US011915799B2

(12) **United States Patent**
Minami et al.

(10) **Patent No.: US 11,915,799 B2**
(45) **Date of Patent: Feb. 27, 2024**

(54) **POLYMER PHYSICAL PROPERTY PREDICTION DEVICE, RECORDING MEDIUM, AND POLYMER PHYSICAL PROPERTY PREDICTION METHOD**

(71) Applicant: **SHOWA DENKO K.K.**, Tokyo (JP)

(72) Inventors: **Takuya Minami**, Tokyo (JP); **Yoshishige Okuno**, Tokyo (JP)

(73) Assignee: **RESONAC CORPORATION**, Tokyo (JP)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 849 days.

(21) Appl. No.: **17/008,999**

(22) Filed: **Sep. 1, 2020**

(65) **Prior Publication Data**

US 2020/0395102 A1 Dec. 17, 2020

Related U.S. Application Data

(63) Continuation-in-part of application No. PCT/JP2019/008704, filed on Mar. 5, 2019.

(30) **Foreign Application Priority Data**

Mar. 9, 2018 (JP) 2018-043171

(51) **Int. Cl.**
G16C 20/30 (2019.01)
C08F 110/02 (2006.01)
G01N 33/44 (2006.01)

(52) **U.S. Cl.**
CPC ***G16C 20/30*** (2019.02); ***C08F 110/02*** (2013.01); ***G01N 33/442*** (2013.01)

(58) **Field of Classification Search**
CPC G01N 33/442; G06N 20/10
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,260,882 A * 11/1993 Blanco C07K 1/00 703/12
6,687,621 B2 2/2004 Schneiderman et al.
(Continued)

FOREIGN PATENT DOCUMENTS

JP H06-271823 9/1994
JP 2004-514902 5/2004
(Continued)

OTHER PUBLICATIONS

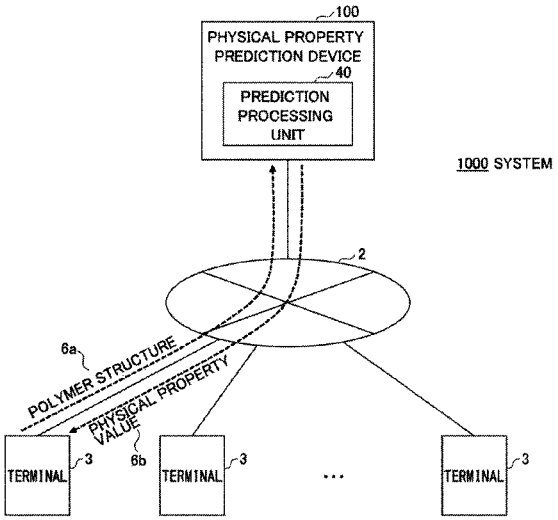
Takuya Minami et al., "Number Density Descriptor on Extended-Connectivity Fingerprints Combined with Machine Learning Approaches for Predicting Polymer Properties" ADMAT, (Undated).
(Continued)

Primary Examiner — Fred M Teskin
(74) *Attorney, Agent, or Firm* — IPUSA, PLLC

(57) **ABSTRACT**

A polymer physical property prediction device includes a processor, and a memory storing program instructions that cause the processor to read a structural unit from a storage unit and use the structural unit to calculate numbers each indicating how many substructures are in a polymer, the polymer being formed of repetition of the structural unit, calculate a number indicating how many atoms are in the structural unit, calculate number densities of the substructures from the numbers of substructures and the number of atoms in the structural unit, construct a regression model that predicts a physical property value by using an experimental value of a physical property of the polymer and the number densities of the substructures, input a polymer structure of which the physical property value is to be predicted, and predict the physical property value corresponding to the input polymer structure by using the regression model.

20 Claims, 16 Drawing Sheets



(56) **References Cited**

U.S. PATENT DOCUMENTS

2015/0310162	A1	10/2015	Okuno et al.
2018/0018408	A1	1/2018	Matsumura et al.
2019/0310207	A1	10/2019	Guzman Cardozo
2020/0051671	A1	2/2020	Maxson et al.
2020/0082913	A1	3/2020	Mir et al.

FOREIGN PATENT DOCUMENTS

JP	5946045	7/2016
JP	2016-180660	10/2016
JP	2018-010428	1/2018
WO	02/077772	10/2002
WO	2008/116495	10/2008
WO	2015/032505	3/2015

OTHER PUBLICATIONS

Takuya Minami et al., "Number Density Descriptor on Extended-Connectivity Fingerprints Combined with Machine Learning Approaches for Predicting Polymer Properties" MRS Advances © 2018 Materials Research Society DOI: 10.1557/adv.2018.454 Downloaded from https://www.cambridge.org/core. IP address: 103.9.96.234, on May 22, 2018 at 08:40:49.

2018 MRS Spring Meeting & Exhibit Apr. 2-6, 2018 Phoenix, Arizona—Session Details.

D. Rogers et al., "Extended-Connectivity Fingerprints" American Chemical Society, J. Chem. Info. Model., vol. 50, No. 5, 2010, pp. 742-754.

International Search Report dated May 28, 2019 with respect to PCT/JP2019/008704.

Journal of Computer Aided Chemistry [online], 2009, vol. 10, [retrieval date May 15, 2019], Internet: < URL:https://www.istage.jst.go.jp/article/jcac/10/0/10_0_30/_article/-char/ja> <DOI:10.2751/jcac.10.30>, pp. 30-37, (Goto, Shun et al., "Development of Predictive Models and Reverse Analysis Method for Polymer Design"), with Partial English Trans.

Extended European Search Report dated Nov. 3, 2021 with respect to the corresponding European patent application No. 19763785.3.

Office Action dated May 27, 2021 issued with respect to the related U.S. Appl. No. 16/666,643.

* cited by examiner

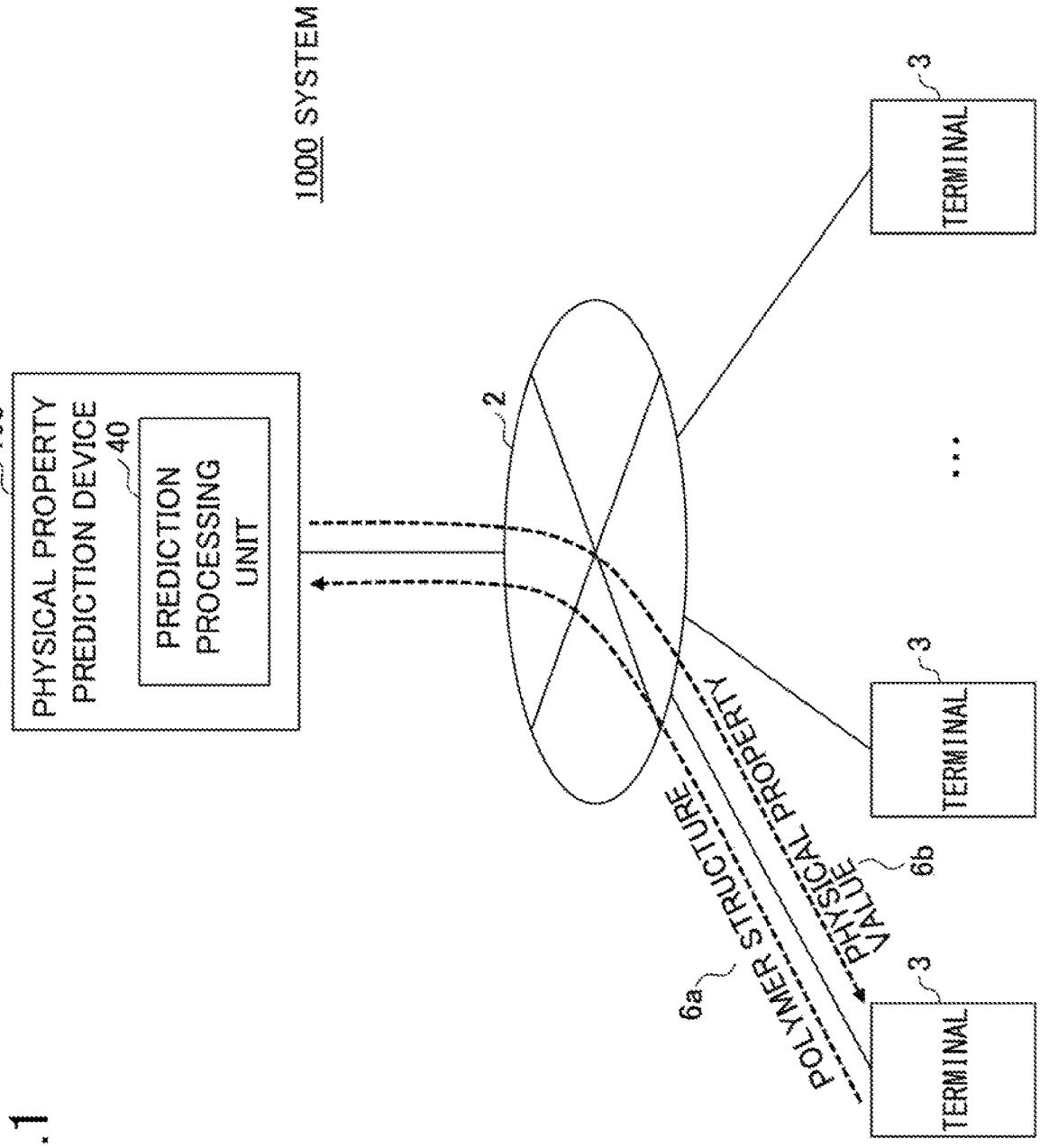
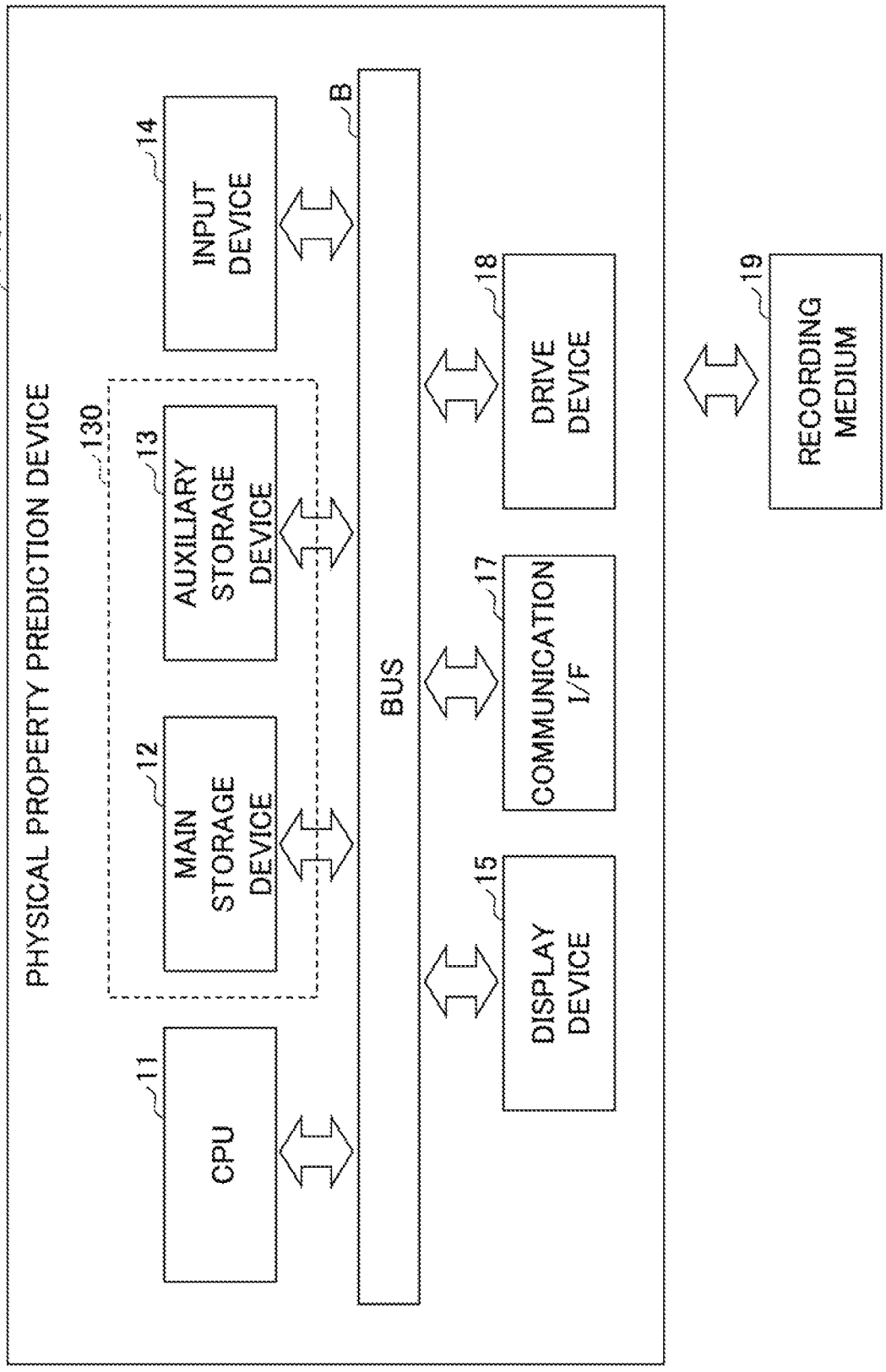

1000 SYSTEM
100
PHYSICAL PROPERTY PREDICTION DEVICE
40
PREDICTION PROCESSING UNIT
2
3
TERMINAL
3
TERMINAL
3
TERMINAL
...
6a
POLYMER STRUCTURE
6b
PHYSICAL PROPERTY VALUE

FIG.1

FIG.2


100
PHYSICAL PROPERTY PREDICTION DEVICE
130
11
CPU
12
MAIN STORAGE DEVICE
13
AUXILIARY STORAGE DEVICE
14
INPUT DEVICE
15
DISPLAY DEVICE
17
COMMUNICATION I/F
18
DRIVE DEVICE
19
RECORDING MEDIUM
B
BUS

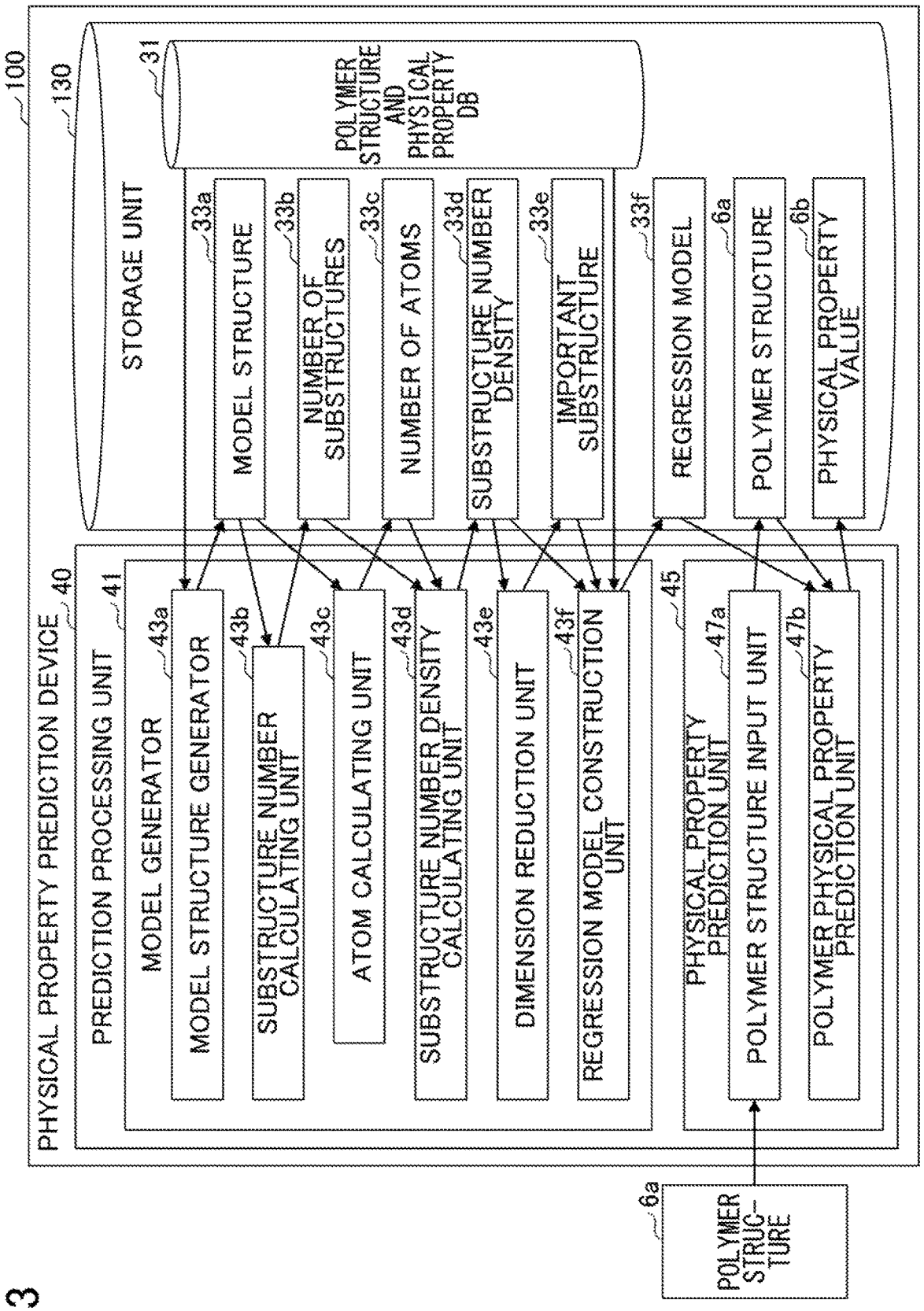

100
130
STORAGE UNIT
31
POLYMER STRUCTURE AND PHYSICAL PROPERTY DB
33a
MODEL STRUCTURE
33b
NUMBER OF SUBSTRUCTURES
33c
NUMBER OF ATOMS
33d
SUBSTRUCTURE NUMBER DENSITY
33e
IMPORTANT SUBSTRUCTURE
33f
REGRESSION MODEL
6a
POLYMER STRUCTURE
6b
PHYSICAL PROPERTY VALUE
PHYSICAL PROPERTY PREDICTION DEVICE
40
PREDICTION PROCESSING UNIT
41
MODEL GENERATOR
43a
MODEL STRUCTURE GENERATOR
43b
SUBSTRUCTURE NUMBER CALCULATING UNIT
43c
ATOM CALCULATING UNIT
43d
SUBSTRUCTURE NUMBER DENSITY CALCULATING UNIT
43e
DIMENSION REDUCTION UNIT
43f
REGRESSION MODEL CONSTRUCTION UNIT
45
PHYSICAL PROPERTY PREDICTION UNIT
47a
POLYMER STRUCTURE INPUT UNIT
47b
POLYMER PHYSICAL PROPERTY PREDICTION UNIT
6a
POLYMER STRUCTURE

FIG.3

FIG.4

31
POLYMER STRUCTURE AND PHYSICAL PROPERTY DB
POLYMER NAME
REPEATING UNIT
REFRACTIVE INDEX
GLASS TRANSITION POINT
poly(ethylene)
CC
1.472
poly(propylene)
C(C)C
1.469
poly(butene)
C(C)C(C)
1.474
poly(vinyl alcohol)
C(O)C
1.546
poly(ethyleneoxide)
CCO
1.4638
32
POLYMER NAME
REPEATING UNIT
NUMBER OF ATOMS
SUBSTRUCTURE NUMBER DENSITY

FIG.5

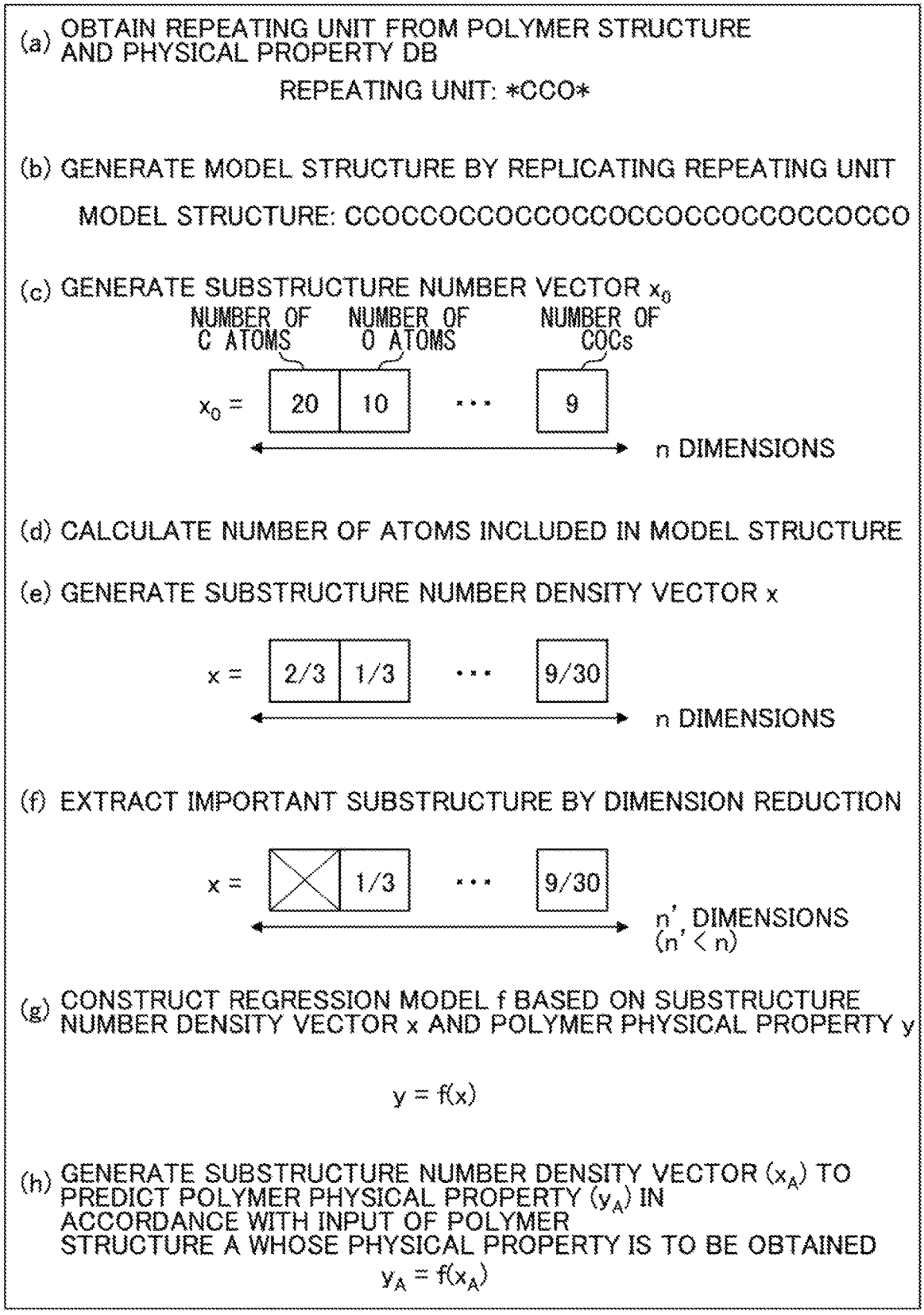

(a) OBTAIN REPEATING UNIT FROM POLYMER STRUCTURE AND PHYSICAL PROPERTY DB
REPEATING UNIT: *CCO*
(b) GENERATE MODEL STRUCTURE BY REPLICATING REPEATING UNIT
MODEL STRUCTURE: CCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCO
(c) GENERATE SUBSTRUCTURE NUMBER VECTOR x_0
NUMBER OF C ATOMS
NUMBER OF O ATOMS
NUMBER OF COCs
x_0 =
20
10
9
n DIMENSIONS
(d) CALCULATE NUMBER OF ATOMS INCLUDED IN MODEL STRUCTURE
(e) GENERATE SUBSTRUCTURE NUMBER DENSITY VECTOR x
x =
2/3
1/3
9/30
n DIMENSIONS
(f) EXTRACT IMPORTANT SUBSTRUCTURE BY DIMENSION REDUCTION
x =
1/3
9/30
n', DIMENSIONS
(n' < n)
(g) CONSTRUCT REGRESSION MODEL f BASED ON SUBSTRUCTURE NUMBER DENSITY VECTOR x AND POLYMER PHYSICAL PROPERTY y
y = f(x)
(h) GENERATE SUBSTRUCTURE NUMBER DENSITY VECTOR (x_A) TO PREDICT POLYMER PHYSICAL PROPERTY (y_A) IN ACCORDANCE WITH INPUT OF POLYMER STRUCTURE A WHOSE PHYSICAL PROPERTY IS TO BE OBTAINED
$y_A = f(x_A)$

FIG.6

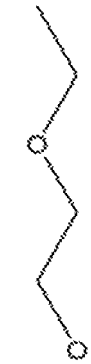
	REPEATING UNIT	MODEL	NUMBER OF CARBON ATOMS	CARBON ATOM NUMBER DENSITY
STRUCTURE A			2	2/3
STRUCTURE B			4	2/3

FIG.7

	REPEATING UNIT	MODEL	COC NUMBER DENSITY
NO REPLICATION			0
REPLICATION (PRESENT EMBODIMENT)			1/6

FIG.8A

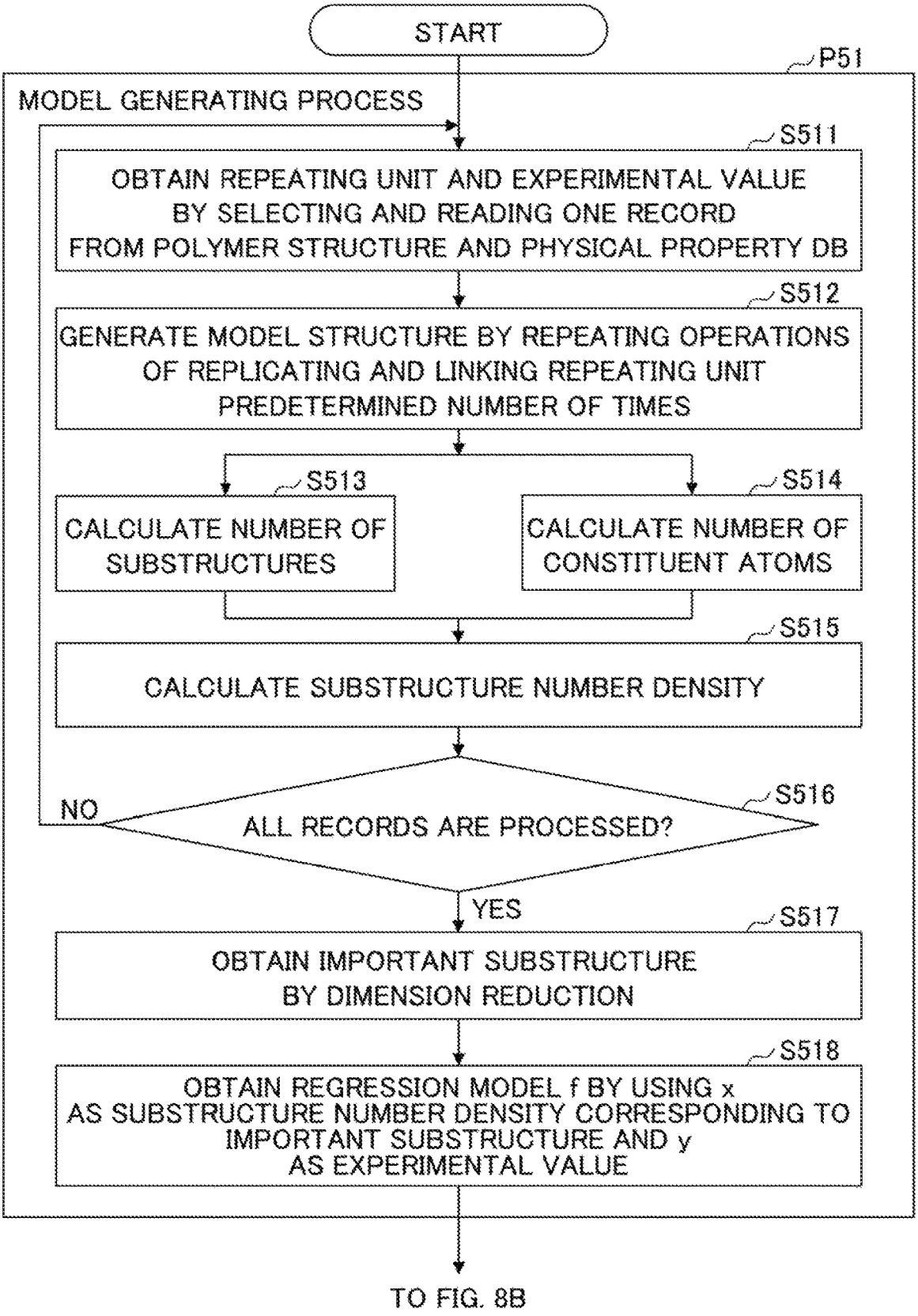

START
P51
MODEL GENERATING PROCESS
S511
OBTAIN REPEATING UNIT AND EXPERIMENTAL VALUE BY SELECTING AND READING ONE RECORD FROM POLYMER STRUCTURE AND PHYSICAL PROPERTY DB
S512
GENERATE MODEL STRUCTURE BY REPEATING OPERATIONS OF REPLICATING AND LINKING REPEATING UNIT PREDETERMINED NUMBER OF TIMES
S513
CALCULATE NUMBER OF SUBSTRUCTURES
S514
CALCULATE NUMBER OF CONSTITUENT ATOMS
S515
CALCULATE SUBSTRUCTURE NUMBER DENSITY
S516
NO
ALL RECORDS ARE PROCESSED?
YES
S517
OBTAIN IMPORTANT SUBSTRUCTURE BY DIMENSION REDUCTION
S518
OBTAIN REGRESSION MODEL f BY USING x AS SUBSTRUCTURE NUMBER DENSITY CORRESPONDING TO IMPORTANT SUBSTRUCTURE AND y AS EXPERIMENTAL VALUE
TO FIG. 8B

FIG.8B

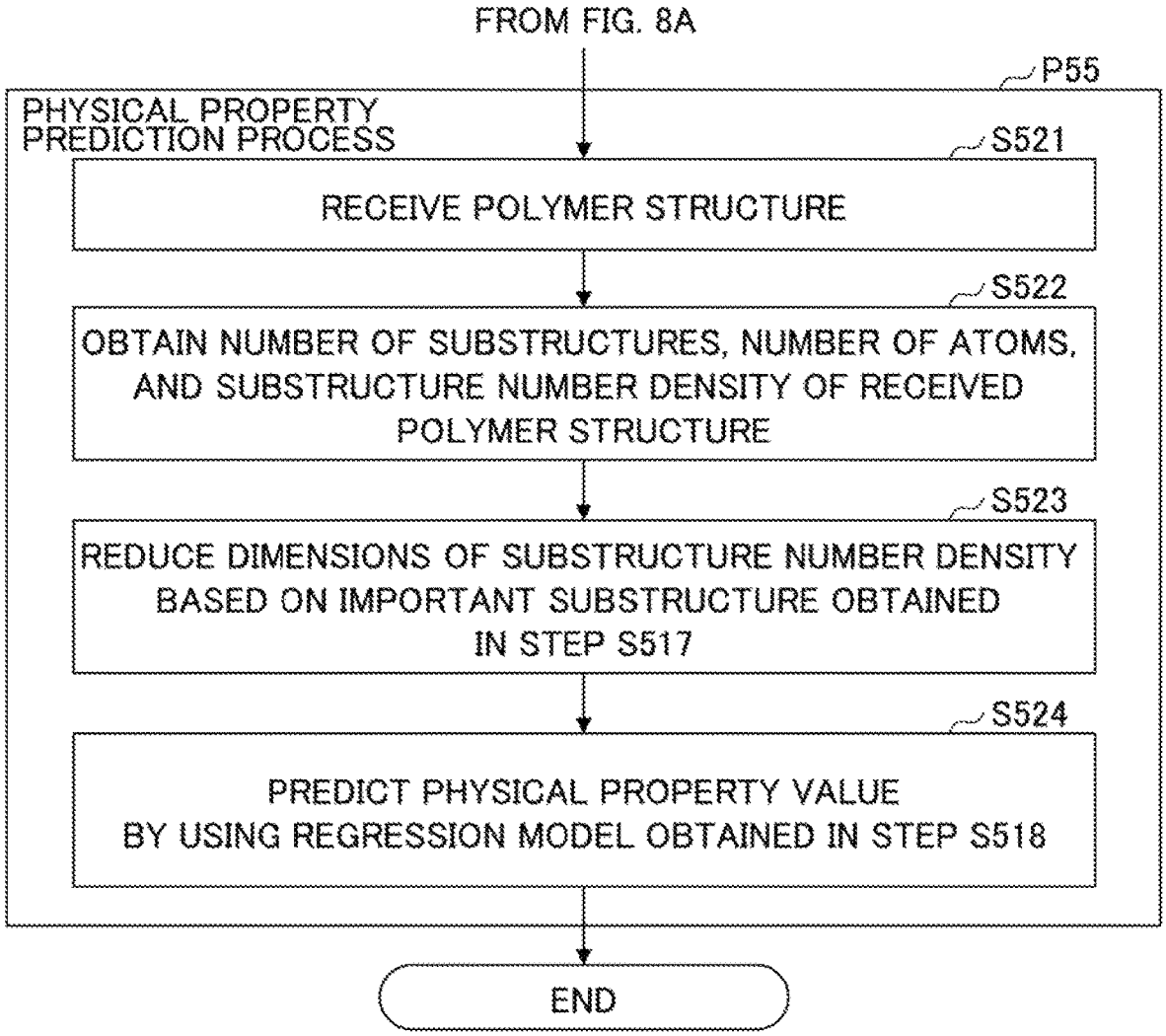

FROM FIG. 8A
P55
PHYSICAL PROPERTY PREDICTION PROCESS
S521
RECEIVE POLYMER STRUCTURE
S522
OBTAIN NUMBER OF SUBSTRUCTURES, NUMBER OF ATOMS, AND SUBSTRUCTURE NUMBER DENSITY OF RECEIVED POLYMER STRUCTURE
S523
REDUCE DIMENSIONS OF SUBSTRUCTURE NUMBER DENSITY BASED ON IMPORTANT SUBSTRUCTURE OBTAINED IN STEP S517
S524
PREDICT PHYSICAL PROPERTY VALUE BY USING REGRESSION MODEL OBTAINED IN STEP S518
END

FIG.9

33k

PHYSICAL PROPERTY	REGRESSION MODEL	IMPORTANT SUBSTRUCTURE	...
REFRACTIVE INDEX	$f_{m1}(x)$	$v_2, v_7, \cdots, v_p$	...
GLASS TRANSITION POINT	$f_{m2}(x)$	$v_3, v_4, \cdots, v_q$	...
⋮	⋮	⋮	⋮

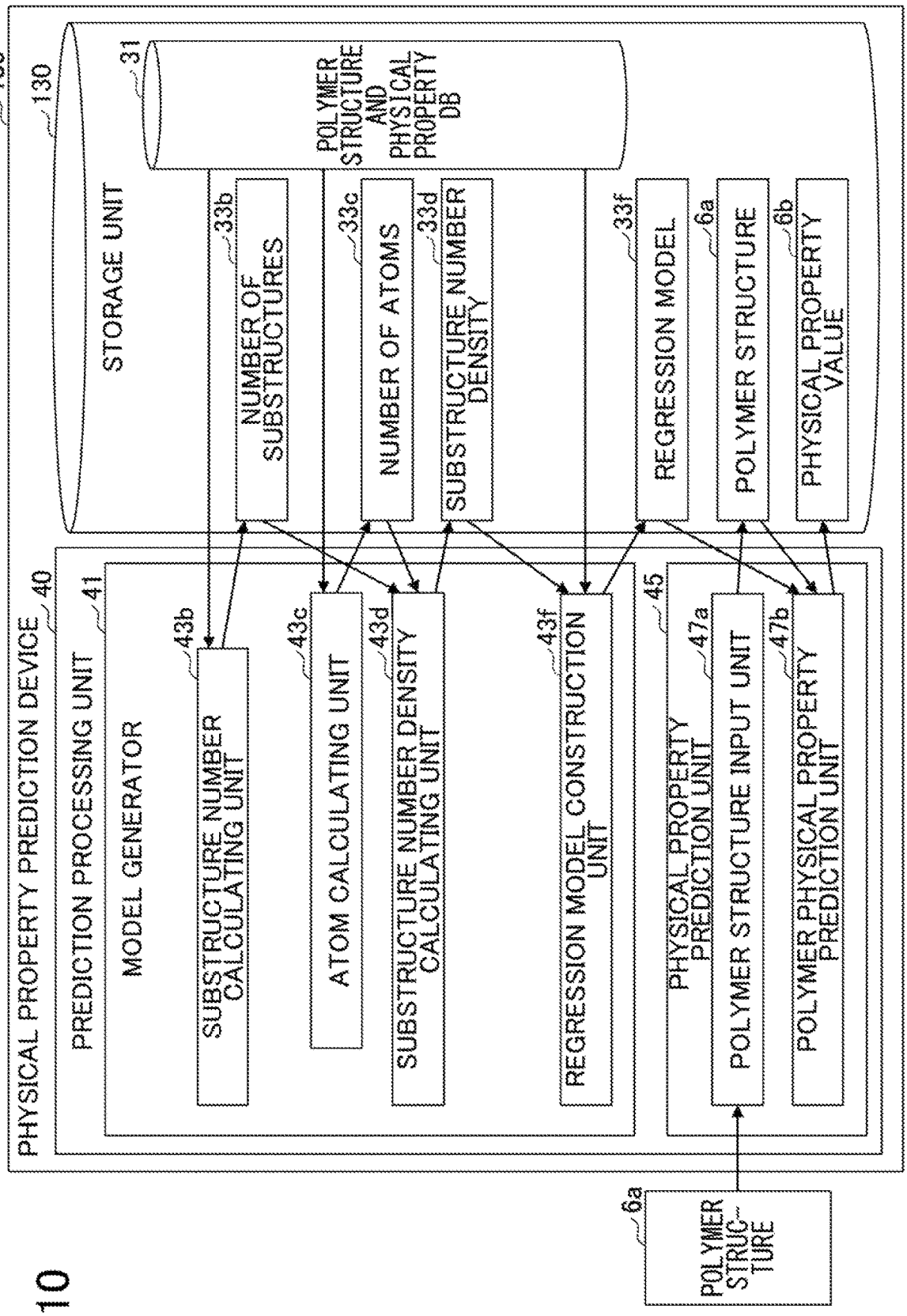

100
130
STORAGE UNIT
31
POLYMER STRUCTURE AND PHYSICAL PROPERTY DB
33b
NUMBER OF SUBSTRUCTURES
33c
NUMBER OF ATOMS
33d
SUBSTRUCTURE NUMBER DENSITY
33f
REGRESSION MODEL
6a
POLYMER STRUCTURE
6b
PHYSICAL PROPERTY VALUE
PHYSICAL PROPERTY PREDICTION DEVICE
40
PREDICTION PROCESSING UNIT
41
MODEL GENERATOR
43b
SUBSTRUCTURE NUMBER CALCULATING UNIT
43c
ATOM CALCULATING UNIT
43d
SUBSTRUCTURE NUMBER DENSITY CALCULATING UNIT
43f
REGRESSION MODEL CONSTRUCTION UNIT
45
PHYSICAL PROPERTY PREDICTION UNIT
47a
POLYMER STRUCTURE INPUT UNIT
47b
POLYMER PHYSICAL PROPERTY PREDICTION UNIT
6a
POLYMER STRUC-TURE

FIG.10

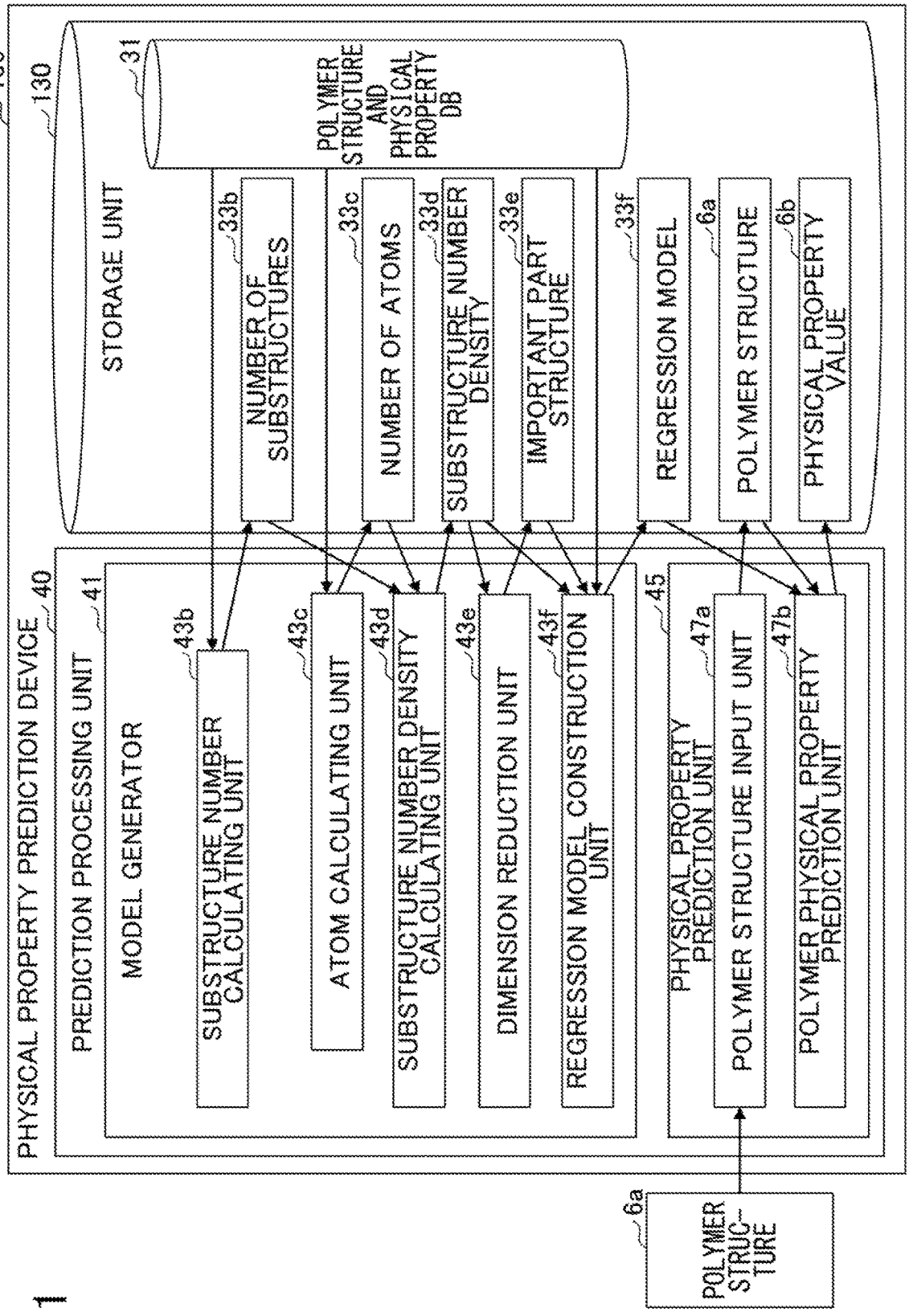

100
130
STORAGE UNIT
31
POLYMER STRUCTURE AND PHYSICAL PROPERTY DB
33b
NUMBER OF SUBSTRUCTURES
33c
NUMBER OF ATOMS
33d
SUBSTRUCTURE NUMBER DENSITY
33e
IMPORTANT PART STRUCTURE
33f
REGRESSION MODEL
6a
POLYMER STRUCTURE
6b
PHYSICAL PROPERTY VALUE
PHYSICAL PROPERTY PREDICTION DEVICE
40
PREDICTION PROCESSING UNIT
41
MODEL GENERATOR
43b
SUBSTRUCTURE NUMBER CALCULATING UNIT
43c
ATOM CALCULATING UNIT
43d
SUBSTRUCTURE NUMBER DENSITY CALCULATING UNIT
43e
DIMENSION REDUCTION UNIT
43f
REGRESSION MODEL CONSTRUCTION UNIT
45
PHYSICAL PROPERTY PREDICTION UNIT
47a
POLYMER STRUCTURE INPUT UNIT
47b
POLYMER PHYSICAL PROPERTY PREDICTION UNIT
6a
POLYMER STRUCTURE

FIG.11

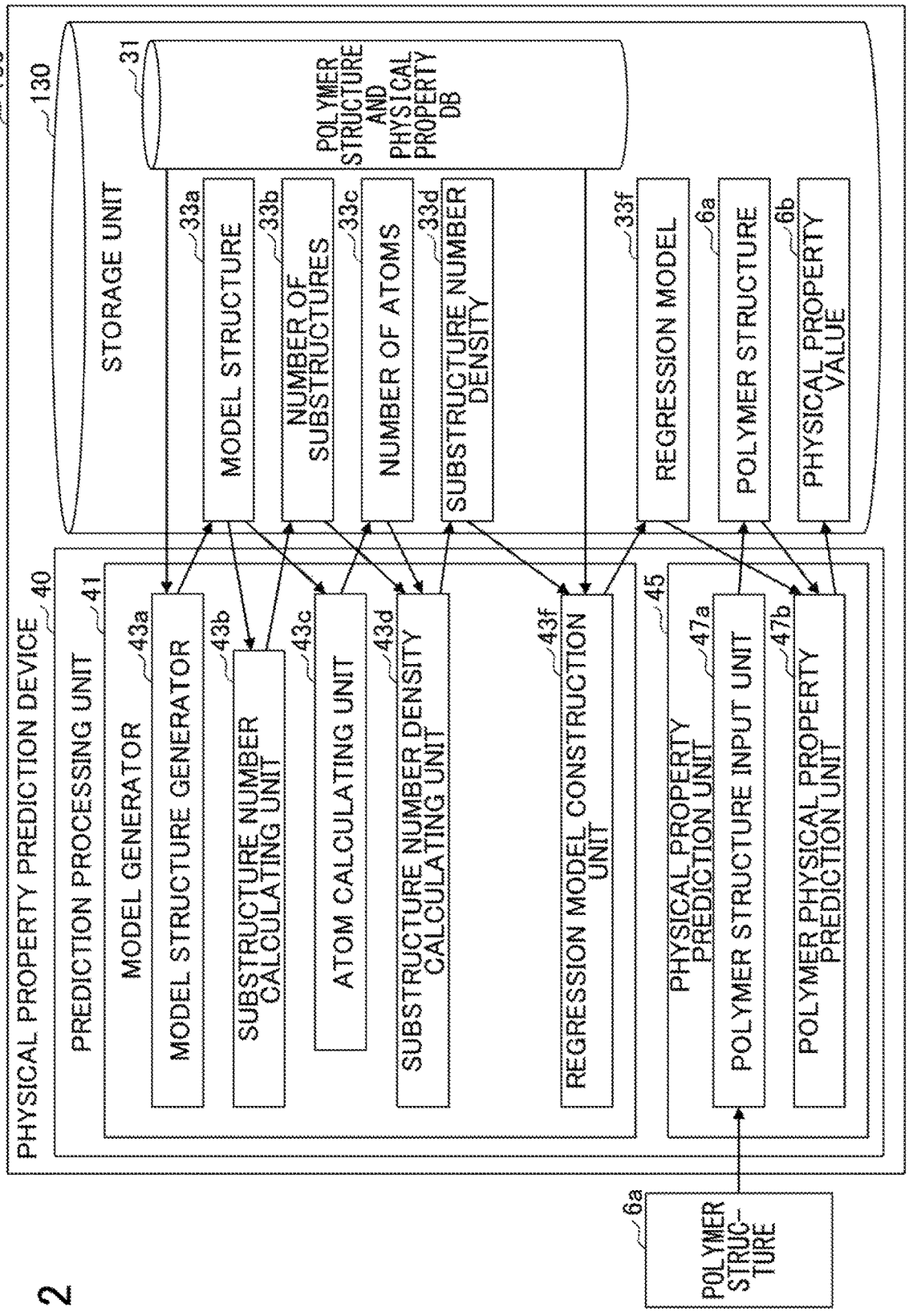

100
130
STORAGE UNIT
31
POLYMER STRUCTURE AND PHYSICAL PROPERTY DB
33a
MODEL STRUCTURE
33b
NUMBER OF SUBSTRUCTURES
33c
NUMBER OF ATOMS
33d
SUBSTRUCTURE NUMBER DENSITY
33f
REGRESSION MODEL
6a
POLYMER STRUCTURE
6b
PHYSICAL PROPERTY VALUE
PHYSICAL PROPERTY PREDICTION DEVICE
40
PREDICTION PROCESSING UNIT
41
MODEL GENERATOR
43a
MODEL STRUCTURE GENERATOR
43b
SUBSTRUCTURE NUMBER CALCULATING UNIT
43c
ATOM CALCULATING UNIT
43d
SUBSTRUCTURE NUMBER DENSITY CALCULATING UNIT
43f
REGRESSION MODEL CONSTRUCTION UNIT
45
PHYSICAL PROPERTY PREDICTION UNIT
47a
POLYMER STRUCTURE INPUT UNIT
47b
POLYMER PHYSICAL PROPERTY PREDICTION UNIT
6a
POLYMER STRUCTURE

FIG.12

RELATED ART

FIG.13

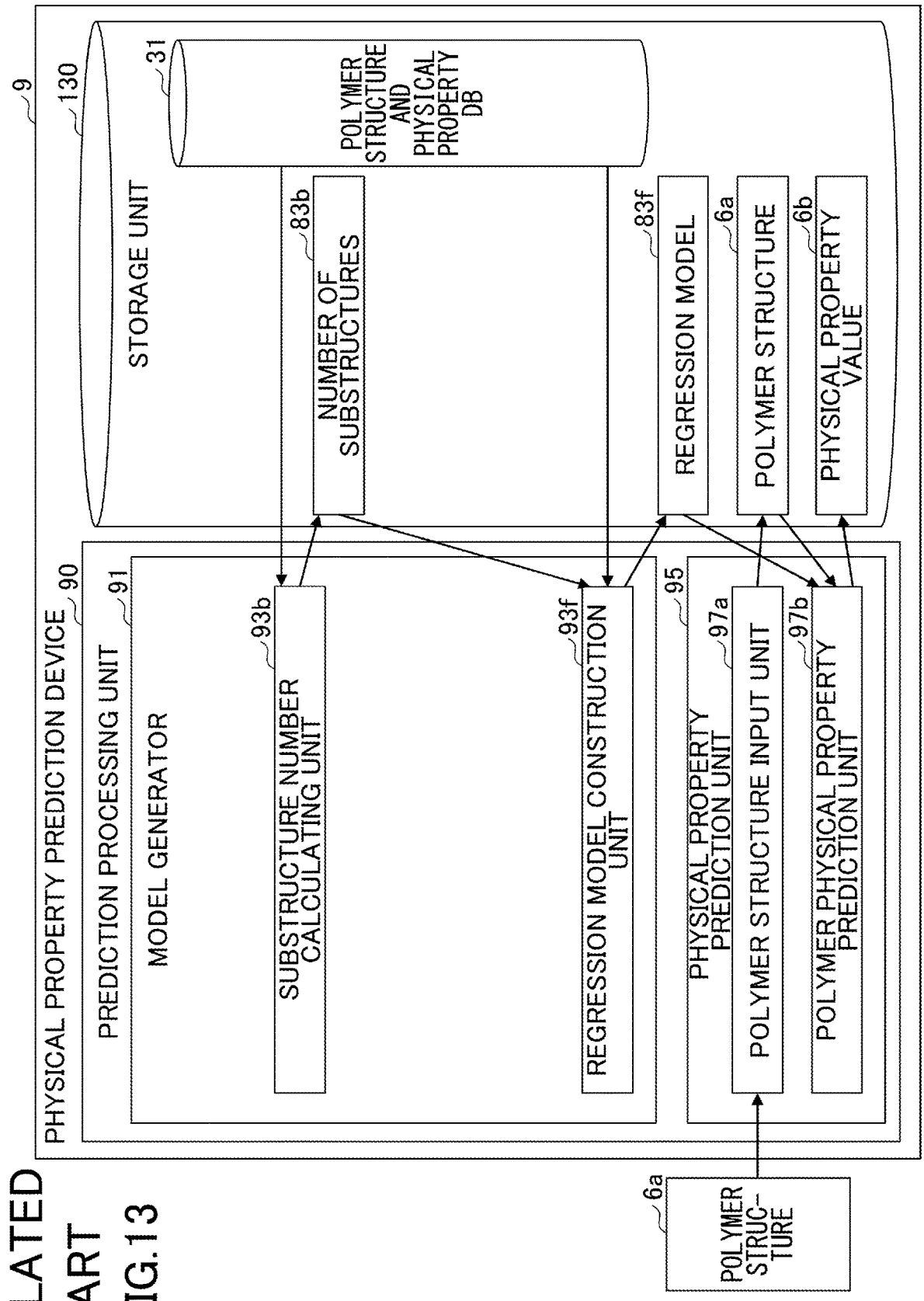


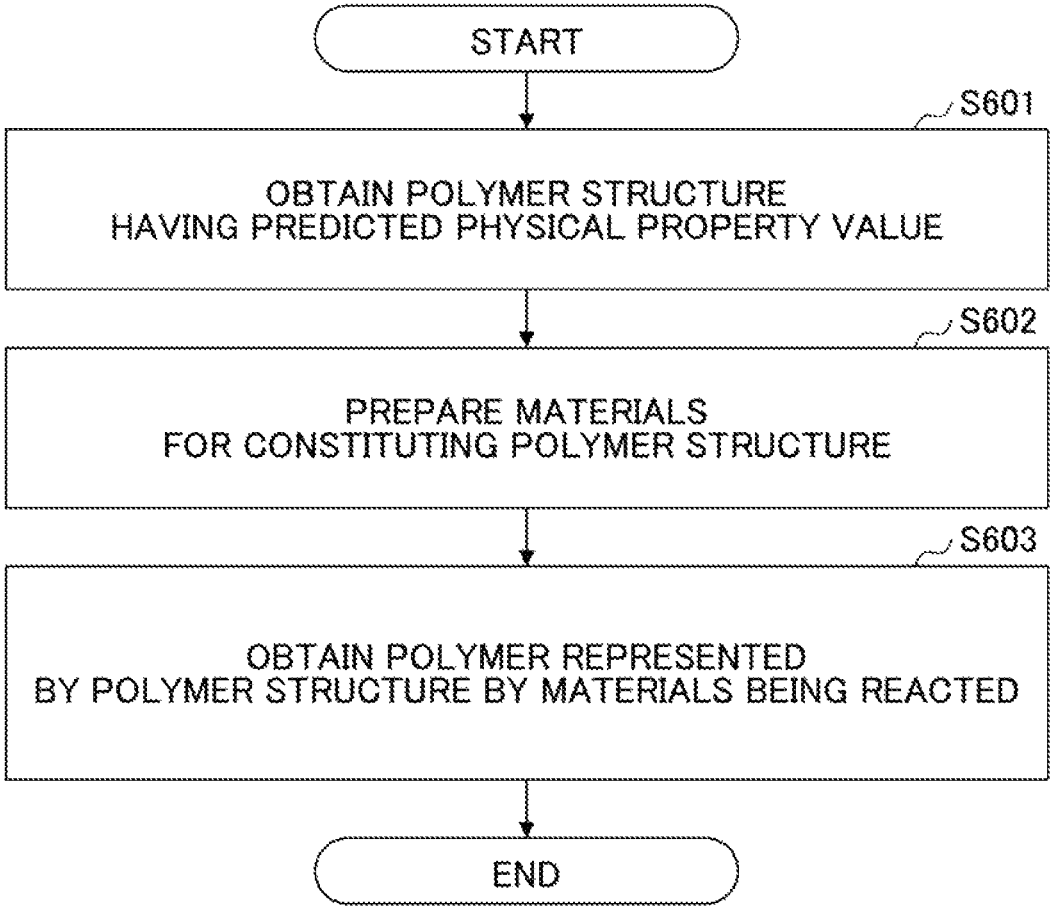
FIG.14

	COMPARATIVE EXAMPLE 1	EXAMPLE 1	EXAMPLE 2	EXAMPLE 3	EXAMPLE 4
R^2	−0.063	0.890	0.900	0.932	0.950
RMSE	0.084	0.027	0.026	0.021	0.018

FIG.15

	PHYSICAL PROPERTY	NUMBER OF DATA	R^2
EXAMPLE 4	REFRACTIVE INDEX	51	0.95
EXAMPLE 5	GLASS TRANSITION POINT [K]	417	0.84
EXAMPLE 6	BAND GAP [eV]	284	0.84
EXAMPLE 7	RELATIVE PERMITTIVITY	264	0.76
EXAMPLE 8	COEFFICIENT OF LINEAR EXPANSION [10^{-5}/K]	54	0.64
EXAMPLE 9	DENSITY [g/cm^3]	51	0.95

FIG.16


START
S601
OBTAIN POLYMER STRUCTURE HAVING PREDICTED PHYSICAL PROPERTY VALUE
S602
PREPARE MATERIALS FOR CONSTITUTING POLYMER STRUCTURE
S603
OBTAIN POLYMER REPRESENTED BY POLYMER STRUCTURE BY MATERIALS BEING REACTED
END

POLYMER PHYSICAL PROPERTY PREDICTION DEVICE, RECORDING MEDIUM, AND POLYMER PHYSICAL PROPERTY PREDICTION METHOD

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part application of International Application No. PCT/JP2019/008704 filed on Mar. 5, 2019, which is based upon and claims priority to Japanese Priority Application No. 2018-043171 filed on Mar. 9, 2018, the entire contents of which are hereby incorporated by reference.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to a device that predicts the physical properties of a polymer (hereinafter "polymer physical property prediction device", a recording medium, and a polymer physical property prediction method.

2. Description of the Related Art

In recent years, machine learning has been used to predict physical properties of organic compounds such as polymers by using a computer.

Conventional techniques for predicting physical properties of polymers use various structure descriptors configured by humans to predict physical properties. Techniques for mechanically identifying substructure patterns of molecules and calculating structure descriptors based on algorithms have also been proposed.

Conventionally, it is necessary for a human to determine a structural pattern, such as the number of carbon atoms, that serves as a structure descriptor of a polymer in advance. With respect to this, there is a method for automatically extracting substructures of low-molecular weight organic molecules, but there is a problem that the accuracy of predicting the physical properties of a polymer is low. Thus, the method has not been applied to polymer structures.

RELATED-ART DOCUMENTS

Patent Documents

[Patent Document 1] Japanese National Publication of International Patent Application No. 2004-514902

Non-Patent Document

[Non-Patent Document 1] David Rogers et al., "Extended-Connectivity Fingerprints", American Chemical Society, J. Chem. Info. Model., Vol. 50, No. 5, 2010, pp. 742-754.

SUMMARY OF THE INVENTION

According to one aspect, a polymer physical property prediction device includes a processor, and a memory storing program instructions that cause the processor to read a structural unit from a storage unit and use the structural unit to calculate numbers each indicating how many substructures are in a polymer, the polymer being formed of repetition of the structural unit, calculate a number indicating how many atoms are in the structural unit, calculate number densities of the substructures from the numbers of substructures and the number of atoms in the structural unit, construct a regression model that predicts a physical property value by using an experimental value of a physical property of the polymer and the number densities of the substructures, the experimental value being obtained from the storage unit, input a polymer structure of which the physical property value is to be predicted, and predict the physical property value corresponding to the input polymer structure by using the regression model.

A means to solve the above-described problem can be a program or a method for predicting the physical properties of a polymer.

According to at least one embodiment of the present invention, the physical properties of a polymer can be accurately predicted.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. **1** is a drawing illustrating an example of a system configuration according to the present embodiment;

FIG. **2** is a drawing illustrating an example of a hardware configuration of a physical property prediction device;

FIG. **3** is a drawing illustrating a first functional configuration example of the physical property prediction device;

FIG. **4** is a drawing illustrating an example of a data configuration of a polymer structure and physical property DB;

FIG. **5** is a drawing for describing an overview of a prediction process according to the present embodiment;

FIG. **6** is a drawing illustrating an example of the number density;

FIG. **7** is a drawing for describing a difference caused when replication of a repeating unit is performed and no replication of a repeating unit is performed;

FIG. **8**A is a flowchart for describing the prediction process;

FIG. **8**B is a flowchart for describing the prediction process;

FIG. **9** is a drawing illustrating an example of a data configuration of a regression model management table;

FIG. **10** is a drawing illustrating a second functional configuration example of the physical property prediction device;

FIG. **11** is a drawing illustrating a third functional configuration example of the physical property prediction device;

FIG. **12** is a drawing illustrating a fourth functional configuration example of the physical property prediction device;

FIG. **13** is a drawing illustrating a functional configuration example of a physical property prediction device based on an existing technique;

FIG. **14** is a drawing illustrating an example of evaluation results;

FIG. **15** is a drawing illustrating an example of evaluation results; and

FIG. **16** is a drawing illustrating an example of a manufacturing method for manufacturing a polymer.

DESCRIPTION OF THE EMBODIMENTS

In the following, an embodiment of the present invention will be described based on the drawings. First, an example of a system configuration according to the present embodiment will be described. FIG. **1** is a drawing illustrating the

example of the system configuration according to the present embodiment In FIG. **1**, a system **1000** includes a physical property prediction device **100** and multiple terminals **3**, and each of the terminals can be connected through a network **2**.

The terminal **3** is an information processing terminal used by a user, is connected to the physical property prediction device **100**, and utilizes processing performed by a prediction processing unit **40**.

The physical property prediction device **100** is an information processing device including the prediction processing unit **40** that predicts a physical property of a polymer. In the physical property prediction device **100**, the prediction processing unit **40** performs machine learning using a structure of the polymer and data indicating the physical property of the polymer that are provided in advance to generate a regression model. In response to receiving a polymer structure **6***a* from the terminal **3**, a physical property value is predicted using the generated regression model with respect to the received polymer structure **6***a*, and a predicted physical property value **6***b* is sent to the terminal **3**.

A configuration of the system **1000** is not limited to the example of FIG. **1**. As an example, the physical property prediction device **100** may be utilized in a stand-alone manner. In this case, when the polymer structure **6***a* is input by a user operation of an input device **14** (FIG. **2**), the physical property value **6***b* is output to a display device **15** (FIG. **2**).

The physical property prediction device **100** according to the present embodiment has a hardware configuration as illustrated in FIG. **2**. FIG. **2** is a drawing illustrating an example of the hardware configuration of the physical property prediction device. In FIG. **2**, the physical property prediction device **100** is an information processing device controlled by a computer and includes a central processing unit (CPU) **11**, a main storage device **12**, an auxiliary storage device **13**, the input device **14**, the display device **15**, a communication interface (I/F) **17**, and a drive device **18**, which are connected to a bus B.

The CPU **11** corresponds to a processor that controls the physical property prediction device **100** according to a program stored in the main storage device **12**. As the main storage device **12**, a random access memory (RAM) and a read only memory (ROM) are used to store or temporarily store a program executed by the CPU **11**, data necessary for processing by the CPU **11**, and data obtained by processing by the CPU **11**, for example.

A hard disk drive (HDD) is used for the auxiliary storage device **13**, for example, to store data such as a program for executing various processing. A portion of the program stored in the auxiliary storage device **13** is loaded into the main storage device **12** and is executed by the CPU **11** to achieve various processing. Hereinafter, the main storage device **12**, the auxiliary storage device **13**, and an external storage device accessible by the physical property prediction device **100** will be collectively referred to as a storage unit **130**.

The input device **14** includes a mouse and a keyboard and is used by a user to input various information required for processing by the physical property prediction device **100**. The display device **15** displays various necessary information under control of the CPU **11**. The input device **14** and the display device **15** may be a user interface, such as an integrated touch panel. The communication I/F **17** performs communication through a network, such as a wired or wireless network. The communication by the communication I/F **17** is not limited to the wired or wireless communication. The drive device **18** interfaces a recording medium **19** set in the drive device **18** (e.g., a compact disc read-only memory (CD-ROM)) with the physical property prediction device **100**.

A program for implementing the processing performed by the physical property prediction device **100** is provided to the physical property prediction device **100** through the recording medium **19**, such as a CD-ROM. A program for implementing various processing according to the present embodiment, which will be described later, is stored in the recording medium **19**, and the program stored in the recording medium **19** is installed in the physical property prediction device **100** through the drive device **18**. The installed program can be executed by the physical property prediction device **100**.

The recording medium **19** for storing the program is not limited to a CD-ROM, but may be one or more non-transitory and tangible media having a computer readable data structure. As a computer readable recording medium in addition to a CD-ROM, a portable storage medium, such as a digital versatile disk (DVD) and a USB memory, or a semiconductor memory, such as a flash memory, may be used.

[First Functional Configuration Example]

FIG. **3** is a drawing illustrating a first functional configuration example of the physical property prediction device. In FIG. **3**, the physical property prediction device **100** includes the prediction processing unit **40**. The prediction processing unit **40** corresponds to a prediction process performed by the CPU **11** executing a corresponding program.

The storage unit **130** stores a polymer structure and physical property DB **31**, a model structure **33***a*, the number of substructures **33***b*, the number of atoms **33***c*, a substructure number density **33***d*, an important substructure **33***e*, a regression model **33***f*, a polymer structure **6***a*, a physical property value **6***b*, for example.

The prediction processing unit **40** obtains the regression model **33***f* for predicting the physical properties of the polymer structure **6***a* specified by a user by performing machine learning, predicts the physical properties of the polymer structure **6***a* by using the obtained regression model **33***f*, and outputs the obtained physical property value **6***b*. The prediction processing unit **40** includes a model generator **41** and a physical property prediction unit **45** as processing units.

The model generator **41** learns the physical property of the polymer structure in which a structural unit is repeatedly replicated and linked using the polymer structure and physical property DB**31** and determines the regression model **33***f*. A structural unit forming a polymer by repetition is called a "repeating unit". In the polymer structure and physical property DB **31**, a repeating unit and an experimental value of a physical property experimentally obtained in advance are stored in association with each polymer.

The model generator **41** includes a model structure generator **43***a*, a substructure number calculating unit **43***b*, an atom calculating unit **43***c*, a substructure number density calculating unit **43***d*, a dimension reduction unit **43***e*, and a regression model construction unit **43***f* as processing units. Either or both of the model structure generator **43***a* and the dimension reduction unit **43***e* can be removed.

The model structure generator **43***a* obtains the repeating unit from the polymer structure and physical property DB **31** and repeats the process of replicating and linking the obtained repeating unit a predetermined number of times to generate the model structure **33***a*. The repeating unit includes a substructure. The repeating unit may also be an

entire of a substructure. The repeating unit are linked multiple times to represent a polymer.

The generated model structure **33***a* is stored in the storage unit **130**. The repeating unit and the model structure **33***a* may be represented in a predetermined form. The first functional configuration example describes the repeating unit and the model structure **33***a* in the simplified molecular input line entry syntax (SMILES) notation, but the notation is not limited to this.

The substructure number calculating unit **43***b* calculates the number of substructures **33***b* based on the model structure **33***a*. The numbers of substructures **33***b* are represented by an n-dimensional vector and stored in the storage **130**. The atom calculating unit **43***c* calculates the number of atoms **33***c* based on the model structure **33***a*. The obtained number of atoms **33***c* is stored in the storage unit **130**.

The substructure number density calculating unit **43***d* calculates the substructure number density **33***d* by using the numbers of substructures **33***b* and the number of atoms **33***c*. The substructure number density **33***d* can be obtained by the following equation:

$$\text{substructure number density} = \text{number of substructures} \div \text{number of atoms}$$

The obtained substructure number densities **33***d* are represented by a vector of the same n-dimensional vector as the vector of the substructure number **33***b* and are stored in the storage unit **130**.

The dimension reduction unit **43***e* constructs a temporal regression model by using the substructure number density **33***d* and the experimental value, and reduces the dimensions of the temporal regression model based on a predetermined dimension reduction method to obtain the important substructure **33***e*. The important substructure **33***e* is stored in the storage unit **130**. The important substructure **33***e* is data specifying an element (e.g., a vector) that will contribute to the prediction accuracy of the regression model **33***f* to be constructed in the substructure number density **33***d*.

For the dimension reduction method, a procedure corresponding to a regression analysis method to be used is determined in advance. For example, when the lasso regression is used, the dimensions can be reduced by removing a coefficient term of 0. When the random forest regression is used, the dimensions can be reduced by calculating the importance of explanatory variables and excluding variables that have the importance lower than or equal to a threshold. In the stepwise method, the dimensions can be reduced by searching for combinations of explanatory variables having highest prediction accuracy by adding or reducing an explanatory variable one by one when a temporal regression model is constructed.

The regression model construction unit **43***f* constructs the regression model **33***f* by performing a regression analysis based on the important substructure **33***e* obtained by the dimension reduction unit **43***e* by using the substructure number density **33***d* and the experimental value. The regression model **33***f* is stored in the storage unit **130**.

In the dimension reduction unit **43***e* and the regression model construction unit **43***f*, the experimental value can be obtained from the polymer structure and physical property DB **31**. The regression model **33***f* obtained by the regression model construction unit **43***f* is expressed in the following equation.

$$y = f(x)$$

By substituting the substructure number density **33***d* for a variable x, a predicted value y of the physical property is obtained.

To the regression analysis for obtaining the temporal regression model and the regression model **33***f*, any one of Lasso regression, random forest regression, a stepwise method, partial linear regression, ridge regression, kernel ridge regression, support vector regression, decision tree regression, Gaussian process regression, and so on can be applied. As an example, a temporal regression model can be obtained by the Lasso regression, and the regression model **33***f* can be obtained by the Gaussian process regression.

Next, the physical property prediction unit **45** predicts the physical properties of the polymer based on the polymer structure **6***a* by using the regression model **33***f* in response to an input of the polymer structure **6***a* specified by a user. The polymer structure **6***a* may specify the repeating unit by the SMILES notation or the like. The physical property prediction unit **45** includes a polymer structure input unit **47***a* and a polymer physical property prediction unit **47***b* as processing units.

The polymer structure input unit **47***a* inputs the polymer structure **6***a* specified by a user and the polymer structure **6***a* is stored in the storage unit **130**. The polymer structure input unit **47***a* may input the polymer structure **6***a* by the communication I/F **17** receiving the polymer structure **6***a* through the network **2** or may input the polymer structure **6***a* in accordance with a user's specification by the input device **14** of the property prediction device **100**.

The polymer physical property prediction unit **47***b* obtains the model structure, the numbers of substructures, and the number of atoms, from the polymer structure **6***a*, to calculate the substructure number density, and predicts the physical property value by using the regression model **33***f* stored in the storage unit **130**. The method of obtaining the model structure, the numbers of substructures, the number of atoms, and the substructure number density based on the polymer structure **6***a* is similar to the method of the model generator **41**.

The predicted physical property value **6***b* is output to the storage unit **130**. When the polymer structure **6***a* is specified through the network **2**, the obtained physical property value **6***b* is sent to the terminal **3** of the user and displayed. When the polymer structure **6***a* is specified by a user operation at the physical property prediction device **100**, the obtained physical property value **6***b* is displayed on the display device **15**.

FIG. **4** is a drawing illustrating an example of a data configuration of the polymer structure and physical property DB. In FIG. **4**, the polymer structure and physical property DB **31** is a database that manages the repeating unit and experimental values of various physical properties, such as a refractive index, for each polymer, and includes items such as a polymer name, the repeating unit, the refractive index, and a glass transition point.

The polymer name indicates specific information specifying the polymer structure. An example in which the repeating unit is represented by the SMILES notation is used for the following description, but the notation of the repeating unit is not limited to the SMILES notation. The refractive index, the glass transition point, and the like are examples of various physical properties of the polymer experimentally obtained in advance, and the experimental values are shown for each name of the physical property.

In this example, a repeating unit "*CC*" is associated with a polymer name "poly(ethylene)" and the experimental values, such as the refractive index "1.472", are shown. Similarly, an experimental value of each physical property is shown for another polymer.

In the model generator **41**, referenced data, generated data, and the like may be managed in a work table **32**. The work table **32** is a table in which calculated values are stored for each repeating unit of the polymer of which the substructure number density **33***d* is obtained, and includes items such as the polymer name, the repeating unit, the number of atoms, and the substructure number density.

The polymer name indicates a polymer name registered in the polymer structure and physical property DB **31**. The number of substructures indicates the number of substructures **33***b* obtained by the substructure number calculating unit **43***b*. The number of atoms indicates the number of atoms **33***c* obtained by the atom calculating unit **43***c*. The substructure number density indicates the substructure number density **33***d* obtained by the substructure number density calculating unit **43***d*.

The substructure number density of the work table **32** and the experimental values of the physical property of the polymer structure and physical property DB **31** are used for the machine learning by the regression analysis. Thus, the work table **32** may be associated with the polymer structure and physical property DB **31** by the polymer name. Here, the data configuration example of the polymer structure and physical property DB **31** and the work table **32** illustrated in FIG. **4** is only one example, but is not limited to this example.

An outline of an entire prediction process performed by the prediction processing unit **40** will be described using an example of the repeating unit. FIG. **5** is a drawing for describing the overview of the prediction process according to the present embodiment. As illustrated in FIG. **5**, in the model generator **41** of the prediction processing unit **40**, the model structure generator **43***a* obtains the repeating unit from the polymer structure and physical property DB **31** (as described in (a) of FIG. **5**). For example, a repeating unit "*CCO*" is obtained. In the following description, "*" indicating a start and an end of a certain structure in the SMILES notation will be omitted.

The model structure generator **43***a* uses the obtained substructure as a repeating unit, and replicates the repeating unit to generate the model structure **33***a* (as described in (b) of FIG. **5**). In this example, the model structure generator **43***a* repeats replicating the repeating unit "CCO" to the right a predetermined number of times, to generate the model structure **33***a* "CCOCOCOCCCOCCCOCCCOCCCOCCCOCC-COCCO", in which the repeating unit "CCO" are linked multiple times, in the memory unit **130**.

The substructure number calculating unit **43***b* generates a substructure number vector x_0 (i.e., the numbers of substructures **33***b*) of the structure model **33***a* (as described in (c) in FIG. **5**). The substructure number vector x_0 is represented in n dimensions by a structure descriptor, such as extended circular finger print (ECFP), atom pair, topological torsion, and neural finger print. In the first functional configuration example, a structure descriptor represented by the number of C atoms, the number of O atoms, the number of COCs, or the like is used for the description. However, the structure descriptor is not limited to this.

In the structure model "CCOCCCOCCCOCCCOCC-COCCCOCCO", an n-dimensional substructure number vector x_0, which consists of the number of C atoms "20", the number of O atoms "10", . . . , and the number of COCs "9", is generated. A structure assigned to each vector, such as "C", "O", "COC", indicates the type of the substructure of the polymer. Here, n is an arbitrary natural number, and is usually about 256 to 2048.

The atom calculating unit **43***c* calculates the number of atoms included in the model structure **33***a* (as described in (d) of FIG. **5**). In the structure model "CCOCOCOCOCC-COCCCOCCCOCCCO", the number of atoms is "30".

The substructure number density calculating unit **43***d* divides the substructure number vector x_0 (i.e., the numbers of substructures **33***b*) obtained by the substructure number calculating unit **43***b* in (c) of FIG. **5** by the number of atoms **33***c* obtained by the atom calculating unit **43***c* in (d) of FIG. **5** to generate the substructure number density vector x (i.e., the substructure number densities **33***d*) in (e) of FIG. **5**.

Specifically, in the structure model "CCOCCCOCCOC-COCCCOCCCOCCCOCCCOCCO", each value of the substructure number vector x_0 is divided by the number of atoms "30" to obtain the substructure number density vector x representing the C atom number density "⅔ (=20/30)", the O atom number density "⅓ (=10/30)", . . . , and the COC density "9/30." The number of dimensions of the substructure number density vector x is n, which is the same number as the number of dimensions of the substructure number vector x_0.

For each record of the polymer structure and physical property DB **31**, (a) of FIG. **5** to (e) of FIG. **5** are repeated. The number of substructures **33***b*, the number of atoms **33***c*, and the substructure number density **33***d* that are obtained may be stored in the work table **32** in association with the polymer name.

Before constructing the regression model **33***f*, a temporal regression model is constructed by the dimension reduction unit **43***e* and the dimension reduction is performed (in (f) of FIG. **5**). This can reduce n dimensions to n' dimensions (n'<n). By the dimension reduction, the important substructure **33***e* that contributes to the prediction accuracy is determined. Here, n' is a natural number smaller than n, and is usually about 16 to 128.

The regression model construction unit **43***f* constructs the regression model **33***f* used in predicting the physical property value based on the important substructure **33***e* (in (g) of FIG. **5**). The regression model **33***f*, whose dimensions are reduced to several tens dimensions, is obtained. The regression model **33***f* is expressed by the following equation (1) that obtains the predicted value y of the physical property of the polymer by the substructure number density vector x (i.e., the substructure number densities **33***d*).

$$y=f(x) \tag{1}$$

When dimensions are not reduced, the number of substructure number density vector x is n, which is the number of dimensions. But after dimensions are reduced, the dimensions of the substructure number density vector x are n'. Because the predicted values are calculated using only the dimensions contributing to the prediction accuracy by using the important substructure **33***e*, the physical property prediction process can be efficiently performed.

The above-described process from (a) to (e) in FIG. **5** corresponds to the model generating process of the model generator **41**. After the model generating process has been performed, in response to receiving the input of the polymer structure **6***a* desired to be processed from the user, the physical property prediction process is performed by the physical property prediction unit **45**.

In response to the input of the polymer structure A of which the physical property is to be obtained, the physical property prediction unit **45** generates the substructure number density vector x_A, and predicts the physical property value y_A (in (h) of FIG. **5**). The polymer structure input unit **47***a* inputs the polymer structure A, and the polymer physi-

cal property prediction unit **47***b* predicts the physical property value y_A by using the regression model **33***f*.

The physical property value y_A with respect to the polymer structure A desired to be processed can be obtained by substituting the substructure number density x_A in the above equation (1).

$$y_A = f(x_A) \quad (2)$$

If the regression model **33***f* is for only one physical property, the value of a predetermined physical property is predicted. If the regression model **33***f* is obtained with respect to each of the various physical properties, the user may specify the polymer structure A and all physical properties or a specific physical property. In this case, the polymer physical property prediction unit **47***b* predicts the physical property value **6***b* by using the regression model **33***f* corresponding to each physical property and the important substructure **33***e*.

The process of calculating the numbers of substructures **33***b* in (c) of FIG. **5** and the process of calculating the number of atoms in (d) of FIG. **5** may be performed in any order. After calculating the number of atoms **33***c* is performed, the numbers of substructures **33***b* may be calculated, or the number of atoms **33***c* and the numbers of substructures **33***b* may be calculated in parallel. In the present embodiment, a "predicted value" of the physical property and a predicted "physical property value" are identical.

In the first functional configuration example, the substructure number density calculating unit **43***d* of the prediction processing unit **40** can recognize whether polymer structures are identical. FIG. **6** is a drawing illustrating an example of the number density. In FIG. **6**, the number of carbons of a polyethylene oxide is illustrated as an example.

Referring to FIG. **6**, when a structure A and a structure B are compared, a notation of the structure B is longer and different from a notation of the structure A. However, polymers represented by the structure A and the structure B are identical.

In an existing technique in which the number of substructures is used as a structure descriptor, the number of carbons of the structure A is "2," the number of carbons of the structure B is "4". Since the numbers of carbons of the structure A and the structure B are different, the polymers are mistakenly recognized as being different. This example indicates that when polymers having different lengths of the repeating unit are compared, the number of substructures cannot appropriately capture structural features of the polymer.

In the first functional configuration example, the substructure number density calculating unit **43***d* obtains the same substructure number density (carbon number density) "⅔" with respect to the structure A and the structure B. Thus, the dimension reduction unit **43***e* and the regression model construction unit **43***f* can correctly recognize that the structure A and the structure B represent the same polymer. Based on these examples, the inventors have found that by using the substructure number density **33***d*, polymers having different lengths of the repeating unit can be treated equivalently.

The inventors have also found that replicating and linking the repeating unit can improve the existing technique having low prediction accuracy. A difference between when replication is performed and when no replication is performed is described in FIG. **7**. FIG. **7** is a drawing for describing the difference between when replication of the repeating unit is performed and when no replication is performed. In FIG. **7**, a polyethylene oxide is used as an example. A repeating unit when replication is performed and a repeating unit when no replication is performed are identical.

When the repeating unit is not replicated, there is no substructure "COC", which is expected to exist, at both ends of the repeating unit. Thus, COC number density is "0". With respect to this, in the first functional configuration example, the repeating unit is replicated and the right side of the repeating unit is linked to an oxygen atom on the left side of the next repeating unit. Two repeating units causes the substructure "COC" to appear and "⅙" is obtained with respect to the number density of "COC", which is expected to exist. In the first functional configuration example, the regression model **33***f* is obtained using the substructure number density **33***d* obtained in such a manner, thereby improving the prediction accuracy of the existing technique.

As described above, the prediction process is performed by the prediction processing unit **40** as illustrated in FIGS. **8**A and **8**B. FIGS. **8**A and **8**B are flowcharts for describing the prediction process. First, a model generating process P**51** is performed by the model generator **41**.

In step S**511**, in the model generator **41**, the model structure generator **43***a* selects and reads one record from the polymer structure and physical property DB **31**, and obtains the repeating unit and the experimental value. In step S**512**, the model structure generator **43***a* repeats an operation of replicating and linking the obtained repeating unit a predetermined number of times to generate the model structure **33***a*.

In step S**513**, the substructure number calculating unit **43***b* calculates the number of substructures **33***b* of the model structure **33***a* by using a predetermined structure descriptor (e.g., ECFP). In step S**514**, the atom calculating unit **43***c* calculates the number of atoms **33***c* of the model structure **33***a*. Step S**513** may be performed after step S**514** has been performed. Alternatively, steps S**513** and S**514** may be performed in parallel.

In step S**515**, the substructure number density calculating unit **43***d* calculates the substructure number density **33***d* by dividing the number of substructures **33***b* by the number of atoms **33***c*. As described above, the numbers of substructures **33***b* and the substructure number densities **33***d* are represented by an n-dimensional vector.

In step S**516**, the model generator **41** determines whether all records of the polymer structure and the physical property DB **31** have been completed. When all records have not been completed (NO in step S**516**), the model generator **41** returns to step S**511** and performs a process similar to the process described above.

When all records have been completed (YES in step S**516**), the dimension reduction unit **43***e* performs a first regression analysis for obtaining a temporal regression model f' by using the substructure number density **33***d* corresponding to the important substructure **33***e* as a variable x' and an experimental value as y'. In step S**517**, the dimension reduction unit **43***e* reduces the dimensions of the temporal regression model f' by a predetermined dimension reduction method and obtains the important substructure **33***e*.

Next, the regression model construction unit **43***f* performs a second regression analysis. In step S**518**, the regression model construction unit **43***f* obtains the regression model f by using the substructure number density **33***d* corresponding to the important substructure **33***e* as the variable x and the experimental value as y. The experimental value y is a value determined by the polymer name and the physical property in the polymer structure and physical property DB **31**. The

regression model f, that is, the regression model **33***f*, is stored in the storage unit **130**.

When the regression model f is obtained for all or some of the physical properties registered in the polymer structure and physical property DB **31**, an experimental value may be obtained for each physical property, and steps **S517** to **S518** may be repeated.

Subsequently, the physical property prediction process **P55** is performed by the physical property prediction unit **45**. When the polymer structure input unit **47***a* receives the polymer structure **6***a* in the prediction processing unit **40** (step **S521**), the model structure, the number of substructures, the number of atoms, and the substructure number density are obtained using the polymer structure **6***a* as the repeating unit (step **S522**). The model structure, the number of substructures, the number of atoms, and the substructure number density of the polymer structure **6***a* are obtained by performing a process similar to the process from step **S512** to step **S515** in the model generating process **P51**.

In step **S523**, the dimensions of the substructure number density of the polymer structure **6***a* are reduced based on the important substructure **33***e* obtained in step **S517**, and in step **S524**, the polymer physical property prediction unit **47***b* predicts the physical property value **6***b* by using the regression model f (i.e., the regression model **33***f* stored in the storage unit **130**) obtained in step **S518**. The obtained physical property value **6***b* is sent to the terminal **3** or displayed on the display device **15**.

When the regression model **33***f* corresponding to each of multiple physical properties is generated, it is desirable to manage the generated regression models **33***f* in a table as illustrated in FIG. **9**. FIG. **9** is a drawing illustrating an example of a data configuration of a regression model management table.

In FIG. **9**, a regression model management table **33***k* is a table for managing the regression model **33***f* for each physical property, and includes items such as the physical property, the regression model, and the important substructure. The physical property represents a name for specifying the physical property, such as the refractive index, and the glass transition point. The regression model represents the regression model **33***f* corresponding to the physical property, constructed in the regression model construction unit **43***f*.

The important substructure specifies a valid vector of the vector of the substructure number density **33***d*, obtained after dimension reduction (i.e., a valid vector after dimension reduction) for using the regression model **33***f*. When the dimension reduction is omitted, the item of the important substructure is omitted.

In the physical property prediction unit **45**, the polymer physical property prediction unit **47***b* obtains the regression model **33***f* and the important substructure **33***e* that are corresponding to the physical property to be predicted from the regression model management table **33***k*.

The inventors have verified the prediction accuracy of various functional configurations described below. According to verification results, the first functional configuration example described above can improve the prediction accuracy the most. Further, the inventors have verified that the prediction accuracy can be sufficiently improved even in a configuration in which either or both of the model structure generator **43***a* and the dimension reduction unit **43***e* are removed.

First, other examples of the functional configuration in which the prediction accuracy is verified will be described below. In the following description, a second functional configuration example is a functional configuration in which both of the model structure generator **43***a* and the dimension reduction unit **43***e* are removed, a third functional configuration example is a functional configuration in which the model structure generator **43***a* is removed, and a fourth functional configuration example is a functional configuration in which the dimension reduction unit **43***e* is removed. In the second functional configuration example to the fourth functional configuration example, processing units similar to the processing units of the first functional configuration example illustrated in FIG. **3** are referenced by the same reference numerals, and the detailed description is omitted.

[Second Functional Configuration Example]

FIG. **10** is a drawing illustrating the second functional configuration example of the physical property prediction device. In FIG. **10**, in the second functional configuration example, the model structure generator **43***a* and the dimension reduction unit **43***e* are removed from the first functional configuration example illustrated in FIG. **3**. Thus, the model structure **33***a* and the important substructure **33***e* are not generated.

In the model generator **41**, the substructure number calculating unit **43***b* and the atom calculating unit **43***c* obtains the number of substructures **33***b* and the number of atoms **33***c* by obtaining the repeating unit from the polymer structure and physical property DB **31**.

Also in the second functional configuration example, the substructure number density calculating unit **43***d* calculates the substructure number density **33***d*. Then, the regression model construction unit **43***f* obtains the regression model **33***f* by using the calculated substructure number density **33***d* and the experimental value of the physical property.

In the physical property prediction unit **45**, when the polymer structure input unit **47***a* inputs the polymer structure **6***a*, the polymer physical property prediction unit **47***b* calculates the number of substructures and the number of atoms by using the polymer structure **6***a* without creating the model structure **33***a*. The polymer physical property prediction unit **47***b* obtains the substructure number density by dividing the obtained number of substructures by the number of atoms, and uses the substructure number density **33***d* to obtain the property value **6***b* by the regression model **33***f*.

In the second functional configuration example, the regression analysis is performed only once. The number of dimensions of the regression model **33***f* is the same number as the number of dimensions of the substructure number densities **33***d*, and the important substructure **33***e* is not determined. However, if the values of the substructure number density **33***d* are identical, even polymers with different lengths of the repeating unit can be treated equivalently, and the prediction accuracy can be sufficiently improved.

[Third Function Configuration Example]

FIG. **11** is a drawing illustrating the third functional configuration example of the physical property prediction device. In FIG. **11**, in the third functional configuration example, the model structure generator **43***a* is removed from the first functional configuration example illustrated in FIG. **3**. Thus, the model structure **33***a* is not generated.

In a manner similar to the second functional configuration example, in the model generator **41**, the substructure number calculating unit **43***b* and the atom calculating unit **43***c* obtain the repeating unit from the polymer structure and physical property DB **31** to obtain the number of substructures **33***b* and the number of atoms **33***c*. The substructure number density calculating unit **43***d* obtains the substructure number density **33***d*. Then, the dimension reduction unit **43***e* obtains the important substructure **33***e* and reduces the dimensions

of the substructure number densities **33***d*, and the regression model construction unit **43***f* outputs the regression model **33***f*.

In the physical property prediction unit **45**, when the polymer structure input unit **47***a* inputs the polymer structure **6***a*, the polymer physical property prediction unit **47***b* calculates the number of substructures and the number of atoms by using the polymer structure **6***a* without generating the model structure **33***a*. Then, the polymer physical, property prediction unit **47***b* obtains the substructure number density by dividing the obtained number of substructures by the number of atoms, reduces the number of dimensions of the substructure number densities based on the substructure number density **33***d*, and outputs the property value **6***b* by using the regression model **33***f*.

In the third functional configuration example, when the values of the substructure number density **33***d* are identical, even polymers with different lengths of the repeating unit can be equivalently treated. Although the model structure **33***a* is not generated, the important substructure **33***e* is determined by the dimension reduction unit **43***e*, thereby improving the prediction accuracy of the regression model **33***f* generated by the regression model construction unit **43***f*.

[Fourth Functional Configuration Example]

FIG. **12** is a drawing illustrating the fourth functional configuration example of the physical property prediction device. In FIG. **12**, in the fourth functional configuration example, the dimension reduction unit **43***e* is removed from the first functional configuration example illustrated in FIG. **3**. Thus, the important substructure **33***e* is not generated.

In the model generator **41**, the model, structure generator **43***a* generates the model structure **33***a*, and the substructure number calculating unit **43***b* and the atom calculating unit **43***c* obtain the number of substructures **33***b* and the number of atoms **33***c* based on the model structure **33***a*.

In the physical property prediction unit **45**, when the polymer structure input unit **47***a* inputs the polymer structure **6***a*, the polymer physical property prediction unit **47***b* generates the model structure **33***a* and calculates the number of substructures and the number of atoms by using the model structure **33***a*. Then, the polymer physical property prediction unit **47***b* obtains the substructure number density by dividing the obtained number of substructures by the number of atoms, and uses the substructure number density **33***d* to obtain the physical property value **6***b* by the regression model **33***f*. The dimensions of the substructure number densities obtained from polymer structure **6***a* are not reduced.

In the fourth functional configuration, when the values of the substructure number density **33***d* are identical, even polymers with different lengths of the repeating unit can be equivalently treated. Although the important substructure **33***e* is not generated, the model structure **33***a* and the substructure number density **33***d* can improve the prediction accuracy of the regression model **33***f* generated by the regression model construction unit **43***f*.

Next, a functional configuration example based on the existing technique will be described as a comparative example.

[Existing Functional Configuration Example]

FIG. **13** is a drawing illustrating a functional configuration example of a physical property prediction device based on the existing technique. In FIG. **13**, a physical property prediction device **9** has a functional configuration based on the existing technique. In the comparative example illustrated in FIG. **13**, a prediction processing unit **90** of the physical property prediction device **9** includes a model generator **91** and a physical property prediction unit **95**.

The model generator **91** is a processing unit that generates a regression model **83***f* and includes a substructure number calculating unit **93***b* and a regression model construction unit **93***f*. The substructure number calculating unit **93***b* obtains the repeating unit from the polymer structure and physical property DB **31** and stores the number of substructures **83***b* in the storage unit **130**. The regression model construction unit **93***f* constructs the regression model **83***f* by performing a regression analysis using the number of substructures **83***b* and an experimental value obtained from the polymer structure and physical property DB **31**, and stores the regression model **83***f* in the storage unit **130**.

The physical property prediction unit **95** is a processing unit that predicts the physical property value by applying the polymer structure **6***a* desired to be processed by a user to the regression model **83***f*, and includes a polymer structure input unit **97***a* and a polymer physical property prediction unit **97***b*. The polymer structure input unit **97***a* stores the polymer structure **6***a* in the storage unit **130** when receiving the polymer structure **6***a*. The polymer physical property prediction unit **97***b* obtains the number of substructures of the polymer structure **6***a* and applies the obtained number of substructures to the regression model **83***f* to obtain the predicted physical property value **6***b*. The physical property value **6***b* is presented to the user.

[Verification of the Prediction Accuracy]

Next, results of verifying the prediction accuracy of the functional configuration example as a comparative example and the first to fourth functional configuration examples will be described. In each functional configuration example, the inventors have verified the prediction accuracy when the refractive index is predicted.

EXAMPLE 1

In Example 1, the second functional configuration example (FIG. **10**) in which the model structure generator **43***a* and the dimension reduction unit **43***e* are removed was used for the verification. In Example 1, the model structure **33***a* is not generated. In addition, the regression analysis is performed only once by the regression model construction unit **43***f*.

In the above-described second functional configuration example, structures of 51 general infinite linear polymers, such as polyethylene and polypropylene, (i.e., repeated units) and experimental values of the refractive index were used as training data. ECFP4 was used as the structure descriptor. Here, ECFP4 indicates ECFP having a radius of 4 and 1024 dimensions. Hereinafter, ECPF4 will be referred to as an ECFP descriptor.

The ECFP descriptor is represented by a vector, each dimension represents a substructure classification, and a value of each dimension corresponds the number of substructures. Software such as RDKit was used to generate the ECFP descriptor. For convenience, the ECFP Descriptor is defined as a substructure number descriptor. Then, a substructure number density descriptor is obtained by dividing each element of the substructure number descriptor by the number of constituent atoms.

As a regression method, the Gaussian process regression was used. The regression model **33***f* was constructed using the substructure number density descriptor as an explanatory variable and the experimental value of the refractive index as an objective variable.

A data set was divided into a training data set and a verification data set by leave-one-out cross-verification, a regression model was constructed after hyperparameter tuning was performed using the training data set, and the refractive index of a polymer that was not learned was predicted using the verification data set. The coefficient of determination R^2 and the root mean squared error (RMSE) were used as evaluation measures of the prediction accuracy. There are various definition equations of R^2, but the following equation was used this time.

$$R^2 \equiv 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \quad \text{[Eq. 1]}$$

Here, y_i is an experimental value of an ith polymer, y_bar is a measured average value, and f_i is the predicted value of the ith polymer.

In the description above, the number of constituent atoms corresponds to the number of atoms of the repeating unit, the substructure number descriptor corresponds to the numbers of substructures **33***b*, and the substructure number density descriptor corresponds to the substructure number densities **33** *d*.

EXAMPLE 2

In Example 2, the third functional configuration example (FIG. **11**) in which the model structure generator **43***a* was removed was used for the verification. Since the model structure **33***a* was not generated, the substructure number density **33***d* was obtained from the repeating unit of the polymer structure and physical property DB **31**. The regression analysis was performed twice by the dimension reduction unit **43***e* and the regression model construction unit **43***f*.

Before constructing the regression model **33***f*, the dimension reduction unit **43***e* constructs a Lasso regression model (corresponding to the temporal regression model) from the substructure number density **33***d* and the experimental value of the physical property of the polymer obtained from the polymer structure and physical property DB **31**, and performs the dimension reduction by excluding variables with coefficients of 0. In the regression model construction unit **43***f*, the regression model **33***f* was constructed by the Gaussian process regression in a manner similar to Example 1. Then, the prediction accuracy of the refractive index was verified by a method similar to the method of Example 1.

EXAMPLE 3

In Example 3, the fourth functional configuration example (FIG. **12**) in which the dimension reduction unit **43***e* is removed was used for the verification. The important substructure **33***e* was not obtained. The regression model **33***f* with the number of dimensions of the substructures number densities **33***d* was constructed.

First, after 10 repeating units were generated as the model structure **33***a*, the substructure number descriptor and the number of atoms were calculated, and the substructure number density descriptor was calculated. After the substructure number density descriptor was calculated, a predicted value of the refractive index was obtained by a method similar to the method of Example 1, and the prediction accuracy was verified.

EXAMPLE 4

In Example 4, the first functional configuration example (FIG. **3**) including the important substructure **33***e* and the dimension reduction unit **43***e* was used for the verification.

First, after 10 repeating units were generated as the model structure **33***a*, the substructure number descriptor and the number of atoms were calculated, and the substructure number density descriptor was calculated. In addition, the dimension reduction unit **43***e* constructed a Lasso regression model (corresponding to the temporal regression model) and constructed the regression model **33***f* by the Gaussian process regression after reducing the dimensions by excluding variables with coefficients of 0. Then, the prediction accuracy of the refractive index was verified by a method similar to the method of Example 1.

COMPARATIVE EXAMPLE 1

In Comparative example 1, the existing functional configuration example (FIG. **13**) in which only the substructure number calculating unit **93***b* and the regression model construction unit **93***f* obtains the regression model **83***f* was used for the verification.

The substructure number calculating unit **93***b* obtained the regression model **83***f* from the number of substructures **83***b* (i.e., the substructure number descriptor) obtained using the repeating unit and the experimental value of the refractive index. The regression model **83***f* was constructed using the Gaussian process regression in a manner similar to the manner of Example 1. The predictive accuracy was verified in a manner similar to the manner of Example 1.

FIG. **14** is a drawing illustrating an example of evaluation results. FIG. **14** illustrates the evaluation results of the prediction accuracy of the refractive index evaluated for each of Comparative example 1 and Examples 1 to 4 as described above. Results of R^2 and RMSE are shown as the evaluation results.

In the evaluation result of R^2, compared with "−0.063" of Comparative Example 1, results of "0.890" in Example 1, "0.900" in Example 2, "0.932" in Example 3, and "0.950" in Example 4 are shown. The closer to "1" R^2 is, the higher the prediction accuracy becomes. Thus, from the evaluation results, it is found that the prediction accuracy can be significantly improved in any of the functional configuration examples according to the present embodiment.

In RMSE, compared with "0.084" in Comparative example 1, results of "0.027" in Example 1, "0.026" in Example 2, "0.021" in Example 3, and "0.018" in Example 4 are shown. The closer to "0" RMSE is, the higher the prediction accuracy becomes. Thus, it is found that the prediction accuracy can be significantly improved in all the functional configuration examples according to the present embodiment in the evaluation results.

From such evaluation results, it can be said that the substructure number density **33***d*, which the inventors have focused on, greatly contributes to the improvement of the prediction accuracy in the present embodiment.

FIG. 15 illustrates an example of results of an evaluation performed by changing the physical property and the number of data in the data set in a manner similar to the manner of Example 4. FIG. 15 illustrates evaluation results of physical properties of Examples 5 to 9 performed by a respective number of data in a manner similar to the manner of Example 4. Results of R^2 are shown as the evaluation results.

In the results of R^2, results of a refractive index, a glass transition point [k], a band gap [eV], a relative permittivity, a coefficient of linear expansion [10^{-5}/K], and a density [g/cm^3] are greater than or equal to 0.6. Thus, it is found that the prediction accuracy can be significantly improved.

Since the physical property value can be predicted in advance using the physical property prediction device **100**, the polymer structure **6***a* to be manufactured can be accurately selected. Thus, it is possible to reduce waste of costs, time, and so on related to polymer manufacturing.

As described above, according to the present embodiment, the physical properties of the polymer can be accurately predicted by generating the structure descriptor based on the polymer structure.

Next, a manufacturing method of a polymer using the desired polymer structure **6***a* and the predicted physical property will be described. FIG. is a drawing illustrating an example of a manufacturing method of the polymer.

From FIG. **16**, the manufacturing method of the polymer in the present embodiment includes a step of obtaining a polymer structure having a property value predicted by the polymer physical property prediction method according to the present embodiment (S**601**) and a step of preparing raw materials for forming the polymer structure (S**602**), and a step of causing the raw materials to react to obtain a polymer represented by the polymer structure (S**603**).

In step S**601**, a polymer structure predicted to have a predetermined physical property value is appropriately selected from the input polymer structures. For example, a polymer structure whose refractive index is within a predetermined range may be selected from the input polymer structures.

In step S**602**, raw materials of monomers, oligomers, and the like are prepared as raw materials for forming the polymer structure selected in step S**601**. For example, if the selected polymer structure is polyethylene, ethylene is prepared.

In step S**603**, the raw materials prepared in S**602** are caused to react to obtain a polymer represented by the polymer structure selected in S**601**. For example, ethylene is polymerized to form polyethylene.

The present invention is not limited to the specifically disclosed embodiments and various variations and modifications can be made without departing from the scope of the claims.

The present application is based on and claims priority to Japanese Patent Application No. 2018-043171, filed on Mar. 9, 2018, the entire contents of which are hereby incorporated herein by reference.

What is claimed is:

1. A polymer physical property prediction device comprising:

a processor; and

a memory storing program instructions that cause the processor to

read a structural unit from a storage unit and use the structural unit to calculate numbers each indicating how many substructures are in a polymer, the polymer being formed of repetition of the structural unit;

calculate a number indicating how many atoms are in the structural unit;

calculate number densities of the substructures from the numbers of substructures and the number of atoms in the structural unit;

construct a regression model that predicts a physical property value by using an experimental value of a physical property of the polymer and the number densities of the substructures, the experimental value being obtained from the storage unit;

input a polymer structure of which the physical property value is to be predicted; and

predict the physical property value corresponding to the input polymer structure by using the regression model.

2. The polymer physical property prediction device as claimed in claim **1**, wherein the program instructions cause the processor to further classify the substructures of the polymer using an algorithm, and the calculated numbers of substructures each indicate a number of each corresponding classified substructure present in the polymer.

3. The polymer physical property prediction device as claimed in claim **1**, wherein the program instructions cause the processor to further remove part of the calculated number densities of the substructures.

4. The polymer physical property prediction device as claimed in claim **1**, wherein the program instructions cause the processor to further generate a model structure by repeatedly using the structural unit obtained from the storage unit.

5. The polymer physical property prediction device as claimed in claim **1**,

wherein the program instructions cause the processor to further

generate a model structure by repeatedly using the structural unit obtained from the storage unit, and

remove part of a structure descriptor representing the calculated number densities of the substructures,

wherein the numbers of substructures and the number of atoms are calculated from the generated model structure.

6. The polymer physical property prediction device as claimed in claim **1**, wherein the program instructions cause the processor to calculate the numbers of substructures and the number of atoms in the polymer structure by using the input polymer structure, calculate the number densities of the substructures from the calculated numbers of substructures and the calculated number of atoms, and apply the calculated number densities to the constructed regression model to predict the physical property value of the polymer structure.

7. The polymer physical property prediction device as claimed in claim **1**, wherein the physical property value is a value of a refractive index or a value of a glass transition point.

8. The polymer physical property prediction device as claimed in claim **1**, wherein the physical property value is a value of a band gap, a value of a relative permittivity, a value of a coefficient of linear expansion, or a value of a density.

9. The polymer physical property prediction device as claimed in claim **1**, wherein the numbers each indicating how many substructures are in the polymer are calculated by counting a substructure in the polymer to obtain a substructure number indicating how many units of the substructure are in the polymer, a number of atoms in the polymer is calculated, and the number densities are calculated by dividing the substructure number by the number of atoms in the polymer.

10. The polymer physical property prediction device as claimed in claim **3**, wherein the program instructions cause the processor to calculate the numbers of substructures and the number of atoms of the polymer structure by using the input polymer structure, calculate the number densities of the substructures from the calculated numbers of substructures and the calculated number of atoms, remove part of the calculated number densities of the substructures, and apply

the number densities whose part is removed to the constructed regression model to predict the physical property value of the polymer structure.

11. The polymer physical property prediction device as claimed in claim **4**, wherein the program instructions cause the processor to generate an input polymer model structure by repeatedly using the input polymer structure, calculate the numbers of substructures and the number of atoms in the input polymer model structure, calculate the number densities of the substructures from the calculated numbers of substructures and the calculated number of atoms, and apply the calculated number densities to the constructed regression model to predict the physical property value of the polymer structure.

12. The polymer physical property prediction device as claimed in claim **5**, wherein the program instructions cause the processor to generate the model structure by repeatedly using the input polymer structure, calculate the numbers of substructures and the number of atoms in the model structure, calculate the number densities of the substructures from the calculated numbers of substructures and the calculated number of atoms, remove the part of the structure descriptor representing the calculated number densities of the substructures, and apply the number densities whose part is removed to the constructed regression model to predict the physical property value of the polymer structure.

13. A non-transitory computer-readable recording medium having stored therein a polymer physical property prediction program for causing a computer to execute a process comprising:

reading a structural unit from a storage unit and using the structural unit to calculate numbers each indicating how many substructures are in a polymer, the polymer being formed of repetition of the structural unit;

calculating a number indicating how many atoms are in the structural unit;

calculating number densities of the substructures from the numbers of substructures and the number of atoms in the structural unit;

constructing a regression model that predicts a physical property value by using an experimental value of a physical property of the polymer and the number densities of the substructures, the experimental value being obtained from the storage unit;

inputting a polymer structure of which the physical property value is to be predicted; and

predicting the physical property value corresponding to the input polymer structure by using the regression model.

14. The recording medium as claimed in claim **13**, wherein the predicting the physical property value further includes calculating the numbers of substructures and the number of atoms in the polymer structure by using the input polymer structure, calculating the number densities of the substructures from the calculated numbers of substructures and the calculated number of atoms, and applying the calculated number densities to the regression model to predict the physical property value of the polymer structure.

15. The non-transitory computer-readable recording medium as claimed in claim **13**, wherein the numbers each indicating how many substructures are in the polymer are calculated by counting a substructure in the polymer to obtain a substructure number indicating how many units of the substructure are in the polymer, a number of atoms in the polymer is calculated, and the number densities are calculated by dividing the substructure number by the number of atoms in the polymer.

16. A polymer physical property prediction method executed by a computer, the method comprising:

reading a structural unit from a storage unit and using the structural unit to calculate numbers each indicating how many substructures are in a polymer, the polymer being formed of repetition of the structural unit;

calculating a number indicating how many atoms are in the structural unit;

calculating number densities of the substructures from the numbers of substructures and the number of atoms in the structural unit;

constructing a regression model that predicts a physical property value by using an experimental value of a physical property of the polymer and the number densities of the substructures, the experimental value being obtained from the storage unit;

inputting a polymer structure of which the physical property value is to be predicted; and

predicting the physical property value corresponding to the input polymer structure by using the regression model.

17. The polymer physical property prediction method as claimed in claim **16**, wherein the predicting the physical property value further includes calculating the numbers of substructures and the number of atoms in the polymer structure by using the input polymer structure, calculating the number densities of the substructures from the calculated numbers of substructures and the calculated number of atoms, and applying the calculated number densities to the regression model to predict the physical property value of the polymer structure.

18. The polymer physical property prediction method as claimed in claim **16**, wherein the numbers each indicating how many substructures are in the polymer are calculated by counting a substructure in the polymer to obtain a substructure number indicating how many units of the substructure are in the polymer, a number of atoms in the polymer is calculated, and the number densities are calculated by dividing the substructure number by the number of atoms in the polymer.

19. A polymer manufacturing method comprising:

reading a structural unit from a storage unit and using the structural unit to calculate numbers each indicating how many substructures are in a polymer, the polymer being formed of repetition of the structural unit;

calculating a number indicating how many atoms are in the structural unit;

calculating number densities of the substructures from the numbers of substructures and the number of atoms in the structural unit;

constructing a regression model that predicts a physical property value by using an experimental value of a physical property of the polymer and the number densities of the substructures, the experimental value being obtained from the storage unit;

inputting a polymer structure of which the physical property value is to be predicted;

predicting the physical property value corresponding to the input polymer structure by using the regression model,

obtaining a polymer structure having the predicted physical property value;

preparing materials for forming the polymer structure; and

obtaining a polymer represented by the polymer structure by reacting the materials.

20. The polymer manufacturing method as claimed in claim **19**, wherein the numbers each indicating how many substructures are in the polymer are calculated by counting a substructure in the polymer to obtain a substructure number indicating how many units of the substructure are in the polymer, a number of atoms in the polymer is calculated, and the number densities are calculated by dividing the substructure number by the number of atoms in the polymer.

* * * * *