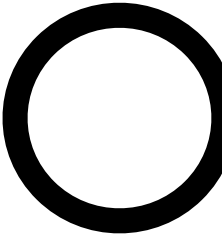
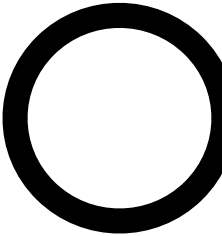
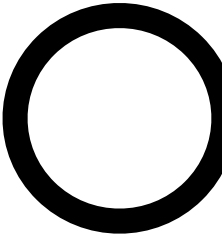
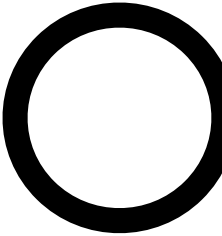
3. E is compared with some ab-initio data (D), in order to estimate the error in the prediction (ε).
   The values of $\mathbf{W}_i$ and $\mathbf{b}_i$ are then **learnt through an optimization algorithm** (e.g., adadelta, rmsprop, …), which minimizes a loss function by cycles of forward propagations, error estimates and weights updates.

Permutation Invariant Polynomials Neural Networks (PIP-NN):







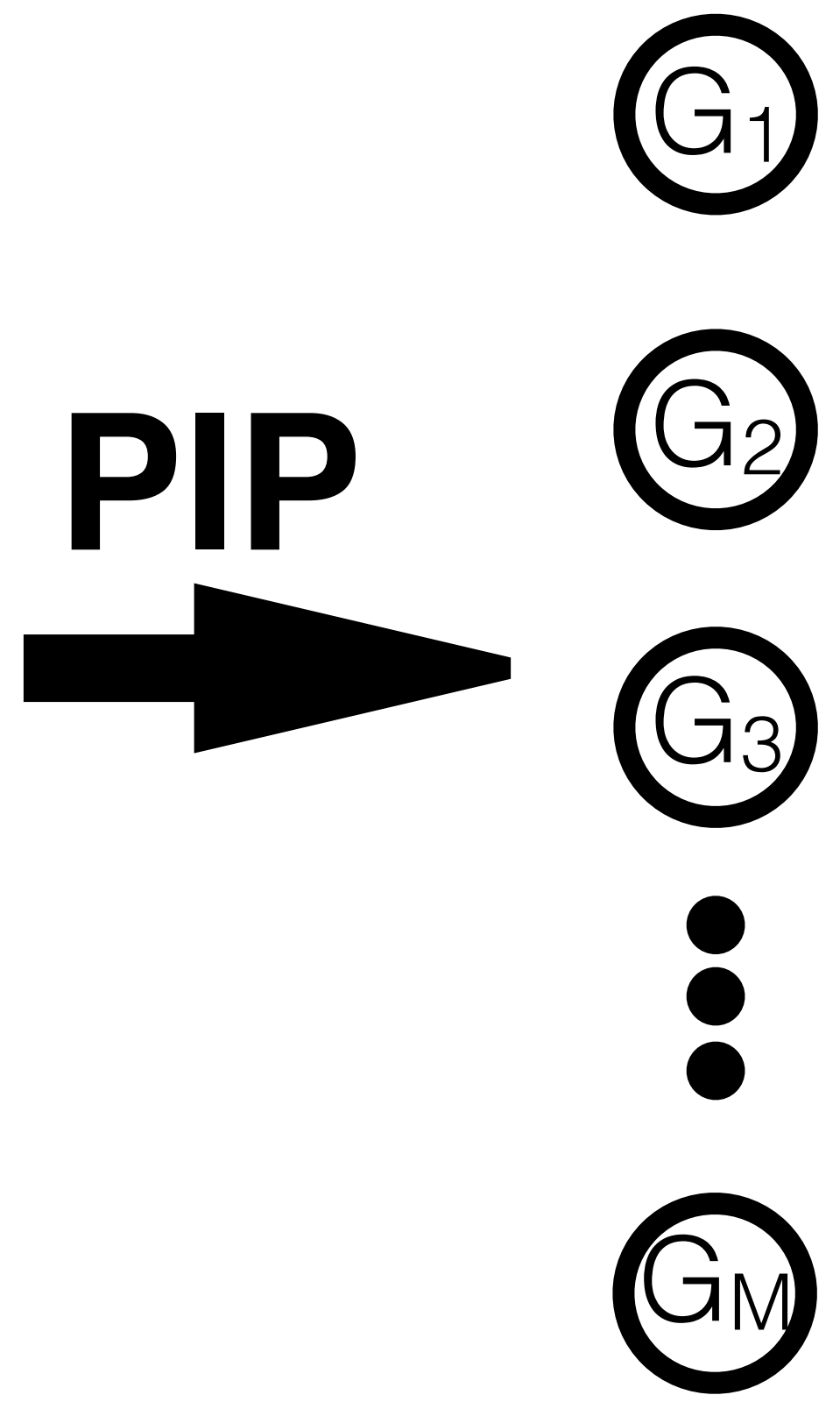


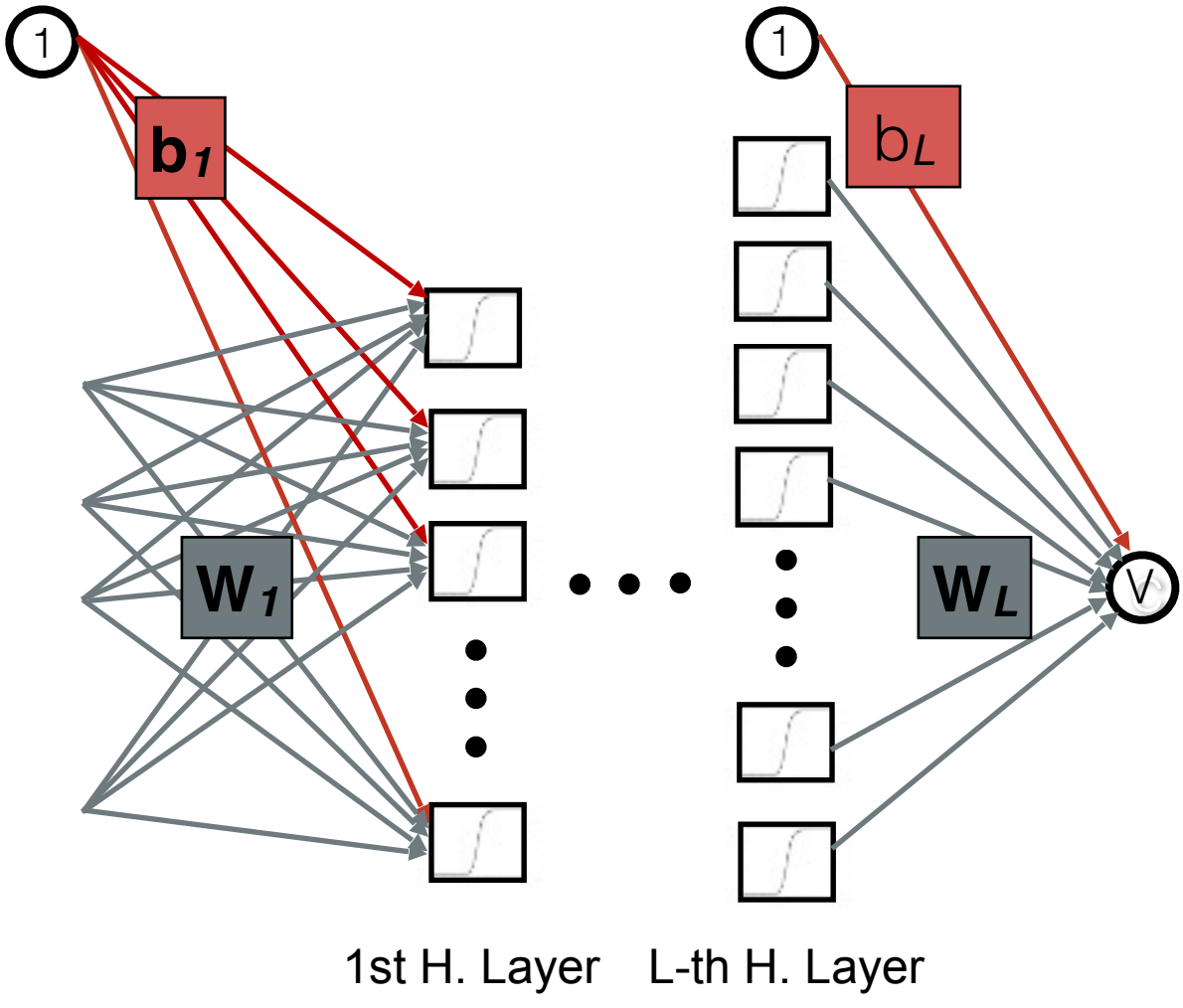

R1

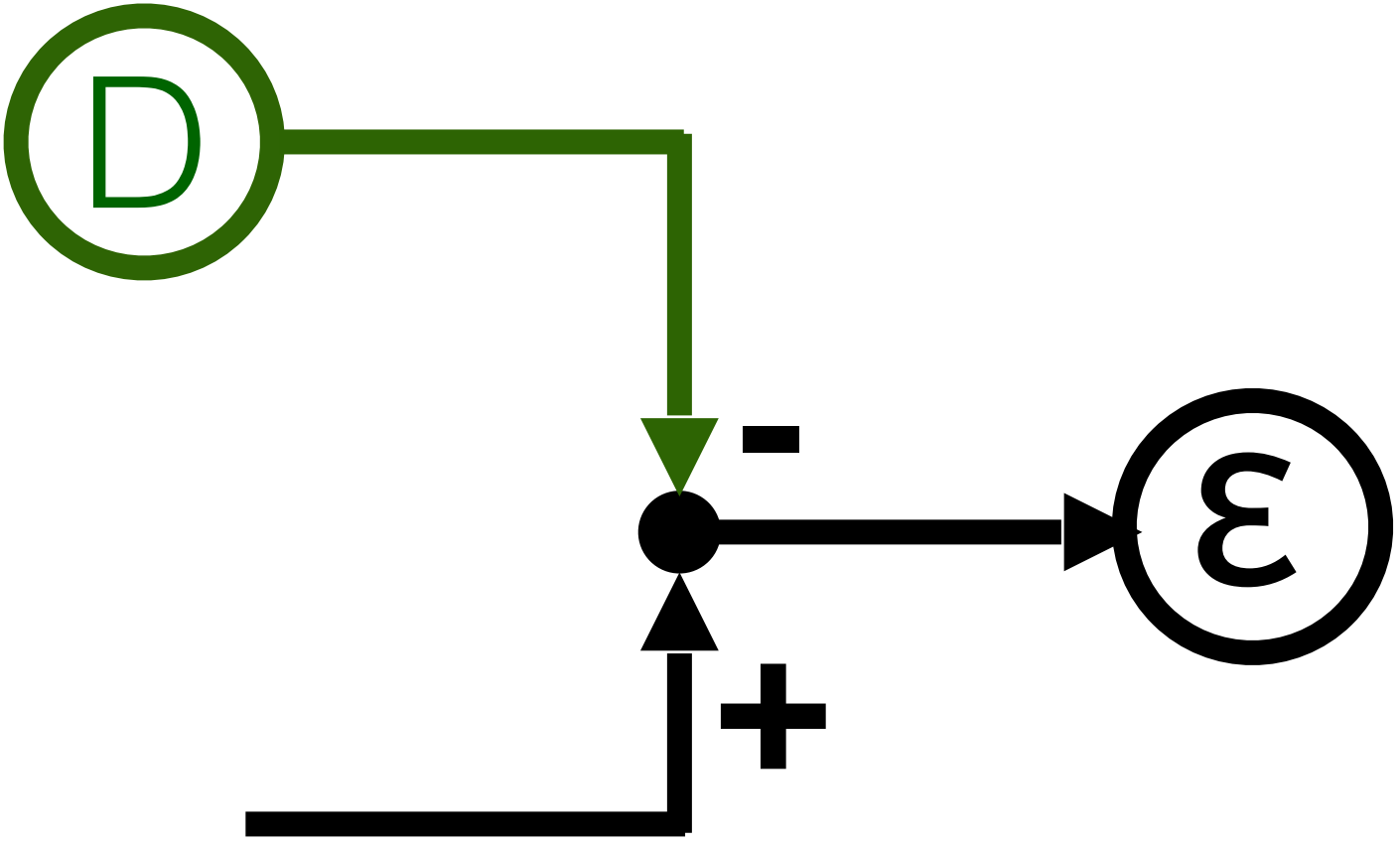

R2


R3


RN


PIP
G1
G2
G3
GM


1
b1
W1
1
bL
WL
V
1st H. Layer
L-th H. Layer


D
-
ε
+

Multi-layer feed-forward Neural Networks (NN) have been adopted **as fitting functional**:

- Easy to implement;
- Easy to train;
- Easy to generalize to new systems;
- Easy to differentiate in **R**;

- Cost effective;
- Easy to be refined;
- Widely tested;
- Easy to be extended to the stochastic case.

15


theano


Lasagne

ANN for PESs: Methodology

ANN for PESs: Methodology

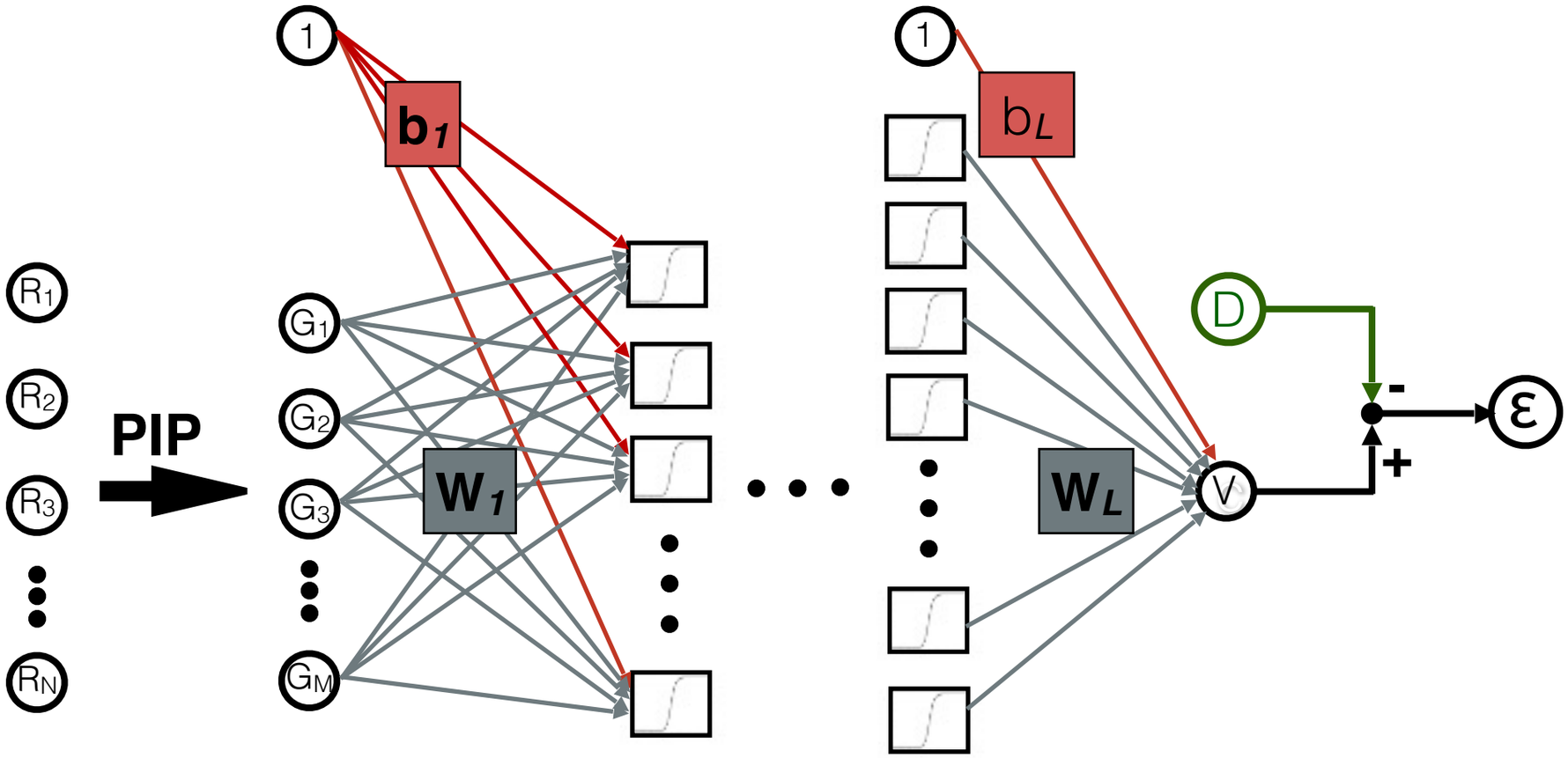
Multi-layer feed-forward Neural Networks (NN) have been adopted **as fitting functional**:

- ✦ Easy to implement;
- ✦ Easy to train;
- ✦ Easy to generalize to new systems;
- ✦ Easy to differentiate in **R**;
- ✦ Cost effective;
- ✦ Easy to be refined;
- ✦ Widely tested;
- ✦ Easy to be extended to the stochastic case.

Permutation Invariant Polynomials Neural Networks (PIP-NN):

theano

Lasagne



3. E is compared with some ab-initio data (D), in order to estimate the error in the prediction (ε).
   The values of $\mathbf{W}_i$ and $\mathbf{b}_i$ are then **learnt through an optimization algorithm** (e.g., adadelta, rmsprop, …), which minimizes a loss function by cycles of forward propagations, error estimates and weights updates.

From Deterministic To Stochastic

In a classic approach to ML, the error is known only at the data points. What can we say about the reliability of the fit at a generic location?

Plus … 3 main concerns about the deterministic PIP-NN deterministic approach …

- ✦ The uncertainty on the data points (Schrödinger Eq. solutions) has not been taken into account;
- ✦ Overfitting Risk;
- ✦ Committee of Neural Networks can produce significantly different results:

No. of fit points N_{pts}	NN		GP	
	1 NN	⟨10 NN⟩	1 GP	⟨10 GP⟩
313	198.00/103.93/87.77	119.11/53.97/43.90	29.09	17.18
625	21.12/12.91/12.03	13.36/7.52/6.53	5.98	3.87
1250	9.29/5.74/4.38	5.74/3.36/2.54	2.17	1.13
2500	4.59/2.43/1.12	2.27/1.23/0.86	1.08	0.62

Fig.1: RMSEs for multiple NN configurations. From: “Neural networks vs Gaussian process regression for representing potential energy surfaces: A comparative study of fit quality and vibrational spectrum accuracy”, J. Chem. Phys. 148, March 2018