



RadonPy

Release of RadonPy: world's first software that fully automates polymer physical property calculations using all-atom classical molecular dynamics simulations

First step toward creating a large physical property database of polymer materials

Key research results

- Launching [RadonPy](#), the first open-source software that successfully automates polymer physical property calculations using the all-atom classical molecular dynamics simulations
- Starting the project for developing a polymer physical property database that contains more than 100,000 types of polymer materials using computational resources, such as the supercomputer Fugaku
- Establishing an industry-academia collaboration consortium for promoting the joint development of RadonPy and database

Overview

Material informatics (MI), a new form of materials research that combines materials data with data science and computational science, is gaining traction. MI applies machine learning to predict new materials with innovative properties and their fabrication methods from a vast design space. Over the past few years, MI technology has spread rapidly in various areas of materials research, and many new materials have been discovered. However, the application of MI for polymer materials⁽¹⁾ is lagging behind that for other materials. Needless to say, data is the most important resource in MI. However, efforts toward creating a comprehensive database of polymer properties to enable data-driven research have been insufficient with very few exceptions, such as the polymer property database [PoLyInfo](#)⁽²⁾ being developed by the National Institute for Materials Science (NISM). If this trend continues, data produced by a university laboratory or a single company will become the standard analysis target for MI of polymer materials.

To overcome this situation, Yoshihiro Hayashi and Ryo Yoshida of the Institute of Statistical Mathematics (ISM) have started to develop RadonPy in collaboration with Junko Morikawa of Tokyo Institute of Technology (Tokyo Tech), Junichiro Shiomi of the University of Tokyo, and a PoLyInfo development team in NIMS. This software fully automates polymer property calculations using an all-atom classical molecular dynamics simulation⁽³⁾ (hereinafter referred to as MD simulation or MD calculation). MD-calculated polymer physical properties fluctuate significantly depending on the setting of calculation conditions, and a large amount of computational resources is generally required; hence, constructing a large-scale database using automatic MD simulations is technically difficult. RadonPy takes as input the chemical structure of the repeating unit of any given polymer (monomer), degree of polymerization⁽⁴⁾, temperature, and other calculation conditions; it automatically computes various physical properties, including thermophysical properties, mechanical properties, and optical properties for various polymer

systems, such as amorphous polymers⁽⁵⁾ and polymer solutions. The released version implements a set of automatic computation functions for 15 types of physical properties. RadonPy implements various functions to automate the entire process of MD simulations. One of the unique features of RadonPy is that it comes with standardized parameter sets and calculation conditions that could be applied to diverse polymer materials, which have been determined by comparing and verifying with experimental data in PoLyInfo. Additionally, a machine learning technique called transfer learning⁽⁶⁾ is used to correct biases and variations between MD-calculated property values and experimental data.

The group is promoting a project to create the world's largest open database of polymer physics that includes over 100,000 types of polymer species using RadonPy. Currently, three universities and 19 companies, in addition to the ISM, are participating in the project, and many researchers are working on the joint development of databases with RadonPy across the boundaries of industry and academia. This initiative is supported by the MEXT Program for Promoting Researches on the Supercomputer Fugaku, and participants fully use the computational resources of the supercomputer Fugaku⁽⁷⁾ to produce and accumulate a large amount of data daily. This project creates a world map of polymer material properties. The Pareto frontier⁽⁸⁾ on which trade-offs between physical properties are generated will be revealed by observing the simultaneous distribution of multiple physical properties. Furthermore, the new materials that form the frontier will be identified during data production. Such exhaustive observations cannot be achieved using only experimental approaches requiring considerable costs, such as in material synthesis. This research is the first step toward opening a new horizon of polymer science.

The research results were published in *npj Computational Materials* (Nature Publishing Group) on 2022/11/8, at 10 (8, at 19 Japan time).

Research background

Data is the most important resource in data-driven research. However, currently available data resources in materials research are quantitatively and qualitatively quite limited. In particular, the development of open data on polymeric materials is lagging behind significantly. Most existing polymer property databases have a small amount of data, and the tools for automatically extracting, forming, and processing data are not maintained (Fig. 1). Furthermore, background information, such as sample preparation conditions and higher-order structures of polymer materials, has not been sufficiently recorded. Creating systematically designed and comprehensive open data that contributes to data-driven polymer research is an urgent issue.

The following three points are responsible for the delay in preparing open data.

- The cost of material synthesis, sample preparation, property measurement, and high-throughput simulation (computational experiments) is high.
- The diversity of design variables to be studied (material species, process conditions, etc.) makes it difficult to incentivize researchers to co-create open data that will serve as a common foundation for the entire research fields.

- There is a high awareness of information confidentiality against competitors, and it is difficult to incentivize researchers to disclose data.

Given this background, the efforts toward creating open data in cooperation with the community have been futile. Since many of these factors are related to cultural issues, we expected that, at least in the short- to medium-term, data that can be produced by a university laboratory or single company will become the standard analysis target for MI.

In MI, an effective strategy to compensate for the lack of data is to integrate data from computational experiments in a data analysis workflow. Currently, the development of large-scale computational property databases for various material systems is underway worldwide. For inorganic and low-molecular-weight compounds, in particular, the development of first-principles computational databases that include tens of thousands to several million materials ([Materials Project](#), [AFLOW](#), [OQMD](#), [QM9](#), etc.), has resulted in dramatic advances in the technical progress and practical development of MI. Meanwhile, for polymer materials, the technical barriers of automating physical property computations and the need for large amounts of computation in molecular simulations have been barriers to the development of databases.

Research content

RadonPy is the world's first open-source software that fully automates polymer property calculations with MD simulations. It takes the repeating unit of a given polymer with its chemical structure, degree of polymerization, temperature, etc. as input, and automatically executes all the processes of MD simulation, including molecular modeling, charge calculations, equilibrium/nonequilibrium MD simulations, automatic determination of equilibration completion, restart scheduling when convergence fails, and physical property calculations in the post-processing stage (Fig. 2). The version published in the first release implements an algorithm that automatically computes 15 types of physical properties, including thermophysical properties, mechanical properties, and optical properties. Computable systems include linear polymers⁽⁹⁾ (homopolymers⁽¹⁰⁾, copolymers⁽¹¹⁾) in an amorphous state⁽⁵⁾ or polymers in solution.

This research was conducted as a pilot study for constructing a large-scale database. In this research, we computed 15 types of physical properties for over 1,000 types of amorphous polymers, and we conducted verification experiments using experimental values from the polymer property database PoLyInfo. We thoroughly investigated the effects of calculation conditions, such as the degree of polymerization, the number of polymer chains in a simulation cell and the type of molecular structure, on the calculated values, as well as the prediction accuracy and performance limits for each property type. The performance of MD-calculated polymer properties has not been systematically verified on such a scale. Furthermore, it has become clear that a machine learning technique called transfer learning exhibits strong learning performance in correcting systematic biases and variations in computational properties (Fig. 3). There is no calculation condition that can be universally applied to various polymer materials. Therefore, mass-produced data by fully automatic computations will inevitably have biases and variations. Furthermore,

biases and variations occur in experimental values due to experimental conditions, unobserved factors related to samples, and characteristics of the measurement system. Machine learning can now bridge the gap between such complex real-world systems and imperfect computational models.

Furthermore, this research clarified the simultaneous distribution and Pareto frontier position of multiple physical properties of polymer materials and the structural characteristics of the polymer materials that exist there (Fig. 4). This research focused on the thermal conductivity of polymer materials in particular. With the increase in heat generation due to the miniaturization and high performance of mobile devices in recent years, there has been a growing demand for new materials that can be applied to insulating resins, molding resins, adhesives, coating agents, etc., for mobile devices. In this study, we were able to identify eight types of amorphous polymers with thermal conductivity exceeding $0.4 \text{ W}/(\text{m} \cdot \text{K})$ (Fig. 5). The achievable thermal conductivity of ordinary amorphous polymers is known to be $0.2\text{--}0.3 \text{ W}/(\text{m} \cdot \text{K})$ at most, but the identified materials significantly exceed this level. Furthermore, we applied the decomposition analysis of heat conduction implemented in RadonPy to clarify the mechanism and design guidelines for increasing the thermal conductivity of amorphous polymers. Results show that polymers with a high density of hydrogen-bonding units exhibit improved thermal conductivity through intermolecular hydrogen bonding and dipole interactions, or alternatively improve thermal conductivity through covalent bonds in polymers with high rigidity and linearity.

Future developments

Currently, the group is promoting a project to create the world's largest polymer physical property database containing over 100,000 types of polymers using RadonPy. In October 2022, an industry-academia collaboration consortium was officially launched for the joint development of databases with RadonPy (unofficial activities started in April 2021). The consortium currently involves the participation of three universities and 19 companies in addition to ISM, and there are nearly 90 researchers who are promoting the joint development of databases with RadonPy across the industry and academia. This initiative is supported by the MEXT Program for Promoting Researches on the Supercomputer Fugaku, and many researchers in the industry and academia fully use the supercomputer Fugaku to produce and accumulate a large amount of data daily. The slogan of the project is to create a world map of polymer material properties. We will clarify the simultaneous distribution and Pareto frontier of multiple physical properties and new materials that form that frontier during data production and accumulation data. In particular, we aim to create data on biodegradable plastics and high-thermal-conductivity polymer materials, thereby enabling the development of new materials that contribute to improved thermal management and help realize a decarbonized, recycling-oriented society. Additionally, one of the objectives of this consortium is to disseminate a model case of co-creation of data that transcends the boundaries of industry and academia to society.

Published paper

Paper title: RadonPy: automated physical property calculation using all-atom classical molecular dynamics simulations for polymer informatics

Authors: Yoshihiro Hayashi¹, Junichiro Shiomi², Junko Morikawa³, Ryo Yoshida^{1,4,5}

Journal: [npj Computational Materials](#)

DOI: 10.1038/s41524-022-00906-4

Date of posting: 2022/11/8, 19 PM (10:00 AM BST)

1 The Institute of Statistical Mathematics, Research Organization of Information and Systems

2 Department of Mechanical Engineering, The University of Tokyo

3 Department of Materials Science and Engineering, School of Materials and Chemical Technology, Tokyo Institute of Technology

4 Graduate University for Advanced Studies, Department of Statistical Science

5 Research and Services Division of Materials Data and Integrated System (MaDIS), National Institute for Materials Science (NIMS)

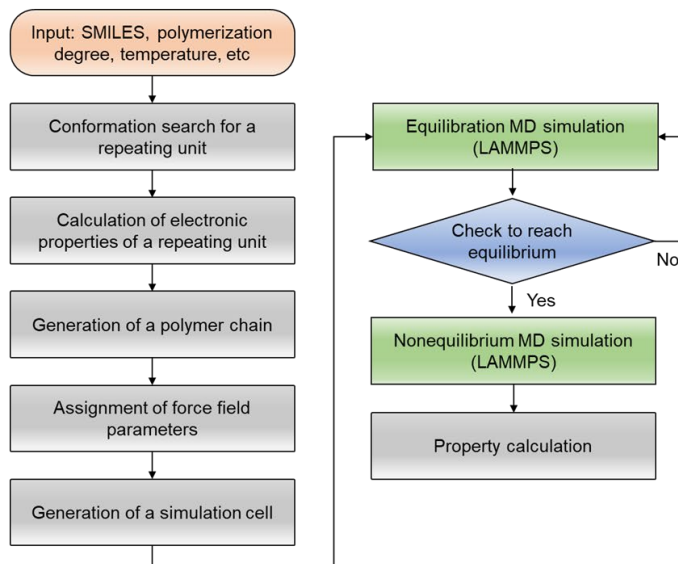
Public website for RadonPy

<https://github.com/RadonPy/RadonPy>

Database	Overview
PoLyInfo (polymer.nims.go.jp)	The world's largest physical property database of polymers that summarizes data extracted from the academic literature. Contains data on approximately 100 physical properties of polymers that are polymerized from about 18,000 types of monomers.
Polymer Genome – Khazana (khazana.gatech.edu)	A platform that provides experimental data extracted from 24 publications and physical property values calculated by ab initio calculations. Contains property data of approximately 1,400 types of polymer / organic materials and approximately 2,600 types of inorganic materials.
Polymer Property Predictor and Database (pppdb.uchicago.edu)	It contains about 260 Flory-Huggins χ parameters and about 210 glass transition temperature data extracted from the academic literature.
NanoMine (materialsmine.org)	Includes composition, process, electron microscopy data, and physical properties of the microstructure of polymer composites.
CROW (polymerdatabase.com)	A repository that contains polymer thermophysical data. Includes experimental data extracted from the academic literature and computed physical property data computed from quantitative structure-activity relationship analysis.
Polymers: A Property Database (poly.chemnetbase.com)	Polymer property data provided as an appendix to the Wiley book "Polymers: A Property Database."

Fig. 1: Representative polymer physical property open-source data. Compared to other applied fields of data science, there is very little data, and most of them do not have tools for automatic data extraction.

RadonPy Workflow of automated all-atom molecular dynamics simulations for polymeric materials



15 properties implemented
(2022/10/01)

- ◆ Thermal conductivity
- ◆ Thermal diffusivity
- ◆ Density
- ◆ Radius of gyration
- ◆ Specific heat capacity C_p
- ◆ Specific heat capacity C_v
- ◆ Compressibility (isothermal)
- ◆ Isentropic compressibility
- ◆ Bulk modulus (isothermal)
- ◆ Isentropic bulk modulus
- ◆ Self-diffusion coefficient
- ◆ Thermal expansion coefficient
- ◆ Linear expansion coefficient
- ◆ Dielectric constant (static)
- ◆ Refractive index

Polymer systems

- ◆ Amorphous
- ◆ Crystalline states
- ◆ Oriented structures

Fig. 2: Overview of RadonPy, software for automatically calculating polymer properties.

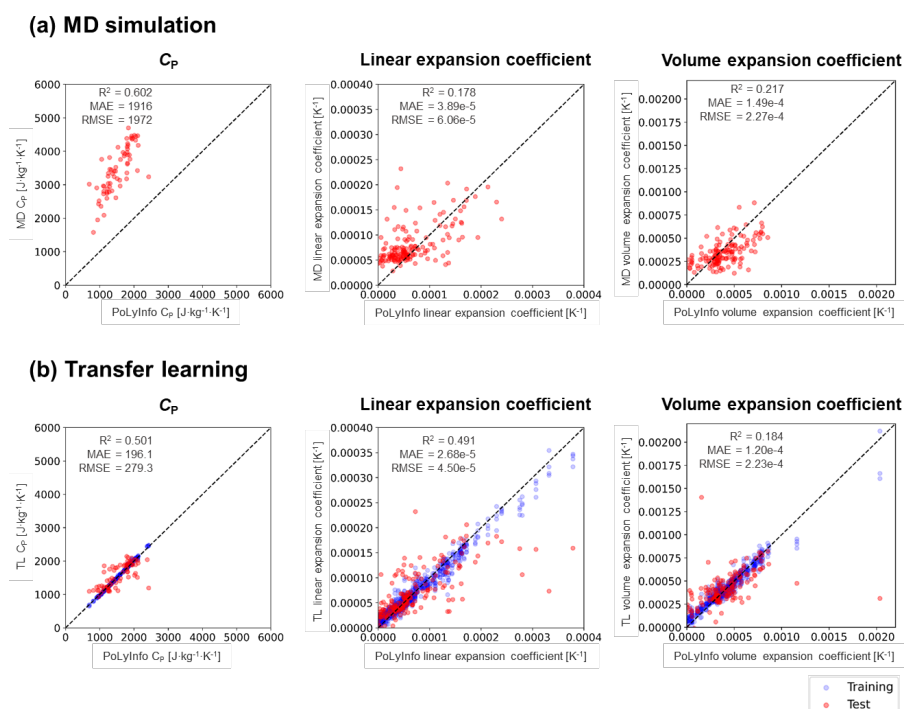


Fig. 3: Calibration of MD-computed physical properties (specific heat capacity at constant pressure (C_p), linear expansion coefficient, volume expansion coefficient). Bias and variations between computed MD values and experimental values (top) were greatly improved by calibration using transfer learning (bottom).

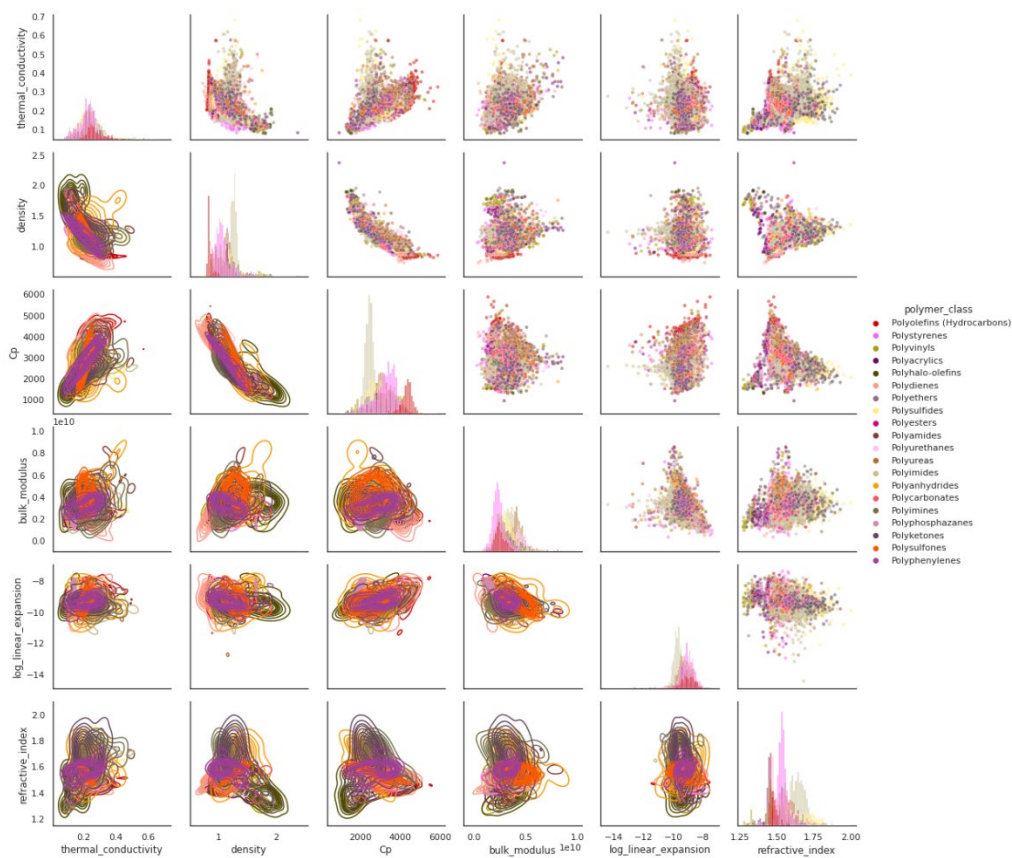


Fig. 4: Simultaneous distribution and Pareto frontier of multiple physical properties (thermal conductivity, density, specific heat capacity at constant pressure (C_p), volume expansion coefficient, linear expansion coefficient, refractive index) of polymer materials that were clarified by high-throughput automatic computation by RadonPy.

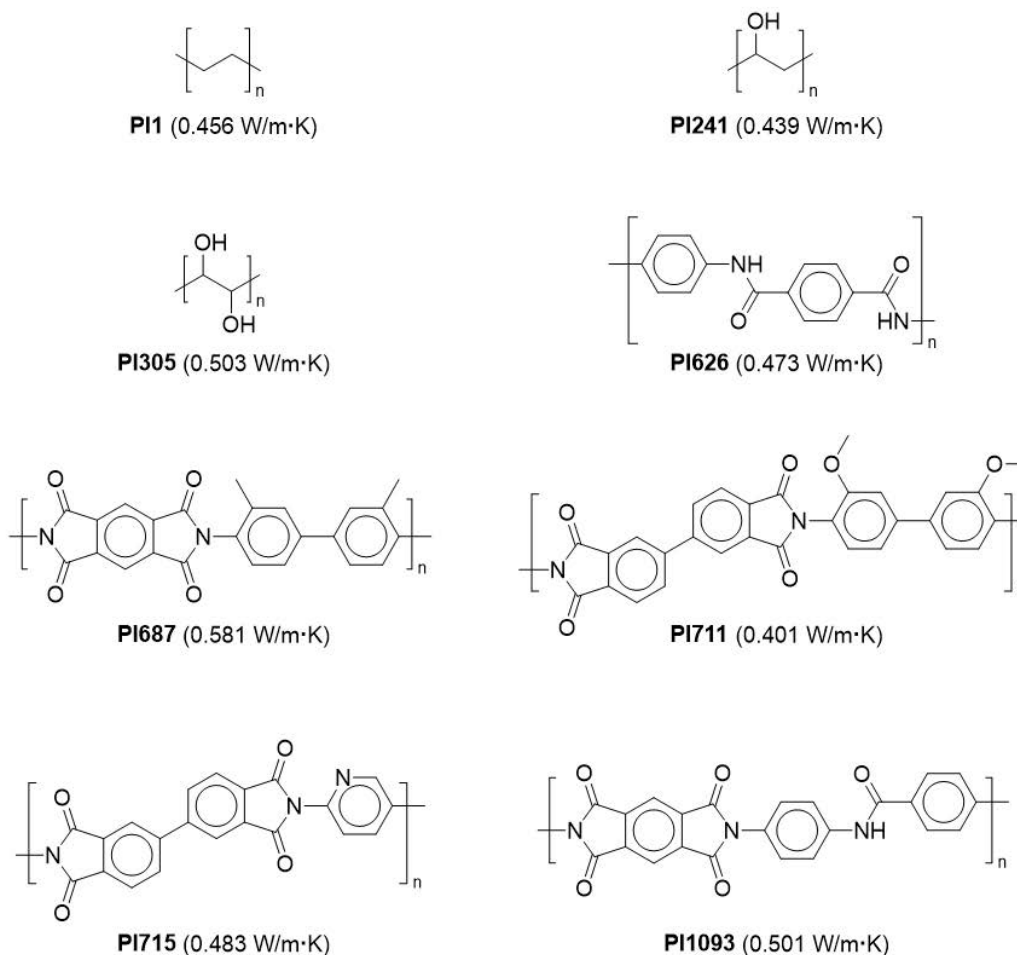


Fig. 5: Eight types of polymers that exhibited a high thermal conductivity of more than 0.4 W/(m · K) in the amorphous state.

Acknowledgments

This research was conducted with support from the MEXT Program for Promoting Researches on the Supercomputer Fugaku “Creation of Data Infrastructure to Transform Data-Driven Polymer Materials Research” (JPMXP1020210314) and JST-CREST “Polymer Thermophysical Properties Materials Informatics”(JPMJCR1913). Part of this research was also conducted using the computational resources of the supercomputer Fugaku (project number: hp210264, hp210213) and the computational resources of the Research Center for Computational Science at the National Institute of Natural Sciences (project numbers: 21-IMS-C126, 22-IMS-C125). We would like to express our gratitude to Isao Kuwajima and Mashi Ishii of the National Institute for Materials Science for providing the polymer physical property database PoLyInfo for this research.

Glossary

(1) Polymer material: Large molecules with a molecular weight of 10,000 or more are generally referred to

as macromolecules or polymer materials. Polymers have a structure in which one or several kinds of constituent elements (repeating units) are repeatedly connected. Materials composed of macromolecules are called macromolecular materials or polymer materials.

(2) PoLyInfo: The world's largest database of polymer physical properties owned by the National Institute for Materials Science. There are records of approximately 100 types of physical properties (e.g., thermophysical properties, electrical properties, mechanical properties), chemical structures, measurement conditions, polymerization methods, etc., collected from the academic literature.

(3) All-atom classical molecular dynamics simulation: Computational experiment that solves Newton's equations of motion (classical mechanics) and determines physical properties and structures from the computed dynamic behavior of atoms, given an interaction potential that acts between atoms

(4) Degree of polymerization: The number of monomer units (repeating units) that make up the polymer

(5) Amorphous polymers: A state of matter with irregularly arranged atoms and no crystal structure is called the amorphous state, and polymers in this state are called amorphous polymers.

(6) Transfer learning: Transfer learning is a general term for statistical machine learning methodologies that compensate for lack of data volume using data or pre-trained models from other related domains. This approach is used when the user wants to build a prediction model with machine learning but the amount of data is small, and it is difficult to learn from scratch. Here, transfer learning from physical property values of MD simulations to experimental values was conducted.

(7) Supercomputer Fugaku: Fugaku is one of the world's fastest supercomputers that was jointly developed by RIKEN and Fujitsu since 2014, and was completed on March 9, 2021. It became the fastest supercomputer in the world in the June 2020 TOP500 list.

(8) Pareto frontier: In multiobjective optimization of multiple properties, there is usually a trade-off relationship in which the performance of other properties declines as one property approaches the optimum level. If a currently obtained solution (e.g., material) cannot further improve multiple physical properties simultaneously, then it is called a Pareto optimal solution. The set of Pareto optimal solutions is called the Pareto frontier.

(9) Linear polymer: A polymer in which monomer units are linearly repeated

(10) Homopolymer: A polymer composed of one type of monomer unit

(11) Copolymer: A polymer composed of multiple types of monomer units

Inquiries regarding this press release

[Research content]

Ryo Yoshida

Professor (Center Director), Data Science Center for Creative Design and Manufacturing, The Institute of Statistical Mathematics, Research Organization of Information and Systems

TEL: +81-50-5533-8534

E-mail: yoshidar@ism.ac.jp

[News, public relations]

URA Station, Planning Office, Operations Planning Headquarters, The Institute of Statistical Mathematics, Research Organization of Information and Systems

10-3 Midoricho, Tachikawa City, Tokyo 190-8562

TEL: +81-50-5533-8580

E-mail: ask-ura@ism.ac.jp