Scaling Resiliency and Analysis Via Machine Learning

Alok Choudhary
Northwestern University

What if a Resilience and Checkpointing Solution Provided

- Improved Resilience via more frequent relevant checkpoints, while
- Reducing the amount of data to be stored by an order of magnitude, and
- Guaranteeing user-specified tolerable maximum error rate for each data point, and
- an order of magnitude smaller mean error for each data set, and
- reduced I/O time by an order of magnitude, and
- Enabling faster restart and faster convergence after restart, while
- Providing data for effective analysis and visualization
Simulation Represents a State Transition Model - What if we analyze the Change in Value?

Observations:
• Variable Values – distribution
• Change in Variable Value – distribution
• Relative Change in Variable Value - distribution

1. Relative change is more predictable.
2. The relative changes in variable values can be learned (ML) and represented in a much smaller state space (compressions).
3. Anomalies are preserved.

A1(t) = 100, A1(t+1) = 110 => change = 10, rel change = 10%
A2(t) = 5, A2(t+1) = 5.5 => change = .5, rel change = 10%

\[ \Delta D_{i,j} = \frac{D_{i,j} - D_{i-1,j}}{D_{i-1,j}} \]
"Incompressible" with Lossy Encoding

NUMARCK Overview

Machine learning based Approach

Transform the data by computing relative changes in ratio from one iteration to the next

Learn the distribution of relative change r using machine learning algorithms and store approximated values

Full checkpoint

Forward Predictive Coding

Data Approximation

C: change ratios

F: Full checkpoint

Traditional checkpointing

Original rlds data

• Highly random

Bspline reconstructed rlds data

• Extreme events missed

~0.35 correlation!

~0.99 correlation!

0.001 RMSE

Distribution Learning

Forward coding

Full checkpointing

Machine learning based Approach

Transform the data by computing relative changes in ratio from one iteration to the next

Learn the distribution of relative change r using machine learning algorithms and store approximated values

F: Full checkpoint

C: change ratios
Examples

FLASH dataset, 0.1% error rate

Different Tolerable Error Rates: Incompressible Ratio (0.1% - 0.5%)

Mean Error Rate

FLASH dataset, 0.1% error rate

 restart from 1 dens
 restart from 1 pres
 restart from 1 temp
 restart from 2 dens
 restart from 2 pres
 restart from 2 temp
 restart from 3 dens
 restart from 3 pres
 restart from 3 temp