Learning Everywhere Resource for BDEC
Compiled by Geoffrey Fox, June 8 2019

General Links
A. BDEC Home Page and set of workshops [1], [2]
B. First BDEC meeting at Bloomington with full links including videos [3]
C. BDEC Bloomington meeting Application Working Group
D. Darpa components of a good demonstrator.
E. MLPerf Consortium [4]
a. MLPerf HPC working group and datasets
   b. MLPerf Deep Time Series working group and Collection of Datasets

Learning Everywhere Projects
A. Deep Learning to Control Plasma Fusion (Bill Tang, Princeton)
a. One slide banner
d. The associated Fusion RNN AI/DL predictive code is open source with a GitHub
   link specified in the paper; and
e. The Gyrokinetic Toroidal Code (GTC) is the exascale-capable fully
   electromagnetic direct numerical simulation (DNS) particle-in-cell (PIC) code that
   will be used to help develop reduced models of pre-disruption classifiers.
   (http://phoenix.ps.uci.edu/GTC/).
f. BDEC Presentation on deep learning for fusion [7]

Learning Everywhere Links
General Discussions
1. Initial Poznan Manifesto for learning everywhere
2. Jeffrey Dean presentation at NIPS 2017 on Machine learning for systems and systems
   for machine learning [8]
3. Microsoft 2018 Faculty Summit presentations on AI for Systems [9], [10]
5. An NSF funded project mainly focused on HPCforML [12], [13]
6. Learning Everywhere papers commenting on Technology, Network Science,
   nanoengineering, biomolecular and computational biology (virtual tissues) with a short
   published paper [14] and longer report [15] plus a presentation at BDEC [16] and at
   IPDPS [17].

Computer Science Issues
7. Papers on Uncertainty Quantification with Machine Learning [18]–[22]
8. Paper on learning index structures for search with machine learning [23]
Experiment Control (MLControl, MLAutotuning)

10. Use of machine learning in experiment control [25]
11. Use of visualization to control smart ensembles of simulation [26]
12. Review of techniques for smart ensembles [27]
13. Use of deep learning in control of material science experiments at light sources [28]

Particle Dynamics

14. Use of Machine Learning to enhance biomolecular simulations [29]; review of different MLandHPC patterns of interest in biomolecular simulations.
15. NIPS 2018 Workshop on machine learning in material science and molecular dynamics [30] with some but not all papers available. There is a follow-up book [31]
16. Review of simulation and early machine learning methods used in macromolecular structure and dynamics [32]
17. Review of machine learning for molecular and materials science [33]
18. MLAutotuningHPC: Use of machine learning to generate an effective Hamiltonian to choose correlated update spins near a critical point [34]
19. MLAutotuningHPC: Use of machine learning to guide molecular dynamics simulations to explore full range of phase space [35]
20. MLAutotuning: Use of Machine Learning to optimize molecular dynamics docking studies [36]
21. MLAutotuning: Efficient exploration of configuration space by adding using machine learning an adaptively computed biasing potential to the original dynamics. [37]–[40]
22. MLAutotuningHPC: Nanoparticle simulations using MLAutotuning to improve performance [41] and MLaroundHPC: learning simulation results [42] (defining surrogates) as a function of defining parameters
23. MLAfterHPC: Software packages for analyzing results of molecular dynamics simulations MDAanalysis [43], [44] and CPPTraj [45], [46]
24. MLAfterHPC: Machine learning to identify crystal structure from molecular dynamics simulations [47]
25. MLAfterHPC: Machine learning (SVM) to analyze structure produced in simulated glassy systems [48]
26. MLAfterHPC: Machine Learning to classify phases of quantum Monte Carlo simulations of the Hubbard model [49]
27. MLAfterHPC: Machine Learning used in materials simulation to aid the design of new materials and to understand properties [50]–[56]
28. MLAfterHPC: Use of Graph Networks to improve prediction of properties of Molecules and Crystals. [57]
29. MLAfterHPC: Using machine learning to analyze results of cosmology simulations to determine parameters of the universe [58] and implemented with Cosmoflow Deep Learning framework [59]
30. MLAfterHPC: Data-mining and manifold learning techniques to detect the slowly varying components of dynamical systems [60]
31. Machine Learning to systematic approach to reduce the dimensionality of a complex molecular system [60]
32. MLaroundHPC: Neural-network representation of DFT potential-energy surfaces [61]–[65]
33. MLaroundHPC: General framework for calculating a many-body coarse-grained potential. [66]
34. MLaroundHPC: Formulate coarse-graining as a supervised machine learning problem and use coarse-graining error and cross-validation to select and compare the performance of different models. [67]
35. MLaroundHPC: Review of the use of neural networks to describe potentials and speed up simulations [68]
36. MLaroundHPC: Use of neural nets to describe potentials and simulation results for Infrared Spectra [69]
37. MLaroundHPC: Deep learning for fast estimate of excitation energy transfer properties (used in solar cells) [70]
38. MLaroundHPC: Machine Learning to learn force calculations during a simulation and avoid repetitive computations [71]
39. MLaroundHPC: Machine Learning used to estimate forces in molecular simulations [38], [72]
40. MLaroundHPC: Investigation of different neural network structures to learn the results of an Ising model simulation near its critical point [73]
41. MLaroundHPC: Machine Learning to give a universal reactive force field that can recover the accuracy of the Schrödinger equation at negligible cost and—because of the locality of the model—leads to an intuitive understanding of the stability and the interactions between molecules. [74]
42. MLaroundHPC: Atom-centered, symmetry-adapted framework to machine-learn the valence charge density based on a small number of reference calculations. [75]
43. MLaroundHPC: The synergies between classical, quantum simulations and ML methods, such as artificial neural networks, have the potential to drastically reshape the way we make predictions in computational structural biology and drug discovery [76]
44. MLaroundHPC: Deep Learning to study compositional and configurational chemical space for molecules of intermediate size. [77]
45. MLaroundHPC: Specifying atom representations for input into machine learning [78]
46. MLaroundHPC: VAMPnets to learn end to end surrogates of molecular dynamics without custom modelling such as collective coordinates (dimension reduction) [79]
47. MLaroundHPC: Use of collective variables (dimension reduction) to study protein dynamics [80]
48. MLaroundHPC: Loop over multiple molecular dynamics and Deep Learning steps to more accurately sample phase for long time computations [81]
49. MLaroundHPC, MLAutotuning: Collective variables to sample molecular dynamics and free energy landscape [82]–[84]
50. MLaroundHPC: Use of machine learning to support long time scale molecular simulations [85]
51. MLaroundHPC: Using Generative Adversarial Networks to produce surrogates of large scale simulations of the effect of gravitational lensing used to study early universe CosmoGAN [86], [87] with supplement [88] on Github

Agent-Based Simulations and Data-driven Approaches to ABM systems

52. MLAutotuning: Smart ensembles for cancer agent based models with PhysiCell. [89]
53. MLAutotuning: Machine-learning and intelligent sampling to build a surrogate meta-model to calibrate agent-based models with data [90]
54. MLAutotuning: Using neural nets to represent cell (agent) behavior based on data for prediction of cancer cell behavior [91]
55. MLAutotuning: Automatic inference of a model of the escape response behavior in a roundworm directly from time series data [92] building on [93], [94]
56. SimulationTrainedML: Using models to fill in gaps in bioinformatics data [95]
57. MLaroundHPC: Using data to predict solutions of complex coupled Agents for metabolic pathway dynamics [96]
58. MLaroundHPC: Use of Deep Learning LSTM to produce surrogates of a one-dimensional biological agent simulation [97] with paper and supplement.
59. MLaroundHPC: Deep Learning for Agent-based Epidemic Forecasting DEFSI [98]
60. MLaroundHPC: Deep Learning RNN and CNN to predict epidemics viewed as time series [99]
61. MLaroundHPC: LSTM Flu epidemic forecasting enhanced by environmental data such as climate [100]

Partial Differential Equations

62. MLaroundHPC: Deep Learning for solving partial differential equations [101], [102] extended to nonlinear systems [103]
63. MLaroundHPC: Deep Learning to find sub-grid processes for Climate prediction [104]
64. MLaroundHPC: Deep Learning to find surrogates for fluid flow simulations [105]
65. MLaroundHPC: Use of equation free modeling [106] for coarse graining combined with manifold learning (dimension reduction) [107]
66. MLaroundHPC: Use of machine learning to improve Extended dynamic mode decomposition for representing Koopman Operator to represent dynamical systems [108]
67. MLaroundHPC: Solving high dimensional (up to 1000’s) partial differential equations using deep learning surrogates [109], [110]
68. MLaroundHPC: Machine Learning surrogates of heart simulations to speed up aortic aneurysm studies [111]

Categories used Above

- **HPCforML**: Using HPC to execute and enhance ML performance, or using HPC simulations to train ML algorithms (theory guided machine learning), which are then used to understand experimental data or simulations.
- **MLforHPC**: Using ML to enhance HPC applications and systems

We further subdivide **HPCforML** as

- **HPCrunsML**: Using HPC to execute ML with high performance
- **SimulationTrainedML**: Using HPC simulations to train ML algorithms, which are then used to understand experimental data or simulations.

We also subdivide **MLforHPC** as

- **MLAutotuning**: Using ML to configure (autotune) ML or HPC simulations.
- **MLafterHPC**: ML analyzing results of HPC as in trajectory analysis and structure identification in biomolecular simulations
- **MLaroundHPC**: Using ML to learn from simulations and produce learned surrogates for the simulations. The same ML wrapper can also learn configurations as well as results.
This differs from SimulationTrainedML as the latter is typically using learned network to predict observation whereas in MLaroundHPC we are using the ML to improve the HPC performance.

- **MLControl**: Using simulations (with HPC) in control of experiments and in objective driven computational campaigns. Here the simulation surrogates are very valuable to allow real-time predictions.

### Citations


