

A SUPPLEMENTARY MATERIAL

A.1 SIMULATION AND TRAINING SETTINGS

A.1.1 DFT

In the DFT simulations, the PYSCF package and its Generalized Kohn-Sham (GKS) function were employed. A convergence limit of 1×10^{-9} and a maximum of 2000 iterations ensured accurate total energy values for unstable molecules. The default optimizer of GKS was utilized, setting the simulation basis to *ccpvdz* and using *LDA,VWN* as the XC functional. Additional simulation settings are detailed in Table 4.

Table 4: **Input variables for in-house DFT simulations.** Coordinates for each element in the molecules are separated by ‘;’. ‘NA’ denotes not applicable, used when there is no other element to bond with.

Material	3D Coordinates (x, y, z; x, y, z) (Å)	Spin	Bond Length (Å)
H	0, 0, 0	1	NA
Li	0, 0, 0	1	NA
Be	0, 0, 0	0	NA
B	0, 0, 0	1	NA
C	0, 0, 0	2	NA
N	0, 0, 0	3	NA
O	0, 0, 0	2	NA
Li — H	0, 0, 0; 0, 0, 3	2	3.0
Li — Li	0, 0, 0; 0, 0, 5	2	5
C \equiv O	0, 0, 0; 0, 0, 2.2	0	2.2
N \equiv N	0, 0, 0; 0, 0, 2.1	2	2.1

A.1.2 FERMINET AND PSIFORMER

In our simulations for FERMINET and PSIFORMER, we employed the code from their GitHub pages (James S. Spencer & Contributors, 2020). The batch size was adjusted to 64 to align with T2E’s training. Training was halted when the mean variance was < 0.001 and *pmove* was < 0.5 , after observing similar values in the referenced papers. This approach allowed for consistent, efficient training across methods, avoiding excessive training time. The default *KFAC* optimizer was used (Martens & Grosse, 2015). Inputs adhered to PySCF style, with only specified settings altered. For molecule simulations, results from the original papers sufficed due to the high similarity in precision and information about their scaling. This decision averted the need for time-intensive in-house simulations yielding similar results. For the molecular simulations, we utilized published results. Given the high similarity in precision and provided scaling information, in-house simulations, requiring significant time, were deemed unnecessary for comparable precision.

A.1.3 VMC

We directly adopted the results from the FERMINET work, aligning our configurations with FERMINET and PSIFORMER. This approach ensured consistency while avoiding redundant recalculations.

A.1.4 T2E

T2E’s simulation settings mirror those of DFT, with the substitution of the Generalized Hartree-Fock method, in alignment with FERMINET and PSIFORMER pretraining. Around 6000 data points were generated for each element and molecule tested. In section 4.2, half of each dataset was randomly selected and combined, totaling 9000 training and 3000 test data points. The datasets encompass coordinates from 0 to 3 at 0.2 intervals. Batch sizes of 64 and 1 were utilized for single-element (section 4.1) and new material/molecule experiments (sections 4.2 and 4.3), respectively.

Table 5: **Input variables for in-house FERMINET and PsiFORMER simulations.** The semicolon in the coordinate separates element coordinates in the molecules. ‘NA’ denotes the absence of another element for bonding.

MATERIAL	3D COORDINATES (X, Y, Z; X, Y, Z)	SPIN	Bond Length (a_0)
H	0, 0, 0	1	NA
Li	0, 0, 0	1	NA
Be	0, 0, 0	0	NA
B	0, 0, 0	1	NA
C	0, 0, 0	2	NA
N	0, 0, 0	3	NA
O	0, 0, 0	2	NA
Li — H	Check Reference	Check Reference	3.0
Li — Li	Check Reference	Check Reference	5
C \equiv O	Check Reference	Check Reference	2.2
N \equiv N	Check Reference	Check Reference	2.1

Lastly, we applied vector augmentation to our embedded text vector, setting the noise standard deviation to 0.5, shift range to 128, and augmentation probability to 0.7.

Table 6: **Simulation variables used for T2E.** The ‘;’ in coordinates separates elements in molecules. ‘NA’ denotes the absence of additional elements for bonding.

Material	3D Coordinates (x, y, z; x, y, z) (Å)	Spin	Bond Length (Å)
H	Varying	1	NA
Li	Varying	1	NA
Be	Varying	0	NA
B	Varying	1	NA
C	Varying	2	NA
N	Varying	3	NA
O	Varying	2	NA
Li — H	0, 0, 0; Varying	2	3.0
Li — Li	0, 0, 0; Varying	2	3.0
N \equiv N	0, 0, 0; Varying	2	2.2
C \equiv O	0, 0, 0; Varying	0	2.2

A.2 DETAILED ANALYSIS OF EXPERIMENTS

A.2.1 FOR SINGLE ELEMENTS

The detailed precision for single element experiments in Table 2 is further elaborated in Table 7. Even upon detailed examination, the accuracy of FERMINET, PsiFORMER, VMC, and T2E closely aligns. T2E stands out by offering reduced parameters and model complexity. As anticipated, DFT consistently ranks last for each element, underscoring its diminishing precision with escalating material complexity. The loss graphs in Figure 6 demonstrate the expected trend: as system complexity augments, the model requires more time to optimize and ensure accurate predictions, owing to the increased particle count in the system.

A.2.2 FOR MOLECULES

Molecule experiments, initially presented in Table 3, are more precisely explored in Table 8. This extended analysis echoes the observations from single-element experiments (section A.2.1), with T2E, FERMINET, and PsiFORMER demonstrating consistent accuracy levels. DFT remains an out-

Table 7: **Total energy predictions with for first and second row elements (H to O) in higher precision, excluding He..** This table outlines total energy values for DFT, FERMINET, and PSI-FORMER from our simulations, alongside VMC and literature reference values, all expressed in Hartrees. 'NA' stands for not available.

ELEMENT	REFERENCE	DFT	T2E	FERMINET	PSIFORMER	VMC
Hydrogen (H)	-0.4998	-0.4774	-0.4992	-0.5001	-0.4999	NA
Lithium (Li)	-7.4780	-7.3413	-7.4324	-7.5017	-7.4634	-7.4780
Beryllium (Be)	-14.6674	-14.4435	-14.5723	-14.6945	-14.6598	-14.6672
Boron (B)	-24.6539	-24.3493	-24.5300	-24.6402	-24.6358	-24.6534
Carbon (C)	-37.8450	-37.4326	-37.6865	-37.8276	-37.8360	-37.8438
Nitrogen (N)	-54.5892	-54.1152	-54.3911	-54.5942	-54.5931	-54.5873
Oxygen (O)	-75.0673	-74.4945	-74.7921	-75.0571	-75.0798	-75.0632

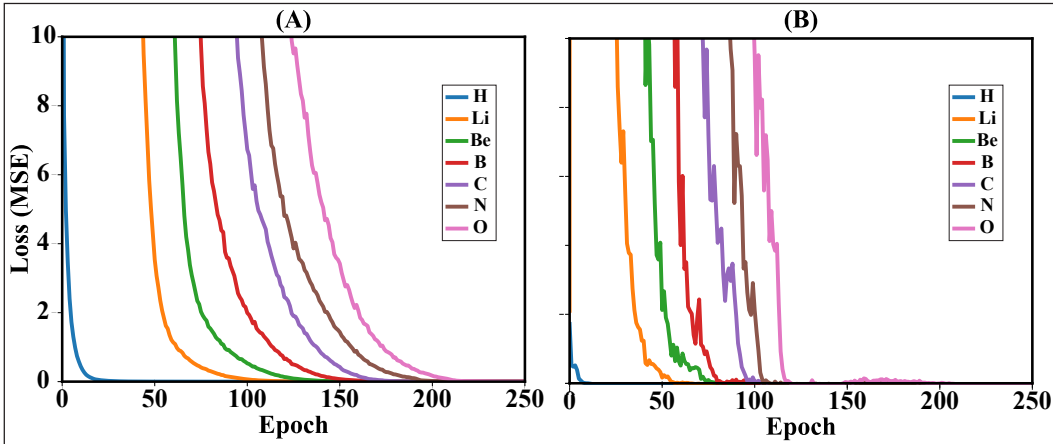


Figure 6: **Training and validation loss graphs for T2E on individual elements.** (A) Training loss. (B) Validation loss.

lier, exhibiting accuracy only for Li — H, analogous to a Be with additional interactions. Notably, DFT fails to reach convergence for $C \equiv O$, underscoring a demand for enhanced optimization cycles within the self-consistent loop.

Table 8: **Enhanced precision in total energy predictions for common molecules.** The table describes total energy estimations from FERMINET and PSIFORMER, referencing previous studies, and DFT computations from PYSCF simulations. All presented values are expressed in Hartrees. Note: "NC" signifies non-convergence.

MOLECULE	REFERENCE	T2E	FERMINET	PSIFORMER	DFT
Li — H	-14.9954	-15.0073	-14.9949	-14.9949	-14.6665
Li — Li	-8.0706	-7.9667	-8.0705	-8.0705	-7.8031
N \equiv N	-109.5423	-110.4302	-109.5418	-109.5418	-108.0186
C \equiv O	-112.9	-113.6524	-113.3247	-113.3247	-102.1189 (NC)

Additionally, Figure 7 illustrates the loss graph for T2E’s molecular performance. As anticipated, performance diminishes with increasing material complexity.

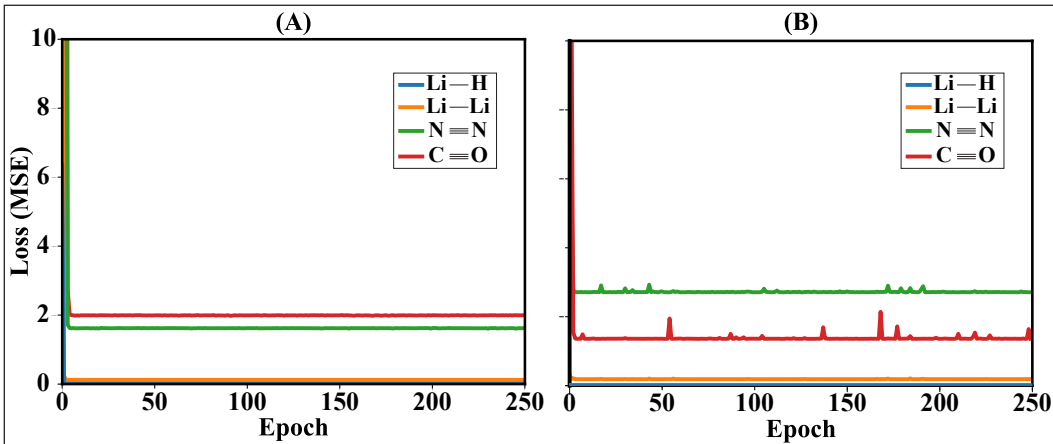


Figure 7: **Train and validation MSE loss graphs of T2E for molecules.** The model was individually trained on each molecule following the usual approach. (A): Training outcomes. (B): Validation outcomes.

A.3 NETWORK FEATURES AND ABLATION STUDIES

During our experiments, insights into T2E’s performance with different architectures and physical information were gained. We observed no benefit, but a mere overfitting, in increasing the number of hidden parameters while keeping the physical information unchanged. Additionally, we saw no useful benefit of regularization terms such as L_1 and L_2 . As depicted in Figure 8, the addition of extra physical information to a pure MLP layer markedly enhances the model’s accuracy. An in-depth analysis of both training and validation data reveals that the omission of external physical information severely restricts the model’s capabilities. The integration of mass number information notably aids the training dataset and provides stability to validation variation, albeit without enhancing accuracy. Further, adding information about valence electrons to the mass number substantially improves the model’s performance on both training and validation datasets. The inclusion of total electron count information does not augment the model’s performance, presumably because this data is already indirectly embedded within the model considering the elements are neutral. The unique and significant role of valence electrons, determined by orbitals, is underscored in enhancing the model’s performance compared to the total electron count.

Lastly, we observed a significant decline in model precision when injecting physical information into just one layer. Notably, the model’s accuracy showed no substantial enhancement with more than two such injections.

A.4 EXTENDED ARCHITECTURAL DETAILS FOR ORBITAL-INFORMED MLP

Orbital-informed MLP part of T2E is illustrated previously in Figure 3. The initial layer takes 768 features from the embedded text and then reduces them to 64 features. After this reduction, a ReLU activation function is applied, followed by dropout regularization. This process is repeated once more before the addition of physical terms. The feature size is then reduced to 34 with the addition of 2 physical parameters, and the same processing is applied, resulting in 8 features just before the final prediction. At this point, physical information is again incorporated into the model, and the prediction is generated.

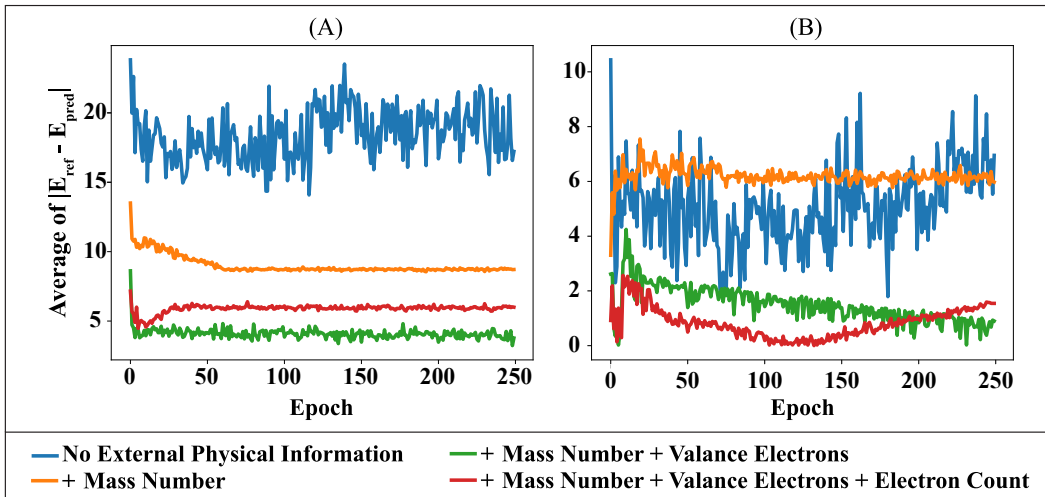


Figure 8: **Train and validation graphs of T2E’s accuracy when different types of physical information are fed into the MLP layer.** The settings were kept consistent with previously given results, with only the injected physical information being altered. Graphs are generated for total energy prediction of B when trained on Li, Be, and C. (A): Training results, (B) Testing results.