



Original Article

Impact of Parallel Computing on Study of Time Evolution of a Quantum Impurity System in Response to a Quench

Nghiem Thi Minh Hoa^{1,2,*}, Dang The Hung^{1,3}, Luong Minh Tuan⁴,
Duong Xuan Nui⁵, Nguyen Duc Trung Kien⁶

¹*PHENIKAA Institute for Advanced Study, PHENIKAA University, Ha Dong, Hanoi, Vietnam*

²*Faculty of Basic Science, PHENIKAA University, Ha Dong, Hanoi, Vietnam*

³*Faculty of Materials Science and Engineering, PHENIKAA University, Ha Dong, Hanoi, Vietnam*

⁴*National University of Civil Engineering, Dong Tam, Hai Ba Trung, Hanoi, Vietnam*

⁵*Vietnam National University of Forestry, Xuan Mai, Chuong My, Hanoi, Vietnam*

⁶*Advanced Institute for Science and Technology, HUST, Bach Khoa, Hai Ba Trung, Hanoi, Vietnam*

Received 11 January 2020

Revised 19 February 2020; Accepted 25 February 2020

Abstract: In an arbitrary system subjected to a quench or an external field that varies the system parameters, the degrees of freedom increases double in comparison to that of an isolated system. In this study, we consider the quantum impurity system subjected to a quench, and measure the corresponding time-evolution of the spectral function, which is originated from the time-resolved photoemission spectroscopy. Due to the large number of degrees of freedom, the expression of the time-dependent spectral function is twice much more complicated than that of the time-independent spectral function, and therefore the calculation is extremely time consuming. In this paper, we estimate the scale of time consumption of such calculation in comparison to that of time-independent calculation, and present our solution to the problem by using parallel computing as implementing both MPI and OpenMP to the calculation. We also discuss the possibility to exploit parallel computing with GPU in the near future, and the preliminary results of time-dependent spectral function.

Keywords: Quantum impurity system, time-dependent spectral function, degrees of freedom, parallel computing, OpenMP, GPU.

1. Introduction

Numerical methods have a great impact on studies of strongly correlated condensed matter systems, where the strong Coulomb interaction between electrons cannot be treated by perturbation

*Corresponding author.

Email address: hoa.ngkiemthiminh@phenikaa-uni.edu.vn

<https://doi.org/10.25073/2588-1124/vnumap.4453>

method. For example, the well-known Kondo effect was shown in the 60s that the first order perturbation gives the wrong ground state [1], while the calculation up to the second order gives the unphysical diverse resistance at low temperature [2], i.e. Kondo problem. And this problem was not solved fully until the study with the numerical renormalization group (NRG) method [3]. Studies of strongly correlated systems now grow diversely into many topics: finding an exotic Kondo effect in certain actinide/lanthanide ions in metal [4], keeping a topological phase by using the spin-orbit coupling [5], and tracking the time revolution of systems as well as finding the nonequilibrium steady-state when systems are subjected to external field [6]. In the studies, a large number of degrees of freedom are involved, serial numerical calculating may take an infeasible long computing-time. Parallel computing is the answer this problem, where a big calculation is divided into many smaller jobs and calculating these jobs is done in parallel. The application programming interfaces created for parallel computers are classified by the assumption they make about the underlying memory architecture: shared memory and distributed memory. While Open Multi-Processing (OpenMP) is the most used in the class of shared-memory, Message Processing Interface (MPI) is the most used in the class of distributed memory.

In this paper, we present a case study showing the impact parallel computing by solving the numerical problem in the time evolution of a strongly correlated impurity system as being subjected to a quench. The outline of the paper is as follows. In Sec. II., we describe the model and the time-dependent NRG formalism to study the time evolution of quantum impurity system following a quench. In Sec. III., we present the numerical problem in calculating the time-dependent spectral function of the impurity system, and the solution by using parallel computing with OpenMP and MPI. In Sec. IV., the success of using parallel computing is shown via the trend of decreasing time-consumption as the number of threads increase in two different Central Processing Units (CPUs), and the comparison between the speedup of real calculations and the prediction by Amdahl's law. From these results, we discuss of the possible use of GPU to accelerate calculations. The time-evolution of the impurity system is represented via the time-dependent spectral function in Sec. V.. The conclusion and outlook are presented in Sec. VI.

2. Model and formalism

2.1. Model

To describe the quantum impurity system subjected to a quench, we consider the following time-dependent Hamiltonian

$$\begin{aligned}
 H(t) = & \sum_{\sigma} \varepsilon_d(t) n_{d\sigma} + U(t) n_{d\uparrow} n_{d\downarrow} \\
 & + \sum_{k\sigma} \varepsilon_k c_{k\sigma}^+ c_{k\sigma} + \sum_{k\sigma} V (c_{k\sigma}^+ d_{\sigma} + d_{\sigma}^+ c_{k\sigma})
 \end{aligned}
 \tag{1}$$

where the quench at time $t=0$ is represented via the change of the local energy level $\varepsilon_d(t) = \theta(-t)\varepsilon_i + \theta(t)\varepsilon_f$ and the Coulomb interaction $U(t) = \theta(-t)U_i + \theta(t)U_f$. $n_{d\sigma} = d_{\sigma}^+ d_{\sigma}$ is the number operator for local electron with spin σ , and ε_k is the kinetic energy of the conduction electrons with constant density of states $\rho(\omega) = \sum \delta(\omega - \varepsilon_k) = 1/2D$ with $D=1$ the half-bandwidth.

The time evolution of the system can be well represented via the time-dependent spectral function, since it exhibits the probability of finding an electron at as a specified energy and time. However, the

time-dependent spectral function involves more degrees of freedom than its time-independent counterpart, one cannot define it easily via Lehmann representation. Therefore, one should define the time-dependent spectral function based on experimental observations. In this paper, we consider the spectral function originated on the time-resolved spectroscopy with the pump-probe technique [7, 8], in which the photoemission-current intensity takes the form

$$I(E, t_{\text{delay}}) = \iint d\omega dt N(E + \omega) e^{-\frac{t^2}{\Delta t^2}} e^{-\omega^2 \Delta t^2} \quad (2)$$

where the probe-pulse shape is taken to be Gaussian, the pulse width is Δt , t_{delay} is the time delay between pump and probe pulses, and the time-dependent spectral function of interest is derived from the lesser Green's function that

$$N(\omega, t) = \int d\tau G^<(t + \frac{\tau}{2}, t - \frac{\tau}{2}) e^{-i\omega\tau} \quad (3)$$

with $G^<(t_1, t_2) = i\langle d^+(t_1), d(t_2) \rangle$, $t_1 = t + \frac{\tau}{2}$ and $t_2 = t - \frac{\tau}{2}$. In this study, we will calculate the time-dependent spectral function, which measures the time-evolution of the occupied density of states.

2.2. Formalism

Using the time-dependent numerical renormalization group (TDNRG) method [9], we have the expression of $N(\omega, t)$ as follows

$$N(\omega, t > 0) = \frac{1}{2\pi i} \times \left(\begin{aligned} & \sum_{m=m_0}^N \sum_{rsq} C_{rs}^m B_{sq}^m \frac{e^{-i(E_q^m - E_r^m)t} - e^{-2i(\omega + E_s^m - E_q^m)t} e^{-2\eta t}}{\omega + E_s^m - \frac{E_q^m + E_r^m}{2} - i\eta} \rho_{rs}^{i \rightarrow f}(m) \\ & - \sum_{m=m_0}^N \sum_{rsq} C_{rs}^m B_{sq}^m \frac{e^{-i(E_q^m - E_r^m)t} - e^{2i(\omega + E_s^m - E_r^m)t} e^{-2\eta t}}{\omega + E_s^m - \frac{E_q^m + E_r^m}{2} + i\eta} \rho_{rs}^{i \rightarrow f}(m) \\ & + \sum_{m=m_0}^N \sum_{rs\eta_1 s_1} C_{rs}^m e^{-2i(\omega - E_r^m + E_s^m)t} e^{-2\eta t} \frac{S_{s\eta_1}^m \sum_q B_{s_1 q}^m \tilde{R}_{q\eta_1}^m S_{\eta_1 r}^m}{\omega - \frac{E_r^m - E_s^m + E_{\eta_1}^m - E_{s_1}^m}{2} - i\eta} \\ & - \sum_{m=m_0}^N \sum_{rs\eta_1 s_1} B_{rs}^m e^{2i(\omega + E_r^m - E_s^m)t} e^{-2\eta t} \frac{S_{s\eta_1}^m \sum_q \tilde{R}_{s_1 q}^m C_{q\eta_1}^m S_{\eta_1 r}^m}{\omega + \frac{E_r^m - E_s^m + E_{\eta_1}^m - E_{s_1}^m}{2} + i\eta} \end{aligned} \right) \quad (4)$$

where $C \equiv d^+$, $B \equiv d$, the matrix elements $C_{rs}^m, B_{rs}^m, E_r^m, S_{sq}^m, \tilde{R}_{rs}^m$, and $\rho_{rs}^{i \rightarrow f}(m)$ are known from the NRG calculations, and η is a positive infinitesimal. For the detail derivation of the expression, we refer readers to our papers [10, 11].

3. Parallel computing

In the last section, we show the time-dependent spectral function originated from the time-resolved photoemission spectroscopy. The calculation for this time-dependent observable is challenging. In the last two terms, since all the four indices r, s, r_1 , and s_1 appear in the denominator, one cannot rewrite the summation over four indices as matrix multiplications for efficient evaluation with BLAS routine. Therefore, one should run all the four loops all together to calculate this expression.

In a specified calculation, the time consumption to calculate the first two terms with three loops in Eq. (4) is 100~200 times faster than that to calculate the last two terms with four loops. While, the trivial time-independent spectral function only involves two loops since the summation over three indices there can normally be recast as matrix multiplications [12, 13], and such calculations only take the time scale of minutes depending on computing systems. With that reference to the time-independent spectral function, calculating the time-dependent spectral function presented in Sec. II., is extremely heavy, and the serial computing is not sufficient.

Parallel computing is the answer the above problem. Two classes of parallel computing are considered in our study: shared memory with Open Multi-Processing (OpenMP) and distributed memory with Message Processing Interface (MPI). In a parallel computing with MPI, every parallel processes works in its own memory space, which is independent from the others. Passing messages between processed is required to transfer data. While, in a parallel computing with OpenMP, parallel computing occurs on every threads, which are able to access to the shared memory. Therefore, different from MPI, OpenMP does not require the overhead of message passing.

In our study, we use the hybrid parallel computing with both shared and distributed memory. The parallel computing with distributed memory is for the two NRG calculations for the matrix elements $C_{rs}^m, B_{rs}^m, E_r^m$, and \tilde{R}_{rs}^m , of two independent Hamiltonian H^i and H^f , which are stored separately in two different processes. Message passing is done to transfer the matrix elements between processes in order to calculate $\rho_{rs}^{i \rightarrow f}(m)$ and S_{sq}^m , which they represent the projection of initial states and density matrices of H^i into the final states of H^f . The parallel computing with shared memory is for the summation with four loops in which the large sum is divided into many smaller jobs. The small jobs are processed in the individual threads independently while the memory is shared among the threads.

4. Speedup

4.1. Time consumption vs. number of threads

As presented in the last section, the use of OpenMP is applied to the summation over four indices in Eq. (4). In this section, we show the efficiency of parallel computing via the trend of time-consumption decreasing with an increasing number of threads. The calculations were done on two different computing systems. In the first system, one node is with two Intel Xeon E5-2680 v3 Haswell CPUs. In each node, there are 24 physical cores, and 48 logical threads thanks to the hyper-threading with folding of two. In the second system, one node is with one Intel Xeon Phi 7250-F Knights Landing CPU. The number of physical cores in each node is 68, and, with the hyper-threading with folding of four, therefore the number of logical threads is 272. The CPU clock is 2.5GHz in the first system, and 1.4GHz in the second system.

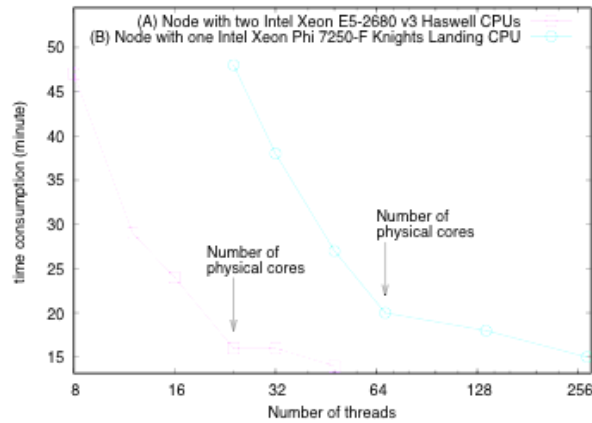


Figure 1. Time consumption of calculation vs. the number of threads in two different types of CPUs.

Figure 1. shows the time-consumptions of the same calculations with one node in each system and with the different number of threads. The decrease of time-consumption with the increasing number of threads is smooth up to the number of physical cores, while running on the further logical threads show a slower decrease of time consumption. The trend is similar in both calculations on the two systems. Besides, even though there are more threads in the KNL CPU than in the Haswell CPU, the CPU clock of KNL is slower than that of Haswell. Therefore, the total time-consumptions of calculations in one single node of each system with the maximum number of threads are similar.

4.2. Amdahl's law

In parallel computing, Amdahl's law predicts the speedup in latency of the execution of a task at fixed workload as follows [14]

$$S_{\text{latency}} = \frac{1}{(1-p) + \frac{p}{s}} \quad (5)$$

In words, it depends on the proportion of execution time that the part benefiting from parallel computing originally occupies, p , and the speedup of that part. If we assume the speedup ideally equals to the number of physical threads, we can predict, with a known value of p , the ideal speedup of a calculation.

Figure 2. shows the prediction of speedup by Amdahl's law and the speedup of real calculations with $p=99.3\%$, which means for every 1000 minutes to calculate the whole workload there are 993 minutes to calculate serially the part benefiting from parallel computing. We can see up to the number of physical core, the speedup of real calculation matches perfectly to the prediction by Amdahl's law. The speedup of real calculations as increasing further the number of threads deviates from the ideal speedup. It is due to the fact of using the logical threads; the speedup does not increase linearly with the number of threads.

However, the parallel computing with OpenMP can only use up to the maximum number of threads in a single node, which is limited, 48 in Haswell CPU and 272 in KNL CPU. While, from the prediction of Amdahl's law, the calculation with large number of proportion benefiting from parallel computing can be even speedup further if the number of threads are more than 1000. Therefore, using the Graphic Processing Unit (GPU) with a large number of cores up to thousands can be the future to our calculation.

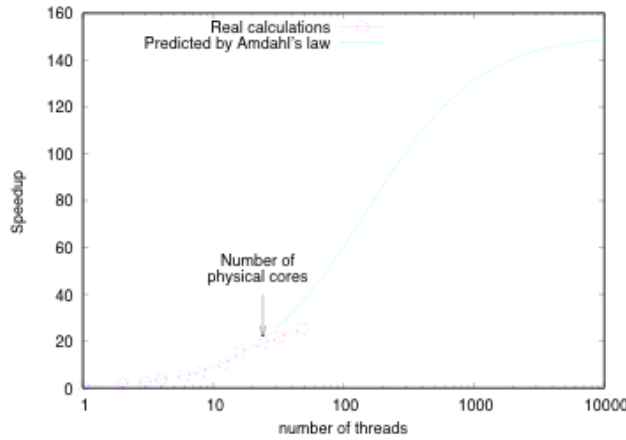


Figure 2. Speedup predicted by Amdahl's law and speedup of real calculations on Haswell CPUs.

5. Preliminary result of time-dependent spectral function

Figure 3. shows our preliminary results of time-dependent spectral function defined in Sec. II. From $t=0$, the quench starts to move the local energy level at the low energy to the higher energy and the Coulomb repulsion is switched to be smaller, therefore the side peak of the spectral function evolves with time gradually accordingly, and the peak at Fermi level is gradually broaden.

Since this observable originates from the time-resolved photoemission spectroscopy, the spectral function here shows the time-dependent occupied density of states. While the inverse photoemission (IPES) gives the unoccupied density of states. Therefore, one may naturally expect the time-resolved IPES can give the time-dependent unoccupied density of states. This interesting observation will be studied in the near future.

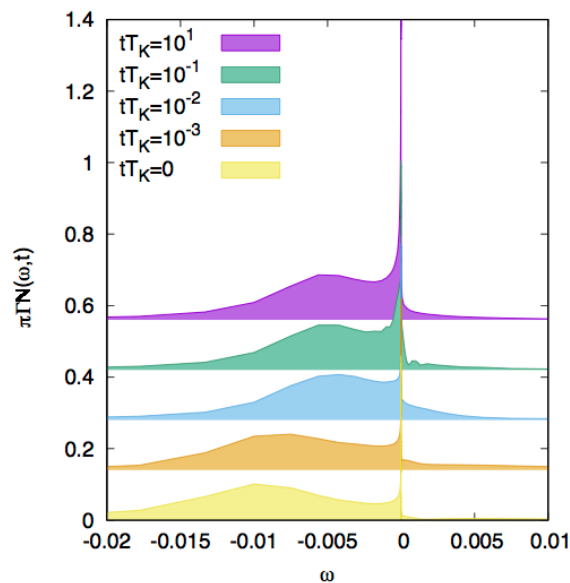


Figure 3. Normalized spectral function at different time.

6. Conclusions

In this paper, we show the computing problem in calculating the time-dependent spectral function originated from the time-reversed photoemission spectroscopy. The problem is due to the sums over four different indices. We solve the problem by mainly using parallel computing with distributed memory, in particular OpenMP. The speedup is shown to be nearly equal to the number of physical threads, while the logical threads gives the slower speedup. We also present the prospective calculation with the use of GPU to speedup further. We note that MPI of the latter versions can also work with shared memory, however, in this paper, we only use MPI for parallel computing with distributed memory.

The preliminary results of time-dependent spectral function are shown to give the time-dependent occupied density of states which can be validated by the time-resolved photomission. We also propose the possible observation of time-dependent unoccupied density of states.

Acknowledgments

We acknowledge the support by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 103.2-2017.353. We acknowledge supercomputer support by the John von Neumann institute for Computing (Jülich).

References

- [1] P.W. Anderson, Localized Magnetic States in Metals, *Physical Review* 124 (1961) 41–53. <https://doi.org/10.1103/PhysRev.124.41>.
- [2] J. Kondo, Resistance Minimum in Dilute Magnetic Alloys, *Progress of Theoretical Physics*. 32 (1964) 37–49. <https://doi.org/10.1143/PTP.32.37>.
- [3] K. Wilson, The renormalization group: Critical phenomena and the Kondo problem, *Reviews of Modern Physics*. 47 (1975) 773. <https://doi.org/10.1103/RevModPhys.47.773>.
- [4] D.L. Cox, A. Zawadowski, Exotic Kondo Effects in Metals: Magnetic Ions in a Crystalline Electric Field and Tunneling Centers, *Advances in Physics* 47 (1998) 599-942. <https://doi.org/10.1080/000187398243500>.
- [5] D. Pesin, L. Balent, Mott physics and band topology in materials with strong spin–orbit interaction, *Nature Physics* 6 (2010) 376–381. <https://doi.org/10.1038/nphys1606>.
- [6] H. Aoki, N. Tsuji, M. Eckstein, M. Kollar, T. Oka, P. Werner, Nonequilibrium dynamical mean-field theory and its applications, *Reviews of Modern Physics* 86 (2014) 779. <https://doi.org/10.1103/RevModPhys.86.779>.
- [7] J.K. Freericks, H.R. Krishnamurthy, T. Pruschke, Theoretical Description of Time-Resolved Photoemission Spectroscopy: Application to Pump-Probe Experiments, *Physical Review Letters* 83 (2009) 808. <https://doi.org/10.1103/PhysRevLett.102.136401>.
- [8] F. Randi, D. Fausti, M. Eckstein, Bypassing the energy-time uncertainty in time-resolved photoemission, *Physical Review B* 95 (2017) 115132. <https://doi.org/10.1103/PhysRevB.95.115132>.
- [9] H.T.M. Nghiem, T.A. Costi, Generalization of the time-dependent numerical renormalization group method to finite temperatures and general pulses, *Physical Review B* 89 (2014) 075118. <https://doi.org/10.1103/PhysRevB.89.075118>.
- [10] H.T.M. Nghiem, T.A. Costi, Time evolution of the Kondo resonance in response to a quench. *Physical Review Letters* 119 (2017) 156601. <https://doi.org/10.1103/PhysRevLett.119.156601>.
- [11] H.T.M Nghiem, H.T. Dang, T.A. Costi, Time-dependent spectral functions of the Anderson impurity model in response to a quench and application to time-resolved photoemission spectroscopy, arXiv:1912.08474. <https://arxiv.org/abs/1912.08474>.

- [12] A. Weichselbaum, J. von Delft, Sum-rule conserving spectral functions from the numerical renormalization group, *Physical Review Letters* 99 (2007) 076402. <https://doi.org/10.1103/PhysRevLett.99.076402>.
- [13] T.A. Costi, V. Zlatić, Thermoelectric transport through strongly correlated quantum dots, *Physical Review B* 81 (2010) 235127. <https://doi.org/10.1103/PhysRevB.81.235127>.
- [14] G.M. Amdahl, Validity of the single processor approach to achieving large scale computing capabilities. *Proceedings of the April 18-20, 1967, Spring joint computer conference*. ACM, 1967, 483-485. <https://doi.org/10.1145/1465482.1465560>.