



OPEN

DATA DESCRIPTOR

Efficient compressed database of equilibrated configurations of ring-linear polymer blends for MD simulations

Katsumi Hagita¹✉, Takahiro Murashima², Masao Ogino³, Manabu Omiya⁴, Kenji Ono⁵, Tetsuo Deguchi⁶, Hiroshi Jinnai⁷ & Toshihiro Kawakatsu²

To effectively archive configuration data during molecular dynamics (MD) simulations of polymer systems, we present an efficient compression method with good numerical accuracy that preserves the topology of ring-linear polymer blends. To compress the fraction of floating-point data, we used the Jointed Hierarchical Precision Compression Number - Data Format (JHPCN-DF) method to apply zero padding for the tailing fraction bits, which did not affect the numerical accuracy, then compressed the data with Huffman coding. We also provided a dataset of well-equilibrated configurations of MD simulations for ring-linear polymer blends with various lengths of linear and ring polymers, including ring complexes composed of multiple rings such as polycatenane. We executed 10^9 MD steps to obtain 150 equilibrated configurations. The combination of JHPCN-DF and SZ compression achieved the best compression ratio for all cases. Therefore, the proposed method enables efficient archiving of MD trajectories. Moreover, the publicly available dataset of ring-linear polymer blends can be employed for studies of mathematical methods, including topology analysis and data compression, as well as MD simulations.

Background & Summary

Molecular dynamics (MD) simulations are powerful tools for elucidating molecular-level behavior not only in biomolecular systems but also in polymer material sciences¹⁻⁴. In MD simulations, coordinate data are recorded for detailed analyses. For such analyses, it is necessary to develop mathematical methods that can accurately evaluate how the linear chain penetrates the ring polymer; this has long been an important problem in the mathematics of topology⁵⁻¹⁴. The relevance of this task is not limited to ring-linear polymer blends^{13,14}; research on knots in proteins¹⁵⁻¹⁸, threading of ring polymers¹⁹⁻²², and cross-linked networks^{23,24} is greatly concerned with the linkage between loops and chains owing to its impact on the material properties. Therefore, public availability of MD coordinate data is expected to promote the development of analysis methods by applied mathematicians.

Recently, there has been increasing attention in the field of polymer materials on mixed systems of ring and linear polymers. This is because recent experimental results have demonstrated the toughness of cross-linked ring-linear polymer blends^{25,26}. Here, ring polymers work as movable cross-linking points to prevent stress concentration^{25,26}. To understand these systems, it is important to first conduct detailed investigations of the equilibrium states of the ring-linear polymer blends. The equilibrium state can be obtained by long-term MD simulations^{13,14} in systems with a large number of ring and linear polymers; however, this is not an easy task.

¹Department of Applied Physics, National Defense Academy, 1-10-20, Hashirimizu, Yokosuka, 239-8686, Japan.

²Department of Physics, Tohoku University, 6-3, Aramaki-aza-Aoba, Aoba-ku, Sendai, 980-8578, Japan. ³Faculty of Informatics, Daido University, 10-3 Takihar-cho, Minami-ku, Nagoya, 457-8530, Japan. ⁴Information Initiative Center, Hokkaido University, Kita 11, Nishi 5, Kita-ku, Sapporo, 060-0811, Japan. ⁵Research Institute for Information Technology, Kyushu University, 744 Motoooka, Nishi-ku, Fukuoka, 819-0395, Japan. ⁶Department of Physics, Ochanomizu University, 2-1-1 Ohtsuka, Bunkyo-ku, Tokyo, 112-8610, Japan. ⁷Institute of Multidisciplinary for Advanced Materials, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai, 980-8577, Japan. ✉e-mail: hagita@nda.ac.jp

Item	Values
Type of ring complex	single, bonded-two-ring, bonded-three-ring, poly [2]catenane, poly [3] catenane
Number of beads in a ring polymer (N_{ring})	80, 120, 160
Number of beads in a linear chain (N_{linear})	10, 20, 40, 80, 160
Ring fraction	0.05, 0.1

Table 1. Parameter conditions.

Thus, it is desirable to improve global efficiency through data sharing and reuse instead of duplicating calculations for multiple groups.

A mechanism for the efficient sharing data with reduced data sizes is important because datasets of MD trajectory data are typically very large. Moreover, compression of floating-point data is a common problem for scientific simulations in high-performance computing (HPC)^{27–36}. Some studies on data compression^{27–29} found that the tailing fraction bits are too random to effectively compress because the tail bits in the fraction part of floating-point values in scientific data are more random than the head bits. Methods to neglect tail bits include error-controlled lossy data-compression methods such as ZFP³⁰, ISABELA³¹, SSEM³², and SZ^{33–36}. Recently, comparisons of compressor performance have been performed using benchmark data in various scientific domains; for example, for ZFP and SZ by Lu *et al.*³⁷, Tao *et al.*³⁸, and Cappello *et al.*³⁹. As a result, SZ is regarded as a standard efficient compressor in HPC research for exascale computing. Note that Di and Cappello⁴⁰ reported that time-trajectory analysis-based compressors^{41–48} become impractical in extremely large-scale particle simulations owing to their limited memory capacity. Thus, we focus on the data compression of snapshots.

For lossy compression of MD trajectory data in polymer systems, the required numerical accuracy (error level) and physical meanings such as preservation of topology should remain unchanged. Moreover, in the bit string of the coordinate data in polymer systems, the bits in the sequence along a chain have similar characteristics to time-series data in scientific simulations. Several authors^{29,49,50} have proposed the Jointed Hierarchical Precision Compression Number - Data Format (JHPCN-DF) method, which is a hierarchical segmented recording based on the required numerical precision (error level).

In this study, we analyze the relationship between the numerical accuracy and topology preservation of polymer MD trajectory data under JHPCN-DF compression with the aim of developing a publicly available database. The examined datasets consist of multiple melt systems with a mixture of ring polymers and linear chains. These datasets were prepared as well-equilibrated initial configurations for subsequent MD simulations in order to measure the rheological⁵¹ and mechanical properties after setting crosslinks. Note that these shared dataset provided the first successful discovery⁵¹ of a viscosity overshoot under biaxial extensional flows. In addition, these datasets are appropriate for the development of more accurate and rigorous mathematical judgment methods⁵², as well as efficient approximation techniques based on primitive path (PP) analysis⁵³. As these datasets provide equilibrium states, they can also be useful for developing further coarse-grained MD models that reproduce these states⁵⁴ and planning neutron scattering experiments to observe ring shapes in ring-linear blends. Moreover, publicly available data of polymer systems can be used as a benchmark dataset in the data-compression research community.

Method

Molecular dynamics simulations of ring-linear polymer blends. We generated a dataset that included all combinations of the parameter conditions shown in Table 1 by performing MD simulations^{13,14}. In all cases, MD simulations with a long length of 10^9 MD steps were performed to obtain a well-equilibrated configuration of ring-linear polymer blends. Figure 1 presents schematics of the ring complexes. The examined system size was approximately 600,000 beads. The box sizes of the periodic boundary condition (PBC) were approximately $(80)^3$ in the scale units. Note that the numbers of ring polymers and linear chains were included in the filename for each binary file.

To obtain equilibrated configurations of ring-linear polymer blends, we performed coarse-grained MD simulations of the Kremer-Grest model⁵⁵. Ring polymers with bead number N_{ring} and linear chains with length N_{linear} were placed in a box with PBCs, where the numbers of ring and linear polymers were M_{ring} and M_{linear} , respectively. The length of each simulation run was 10^9 MD steps with a time step (Δt) of 0.005τ , where τ is a time unit.

In the KG model, the Lennard–Jones (LJ) potential with a cutoff length of r_c was applied to every pair of particles.

$$U_{\text{LJ}}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 - \left(\frac{\sigma}{r_c} \right)^{12} + \left(\frac{\sigma}{r_c} \right)^6 \right]$$

when $r < r_c$, whereas $U_{\text{LJ}}(r) = 0$ when $r \geq r_c$, where r is the distance between the beads, ε is the interaction strength, σ is the scale unit, and r_c is the cutoff length of the interaction. For simplicity, we set $\varepsilon = \sigma = 1$ hereafter.

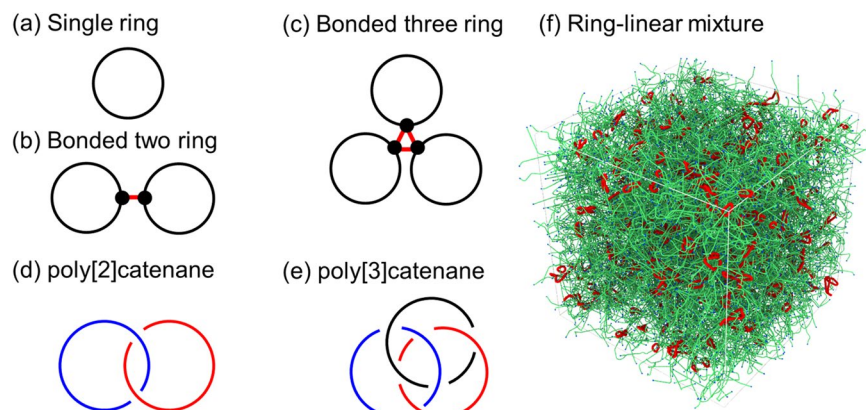


Fig. 1 Schematics of single ring, bonded-rings, poly-catenanes, and ring-linear mixture. The snapshot of the ring-linear mixture with primitive path (PP)⁵³ presentations for $N_{\text{ring}} = N_{\text{linear}} = 160$ with ring fraction 0.1 was rendered by OVITO⁶⁰. In (f), ring polymers and linear chains are shown in red and green, respectively. The ends of linear chains are shown in blue.

To reproduce the excluded volume of chains with minimal computing costs, we set r_c to $2^{1/6}$. For bonded beads, the finite extensible nonlinear elastic (FENE) potential was also applied, where

$$U_{\text{FENE}}(r) = -\frac{k}{2}R_0^2 \ln \left[1 - \left(\frac{r}{R_0} \right)^2 \right]$$

for $r < R_0$ and $U_{\text{FENE}}(r) = 0$ for $r \geq R_0$. Here, k is the spring constant and R_0 is the maximum bond length. The LJ and FENE potentials with $k = 30$ and $R_0 = 1.5$ are widely used to prevent chains from crossing each other. The ring and linear polymers were placed in a box under PBCs with a bead number density of 0.85. Additionally, all ring polymers were unconcatenated. The bead dynamics in our model were described by a Langevin equation with a friction constant (γ) of $0.5 m\tau^{-1}$ and a temperature T . For simplicity, we set the mass of a bead (m) to unity so that T and LJ time ($\tau = \sigma(m/\epsilon)^{1/2}$) became unity. The velocity Verlet algorithm was used for numerical integration of the Langevin equation. In this study, we used LAMMPS⁵⁶ and HOOMD-blue⁵⁷ MD simulation software.

Topology judgement method of chain-penetration into a ring. We evaluated the Gauss Linking Numbers (GLNs) for all ring-linear pairs. However, GLNs cannot be applied to a ring and a linear chain unless the latter is a closed loop. In practice, the ends of linear chains are virtually connected to each other, but we prepared an extra linear chain and connected it to the original linear chain to form a cyclic chain. Details of this method were given in our previous work^{13,14}. To compute GLNs among cyclic chains and ring polymers, we used the Topoly Python package⁵⁸. For a catenated cyclic chain and ring pair, the GLN was equal to 1. Otherwise, $\text{GLN} = 0$. When $\text{GLN} = 1$, we concluded that the linear chain had penetrated the target ring chain.

Efficient compression of floating data. To achieve efficient sharing of lossy and lossless compressed data, the JHPCN-DF method^{29,49,50} was used for hierarchical segmented recording based on the required numerical precision (error level). In essence, the JHPCN-DF framework involves lossless compression with segmented recording; for users who employ parts of the recording, it works as lossy compression. One of the merits of this framework is a substantial reduction of data transfer from big supercomputers to front-end computers for data confirmation through visualization. It should be noted that the part of compression related to the first fraction bits can be regarded as the same as masked data compression²⁸, which was proposed independently by Gomez and Cappello.

The required number of bits in the IEEE 754 format differs for different purposes such as visualization and analysis of scientific data, as shown in Fig. 2. Thus, the required number of bits needs to be properly evaluated for each purpose and simulation target. In scientific simulations using the laws of physics, the first fraction bits are correlated in space and time. However, the tailing fraction bits do not always contribute to visualization and analyses and may instead exhibit random noise-like behavior, which negatively affects data compression^{27–29,49,50}. A higher compression ratio using only the first fraction bits can be observed if the tailing fraction bits can be neglected. Regarding compression efficiency, both data size and ease of use should be considered. For the latter, a simple solution should not change the Application Programming Interface (API). Thus, the conventional binary format with Huffman coding (ex. gzip), and HDF5 can be used as the data API. A combination of zero padding and data compression (such as Huffman coding) can be effective because the size of information in the zero padded bits becomes negligibly small after Huffman coding.

In our implementation^{29,49,50}, the required bit length of each floating-point data was checked for user-specified error levels, such as 0.000001. For the case of IEEE 754 double-precision floating-point data, the stored value of the original variable requires zero padding and a 64-bit integer to record the separated bits necessary to

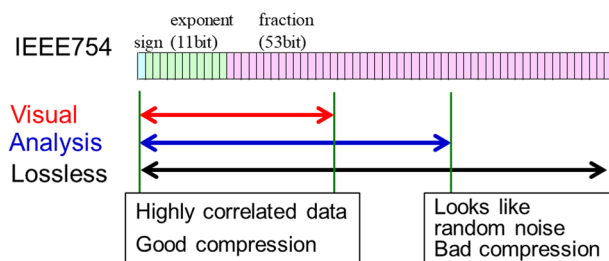


Fig. 2 Required number of bits for visualization and analysis.

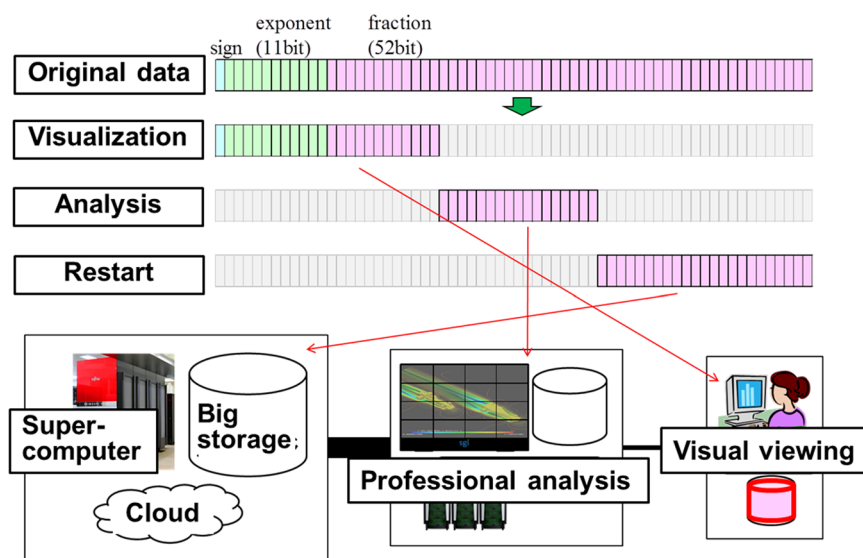


Fig. 3 Example application of separated binary files created within the JHPCN-DF. In this example, the required number of bits was 24 bits and 41 bits for visualization and analysis, respectively. 64 bits of double-precision data were split into three 64-bit recordings: [24 bits + 0-padding (40 bits)], [0-padding (24 bits) + 17 bits + 0-padding (23 bits)], and [0-padding (41 bits) + 23 bits]. Huffman coding reduced the total size of the original 64 bits to less than 64 bits.

reconstruct higher precision data and the original data (lossless). The recordings in the separated binary files using the JHPCN-DF framework are presented in Fig. 3. In this example, 64 bits of double-precision data were split into three parts: [24 bits + 0-padding (40 bits)], [0-padding (24 bits) + 17 bits + 0-padding (23 bits)], and [0-padding (41 bits) + 23 bits]. Before Huffman coding, the total size of the original 64 bits was 192 bits in memory. After Huffman coding, the total size of the original 64 bits became less than 64 bits. For decoding, the OR-operation for the separated data reconstructs original (lossless) data and/or higher precision data. For the example shown in Fig. 3, lossless data can be obtained using the OR-operation for three 64-bit data recordings: OR([24 bits + 0-padding (40 bits)], [0-padding (24 bits) + 17 bits + 0-padding (23 bits)], and [0-padding (41 bits) + 23 bits]).

Data Records

The dataset⁵⁹ consists of 150 systems of ring-linear polymer blends, as shown in Table 1. The datasets are available via the Figshare repository.

Dataset 1. Each filename contains information on the type of ring complex: N_{ring} , M_{ring} , N_{linear} , M_{linear} and f_{ring} . For example, “TwoB_NR120x240_NL20x28800_fr005-D-jhpcndf000001” indicates that the complex was bonded to two ring polymers (as shown in Fig. 3(b)), $N_{\text{ring}} = 120$, $M_{\text{ring}} = 240$, $N_{\text{linear}} = 20$, $M_{\text{linear}} = 28,800$, and $f_{\text{ring}} = 0.05$. The types of ring complex are indicated by “One,” “TwoB,” “ThreeB,” “TwoC,” and “ThreeC,” which correspond to Fig. 3(a–e), respectively. Note that “D-jhpcndf000001” indicates double-precision binary with JHPCN-DF compression and an error level of 0.00001.

Each file contains the following data:

- Size of PBC box (3×8 bytes)
- Positions of beads ($3 \times N_{\text{total}} \times 8$ bytes)

Here, $N_{\text{total}} = N_{\text{ring}} M_{\text{ring}} + N_{\text{linear}} M_{\text{linear}}$. Moreover, $3 \times N_{\text{ring}} \times M_{\text{ring}} \times 8$ bytes in the second line indicates the

Specified error level	Compressed size [bytes]	Compression ratio	Confusion matrix	Error ratio
Lossless	1,058,081,560	—	—	0 (ground truth)
0.1	568,736,963 480,696,320 480,542,720	53.75% 45.43% 45.41%	[[180,543, 6,745], [6,864, 1,374,717,848]]	9.898e-06
0.01	697,258,978 679,249,920 668,119,040	65.89% 64.42% 63.14%	[[186,525, 763], [726, 1,374,723,986]]	1.083e-06
0.001	828,550,734 1,056,235,520 805,140,480	78.31% 99.83% 76.09%	[[187,216, 72], [83, 1,374,724,629]]	1.127e-07
0.0001	985,218,102 1,058,918,400 978,309,120	93.12% 100.08% 92.46%	[[187,283, 5], [10, 1,374,724,702]]	1.091e-08

Table 2. Single-precision binary recording: compressed file size [bytes] and confusion matrix of topology judgments. Regarding the compressed size and compression ratio, the upper, middle, and lower records correspond to the results of JHPCN-DF, SZ compression, and both JHPCN-DF and SZ compression, respectively.

Specified error level	Compressed size [bytes]	Compression ratio	Confusion matrix	Error ratio
Lossless	1,460,836,817	—	—	0 (ground truth)
0.00001	1,133,326,989 1,147,484,160 1,111,470,080	77.58% 78.55% 76.08%	[[187,290, 0], [0, 1,374,724,710]]	0
0.000001	1,219,899,212 1,213,665,280 1,187,983,360	83.51% 83.08% 81.32%	[[187,290, 0], [0, 1,374,724,710]]	0
0.0000001	1,265,078,931 1,256,069,120 1,238,814,720	86.60% 85.98% 84.80%	[[187,290, 0], [0, 1,374,724,710]]	0
Single precision	1,058,081,560	72.43%	[[187,289, 1], [0, 1,374,724,710]]	7.273 e-10

Table 3. Double-precision binary recording: compressed file size [bytes] and confusion matrix of topology judgments. Regarding the compressed size and compression ratio, the upper, middle, and lower records correspond to the results of JHPCN-DF, SZ compression, and both JHPCN-DF and SZ compression, respectively.

positions of the ring polymers. The remaining data indicate the positions of linear chains. In this database, we assumed that the bead order represents the bond connection. N_{ring} beads made a single ring polymer, whereas N_{linear} beads made a linear chain.

In addition, the tailing fraction bits of bead positions were also provided with int64 binary; these are indicated with “D-jhpcndf000001XOR” to denote JHPCN-DF compression and the tailing (XOR) parts. Here, the tailing fraction bits were obtained from the XOR-operation between the original data and the double-precision binary with JHPCN-DF compression.

- Tailing fraction bits of positions of beads ($3 \times N_{\text{total}} \times 8$ bytes)

Technical Validation

Evaluation of segmented recording data. For the double-precision data generated in the MD simulations, we applied JHPCN-DF compression with user-specified error levels of 0.00001, 0.000001, and 0.0000001. For tests of single-precision binary data, single-precision data were obtained by casting from double-precision data. For single-precision binary analysis, we examined cases with user-specified error levels of 0.1, 0.01, 0.001, and 0.0001. Here, 0.0001 was smaller than the limit from the value range, as mentioned below.

Tables 2 and 3 present the size [bytes] and compression ratio of compressed files for single and double-precision binary recording. Here, we employed three methods to achieve the specified error level of the compressed files: (1) “tar” and “gzip -9” for the segmented recording binary file based on JHPCN-DF, (2) “tar” for the “sz”-compressed file of the lossless binary file, and (3) “tar” for the “sz”-compressed file of the segmented recording binary file with JHPCN-DF. Here, we used version 2.1.8.3 of SZ with the Zstd best compression mode³⁶. In the process of generating the compressed files, we monitored the maximum and minimum values of positions: Max = 1981.244394305023 and Min = -1806.817917672729. It should be noted that these values may be inaccurate with single precision. In the case of single precision, from this range and fraction part of 23-bits, as $(\text{Max} - \text{Min})/2^{23}$ was approximately 0.00045, the error level cannot be maintained even for a single-precision binary without JHPCN-DF. According to the obtained compression ratios, the results for all compression methods were similar. For all cases, the combination of JHPCN-DF and the SZ-compressor showed the best performance. It should be noted that the increased size of SZ-compressed files for single-precision data

Specified error level	$N_{\text{linear}} = 10$	$N_{\text{linear}} = 20$	$N_{\text{linear}} = 40$	$N_{\text{linear}} = 80$	$N_{\text{linear}} = 160$
0.1	1.020e-05	9.678e-06	9.475e-06	9.210e-06	9.898e-06
0.01	1.099e-06	1.096e-06	1.026e-06	1.003e-06	1.105e-06
0.001	1.113e-07	1.240e-07	1.296e-07	6.764e-08	6.764e-08
0.0001	9.864e-09	1.409e-08	1.127e-08	1.127e-08	0

Table 4. Single-precision binary recording: N_{linear} -dependence of the error ratio of topology judgments.

Specified error level	$N_{\text{linear}} = 10$	$N_{\text{linear}} = 20$	$N_{\text{linear}} = 40$	$N_{\text{linear}} = 80$	$N_{\text{linear}} = 160$
0.00001	0	0	0	0	0
0.000001	0	0	0	0	0
0.0000001	0	0	0	0	0
Single precision	0	0	5.637e-09	0	0

Table 5. Double-precision binary recording: N_{linear} -dependence of the error ratio of topology judgments.

with a specified error level of 0.0001 may be a result of insufficient detail parameter tuning. Further analysis of this hypothesis is beyond the scope of this paper.

Topology analyses using segmented recording data. As a test for the segmented recording data, we evaluated the GLN for topology judgment regarding penetration of a linear chain into a ring polymer using the method proposed by the authors^{13,14}. This is because the topology is not conserved if the numerical accuracy is poor. The ratio of correct answers of the topology judgment was used as the evaluation index, which was obtained for several user-specified accuracies. Tables 2 and 3 present the confusion matrix and error ratio of the topology judgment for all pairs of ring polymers and linear chains in all systems. Here, the confusion matrix has been effectively employed as a two-class classification problem in machine learning and is given as [[True Positive (TP), False Negative (FN)], [False Positive (FP), True Negative (TN)]], where “Positive” means that the linear chain penetrated into the ring polymer and “True” means that the topology was preserved between lossless compression and the specified error level. The error ratio was defined as $(FP + FN)/(TP + FP + FN + TN)$.

According to the single-precision binary recording in Table 2, increasing the error level (tolerance) increases misjudgment of the topology. This test provides a good example of the relationship between numerical precision and topology judgment errors. However, regarding the original purpose of achieving recording with topology conservation, the single-precision binary format was insufficient. Moreover, the double-precision data in Table 3 exhibited no error in topology judgment with an error level of 0.00001, whereas the single-precision data exhibited two errors. Consequently, we used the JHPCN-DF method with an error level of 0.00001 to develop the publicly available database of well-equilibrated initial configurations of ring-linear polymer blends.

We also investigated the influence of the size of linear chains (N_{linear}) because an incorrect judgment is more likely for shorter linear chains due to the limitation of the topology judgment algorithm between a ring polymer and a linear chain¹³. Tables 4 and 5 present the N_{linear} dependence of the error ratio of topology judgments. If the error ratio can be optimized for this problem, compression with an error level corresponding to N_{linear} is justified.

Code availability

To decode the JHPCN-DF compression, no special attention was required. (The easy-to-use sample code for generating the LAMMPS input data file is attached in the Supporting Information) To encode/segment the data into two parts with JHPCN-DF compression, as shown for the above data, the main part of the reference code⁴¹ is as follows:

```

union fi64{
double f;
uint64_t i64;
};
double fval0,fval1,allowerr,logallo;
int i,ntotal,ival,ival2,sval;
union fi64 fival,fival1;
double *posi_before_compress,*posi_after_compress;
uint64_t *tailing_fraction_bits_posi;

allowerr = 0.00001
logallo = log(allowerr)/log(2.0);
for(i = 0;i < 3*ntotal;i++){
fval0 = posi_before_compress[i];
frexp(fval0,&ival);
ival2 = (int)(-logallo + ival);
sval = (int)(53-ival2);

```

```

if(sval > 52) sval = 53;
do {
sval--;
fival.f = fval0;
fival.i64 = (fival.i64 >> sval);
fival.i64 = (fival.i64 << sval);
fval1 = fival.f;
} while ((fval1 - fval0)*(fval1 - fval0) > allowerr*allowerr);
posi_after_compress[i] = fval1;
fival1.f = fval1;
fival.f = fval0;
tailing_fraction_bits_posi[i] = (fival1.i64 ^ fival.i64);
}

```

For software developers, RIKEN has released the open library “JHPCN-DF” at the following GitHub repository: <https://github.com/avr-aics-riken/JHPCN-DF>.

Received: 19 April 2021; Accepted: 21 December 2021;

Published online: 08 February 2022

References

- Binder, K. Monte Carlo and Molecular Dynamics Simulations in Polymer Science; Oxford University Press: Oxford, UK, 1995.
- Frenkel, D.; Smit, B. Understanding Molecular Simulation: From Algorithms to Applications, 2nd ed.; Academic Press: San Diego, 2002.
- Rapaport, D.C. The Art of Molecular Dynamics Simulation; Cambridge University Press: Cambridge, UK, 2004.
- Gartner, T. E. III & Jayaraman, A. Modeling and Simulations of Polymers: A Roadmap. *Macromolecules* **52**(3), 755–786 (2019).
- Panagiotou, E. The linking number in systems with Periodic Boundary Conditions. *J. Comput. Phys.* **300**, 533–573 (2015).
- Panagiotou, E., Millett, K. C. & Atzberger, P. J. Topological Methods for Polymeric Materials: Characterizing the Relationship Between Polymer Entanglement and Viscoelasticity. *Polymers* **11**, 437 (2019).
- Millett, K. C., Dobay, A. & Stasiak, A. Linear random knots and their scaling behavior. *Macromolecules* **38**, 601–606 (2005).
- Halverson, J. D., Grest, G. S., Grosberg, A. Y. & Kremer, K. Rheology of Ring Polymer Melts: From Linear Contaminants to Ring-Linear Blends. *Phys. Rev. Lett.* **108**, 038301 (2012).
- Jeong, C. & Douglas, J. F. Relation between Polymer Conformational Structure and Dynamics in Linear and Ring Polyethylene Blends. *Macromol. Theory Simul.* **26**, 1700045 (2017).
- Tsalikis, D. G. & Mavrantzas, V. G. Threading of Ring Poly(ethylene oxide) Molecules by Linear Chains in the Melt. *ACS Macro Lett.* **3**, 763–766 (2014).
- Katsarou, A. F., Tsamopoulos, A. J., Tsalikis, D. G. & Mavrantzas, V. G. Dynamic Heterogeneity in Ring-Linear Polymer Blends. *Polymers* **12**, 752 (2020).
- Tsalikis, D. G. & Mavrantzas, V. G. Size and Diffusivity of Polymer Rings in Linear Polymer Matrices: The Key Role of Threading Events. *Macromolecules* **53**, 803–820 (2020).
- Hagita, K. & Murashima, T. Effect of Chain-Penetration on Ring Shape for Mixtures of Rings and Linear Polymers. *Polymer* **218**, 123493 (2021).
- Hagita, K. & Murashima, T. Multi-Ring Configurations and Penetration of Linear Chains through Rings on Bonded Rings and Polycatenanes in Linear Chain Matrices. *Polymer* **223**, 123705 (2021).
- Sułkowska, J. I., Rawdon, E. J., Millett, K. C., Onuchic, J. N. & Stasiak, A. Conservation of complex knotting and slipknotting patterns in proteins. *PNAS* **109**, E1715–E1723 (2012).
- Dabrowski-Tumanski, P. & Sulikowska, J. I. Topological knots and links in proteins. *PNAS* **114**, 3415–3420 (2017).
- Potestio, R., Micheletti, C. & Orland, H. Knotted vs. Unknotted Proteins: Evidence of Knot-Promoting Loops. *PLoS Comput. Biol.* **6**, e1000864 (2010).
- Wüst, T., Reith, D. & Virnau, P. Sequence Determines Degree of Knottedness in a Coarse-Grained Protein Model. *Phys. Rev. Lett.* **114**, 028102 (2015).
- Michieletto, D., Marenduzzo, D., Orlandini, E., Alexander, G. P. & Turner, M. S. Threading Dynamics of Ring Polymers in a Gel. *ACS Macro Lett.* **3**, 255–259 (2014).
- Michieletto, D., Marenduzzo, D., Orlandini, E., Alexander, G. P. & Turner, M. S. Dynamics of self-threading ring polymers in a gel. *Soft Matter* **10**, 5936–5944 (2014).
- Rosa, A., Smrek, J., Turner, M. S. & Michieletto, D. Threading-Induced Dynamical Transition in Tadpole-Shaped Polymers. *ACS Macro Lett.* **9**, 743–748 (2020).
- Landuzzi, F., Nakamura, T., Michieletto, D. & Sakaue, T. Persistence homology of entangled rings. *Phys. Rev. Res.* **2**, 033529 (2020).
- Lang, M. On the Elasticity of Polymer Model Networks Containing Finite Loops. *Macromolecules* **52**, 6266–6273 (2019).
- Panyukov, S. Loops in Polymer Networks. *Macromolecules* **52**(11), 4145–4153 (2019).
- Okumura, Y. & Ito, K. The Polyrotaxane Gel: A Topological Gel by Figure-of-Eight Cross-links. *Adv. Mater.* **13**, 485–487 (2001).
- Yamamoto, K., Nameki, R., Sogawa, H. & Takata, T. Macrocyclic Dinuclear Palladium Complex as a Novel Doubly Threaded [3] Rotaxane Scaffold and Its Application as a Rotaxane Cross-Linker. *Angew. Chem. Int. Ed.* **59**, 2–8 (2020).
- Burtscher, M. & Ratanaworabhan, P. FPC: A high-speed compressor for double-precision floating-point data. *IEEE Trans. Comput.* **58**, 18–31 (2009).
- Gomez, L. A. & Cappello, F. Improving floating point compression through binary masks. *Proc. IEEE Int. Conf. Big Data* 326–331 (2013).
- Hagita, K., Omiya, M., Honda, T. & Ogino, M. Efficient data compression by efficient use of HDF5 format. *Proc. IEEE/ACM Supercomputing (SC14)*, poster 15 (2014).
- Lindstrom, P. Fixed-rate compressed floating-point arrays. *IEEE Trans. Vis. Comput. Graph.* **20**, 2674–2683 (2014).
- Lakshminarasimhan, S. *et al.* Compressing the incompressible with ISABELA: *In-situ* reduction of spatio-temporal data. *Proc. Eur. Conf. Parallel Process.* 366–379 (2011).
- Sasaki, N., Sato, K., Endo, T. & Matsuoka, S. Exploration of Lossy Compression for Application-Level Checkpoint/Restart. *Proc. IEEE Int. Parallel Distrib. Process. Symp.* 914–922 (2015).
- Di, S. & Cappello, F. Fast error-bounded lossy HPC data compression with SZ. *Proc. IEEE Int. Parallel Distrib. Process. Symp.* 730–739 (2016).

34. Tao, D., Di, S., Chen, Z. & Cappello, F. Significantly improving lossy compression for scientific data sets based on multidimensional prediction and error-controlled quantization. *Proc. IEEE Int. Parallel Distrib. Process. Symp.* 1129–1139 (2017).
35. Zou, X. *et al.* Performance Optimization for Relative-Error-Bounded Lossy Compression on Scientific Data. *Proc. IEEE Int. Parallel Distrib. Process. Symp.* 1665–1680 (2020).
36. Argonne National Laboratory, <http://collab.msc.anl.gov/display/ESR/SZ>.
37. Lu, T. *et al.* Understanding and modeling lossy compression schemes on HPC scientific data. *Proc. IEEE Int. Parallel Distrib. Process. Symp.* 348–357 (2018).
38. Tao, D., Di, S., Liang, X., Chen, Z. & Cappello, F. Optimizing lossy compression rate-distortion from automatic online selection between SZ and ZFP. *IEEE Trans. Parallel Distrib. Syst.* **30**, 1857–1871 (2019).
39. Cappello, F. *et al.* Use cases of lossy compression for floating-point data in scientific data sets. *Int. J. High Perform. Comput. Appl.* **33**, 1201–1220 (2019).
40. Di, S. & Cappello, F. Optimization of Error-Bounded Lossy Compression for Hard-to-Compress HPC Data. *IEEE Trans. Parallel Distributed Syst.* **29**, 129–143 (2018).
41. Hagita, K., Takeda, T., Kato, T., Ohtani, H. & Ishiguro, S. Efficient Data Compression of Time Series of Particles' Positions for High-Throughput Animated Visualization. *Proc. IEEE/ACM Supercomputing (SC13)*, poster 8 (2013).
42. Hagita, K., Kato, T., Ohtani, H. & Ishiguro, S. TOKI Compression for Plasma Particle Simulations. *Plasma Fusion Res.* **9**, 3401083 (2014).
43. Ohtani, H. *et al.* Irreversible data compression concepts with polynomial fitting in time-order of particle trajectory for visualization of huge particle system. *J. Phys.: Conf. Ser.* **45**, 1–11 (2013).
44. Yang, D. Y., Grama, A. & Sarin, V. Bounded-error Compression of Particle Data from Hierarchical Approximate Methods. *Proc. IEEE/ACM Supercomputing (SC99)*, Article 32 (1999).
45. Huwald, J., Richter, S., Ibrahim, B. & Dittrich, P. Compressing molecular dynamics trajectories: Breaking the one-bit-per-sample barrier. *Comput. Chem.* **37**, 1897–1906 (2016).
46. Han, Y., Sun, W. & Zheng, B. COMPRESS: A Comprehensive Framework of Trajectory Compression in Road Networks. *Proc. ACM Trans. Database Sys.* **11** (2017).
47. Tomasi, M. Polycomp: Efficient and configurable compression of astronomical timelines. *Astron. Comput.* **16**, 88–98 (2016).
48. Dvořák, J., Maňák, M. & Váša, L. Predictive compression of molecular dynamics trajectories. *J. Mol. Graph. Model.* **96**, 107531 (2020).
49. Liu, L. & Ogino, M. Performance evaluation of efficient data compression JHPCN-DF for large-scale structural analysis, *Mech. Eng. Lett.* **2**, 16–00119 (2016).
50. Liu, L., Ogino, M. & Hagita, K. Efficient Compression of Scientific Floating-Point Data and An Application in Structural Analysis. *Trans. J. Soc. Comput. Eng. Sci.* **2017**, 20170002 (2017).
51. Murashima, T., Hagita, K. & Kawakatsu, T. Viscosity Overshoot in Biaxial Elongational Flow: Coarse-Grained Molecular Dynamics Simulation of Ring-Linear Polymer Mixtures. *Macromolecule* **54**, 7210–7225 (2021).
52. Uehara, E. private communication.
53. Sukumaran, S. K., Grest, G. S., Kremer, K. & Everaers, R. Identifying the primitive path mesh in entangled polymer liquids. *J. Polym. Sci. Part B: Polym. Phys.* **43**, 917–933 (2005).
54. Ohkuma, T. private communication.
55. Kremer, K. & Grest, G. S. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *J. Chem. Phys.* **92**, 5057–5086 (1990).
56. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
57. Anderson, J. A., Glaser, J. & Glotzer, S. C. HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations *Comput. Mat. Sci.* **173**, 109363 (2020).
58. Dąbrowski-Tumański, P., Rubach, P., Niemyska, W., Greń, B. & Sulkowska, J. I. Topoly: Python package to analyze the topology of polymers. *Brief. Bioinformatics* **bbaa196**. <https://doi.org/10.1093/bib/bbaa196> (2020).
59. Hagita, K. *et al.* Efficient Compressed Database of Equilibrated Configurations of Ring-Linear Polymer Blends for MD Simulations. *figshare* <https://doi.org/10.6084/m9.figshare.c.5376578> (2021).
60. Stukowski, A. Visualization and analysis of atomistic simulation data with OVITO – the Open Visualization Tool. *Modelling Simul. Mater. Sci. Eng.* **18**, 015012 (2010).

Acknowledgements

One of the authors (K.H.) gratefully acknowledge Prof. Yutaka Ishikawa at RIKEN Center for Computational Science for fruitful discussions for lossy compression of floating-point data in scientific simulations. The authors are partially supported by the Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures (JHPCN) and the High-Performance Computing Infrastructure (HPCI) in Japan: hp200048, hp200168 and hp210132. This work was partially supported by JSPS KAKENHI, Japan, grant nos.: JP18H04494, JP19H00905, and JP20H04649, and JST CREST, Japan, grant nos.: JPMJCR1993 and JPMJCR19T4.

Author contributions

The manuscript was written with contributions from all the authors. The preparation of datasets and analyses were mainly performed by K.H. and T.M. The developments of the JHPCN-DF compression were mainly contributed by K.H., M.Ogino, M.Omiya, and K.O. The discussion of the future prospects of the presented datasets of ring-linear polymer blends was mainly made by K.H., T.M., T.D., H.J. and T.K. All authors have given their approval of the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-022-01138-3>.

Correspondence and requests for materials should be addressed to K.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2022