and T_2 are neither symmetric nor semi-positive and the transfer matrix cannot be symmetrized.

Free energy: The free energy per site can be obtained directly from the partition function. In the thermodynamic limit, it is given by

$$\begin{aligned}
 F &= -\lim_{N \rightarrow \infty} \frac{1}{\beta N} \ln Z = -\lim_{N \rightarrow \infty} \lim_{\epsilon \rightarrow 0} \frac{1}{\beta N} \ln \text{Tr} T_M^{N/2} \\
 &= -\lim_{N \rightarrow \infty} \lim_{\epsilon \rightarrow 0} \frac{1}{\beta N} \ln \sum_l \langle \psi_l^L | T_M^{N/2} | \psi_l^R \rangle \\
 &= -\lim_{N \rightarrow \infty} \lim_{\epsilon \rightarrow 0} \frac{1}{\beta N} \ln \lambda_{max}^{N/2} \sum_l (\lambda_l / \lambda_{max})^{N/2}, \\
 &= -\lim_{\epsilon \rightarrow 0} \frac{\ln \lambda_{max}}{2\beta}
 \end{aligned} \tag{13}$$

where λ_l is the eigenvalue of T_M , and $|\psi_l^R\rangle$ and $\langle \psi_l^L|$ are the corresponding right and left eigenvectors, respectively. We use λ_{max} to denote the largest eigenvalue of T_M . The normalization condition $\langle \psi_l^L | \psi_l^R \rangle = 1$ is assumed hereafter. In obtaining the above formula, the property that the order of the limits $N \rightarrow \infty$ and $\epsilon \rightarrow 0$ is interchangeable [27,28] is used. The free energy is determined solely by the *largest eigenvalue* of the quantum transfer matrix T_M . In principle, one can obtain all thermodynamic quantities from various derivatives of F , i.e. only the largest eigenvalue of T_M is needed to study the full thermodynamics of a system. This is an important property in the transfer-matrix theory, which allows us to adopt the idea of the density-matrix renormalization to approximately find λ_{max} .

Conservation law: The local Hamiltonian $h_{i,i+1}$ conserves the total spin at sites i and $i+1$, i.e. $s_k^i + s_k^{i+1} = s_{k+1}^i + s_{k+1}^{i+1}$. In terms of the spins σ introduced before, the same relation holds if one moves in the space direction: $\sigma_k^i + \sigma_{k+1}^i = \sigma_k^{i+1} + \sigma_{k+1}^{i+1}$. From this local conservation law it can be shown that the total spin, $\sum_k \sigma_k^i$, is also conserved in the Trotter space: $\sum_k \sigma_k^i = \sum_k \sigma_k^{i+1}$. Thus T_M is block-diagonal according to the value of $\sum_k \sigma_i^k$. For the spin-1/2 Heisenberg model, it was shown rigorously that the largest eigenvector is nondegenerate and lies in the $\sum \sigma_i^k = 0$ subspace, irrespective of the sign of J and the value of B [29]. For larger spins, it was found numerically that the largest eigenvectors of T_M are also in the $\sum_k \sigma_k^i = 0$ subspace [13], but a rigorous proof for this is still not available.

The above discussion can be readily extended to an interacting electron model, for example the Hubbard model. In this case charge is conserved in

addition to the spin conservation. If we use $n_{k\sigma}^i$ to denote the number of electrons with spin σ at site i and time slice k , then the spin and charge conservation laws for this system are given by the equation $n_{k\sigma}^i + n_{k\sigma}^{i+1} = n_{k+1\sigma}^i + n_{k+1\sigma}^{i+1}$ in the real space. In the Trotter space, the corresponding conservation law can be expressed as $\tilde{n}_{i\sigma}^k + \tilde{n}_{i\sigma}^{k+1} = \tilde{n}_{i+1\sigma}^k + \tilde{n}_{i+1\sigma}^{k+1}$, where $\tilde{n}_{k\sigma}^i = (-1)^{i+k}(n_{k\sigma}^i - 1/2)$. When $i+k$ is odd, this transformation between n_{σ}^i and \tilde{n}_{σ}^i is equivalent to a particle-hole transformation: $c_{i\sigma}^k \leftrightarrow (c_{i\sigma}^k)^\dagger$.

Form of local transfer matrix τ : Before we discuss the properties of thermodynamic quantities, let us illustrate the symmetry property of the local transfer matrix explicitly using the spin- $\frac{1}{2}$ Heisenberg model (1). To obtain $\tau_{1,2}$, we first calculate the matrix elements of $v^{1,2}$ by diagonalizing the local Hamiltonian $h_{1,2}$. If the basis states of the two spins are ($| \uparrow\uparrow \rangle$, $| \uparrow\downarrow \rangle$, $| \downarrow\uparrow \rangle$, $| \downarrow\downarrow \rangle$), then it is simple to show that

$$v^{1,2} = \begin{pmatrix} e^{-(\frac{J}{4}-\frac{B}{2})\epsilon} & 0 & 0 & 0 \\ 0 & e^{\frac{J\epsilon}{4}} \cosh \frac{J\epsilon}{2} & -e^{\frac{J\epsilon}{4}} \sinh \frac{J\epsilon}{2} & 0 \\ 0 & -e^{\frac{J\epsilon}{4}} \sinh \frac{J\epsilon}{2} & e^{\frac{J\epsilon}{4}} \cosh \frac{J\epsilon}{2} & 0 \\ 0 & 0 & 0 & e^{-(\frac{J}{4}+\frac{B}{2})\epsilon} \end{pmatrix}, \quad (14)$$

$v^{1,2}$ is a block-diagonal and symmetric matrix. Substituting this equation into (11), we then obtain

$$\tau_{1,2} = \begin{pmatrix} -e^{\frac{J\epsilon}{4}} \sinh \frac{J\epsilon}{2} & 0 & 0 & 0 \\ 0 & e^{\frac{J\epsilon}{4}} \cosh \frac{J\epsilon}{2} & e^{-(\frac{J}{4}-\frac{B}{2})\epsilon} & 0 \\ 0 & e^{-(\frac{J}{4}+\frac{B}{2})\epsilon} & e^{\frac{J\epsilon}{4}} \cosh \frac{J\epsilon}{2} & 0 \\ 0 & 0 & 0 & -e^{\frac{J\epsilon}{4}} \sinh \frac{J\epsilon}{2} \end{pmatrix}. \quad (15)$$

One can immediately see that $\tau_{1,2}$ is still block-diagonal. But it is asymmetric when B is non-zero.

Thermal average of local quantities: As mentioned above, all thermodynamic quantities can be calculated from the derivatives of F . However, for the internal energy or any other quantity which can be written as a sum of local operators, we can evaluate the thermal average of these quantities directly from the largest eigenvalue and the corresponding eigenvectors $\{|\psi_{max}^R\rangle, \langle\psi_{max}^L|\}$ of T_M . This can eliminate the error resulting from the numerical derivatives.

Let us consider the thermal average of a local operator, $A_{1,2}$ which is defined at sites 1 and 2 and commutes with the local Hamiltonian $h_{1,2}$. Following the steps leading to (7), it can be shown that

$$\langle A_{1,2} \rangle = \lim_{N \rightarrow \infty} \frac{1}{Z} \text{Tr} (A_{1,2} e^{-\beta H})$$

$$\begin{aligned}
&= \lim_{N \rightarrow \infty} \lim_{\epsilon \rightarrow 0} \frac{1}{Z} \text{Tr} \left(\mathcal{T}_M(A_{1,2}) \mathcal{T}_M^{N/2-1} \right) \\
&= \lim_{\epsilon \rightarrow 0} \frac{\langle \psi_{max}^L | \mathcal{T}_M(A_{1,2}) | \psi_{max}^R \rangle}{\lambda_{max}}
\end{aligned} \tag{16}$$

where

$$\mathcal{T}_M(A_{1,2}) = (\tilde{\tau}_{1,2}(A_{1,2}) \tau_{3,4} \cdots \tau_{2M-1,2M}) T_2, \tag{17}$$

and the elements of $\tilde{\tau}_{1,2}(A_{1,2})$ are given by

$$\langle \sigma_1^1 \sigma_2^1 | \tilde{\tau}_{1,2}(A_{1,2}) | \sigma_1^2 \sigma_2^2 \rangle \equiv \langle s_1^1, s_1^2 | A_{1,2} v^{1,2} | s_2^1, s_2^2 \rangle. \tag{18}$$

The internal energy is the thermal average of the Hamiltonian H . Using the translation symmetry, the internal energy per site is equal to the thermal average of $h_{1,2}$. Thus

$$U = \langle h_{1,2} \rangle = \lim_{\epsilon \rightarrow 0} \frac{\langle \psi_{max}^L | \mathcal{T}_M(h_{1,2}) | \psi_{max}^R \rangle}{\lambda_{max}} \tag{19}$$

Similarly the uniform magnetization is given by

$$M_z = \lim_{N \rightarrow \infty} \frac{1}{N} \langle \sum_i S_i^z \rangle = \lim_{\epsilon \rightarrow 0} \frac{\langle \psi_{max}^L | \mathcal{T}_M(\frac{S_1^z + S_2^z}{2}) | \psi_{max}^R \rangle}{\lambda_{max}} \tag{20}$$

Specific heat and spin susceptibility: From the results for U and M_z , one can evaluate the specific heat

$$C_v = \frac{\partial U}{\partial T}, \tag{21}$$

and the magnetic susceptibility

$$\chi = \frac{\partial M_z}{\partial B}, \tag{22}$$

by numerical derivatives. The magnetic susceptibility is simply given by M_z/B for sufficiently small magnetic field since $M_z = 0$ at $B = 0$. Therefore, the numerical errors for χ at $B = 0$ are generally smaller than those at $B \neq 0$ or those for the specific heat.

Correlation length: The correlation length at finite temperatures can also be calculated from the eigenvalues of \mathcal{T}_M . To do this let us consider the correlation function of the fluctuation of a quantity around its mean value, $\delta A_i = A_i - \langle A_i \rangle$, between sites 1 and $r+1$. For even r , the thermal average of $\delta A_1 \delta A_{r+1}$ in the thermodynamic limit $N \rightarrow \infty$ is found to be

$$\begin{aligned}
\langle \delta A_1 \delta A_{r+1} \rangle &= \frac{\langle \psi_{max}^L | \mathcal{T}_M(\delta A_1) \mathcal{T}_M^{r/2-1} \mathcal{T}_M(\delta A_{r+1}) | \psi_{max}^R \rangle}{\lambda_{max}^{r/2+1}} \\
&= \sum_l \frac{\langle \psi_{max}^L | \mathcal{T}_M(\delta A_1) | \psi_l^R \rangle \langle \psi_l^L | \mathcal{T}_M(\delta A_{r+1}) | \psi_{max}^R \rangle}{\lambda_{max} \lambda_l} \left(\frac{\lambda_l}{\lambda_{max}} \right)^{r/2}.
\end{aligned} \tag{23}$$

In the limit $r \rightarrow \infty$, the above formula becomes

$$\begin{aligned} & \lim_{r \rightarrow \infty} \langle \delta A_1 \delta A_{r+1} \rangle \\ &= \lim_{\epsilon \rightarrow 0} \frac{\langle \psi_{max}^L | \mathcal{T}_M(\delta A_1) | \psi_\alpha^R \rangle \langle \psi_\alpha^L | \mathcal{T}_M(\delta A_{r+1}) | \psi_{max}^R \rangle}{\lambda_{max} \lambda_\alpha} e^{-(\xi^{-1} - i\kappa)r}, \end{aligned} \quad (24)$$

where $\lambda_\alpha = |\lambda_\alpha| e^{2i\kappa}$ is the largest eigenvalue of \mathcal{T}_M satisfying the condition

$$\langle \psi_{max}^L | \mathcal{T}_M(\delta A_1) | \psi_\alpha^R \rangle \langle \psi_\alpha^L | \mathcal{T}_M(\delta A_{r+1}) | \psi_{max}^R \rangle \neq 0. \quad (25)$$

If the vector $\mathcal{T}_M(\delta A_1) | \psi_{max}^R \rangle$ is in the same invariant subspace as $| \psi_{max}^R \rangle$ (for example, when $A_i = S_i^z$), then λ_α is usually the second largest eigenvalue of \mathcal{T}_M in that subspace. If, on the other hand, the vector $\mathcal{T}_M(\delta A_1) | \psi_{max}^R \rangle$ is in a different subspace $| \psi_{max}^R \rangle$, (for example, when $A_i = S_i^x$), then λ_α is the largest eigenvalue of \mathcal{T}_M in the subspace which contains $\mathcal{T}_M(\delta A_1) | \psi_{max}^R \rangle$.

In (24), ξ is the thermal correlation length determined by λ_{max} and λ_α

$$\xi^{-1} = \lim_{\epsilon \rightarrow 0} \frac{1}{2} \ln \left| \frac{\lambda_{max}}{\lambda_\alpha} \right| = - \lim_{r \rightarrow \infty} \frac{1}{r} \ln \langle \delta A_1 \delta A_{r+1} \rangle, \quad (26)$$

and κ is the characteristic wave vector of the most dominant fluctuations which is determined by the phase of λ_α

$$\kappa = \lim_{\epsilon \rightarrow 0} \frac{1}{2} \arg \left(\frac{\lambda_\alpha}{\lambda_{max}} \right) + n\pi \quad (n = 0 \text{ or } 1). \quad (27)$$

The value of κ is not uniquely determined because \mathcal{T}_M crosses over two sites and cannot distinguish between κ and $\kappa + \pi$. This ambiguity in determining the value of κ can be removed by considering the behavior of κ at high temperature and the continuity of κ . At high temperature λ_α is generally real, κ is therefore either 0 or π . From physical considerations, one can readily determine whether κ is 0 or π .

DMRG implementation for \mathcal{T}_M

The transfer matrix \mathcal{T}_M , rotated by $\pi/2$ in the rest of this section, is shown again in Fig. 2. It corresponds to a one-dimensional quantum operator con-

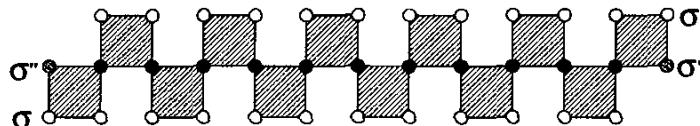


Fig. 2. One-dimensional structure of the transfer matrix \mathcal{T}_M along the Trotter direction which is taken horizontal here. \mathcal{T}_M is formed by connecting local transfer matrices τ at solid circles. At each solid circle, the common states of two τ are summed over. Either the upper open circles or the lower ones individually construct the Hilbert space to represent \mathcal{T}_M . Two shaded circles represent the same states σ'' to be summed over to form a transfer-matrix ring.

sisting of the local transfer matrices. As such it can be treated as the usual quantum spin chains. There is a difference, however, to the usual Hamiltonians which are sums of local terms. In T_M , the local transfer matrices are connected at the sites shown as solid circles in Fig. 2 and one has several possibilities of separating T_M into two parts and enlarging them in the density-matrix renormalization procedure. In each iteration of the infinite-system algorithm, two time slices are added, i.e. M is increased by one. It turns out that the temperature is lowered as $T = 1/\epsilon M$ for given ϵ . Moreover the asymmetry of the transfer matrix T_M intrinsically also leads to a non-symmetric reduced density matrix. Below we discuss these aspects.

Superblock I: This construction, slightly different from the one proposed for systems which are spatial reflection symmetric [7], is generally suitable for (quasi) one-dimensional systems with local interactions. The two typical configurations of the superblock are shown in Fig. 3. The superblock consists of the two blocks in the dashed frames, which we call renormalized blocks, plus two additional time slices. The system contains a renormalized block and one slice on the right. The rest is thus its environment. We use n_s and n_e to label

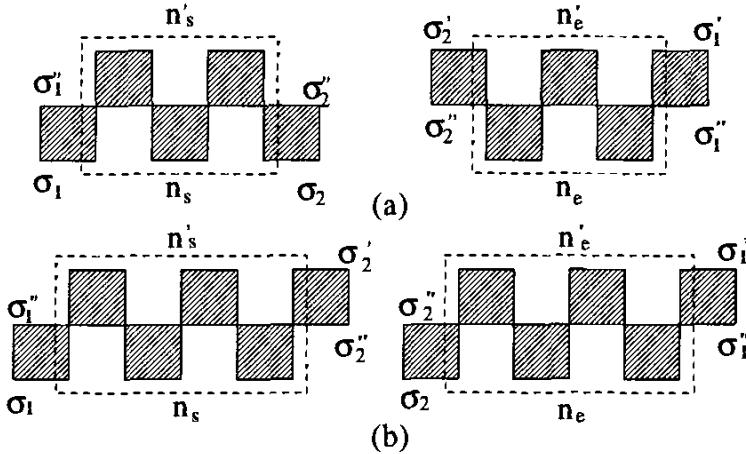


Fig. 3. Configurations of the superblock: (a) $M = 5$ (odd) and (b) $M = 6$ (even). The left and right transfer matrices are connected by summing over the internal states σ_1'' and σ_2'' at slices σ_1 and σ_2 to form a periodic local transfer matrix chain for T_M .

the basis states of the renormalized blocks in the system and the environment, respectively. The states of two slices are represented by σ_1 , σ_1'' , σ_1' and σ_2 , σ_2'' , σ_2' correspondingly. Then the elements of the left transfer matrices are denoted as $S_o(\sigma_1'', n_s, \sigma_2'; \sigma_1, n_s, \sigma_2)$ or $S_e(\sigma_1', n_s, \sigma_2''; \sigma_1'', n_s, \sigma_2)$ and those of the right ones as $E_o(\sigma_2', n_e, \sigma_1'; \sigma_2'', n_e, \sigma_1')$ or $E_e(\sigma_2'', n_e, \sigma_1'; \sigma_2, n_e, \sigma_1'')$, where the subscripts o or e correspond to M being odd or even for T_M . Now let us see how $S_{e,o}$ and $E_{e,o}$ can be constructed and renormalized in an iterative way.

- Initially, $M = 2$ is *even*, the number of slices is $2M = 4$, and T_M contains four local transfer matrices τ . One then has $S_e(\sigma_1', \sigma_1'', \sigma_2'; \sigma_1'', \sigma_2, \sigma_2'') = \sum_{\sigma''} \tau(\sigma_1', \sigma_1'' | \sigma_1'', \sigma_2'') \tau(\sigma_2'', \sigma_2' | \sigma_2, \sigma_2'')$ and $E_e = S_e$.

- When we construct T_M with an *odd* M , one slice is added to the right of the renormalized system block in S_e and one to the left of the renormalized environment block in E_e . The elements of the enlarged matrices, respectively, are given by

$$\begin{aligned}\mathcal{S}_o(\sigma''_1, \tilde{n}'_s, \sigma''_2; \sigma_1, \tilde{n}_s, \sigma_2) &= \sum_{\sigma''} \mathcal{S}_e(\sigma'_1, n'_s, \sigma''; \sigma''_1, n_s, \sigma) \tau(\sigma'', \sigma''_2 | \sigma, \sigma_2) \\ \mathcal{E}_o(\sigma'_2, \tilde{n}'_e, \sigma''_1; \sigma''_2, \tilde{n}_e, \sigma_1) &= \sum_{\sigma''} \tau(\sigma'_2, \sigma' | \sigma''_2, \sigma'') \mathcal{E}_e(\sigma'', n'_s, \sigma''_1; \sigma, n_s, \sigma_1).\end{aligned}$$

As a consequence, the basis states of the enlarged renormalized blocks become $|\tilde{n}_s\rangle = |n_s\rangle \otimes |\sigma\rangle$ and $|\tilde{n}_e\rangle = |\sigma\rangle \otimes |n_e\rangle$.

- When \mathcal{T}_M with even M is constructed, \mathcal{S}_e and \mathcal{E}_e can be considered to be identical. Then one can actually carry out the manipulations on \mathcal{S}_e only. As illustrated in Fig. 4 (a), this involves the enlargement via

$$\mathcal{S}_e(\sigma''_1, \tilde{n}'_s, \sigma'_2; \sigma_1, \tilde{n}_s, \sigma''_2) = \sum_{\sigma''} \mathcal{S}_o(\sigma''_1, n'_s, \sigma''; \sigma_1, n_s, \sigma) \tau(\sigma', \sigma'_2 | \sigma'', \sigma''_2).$$

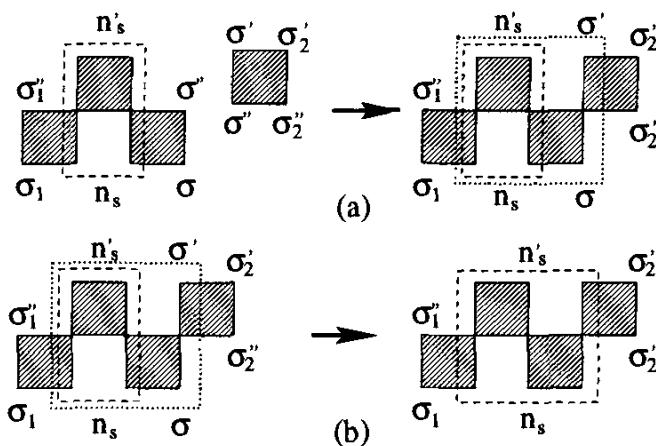


Fig. 4. (a) The renormalized block in the dashed frame is enlarged, by adding one slice $(\sigma, \sigma'', \sigma')$, into the one in the dotted frame and the basis $|n_s\rangle$ into $|\tilde{n}_s\rangle = |n_s\rangle \otimes |\sigma\rangle$.
(b) The block in the dotted frame is further renormalized into the one in the dashed frame and $|\tilde{n}_s\rangle$ into $|n_s\rangle$

- As soon as the number of states in $|\tilde{n}_\gamma\rangle$ ($\gamma = s, e$) exceeds m , the enlarged transfer matrices $\mathcal{S}_{e,o}$ and \mathcal{E}_o are renormalized by

$$\mathcal{A}_\alpha(n'_\gamma, n_\gamma) = \sum_{\tilde{n}'_\gamma, \tilde{n}_\gamma} O_\gamma^l(n'_\gamma, \tilde{n}'_\gamma) \mathcal{A}_\alpha(\tilde{n}'_\gamma, \tilde{n}_\gamma) O_\gamma^r(\tilde{n}_\gamma, n_\gamma) \quad (28)$$

where $\mathcal{A}_\alpha(\tilde{n}'_\gamma, \tilde{n}_\gamma)$ stand for the enlarged matrices: $\mathcal{S}_e(\sigma'_1, \tilde{n}'_s, \sigma''_2; \sigma''_1, \tilde{n}_s, \sigma_2)$, or $\mathcal{S}_o(\sigma''_1, \tilde{n}'_s, \sigma''_2; \sigma_1, \tilde{n}_s, \sigma_2)$, or $\mathcal{S}_e(\sigma'_2, \tilde{n}'_e, \sigma'_1; \sigma''_2, \tilde{n}_e, \sigma_2)$ for $\alpha = e$ or o , while $\mathcal{A}_\alpha(n'_\gamma, n_\gamma)$ with $n_\gamma, n'_\gamma \in [1, m]$ are the renormalized ones which are used to construct \mathcal{T}_M . An example for the renormalization of \mathcal{S}_e is shown in Fig. 4(b). The transformation matrices $O_\gamma^{l,r}$ are constructed from the m largest left and right eigenvectors of the reduced density matrices for the system ($\gamma = s$) or the environment ($\gamma = e$). The reduced density matrices are represented in terms of both left and right eigenvectors of \mathcal{T}_{M-1} and are discussed later on.

- The transfer matrix \mathcal{T}_M for the superblock is formally written as

$$\begin{aligned}\mathcal{T}_M(\sigma'_1, n'_e, \sigma'_2, n'_s; \sigma_1, n_e, \sigma_2, n_s) \\ = \begin{cases} \sum_{\sigma''_1, \sigma''_2} \mathcal{S}_o(\sigma''_1, n'_s, \sigma''_2; \sigma_1, n_s, \sigma_2) \times \mathcal{E}_o(\sigma'_2, n'_e, \sigma'_1; \sigma''_2, n_e, \sigma''_1), \\ \sum_{\sigma''_1, \sigma''_2} \mathcal{S}_e(\sigma''_1, n'_s, \sigma''_2; \sigma_1, n_s, \sigma_2) \times \mathcal{E}_e(\sigma'_2, n'_e, \sigma'_1; \sigma_2, n_e, \sigma_1) \end{cases} \quad (29)\end{aligned}$$

for M odd or even as shown in Fig. (3).

In practical computations, one does not have to construct T_M explicitly. Rather, instead of direct matrix-vector manipulations between T_M and $|\psi_{max}\rangle$, one can successively operate with $\mathcal{S}_{e,o}$ and $\mathcal{E}_{e,o}$ on $|\psi_{max}\rangle$. In this case, a high computational performance can be achieved by an efficient optimization [31]. Moreover, one can also easily make a parallelization with respect to the indices of the states for two spins (σ_1'' and σ_2'') connecting left and right transfer matrices.

Superblock II: An alternative construction of superblocks was proposed in [13] to improve the numerical performance and is shown in Fig. 5(a).

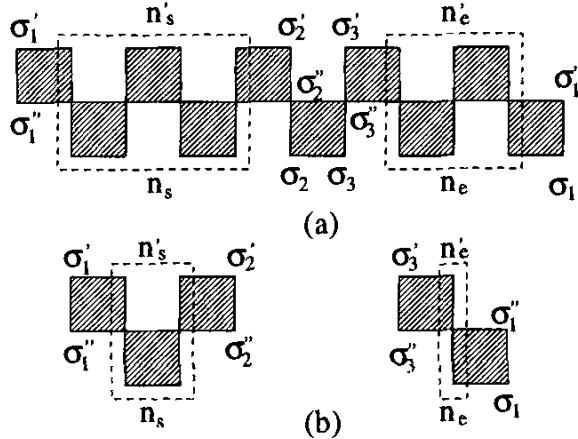


Fig. 5. An alternative construction of the superblock. (a) A superblock configuration for $M = 5$ is constructed in the basis $|\sigma_1\rangle \otimes |n_s\rangle \otimes |\sigma_2\rangle \otimes |\sigma_3\rangle \otimes |n_e\rangle$; (b) Initialization with $M = 3$.

- Initially, the (renormalized) block in the system contains two slices and the one in the environment does only one slice as shown in Fig. 5(b). The renormalized system and environment blocks are connected by summing over the three additional spins σ_1'' , σ_2'' and σ_3'' to form T_M as seen in Fig. 5(a).
- In each iteration, M is increased by one and the number of slices in the two renormalized blocks always differs by one. Because of this, only the larger renormalized block is enlarged and then renormalized. This is a kind of alternative renormalization procedure between the renormalized system and environment blocks as illustrated in Fig. 6. Moreover, this transfer matrix

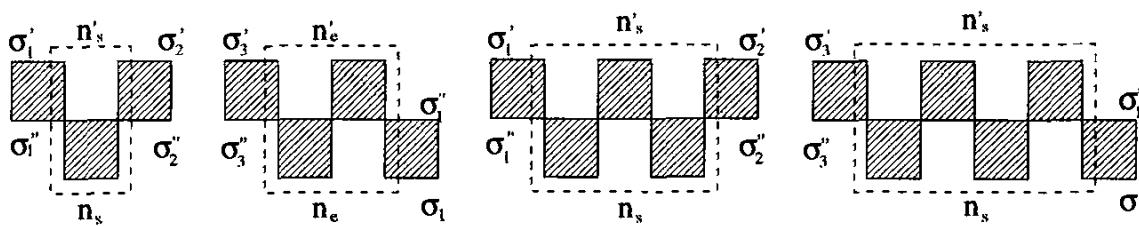


Fig. 6. An alternative renormalization procedure. From left to right: $\mathcal{S}_o(M = 3) \rightarrow \mathcal{E}_e(M = 4) \rightarrow \mathcal{S}_o(M = 5) \rightarrow \mathcal{E}_e(M = 6)$, The renormalized block on the left side of each arrow is enlarged and then renormalized to the one on the right side.

containing the larger renormalized block can also be directly taken as another transfer matrix for the construction of T_M in the next iteration, i.e. $\mathcal{S}_e(M) =$

$\mathcal{S}_o(M - 1)$ and $\mathcal{E}_o(M) = \mathcal{E}_e(M - 1)$. The corresponding bases are given by

$$|n_e(M)\rangle \leftarrow |\tilde{n}_e(M)\rangle = |n_s(M - 1)\rangle \otimes |\sigma\rangle \quad (30)$$

$$|n_s(M)\rangle = |n_s(M - 1)\rangle \quad (31)$$

for even M , and

$$|n_e(M)\rangle = |n_e(M - 1)\rangle \quad (32)$$

$$|n_s(M)\rangle \leftarrow |\tilde{n}_s(M)\rangle = |n_e(M - 1)\rangle \otimes |\sigma\rangle \quad (33)$$

for odd M . In the above expressions, the left arrows indicate renormalization.

The advantage of forming such a type of superblock is that the transfer matrix \mathcal{T}_M in this case can always be factorized into a product of two sparse matrices which are block diagonal with respect to $|n_s\rangle$ and $|\sigma_3\rangle \otimes |n_e\rangle$, respectively. To treat these two sparse matrices instead of \mathcal{T}_M itself allows one to save both computer memory space and CPU time.

Methods for calculating λ_{max} : There are several methods for computing λ_{max} and other extreme eigenvalues and eigenvectors of \mathcal{T}_M . The simplest one is the so-called power method. This method starts from an arbitrary initial vector $|\psi_0^R\rangle$ which is not orthogonal to $|\psi_{max}^R\rangle$. By applying \mathcal{T}_M iteratively to the vector which is generated in the previous iteration, i.e. $|\psi_K^R\rangle = \mathcal{T}_M|\psi_{K-1}^R\rangle$, it will finally project out $|\psi_{max}^R\rangle = |\psi_K^R\rangle$ for sufficiently large K . The value of K needed depends on the ratio between the second-largest eigenvalue λ_2 and the largest eigenvalue $\eta = |\lambda_2/\lambda_{max}|$. The smaller η is, the smaller K is needed. If δ is the relative error required for λ_{max} , then K required is of order $\ln \delta / \ln \eta$. For the spin one-half Heisenberg model, we find that the value of K needed for producing an eigenvalue with a relative error less than 10^{-16} is generally less than 20, but it increases with decreasing temperature since η or the correlation length increases with decreasing temperatures. The left eigenvector $\langle\psi_{max}^L|$ can be similarly determined for both constructions of the superblock. When a Hamiltonian has a reflection symmetry, one can obtain the left eigenvector directly from the right one by a symmetric construction of the superblock [7].

Other methods for determining λ_{max} include the implicitly restarted Arnoldi method and the look-ahead Lanczos method. These are more efficient than the power method and can be used to evaluate not only the largest eigenvalue but also the second-largest or other eigenvalues of \mathcal{T}_M .

Both the implicitly restarted Arnoldi method and the look-ahead Lanczos method are well developed and tested. The source codes for both can be down-loaded freely from the internet [30].

Reduced density matrix: The density matrix can generally be defined as a semi-positive matrix whose trace is equal to the partition function up

to a constant factor. For a Hamiltonian system, the thermodynamic density matrix is known to be

$$\rho_{th} = e^{-\beta H}. \quad (34)$$

The reduced density matrix for the system block is defined from ρ_{th} by integrating out all degrees of freedom of the environment block

$$\rho_{th,s} = \text{Tr}_e e^{-\beta H}, \quad (35)$$

where s and e denote the system and environment blocks, respectively. It is straightforward to show that the traces of ρ_{th} and $\rho_{th,s}$ are equal to the partition function Z . At zero temperature, $\beta \rightarrow \infty$, $\rho_{th,s}$ is equal to $\text{Tr}_e(|\psi_0\rangle\langle\psi_0|)$ up to a normalization constant, where $|\psi_0\rangle$ is the ground state eigenstate of the Hamiltonian.

In the Trotter space, the density matrix for the superblock can be defined as

$$\rho = \mathcal{T}_M^{N/2}. \quad (36)$$

This is an extension of the thermodynamic density matrix ρ_{th} in the Trotter space. Similarly, by integrating out all degrees of freedom of the environment, we define the reduced density matrix of the system block as

$$\rho_s = \text{Tr}_e \mathcal{T}_M^{N/2}. \quad (37)$$

In the thermodynamic limit, ρ_s is equal to $\text{Tr}_e(|\psi_{max}^R\rangle\langle\psi_{max}^L|)$ up to a normalization constant. This definition of the reduced density matrix is similar to that of $\rho_{th,s}$. But unlike $\rho_{th,s}$, ρ_s is non-symmetric. Physically, ρ_s should be a semi-positive matrix. If both T_1 and T_2 are symmetric and T_1 or T_2 is semi-positive, we can rigorously prove that all the eigenvalues of ρ_s are larger than or equal to zero. A general proof for the semi-positivity of ρ_s is still not available although we found numerically that the eigenvalues of ρ_s are semi-positive definite when the Trotter space is not truncated.

Suppose that ω_l is an eigenvalue of ρ_s and $|\phi_{s,l}^{R,L}\rangle$ are the corresponding right and left eigenvectors. Then one can write

$$\omega_l = \sum_j \langle\psi_{max}^L|\phi_{s,l}^R\rangle|\psi_{e,j}^R\rangle\langle\phi_{s,l}^L|\langle\psi_{e,j}^L|\psi_{max}^R\rangle, \quad (38)$$

where $\{|\psi_{e,j}^R\rangle, \langle\psi_{e,j}^L|\}$ is a complete set of basis vectors of the environment block. In obtaining (38), the normalization conditions, $\langle\phi_{s,l}^L|\phi_{s,l'}^R\rangle = \delta_{l,l'}$, $\langle\phi_{e,j}^L|\phi_{e,j'}^R\rangle = \delta_{j,j'}$, and $\langle\psi_{max}^L|\psi_{max}^R\rangle = 1$ are used. (38) indicates that ω_l is the projection probability of the state $|\phi_{s,l}^R\rangle\langle\phi_{s,l}^L|$ in $|\psi_{max}^R\rangle\langle\psi_{max}^L|$. Thus the optimal states which should be retained are the eigenvectors of the reduced density matrix ρ_s corresponding to the largest eigenvalues ω_l .

Variational principle: The zero-temperature DMRG method satisfies the variational principle because one is dealing with a hermitian operator H and works in a truncated Hilbert space. Thus the ground-state energy E_0 obtained using this method is always an upper bound for the true value, i.e. $E_0 \geq E_0^{\text{True}}$. The systematic error for the ground-state energy decreases as the number of retained states increases. The truncation error is generally smaller than the true systematic error of the result.

At finite temperature, the TMRG method also uses a small set of basis vectors to approximate the full Hilbert space. We expect that the value of λ_{\max} obtained by the TMRG is a lower bound to the true value, namely $\lambda_{\max} \leq \lambda_{\max}^{\text{True}}$. Substituting this inequality into (13), we then have

$$F(T) \geq F^{\text{True}}(T), \quad (39)$$

which can be taken as a generalized variational principle.

Numerical aspects

Reduced density matrix: In practical TMRG calculations, due to round-off errors and especially the truncation of the Trotter space, some eigenvalues of the reduced density matrices $\rho_{s,e}$ may become complex, which violates the semi-positivity of $\rho_{s,e}$. However, we found that the moduli of the complex eigenvalues are generally small and are supposed to be zero if no truncation was made. There exist some ways to cure this violation of the semi-positivity. Here we discuss an economic way which we have used extensively in the numerical calculations as follows.

First, since ρ_s is real, complex eigenvalues and corresponding eigenvectors always appear in pairs. For instance, let us consider

$$\begin{cases} \rho_s |\phi_{s,l}^R\rangle = \omega_l |\phi_{s,l}^R\rangle \\ \rho_s^T |\phi_{s,l}^L\rangle = \omega_l^* |\phi_{s,l}^L\rangle \end{cases}; \quad \begin{cases} \rho_s |\phi_{s,l+1}^R\rangle = \omega_{l+1} |\phi_{s,l+1}^R\rangle \\ \rho_s^T |\phi_{s,l+1}^L\rangle = \omega_{l+1}^* |\phi_{s,l+1}^L\rangle \end{cases}, \quad (40)$$

where the superscripts T and $*$ represent the transpose and complex conjugate, respectively. When $\omega_{l+1} = \omega_l^*$, thus $|\phi_{s,l+1}^{R,L}\rangle = |\phi_{s,l}^{R,L}\rangle^*$.

Secondly, as we expected, the imaginary part is usually negligible. Then $|\phi_{s,l+1}^{L,R}\rangle$ and $|\phi_{s,l}^{L,R}\rangle$ can be regarded as two “degenerate” left/right eigenvectors corresponding to $\bar{\omega}_l = \text{Re } \omega_l = \text{Re } \omega_{l+1}$. In this sense, we can combine these two complex eigenvectors $|\phi_{s,l+1}^{L,R}\rangle$ and $|\phi_{s,l}^{L,R}\rangle$ into two real ones $|\bar{\phi}_{s,l+1}^{L,R}\rangle$ and $|\bar{\phi}_{s,l}^{L,R}\rangle$.

Finally, ρ_s is block diagonalized by a standard library routine such as DGEEV from the Lapack subroutines. We note that the complex eigenvector pair $|\phi_{s,l}^{R,L}\rangle$ output by the subroutine is generally normalized to a *complex number*, i.e.

$$\langle \phi_{s,l}^L | \phi_{s,l}^R \rangle = \langle \phi_{s,l+1}^L | \phi_{s,l+1}^R \rangle = e^{i\alpha}. \quad (41)$$

We simply choose

$$\begin{cases} \bar{\phi}_{s,l}^L = \operatorname{Re} \phi_{s,l}^L \\ \bar{\phi}_{s,l+1}^L = \operatorname{Im} \phi_{s,l}^L \end{cases}; \quad \begin{cases} \bar{\phi}_{s,l}^R = \operatorname{Re} e^{-i\alpha} \phi_{s,l}^R \\ \bar{\phi}_{s,l+1}^R = \operatorname{Im} e^{-i\alpha} \phi_{s,l}^R \end{cases}. \quad (42)$$

This procedure does not suffer from round-off errors destroying the orthogonality between the right and left basis of ρ_s and has been used for (frustrated) spin as well as electron systems. The above discussions for ρ_s are also valid for ρ_e which is defined in a similar way and has to be used for the first construction of T_M shown in Fig. 3.

Comparison to exact results: The spin one-half Heisenberg model in one dimension is integrable by Bethe ansatz. Many of its thermodynamic properties, for example the specific heat and the spin susceptibility, can be evaluated by solving the Bethe-ansatz equations. This model therefore provides a good opportunity for testing the TMRG method.

Figures 7 and 8 show the TMRG results for the specific heat C_v/T , the spin susceptibility χ , and the correlation length ξ of the spin one-half Heisenberg antiferromagnet at zero temperature. The exact results are also shown in these figures for comparison. The agreement between the TMRG results

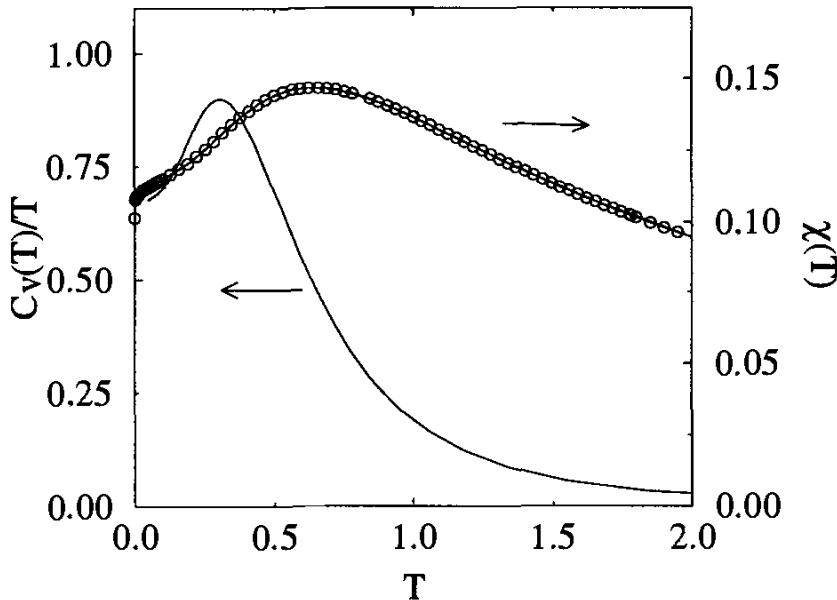


Fig. 7. Zero-field TMRG results for C_v/T and χ of the spin one-half Heisenberg model, calculated with $\epsilon = 0.1$ and $m = 81$. The circles are the exact results.

and the exact ones is apparently very good. For the results shown in these figures, $m = 81$ and $\epsilon = 0.1$ are used. The errors of the TMRG results are less than 10^{-4} even at very low temperatures. More accurate results can be obtained if more states are retained or if a systematical extrapolation with respect to both m and ϵ is taken.

A more detailed discussion in this respect can be found in [13].

Error analysis: In the TMRG calculation, we always deal with spatially infinite systems. This is different from zero-temperature and classical-system

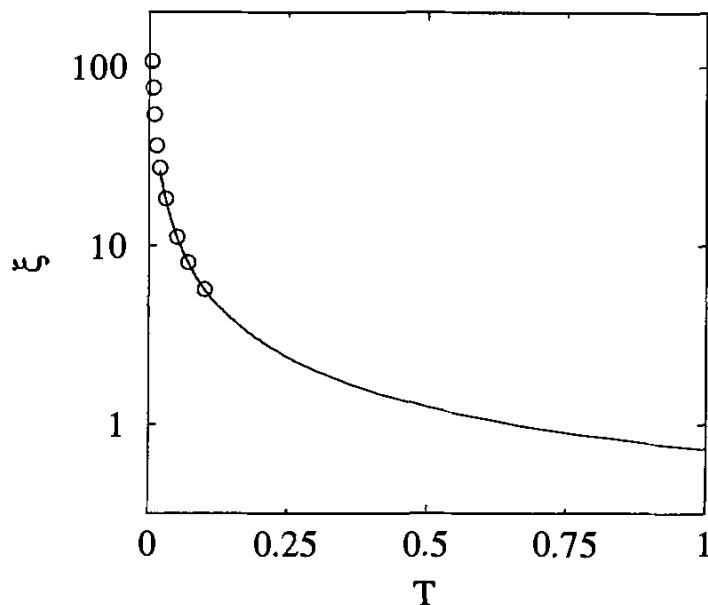


Fig. 8. TMRG result for the thermal correlation length of the spin one-half Heisenberg model at zero field, calculated with $\epsilon = 0.1$ and $m = 81$. The circles are thermal Bethe-ansatz results.

calculations. On the other hand, in addition to the error resulting from the basis truncation, there is a systematic error caused by the Trotter-Suzuki decomposition.

This error as shown in (4), is of the order ϵ^2 . At high temperatures, or more precisely when the size M of superblock is large, the truncation error is very small, the error comes only from the Trotter-Suzuki approximation and varies as ϵ^2 for fixed m as shown in Fig. 9. At low temperature, the accumulated truncation error becomes important and the total error may not decrease with decreasing ϵ . This is also shown in Fig. 9.

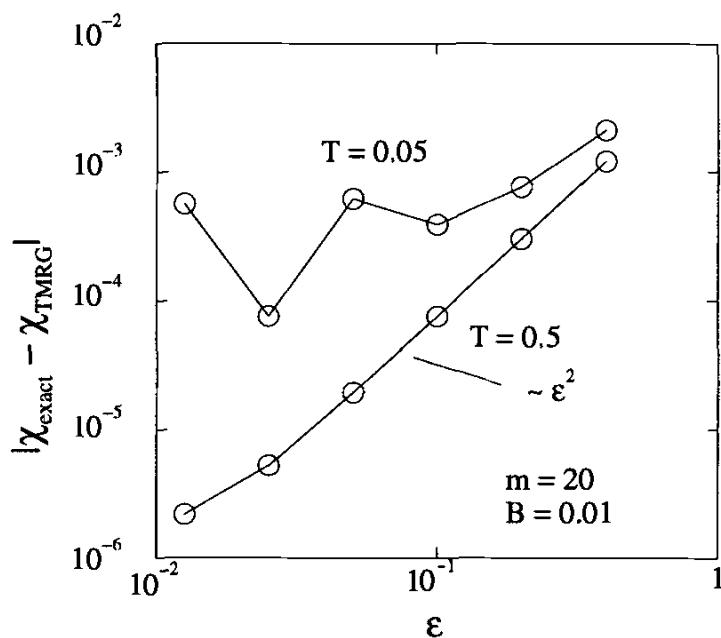


Fig. 9. Error in the susceptibility as a function of the parameter ϵ at two different temperatures.

In Fig. 10, the error is shown as a function of m . Given ϵ , we find that the systematical errors drop exponentially with increasing m initially and reach a value which is only determined by the Trotter-Suzuki approximation when m becomes large enough. This is similar to the zero-temperature DMRG method. Physically this is because at finite temperatures the correla-

tion length is always finite and there is a gap between the largest eigenvalue and the next largest eigenvalues of the transfer matrix T_M . If we assume that the ratio between the largest eigenvalue and the next largest eigenvalues of T_M is $\exp(\Delta_T)$, the error is expected to drop as $\exp(-\Delta_T m)$. Since Δ_T decreases with decreasing temperature, the systematical error drops faster at high temperature than at low temperature.

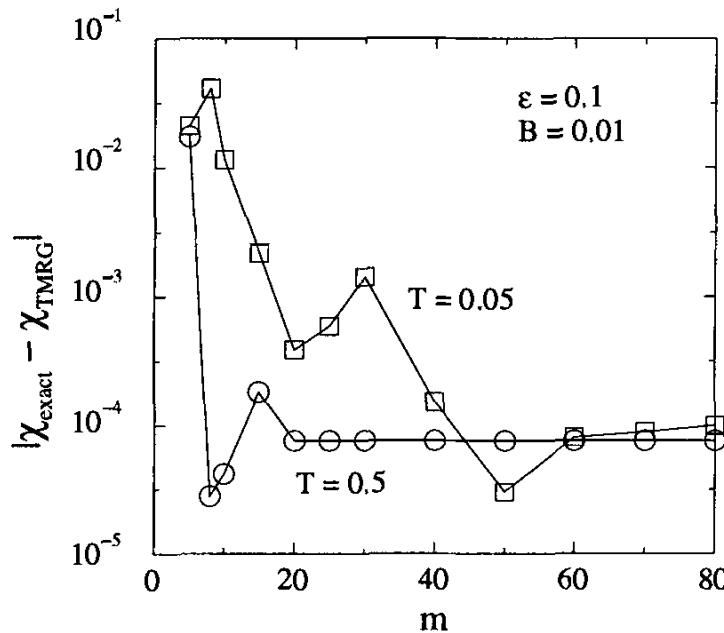


Fig. 10. Error in the susceptibility as a function of the number m of kept states at two different temperatures.

From the above discussion, we see that reducing the value of ϵ (or equivalently increasing the initial temperature at $2M = 4$) would reduce the error of the Trotter-Suzuki approximation, but the accumulated truncation error at low temperature will become large. Thus simply choosing a very small ϵ in the calculation would not improve the results at low temperatures.

The truncation errors for the second-largest eigenvalue of T_M are generally larger than those of λ_{max} . Thus the correlation lengths determined from (26) are generally expected to be less accurate than the free energy or other thermodynamic quantities.

2 Momentum-Space DMRG

The standard DMRG was originally proposed to treat quantum lattice models with short-range interactions in real space. It has been used successfully to investigate the low-energy properties of many quasi one-dimensional systems. However, its application in 2D is still not satisfactory. This has been the motivation for the extension of the DMRG method to momentum space. In principle, the standard DMRG can be applied directly in momentum space. However, as the interactions are non-diagonal in the momentum space and the DMRG iterations involve operator manipulations of the order N^3 , a straightforward implementation is not possible for realistic calculations. In order to resolve this difficulty, so-called composite operators were introduced in [21].

Then the number of operators whose matrix elements are stored and updated is reduced from the order N^3 to the order N . This makes the DMRG iteration in momentum space practically possible. Since momentum is now a good quantum number, the matrices becomes more sparse and many more states can be kept in momentum space than in the real space.

In this section, we first use the Hubbard model to demonstrate how to apply the DMRG in momentum space. Later we will discuss how to extend the discussions to other interacting electron systems.

DMRG implementation

The Hubbard model is defined by the Hamiltonian

$$H = -t \sum_{\langle ij \rangle \sigma} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow}, \quad (43)$$

where $\langle ij \rangle$ means summation over nearest neighbors. In momentum space, this model is

$$H = \sum_{k\sigma} \epsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \frac{U}{N} \sum_{k_1, k_2, k_3} c_{k_1\uparrow}^\dagger c_{k_2\uparrow} c_{k_3\downarrow}^\dagger c_{k_1-k_2+k_3\downarrow}, \quad (44)$$

where ϵ_k is the energy dispersion of the electrons and periodic boundary conditions are assumed. With each point in k -space one can associate four degrees of freedom, namely the states $\{ |0\rangle, c_{k\uparrow}^\dagger |0\rangle, c_{k\downarrow}^\dagger |0\rangle, c_{k\uparrow}^\dagger c_{k\downarrow}^\dagger |0\rangle \}$, where $|0\rangle$ denotes the vacuum. A k -point with these four states can be taken as the basic unit. However, in calculations it is more convenient to treat the spin degree of freedom as an extra spatial coordinate and to consider a momentum-spin point $(k\sigma)$, which has only two degrees of freedom $\{|0\rangle, c_{k\sigma}^\dagger |0\rangle\}$, as a basic unit site. In this case, the spin rotation symmetry is not manifest, but the total number of degrees of freedom of a superblock is reduced and more states can be retained in the truncation of the Hilbert space. This can reduce truncation errors and save computer time.

As in real space, we define a superblock as an assembly of a system block A , an environment block B , and two unit sites $(k_1\sigma_1)$ and $(k_2\sigma_2)$. To carry out the DMRG iteration, we need to order the $(k\sigma)$ points. As all $(k\sigma)$ points are connected by interactions, the ordering of $(k\sigma)$ is not as important as in real space. However, the rate of convergence for the ground-state energy does depend on the ordering. A rule of thumb for the ordering of the $(k\sigma)$ points is that those $(k\sigma)$ pairs which have the strongest interaction should be arranged as close as possible in the ordered $(k\sigma)$ chain. In the Hubbard model, however, the interaction strength does not depend on k and one may choose any ordering.

The Hubbard interaction is local in real space but non-local in momentum space — it contains terms which link two or more k points in momentum space. The summation in the second term of (44) contains N^3 terms. To evaluate the matrix elements of the Hamiltonian, we need to evaluate the matrix

elements of these N^3 terms. If one treated them separately, the calculation would be very difficult, if not impossible. However, after a careful analysis, one finds that the three-fold summations in the second term of H can actually be reduced to a summation of some terms, whose number is the order of N . To do this, one has to define the following composite operators in block A

$$\begin{aligned} a_0(p\sigma) &= c_{p\sigma}\delta_{(p\sigma)\in A}, \\ a_1(p\sigma) &= \sum_q a_0^\dagger(q\sigma)a_0(p+q\sigma), \\ a_2(p) &= \sum_q a_0^\dagger(q\uparrow)a_0(p+q\downarrow), \\ a_3(p\sigma) &= \sum_{q_1q_2} a_0^\dagger(q_1\bar{\sigma})a_0(q_2\bar{\sigma})a_0(p+q_1-q_2\sigma)\delta_{(p\sigma)\notin A}, \\ a_4(p) &= \sum_q a_0(q\downarrow)a_0(p-q\uparrow), \end{aligned} \quad (45)$$

where $\bar{\sigma} = -\sigma$. In block B , the corresponding operators can be obtained by changing (a, A) into (b, B) . In each block, the total number of composite operators is $6N$.

To simplify the discussion below, we define $A' = A \oplus (k_1\sigma_1)$ and $B' = B \oplus (k_2\sigma_2)$. The Hamiltonian can formally be written as $H = H_{A'} + H_{B'} + H_{A'B'}$, where $H_{A'}$ (or $H_{B'}$) contains all interactions among electrons belonging to A' (or B'), and $H_{A'B'}$ represents the interaction between A' and B' . $H_{A'}$ can be iteratively obtained from H_A through the equation

$$H_{A'} = H_A + n_{k\sigma}\epsilon_k + \frac{U}{N} \left[n_{k\sigma}a_1(0, \bar{\sigma}) + c_{k\sigma}^\dagger a_3(k\sigma) + a_3^\dagger(k\sigma)c_{k\sigma} \right]. \quad (46)$$

In a DMRG iteration, we can take $H_{A'}$ (similarly $H_{B'}$) as a single operator to be stored and updated. The composite operators in A' can also be obtained from those in A via

$$\begin{aligned} a'_1(p\sigma') &= a_1(p\sigma') + \delta_{\sigma,\sigma'} \left[a_0^\dagger(k-p\sigma)c_{k\sigma} + c_{k\sigma}^\dagger a_0(k+p\sigma) + n_{k\sigma}\delta_{p,0} \right], \\ a'_2(p) &= a_2(p) + \delta_{\sigma,\uparrow}c_{k\uparrow}^\dagger a_0(k+p\downarrow) + \delta_{\sigma,\downarrow}a_0^\dagger(k-p,\uparrow)c_{k\downarrow}, \\ a'_3(p\sigma') &= a_3(p\sigma') + \delta_{\sigma,\sigma'}a_1(p-k\bar{\sigma})c_{k\sigma} \\ &\quad + \delta_{\sigma',\bar{\sigma}} \left[n_{k\sigma}a_0(p\bar{\sigma}) - \sigma c_{k\sigma}^\dagger a_4(p+k) - c_{k\sigma}\tilde{a}_2(p-k) \right], \\ a'_4(p) &= a_4(p) + \sigma a_0(p-k,\bar{\sigma})c_{k\sigma}, \end{aligned}$$

where $\tilde{a}_2(p) = a_2(p)$ if $\sigma = \uparrow$ and $a_2^\dagger(p)$ if $\sigma = \downarrow$. With the above definitions we can show that

$$\begin{aligned} H_{A'B'} &= \frac{U}{N} \sum_p \left\{ \sum_{\sigma'} \left[\frac{1}{4} a'_1(p\sigma') b'_1(-p\bar{\sigma}') + b'^\dagger_0(p\sigma') a'_3(p\sigma') \right. \right. \\ &\quad \left. \left. + (a' \longleftrightarrow b') \right] + a'^\dagger_4(p) b'_4(p) - b'_2(p) a'^\dagger_2(p) \right\} + h.c.. \end{aligned} \quad (47)$$

Thus, in the DMRG iterations, only the matrix elements of H_A , H_B and the composite operators in both subblocks need to be evaluated and stored.

The truncation of the Hilbert space in the DMRG procedure may cause the loss of the symmetries of the problem. For example, in real space DMRG, the momentum conservation is lost. In momentum space, however, this symmetry is preserved. Combined with the conservation of the number of up spins N_\uparrow and the number of down spins N_\downarrow , the momentum conservation can be used to block-diagonalize the Hamiltonian. This does not only save computer time, but also allows one to keep many more eigenstates in the truncation as the number of non-zero matrix elements is now significantly reduced. Clearly the basis states in each block can be classified by three quantum numbers ($N_\uparrow, N_\downarrow, P$), where P is the total momentum. For the composite operators defined in (45) it can be shown that their matrix elements are non-zero only when the difference between the momenta of the initial and the final state is p .

In the DMRG method, the basis states for both the system A and environment B are incomplete. If the interaction between A and B , i.e. H_{AB} , is non-zero, the matrix elements of H_{AB} will be less accurately approximated than the terms defined just in each subblock after the truncation of Hilbert space. In a one-dimensional-space calculation, one can choose A and B so that no interactions exist between A and B , i.e. $H_{AB} = 0$, if the interaction is short ranged. In real space in 2D or in momentum space in any dimension, however, the interaction between A and B is always non-zero, no matter how A and B are constructed. Thus in general results obtained by the DMRG method in these cases will not be as accurate as, for example, the ground-state energy of the 1D spin-one Heisenberg model [32].

Initialization

In momentum space, lattices of different size have different k points. Thus the size of the lattice has to be fixed at the beginning. This means that only the finite-system algorithm of the DMRG can be used in momentum space. To use the finite-size approach, however, one has to first build up a series of initial system and environment blocks. Since one cannot use the standard infinite-system algorithm, one has to look for a different procedure. A simple way is to use the conventional renormalization-group method [3,4]. The initial blocks C_n for system ($C = A$) as well as environment ($C = B$) are built up as follows:

1. Start from a small block C_1 , which can be handled without truncation of basis states.
2. Add a new ($k\sigma$) point to C_1 to form a new system block C_2 .
3. Diagonalize the Hamiltonian in the Hilbert space spanned by C_2 .
4. Truncate the Hilbert space to retain the m lowest energy eigenstates.
5. Replace C_2 by C_1 and repeat the steps 2-5 until all the initial system blocks required are established.

All $(N_\uparrow, N_\downarrow, P)$ subspaces that a block can have, are determined solely by the $(k\sigma)$ points the block contains. In the above initialization procedure, not all $(N_\uparrow, N_\downarrow, P)$ subspaces in a block need be considered. If a subspace with the quantum numbers $(N_{1,\uparrow}, N_{1,\downarrow}, P_1)$ in a system (environment) block cannot be matched to a subspace with quantum numbers $(N_{2,\uparrow}, N_{2,\downarrow}, P_2)$ in the corresponding environment (system) block such that $(N_{1,\uparrow} + N_{2,\uparrow}, N_{1,\downarrow} + N_{2,\downarrow}, P_1 + P_2)$ is equal to the quantum numbers of the targeted ground state, then this $(N_{1,\uparrow}, N_{1,\downarrow}, P_1)$ subspace in the system (environment) block will have no contribution to the ground state and can be discarded.

After the above initialization step, one can then use the finite-system algorithm of the DMRG method to find out the eigenvalue and eigenvector of the ground state for a given filling factor and momentum.

Results

In the following, we discuss the performance of the momentum-space DMRG comparing some preliminary results, for the Hubbard model, with those results from real-space DMRG results in one dimension and from various calculation in two dimensions.

Table 1 compares the ground-state energies obtained by the DMRG in momentum space as well as in real space at half-filling for 16 sites. For a

Table 1. Ground-state energies for a 16-sites Hubbard chain at half-filling, calculated by DMRG in momentum space and in real space

m	Momentum space		Real space	
	U=1	U=4	U=1	U=4
400	-1.02925	-0.51316	-1.02958	-0.575896
600	-1.02944	-0.53574	-1.02969	-0.575900
800	-1.02952	-0.53724	-1.02972	-0.575901
1000	-1.02958	-0.54562		
1200	-1.02959	-0.55218		

finite chain, no exact diagonalization results are available. When 800 states were kept in the real-space calculation, absolute errors for the ground states were found to be of the order of 10^{-6} for $U = 4$ as examined by varying the number of kept states. Relative to this results, an error for the ground-state energy obtained by the momentum-space DMRG with $m = 1000$ is 5% when $U = 4$. When U is not too small, we found that the ground-state energy is lower in real space than in momentum space if the same number of states is kept. This means that for the Hubbard model the DMRG method works better in real space than in momentum space in one dimension. This is not surprising because the Hubbard interaction is a local interaction and at half filling the electrons tend to be localized in space.

Table 2 compares the momentum-space results with those from the exact diagonalization and various approaches such as quantum Monte-Carlo calculations (QMC), exact cluster diagonalization (CD) and stochastic diagonalization (SD) for square lattices. The largest lattice calculated by using momentum-space DMRG had 12×12 sites. Comparison with the exact results for the 4×4 lattice shows that the DMRG results are very accurate when U is small. The relative error for the DMRG result with $U = 2$ is 3×10^{-4} . The

Table 2. Ground-state energies for Hubbard square lattices obtained by various methods: DMRG ($m=1000$), Exact (4×4), CD (6×6), QMC and SD with N electrons.

$L_x \times L_y$	U	N	Exact [33]/CD [34]	DMRG	QMC [35]	SD [35]
4×4	2	16	-18.01757	-18.012		
4×4	4	14	-15.74459	-15.673		
4×4	4	16	-13.62185	-13.571	-13.6	-13.59
4×4	8	16	-8.46887	-8.263	-8.48	
6×6	4	26	41.49	41.108	41.98	40.77
8×8	4	10		-34.325	-34.3	-34.31
8×8	4	18		-54.394	-54.6	-54.37
8×8	4	26		-66.098	-66.8	-66.05
12×12	4	18		-64.107		

momentum-space DMRG works better in the weak-coupling limit because the electron basis states used here are plane waves. When $U=0$, the ground state is a filled Fermi sea of non-interacting electrons, and the momentum-space method gives the exact ground-state energy even if only one state is kept. For large U , the DMRG results are not as good as in the weak-coupling limit. However, they are still comparable with those obtained by the quantum Monte Carlo methods. Both the DMRG and stochastic diagonalization methods satisfy the variational principle while the projected quantum Monte Carlo does not satisfy it due to the importance of sampling. For an 8×8 lattice, the ground-state energies obtained by the DMRG method are systematically lower (hence better) than the stochastic diagonalization results.

As shown above, the relative error for the ground-state energy obtained by the momentum-space DMRG with $m = 1000$ is 5%, when $U = 4$ and the chain length is 16. The value of this error is higher than that in two dimensions with the same value of U and lattice size. Therefore one can expect that the momentum-space DMRG works better for the Hubbard model in two dimensions than one. Physically this is because the relative contribution of the kinetic energy term to the ground-state energy, which is treated rigorously in the momentum-space DMRG method, compared with the potential energy, is larger in two dimensions than in one.