Supplementary Information for "A Universal Framework for Featurization of Atomistic Systems"

0.1 Code and Saved Models

Please checkout and install the "MCSH_paper1" branch of $AMPTorch$ to try any of the saved models and test scripts in this study. The code can be found here: https://github.com/ulissigroup/amptorch/tree/MCSH_paper1. The test scripts can be found here: https://github.com/ray38/GMP_AmpTorch_Tests. Tutorials for regenerating all the test models are all included in the repo.

0.2 MD17 Aspirin Examples

0.2.1 Standard Training Procedure

Cutoffs for both the BP scheme and the GMP scheme are set to 10 Å. The neural networks are all trained using the same procedure: $lr = 1e^{-3}$ for 6000 epochs. For the BP vs. GMP comparison example on aspirin MD data, the batch size is chosen to be 256 images. For the force training example, the batch size is chosen to be 32 images.

0.2.2 Behler-Parrinello + HDNN Comparison Test Setups

12 sets of Behler-Parrinello feature are selected, as listed below

Set	G2		G4			$N_{feature}{}^a$
	η	R_s	η	ζ	γ	
1	[0.05, 0.0965, 0.1864, 0.3598, 0.6947, 1.3413 , 2.5897, 5.0]	[0, 1.5]	[0.001, 0.01, 0.03]	[1.0, 2.0, 4.0]	[1.0, -1.0]	156
2	[0.05, 0.0965, 0.1864, 0.3598, 0.6947, 1.3413 , 2.5897, 5.0]	[0, 1.5]	[0.01, 0.03]	[1.0, 4.0]	[1.0, -1.0]	96
3	[0.05, 0.0965, 0.1864, 0.3598, 0.6947, 1.3413 , 2.5897, 5.0]	[0]	[0.01]	[1.0, 4.0]	[1.0, -1.0]	48
4	[0.05, 0.0965, 0.1864, 0.3598, 0.6947, 1.3413 , 2.5897, 5.0]	[0]	[0.001, 0.01, 0.03]	[1.0, 2.0, 4.0]	[1.0, -1.0]	132
5	[0.05, 0.0965, 0.1864, 0.3598, 0.6947, 1.3413 , 2.5897, 5.0]	[0]	[0.01, 0.03]	[1.0, 4.0]	[1.0, -1.0]	72
6	[0.05, 0.0965, 0.1864, 0.3598, 0.6947, 1.3413 , 2.5897, 5.0]	[0, 1.5]	[0.01]	[1.0, 4.0]	[1.0, -1.0]	72
7	[0.05, 0.2324, 1.0772, 5.]	[0, 1.5]	[0.001, 0.01, 0.03]	[1.0, 2.0, 4.0]	[1.0, -1.0]	132
8	[0.05, 0.2324, 1.0772, 5.]	[0, 1.5]	[0.01, 0.03]	[1.0, 4.0]	[1.0, -1.0]	72
9	[0.05, 0.2324, 1.0772, 5.]	[0]	[0.01]	[1.0, 4.0]	[1.0, -1.0]	32
10	[0.05, 0.2324, 1.0772, 5.]	[0]	[0.001, 0.01, 0.03]	[1.0, 2.0, 4.0]	[1.0, -1.0]	120
11	[0.05, 0.2324, 1.0772, 5.]	[0]	[0.01, 0.03]	[1.0, 4.0]	[1.0, -1.0]	60
12	[0.05, 0.2324, 1.0772, 5.]	[0, 1.5]	[0.01]	[1.0, 4.0]	[1.0, -1.0]	48

Table 2: List of tested Behler-Parrinello feature sets. η and R_s for G2 functions are used combinatorially, same as η, ζ and γ for G4 functions. Moreover, there are 3 types of elements (C, H, O) for this dataset. Therefore, feature set 1 has $3(elements) \times 8(\eta) \times 2(R_s) + 6(possible\, element\, pairs) \times 3(\eta) \times 3(\zeta)s \times 2(\gamma) = 156$ features per atom. [a]Number of features per atom.

The list of test results with BP + HDNN models are shown below in Table 3

Set	$N_{feature}{}^a$	MAE train (meV)	MAE test (meV)	Time (ms/image)
1	156	73.0 ± 3.0	98.8 ± 2.7	22.1 ± 0.5
2	96	74.2 ± 1.4	98.8 ± 1.7	13.3 ± 0.4
3	48	71.8 ± 1.9	97.3 ± 1.8	8.2 ± 0.1
4	132	73.5 ± 1.6	100.3 ± 2.5	21.2 ± 0.3
5	72	73.2 ± 1.2	99.6 ± 1.5	12.1 ± 0.1
6	72	74.4 ± 1.6	97.4 ± 2.2	9.1 ± 0.4
7	132	73.3 ± 1.4	99.9 ± 1.5	20.9 ± 0.1
8	72	71.6 ± 1.5	97.8 ± 2.4	12.1 ± 0.1
9	32	70.5 ± 2.2	97.3 ± 1.9	8.1 ± 0.2
10	120	73.5 ± 1.5	101.2 ± 1.4	20.4 ± 0.0
11	60	71.7 ± 2.1	99.3 ± 1.6	11.8 ± 0.1
12	48	72.4 ± 1.7	96.9 ± 1.3	8.2 ± 0.1

Table 3: Performance test results of the tested GMP + HDNN setups. The values are the average values of the 10 trials, and the uncertainties are estimated by their standard deviation. ^{a}Number of features per atom.

0.2.3 GMP+SNN Comparison Test Setups

The probe of GMP has two parts: groups of MCSHs for probing angular features and radial gaussian for probing radial features. In this work, all groups of MCSH up to the indicated order are included, and the number of possible groups for each order are listed below in Table 4:

Order	Number of Possible Groups	Total Number of Groups up to This Order
0	1	1
1	1	2
2	2	4
3	3	7
4	4	11
5	5	16
6	7	23
7	8	31
8	10	41

Table 4: List of possible groups of MCSHs for each order. Note that the number of possible groups

We combine the radial probes with the lists of radial Gaussians combinatorially to obtain the full list of probes/features. The list of widths (standard deviations) of the Gaussians used in this test are manually picked and listed below in Table 5:

Number of Gaussians	List of Sigmas
1	[0.25]
2	[0.25, 2.0]
3	[0.25, 1.0, 2.0]
4	[0.25, 0.75, 1.5, 2.0]
5	[0.25, 0.5, 1.0, 1.5, 2.0]
6	[0.25, 0.5, 0.75, 1.0, 1.5, 2.0]

Table 5: List of Sigmas used in Test 1.

Therefore, when there are 5 Gaussians with MCSH up to order 6, there are $23 \times 5 = 115$ descriptors per atom. The complete list of test results with GMP+SNN is give in Table 6.

Num. Gaussians[a]	MCSH order[b]	$N_{feature}$[c]	MAE train (meV)	MAE test (meV)	Time (ms/image)
1	0	1	180.1 ± 0.5	180.7 ± 0.9	2.7 ± 0.2
	1	2	146.5 ± 0.9	146.9 ± 1.1	2.8 ± 0.2
	2	4	133.9 ± 0.9	134.6 ± 1.1	3.0 ± 0.2
	3	7	116.8 ± 1.3	119.5 ± 1.6	3.1 ± 0.2
	4	11	114.7 ± 1.6	117.2 ± 1.3	3.4 ± 0.2
	5	16	104.1 ± 2.1	107.8 ± 1.7	3.8 ± 0.2
	6	23	103.2 ± 1.1	107.1 ± 1.3	4.1 ± 0.1
	7	31	95.2 ± 1.1	99.4 ± 1.6	4.8 ± 0.2
2	0	2	162.4 ± 2.1	163.5 ± 2.1	2.7 ± 0.1
	1	4	117.5 ± 1.5	120.5 ± 1.7	2.9 ± 0.2
	2	8	100.0 ± 0.8	105.9 ± 1.5	3.2 ± 0.2
	3	14	88.3 ± 1.9	98.1 ± 1.4	3.5 ± 0.2
	4	22	87.1 ± 1.8	97.7 ± 2.0	3.9 ± 0.1
	5	32	84.3 ± 0.9	97.0 ± 1.2	4.7 ± 0.2
	6	46	80.7 ± 1.0	94.9 ± 1.2	5.9 ± 0.4
	7	62	74.8 ± 1.2	88.1 ± 1.2	7.1 ± 0.3
3	0	3	142.7 ± 2.1	144.8 ± 1.6	2.7 ± 0.2
	1	6	89.2 ± 1.0	92.5 ± 0.9	3.0 ± 0.4
	2	12	78.2 ± 0.8	84.1 ± 0.7	3.3 ± 0.2
	3	21	74.1 ± 0.9	83.8 ± 1.4	4.0 ± 0.3
	4	33	72.3 ± 0.8	83.6 ± 0.9	4.9 ± 0.3
	5	48	67.9 ± 1.4	81.1 ± 2.2	6.0 ± 0.3
	6	69	65.5 ± 1.6	80.0 ± 1.7	7.7 ± 0.2
	7	93	60.4 ± 0.8	76.3 ± 1.7	9.8 ± 0.4
4	0	4	120.3 ± 1.7	121.6 ± 1.6	2.9 ± 0.2
	1	8	77.2 ± 1.3	79.8 ± 1.0	3.0 ± 0.2
	2	16	69.9 ± 0.9	75.3 ± 0.9	3.4 ± 0.0
	3	28	66.2 ± 2.2	75.3 ± 1.8	4.4 ± 0.2
	4	44	64.5 ± 0.9	75.8 ± 1.6	5.6 ± 0.3
	5	64	61.8 ± 0.6	75.2 ± 1.0	7.1 ± 0.4
	6	92	57.8 ± 1.0	73.0 ± 1.2	9.5 ± 0.3
	7	124	54.5 ± 0.7	70.8 ± 1.2	11.8 ± 0.5
5	0	5	92.3 ± 1.3	93.3 ± 1.3	2.9 ± 0.1
	1	10	62.6 ± 1.1	65.1 ± 1.0	3.1 ± 0.2
	2	20	62.0 ± 1.9	66.3 ± 1.8	3.6 ± 0.0
	3	35	59.4 ± 1.4	67.1 ± 1.6	4.7 ± 0.1
	4	55	56.1 ± 1.1	65.3 ± 1.1	6.4 ± 0.3
	5	80	50.9 ± 0.5	61.5 ± 0.9	8.3 ± 0.3
	6	115	47.3 ± 0.8	60.1 ± 1.1	10.8 ± 0.3
	7	155	43.8 ± 0.8	57.6 ± 1.4	14.0 ± 0.4
6	0	6	85.3 ± 1.8	86.5 ± 1.8	2.9 ± 0.2
	1	12	60.6 ± 1.5	62.9 ± 1.3	3.5 ± 0.2
	2	24	58.3 ± 1.2	62.5 ± 1.4	4.2 ± 0.3
	3	42	54.6 ± 1.0	61.5 ± 0.7	5.3 ± 0.3
	4	66	51.7 ± 0.8	60.7 ± 1.4	7.0 ± 0.4
	5	96	47.3 ± 0.9	58.7 ± 1.2	9.1 ± 0.4
	6	138	44.0 ± 0.7	56.4 ± 0.9	12.2 ± 0.6
	7	186	41.3 ± 0.9	55.7 ± 0.9	15.8 ± 0.6

Table 6: Performance test results of the full sets of tested GMP + SNN setups. The values are the average values of the 10 trials, and the uncertainties are estimated by their standard deviation. [a]Number of possible Gaussian functions used to construct the descriptor probes. [b]The highest MCSH order used to construct the probes. For example, when highest order is 2, that means all groups from MCSH of order 0, 1 and 2 are used to construct the probes. [c]Number of features per atom.

Model	Sigmas	$N_{feature}$
$GMP(2,4) + SNN(50,3)$	[0.25, 0.75, 1.5, 2.0]	16
$GMP(3,12) + SNN(50,3)$	[0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1.0, 1.25, 1.5, 1.75, 2.0]	84

Table 7: Setups for the GMP+SNN models used in the force training example

0.2.4 Force Training Example

The sigmas of the radial probe Gaussians are listed in Table 7

0.3 QM9 Example

0.3.1 Per-element Bias

A per-element bias is added to the SNN model to improve performance. Conceptually, this is equivalent to fitting to formation energies rather than absolute energies. The total energy of an atom is the model predicted energy plus the per-element bias of the specific atom type. To determine the bias, a linear model is applied. The number of atoms for each element types are counted for all the images in the training set, and they are the independent variable. The corresponding energies for each system is the dependent variable. For example, the per-element bias of the trials with 100K training images is shown below in Table 8:

Atom Type	Per-element Bias (meV)
H	-2795.2721
C	-6217.7719
N	-4552.4431
O	-4432.6761
F	-4075.4931

Table 8: Per-element Bias of each atom type found by the linear model, for the 100K training set

0.3.2 Training Procedure

With the per-element bias determined, GMP+SNN models are fitted to the atomization energy minus the per-element biases. The model setups are given in Table 9. The cutoff distance is always 15 Å, so that the largest radial probe takes a negligible value of $8 \times 10^{-}4$ at the cutoff. The models are trained for 12,000 epochs with learning rate decrease by factor of 2 every 2,000 epochs, from 1^{-2} to 3^{-4}. The batch size is set to be 32 images.

0.3.3 Transfer Learning to New Element

For this example, we used the same procedure as above, with the caveat that the per-element biases are not fitted using a linear model, but directly pulled from the 100k molecule trial. For more detail please refer to the test scripts.

Shown in Figure 6 is the error distribution of the system containing F atoms in the basis of error per F atom. Shown in Figure 7 are the error distributions from different trials with different random seeds.

Model	Sigmas	$N_{feature}$	$N_{parameters}$
$GMP(1,8)+SNN(15,5)$	[0.02,0.2,0.4,0.69,1.1,1.66,2.66,4.4]	16	1111
$GMP(2,8)+SNN(25,5)$	[0.02,0.2,0.4,0.69,1.1,1.66,2.66,4.4]	84	3001
$GMP(2,13)+SNN(60,5)$	[0.02,0.12,0.24,0.36,0.5,0.69,0.92,1.2,1.52,2.0, 2.66,3.5,5.0]	52	14701
$GMP(2,37)+SNN(100,5)$	[0.02,0.05,0.08,0.12,0.16,0.2,0.24,0.28,0.32,0.36, 0.4,0.45,0.5,0.56,0.62,0.69,0.76,0.84,0.92,1.01, 1.1,1.2,1.3,1.4,1.52,1.66,1.82,2.0,2.42,2.66, 2.92,3.2,3.5,3.9,4.4,5.0]	148	45701

Table 9: Setups for the GMP+SNN models used in the QM9 examples. Cutoff distance is always 15 Å. $N_{parameters}$ is the number of trainable parameters of the neural network model.

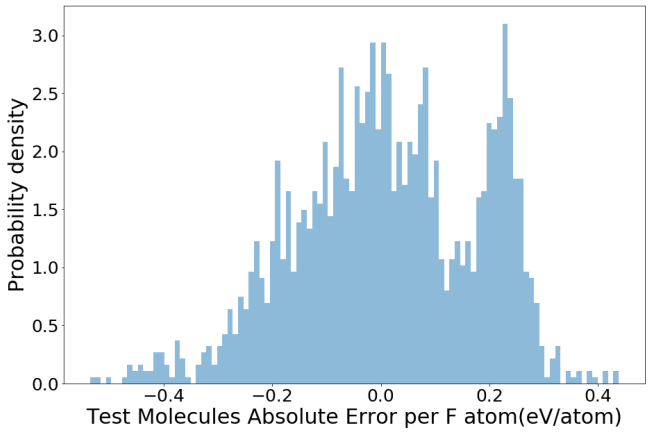

Figure 6: Error distribution of F-containing molecules per F atom.

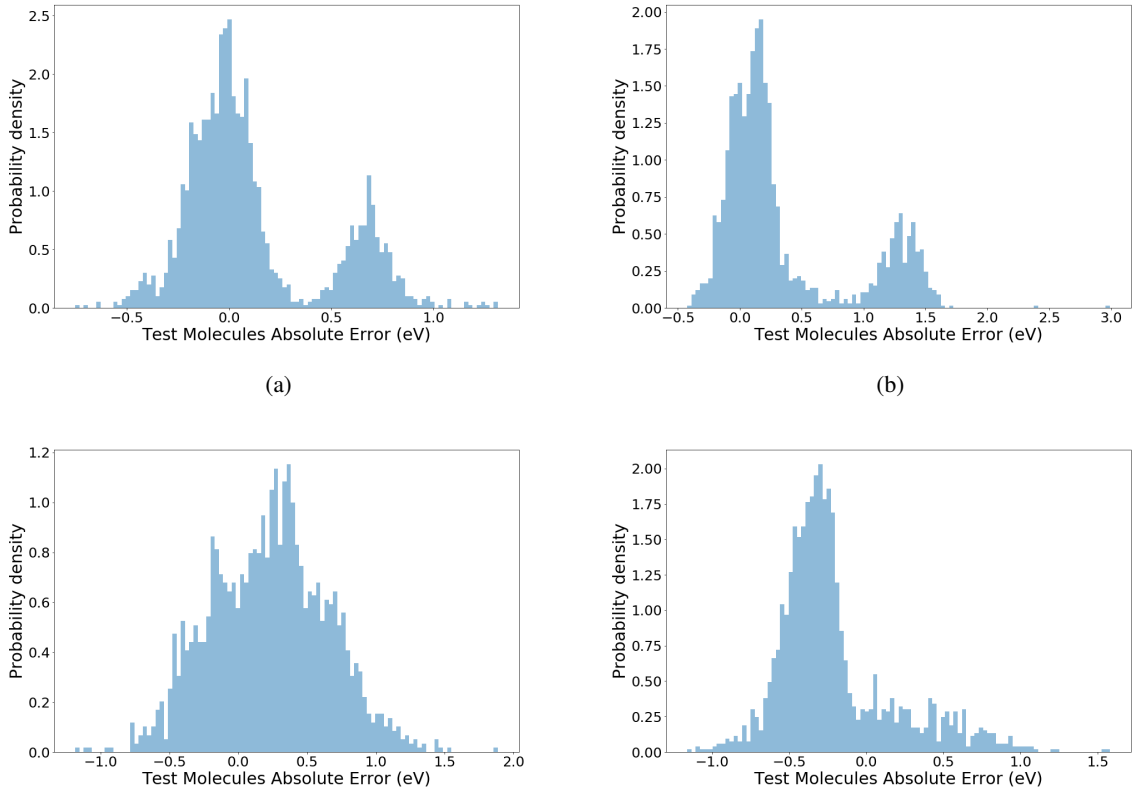

Figure 7: Error distribution of prediction for F-containing molecules with different random seed for the model trained on molecules without F.

Model	Sigmas	$N_{feature}$	$N_{parameters}$
GMP(3,13)+SNN(80,6)	[0.02,0.16,0.32,0.5,0.76,1.1,1.52,2.2,3.2,5.0]	91	34161
GMP(4,19)+SNN(150,7)	[0.02,0.08,0.16,0.24,0.32,0.4,0.5,0.62,0.76,0.92,1.1, 1.3,1.52,1.82,2.2,2.66,3.2,3.9,5.0]	209	146701
GMP(5,37)+SNN(250,8)	[0.02,0.05,0.08,0.12,0.16,0.2,0.24,0.28,0.32,0.36, 0.4,0.45,0.5,0.56,0.62,0.69,0.76,0.84,0.92,1.01, 1.1,1.2,1.3,1.4,1.52,1.66,1.82,2.0,2.42,2.66, 2.92,3.2,3.5,3.9,4.4,5.0]	592	528501

Table 10: Setups for the GMP+SNN models used in the OC20 examples. Cutoff distance is always 15 Å. $N_{parameters}$ is the number of trainable parameters (weights) of the neural network model.