2020 RIKEN-NRC HPC Workshop

JST: October 20 and 21 – 08:00-11:30
EDT: October 19 and 20 – 19:00-22:30

AS OF 2020-10-19

Zoom MC and Timekeeper: Justin Hickey, NRC

Day 1

EDT (19)	JST (20)	Program	Session Chairs
19:00-	08:00-	**Welcome** Michihiko Minoh, Executive Director, RIKEN	Joel Martin (NRC), Kengo Nakajima (R-CCS)
19:10-	08:10-	**NRC opening remarks** Carolyn Watters, Chief Digital Research Officer, NRC	
19:20- 20:05	08:20- 09:05	**Overview of Fugaku and virtual tour** Satoshi Matsuoka, Director, Center for Computational Science, RIKEN	Mitsuhisa Sato (R-CCS)
20:05- 20:35	09:05- 09:35	**NRC HPC activities overview** Joel Martin, Senior Director, Digital Technologies Research Centre, NRC	
20:40 - 21:10	09:40 - 10:10	**"RISM for HPC" – a new software tool to study biomolecules in complex solutions** Sergey Gusarov, Research Officer, Theory and Modelling, Nanotechnology Research Centre, NRC	Li-Lin Tay (NRC Metrology)
21:10 - 21:40	10:10 - 10:40	**All-atom molecular dynamics simulations of Spike protein on the surface of SARS-CoV-2 in solution** Yuji Sugita, Team Leader, Computational Biophysics, Center for Computational Science, RIKEN	
21:40 - 22:10	10:40 - 11:10	**EGSnrc Monte Carlo simulation of radiation transport** Frederic Tessier, Research Officer, Ionizing Radiation Standards, Metrology Research Centre, NRC	
22:10 - 22:30	11:10 - 11:30	**Discussions** All participants will be invited to ask questions/make comments. Should there be a need, a separate virtual room for a breakout session will be available.	Kengo Nakajima (R-CCS)
22:30	11:30	**End**	

Day 2

EDT (20)	JST (21)	Program	Session Chairs
19:00-	08:00-	**Day 2 Welcome** Shigeru Sasaki, Head of Japan Office, NRC	Kento Sato (R-CCS)
19:05 - 19:35	08:05 - 08:35	**Prediction and countermeasure for droplet/aerosol Infection under the indoor environment for the fight against COVID-19** Makoto Tsubokura, Team Leader, Complex Phenomena Unified Simulation, Center for Computational Science, RIKEN	
19:35	08:35	**Agent-based simulation of COVID-19 infection cluster**	

- 20:05	- 09:05	Nobuyasu Ito, Team Leader, Discrete Event Simulation, Center for Computational Science, RIKEN	
20:05 - 20:35	09:05 - 09:35	**Large-scale natural language processing** Jackie Lo, Research Officer, Multilingual Text Processing, Digital Technologies Research Centre, NRC	
20:40 - 21:10	09:40 - 10:10	**Quantum computing & materials science** Seiji Yunoki, Team Leader, Computational Materials Science, Center for Computational Science, RIKEN	Michael Schuurman (NRC Security and Disruptive Technologies)
21:10 - 21:40	10:10 - 10:40	**Extensive deep neural networks for large scale inference from small scale quantum simulation** Isaac Tamblyn, Research Officer, Deep Learning, Security and Disruptive Technologies Research Centre, NRC	
21:40 - 21:50	10:40 - 10:50	**Closing of plenary session** NRC: Joel Martin, Senior Director, Digital Technologies Research Center RIKEN: Satoshi Matsuoka, Director, Center for Computational Science	Kengo Nakajima (R-CCS)
21:50 - 22:30	10:50 - 11:30	**Special Session: Discussion on the topic of Quantum computing (Optional)** NRC: Phil Kaye, Deputy Director, Quantum Sensing Challenge Program RIKEN: Yasunobu Nakamura, Director, Planning Office for the Quantum Computing Center	Phil Kaye (NRC)
22:30	11:30	**End**	

Presentation Abstracts

"RISM for HPC" – a new software tool to study biomolecules in complex solutions

Sergey Gusarov, Research Officer, Theory and Modelling, Nanotechnology Research Centre, NRC

Integral equation theory of molecular liquids based on statistical mechanics and, in particular its version known as Reference Interaction Site Model (RISM) provides a firm platform to handle complex chemical and biomolecular systems in solution. It enables, at modest computational cost, the calculations of thermodynamics, electronic properties and molecular solvation structure of a solute molecule in a given molecular liquid or mixture. Using 3D-RISM, one can study chemical reactions, including reaction coordinates and transition state search, with the molecular solvation described from the first principles. The method yields all of the features available by using other solvation approaches.

Similar to explicit solvent simulations, 3D-RISM properly accounts for chemical peculiarities of both solute and solvent molecules, such as hydrogen bonding and hydrophobic forces, by yielding the 3D site density distributions of the solvent. Moreover, it readily provides, via analytical expressions, all of the solvation thermodynamics, including the solvation free energy potential, its energetic and entropic decomposition, and partial molar volume and compressibility which can be used to combine them with fine scale approaches, e.g. Quantum Chemistry, Molecular Mechanics, and Molecular Dynamics. On the other hand, solvent distribution functions allowed us to combine this methodology with higher scale approaches such as Dissipative Particle Dynamics. The almost linear scalability defined by Fast Fourier Transform makes it possible to use 3D-RISM-KH in combination with other efficient methods without a dramatic increase in computational cost.

Recently NRC-NANO released the new software tool ("RISM for HPC") realizing this methodology. It combines many new efficiency improvements and fully parallelized. Because of new developments the tool could be used to study the nano-morphology of biomolecules including ligand protein interactions and potential of mean force (PMF) of biomolecules in physiological solution.

All-atom molecular dynamics simulations of Spike protein on the surface of SARS-CoV-2 in solution

Yuji Sugita, Team Leader, Computational Biophysics, Center for Computational Science, RIKEN

The spike protein of SARS-CoV-2 (S-protein) consists of a trimeric polypeptide protein with glycosylated residues on the surface. It triggers the virus entry into a host cell. Because of their functional importance, S-protein structure and function have been extensively investigated by many experimental and theoretical scientists. Here, we performed all-atom molecular dynamics (MD) simulations of a fully glycosylated S-protein in Up and Closed forms in solution. By analyzing the simulation trajectories, we aim to elucidate key inter-domain interactions involving residue pairs and residue-glycan pairs. In addition, to understand molecular mechanisms underlying structural transitions between Up and Closed forms, we also carried out target MD simulations from Closed to Up and from Up to Closed. In addition, we also used one of the enhanced conformational sampling methods, gREST (generalized Replica Exchange Solute Tempering) method starting from Closed form. These simulations contribute to our understanding of conformational stability and functional motions of S-protein, which are relevant for vaccines and antiviral drugs developments.

EGSnrc Monte Carlo simulation of radiation transport

Frederic Tessier, Research Officer, Ionizing Radiation Standards, Metrology Research Centre, NRC

Researchers dealing with radiation in any of its many scientific applications know today that Monte Carlo simulation is an outstanding way to solve most radiation transport problems, in exquisite detail. However, what this method actually entails is not always obvious, hence the presentation begins with a fundamental review of the Monte Carlo approach, using simple examples to reveal how it leverages random numbers to calculate meaningful quantities, efficiently. Given modern computer resources, Monte Carlo supports cutting-edge radiation dosimetry, proves invaluable in radiation safety analyses, and is rapidly dislodging analytical methods in planning complex radiotherapy cancer treatments in the clinic. Of the few existing general-purpose computer codes for ionizing radiation Monte Carlo simulation, the EGSnrc software developed by the National Research Council of Canada stands out as the "gold standard" both in terms of accuracy and speed. It models the propagation and interactions of electrons and photons in virtually any homogeneous material, for energies ranging from 1 keV to 1 GeV.

Historically, the development of EGSnrc has been motivated by clinical radiation therapy and thus it implements rigorous interaction models and algorithms to reach a nominal accuracy of 0.1%. But given its generic physics engine and elaborate geometry definition language, it is readily brought to bear on a wide range of applications, showcased through compelling simulation examples: medical devices, safety and and security equipment, industrial radiation processing facilities, radiation detectors, and fundamental questions in radiation physics. To conclude, we contemplate the large scale calculations that can be tackled by EGSnrc on a machine such as the Fugaku supercomputer, and the kind of collaboration that can be envisioned between NRC and RIKEN to redefine what can be achieved with Monte Carlo simulation of radiation transport.

Prediction and countermeasure for droplet/aerosol Infection under the indoor environment for the fight against COVID-19

Makoto Tsubokura, Team Leader, Complex Phenomena Unified Simulation, Center for Computational Science, RIKEN

Virus droplet infection caused by sneezing, coughing, or talking is strongly influenced by the flow, temperature and humidity of the air around an infected person and potential victims. Especially in the case of the new coronavirus, possibility of aerosol infection by atomized droplets is suggested, in addition to the usual droplet infection. Because smaller aerosol particles drift in the air for a longer time, it is imperative to predict their scattering route and to estimate how surrounding airflow affects the infection. Then the risk of droplet infection can be properly assessed and effective measures to reduce infection can be proposed. In this project, massively parallel coupling simulation of virus droplet scattering, with airflow and heat transfer under the indoor environment such as inside commuter trains, offices, classrooms, and hospital rooms are conducted. By taking into account the characteristics of the virus, its infection risk of virus droplets is assessed under various conditions. Then countermeasures to reduce the risk are proposed from a viewpoint of controlling the air flow. Complex Unified Simulation framework called CUBE, developed at RIKEN R-CCS and implemented on the supercomputer Fugaku, is mainly used, which makes it possible to execute the world-largest and highly accurate virus droplet simulation ever conducted. As the output of this project, the risk of droplet infection under the indoor environment is quantitatively evaluated, and specific countermeasures to reduce the infection risk is proposed in terms of effective ways of opening/closing windows, use of air conditioning, and placement of

partitions. These outputs from the simulation can protect the living and working environment from virus droplet infection, and contribute to earlier recovery of our socio-economic activities.

This project is supported by the Ministry of Education, Culture, Sports, Science and Technology and RIKEN R-CCS, under the collaboration with RIKEN, Kyoto Institute of Technology, Kobe University, Osaka University, Toyohashi University of Technology, Kajima Corporation, and Daikin Industries, Ltd.

Agent-based simulation of COVID-19 infection cluster

Nobuyasu Ito, Team Leader, Discrete Event Simulation, Center for Computational Science, RIKEN

Formation of the COVID19 infection cluster and effectiveness of contact trace to suppress it are analyzed with computer simulation using an agent-based model. Simulations are made for systems with N agents in a closed system contacting with each other randomly among agents except ones in symptomatic state. Initially, there are one just-infected agent in pre-symptomatic state and (N-1) susceptible agents. Typical value of R0 used in the simulations are 2.5. For the contact trace, a privacy-preserving proximity trace is simulated. The followings are observed: (1) Time delay between symptomatic to alert, and basic reproduction number influence strongly on peak value of infected rate. (2) Rate of asymptomatic agent affects weakly on it. (3) Effect of contact alert does not change if alert interval, that is, time interval of contact alert, and isolating time of alert receivers are longer than about one week. Contact alert does not work if these times are less than four days. (4) Effect of contact alert is proportional to square of rate of agent with contact alert application.

Large-scale natural language processing

Jackie Lo, Research Officer, Multilingual Text Processing, Digital Technologies Research Centre, NRC

The multilingual text processing team and the text analytics team at the National Research Council Canada (NRC) carry out research on a wide range of natural language processing (NLP) tasks. The data volume and the scale of these NLP systems are so enormous that research work could not possibly carried out without the access of the high performance computing cluster. For instance, we work on product-level machine translation systems using the transformer models that serve the Canadian parliamentary, federal government and public safety departments. We also are the top leader in international competitions on automatic machine translation evaluation and semantic textual similarity for more than 20 tested translation directions by using massive pretrained language model. Our change point detection on social media data stream is also employed by the Global Public Health Intelligence Network (GPHIN) that is able to detect the outbreak of a pneumonia-like disease before the first formal report about the novel coronavirus was made to the World Health Organization. Other NRC NLP research include text style transfer, hypernym discovery, Indigenous languages revitalization, etc. In this presentation, we give an overview on these research and discuss how the high performance computing cluster enable us to develop and excel in these large scale NLP research.

Quantum computing & materials science

Seiji Yunoki, Team Leader, Computational Materials Science, Center for Computational Science, RIKEN

As R. Feynman was originally suggested in 1982, quantum many-body systems such as quantum materials are one of the most promising classes of applications for quantum computing. In this talk, I would like to first briefly introduce some of the research activities in RIKEN R-CCS on HPC for quantum simulations and quantum

computing, including massively parallel quantum computer simulator [1]. I will then discuss quantum algorithms of quantum simulations for near-term quantum computers such as NISQ devices, focusing on quantum-classical hybrid algorithms with parametrized quantum circuits [2,3] and beyond the variational quantum circuit ansatz [4].

[1] “Massively parallel quantum computer simulator, eleven years later”, H. D. Raedt, F. Jin, D. Willsch, M. Willsch, N. Yoshioka, N. Ito, S. Yuan, and K. Michielsen, Compt. Phys. Comm. **237**, 47 (2019).

[2] “Symmetry-adapted variational quantum eigensolver”, K. Seki, T. Shirakawa, and S. Yunoki, Phys. Rev. A **101**, 052340 (2020).

[3] “Discretized quantum adiabatic process for free fermions and comparison with the imaginary-time evolution”, T. Shirakawa, K. Seki, and S. Yunoki, arXiv:2008.09361.

[4] “Quantum power method by a superposition of time-evolved states”, K. Seki and S. Yunoki, arXiv:2008.03661.

Extensive deep neural networks for large scale inference from small scale quantum simulation

Isaac Tamblyn, Research Officer, Deep Learning, Security and Disruptive Technologies Research Centre, NRC

We present a physically-motivated topology of a deep neural network that can efficiently infer extensive parameters (such as energy, entropy, or number of particles) of arbitrarily large systems, doing so with scaling. We use a form of domain decomposition for training and inference, where each sub-domain (tile) is comprised of a non-overlapping focus region surrounded by an overlapping context region. The size of these regions is motivated by the physical interaction length scales of the problem. We demonstrate the application of EDNNs to three physical systems: the Ising model and two hexagonal/graphene-like datasets. In the latter, an EDNN was able to make total energy predictions of a 60 atoms system, with comparable accuracy to density functional theory (DFT), in 57 milliseconds. Additionally EDNNs are well suited for massively parallel evaluation, as no communication is necessary during neural network evaluation. We demonstrate that EDNNs can be used to make an energy prediction of a two-dimensional 35.2 million atom system, over 1.0 μm2 of material, at an accuracy comparable to DFT, in under 25 minutes. Such a system exists on a length scale visible with optical microscopy and larger than some living organisms.