WASP: Intelligent Storage for Gridded Numerical Data

Lawrence Frank, John Clyne, Tom Lesperance, Alan Norton, and Scot Pearse

Problem

High performance computing technologies enable numerical modeling of a broad gamut of scientific phenomena, resulting in the production of vast quantities of numerical data. Concurrently, advances in modern digital imaging methods are revolutionizing a wide range of scientific disciplines. This is particularly true in the biological and geosciences, two seemingly very different disciplines. But these advanced capabilities come at a cost: Increasing data size and complexity require more sophisticated methods for data management, analysis and visualization.

The NSF funded **Wavelet-enabled progressive data Access and Storage Protocol (WASP)** project will provide a common software framework for supporting a multi-scale progressive data refinement method based upon the representation of the data as a wavelet expansion[1], and enabling interactive exploration of large data sets for the bio- and geoscience communities.

Progressive Data Access

Efficient management of large data for visualization, and storage supporting scientific discovery is an increasingly non-trivial problem as the size and complexity of imaging and simulation data increases. This project addresses a critical technical challenge that limits the explorative capabilities and the utility of large data: the implementation of a general progressive access, multi-scale data representation for the efficient handling of large, complex, multi-variate data sets.

Progressive data access is the enabling technology behind ubiquitous digital mapping technologies such as Google Maps™; coarsened approximations of high resolution maps are presented when the viewpoint is far away. As the viewer moves closer to an area of interest more and more detail is revealed. This project will develop and release an open source C/C++ WASP software toolkit that affords structured scientific data sets, and cDe Moivre'sorresponding applications, similar, yet more sophisticated and powerful, data access capabilities as those employed by consumer mapping display technologies.

- Lawrence Frank (lfrank@ucsd.edu) and Tom Lesperance are with the Center for Scientific Computation in Imaging, UCSD.

  http://csci.ucsd.edu

- John Clyne (clyne@ucar.edu), Alan Norton, and Scott Pearse are with the Computational and Information Systems Lab., NCAR.

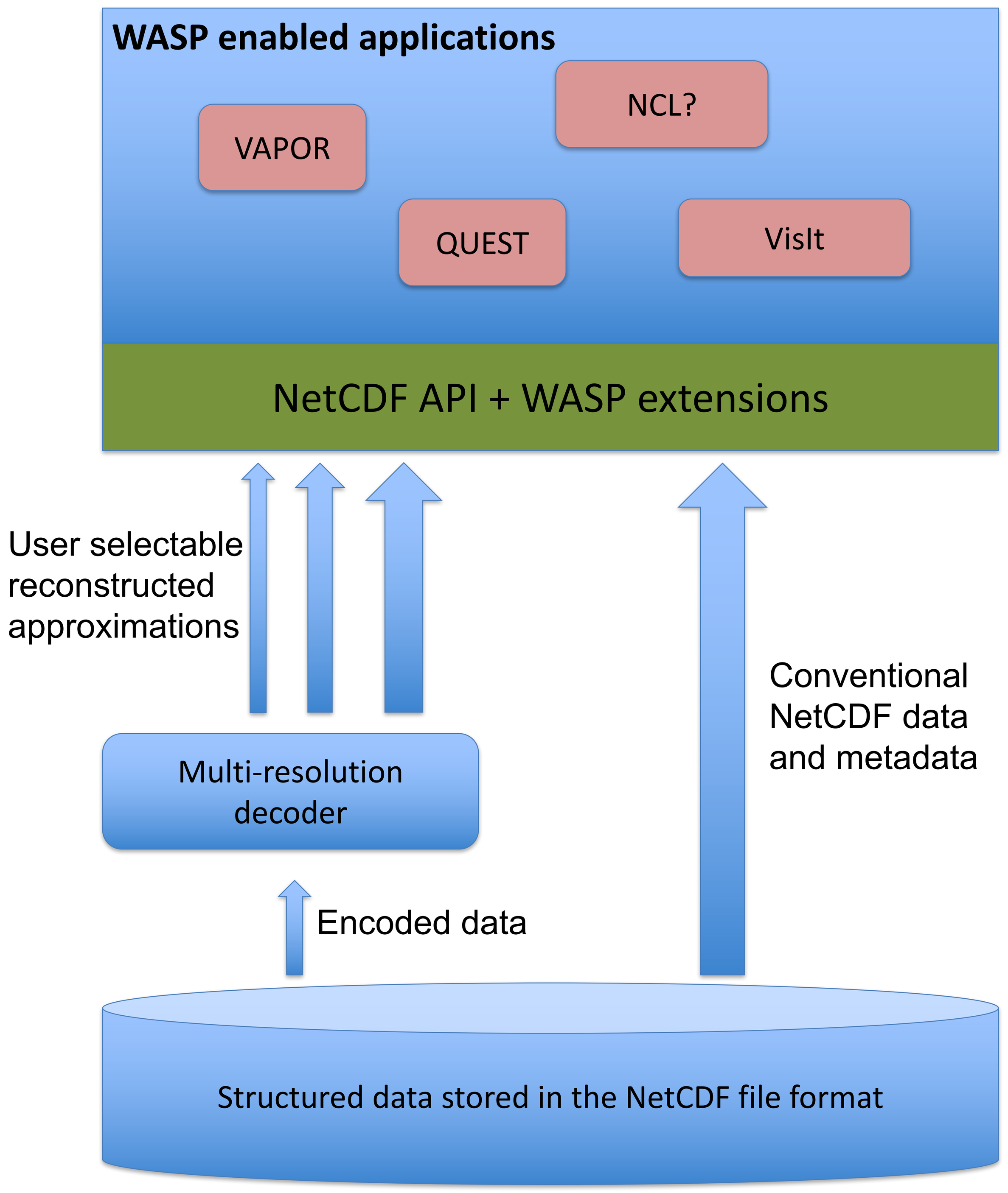
  http://www.cisl.ucar.edu

WASP project goals

The WASP project expands upon the successful, open source VAPOR (www.vapor.ucar.edu) project, an NSF funded geosciences data visualization and analysis package with thousands of users worldwide. VAPOR enables the interactive exploration of the highest resolution numerical simulation outputs arising from climate, weather, and other disciplines. The cornerstone of this ability is an intelligent wavelet based storage format that allows the researcher to progressively coarsen or refine data, as needed, thus trading off speed (interactivity) for fidelity. The goals of the WASP project are to enhance and reuse the first generation VAPOR storage model, both improving its utility for the existing VAPOR user community, and generalizing the software to make it suitable for other science domains, such as the bioimaging community. Specific objectives are to:

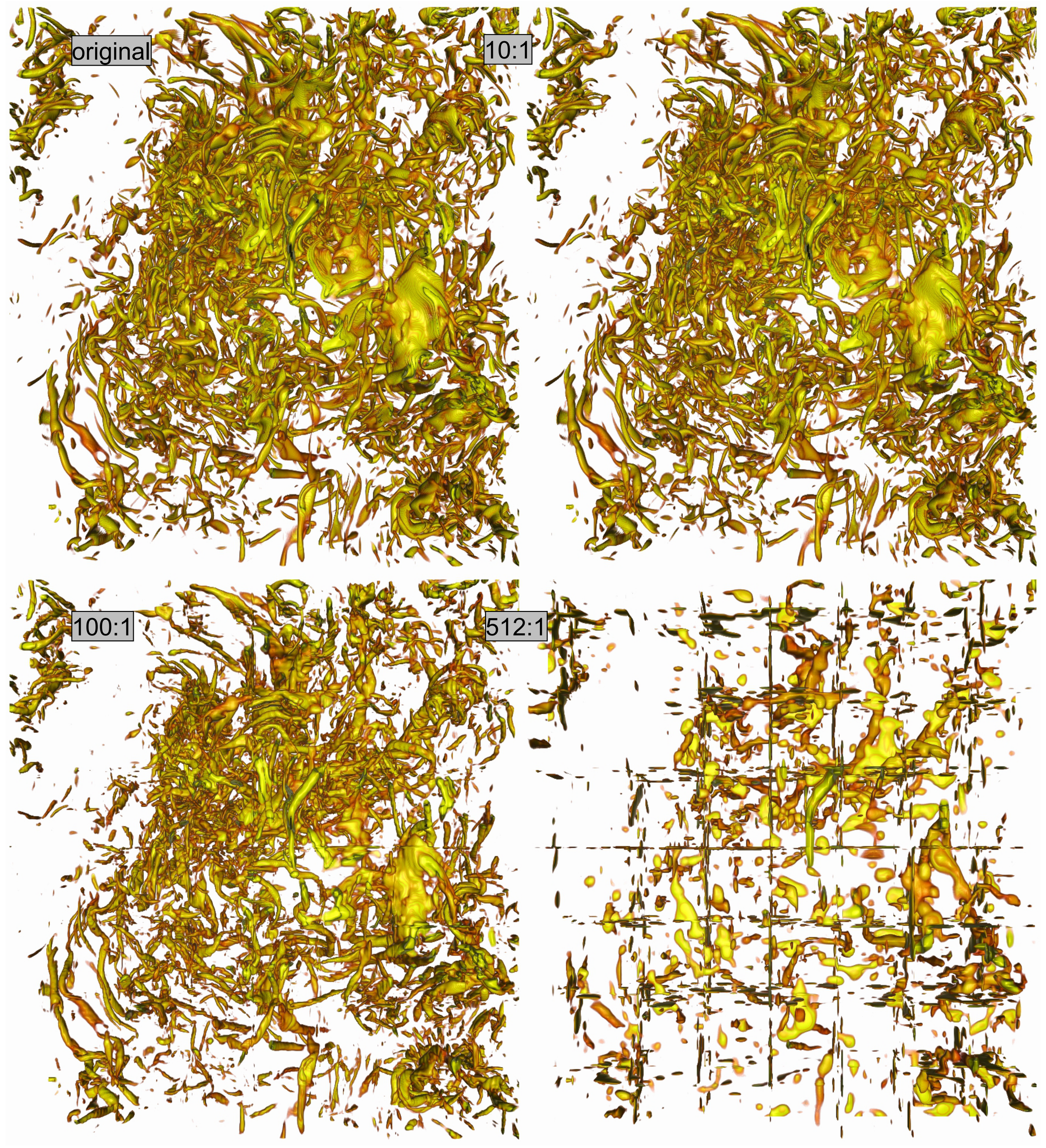
1. **Develop open NetCDF file format conventions for progressive data access using wavelets**
2. **Develop an open source toolkit (API) in support of these storage conventions**
3. **Integrate the new toolkit into the extant VAPOR geoscience visualization tools**
4. **Integrate new the toolkit into the STK biological imaging visualization and analysis tools**

WASP functional diagram



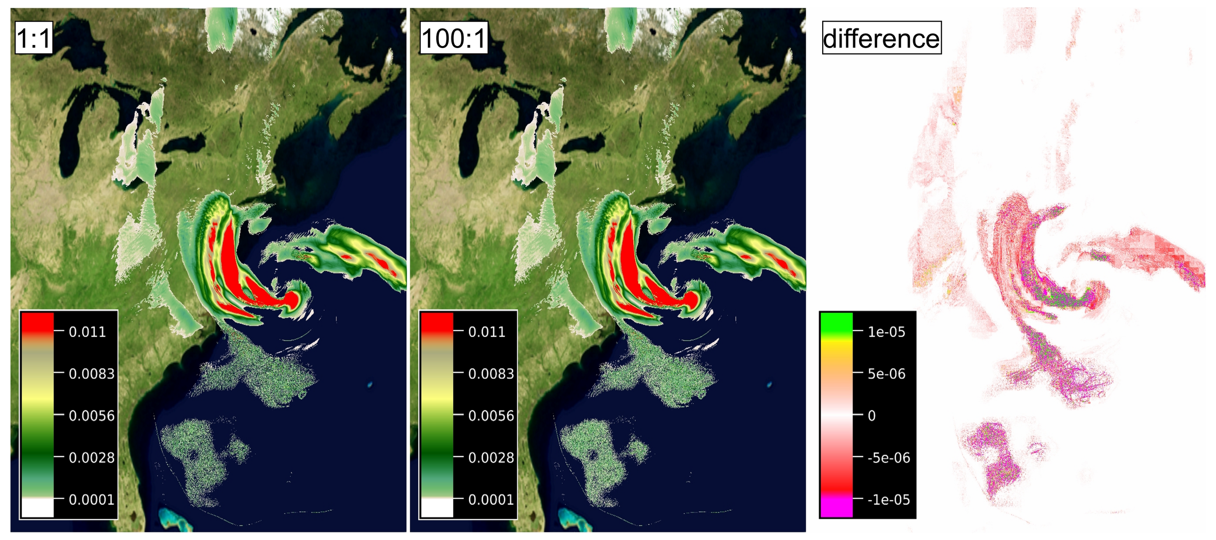
Turbulence simulations

Here we show qualitative results volume rendering data arising from high resolution numerical simulations that have been compressed by discarding the smallest wavelet coefficients. The 2048^3 simulation explores particle-turbulence interactions in conditions which mimic cumulus cloud cores[3]. Shown is the enstrophy field derived by taking the curl of compressed velocity field components using 6th-order finite differences. Compression factors are indicated in the top-left corner of each sub-figure. The original uncompressed variables occupy 32GBs per time step.



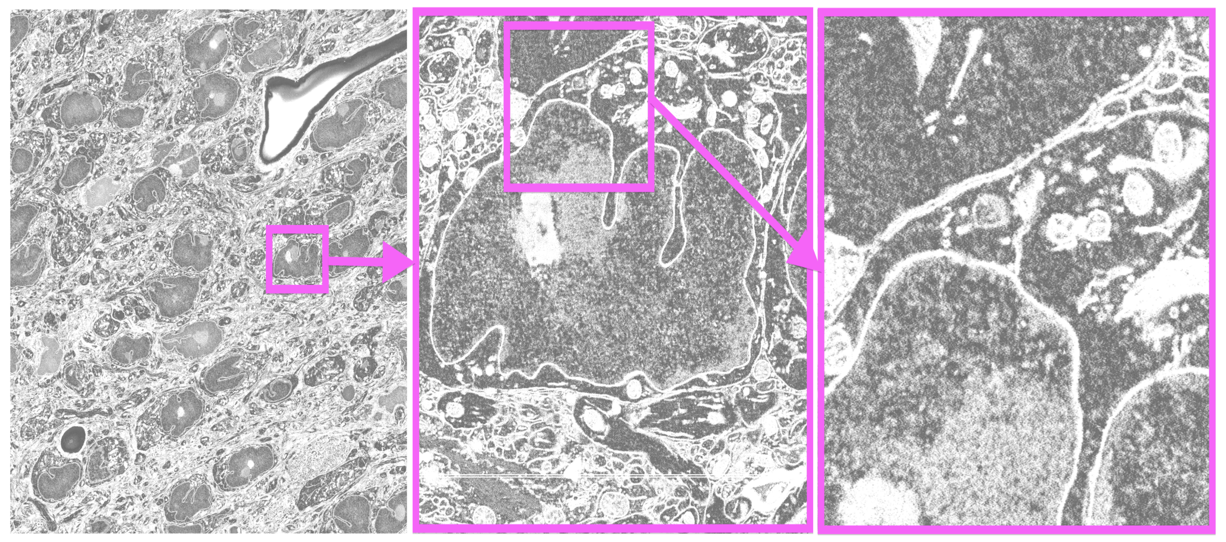
NWP: Hurricane Sandy

Here we show results for a very high WRF resolution simulation of Hurricane Sandy, computed on a 0.5km grid (~5000x5000x150 grid points)[4]. Rain water mixing ratio summed over a vertical column. Original (left), compressed 100:1 (middle), and difference between fields (right).



Microscopy imaging

WASP has applications beyond geoscience and even numerical simulation. The figure below illustrates the power of progressive data access applied to a massive microscopy data set acquired at the National Center for Microscopy and Imaging Research. The raw data consist of 32,000 by 24,000 by 703 grid points of 16bit quantities, occupying over 1 Terabyte of storage and thus virtually impossible to interactively explore in their raw form. However, with WASP an isolated region-of-interest can be rapidly identified and isolated using a coarsened representation, and then examined in detail using the highest fidelity data available. Data courtesy of M. Ellisman, UCSD.



Performance

The table below gives an indication of the speedups possible when visualizing a 0.5km Hurricane Sandy simulation, shown in the previous panel, using a commodity laptop, and a high-end graphics workstation. All times are reported in seconds.

Reduction / Device	Coefficient Resolution	1:1 1:1	1:10 1:1	1:100 1:1	1:1 1:8	1:10 1:8	1:100 1:8	1:10 1:64	1:100 1:64
Laptop	Read	352	57	9	358	52	8	49	7
	Transform	14	14	19	3	2	2	0	0
	Display	n/a	n/a	n/a	5	5	5	1	1
	Total	366	71	28	365	59	15	51	8
Workstation	Read	244	49	7	188	46	5	33	5
	Transform	194	191	194	28	26	26	4	5
	Display	n/a	n/a	n/a	n/a	n/a	n/a	2	2
	Total	437	241	200	216	72	31	39	11

Download WASP prototype today!

A working prototype of WASP is available for download from:
www.waspdata.org

WASP is also available as part of VAPOR Version 3:
www.vapor.ucar.edu

References

1. S. Mallat. A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way, 2008.
2. J. Clyne and A. Norton. Progressive Data Access for Regular Grids, in High Performance Visualization, 2012
3. Data source: Peter J. Ireland and Lance R. Collins (Cornell University)
4. Data source: Peter Johnsen (Cray), Mel Shapiro (NCAR), and Tom Galerno (NCAR)

This work is supported by NCAR's Computational & Information Systems Lab. (CISL), UCSD Center for Scientific Computation in Imaging (CSCI) and NSF Grants and ACI-1440412, and the Korean Institute for Science and Information Technology (KISTI).