

# Scaling FLOP/SYNC-intensive HP Data Analytics (HPDA)

1. Need for end-to-end HPDA for CS&E
2. HPDA can be FLOPS/SYNC-intensive
3. Challenges:
  - algorithms
  - productivity

# 1. End-to-end HPDA in CS&E

TRADITIONAL

Check-pointing

Sampling

Visualization

Data-assimilation

Trajectory analysis

3D/time correlations

END-TO-END

Uncertainty quantification

Coupled sampling

Adjoint-based assimilation

Pattern recognition

Model reduction

Streaming

## 2. FLOPS/SYNC-intensive HPDA

### METHODS

Nearest-neighbors

Kernel methods

Logistic regression

Support vectors

Graphical models

Deep learning

### ALGORITHMS

Linear algebra

Optimization

Geometry

Sampling

Indexing/searching

Graphs/Trees

now: hadoop, async stochastic gradient,

but: jacobi vs N-body

# N-body methods

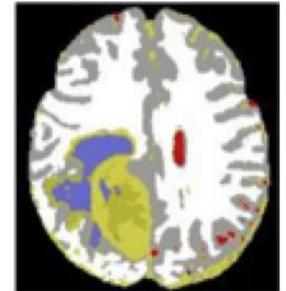
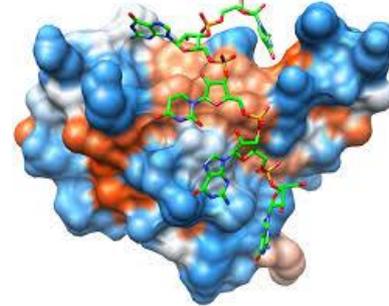
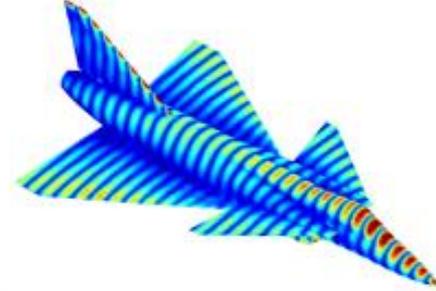
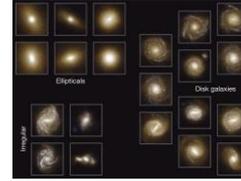
$$O(N^2) \rightarrow O(N)$$

Gravity & Coulomb  
Waves & Scattering  
Fluids & Transport

Graphics  
Machine learning  
Kriging  
Image analysis

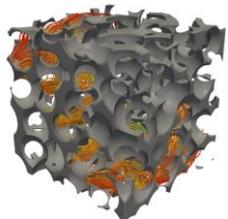
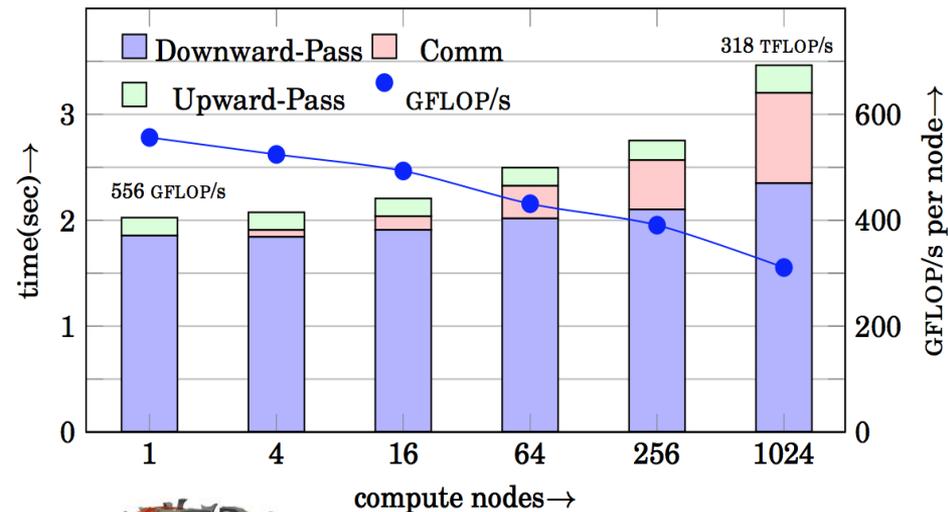
simulation

data analysis

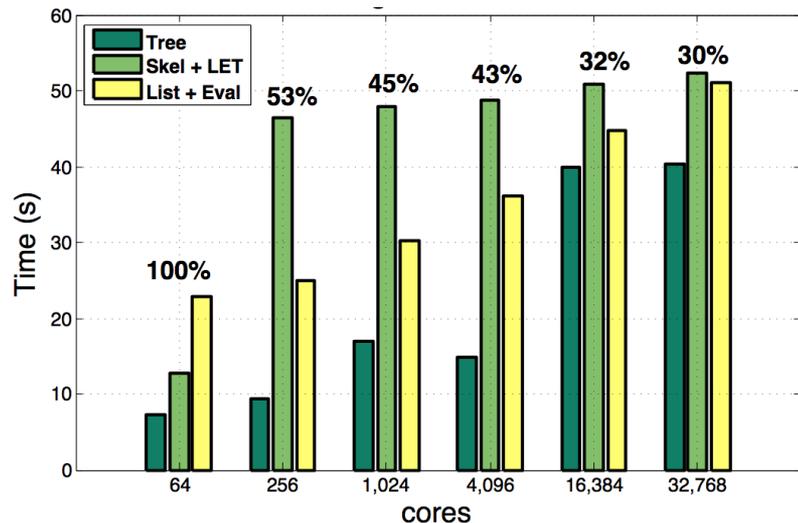


# Scaling TACC's STAMPEDE 16 sandy bridge/ node

## FLUIDS 12B / 3D ~300GB



## CLASSIFICATION 1B / 128D ~500GB



0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9

# 3. Challenges productivity/reproducibility/performance gaps

Next generation HPDA algorithms

Large evolving design space

Expanding complexity of

Algorithms / APIs / Hardware

No parallel machine model

Scheduling / Streaming

End-to-end scalability

C++/MPI+X/BLAS/PETSc VS Java/Hadoop/SQL/SPARK

## MPI+X

Open{MP,CL,ACC}

Pthreads, TBB

CUDA/SSE/AVX

PREFETCHING

NVRAMS

FPGAS