We thank the reviewers for the thoughtful feedback in these difficult times caused by the global COVID-19 pandemic.

1. LCAO is not really a basic assumption (l38), but a first-order approximation, and ends up being nonlinear (R2). We agree that LCAO serves as a first-order approximation; in the revised manuscript, we have corrected line l38 and the other relevant sentences. However, in QM9, the energy is calculated using LCAO (6-31G); therefore, when QM9 is used for training, the model must be based on LCAO, and QDF achieved high extrapolation performance. If the dataset included "experimental data," a nonlinearity would be entailed.

2. The HK map is nonlinear and modeled with a feedforward DNN, which is not an adequate conclusion without evidence (R2). The feedforward DNN is a choice for modeling the nonlinearity of the HK map from ρ to V; indeed, the HK map has been modeled using kernel methods [23]. As you may be aware, our HK map is modeled in an LDA (not a GGA) fashion. We emphasize that even this LDA-like HK map achieved high extrapolation performance.

3. Results on full QM9 (R2). QDF requires substantially more training time and memory than those required by a GCN, because of the large number of GTOs, grid field points, and the alternate learning of two loss functions. Therefore, we could not train QDF on all 130,000 molecules included in QM9 using our current computational environment. Training QDF on full QM9 necessitates the efficient use of dozens of GPUs. We will address this in future work.

4. The comparison of LCAO to GCN seems somewhat forced. QDF could be presented without this comparison (R2). Considering the NeurIPS community, it is important to provide the comparison between LCAO and GCN with regard to the recent trend of molecular GCNs. Of course, QDF can be proposed without a comparison to GCN. In fact, we have another work involving the application of QDF to materials informatics that will be published as a physics paper. It would not make much sense to describe LCAO in detail and compare LCAO to GCN in that paper.

5. Spherical harmonics in GTO (R2). Thanks for this critical comment. Yes, we omitted the spherical harmonics in the GTO for simplicity. As you have commented, we simplified some descriptions so they can be easily understood by readers in the NeurIPS community who are not familiar with physics and chemistry. We emphasize that even though the simplified GTO was employed, QDF achieved high extrapolation performance. The improvement of the GTO by considering spherical harmonics and other factors is an important future work.

6. Eikenberg 2018 (R2). Thank you for bringing this important related study. We did not cite Eikenberg 2018; however, we have already cited some ML studies that use GTOs for modeling the electron density as [21,26]. In the revised manuscript, we have additionally cited Eikenberg 2018 and discuss these related studies in detail.

7. An extrapolation evaluation of splitting according to heavy atoms (R2). Thanks for the constructive and beneficial feedback. In this rebuttal period, we split QM9 according to the heavy atoms and evaluated the extrapolation performance. After training the model using data regarding molecules with less than 7 heavy atoms (# of samples is 3,000), the interpolation error is 0.10 eV; further, the extrapolation error for molecules with 8 and 9 heavy atoms (# of samples is 127,000) is 0.19 eV. As you mentioned, QM9 has a substantial bias (97 % of QM9 corresponds to 8 and 9 heavy atoms); however, our QDF was successful in performing the extrapolation.

8. The phrases of "ML researchers" and "ML community" (R2). We apologize for using these somewhat confusing phrases. To prevent confusion among readers, we have excluded these phrases from the revised manuscript.

9. QDF is only compared on a single task (R3). We used a single benchmark dataset; however, we evaluated the extrapolation performances for three different energy properties in QM9. For predicting other properties, such as the HOMO and LUMO, QDF needs to capture the excited states of molecules, which is different from the ground-state energy and will be more difficult to predict. We will perform model extension in this regard in future work.

10. The computational complexity of QDF (R4). As we had mentioned in Response 3, considering the grid field points of all molecules entails a high memory cost. However, once the model is trained (e.g., using a GPU, the model can be trained for 10,000 molecules within 24 hours), the prediction can be performed within a second for a molecule.

11. Why is the dimensionality the same as the number of basis functions, and what do these dimensions represent (R4)? An initial assumption in LCAO is that the number of molecular orbitals N is equal to the number of atomic orbitals (or basis functions) in the linear combination. This is why the number of orbitals N corresponds to the dimensionality of the vector on a position r, i.e., $\boldsymbol{\psi}(r) \in \mathbb{R}^N$, and this N determines the computational approximation and accuracy. In the revised manuscript, we have clarified this, particularly for readers in the NeurIPS community.

12. QDF can be more broadly placed in the genre of "ML for quantum chemistry" (R4). Yes, QDF is an ML model for quantum chemistry and can be viewed as one of the "physics-oriented" approaches, such as the physics-informed, Hamiltonian (graph), and Fermionic neural networks (PINN, HNN, and FNN) [Raissi 2019, Pun 2019, Greydanus 2019, Sanchez-Gonzalez 2019, Pfau 2019]. Among these, the HNN, which was proposed in the NeurIPS community, involves physical considerations in training the neural network and addresses some classical mechanics problems. In that sense, our QDF is also such a neural network, but it focuses on solving quantum chemistry problems.