

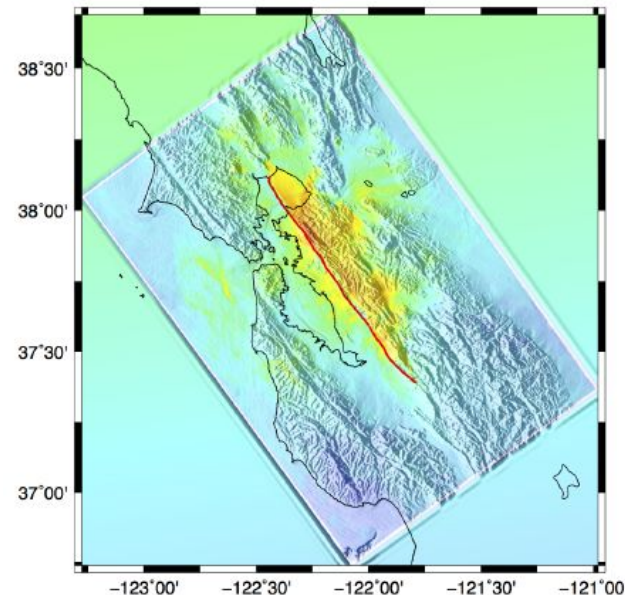
Transitioning an Earthquake Simulation Application to Using HDF5

Houjun Tang

Lawrence Berkeley National Lab

ECP-EQSIM

- High-Performance, Multidisciplinary Simulation for Regional-Scale Earthquake Hazard and Risk Assessments
- Provide the first **strong coupling** and **linkage** between simulations of earthquake *hazards* (ground motions) and *risk* (structural system demands).
- **SW4**, main code to simulate seismic wave propagation.



SW4 Existing I/O

- **Input**

- Material model and topography: `rfile`, custom binary format.
- Forcing function: `SRF`, ASCII format.
- Locations for output data: input file, ASCII format.

- **Output**

- Time-series data: `USGS` ASCII format, or `SAC` binary format.
 - × A large number of small files (**10k+**, a few **MB** each).
- Image: custom binary format.
 - × Multiple small files.
- Checkpoint: custom binary format.
 - × One single file, ~40 TB.

SW4 New HDF5 Implementation - Input

- Material model and topography: **sfile** (w/ new curvilinear model)
- Forcing function: **SRF-HDF5**.
- Locations for output data: **SAC-HDF5**.
- **HDF5 implementation reduces**
 - *Storage space*
 - ✓ **sfile** is **1/2** the size of **rfile**
 - ✓ **SRF-HDF5** is **1/3** of **SRF**.
 - *Read time*
 - ✓ **sfile** is **2.7x** faster than **rfile** with 12,288 MPI ranks
 - ✓ **SRF-HDF5** TBD.
 - *Human error*
 - ✓ No need to keep a long list (>1k) of locations in a text file.

SW4 New HDF5 Implementation - Output

- Time-series data: **SAC-HDF5**.
- Image: *TBD*
- Checkpoint: *TBD*
- **HDF5 implementation reduces**
 - *Number of files*
 - ✓ A single `SAC-HDF5` file compared with >10k files
 - ✓ Better portability, easy to share, and access.
 - *Storage space*
 - ✓ `SAC-HDF5` is **1/5** the size of `USGS(ASCII)`, ~ same size of `SAC(binary)`.
 - *Write time*
 - ✓ Small scale results are good, large scale *TBD*.

Thanks!

Questions?

