We thank the reviewers for their insightful comments. Generally, we noticed that we might not have explained clear enough that generating 3d molecular geometries is an independent, quite novel task and not to be confused with generating molecular graph connectivity. We added the following clarification to the ms: *"Note, that we explicitly do* not *aim to generate molecular graphs, but the atomic types and positions, which include the full information necessary to solve the electronic problem in the Born-Oppenheimer approximation. Thus, we avoid abstract concepts such as bonds and rings, that are not rooted in quantum mechanics and rely heavily on heuristics. We use these constructs only for evaluation and comparison to graph-based approaches."*

**Rev. #2** ***"Auxiliary tokens"*** We agree that our notation of the focus token was confusing. While explicitly mentioned in Sec. 3, we implicitly treat it as an atom with a special type in Sec. 4. We unified the notation to avoid confusion and clarified the significance of the focus token, as it constrains the space to place the next atom to a small region. The method would be impractical without it, as the space spanned by the grid would have to grow with the molecule. Note that we do only place atoms, not bonds (i.e. edges), as these are artificial constructs based on heuristics, e.g. overlapping covalent radii. Thus, we use bonds and rings only to visualize and compare to graph-based methods. ***"Incremental"*** To our knowledge, we present the *first generative neural network for molecular 3d structures with arbitrary composition*. While SchNet is used as the encoder for partial molecules, the overall G-SchNet architecture is novel. We added further baselines comparing to generation of graphs (Tab. 1) and geometries (Mansimov et al., see Rev #4). ***"Data filtering"*** We bias the distribution sampled from our network by fine-tuning on a subset of molecules with the desired property. This is *transfer learning*, not a heuristic. ***Improvements*** We added baseline methods (Tab. 1) and additional tasks from QM9 for targeted discovery (Fig.1 a-c)

**Rev. #3** ***Improvements*** The focus token is essential to our architecture, as explained above and clarified in the ms. We have added an ablation study for the origin token (Fig. 1 d-g). Without the origin token, the validity of generated molecules significantly drops (almost by 20%) and the amount of generated structures faithful to the training data also decreases substantially.

**Rev. #4** ***"Relaxation"*** Added to ms: *"Unlike molecular graphs which implicitly assume a structural equilibrium, G-SchNet is not restricted to a particular configuration and instead learns to generate equilibrium molecules purely by training on the QM9 data set. In order to assess the quality of our model, we need to compare the generated 3d structures to their relaxed counterparts, i.e. the closest local minimum on the potential energy surface. Those are found by minimizing the energy w.r.t. the atom positions."* ***"Related work"*** The reviewer is correct that Mansimov et al. "translate" from graphs to 3d geometry while G-SchNet generates 3d structures from scratch. Their approach could be coupled with graph generative models but these would still be inherently limited to targeting properties that do not crucially depend on the spatial configuration of molecules. We expanded the related work accordingly. We also added Mansimov et al. as a further baseline, noting that the comparison is to be taken with a grain of salt due to the different approach: While Mansimov et al. achieve median RMSD of 0.39Å (0.37Å with force field post-processing) when translating graphs from QM9, the generated structures from G-SchNet contained in the QM9 test set reach a median RMSD of 0.21Å. ***Improvements*** We clarified the relaxation procedure (see above) and added an ablation study for the origin token (Fig. 1 d-g).

*Table 1: Baseline comparisons added to ms (compressed to fit rebuttal).*

| | **G-SchNet** 3d structure | **CGVAE***[21] graph | **GraphVAE***[59] graph | **NeVAE***[60] graph | **LSTM***[21] SMILES | **CVAE***[18] SMILES | **GVAE***[31] SMILES |
|---|---|---|---|---|---|---|---|
| **valid** | 77.15% | 100.00% | 61.00% | 98.00% | 94.78% | 10.00% | 30.00% |
| **novel** | 87.47% | 94.30% | 85.00% | 100.00% | 82.98% | 90.00% | 95.44% |
| **unique** | 91.87% | 98.57% | 40.90% | 99.86% | 96.94% | 64.50% | 9.30% |

*statistics are taken from Fig. 3 in [21]

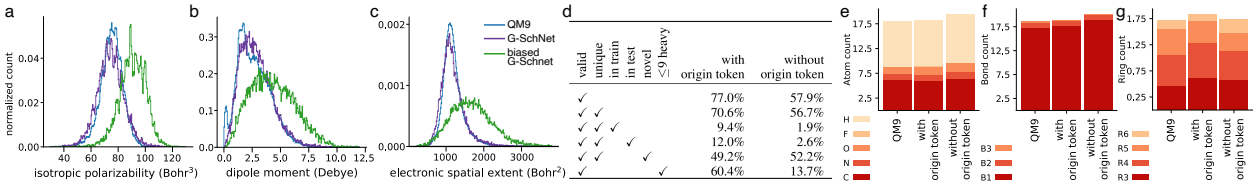


*Figure 1: Additional experiments added to ms: more targets on QM9 (**a-c**) and ablation study (**d-g**) (compressed to fit rebuttal). (**a-c**) Distribution of three quantum-chemical properties for molecules from QM9 (blue), generated by an unbiased G-SchNet (purple), and generated by G-SchNets biased towards larger values of the respective property (green). (**d-g**) Ablation study on the effect of the origin token.*