

# New HOPS: a discussion

*Geoff Crew, MIT Haystack Observatory  
Hilo, Hawaii, December 2, 2019*

# The ngEHT Mandate

- Picking up from where M. Janssen left off from Monday...
- HOPS performed adequately for 2017 data reduction
  - consistency with AIPS and CASA was established
  - the HOPS-based pipeline was adopted for production work\*
- The ngEHT funding via MSRI-I calls for updates to HOPS:
  - 4 years of development to become “shovel ready”<sup>o</sup>
  - N stations more than doubles (e.g. to 20-30)
  - Bandwidth quadruples at some sites (e.g. 8 GHz)
  - Simultaneous 230 / 345 GHz observing perhaps
  - Anything else that gets decided in then next 4 years

# So What's Broken?

- We've (almost) come to the end of what can be fixed with “bandaids”
- Maximum # of channels is 64 and that's hard to fix (a-zA-Z0-9%\$)
- No full complex bandpass corrections (only per-channel phase+delay)
- single-baselines are both a virtue and a problem:
  - HOPS finds more fringes than AIPS or CASA
  - station-based phases and delays are not readily accessible
- Fourfit is a one-shot process; multi-step processing not supported
- User interface is a challenge:
  - control file syntax is a bit arcane
  - all you get is one fourfit plot that either works or is garbage

# An Embedded History of VLBI

- HOPS code has 30+ years of history in it
  - was coded in C, but reads like the Fortran it was ported from
  - not modular except for a few of the i/o libraries
  - was written for hardware correlators
  - was written for computers that no longer exist
  - little endianism won out over big endianism (apparently)
  - was (successfully) successfully adapted to DiFX (but not e.g. SFXC)
- Plotting and results are not independently generated
  - amplitude and SNR come as side-effects of what you plotted
  - PGPLOT is maybe ok today, but not really supported anymore



# Some New Features

- Global fringe solutions (and station based-quantities)
- Complex bandpass
- A more human-friendly interface (e.g. Python as CASA does)
- Allow distributed computing and/or parallelization (threads, OpenMPI)
- Insert hooks to allow plug-in modules for customizations as needed
- Allow a strategy for iterative calibration and fringing
- Improved data formats (internal in-memory as well as disk storage)
- Enable better exchange with other analysis packages:
  - FITS-IDI? (or HDF5 or whatever else comes along?), CASA MS?, ...
  - (either enables better use of HOPS with simulated data)
- A more flexible/interactive plotting system
  - single summary is fine when everything is working
  - provide real support for investigation of problems

# Likely Changes, Part 1

- Maintain existing tools “as is” for serious regression (probably patched)
- Arbitrary number of channels; eliminate internal magic sizing numbers
- New control file format (e.g. use python or some config module)
- New internal data formats (rationalized, new structures or objects)
- New disk data formats: “mk4” ➡ “hops”
  - machine/compiler independent little-endian (not big-endian)
  - rationalized data types (as with internal formats, optimized for disk i/o)
  - new root file format (ovex is ancient history, and current root is artificial)
  - preserve the current m4py-type capability
  - allow translator tools to exchange with “hops”, “mk4” and other formats

# Likely Changes, Part 2

- Basically: FIX what is broken
- Not gratuitously break the current pipelines, but allow simplification
- Most likely to be implemented in a mix of C, C++ or Python
- Provide a more canonical adaptation to unix/linux environments
- *Implement what is most important to have available in 4 years*

# So, We're Looking for Input

- Don't bother saying "just use my package"
- Tell us what you think is missing (Me, John Barrett, Lindy Blackburn)
- Provide input on priorities:
  - "gotta haves" vs "things to enable and implement in the future"
  - there is really not a lot of support for this
  - community partners may be interested in making some enhancements