

## Chapter 7

# Discrete-Event Simulation Research Team

### 7.1 Members

Nobuyasu Ito (Team Leader)  
 Yohsuke Murase (Research Scientist)  
 Naoki Yoshioka (Research Scientist)  
 Daigo Umemoto (Guest Researcher)  
 Tomio Kamada (Guest Researcher)  
 Takeshi Uchitane (Guest Researcher)

### 7.2 Research Activities

Discrete-event simulation comprise various kinds of models, for example, particles, agents, automata, games and so on, and their applications are from material and biomedical sciences to ecological and environmental problems. Furthermore, computer systems executing various simulations including discrete-event ones are also described as discrete-event systems and therefore design and construction of new computer requires discrete-event simulations. Discrete-event simulation research team (DESRT), only one research hub of discrete-event treatment in the Riken Center for Computational Science, has been challenging simulations for natural and social phenomena, and for future computer. Social designs and controls have becoming the more interesting target since so-called the "big data sciences" became popular. Continuous simulations are also becoming building block of discrete-event simulations. The DESRT aims to cultivate such applications of supercomputers.

One characteristic feature of discrete-event simulations is, of course, their discrete nature. Interpolation of two results from two parameter sets is not straightforward, but verification using results of simulations for parameter sets between them is expected. Discrete-event models often comprise with tens, hundreds, thousands, or more parameters which are often, again, discrete. So necessary number of simulations usually very large, and exhaustive parameter search in brute force way is not feasible even with a big supercomputer. One promising way is to use a Bayesian way: do best of what we observed. So-called the AI methods like multivariate statistical analysis, agent-based modeling, genetic method, and machine learning are expected to be powerful for the purposes.

Activities of the DESRT in the year 2019 are the following:

- development and support of job-management application software named OACIS and CARAVAN[1,2,24]
- simulation analysis of social phenomena[3-5,6-18]
- simulation analysis of nonlinear nonequilibrium physical phenomena[19]
- development an application software for quantum-computer simulation[20,21,25]

Some of their highlight achievements are in the following subsections.

class	capability computing		←	→	capacity computing	
	A	B	C	D		
manual execution	easy	tedious	troublesome	desperate		
number of simulations	$\sim 10^1$	$\sim 10^4$	$\sim 10^7$	$\sim 10^{10}$ and more		
(node · second ) per simulation	$\sim 10^{10}$ and more	$\sim 10^7$	$\sim 10^3$	$\sim 10^0$		
output per simulation	$\sim$ TB and more	$\sim$ GB	$\sim$ MB	$\sim$ KB		

Table 7.1: A classification of jobs in terms of necessary number of simulations. The OACIS is designed for jobs for class A, B and C, and the CARAVAN for C and D.

## 7.2.1 Job management applications, OACIS and CARAVAN

A characteristic feature of discrete-event models is their large and complex parameter spaces, and they often show qualitatively different various behaviors with different parameters. Supercomputers help to overcome such difficulty: their performance with extreme parallelism allows us to simulate huge number of parameter sets.

There are two kinds of computer jobs: one pursues computing capability and the other capacity. Discrete-event simulations are in the capacity computing, and many simulations are necessary to be prepared and executed efficiently. Table 7.1 shows a classification of computer jobs in terms of their capacity. But manual operation and orchestration of thousands and more jobs are hard, and application tools for the purpose are usually used. The OACIS and CARAVAN are applications developed and released in our team. OACIS since 2012 can manage most jobs interactively through web browser up to millions. CARAVAN since 2017 handles more jobs assuming that input and output size of each job be small. In this year 2019, eight releases of the OACIS[24] were made open for users, together with user-support activities and one tutorial meeting.

## 7.2.2 Traffic simulation and analysis

The DESRT has been developing a car-traffic simulator of the Kobe city. It is used to see how computer simulations contribute to analyze, predict and optimize modern traffic issues like efficiency, congestion and pollution. After reproducing and analyzing current Kobe traffic in the previous year, simulations of artificial traffic is now challenged. For one traffic situation of the Kobe city, thousands of parameter samples were needed to be simulated to achieve statistically significant analysis. So we had developed a simulator executing various artificial situations simultaneously using K computer. It is prepared by combining the CARAVAN and the simulator for each parameter. From these various simulation parameters, simulation results sometimes very smooth, and sometimes very congested. To identify and classify various outcome from the simulations, deep autoencoder using neural network models is prepared and trained to learn the simulation results(Fig. 7.1). After training, activities of ten neurons in the inner-most layer, which control the output layer activity, are classified using the K-mean methods. Five classes are identified, which turned out to correspond to congestion patterns.

One question arises from this success of the traffic classification using neural network. It is about behavior of the traffic model used in the simulation. From the fact that the neural model learned the traffic pattern, regime diagram of the model will be favorable for the neural learning. To see the regime-diagram morphology, a simplified model of car traffic on road network is proposed.

It is a flow model on directed graph  $G = (V, E)$  with a vertex set  $V = \{i, i = 1, 2, 3, \dots, N\}$  and an edge set  $E = \{e_k = (i_k, j_k), i_k \in V, j_k \in V, k = 1, 2, 3, \dots, M\}$  is treated in the following. Traffic flow of each edge with car density  $\rho$  is assumed to be governed by a single-peaked fundamental diagram,  $f(\rho)$ , where  $f(\rho) = 0$  for  $\rho \leq 0$  or  $\rho \geq \rho_m$  and  $f(\rho)$  takes maximum value  $f_m$  at  $\rho_p$  ( $0 < \rho_p < \rho_m$ ). This  $f(\rho)$  increases monotonically continuously in  $0 < \rho < \rho_{mp}$  from 0 to  $f_m$ , and decreases monotonically continuously in  $\rho_p < \rho < \rho_m$  from  $f(\rho_m + 0)$  to 0 (Fig. 7.2). The edge is in free-run regime for  $0 < \rho < \rho_p$ , and it is in congested regime for  $\rho_p < \rho < \rho_m$ . Therefore a fundamental diagram here is characterized by three parameters  $\rho_p$ ,  $\rho_m$  and  $f_m$ . Each edge may have different FD, so the FD parameters of an edge  $e_k$  are denoted by  $\rho_{p,k}$ ,  $\rho_{m,k}$  and  $f_{m,k}$ .

Flow on this graph G obeying the FDs is considered. A divergent  $d_i$  (or sink if  $d_i < 0$ ) is assumed on each vertex  $i$ , and a positive flux  $f_k$  on each edge  $e_k$ . This flow is conserving in a sense that in-coming flux sum plux divergent and out-going flux sum are equal for all vertices(Fig. 7.3(a) ) and this relation is named as a "conserving condition". A flux  $f_k$  obeying a FD of a edge  $e_k$  so it should be in  $0 \leq f_k \leq f_m$  and this relation is named as "FD constraint".

Here an edge corresponds to a road segment, and a vertex to a crossing and/or parking. Only a steady traffic flow is considered with coarse-grained time scale averaged over movement of each car and signal change. A set of

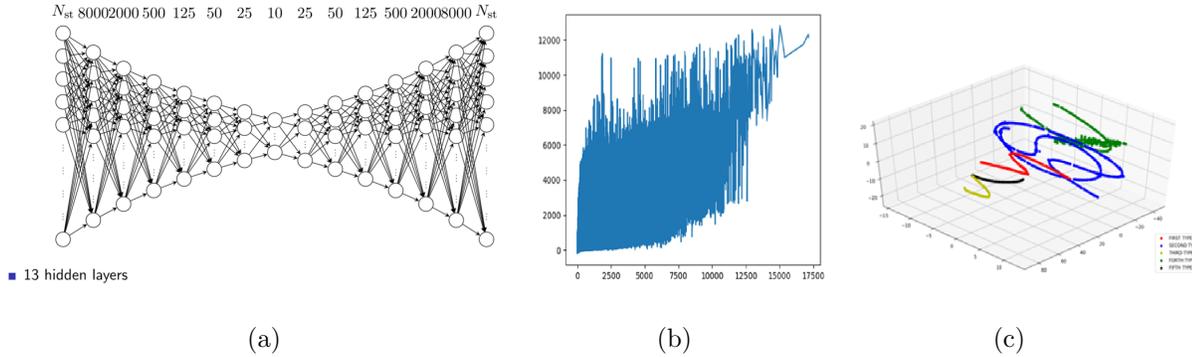


Figure 7.1: (a) A deep autoencoder network used to learn simulation results is shown. Each of left-most input layer and right-most has more than 30,000 neurons corresponding to road segments in Kobe traffic simulation. Quarter or four-times neurons are allocated for next layer, and the inner-most layer has ten neurons. (b) An example of training results Traffic flux for each road segment is shown. Horizontal axis shows traffic flux given in the input layer and vertical axis shows corresponding output from the output layer. The output signals reproduce statistically corresponding input flux and the network learned the simulation results. (c) K-mean classification results of the inner-most neuron activities is shown in three-dimensional plot. Five classes are identified.

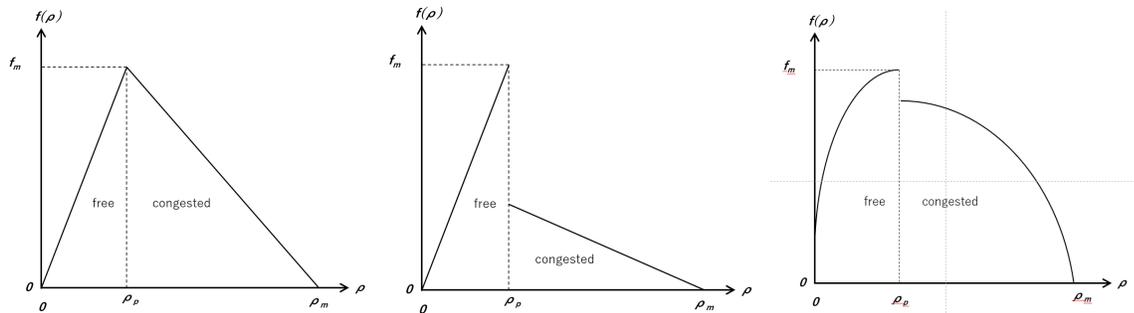


Figure 7.2: Some examples of fundamental diagram  $f(\rho)$  of each edge.

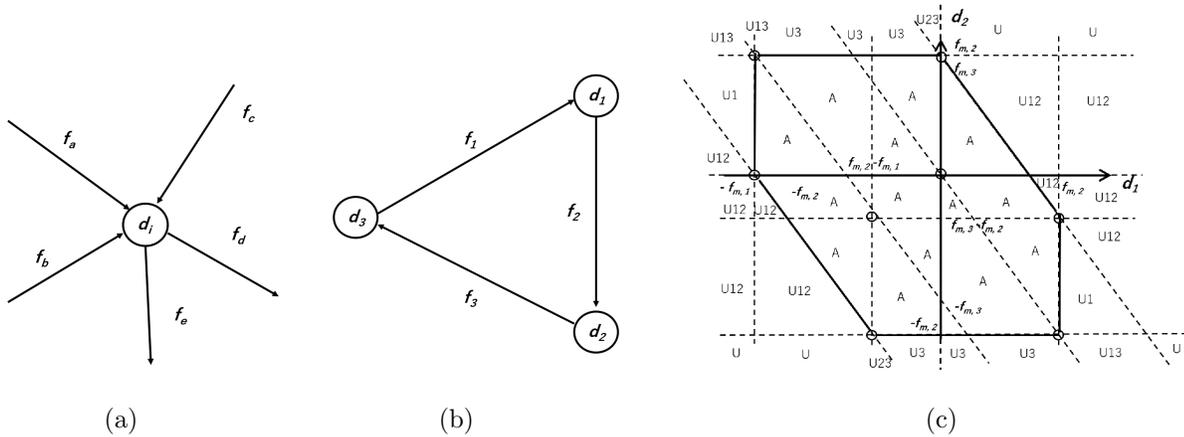


Figure 7.3: (a) Flow at vertex  $i$ . This flow and divergent satisfies a relation  $f_a + f_b + f_c + d_i = f_d + f_e$ . Even if a given graph and a set of divergence is accepted, it does not guarantee a possibility of routing for all cars. But if it is unacceptable, it is not possible to satisfy demand of all cars. (b) shows an example of a simple graph with three vertices with three edge,  $N = 3$  and  $M = 3$ . (c) shows regime diagram of this model for a graph (b). Area marked with "A" is acceptable and all cars can finish their drives. Area "U" is unacceptable and some cars cannot reach to their destinations, and number(s) with the "U" denote(s) overcapacity edge(s).

divergence  $\{d_i\}$  is called acceptable("A" regime), if a set of flow  $\{f_k\}$  satisfying the conserving condition and the FD constraint exists. Otherwise, it is called unacceptable("U" regime) and smooth traffic cannot be expected. The U regime can be classified further depending of which part of the considering graph is not unacceptable. It may be just one edge or some edges, or all edges. This classification is denoted by showing unacceptable edge(s), for example, U1, U23 and U123.

The FD constraint forms convex polyhedral domain and therefore each regime of this model is convex polyhedral, though some regimes will continue to infinity. As an example, regime diagram of a simple graph Fig. 7.3(b) with  $N = 3$  and  $M = 3$  and  $G = (V = \{1,2,3\}, E = \{e_1 = (3,1), e_2 = (1,2), e_3 = (2,3)\})$  is shown in Fig. 7.3(c). From the conserving condition,  $d_1 = f_2 - f_1$ ,  $d_2 = f_3 - f_2$  and  $d_3 = f_1 - f_3 = -d_1 - d_2$ . So the parameter space is two dimensional and here we take  $(d_1, d_2)$ . From the FD constraint, an acceptable regime in this parameter space is an area covering a point  $(d_1, d_2)$  when  $f_1, f_2$  and  $f_3$  take all acceptable flux values. Each of  $f_i (i \in V)$  may take any value between 0 and  $f_{m,i}$  independently, and therefore the acceptable regime is in a set of points which linearly interpolate the following eight points:  $(0, 0)$ ,  $(f_{m,2}, -f_{m,2})$ ,  $(f_{m,2}, f_{m,3} - f_{m,2})$ ,  $(0, f_{m,3})$ ,  $(-f_{m,1}, 0)$ ,  $(f_{m,2} - f_{m,1}, -f_{m,2})$ ,  $(f_{m,2} - f_{m,1}, f_{m,3} - f_{m,2})$  and  $(-f_{m,1}, f_{m,3})$ . When one or two of  $f_i$  are out of the range, corresponding unacceptable regime is named as  $U_i$  or  $U_{ij}$  using numbers of outranged edge,  $i$  and  $j$ .

This graph flow model is a simple one which will be a crude approximation of network traffic. But it will capture a basic feature of network traffic like jam and grid-lock. Regime diagram of this model will be a good starting point to design and control car traffic system using neural network AI.

### 7.2.3 Quantum computer simulation

Quantum information processing is expected to be a next breakthrough beyond current computers. Its growth is now reaching to a promising stage of device developments, and now a realistic design optimization will be a current issue. Simulation of quantum information processing will now be necessary to design a better quantum computers. The DESRT, together with global collaboration, had reached to a 48 qubits simulation which was the world-record scale<sup>1</sup>, and now we have developed further optimized algorithm for massively parallel computers. It is a way to prepare an optimal memory allocation, and it turned out to accelerate two to three times faster simulation. It was achieved using two optimization techniques which are proposed in order to improve performance of the simulator. The "page method" reduces unnecessary copies in each node. It is found that this method makes approximately 17% speed-up maximum. Initial permutation of qubits is also studied

<sup>1</sup> Hans De Raedt, Fengping Jin, Dennis Wilsch, Madita Nocon, Naoki Yoshioka, Nobuyasu Ito, Shengjun Yuan and Kristel Michielsens, "Massively parallel quantum computer simulator, eleven years later," Computer Physics Communications **237** (2019) 47-61.

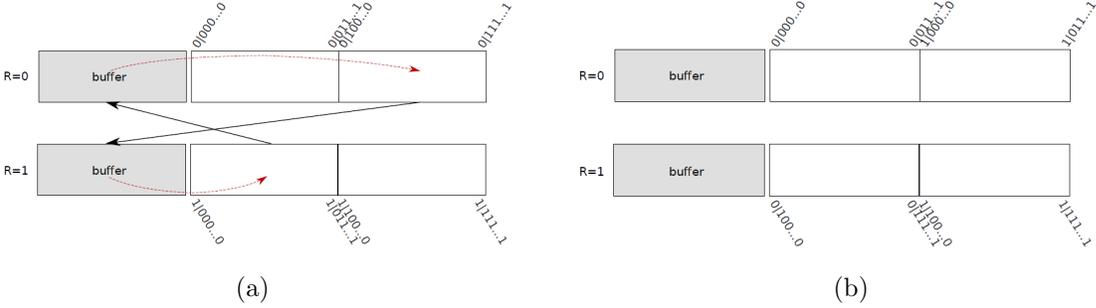


Figure 7.4: Naive data transfer procedure in memory of node R=0 and 1, without using the page method, is shown schematically. In this figure,  $i|jkl \dots z \rangle$  denotes a coefficient of a state  $|ijkl \dots z \rangle$ . (a) Firstly, transferring halves of data were transferred into buffering areas in target nodes(black arrows), then data in buffering areas are copied in target areas(red broken arrows). (b) After (a) procedures, two-qubits interaction for  $|0jkl \dots z \rangle$  and  $|1jlk \dots z \rangle$  is simulated by operations on each node.

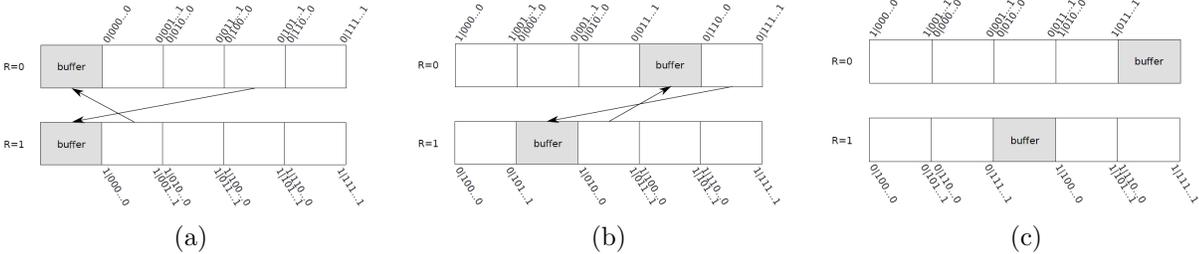


Figure 7.5: Data transfer procedure in memory of node R=0 and 1, using the page method, is shown schematically. (a) Firstly, transferring quarters of data were transferred into buffering areas in target nodes(black arrows) (b) After (a) procedures, data transferred areas are used as buffering areas and transferring quarters of data were transferred into this new buffering areas in target nodes (black arrows) (c) Then two-qubits interaction for  $|0jkl \dots z \rangle$  and  $|1jlk \dots z \rangle$  is simulated by operations on each node.

how it affects performance of the simulator. It is found that a simple permutation in ascending order of the number of operations for each qubit is sufficient in the case of simulations of quantum adder circuits.

On present classical computer, limitation of quantum-computer simulations treating quantum state exactly up to numeric precision of complex coefficients comes mainly from memory capacity. For  $N$  qubits simulation,  $2^N$  complex coefficients are necessary. If double precision for complex coefficients is used, necessary memory storage is  $16 \times 2^N = 2^{N+4}$  byte. A 45 qubits simulation, for example, uses 0.5EB storage. Massively parallel supercomputer can treat this 0.5EB in distributed main memory, and it was a key to achieve 45 qubits in double precision, 46 qubits in single, 47 qubits in half, and 48 qubits in one-byte precision. At every quantum gate operation or qubit interaction, simulator need to reorganize these coefficients so that all interacting coefficients are on the same node using inter-node data transfer. When interacting qubits are all on the same node, inter-node transfer is not necessary. When they are not, inter-node transfer of half of the coefficients for two-qubits interaction and three fourth for three-qubits interaction is necessary. Both of the two techniques optimize these data transfer.

The page method is a technique to omit in-node data transfer after inter-node transfer, and memory usage for data transfer reduced to half compared with a naive transfer described in Fig. 7.4. Procedure with the page method is described in Fig. 7.5. This page method usually reduces execution time because it reduces transfer from buffering area, but there is one drawback: it increases number of transfer operation. Though totally transferring data sizes are the same, increase of transfer call may cause additional waiting time for inter-node latency. Performance comparison on the K computer is given in Fig. 7.6. Slow-down around 41 qubits( $N=41$ ) will be caused by the increase of transfer call.

In this quantum computer simulation, data transfer cost is dominant. It is observed in Fig. 7.6(a), for example. In this simulation, each node holds up to 29 qubits using 8GB memory. And elapse time increases

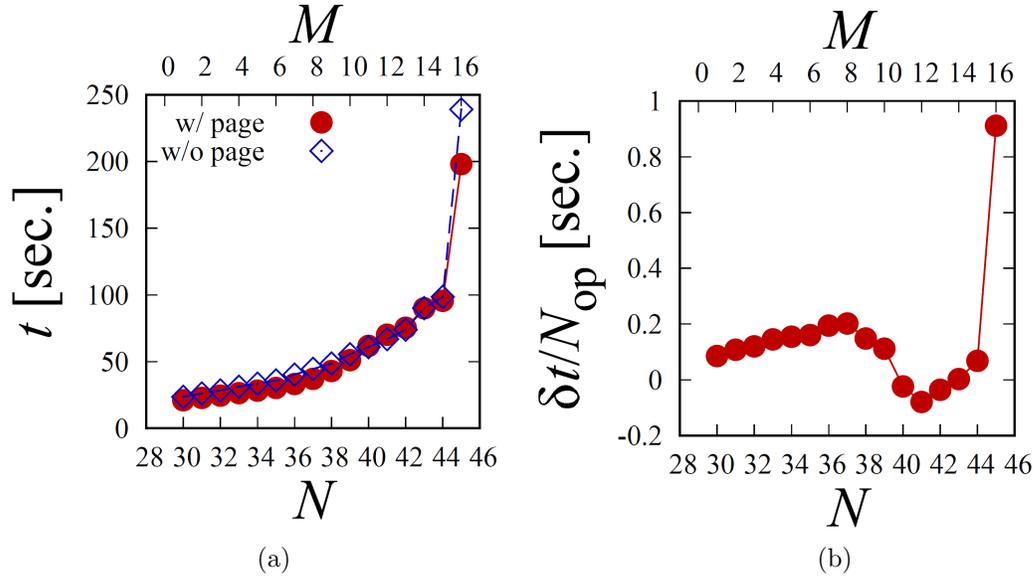


Figure 7.6: Performance of the page method is plotted. Elapse time of one Hadamard operation for each of  $N$  qubits is shown in cases with and without the page method.  $M$  on the top indicate number of qubits allocated over nodes, which implies that  $2^M$  nodes are used for simulation. (a) Elapse time with and without the page method is plotted. (b) Difference of the elapse time per Hadamard operation without and with the page method is plotted. Generally, this difference is positive and the page method improves the performance except around  $N = 41$ .

monotonically for the more qubits. Therefore, qubit allocation for the memory address and node number which minimizes inter-node transfer is expected. Transfer cost for one qubit allocations can be estimated easily, but number of possible allocation is factorial of qubit number. This transfer cost, which will be dominant cost for quantum computer simulation using huge number of nodes of massively parallel supercomputer, changes few times to few tens time depending of initial qubit allocation. Examples of the transfer-cost distribution is shown in Fig. 7.7. One way to find efficient allocation is to make a stochastic optimization. Instead, a simple method named "use-frequency method" is proposed. It is to allocate local address to node number from most frequently used qubit to least used one. This method seems to achieve relatively good performance in our simulation benchmark, and it is also observed in Fig. 7.7.

## 7.2.4 Other activities

In the year of 2019, following subjects were studied in this team:

1. game theoretic study on cooperation and collaboration[8-10,16,18]
2. social network[3-5,15,17]
3. presursor of breakdown under subcritical stress[19]

## 7.3 Schedule and Future Plan

From the research activities of DESRT so far, following problems are becoming clearer:

1. Simulation models, typically social ones, comprise with large input parameters and output numbers, and their behaviors are strongly nonlinear with various regimes.
2. Social "big data" are often not big enough to picture details. It is clearly observed from our multivariate analysis of the traffic data. Thousands of samples are necessary to get minor traffic factors, but such repetitions are not expected in the real traffic. Weather, economics, calendar, accidents and other factors varies every day.

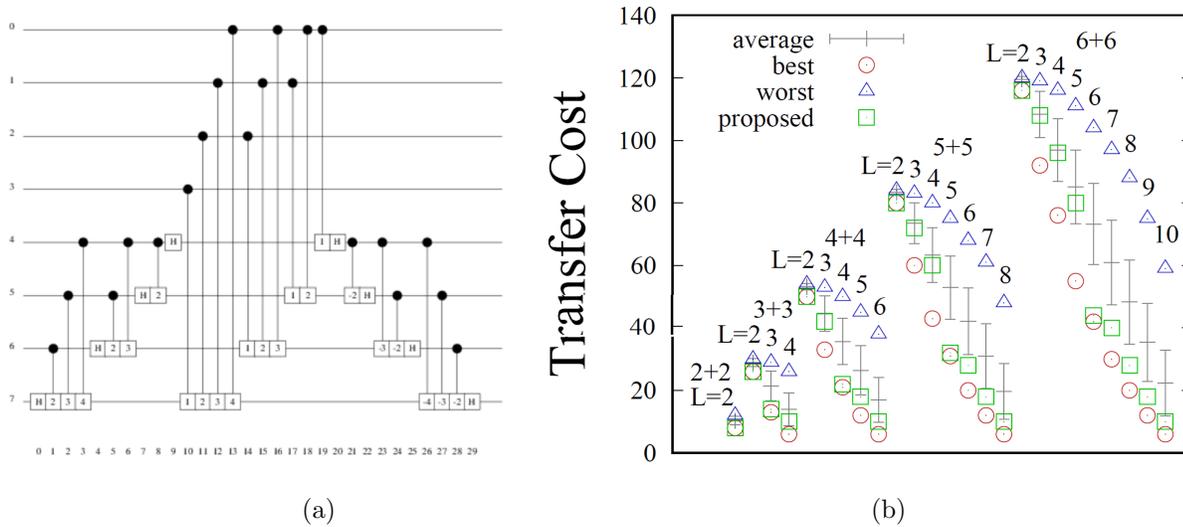


Figure 7.7: Distribution of total transfer costs of quantum addition of two registers, 2+2 to 6+6, obtained by enumerating all permutation  $L$  denotes number of in-node qubits. For example, in a case of  $L = 7$  of 6+6, totally 7 qubits are allocated to in-node address and  $6 + 6 - 7 = 5$  qubits to node number using  $2^5 = 32$  nodes. So there are  $12 = 4.8 \times 10^8$  ways of qubits allocation. Transfer cost is proportional to total data size of inter-node transfer. The best, worst, and average transfer costs are plotted for each case of the number of qubits per register and the number of local qubits. The empty squares are the results of the use-frequency method.

3. Development of multiscale social and economic models and simulations.
4. Bayesian AI will be useful to cultivate diverse behaviors of social phenomena.
5. Quantum information processing may become in near future.

## 7.4 Publications

### 7.4.1 Articles

- [1] Itsuki Noda, Yohsuke Murase, Nobuyasu Ito, Kiyoshi Izumi, Hiromitsu Hattori, Tomio Kamada, Hideyuki Mizuta "Project CASSIA - Framework for Exhaustive and Large-Scale Social Simulation," In: Sato M. (eds) Advanced Software Technologies for Post-Peta Scale Computing. Springer, Singapore, p271-299 (2019)
- [2] Yohsuke Murase, Hiroyasu Matsushima, Itsuki Noda, Tomio Kamada - CARAVAN: a framework for comprehensive simulations on massive parallel machines-"In: Lin D., Ishida T., Zambonelli F., Noda I. (eds) Massively Multi-Agent Systems II. MMAS 2018. Lecture Notes in Computer Science, vol 11422. Springer, Cham (2019).
- [3] Janos Kertesz, Janos Torok, Yohsuke Murase, Hang-Hyun Jo, Kimmo Kaski "Multiplex Modeling of the Society," A book chapter in "Multiplex and Multilevel Networks," edited by S. Battiston, G. Caldarelli, and A. Garas
- [4] Yohsuke Murase, Hang-Hyun Jo, Janos Torok, Janos Kertesz, Kimmo Kaski, "Structural transition in social networks: The role of homophily," Scientific Reports, 9, 4310 (2019)
- [5] Yohsuke Murase, Hang-Hyun Jo, Janos Torok, Janos Kertesz, Kimmo Kaski, "Sampling networks by nodal attributes," Physical Review E, 99, 052304 (2019)

### 7.4.2 Invited talks

- [6]33rd Workshop "Recent Developments in Computer Simulation Studies in Condensed Matter Physics" (The University of Georgia, Athens, U. S. A., February 17-21, 2020) Nobuyasu Ito, "Phase diagram of a traffic simulation model" (Kurt Binder Talk, Feb. 21).
- [7]X Brazilian Meeting on Simulational Physics (BMSP) (Belo Horizonte and Ouro Preto, Brazil, July 15-19, 2019) Nobuyasu Ito, "Simulation of Quantum Computer" (Jul.17, Invited).

- [8] Y. Murase et al., "Conservation of population size is required for self-organized criticality in evolution models," Roles of Heterogeneity in Non-equilibrium collective dynamics (RHINO 2019)
- [9] Y. Murase et al., "Computational Ethics for the Tragedy of the Commons," Self-Organization and Complexity in Social System (Satellite Session of CCS2019)
- [10] Y. Murase et al., "A large-scale search for successful strategies for social dilemma using the K-computer," The 5th Workshop on Self-Organization and Robustness of Evolving Many-Body Systems.

### 7.4.3 Oral talks

- [11] Nobuyasu Ito, "Social simulation with the Fugaku" (Invited, December 14), the 5th workshop on self-organization and robustness of evolving many-body systems (Kobe, December 13-14)
- [12] Nobuyasu Ito, "Social simulations and supercomputer" (Jul. 5), The big data on economy, science and technology Society Conference 2019 (July 5 - 7, 2019, Buenos Aires, Argentina)
- [13] Nobuyasu Ito, "Regime diagram of a traffic simulation model" (January 24, OS13-6), 25th International Symposium on Artificial Life and Robotics (AROB25th 2020) and 5th International Symposium on BioComplexity (ISBC5) (January 22 - 24, 2020, B-Con PLAZA, Beppu, Japan)
- [14] Luning Zhang, Naoki Yoshioka, Daigo Umemoto and Nobuyasu Ito, "Application of deep neural network to traffic simulation" (January 24, OS13-7), 25th International Symposium on Artificial Life and Robotics (AROB25th 2020) and 5th International Symposium on BioComplexity (ISBC5) (January 22 - 24, 2020, B-Con PLAZA, Beppu, Japan)
- [15] Y. Murase et al., "Sampling Networks by Nodal Attributes" NetSci 2019
- [16] Y. Murase et al., "Computational Ethics for the Tragedy of the Commons" FSP2019: Frontiers of Statistical Physics
- [17] Yohsuke Murase, "Network extraction from node attribution," Network Science Seminar (2019).
- [18] Y. Murase et al., "Successful strategies in the Tragedy of the Commons" Conference on Complex Systems 2019
- [19] Naoki Yoshioka, Ferenc Kun, and Nobuyasu Ito, "Jump distance of epicenters in thermally induced cracking of fiber bundles," FSP2019: Frontiers of Statistical Physics (Tokyo, Japan, Jun. 7-8, 2019)

### 7.4.4 Poster presentation

- [20] Supercomputing 19 (Nov. 17-22, 2019, Colorado Convention Center, Denver, U.S.A.) Naoki Yoshioka, Hajime Inaoka, Nobuyasu Ito, Fengping Jin, Kristel Michielsen and Hans De Raedt, "Optimization for Quantum Computer Simulation" (Nov. 19 Poster 69).

### 7.4.5 Other presentations

- [21] Nobuyasu Ito, "Simulation of quantum computer" (Osaka university, 2019, October 8).
- [22] Nobuyasu Ito, "Prediction science harmonizing global community and individuals" (May 24, 2019, Tokyo) the 6th forum of future strategy "Beyond human - machine taking over human".

### 7.4.6 Award

- [23] Yohsuke Murase, "Theoretical study on social activities and application of supercomputers", Outstanding performance award in social science of NTT Docomo mobile science

### 7.4.7 Software

- [24] eight releases of OACIS (v3.4.0 ~ v3.7.1)
- [25] Quantum computer simulator "braket", <https://github.com/naoki-yoshioka/braket>