UW-MADISON HTCONDOR INSAR WORKFLOW

Elena C. Reinisch, Ph.D. Student

UW-Madison Department of Geoscience

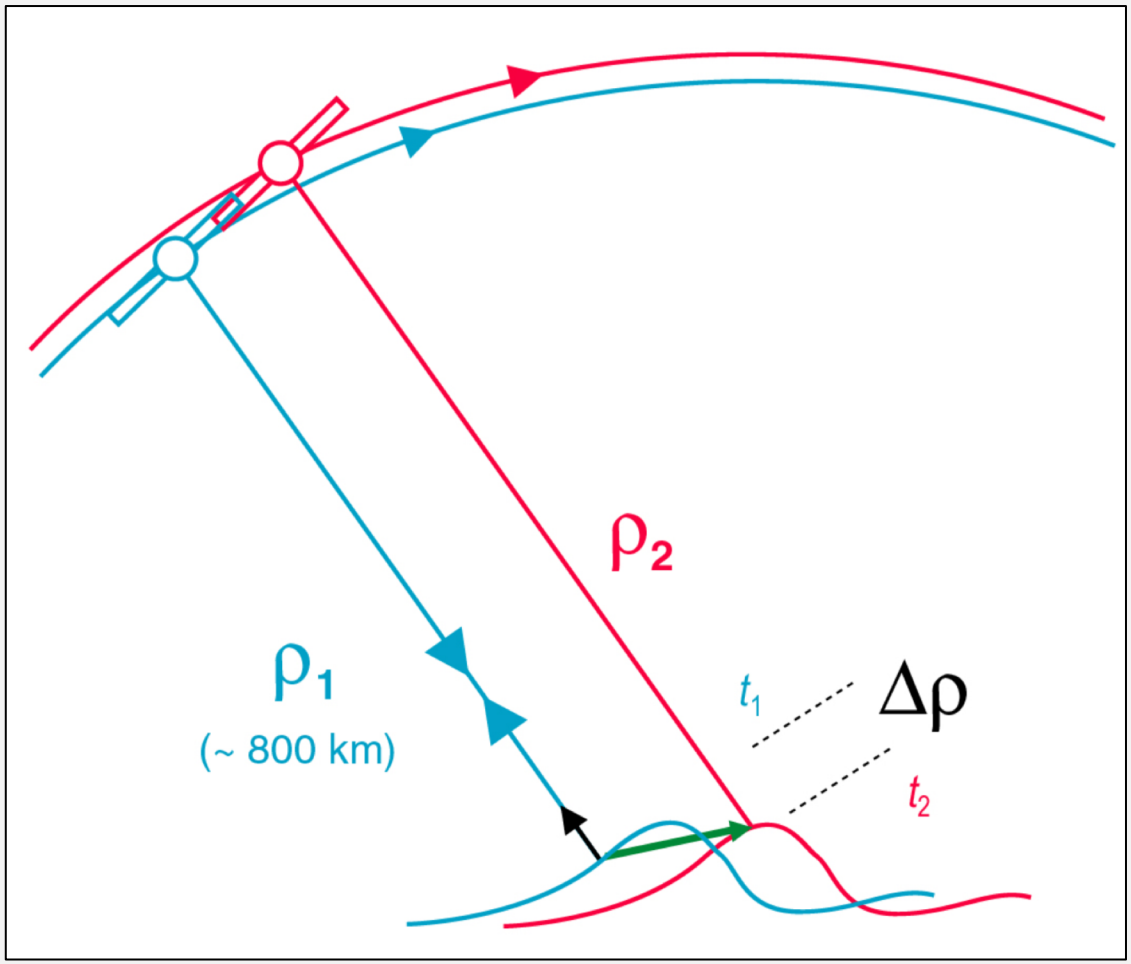
May 24, 2018

WHY MONITOR DEFORMATION?

Tracking deformation patterns in space and time can lead to a better understanding of geophysical processes

- earthquakes
- sinkholes
- volcanoes
- geothermal areas

Interferometric Synthetic Aperture Radar (InSAR)



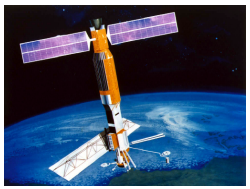
Range Change

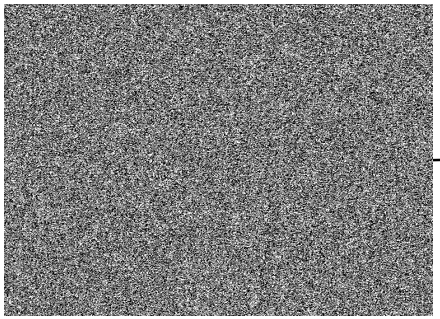
$$\Delta\rho = \frac{\lambda}{2}\left[\phi(t_2) - \phi(t_1)\right]$$

$$\Delta\rho = -\mathbf{u} \bullet \hat{\mathbf{s}}$$

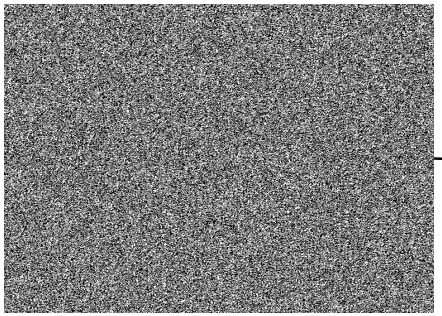
- First image at (epoch) t_1
- Second image at t_2
- Phase shift $\Delta\phi \rightarrow$ Range change $\Delta\rho$
- Component of ground displacement **u** along line of sight **s** from the ground to the radar

Satellite Observation



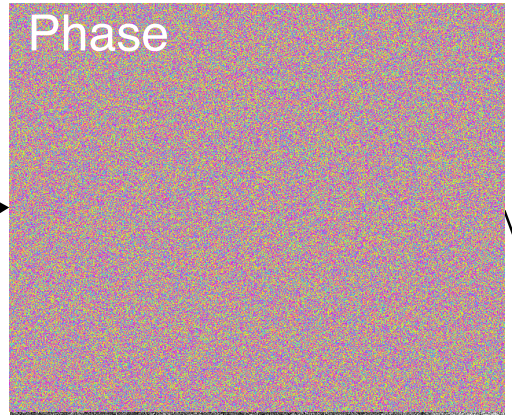


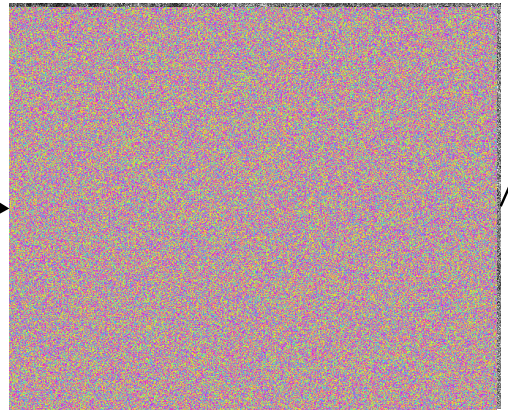
First Time

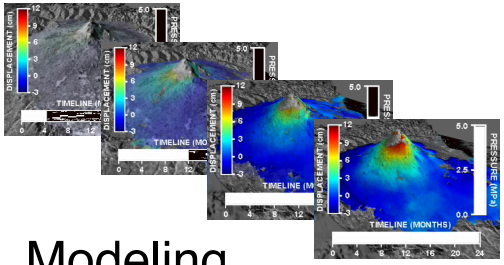


Next Time

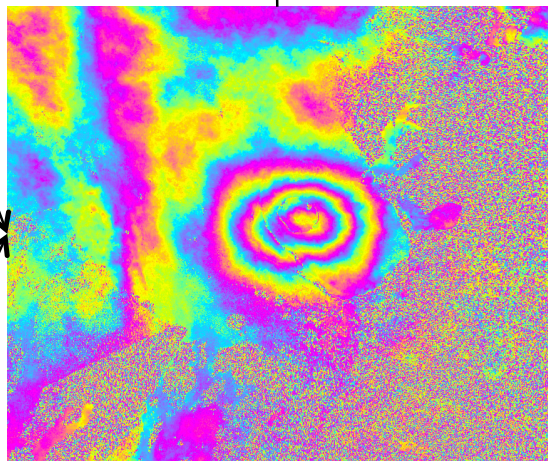
Magic of imaging







Modeling



Magic of interferometry

Figure courtesy of Paul Rosen, “Principles and Theory of Radar Interferometry”. UNAVCO Short Course, Aug. 4, 2014

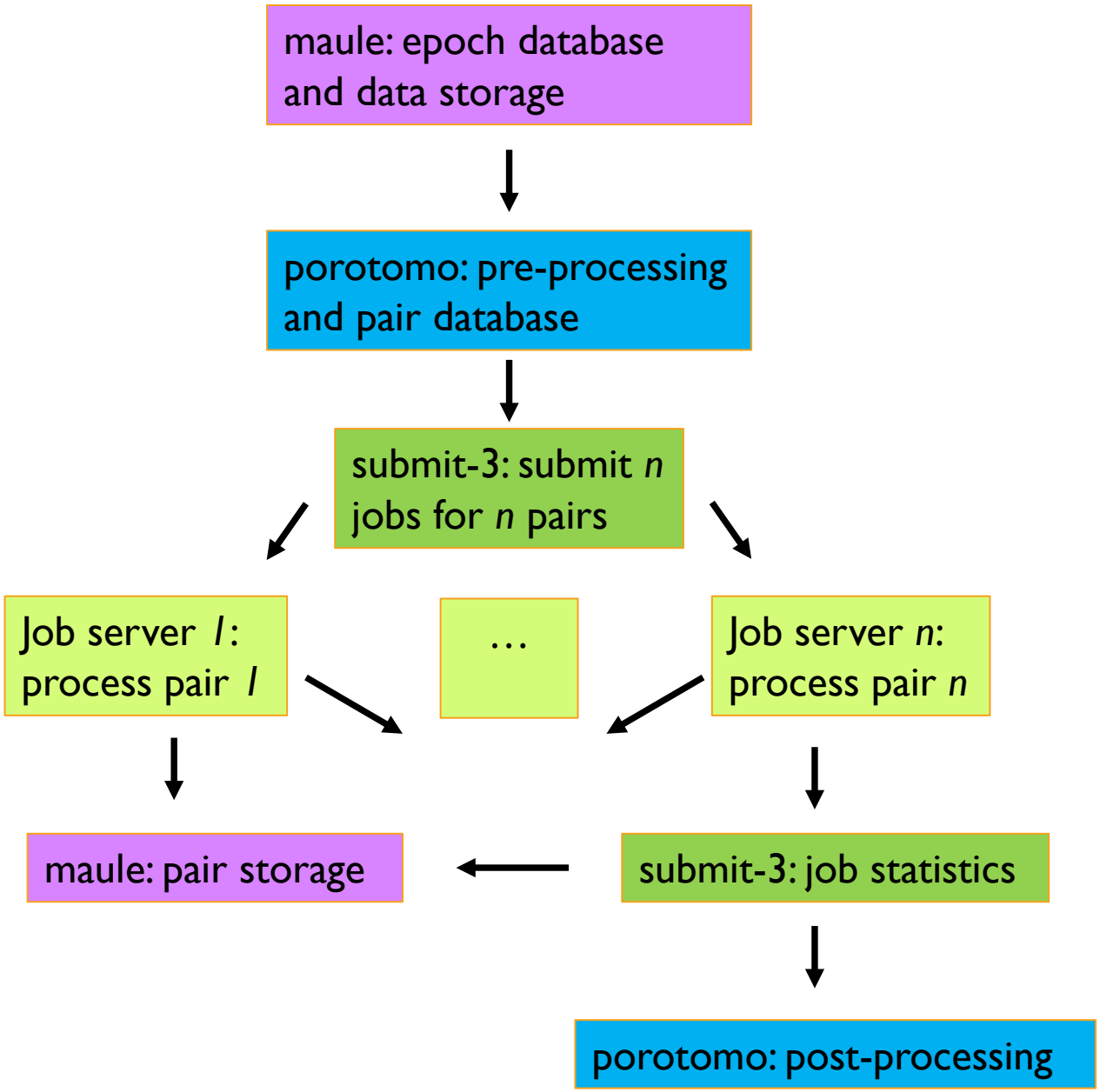
OBJECTIVE

Use high throughput computing resources to independently process *many pairs at a time* in a *consistent* and *user-friendly* manner

- GMT5SAR (InSAR processing software) [Sandwell et al., 2011]
- UW-Madison's Center for High Throughput Computing (CHTC)

Sandwell, D., Mellors, R., Tong, X., Wei, M., & Wessel, P., 2011. GMTSAR: An InSAR Processing System Based on Generic Mapping Tools, UC San Diego: Scripps Institution of Oceanography.

PROCESS AT A GLANCE



COMPATIBILITY ACROSS SERVERS

- Define consistent set of environment variables in setup.sh scripts and login files
- Consistent directory scheme across storage and computational servers
  (e.g., /[disk]/insar/[sat]/[trk]/raw)
- Direct SSH between servers

EXAMPLE SETUP.SH SCRIPT

manage permissions

set path to repo

add paths to GMT5, GMT5SAR, and workflow scripts

set variables for server credentials

set variables for SAR database credentials

setup_maule.sh (~/tmp) - VIM1

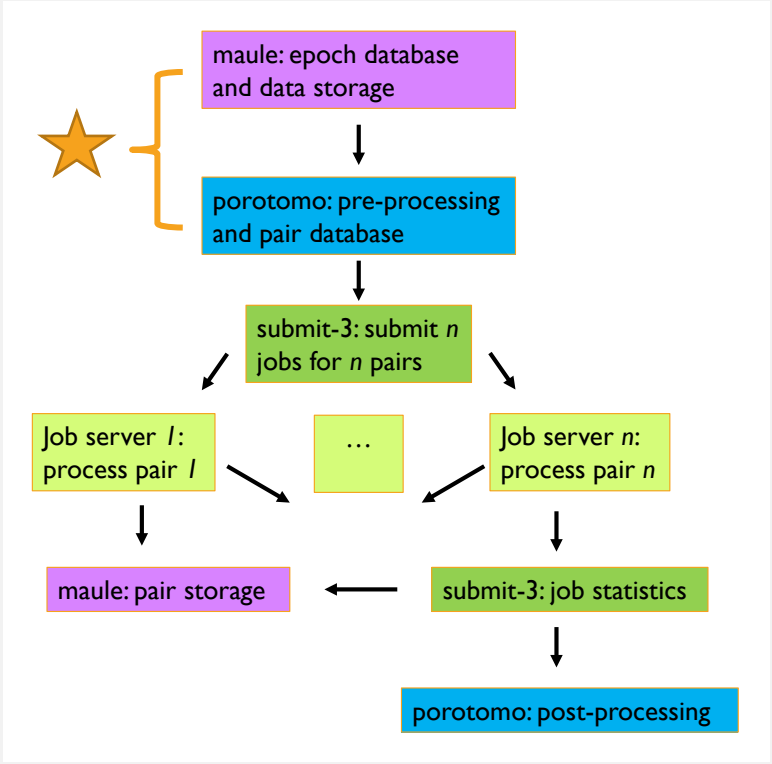
setup_maule.sh buffers

```
1  # Make files readable and writable to others using workflow
2  umask 002
3
4  # set location of Workflow repo
5  export bin_htcondor_home='/home/ebaluyut/bin_htcondor'
6
7  # set paths to GMT/GMTSAR
8  export PATH=/opt/gmt5/bin:$PATH
9  export PATH=/opt/gmt5/lib64:$PATH
10 export LD_LIBRARY_PATH=/opt/gmt5/lib64:$LD_LIBRARY_PATH
11 export PATH=/home/ebaluyut/SSARA-master:$PATH
12
13 # set path to workflow metadata directory
14 export PATH=/home/ebaluyut/gmtsar-aux:$PATH
15
16 # set paths to workflow scripts
17 export PATH=${bin_htcondor_home}:$PATH
18
19 # For compilation only
20 export LIBRARY_PATH=/opt/gmt5/lib64:$LIBRARY_PATH
21
22 export submit3='ebaluyut@submit-3.chtc.wisc.edu'
23 export ice='ebaluyut@ice.geology.wisc.edu'
24 export t31='ebaluyut@ice.geology.wisc.edu:/mnt/t31'
25
26 # export SAR credentials
27 export unavuser='*****'
28 export unavpass='*****'
29
30 export asfuser='*****'
31 export asfpass='*****'
32
33 export eossouser='*****'
34 export eossopass='*****'
35
36 #set Python Path
37 export PYTHONPATH=/home/ebaluyut/SSARA-master
38 export PYTHONPATH=${bin_htcondor_home}:$PYTHONPATH
39
```

NORMAL setup_maule.sh sh 97% 38: 50

SUMMARY OF INITIAL STEPS (1-2)

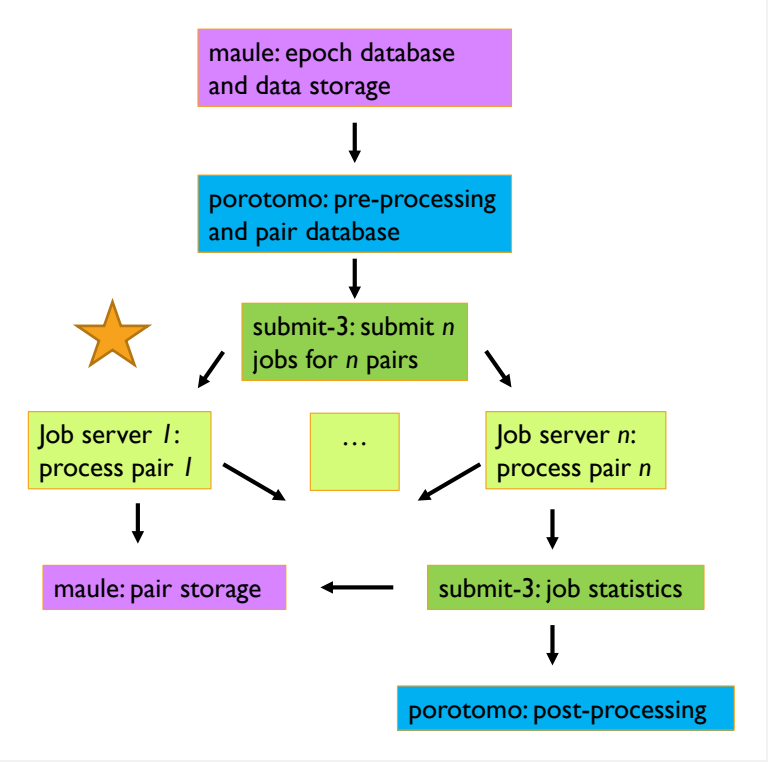
- Download raw data
- Prepare data for jobs
- Update databases
- Generate sub-list of pairs to process based on input criteria
- Copy sub-list of pairs to submit-3 server



#mast	slav	orb1	orb2	doy_mast	doy_slav	dt	trk	orbdir	swath	wv	bpar	bperp
20141103	20141127	3111	3461	306.076679666951009	330.076605207008981	24	T64	A	F2	0.0555	0.0	0.0
20141103	20141221	3111	3811	306.076679666951009	354.076473717008980	48	T64	A	F2	0.0555	22.5	-13.6
20141103	20150207	3111	4511	306.076679666951009	37.076645720630999	96	T64	A	F2	0.0555	36.5	42.7
20141103	20150502	3111	5736	306.076679666951009	121.076640926082007	180	T64	A	F2	0.0555	-22.1	-58.6
20141103	20150526	3111	6086	306.076679666951009	145.076688603830007	204	T64	A	F2	0.0555	-55.6	-93.5

3. SUBMIT JOB (ON SUBMIT-3)

- For each pair in sub-list of pairs:
  - submit job
  - with each job, transfer:
    - Portable (compressed) versions of software and orbit files
    - setup.sh script
    - SSH keys
- User friendly
  - "make run"



EXAMPLE SUBMIT FILE

TSX_T53_In20180112_20180319.sub+

```
universe = vanilla
# Name the log file:
log = TSX_T53_In20180112_20180319.log

# Name the files where standard output and error should be saved:
output = TSX_T53_In20180112_20180319-process.out
error = TSX_T53_In20180112_20180319-process.err

# Specify your executable (single binary or a script that runs several
#  commands), arguments, and a files for HTCondor to store standard
#  output (or "screen output").
executable = run_pair.sh
arguments = "TSX T53 20180112 20180319 ebaluyut strip_008 brady_dem_3dep_10m_kf.grd 80 -119.03 -118.99 39.78 39.82 brady"
output = TSX_T53_In20180112_20180319.out

# Specify that HTCondor should transfer files to and from the
#  computer where each job runs. The last of these lines *would* be
#  used if there were any other files needed for the executable to run.
transfer_input_files = run_pair.sh, bin_htcondor.tgz, setup.sh, GMT5SAR.tgz,  gmtsar_dependencies.tgz, ssh.tgz,  GMT.tgz

# additional requirements
# this will allow us to make sure that you don't have too many jobs transferring data from SSEC at the same time
concurrency_limits = SSEC_FTP

# Tell HTCondor what amount of compute resources
#  each job will need on the computer where it runs.
# It's still important to request enough computing resources. The below
#  values are a good starting point, but consider your file sizes for an
#  estimate of "disk" and use any other information you might have
#  for "memory" and/or "cpus".
request_cpus = 1
request_memory = 3GB
request_disk = 12GB
queue
```

TRANSFERRED SETUP.SH SCRIPT

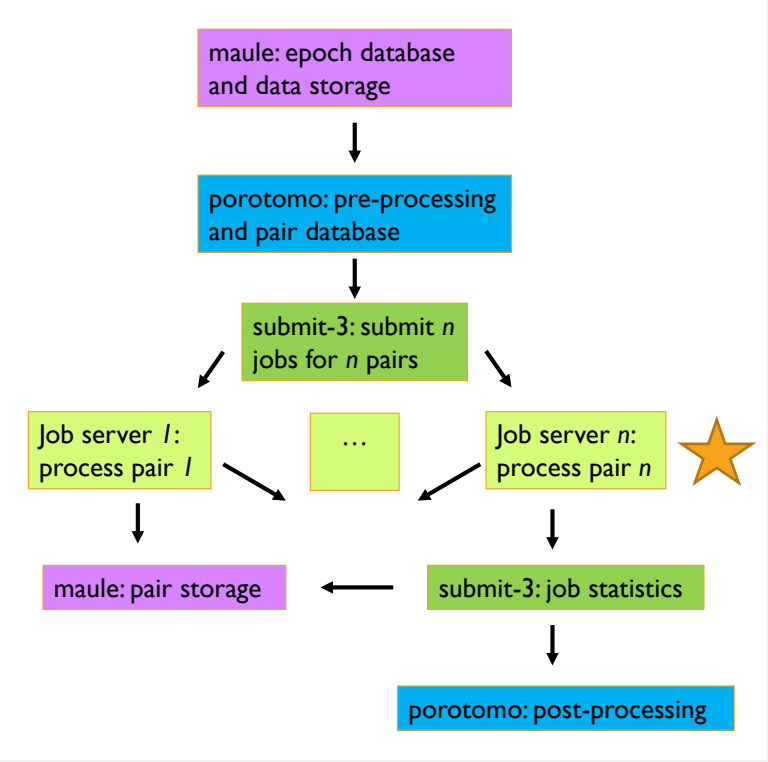
```
#GMT and GMTSAR
umask 002

export OMP_NUM_THREADS=1
export PATH=`pwd`/bin_htcondor:$PATH                    # scripts from Kurt Feigl
export PATH=`pwd`/this_gmt/bin:$PATH                # GMT    points to a symbolic link
export PATH=`pwd`/gmtsar/bin:$PATH                  # GMTSAR points to a symbolic link
export PATH=`pwd`/GMT5SAR5.2/bin:$PATH                 # GMTSAR points to a symbolic link
export PATH=`pwd`/gmtsar_dependencies/bin:$PATH # needed for libraries
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:`pwd`/this_gmt/lib64
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:`pwd`/gmtsar_dependencies/lib

# save account name for maule
export maule=ebaluyut@maule.ssec.wisc.edu
```

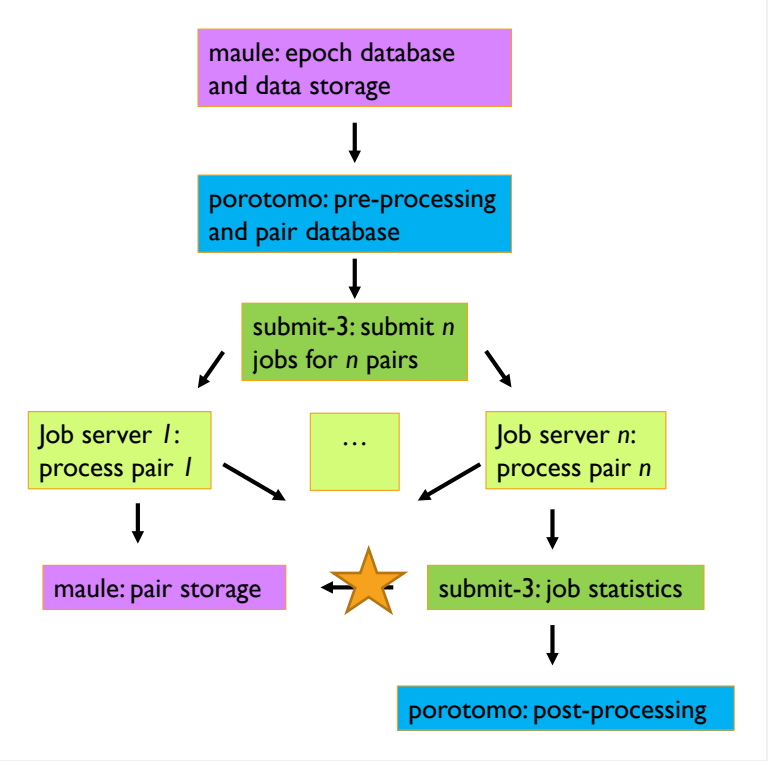
4. PROCESS PAIR (ON JOB SERVER)

- On each job server (non-interactive):
  - Setup software and SSH directories
  - Copy data files from maule
  - Process pair
  - Transfer tarred result to maule



5. GATHER JOB STATISTICS (ON SUBMIT-3)

- After all pairs have finished, gather statistics on jobs (submit-3)
- Transfer files to chosen server
  - maule - permanent storage
  - porotomo - post-processing



LESSONS LEARNED SO FAR

- Handling different levels of computational proficiency
  - Makefiles
  - Well-documented code
  - Comprehensive user manual

- Improving workflow efficiency with the help of CHTC
  - Portable software
  - Direct SSH

- Version tracking and software consistency
  - Git repository

SUCCESS AND FUTURE WORK

- Many successful runs for a variety of satellites:
  - S1-A
  - ERS-2
  - ALOS
  - TSX/TDX
  - Envisat

- Many successful sites:
  - Geothermal fields: Brady Hot Springs, NV; East Mesa, CA; Don Campbell, NV; Dixie Meadows, NV; Tungsten, NV; Coso, CA
  - Laguna del Maule Volcanic Field, Chile

- Continuing work on automating processes, tutorials, etc.
- Software publicly available on GitHub: https://github.com/ecreinisch/bin_htcondor
- Workflow manual available at https://elenacreinisch.com/page/software

ACKNOWLEDGEMENTS

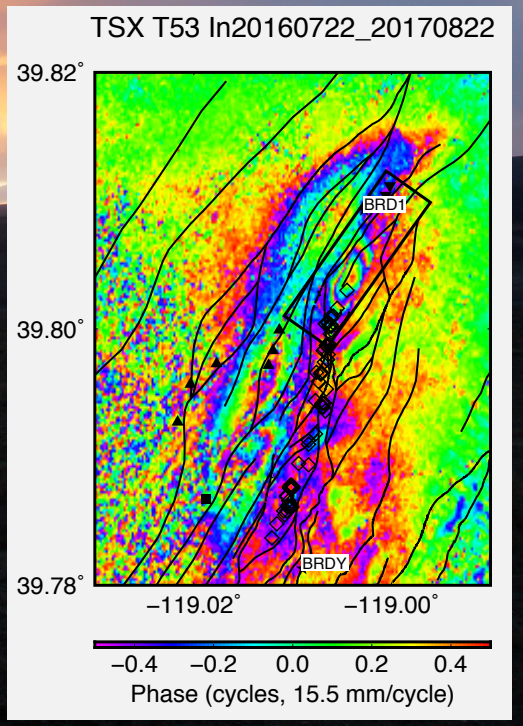
- Thank you to the following people for giving helpful discussions, and advice, and for testing the workflow:
  - Kurt Feigl, Sam Batzli, Hélène Le Mével, Dougal Hansen, Tabrez Ali

- Thank you to UW-Madison's CHTC and specifically their coordinators, Lauren Michael and Christina Koch

- Thank you to the GMT5SAR community for all of the collaboration

- This material is based upon the work support by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1256259

THANK YOU!
QUESTIONS?



Brady Hot Springs Geothermal field, NV