JCTC Journal of Chemical Theory and Computation

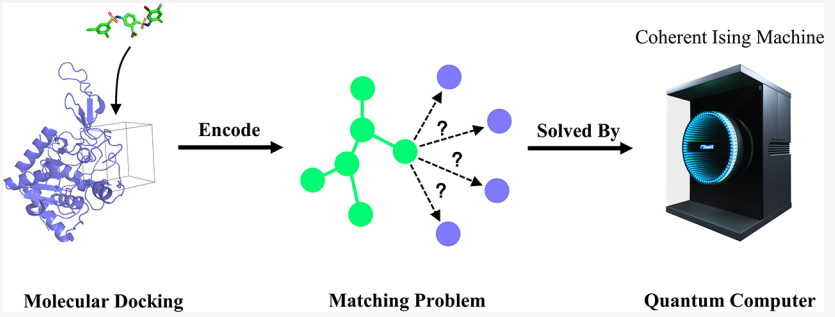
pubs.acs.org/JCTC Letter

Encoding Molecular Docking for Quantum Computers

Jinyin Zha, Jiaqi Su, Tiange Li, Chongyu Cao, Yin Ma, Hai Wei, Zhiguo Huang, Ling Qian, Kai Wen,* and Jian Zhang*

Cite This: https://doi.org/10.1021/acs.jctc.3c00943 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Molecular docking is important in drug discovery but is burdensome for classical computers. Here, we introduce Grid Point Matching (GPM) and Feature Atom Matching (FAM) to accelerate pose sampling in molecular docking by encoding the problem into quadratic unconstrained binary optimization (QUBO) models so that it could be solved by quantum computers like the coherent Ising machine (CIM). As a result, GPM shows a sampling power close to that of Glide SP, a method performing an extensive search. Moreover, it is estimated to be 1000 times faster on the CIM than on classical computers. Our methods could boost virtual drug screening of small molecules and peptides in future.

Molecular docking[1] is a widely used *in silico* technique in virtue screening, lead optimization, and mechanism study. It determines the pose and binding free energy (ΔG_{Bind}) between a ligand and a protein. Suppose the ligand consist of n atoms (a_1, a_2, ..., a_n), and the coordinate of a_i is $\mathbf{r}_i$. molecular docking is to minimize

$$\Delta G_{\text{Bind}}(a_1, a_2, ...,a_n, \mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n)\mathbf{r}_i \in \mathbf{D} \quad (1)$$

Domain **D** is a docking box restricting the pose of the ligand. Equation 1 is usually called a "scoring function". However, minimizing (1) is nondeterministic polynomial-time (NP)-hard,[2] suggesting that its precise solution requires vast enumeration of $\mathbf{r}_i$. To accelerate molecular docking, software like AutoDockFR[3] and GOLD[4] apply approximation algorithms like simulated annealing. Glide[5] applied enumeration but in a stepwise fashion, that different scoring functions, from rough to precise, are in order used to narrow the solution space step by step into the final docking poses. Recently, deep learning models,[6] especially generative models, have been explored for molecular docking. Due to a direct prediction of binding pose, deep learning models could be much faster, although their reliability might require further examinations. Besides algorithmic approaches, attempts have also been made on hardware, for example, using GPU[7,8] for highly parallel molecular docking. However, current status of molecular docking is still hard to screen billion-level databases like ZINC,[9] so that there is still an urgent need to develop new methods to boost molecular docking.

Quantum computers (QCs) like D-WAVE[10,11] (a quantum annealer) and the Coherent Ising Machine (CIM)[12−18] are new approaches to accelerate the solving of NP-hard problems. A 100,000 qubit CIM has been reported to solve MAX-CUT[15] 1000 times faster than advanced digital computers. Recently, they have also been explored in NP-hard problems in science like compressed sensing,[19] RNA folding,[20] molecular unfolding,[21] and protein protonation.[22] To solve an NP-hard problem with QCs, one way is to encode the problem into a quadratic unconstrained binary optimization (QUBO) model,[23,24] which is to minimize a quadratic form ($\mathbf{x}^{\mathrm{T}}Q_{r\times r}\mathbf{x}$), where $\mathbf{x}$ is r binary decision variables and Q is a coefficient matrix. In QCs, $\mathbf{x}$ is represented by the state of r qubits while the coefficient matrix $Q_{r\times r}$ is translated to interactions among qubits. The qubit system is evolved to reach a ground-state Hamilton, which corresponds to minima of the QUBO instance (see section S1 of the Supporting Information for

Received: August 27, 2023
Revised: December 1, 2023
Accepted: December 8, 2023

https://doi.org/10.1021/acs.jctc.3c00943
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

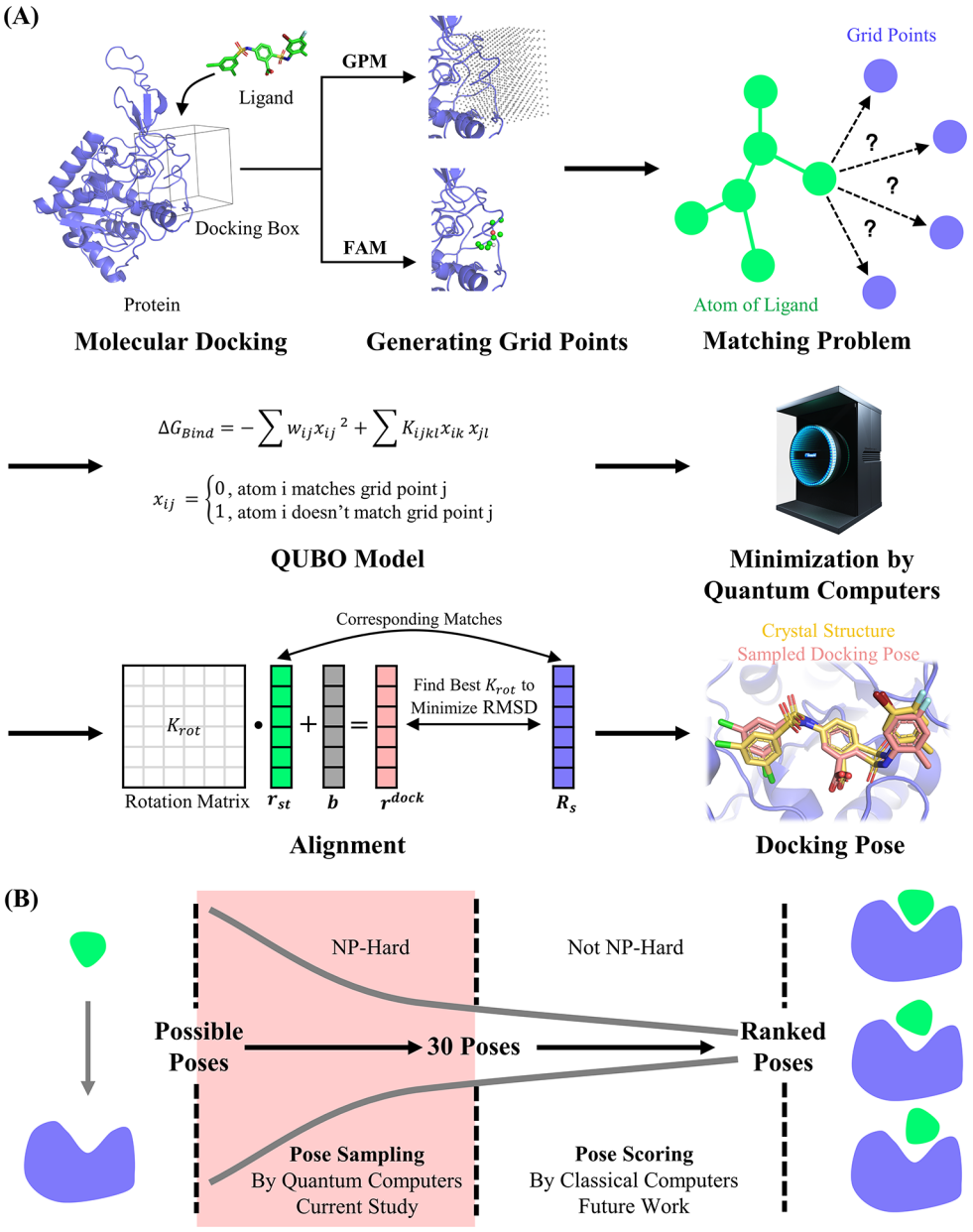


Figure 1. (A) Workflow of Grid Point Matching (GPM) and Feature Atom Matching (FAM). (B) Current study focuses on accelerating pose sampling in molecular docking with GPM and FAM.

more details). For example, in MAX-CUT, there are n fully connected vertexes (p_1, p_2, ..., p_n) and the weight of edge between p_i and p_j is m_{ij}. The goal is to split the vertexes into two sets (A and B), so that the total edge weights linking vertexes in different sets reaches a maximum. It is obvious that the decision variable x_i could be whether p_i belongs to A or B. Therefore, MAX-CUT could be written in eq 2 as a quadratic form. Minimization of f by QCs could yield a best group of x_i, which points to the division of vertexes.

$$f = \sum_{i=1}^{n} \sum_{j=i+1}^{n} m_{ij}(x_i - x_j)^2$$

$$= \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \sum_{j=2}^{n} m_{1,j} & -2m_{12} & \cdots & -2m_{1,n-1} & -2m_{1,n} \\ 0 & \sum_{j=3}^{n} m_{2,j} & \cdots & -2m_{2,n-1} & -2m_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \sum_{j=n}^{n} m_{n-1,j} & -2m_{n-1,n} \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}$$

$$x_i = \begin{cases} 1, & p_i \in A \\ 0, & p_i \in B \end{cases} \tag{2}$$

https://doi.org/10.1021/acs.jctc.3c00943
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Therefore, this Letter focuses on how to encode molecular docking (eq 1) into QUBO models (Figure 1A). In brief, we first translate molecular docking into a matching problem between ligand atoms and space positions and then describe the matching problem with a QUBO instance. In detail, the initial difficulty of the encoding is how to convert the continuous docking box **D** into a discrete domain of binary decision variables. Inspired by grid methods, first, we discretize **D** into N grid points (g_1, g_2, ..., g_N), to translate molecular docking into a matching problem between atoms of ligands and grid points. If a_i matches grid point g_{s_i}, the distance between a_i and a_j is d_{ij}, and the distance between g_{s_i} and g_{s_j} is D, ΔG_{Bind} in eq 1 could be estimated as

$$\Delta G_{\text{Bind}}(a_1, a_2, ...,a_n, \mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n) \approx \Delta G'_{\text{Bind}}(a_1, a_2, ...,a_n, g_{s_1}, g_{s_2}, ..., g_{s_n}) = \sum_{i=1}^{n} w_{a_i g_{s_i}}$$

$$s_i = 1, 2, ..., N \quad s.t.\ |d_{ij} - D_{s_i s_j}| \le c_{\text{dist}} \tag{3}$$

where $w_{a_i g_{s_i}}$ defines the fitness when a_i is placed around g_{s_i}, and could be viewed as an intermolecular term in ΔG_{Bind}, while the constraint term retains the shape of the ligand and could be regarded as an intramolecular term in ΔG_{Bind}. Note that due to discretization of the solution space ($(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n) \rightarrow (g_{s_1}, g_{s_2}, ..., g_{s_n})$), $\Delta G'_{\text{Bind}}$ is a rough estimation of ΔG_{Bind}.

Second, eq 3 is converted into a QUBO model. x_{ij} is applied as the binary decision variable to indicate whether a_i and g_j is matched. Therefore, eq 3 could be interpreted as a quadratic form in (4). Note that an additional constraint term is included in order to restrict that one atom should match, at most, one grid point. However, one gird point could match to more than one atom because of the discretization of the docking box **D**.

$$\Delta G'_{\text{Bind}}(a_1, a_2, ...,a_n, g_{s_1}, g_{s_2}, ..., g_{s_n}) = \Delta G'_{\text{Bind}}(a_1, a_2, ...,a_n, g_1, g_2, ..., g_N, x_{11}, x_{12}, ..., x_{nN}) = \sum_{i=1}^{n}\sum_{j=1}^{N} w_{a_i g_j} x_{ij}^{\ 2}$$

$$x_{ij} = \begin{cases} 1, & a_i \text{ matches } g_j \\ 0, & a_i \text{ does not match } g_j \end{cases}$$

$$s.t.\ |d_{ik} - D_{jl}| \le c_{\text{dist}} \text{ and } \sum_{i=1}^{n} x_{ij} \le 1 \tag{4}$$

Unlike the QUBO in eq 2, constraints exist in the current QUBO model. Therefore, we then converted constraints into two quadratic Lagrange terms in eq 5. K_{dist}, K_{mono}, and c_{dist} are three parameters that require parametrization.

$$\begin{aligned}\Delta G''_{\text{Bind}}(a_1, a_2, ...,a_n, g_1, g_2, ..., g_N, x_{11}, x_{12}, ..., x_{nN}) &= \sum_{i=1}^{n}\sum_{j=1}^{N} w_{a_i g_j} x_{ij}^{\ 2} + K_{\text{dist}} \sum_{i=1}^{n}\sum_{j=1}^{N}\sum_{k=1}^{n}\sum_{l=j+1}^{N} u_{ijkl} x_{ij} x_{kl} \\ &+ K_{\text{mono}} \sum_{i=1}^{n}\sum_{j=1}^{N}\sum_{k=1}^{n}\sum_{l=j+1}^{N} v_{ijkl} x_{ij} x_{kl} \\ &= \begin{pmatrix} x_{11} \\ x_{12} \\ ... \\ x_{nN} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} w_{a_1 g_1} & K_{\text{dist}} u_{1112} + K_{\text{mono}} v_{1112} & \cdots & K_{\text{dist}} u_{11nN} + K_{\text{mono}} v_{11nN} \\ 0 & w_{a_1 g_2} & \cdots & K_{\text{dist}} u_{12nN} + K_{\text{mono}} v_{12nN} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{a_n g_N} \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ ... \\ x_{nN} \end{pmatrix}\end{aligned}$$

$$u_{ijkl} = \begin{cases} 1, & |d_{ik} - D_{jl}| > c_{\text{dist}} \\ 0, & |d_{ik} - D_{jl}| > c_{\text{dist}} \end{cases} \qquad v_{ijkl} = \begin{cases} 1, & j = l \text{ and } i \ne k \\ 0, & \text{otherwise} \end{cases}$$

$$x_{ij} = \begin{cases} 1, & a_i \text{ matches } g_j \\ 0, & a_i \text{ does not match } g_j \end{cases} \tag{5}$$

The QUBO model could then be solved by QCs, as mentioned above. Due to their stochastic nature, QCs could output different results within different runs. These results are taken as sampled poses of ligand binding. Those poses could be further ranked via comparison of estimated ΔG_{Bind}.

Finally, solved matches are converted to docking poses. As a proof of concept, we only consider here the rigid docking, where the shape of the ligand is maintained. Therefore, the conversion is realized via an alignment. Suppose the standardize coordinate of a_i is $\mathbf{r}_i^{\text{st}}$ and it matches to grid point g_{s_i}, whose coordinate is $\mathbf{R}_{s_i}$, a Kabsch RMSD rotation matrix is calculated to be

$$K_{\text{rot}} = \arg\min_{K} \left[K \begin{pmatrix} \mathbf{r}_1^{\text{st}} \\ \mathbf{r}_2^{\text{st}} \\ ... \\ \mathbf{r}_n^{\text{st}} \end{pmatrix} + \mathbf{b} - \begin{pmatrix} \mathbf{R}_{s_1} \\ \mathbf{R}_{s_2} \\ ... \\ \mathbf{R}_{s_n} \end{pmatrix} \right], \text{ where } \mathbf{b} = \frac{\sum_{i=1}^{n} \mathbf{R}_{s_i}}{n} - \frac{\sum_{i=1}^{n} \mathbf{R}_i^{\text{st}}}{n} \tag{6}$$

b is a column vector unifying the center of ligand atoms and their matched grid points. The final docking pose is calculated by eq 7. The alignment is realized by the Prody[25,26] package.

$$\begin{pmatrix} \mathbf{r}_1^{\text{dock}} \\ \mathbf{r}_2^{\text{dock}} \\ ... \\ \mathbf{r}_n^{\text{dock}} \end{pmatrix} = K_{\text{rot}} \begin{pmatrix} \mathbf{r}_1^{\text{st}} \\ \mathbf{r}_2^{\text{st}} \\ ... \\ \mathbf{r}_n^{\text{st}} \end{pmatrix} + \boldsymbol{b} \tag{7}$$

To quantify $w_{a_i g_j}$ in eq 5, we introduced two encoding methods, differed by ways to generate grid points. The first is called Grid Point Matching (GPM), whose grid points are directly generated in a predefined docking box D with a gap of 2 Å. $w_{a_i g_j}$ is defined as the van der Waals energy when the atomic type of a_i is placed on g_i and is precalculated and saved on each g_i using AutoGrid in AutoDockFR.[3] In fact, in the scoring function of AutoDockFR, electrostatic and solvation terms are also included. They are neglected here due to poorer performances (see section S2 of the Supporting Information),

https://doi.org/10.1021/acs.jctc.3c00943
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

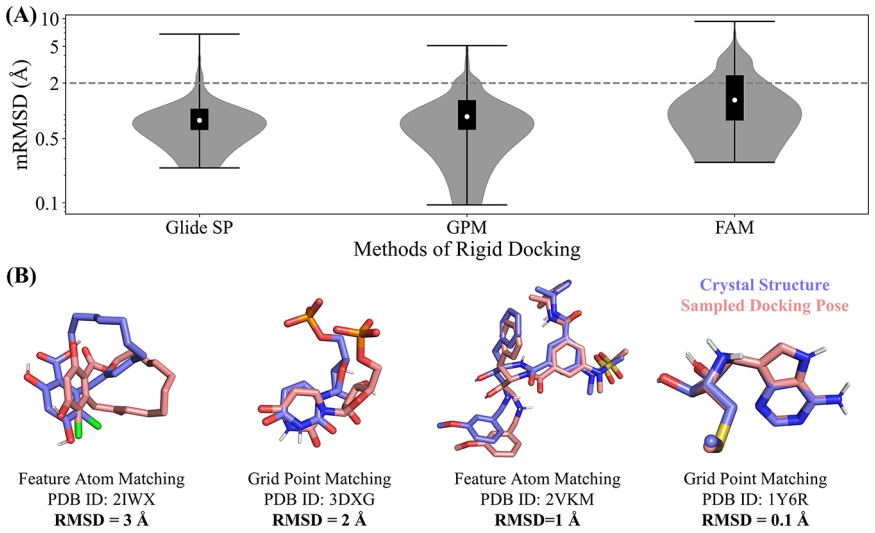


Figure 2. Performance and computational cost of Grid Point Matching and Feature Atom Matching. (A) Distribution of minimum RMSD in sampling (mRMSD) by three sampling methods including Glide SP, Grid Point Matching, and Feature Atom Matching. The gray dashed line represents the cutoff value of good docking poses (<2 Å). On each violin plot, the upper and lower lines define the range of values; the box defines values between 0.25 and 0.75 quartiles; the white dot points the median. GPM stands for "Grid Point Matching", while FAM stands for "Feature Atom Matching". (B) Examples of sampled docking poses (red) compared with poses in crystal structures (blue).

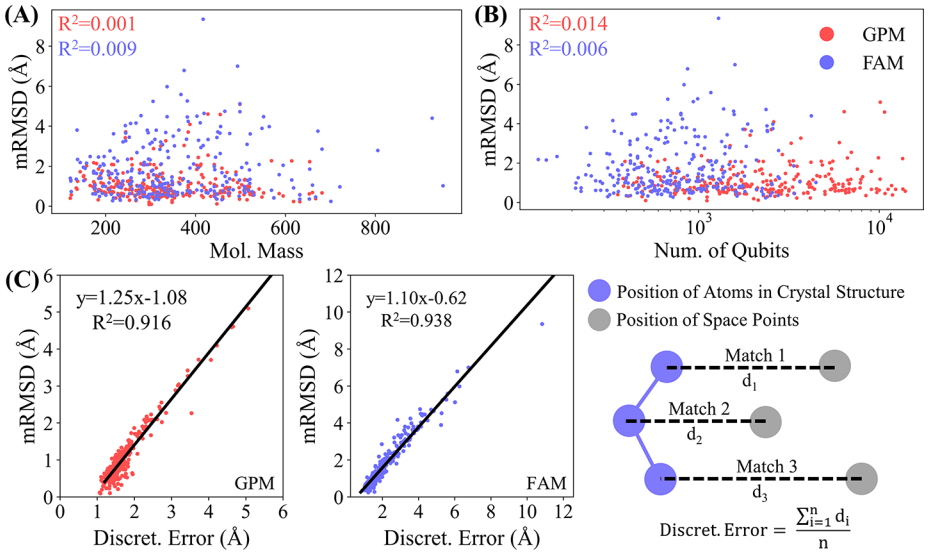


Figure 3. Correlation between mRMSD and molecular mass (A), number of qubits (B), and discretized error (C). The coefficients of determination (R^2) are listed on the top-left. The definition of Discret. Error is given in the bottom right, which is the mean of distances between the coordinate of a ligand atom in crystal structure and the coordinate of its matched space point. GPM stands for "Grid Point Matching", while FAM stands for "Feature Atom Matching".

possibly accounted for by accumulation of error in discrete space. Matches whose $w_{a_i g_j}$ is above 0 are droped for reducing the consumption of qubits. The second process is called Feature Atom Matching (FAM). In brief, gird points generated with a gap of 1 Å in the predefined docking box are coarse-grained into 3 types of "feature atoms" (FAs) using the AutoSite[27] algorithm (see section S3 of the Supporting Information for more details). A FA could be a neutral C, a H-bond-donor H or a H-bond-acceptor O. $w_{a_i g_j}$ in FAM is defined as $|\chi_{a_i} - \chi_{g_j}| - 0.5$, where χ_{a_i} and χ_{g_j} is the Pauling's electronegativity[28] of the ligand atom and FAs, respectively. The definition shows the similarity of the wiliness to form H-bonds.

It has been mentioned before that the scoring functions ($\Delta G'_{\text{Bind}}$) of GPM and FAM are a rough estimation of ΔG_{Bind}. In fact, we found eq 5 poor in ranking poses (see section S4 of the Supporting Information). Therefore, we split QC-based molecular docking into pose sampling and pose scoring, as shown in Figure 1B. Pose sampling is to sample possible solutions of the docking pose, while the following pose scoring is to rank these poses. Comparatively, pose sampling is NP-hard while pose scoring is not. Also, it only requires a rough scoring function. Therefore, in our workflow, GPM and FAM are used for pose sampling while the pose scoring could be left for classical computers.

Parameterization and benchmarking of both methods focus on their power to sample poses close to crystal structures. In

https://doi.org/10.1021/acs.jctc.3c00943
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

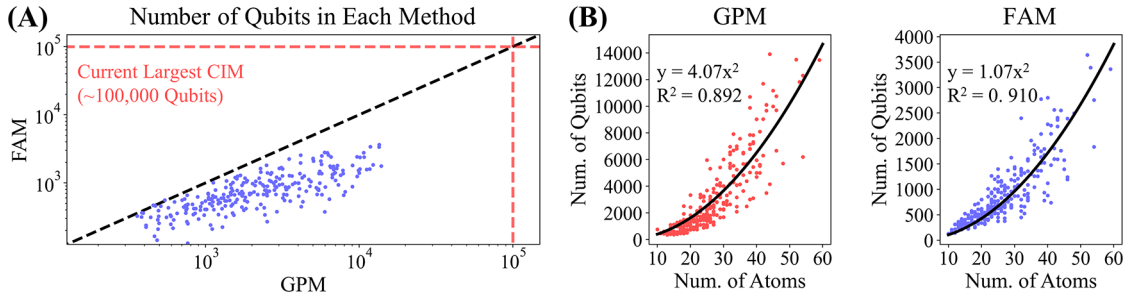


Figure 4. Computational cost on the CIM. (A) Comparison of number of qubits acquired by Grid Point Matching and Feature Atom Matching. The current largest CIM (~100,000 qubits) is drawn in red dashed lines. GPM stands for "Grid Point Matching", while FAM stands for "Feature Atom Matching". (B) Relationship between number of qubits and number of ligand atoms. Data are plot in red (GPM) or blue (FAM) scatters, while the fitted curves are drawn in black.

detail, redocking tests,[29] by which the ligand from a protein−ligand complex is docked back to its binding site, are performed. For each case, 30 poses were sampled, and the minimum root-mean-square distance (mRMSD) between sampled poses and the pose in the crystal structure was used to indicate the sampling power. Obviously, a smaller value represents a stronger ability for pose sampling. Cubic docking boxes for redocking tests are defined with the same center of the ligand, and their lengths are set at the largest distance among ligand atoms plus 8 Å. The parametrization was done on Astex Diversity Set (ADS) to minimize average mRMSD while the benchmarking was carried on CASF-2016.[30,31] Data cleansing, results of parametrization,[29,32] and further analysis of parameters could be found in sections S5 and S6, Table S1, and Table S2 of the Supporting Information. To realize such large amounts of work, a QC simulator built by the pyqubo[33] package was adapted here.

Sampling power evaluated by redocking tests is shown in Figure 2. GPM could sample high-quality poses (mRMSD < 2 Å) for 225 of 257 cases (87.5%), with an average mRMSD of 1.1 Å and a maximum mRMSD around 5 Å. While in FAM, high-quality poses are sampled only in 173 cases (67.3%), with a larger mRMSD in average and on maximum to be 1.8 and 9.4 Å, respectively. A better performance in GPM was not surprising due to less discretization of the docking box. The performance of GPM and FAM is not sensitive to grid positions (see section S7 and Table S3 of the Supporting Information for more details). To further quantify the sampling power, we compared our algorithms with the rigid sampling technique in Glide SP[5] (also by the redocking test described above), which is an exhaustive search so that it is believed to show a gold performance. Glide SP[5] could sample high-quality poses for 240 cases (93.4%) in CASF-2016, with an average mRMSD of 1.0 Å and a maximum mRMSD of 6.8 Å. Therefore, GPM has shown a close sampling power to Glide SP, ensuring its loading on a complete docking workflow (pose sampling + pose scoring).

We further seek the main factors affecting sampling performance in both methods. As shown in Figure 3. The mRMSD in sampling has almost no correlation with mass of the ligand (R^2 = 0.001 in RPM and R^2 = 0.009 in FAM) or number of qubits (R^2 = 0.014 in GPM and R^2 = 0.006 in FAM), suggesting our methods are not interfered with by the size of the ligands. However, mRMSD has a good linear relationship with discretization error in both techniques (R^2 = 0.938 in GPM and R^2 = 0.916 in FAM). Also, both regression coefficients are close to 1. This suggests that discretization of the docking box might mostly explain the mRMSD in both methods. Therefore, supposing that the box is less discretized and more qubits are used, molecular docking by QC could show a better performance.

Finally, we analyzed the computational costs of loading both algorithms on the CIM. The CIM can construct fully connected graphs, while others can only support local graphs with limited connections, making it suitable for mapping both algorithms. As shown in Figure 4A, the largest requirement on CASF is 13,908 qubits while that of FAM is 3,640, which is 1−2 magnitude smaller than the largest number of qubits of an experimental CIM[15] (~100,000 qubits), leaving space for further developing flexible docking methods. A quadric relationship was further estimated for both methods between the numbers of qubits and the numbers of ligand atoms (heavy atoms and polar Hs) in Figure 4B, which suggests the limit of input ligand atoms at about 156 (GPM) and 305 (FAM), roughly corresponding to peptides of 15 and 30 residues, respectively. Therefore, GPM and FAM are suitable for docking small molecules and peptides but not protein−protein docking. In addition, since the running time of MAX-CUT on the CIM is 593 us,[15] the running time of pose sampling is estimated to be in milliseconds, outperforming classical computers within 3 magnitudes.

To conclude, molecular docking,[34,35] and more precisely the pose sampling process in it, is NP-hard[2] and is burdensome for classical computers, which limits the speed and amount of virtue screening. QCs[11,12] are reliable tools to efficiently solve NP-hard problems. In fact, QCs have already been used to solve molecular unfolding,[21] which is a preparation step before molecular docking. However, QCs could not solve pose sampling or molecular docking directly. Solutions include using Gaussian Boson Sampling[36] or encoding the problem into QUBO models. Here, we proposed Grid Point Matching (GPM) and Feature Atom Matching (FAM). They first translate pose sampling into a matching problem between atoms and space positions and then into QUBO models. The solved matches are converted back to docking poses via an alignment. After extensive benchmarking of their sampling power, we have shown that GPM has a close performance with Glide SP, and the performance could be even better when a more computational resource is applicable. In the future, we are going to load GPM and FAM on a complete docking workflow on CIMs to perform virtue screening in billion-level databases like ZINC.[9]

https://doi.org/10.1021/acs.jctc.3c00943
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

■ ASSOCIATED CONTENT

Data Availability Statement

The Astex Diversity Set could be download from the Cambridge Crystallographic Data Centre at https://www.ccdc.cam.ac.uk/support-and-resources/Downloads/. The CASF-2016 is fetched from the PDBbind-CN database and is available at http://www.pdbbind.org.cn/download.php. The source code for both algorithms and a notebook for tutorial could be found at https://github.com/JinyinZha/QDock. The AutoDockFR 1.0 package used for the methods could be downloaded from https://anaconda.org/HCC/adfr-suite/files.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.3c00943.

Section S1 is a detailed description of solving QUBO Models with QCs. Section S2 explains why only van der Waals term is included in GPM. Section S3 introduces the AutoSite algorithm. Section S4 evaluates the docking power pf GPM and FAM. Section S5 introduces procedures of data collection. Section S6 introduces the parametrization procedures and results. Section S7 discusses the sensitivity of GPM and FAM. Table S1 lists the training set. Table S2 lists the test set. Table S3 lists the performance of Feature Atom Matching in 10 runs. (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Kai Wen − *Beijing QBoson Quantum Technology Co., Ltd., Beijing 100015, China*; Phone: +86-10-53392661; Email: wenk@boseq com; Fax: +86-10-53392661

Jian Zhang − *Medicinal Chemistry and Bioinformatics Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China*; orcid.org/0000-0002-6558-791X; Phone: +86-21-63846590; Email: Jian.zhang@sjtu.edu.cn; Fax: +86-21-64154900

Authors

Jinyin Zha − *Beijing QBoson Quantum Technology Co., Ltd., Beijing 100015, China; Medicinal Chemistry and Bioinformatics Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China*

Jiaqi Su − *Beijing QBoson Quantum Technology Co., Ltd., Beijing 100015, China*

Tiange Li − *Beijing QBoson Quantum Technology Co., Ltd., Beijing 100015, China*

Chongyu Cao − *Beijing QBoson Quantum Technology Co., Ltd., Beijing 100015, China*

Yin Ma − *Beijing QBoson Quantum Technology Co., Ltd., Beijing 100015, China*

Hai Wei − *Beijing QBoson Quantum Technology Co., Ltd., Beijing 100015, China*

Zhiguo Huang − *China Mobile (Suzhou) Software Technology Company Limited, Suzhou 215163, China*

Ling Qian − *China Mobile (Suzhou) Software Technology Company Limited, Suzhou 215163, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.3c00943

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by the National Key R&D Program of China (2023YFF1205103), the National Natural Science Foundation of China (81925034), the Innovation Program of Shanghai Municipal Education Commission (2019-01-07-00-01-E00036), the Starry Night Science Fund of Zhejiang University Shanghai Institute for Advanced Study (SN-ZJU-SIAS-007), the innovative research team of high-level local universities in Shanghai (SHSMU-ZDCX20212700), the open fund of state key laboratory of Pharmaceutical Biotechnology, Nanjing University (KF-202204), the GuangCi Professorship Program of Ruijin Hospital Shanghai Jiao Tong University School of Medicine, and the Key Research and Development Program of Ningxia Hui Autonomous Region (2022CMG01002) .

■ REFERENCES

(1) Saikia, S.; Bordoloi, M. Molecular Docking: Challenges, Advances and its Use in Drug Discovery Perspective. *Curr. Drug Targets* **2019**, *20*, 501−521.

(2) Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12964−75.

(3) Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* **2016**, *11*, 905−919.

(4) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Bio.* **1997**, *267*, 727−748.

(5) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(6) Wang, Y.; Shen, Y.; Chen, S.; Wang, L.; Ye, F.; Zhou, H. *ICLR 2023*; ICLR, 2023; pp https://openreview.net/pdf?id=ySCL-NG_I3.

(7) Yu, Y. J.; Cai, C.; Wang, J. Y.; Bo, Z. H.; Zhu, Z. D.; Zheng, H. Uni-Dock: GPU-Accelerated Docking Enables Ultralarge Virtual Screening. *J. Chem. Theory Comput.* **2023**, *19*, 3336−3345.

(8) Huang, Y. P.; Zhang, H.; Jiang, S. Y.; Yue, D. J.; Lin, X. H.; Zhang, J.; Gao, Y. Q. DSDP: A Blind Docking Strategy Accelerated by GPUs. *J. Chem. Inf. Model* **2023**, *63*, 4355−4363.

(9) Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. ZINC20-A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 6065−6073.

(10) Cohen, E.; Tamir, B. D-Wave and predecessors: From simulated to quantum annealing. *Int. J. Quantum. Inf.* **2014**, *12*, 1430002.

(11) Yarkoni, S.; Raponi, E.; Back, T.; Schmitt, S. Quantum Annealing for Industry Applications: Introduction and Review. *Rep. Prog. Phys.* **2022**, *85*, 104001.

(12) Wang, Z.; Marandi, A.; Wen, K.; Byer, R. L.; Yamamoto, Y. Coherent Ising machine based on degenerate optical parametric oscillators. *Phys. Rev. A* **2013**, *88*, 063853.

(13) McMahon, P. L.; Marandi, A.; Haribara, Y.; Hamerly, R.; Langrock, C.; Tamate, S.; Inagaki, T.; Takesue, H.; Utsunomiya, S.; Aihara, K.; Byer, R. L.; Fejer, M. M.; Mabuchi, H.; Yamamoto, Y. A fully programmable 100-spin coherent Ising machine with all-to-all connections. *Science* **2016**, *354*, 614−617.

(14) Inagaki, T.; Haribara, Y.; Igarashi, K.; Sonobe, T.; Tamate, S.; Honjo, T.; Marandi, A.; McMahon, P. L.; Umeiki, T.; Enbutsu, K.; Tadanaga, O.; Takenouchi, H.; Aihara, K.; Kawarabayashi, K.-i.;

Inoue, K.; Utsunomiya, S.; Takesue, H. A coherent Ising machine for 2000-node optimization problems. *Science* **2016**, *354*, 603−606.
(15) Honjo, T.; Sonobe, T.; Inaba, K.; Inagaki, T.; Ikuta, T.; Yamada, Y.; Kazama, T.; Enbutsu, K.; Umeiki, T.; Kasahara, R.; Kawarabayashi, K.-i.; Takesue, H. 100,000-spin coherent Ising machine. *Sci. Adv.* **2021**, *7*, No. eabh0952.
(16) Lu, B.; Fan, C. R.; Liu, L.; Wen, K.; Wang, C. Speed-up coherent Ising machine with a spiking neural network. *Opt. Express* **2023**, *31*, 3676−3684.
(17) Honjo, T.; Inaba, K.; Inagaki, T.; Ikuta, T.; Yamada, Y.; Takesue, H. A coherent Ising machine based on a network of 100,000 degenerate optical parametric oscillator pulses. *IEEE Conf. Nanotechnol.* **2022**, 405−408.
(18) Marandi, A.; Wang, Z.; Takata, K.; Byer, R. L.; Yamamoto, Y. Network of time-multiplexed optical parametric oscillators as a coherent Ising machine. *Nat. Photonics* **2014**, *8*, 937−942.
(19) Aonishi, T.; Mimura, K.; Okada, M.; Yamamoto, Y. L0 regularization-based compressed sensing with quantum-classical hybrid approach. *Quant. Sci. Technol.* **2022**, *7*, 035013.
(20) Zaborniak, T.; Juan, G.; Mueler, H.; Jabbari, H.; Stege, U. *IEEE International Conference on Quantum Computing and Engineering*; IEEE, 2022; pp 174−185.
(21) Mato, K.; Mengoni, R.; Ottaviani, D.; Palermo, G. Quantum molecular unfolding. *Quant. Sci. Technol.* **2022**, *7*, 035020.
(22) Hu, H. Quantum Computer Simulation of Protein Protonation. *J. Chem. Theory Comput.* **2023**, *19*, 5671−5676.
(23) Pambudi, D.; Kawamura, M. Constructing the Neighborhood Structure of VNS Based on Binomial Distribution for Solving QUBO Problems. *Algorithms* **2022**, *15*, 192−206.
(24) Date, P.; Patton, R.; Schuman, C.; Potok, T. Efficiently embedding QUBO problems on adiabatic quantum computers. *Quantum Inf. Process.* **2019**, *18*, 117−147.
(25) Bakan, A.; Dutta, A.; Mao, W.; Liu, Y.; Chennubhotla, C.; Lezon, T. R.; Bahar, I. Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinformatics* **2014**, *30*, 2681−3.
(26) Bakan, A.; Meireles, L. M.; Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **2011**, *27*, 1575−7.
(27) Ravindranath, P. A.; Sanner, M. F. AutoSite: an automated approach for pseudo-ligands prediction-from ligand-binding sites identification to predicting key ligand atoms. *Bioinformatics* **2016**, *32*, 3142−3149.
(28) Politzer, P.; Murray, J. S. Electronegativity: A continuing enigma. *J. Phys. Org. Chem.* **2023**, *36*, e44066.
(29) Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T. M.; Mortenson, P. N.; Murray, C. W. Diverse, High-Quality Test Set for the Validation of Protein-Ligand Docking Performance. *J. Med. Chem.* **2007**, *50*, 726−741.
(30) Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R. Comparative Assessment of Scoring Functions: The CASF-2016 Update. *J. Chem. Inf. Model* **2019**, *59*, 895−913.
(31) Li, Y.; Su, M.; Liu, Z.; Li, J.; Liu, J.; Han, L.; Wang, R. Assessing protein-ligand interaction scoring functions with the CASF-2013 benchmark. *Nat. Protoc.* **2018**, *13*, 666−680.
(32) Martinez-Cantin, R. BayesOpt: A Bayesian Optimization Library for Nonlinear Optimization, Experimental Design and Bandits. *J. Mach. Learn. Res.* **2014**, *15*, 3735−3739.
(33) Zaman, M.; Tanahashi, K.; Tanaka, S. PyQUBO: Python Library for Mapping Combinatorial Optimization Problems to QUBO Form. *Ieee. T. Comput.* **2022**, *71*, 838−850.
(34) Li, S.; Zhang, J. M.; Lu, S. Y.; Huang, W. K.; Geng, L.; Shen, Q. C.; Zhang, J. The Mechanism of Allosteric Inhibition of Protein Tyrosine Phosphatase 1B. *PLoS One* **2014**, *9*, No. e97668.
(35) Zhang, X. J.; Zhang, O.; Shen, C.; Qu, W. L.; Chen, S. C.; Cao, H. Q.; Kang, Y.; Wang, Z.; Wang, E. C.; Zhang, J. T.; Deng, Y. F.; Liu, F. R.; Wang, T. Y.; Du, H. Y.; Wang, L. C.; Pan, P. C.; Chen, G. Y.; Hsieh, C. Y.; Hou, T. J. Efficient and accurate large library ligand docking with KarmaDock. *Nat. Comput. Sci.* **2023**, *3*, 789−804.
(36) Banchi, L.; Fingerhuth, M.; Babej, T.; Ing, C.; Arrazola, J. M. Molecular docking with Gaussian Boson Sampling. *Sci. Adv.* **2020**, *6*, No. eaax1950.

https://doi.org/10.1021/acs.jctc.3c00943
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX