Efficient screening of chemicals is essential for exploring new materials. However, the search space is astronomically large, making calculations with conventional computers infeasible. For example, an $N$-component system of organic molecules generates $>10^{60N}$ candidates. Here, a quantum-inspired annealing machine is used to tackle the challenge of the large search space. The prototype system extracts candidate chemicals and their composites with desirable parameters, such as melting temperature and ionic conductivity. The system can be at least $10^4$–$10^6$ times faster than conventional approaches. Such dramatic acceleration is critical for exploring the enormous search space in virtual screening of materials.

1. Introduction

Finding new functional materials with improved performance is critical, especially in developing next-generation energy-related devices. Machine learning, which can handle the big data from materials science and identify statistical trends, has become essential for material exploration. Trained models can predict the performance ($y$) of diverse materials from their structures ($x$).

The predicted values ($y_{\text{pred}}$) from a trained model ($f_{\text{ML}}$) have become increasingly reliable. Various parameters, including basic molecular properties (e.g., mechanical strength and melting point), conductivity, photoconversion efficiency, and synthetic yields, have been predicted successfully. If an appropriate model is prepared, the prediction will become as accurate as those from human researchers and molecular simulations. Another major advantage of machine learning is its high prediction speed (e.g., $10^{-3}$ s per condition), allowing researchers to screen chemicals much faster than with real experiments and first-principles calculations.

A key challenge in materials informatics is the astronomically large search space for candidate materials. Even for small organic molecules, there are more than $10^{60}$ candidates. Current virtual compound screening using a machine learning function, $y_{\text{pred}} = f_{\text{ML}}(x)$, may not finish before the end of the universe (Figure 1). Furthermore, the search space increases exponentially when a composite system of multiple chemicals is considered ($10^{60N}$ candidate composites for an $N$-component system), and most conventional materials and devices consist of multiple components.

Various approaches have been proposed to solve the massive search space problem, such as defining an inverse function $x = f_{\text{ML}}^{-1}(y)$, which predicts structures or experimental conditions from a target parameter $y$. However, the function is intrinsically hard to define because of the uniqueness of the solution problem. Furthermore, the preparation of chemical structures itself is still a challenge even with cutting-edge deep learning techniques due to the current limitations of model complexity and computing power. Other molecule generation techniques, such as Bayesian approaches and deep reinforcement learning, also tend to have similar limitations, such as excessive search space compared to the limited computing power. Most studies have focused on single-component systems and the methodology for multiple-component systems has not been developed satisfactorily.

A practical approach to exploring materials is to narrow the search space by filtering candidates manually. Based on the knowledge of human researchers, only feasibly useful structures are selected. Typically, $10^5$–$10^6$ candidate chemicals would be selected to allow calculations to be finished quickly. However, filtering based on existing knowledge may overlook hidden potential candidates. The approach may fail when treating $N$-component systems. If there are $10^M$ chemical candidates, their available combinations will be $10^{M+N}$, excluding their weight compositions (i.e., $N – 1$ additional independent variables).

Here, we introduce a quantum physics inspired annealing machine to tackle the challenge of the huge search space. The main limitation of conventional approaches is the insufficient computing power compared with the search space. Quantum physics based or inspired computing solves the exploration problems efficiently due to massive parallelization that cannot be achieved by conventional computers.
A newly developed system was at least $10^4$–$10^8$ times faster than conventional approaches, helping to solve the combinatorial explosion problem in materials science.

2. Molecular Exploration System Framework

Quantum computers can solve some types of problems quickly, including combinatorial optimization and molecular quantum processes.\[18,21\] The superposition principle is useful for exploring new materials from a vast search space efficiently (Figure 2a). Gate-based quantum computers can solve various problems, although they currently have many technical problems, such as long-time coherence.\[18,21\] In contrast, quantum annealing machines have already been commercialized.\[23\] They are used to solve the Ising model (Equation (1)).\[17,23,24\] Previously, annealers have been used in limited cases to explore materials (e.g., orders of nanomaterials).\[17\] In this work, our prototype system was used to explore general organic materials with diverse properties.

The architecture of our material screening system is shown in Figure 2a. First, structural information was converted into binary vectors using a mapping function, $f_{\text{map}}$ (Figure 2b). Binarization was needed because quantum annealers accept only binaries.\[24\] A machine learning function, $f_{\text{ML}}$, was prepared in the form of the Ising model. An experimental database of target materials was used to train the model.

The machine learning function was a linear model considering quadratic terms ($x_i x_j, i \neq j$). Annealers can solve the problem of $x_{\text{ideal}} = \arg \max f_{\text{ML}}(x)$ (or argmin) promptly (around $10^{-4}$ to $10^{-8}$ s, Figure 2c).\[22,24\] The obtained $x_{\text{ideal}}$ could give $y_{\text{pred}}$, which is close to the global maximum (or minimum) of the system (Equation (1)).\[24\]

$$y_{\text{pred}} = f_{\text{ML}}(x) = \sum_{i \neq j} J_{ij} x_i x_j + \sum_i h_i x_i$$

Conventional computers may not be able to solve the problem when the dimension of $x$ becomes higher than around 100 (i.e., $2^{100} \approx 10^{30}$-dimensional search space).\[22\] Finally, chemical structures were proposed as real-world information based on an inverse mapping function, $f_{\text{map}}^{-1}$.

![Image](image_url)`
For the annealer, we used a digital annealing unit, which was developed based on quantum annealing.[22] The machine is a digit simulator, but the hardware was specially designed to solve versatile quadratic unconstrained binary optimization problems and can treat up to 8192 fully connected binary variables with massive parallelization.[23] In contrast, the commercialized quantum annealer D-Wave still has a strict limitation on the input variables and binary interactions; there are typically fewer than 100 available variables for achieving fully connected bits.[9,23,24]

3. Exploration with a Single-Component System

First, we extracted high-melting-temperature molecules with the annealer. An open experimental database of organic molecules[25] was used for machine learning (about 3000 chemicals, Table S1, Supporting Information). The structural information was converted into 2048-dimensional molecular fingerprints \( x_{FP} \).[26] The vector was compressed to \( n \)-dimensional vectors to obtain a binary input, \( x \) (0 < \( n < 2048 \), Figure 2b, 3a and Figure S1, S2, Supporting Information). Compression was needed because \( nC_2 \) interactions must be added as quadratic terms \( x_i x_j \) (i.e., \( 2048C_2 = 2096 \)128 terms with \( n = 2048 \) are too many for regression). After adding the terms, the statistical relationships between the input and z-scores of the melting temperatures \( y \) were inputted to train a linear regression model (Figure 2a and Equation (1)). The regression determined the coefficients \( j_{ij} \) and \( h_i \).

Regularized models[27] were used for regressions to avoid multicollinearity problems. Ridge (i.e., L2 norm) models gave better prediction performance than did Lasso models (i.e., L1 norm, Figure S3, Supporting Information). The sparsity obtained by the Lasso model may have discarded too much of the contribution of quadratic terms \( x_i x_j \) in Equation (1), while the interaction terms were essential to improve the model complexity. Because regression algorithms were developed for continuous variables,[27] a specialized algorithm and theory for binary vectors, including the selection of regression norms, will be needed in future work.

For regression, the database was split randomly into the training and testing datasets. The training dataset was used for regression, and the testing dataset was used only for prediction. The \( R^2 \) score for the testing dataset was as high as 0.5 with a dimension of \( x = 576 \) (Figure 3b; results with other conditions are shown in Figure S3, Supporting Information). Root mean square error was calculated to be 0.57 (corresponding to about 50 °C). The result was comparable to a former report, which aimed state-of-the-art accuracy for a melting point prediction task using the same dataset.[28] The lower \( R^2 \) score than the training dataset (0.75) was certainly caused by the difficulty of the task itself, not by the quadratic model (i.e., a huge variety of chemical structures compared to the inputted records).[28] The regression scores were higher than those obtained with random forest regression models, which are regularly used to treat nonlinear systems (Figure S4, Supporting Information).[1,5] Therefore, the proposed quadratic model itself was sufficient to predict chemical properties from binary vectors.

![Figure 3](https://www.advancedsciencenews.com/doi/figure-pdf/10.1002/adsi.20000209)

**Figure 3.** a) Scheme of preparing \( x \) and \( y \) for the melting point database. b) Regression result of melting point (as a z-score, \( y \)). The database was split into the training (90%) and testing (10%) datasets, for which the \( R^2 \) scores were 0.75 and 0.51, respectively. Ridge regression was conducted with a regularization strength \( \alpha = 0.1 \). The dimension of \( x \) was 576, which corresponds to a search space of \( 2^{576} \approx 10^{170} \) chemical structures in maximum.

© 2021 The Authors. Advanced Intelligent Systems published by Wiley-VCH GmbH
It was essential to obtain $x_{\text{ideal}}$ yielding the highest predicted performance, $y_{\text{pred}}$, from Equation (1). A digital annealer was used to solve the problem (Figure 4a). As a control, simulated annealing on a conventional computer was examined. The annealing time with the annealer was only around 2.5 s, regardless of dimension $n$. In contrast, the required time with the conventional machine increased dramatically and exceeded 300 s with $n = 1250$. $y_{\text{pred}}$ obtained in the control experiment was substantially smaller than in the annealer with $n > 100$, demonstrating that the control could not reach the global (Figure 4a). An impractically long time would be required to match the results of the annealer with the conventional simulator.

The digital annealer has also been shown to be faster than other computational methods such as parallel tempering Monte Carlo.\(^{[22]}\) The speedup was achieved by a unique hardware design to solve the quadratic problems (comprehensive comparisons against the conventional methods are shown in our previous study).\(^{[22]}\) Although the annealing time was longer than for real quantum annealers (around milliseconds for a total annealing process, excluding server connection time),\(^{[23]}\) several seconds seems practical for most applications. It should also be noted that the quantum machines still tend to suffer from solving high dimensions (>100), whereas a digital annealer can treat fully connected 1024 bit problems.\(^{[24]}\)

After decompressing $x_{\text{ideal}}$ to the fingerprint form, $x_{\text{ideal,FP}}$, Tanimoto similarity scores\(^{[29]}\) were calculated to compare the fingerprints of the ideal compound and those recorded in the test dataset (Figure 4b). Experimental $y$ became statistically higher when the high similarity scores increased (Figure 4c, red plots). The increase indicated that $x_{\text{ideal}}$ found by the annealer

![Figure 4. a) Annealing time and best $y$ found as a function of the dimension of $x$. Linear regression models were prepared with different dimensions. $x_{\text{ideal}}$ giving the highest $y_{\text{pred}}$ was explored by the digital annealer and a conventional computer as a control. Annealing time by the digital annealer was almost constant because it virtually solved the 1024-dimensional problems regardless of the conditions (see the Supporting Information for details). b) Scheme to extract candidate compounds using similarity scores with $x_{\text{ideal}}$. c) Relationship between $y$ and similarity to a target fingerprint. In the upper panel, 20% of the test data were compared with $x_{\text{ideal}}$. In the bottom panel, the test dataset was compared with the existing fingerprint, which gave the highest $y$ in the training dataset. The analysis was conducted with a dimension of 248. d) Statistical trends for the extraction performance. Averages $y$ of the top 5% similar compounds are plotted as a function of dimension. Error bars show the standard deviation of the fivefold cross-validation results (see Supporting Information for details).]
maintained essential and universal information to achieve desirable performance. In contrast, there was no significant relationship between y and the similarity scores when the fingerprints of an existing compound with the highest y in the training dataset were compared (Figure 4c, blue plots). This was because the structural information of only one compound was too specific and narrow, and may have overlooked some essential structural information to achieve the desired performance. We calculated the average y of the top 5% most similar compounds (Figure 4d). Statistically, the annealing approach was preferable with various dimensions n rather than with existing fingerprints. Most of the extracted chemicals displayed z-scores > 1, corresponding to melting points of around 200 °C.

The proposed framework was useful for extracting chemicals with other desirable parameters, such as low melting temperature, high solubility parameter, high glass transition temperature, and high photoconversion efficiency in organic thin-film solar cells (Figure S5—S10, Supporting Information). This versatile framework could be used for a wide range of materials and properties, not only for organic molecules, but also for inorganic compounds. During the annealing processes, additional restriction conditions using the fingerprints of existing chemicals also improved the extraction performance. The restrictions increased the structural feasibility of xideal, whereas only y was pursued without the restriction (Figure S5, Supporting Information for further information).

4. Generating New Chemicals from Ideal Fingerprints

Instead of comparing the similarity scores with the existing database, directly generating molecules from xideal is preferable for accessing the vast chemical search space. With a large dimension n (e.g., > 1000), x can express a sufficiently large number of chemicals, 21000 ≈ 10300. Therefore, if appropriate mapping functions (fMAP and fMAP−1) are given, the annealing framework may be able to explore arbitrary organic molecules. However, determining the functions, especially the inverse functions, is still unwieldy, regardless of the remarkable progress in cheminformatics and machine learning.[1,12,14]

Here, we used a more straightforward approach to preparing chemicals directly from xideal. After decompressing xideal to the 2048-dimensional fingerprint, xideal,FP, chemical structures were generated from the fingerprint fragments (Figure 5a). A fingerprint bit xFP,i represents a specific fragment of a chemical structure. In the example in Figure 5a, bit xFP,0 indicates whether an aromatic ring is available in a molecule. The bit is 1 if the ring is present or else 0.

From the melting temperature database, chemical fragments were extracted by the breaking of the retrosynthetically interesting chemical substructures (BRICS) algorithm.[30] Then, some fragments sharing the same bit of the ideal fragment were extracted and reconnected to make molecules according to BRICS (Figure 5b, c, and Figure S11, Supporting Information).

The predicted y of the generated chemicals was statistically higher than those of the chemicals of the original database and randomly generated chemicals (control, Figure 5b). The average scores were 1.80, 0, and 0.71 for the proposed approach, original database, and random generation, respectively. Large values of y were even obtained from the ideal fingerprint (Figure 5c and Figure S11, Supporting Information). Our high-throughput molecular generation system will help researchers design new materials with desirable properties.

5. Optimizing the Composition of Li+-Conducting Electrolytes

Finally, we extended the annealing framework to find the optimal compositions of multiple chemicals forming composite materials.
The search space increases exponentially with \( N \)-component systems \((N > 1)\). We focused on finding the best composition of organic \( \text{Li}^+ \)-conducting electrolytes consisting of up to six chemicals (Figure 6a and Figure S12, Supporting Information).

\( \text{Li}^+ \)-conducting electrolytes were selected because of the essential role of \( \text{Li}^+ \) batteries in society and the demand for higher conductivity to improve outputs.\(^{[31]}\) Most electrolytes consist of at least two components, namely, a solvent and salt.\(^{[3,32–34]}\) Additional chemicals are often added to improve viscosity, conductivity, stability, and other battery properties.\(^{[32,34]}\) The search space becomes even larger when polymeric electrolytes are considered as essential components for next-generation solid-state batteries and conventional gel electrolytes.\(^{[3,33]}\)

We previously explored organic \( \text{Li}^+ \)-conducting electrolytes using machine learning.\(^{[3]}\) The largest database of \( \text{Li}^+ \)-conducting electrolytes covered both monomeric and polymeric conductors (around unique 1200 records near room temperature, Table S1 and Figure S13, Supporting Information). The prediction accuracy of a model trained with the database was as accurate as that of experienced researchers\(^{[3]}\) and the machine learning model identified new polymeric conductors. However, a narrow search space \((<10,000)\) was used during the screening and it is likely that candidates in the broader search space were overlooked.

Composition information about electrolytes was converted into binary vectors. For simplicity, up to six chemical

![Figure 6. a) Scheme to treat composite materials. b) Results for exploring optimal composites by the annealing, random sampling, and Bayes optimization approaches. The average scores for five independent trials are shown in the cases of the anneal and random approaches. Error bars represent standard deviations. Only one trial was conducted for the Bayes, because of the excessively long calculation time. The dimension of \( x \) was 398. c) Probability distribution of the composites extracted by the three approaches. d) Example structures of the extracted chemicals.](image-url)
components\(^3\) were considered for one composite. Weight ratios of the components were also included in \(x\). The continuous values were converted into binaries by a unary method (see Supporting Information for details). Then, regression was conducted to predicted room temperature conductivity. A linear model yielded high \(R^2\) scores of 0.88 and 0.68 for the training and testing datasets, respectively (Figure S13b, Supporting Information).

After obtaining \(x_{\text{ideal}}\) with the annealer from the regression model, the decompressed vector was compared with the fingerprints of a chemical compound database. There were around 10,000 candidate chemicals in the database. Therefore, the search space was as big as \(10^{10^4}\), ignoring weight ratios (i.e., an additional five independent variables).

Compounds were extracted from the database by comparing their fingerprints with those of the ideal compounds. The selections were made by a random process, where high-similarity compounds were extracted preferentially according to a Gaussian distribution. The weight ratio was calculated by restoring the corresponding binaries from the Gaussian distribution. The weight ratio was calculated by

\[
\text{weight ratio} = \frac{\text{weight of binary}}{\sum \text{weights of binaries}}
\]

These selection processes were executed five times to ensure statistical validity. On average, it took only around 3 s to obtain composites with \(y_{\text{pred}} > 1.3\). After finishing the 100 iterations, the maximum value increased to 1.4, which took about 14 s.

However, 10\(^5\) iterations were not sufficient to achieve \(y_{\text{pred}} = 1.3\) when the compounds and weight ratios were selected by a completely random process (Figure 6b). The average time to reach the maximum \(y_{\text{pred}} = 1.2\) was 15,600 s, whereas the annealing method took less than 0.85 s. Therefore, the latter was at least 15,600/0.85 \(\cong\) 18,000 times faster than the standard approach.

Bayesian optimization, a conventional approach to explore optimal conditions,\(^1,35\) did not find the best combination (Figure 6b). The best score was only 0.88 after 6000 iterations, although the method took an exceptionally long time of 1,240,000 s. The poor result was attributed to the nonsmoothness of the function and search space that was too large (see Figure S14 and Supporting Discussion, Supporting Information). It should also be noted that even recent deep learning techniques, such as reinforcement learning, still tend to suffer from multicomponent systems because the calculation normally takes at least several minutes even for a one-component system\(^16\) and the time will increase exponentially with \(n\)-components.

The annealing approach improved the statistical distribution of \(y_{\text{pred}}\) greatly. For instance, 7% of the proposed electrolytes provided \(y_{\text{pred}}\) of over 1.2, whereas in the random approach only 0.0002% did (Figure 6c). The speedup ratio for this specific task was as large as \(4 \times 10^7\). The proposed system finished the task in only 14 s, whereas the control method would have needed 5.5 \(\times 10^5\) s \(\cong\) 6 days. The Bayes method could not yield an electrolyte of \(y_{\text{pred}} > 1.2\). A simple comparison of the iteration rates (7.2 and 0.011 compounds per seconds for the annealing and Bayes, respectively) indicated that the annealing would be at least \(4 \times 10^4\) \(\times\) (7.2/0.0048) \(=\) 10\(^8\) times faster than the Bayes approach.

From the viewpoint of actual experiments, a sufficiently large number of candidates are needed. The materials prescreened by computers are still candidates extracted from \(f_{\text{ML}}\). Because the predictions are not perfect, most candidates would be excluded during the next screening steps (e.g., human-based composition check and real experiments). The poor extraction efficiency and excessively long calculation time of the conventional methods may not be useful for practical projects.

The quality of the extracted compounds obtained by annealing was also higher than that of the controls (Figures 6d and Figure S15–S17, Supporting Information). About half of the chemicals extracted by annealing were experimentally examined as electrolyte components for Li\(^+\) batteries\(^1,32\) (shown in yellow, Figure S15, Supporting Information). In contrast, 75% or more of the compounds extracted by the controls were unconventional (Figures S16 and S17, Supporting Information). Experimentally, most of the compounds proposed by the random approach would not provide ionic conductivity because the components were far from salt- or solvent-like structures (e.g., no polar solvents in a composite, Figure S16 and S17, Supporting Information).

High prediction scores could be obtained by \(f_{\text{ML}}\) even with noisy composites, whose structures were dissimilar from that of electrolytes, because of the lack of training data during machine learning. We trained the model only with Li\(^+\)-conducting electrolytes, not with non-Li\(^+\)-conducting composites. This was because current machine learning models can process only specific tasks in narrow fields due to the limited complexity of the model and computing power,\(^15\) and adding too broad a context may decrease the prediction accuracy.\(^16\) The preparation of databases, which is generally a manual process,\(^13,3\) is also hugely time-consuming. Therefore, the inaccurate predictions by \(f_{\text{ML}}\) are currently unavoidable and should be avoided during the candidate selection step.

The ratio of conventional and unconventional compounds in proposals is crucial during screening. If there are too many unconventional compounds, the proposed compounds will be “noisy” (Figures S16 and 17, Supporting Information), whereas suggesting too many conventional compounds hinders the discovery of new materials. The ratio of 50% provided by annealing may be acceptable during the exploration of potential candidates. The balance can be tuned by adding a restriction condition during annealing in a similar way to the single-component system (Figure S5, Supporting Information; future challenges are described in the Supporting Discussion part).

### 6. Conclusion

We constructed a rapid extraction system for chemicals using machine learning and a quantum-inspired annealing machine, specialized digital hardware to treat the Ising model.\(^22\) The framework allowed rapid exploration of chemicals and compositions, and the extraction was at least \(10^4\)–\(10^8\) times faster than conventional methods. Increasing the speed of prediction methods is essential to tackle the combinatorial explosion problem in chemical screening (e.g., \(10^{2000}\) candidates for an \(N\)-component system of small organic molecules). The structures extracted by our method were higher quality than those extracted by the control methods because the annealer could extract the essential features of desirable compounds. The screening system will be critical to dramatically shortening the time required to optimize versatile, functional materials and devices of organic (small and
macromolecules) and inorganic systems. Further improving the exploration efficiency by optimizing mapping functions ($f_{MAP}$ and $f_{MAP}^{-1}$), expanding materials databases for higher prediction accuracy, and inventing new regression and exploration algorithms (e.g., suppress overestimation of predicted values during screening; see Supporting Information for further discussion) will be our next challenges.

Supporting Information
Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements
This work was partially supported by Grants-in-Aid for Scientific Research (Nos. 17H03072, 18H05151, 20H05298, and 19K15638) from MEXT, Japan. The work was also partially supported by the Research Institute for Science and Engineering, Waseda University.

Conflict of Interest
The authors declare no conflict of interest.

Keywords
lithium-ion batteries, machine learning, materials informatics, organic functional materials, quantum-inspired annealing

Received: September 14, 2020
Revised: November 23, 2020
Published online: January 8, 2021

[1] R. Ramprasad, R. Batra, G. Piliania, A. Mannodi-Kanakkithodi, C. Kim, *Npj Comput. Mater.* 2017, 3, 54.
[2] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* 2018, 559, 547.
[3] K. Hatakeyama-Sato, T. Tezuka, M. Umeki, K. Oyaizu, *J. Am. Chem. Soc.* 2020, 142, 3301.
[4] A. Mannodi-Kanakkithodi, A. Chandrasekaran, C. Kim, T. D. Huan, G. Pilania, V. Botu, R. Ramprasad, *Npj Comput. Mater.* 2018, 4, 2639.
[5] M. H. S. Segler, M. Preuss, M. P. Waller, *Nature* 2018, 555, 604.
[6] R. S. Bohacek, C. McMartin, W. C. Guida, *Med. Res. Rev.* 1996, 16, 3.
[7] P. Kirkpatrick, C. Ellis, *Nature* 2004, 432, 823.
[8] F. C. Adams, G. Laughlin, *Rev. Mod. Phys.* 1997, 69, 337.
[9] J. N. Kumar, Q. Li, K. Y. T. Tang, T. Buonassisi, A. L. Gonzalez-Oyarce, J. Ye, *Npj Comput. Mater.* 2019, 5, 73.
[10] B. Sanchez-Lengeling, A. Aspuru-Guzik, *Science* 2018, 361, 360.
[11] K. Kim, S. Kang, J. Yoo, Y. Kwon, Y. Nam, D. Lee, I. Kim, Y. S. Choi, Y. Jung, S. Kim, W. J. Son, J. Son, H. S. Lee, S. Kim, J. Shin, S. Hwang, *Npj Comput. Mater.* 2018, 4, 67.
[12] K. Hatakeyama-Sato, K. Oyaizu, *Commun. Mater.* 2020, 1, 49.
[13] R. Gomez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernandez-Lobato, B. Sanchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, A. Aspuru-Guzik, *ACS Cent. Sci.* 2018, 4, 268.
[14] S. Wu, G. Lambard, C. Liu, H. Yamada, R. Yoshiida, *Mol. Inform.* 2020, 39, e1900107.
[15] M. Olivecrona, T. Blaschke, O. Engkvist, H. Chen, *Cheminform.* 2017, 9, 48.
[16] K. Kitai, J. Guo, S. J., S. Tanaka, K. Tsuda, J. Shiomi, R. Tamura, *Phys. Rev. Res.* 2020, 2, 013319.
[17] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, S. Lloyd, *Nature* 2017, 549, 195.
[18] A. D. King, J. Carrasquilla, J. Raymond, I. Ozfidan, E. Andriyash, A. Berkley, M. Reis, T. Lanting, R. Harris, F. Altomare, K. Boothby, P. I. Bunyk, C. Enderud, A. Frechette, E. Hoskinson, N. Ladizinsky, T. Oh, G. Poulin-Lamarre, C. Rich, Y. Sato, A. Y. Smirnov, L. J. Swenson, M. H. Volzmann, J. Wittaker, J. Yao, E. Ladizinsky, M. W. Johnson, J. Hilton, M. H. Amin, *Nature* 2018, 560, 456.
[19] O. A. von Lilienfeld, *Angew. Chem., Int. Ed.* 2018, 57, 4164.
[20] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunswoth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegeker, M. P. Harrison, J. Hartmann, A. Ho, M. Hoffmann, T. Huang, et al., *Nature* 2019, 574, 505.
[21] M. Aramon, G. Rosenberg, E. Valiante, T. Miyazawa, H. Tamura, H. G. Katzgraber, *Front. Phys.* 2019, 7, 48.
[22] E. Gibney, *Nature* 2017, 541, 447.
[23] M. W. Johnson, M. H. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, J. A. Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, C. J. Truncik, S. Uchaikin, J. Wang, B. Wilson, G. Rose, *Nature* 2011, 473, 194.
[24] Jean-Claude Bradley Open Melting Point Dataset, https://doi.org/10.6084/m9.figshare.1031638 (accessed: June 2020).
[25] D. Rogers, M. Hahn, *J. Chem. Inf. Model.* 2010, 50, 742.
[26] A. M. Bruckstein, D. L. Donoho, M. Elad, *SIAM Rev.* 2009, 51, 34.
[27] I. V. Tetko, Y. Sushko, S. Novotarskyi, L. Patiny, I. Kondratov, A. M. Bruckstein, D. L. Donoho, M. Elad, *J. Chem. Inf. Model.* 2014, 54, 3320.
[28] D. J. Rogers, T. T. Tanimoto, *Science* 1960, 132, 1115.
[29] J. Degen, C. Wegscheid-Gerlach, A. Zaliani, M. Rarey, *ChemMedChem* 2008, 3, 1503.
[30] J. B. Goodenough, K. S. Park, *J. Am. Chem. Soc.* 2013, 135, 1167.
[31] M. Ue, K. Ida, S. Mori, *J. Electrochem. Soc.* 1994, 141, 2989.
[32] Z. Xue, D. He, X. Xie, *J. Mater. Chem. A* 2015, 3, 19218.
[33] M. Marcinek, J. Szydek, M. Marczewski, M. Piszcz, L. Niedzicki, M. Kalita, A. Plewa-Marczewska, A. Bitner, P. Wieczorek, T. Trzciak, M. Kasprzyk, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, et al., *Front. Phys.* 2019, 7, 48.
[34] M. Aramon, G. Rosenberg, E. Valiante, T. Miyazawa, H. Tamura, H. G. Katzgraber, *Front. Phys.* 2019, 7, 48.
[35] E. Gibney, *Nature* 2017, 541, 447.
[36] M. W. Johnson, M. H. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, J. A. Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, C. J. Truncik, S. Uchaikin, J. Wang, B. Wilson, G. Rose, *Nature* 2011, 473, 194.