Supplementary Material

Xiangyun Lei, Andrew J. Medford

February 4, 2021

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: here:https://github.com/ray38/GMP_AmpTorch_Tests. Tutorials for regenerating all the test models are all included in the repo.

2 MD17 Aspirin Examples

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.

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 1: 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 2

Set	$N_{feature}{}^{a}$	MAE train (meV)	MAE test (meV)	Time (ms/image)
1	156	59.9 ± 0.7	87.5 ± 0.7	19.2 ± 0.3
2	96	58.3 ± 0.8	83.5 ± 1.0	11.6 ± 0.2
3	48	59.9 ± 0.6	86.0 ± 1.2	7.8 ± 0.4
4	132	59.6 ± 1.0	88.1 ± 1.9	18.8 ± 0.4
5	72	58.4 ± 0.5	86.6 ± 1.1	11.1 ± 0.2
6	72	60.1 ± 1.5	82.7 ± 2.0	8.0 ± 0.3
7	132	59.6 ± 0.8	88.0 ± 1.3	18.7 ± 0.1
8	72	57.6 ± 0.6	84.3 ± 0.7	11.1 ± 0.3
9	32	59.2 ± 1.0	85.5 ± 1.5	7.5 ± 0.3
10	120	59.4 ± 1.0	88.1 ± 1.6	18.5 ± 0.3
11	60	58.7 ± 0.6	87.3 ± 1.2	10.9 ± 0.3
12	48	58.0 ± 0.7	82.2 ± 1.2	7.6 ± 0.2

Table 2: 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.

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 3:

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 3: 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 4:

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 4: 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 5.

Num. Gaussians[a]	MCSH order[b]	$N_{feature}$[c]	MAE train (meV)	MAE test (meV)	Time (ms/image)
1	0	1	175.7 ± 2.9	176.1 ± 2.2	2.6 ± 0.1
	1	2	146.2 ± 0.6	146.5 ± 1.1	2.7 ± 0.1
	2	4	134.2 ± 3.5	135.2 ± 3.9	2.8 ± 0.2
	3	7	114.1 ± 1.0	117.8 ± 1.1	3.0 ± 0.1
	4	11	109.0 ± 1.0	112.8 ± 1.0	3.2 ± 0.1
	5	16	92.4 ± 1.1	97.5 ± 1.3	3.5 ± 0.1
	6	23	88.8 ± 1.2	94.3 ± 1.4	4.0 ± 0.1
	7	31	78.0 ± 1.0	84.7 ± 1.4	4.5 ± 0.1
2	0	2	158.7 ± 0.4	160.1 ± 1.2	2.7 ± 0.1
	1	4	116.1 ± 0.9	118.6 ± 1.0	2.8 ± 0.1
	2	8	94.6 ± 1.4	100.6 ± 1.4	3.0 ± 0.1
	3	14	77.2 ± 0.9	86.3 ± 1.3	3.4 ± 0.1
	4	22	73.4 ± 0.9	84.3 ± 1.2	3.8 ± 0.1
	5	32	68.4 ± 1.3	80.7 ± 1.8	4.5 ± 0.1
	6	46	66.5 ± 0.9	79.6 ± 1.8	5.4 ± 0.1
	7	62	58.4 ± 0.6	71.2 ± 1.1	6.6 ± 0.1
3	0	3	140.0 ± 0.6	141.8 ± 0.8	2.8 ± 0.1
	1	6	83.4 ± 1.4	86.3 ± 1.6	2.9 ± 0.1
	2	12	68.0 ± 0.7	74.2 ± 0.7	3.3 ± 0.1
	3	21	60.5 ± 0.4	69.7 ± 0.5	3.8 ± 0.1
	4	33	57.3 ± 0.7	67.9 ± 1.0	4.4 ± 0.1
	5	48	52.8 ± 0.7	64.4 ± 1.3	5.6 ± 0.3
	6	69	50.2 ± 0.7	62.5 ± 1.0	6.8 ± 0.2
	7	93	45.4 ± 0.6	57.8 ± 1.1	8.5 ± 0.1
4	0	4	116.3 ± 0.9	117.9 ± 1.0	2.8 ± 0.1
	1	8	68.3 ± 0.9	71.1 ± 1.0	3.0 ± 0.1
	2	16	59.5 ± 0.6	65.5 ± 0.7	3.4 ± 0.1
	3	28	50.6 ± 0.7	58.5 ± 0.8	4.0 ± 0.1
	4	44	48.6 ± 0.8	57.7 ± 1.0	5.1 ± 0.2
	5	64	45.0 ± 0.6	55.5 ± 1.2	6.3 ± 0.2
	6	92	42.5 ± 0.6	54.0 ± 0.9	8.1 ± 0.1
	7	124	40.7 ± 0.5	53.7 ± 0.7	10.3 ± 0.1
5	0	5	83.5 ± 0.8	85.2 ± 0.5	2.7 ± 0.1
	1	10	54.4 ± 0.5	56.6 ± 0.6	3.0 ± 0.1
	2	20	46.9 ± 0.7	51.1 ± 0.7	3.5 ± 0.1
	3	35	44.2 ± 0.5	50.6 ± 0.7	4.5 ± 0.1
	4	55	40.8 ± 0.5	48.8 ± 0.9	5.7 ± 0.2
	5	80	37.8 ± 0.5	47.3 ± 0.8	7.4 ± 0.2
	6	115	35.5 ± 0.6	45.6 ± 0.8	9.6 ± 0.2
	7	155	32.6 ± 0.5	43.2 ± 0.9	12.3 ± 0.2
6	0	6	71.7 ± 0.5	73.0 ± 0.5	2.9 ± 0.0
	1	12	51.2 ± 0.6	53.6 ± 0.7	3.2 ± 0.1
	2	24	45.5 ± 0.5	50.1 ± 0.9	3.8 ± 0.1
	3	42	38.8 ± 0.5	44.9 ± 0.6	4.9 ± 0.1
	4	66	36.8 ± 0.7	44.2 ± 0.9	6.4 ± 0.2
	5	96	34.1 ± 0.6	42.9 ± 0.9	8.3 ± 0.2
	6	138	32.6 ± 0.5	42.4 ± 0.8	10.8 ± 0.2
	7	186	30.6 ± 0.5	41.6 ± 0.9	14.2 ± 0.2

Table 5: 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 6: Setups for the GMP+SNN models used in the force training example

2.4 Force Training Example

The sigmas of the radial probe Gaussians are listed in Table 6

3 QM9 Example

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 shown below in Table 7:

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

Table 7: Per-element Bias of each atom type found by the linear model, for the 100K training set

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 8. 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.

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 linear model, but directly pulled from the 100k molecule trial. For more detail please refer to the test scripts.

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 8: 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.

4 OCP S2EF Dataset Setups

The per-element biases are determined by the 200K training set, and used for all trials. The biases are listed in Table 9. The model setups are given in Table 10. The models were trained for roughly 10,000 epochs with learning rate decrease from 5^{-3} to 2^{-4}, and batch size of 32 images.

Type	Per-element Bias (eV)	Type	Per-element Bias (eV)	Type	Per-element Bias (eV)
H	-3.459224543285167	B	-6.131687348061676	C	-8.409056118016434
N	-8.482527946982596	O	-6.625329843626161	Na	-1.6272161266181353
Al	-3.685546100708493	Si	-5.366115554333206	P	-5.356124619536539
S	-4.701875825379599	Cl	-3.206133340067741	K	-1.6639990985033588
Ca	-2.564389194029974	Sc	-6.857051683285094	Ti	-7.83797523757251
V	-8.469671104728693	Cr	-8.525323765104233	Mn	-7.949821377229276
Fe	-7.29185704318087	Co	-6.339343605830223	Ni	-4.983740855397919
Cu	-3.0599559655471738	Zn	-0.7636980492562433	Ga	-2.645814340187994
Ge	-4.137512388812219	As	-4.578400063883111	Se	-3.9030270048270483
Rb	-1.501116295377563	Sr	-2.4817501066487404	Y	-7.159640206699965
Zr	-8.657410352144826	Nb	-9.692635800596257	Mo	-9.917990868764402
Tc	-9.376118112224386	Ru	-8.487414762933673	Rh	-6.88580158940369
Pd	-4.878716025523049	Ag	-2.002845859469274	Cd	-0.3521934743778097
In	-2.1579404553381734	Sn	-3.489109201477598	Sb	-3.7507685121733054
Te	-3.1467202546325685	Cs	-1.6714210826933673	Hf	-10.015796029262765
Ta	-11.220753704243743	W	-11.738519709357998	Re	-11.149426336972052
Os	-10.243098652486818	Ir	-8.378935731902352	Pt	-5.868231040118616
Au	-2.7985176919629122	Hg	0.12859186837374637	Tl	-1.9054477570253923
Pb	-3.1333889952396787	Bi	-3.430346703339706		

Table 9: Per-element Bias of each atom type found by the linear model, with the 200K training set for the OC20 dataset

Model	Sigmas	$N_{feature}$	$N_{parameters}$
GMP(2,10)+SNN(50,5)	[0.02,0.16,0.32,0.5,0.76,1.1,1.52,2.2,3.2,5.0]	40	10151
GMP(4,19)+SNN(350,5)	[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	445201

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.