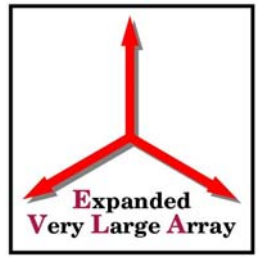
Computing: Off-line Processing

J. McMullin

Off-line: Key Interrogatives

- Why
- What & When
  - Priorities, Timelines
- Who & Where
  - Organization, Resources
- How
  - Process

Why AIPS++



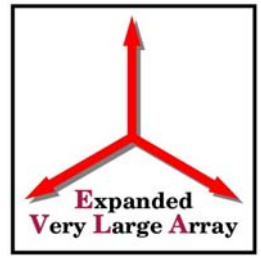
- ALMA has adopted this as the baseline off-line reduction package
- Existing packages not easily scalable/extensible to demands of future instruments
- AIPS++ has not fulfilled its mission –

  Can it?

Why: Technical Review March 03

- Technical Review indicated no fundamental flaws with AIPS++ core or architecture, but found defects in process and management methodology
  - Progress made past 6 months, in right direction
  - Strengthen Project Scientist and Project Manager
  - Focus on end-to-end Use Cases targeted toward projects
    - Base on current instruments, e.g. VLA
    - Aim for key projects ALMA, EVLA
  - Short-term goals performance & robustness
  - Pursue technology upgrade
    - Choices 10 years old, some better options now
    - Will need to meet performance goals (e.g. GUIs)
    - Proof-of-concept based on ALMA ACS

ALMA Sequence



- Requirements
- Audit
- 3 Phase Testing
- Benchmarks
- Testing Plan

What and When

Items marked with an 'X' are complete;

Items marked in red are specific to ALMA

R1 – October 1, 2003

	Audit – Functionality Improvements	Requirement
X	Continuum Subtraction	(5.2-5)
X	Data Calibration Enhancements	(4.1-3)
X	Viewing cursor position enhancements	(7.1-5)
X	Blinking prototype	(7.1-3)
	MS concatenate/split capability	(3.1-20.1:5)
	Atmospheric correction:specified zenith opacity	(4.2-3)
X-Add	Automated Benchmark; web publication	(1.1-4)
X	Single Dish Imaging: selectable weighting	(5.2-6)

X Performance

Imaging comparisons will improve to within a factor of 2 for Gildas (256x256) and AIPS (up to 2Kx2K).

Robustness

All developer's have 25% time allocated for testing/defect resolution.
Testing process (project testing, NAUG, SSR).

Infrastructure

MeasurementSet GAIN subtable design/review
MeasurementSet revision design/review
X Multiple spectral window imaging (enhancement for integrated testing)
Drop Region specification design (required for 5.2-5, etc)
vlafiller enhancements (required for 4.2-2, etc)

ALMA Prototype Pipeline

Initial Phase A (Report due 15-October-2003)

R1.1 – April 1, 2004

Audit – Functionality Improvements	Requirement
Calibration interpolation	(4.1-5)
Plot selection enhancements	(7.2-1)
Viewer overlay improvements	(7.2-6.3)
MS concatenation/split capability	(3.1-20.1:5)
Msplot zoom enhancements	(7.2-2)
Calibration enhancements	(4.3-4)
Calibration enhancements	(4.3-5)
Statistical editing of data	(4.1-6.7.1:4)
Data selection improvements	(7.3-2)
Editing of calibration solutions	(4.1-12)
General quantity editing of data	(4.1-4.1:9)
Translation from/to units of temperature	(6.1-3.2; external)

X Performance

Imaging comparisons will improve to within a factor of 2 of AIPS over a range of image sizes (up to at least 4K x 4K).
Calibration comparisons (core calibration, i.e., rf passband, phase and amplitude) will improve to within a factor of 2 for Gildas and AIPS (<100 solution intervals).

Robustness

All developer's have 25% time allocated for testing/defect resolution.
Testing process (project testing, NAUG, SSR).

Infrastructure

Flagging tool integration/design (for Req:4.1-4,4.1-6)
Common Framework Design and Implementation begin 02-Dec-2003.

ALMA Prototype Pipeline

Completion of Phase B and C

What: Performance -- Strategy

- Finish AIPS++/IRAM Phase 3 (performance) test
- Set up automated, web-accessible, performance regression tests of AIPS++ against AIPS, Gildas, and Miriad
  - Start simple, then extend to more complex data (SSR requirements, guidelines, & expectations, for automated benchmark effort at: http://www.aoc.nrao.edu/~dshepher/alma/ssr.bm.rqmts.pdf)
- Systematically work through performance problems in importance order
  - Resolution of some issues will require scientific input (e.g., when is an inexact polarization calculation OK)

→ Summer 2004 (CDR2)

AIPS++/IRAM Phase 3

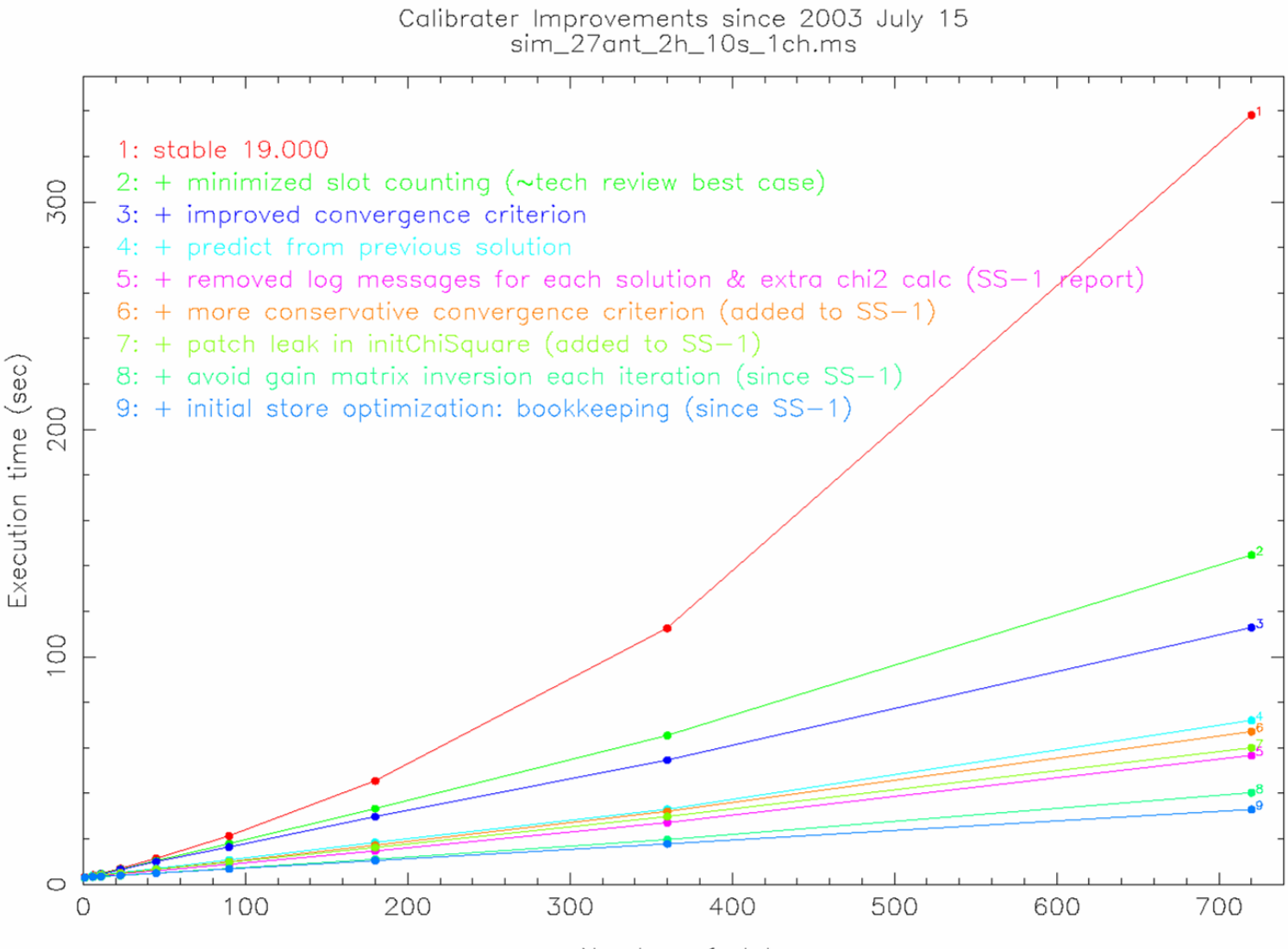
	GILDAS/CLIC	AIPS++	A/G	Comments
Filler	1873	10939	5.8	
Init (write header info)	385		n/a	
Fill model/corr data cols.		2140	n/a	
PhCor (Check Ph-corr data)	889	3484	3.9	(Glish)
RF (Bandpass cal)	5572	2298	0.4	
Phase (Phase cal)	3164	1111	0.4	
Flux (Absolute flux cal)	1900	2093	1.2	(Glish)
Amp (Amplitude cal)	2242	614	0.3	
Table (Split out calib src data)	1200	5150	4.3	(FITS write)
Image	332	750	2.3	
Total	17600s	28600s	1.6	

- Caveats: Gildas compilation, memory issues, core algorithms similar – main performance differences in non-comparable steps, AIPS++ implementation via scripting language

Performance: Calibrater

Improvements checked in and available now.

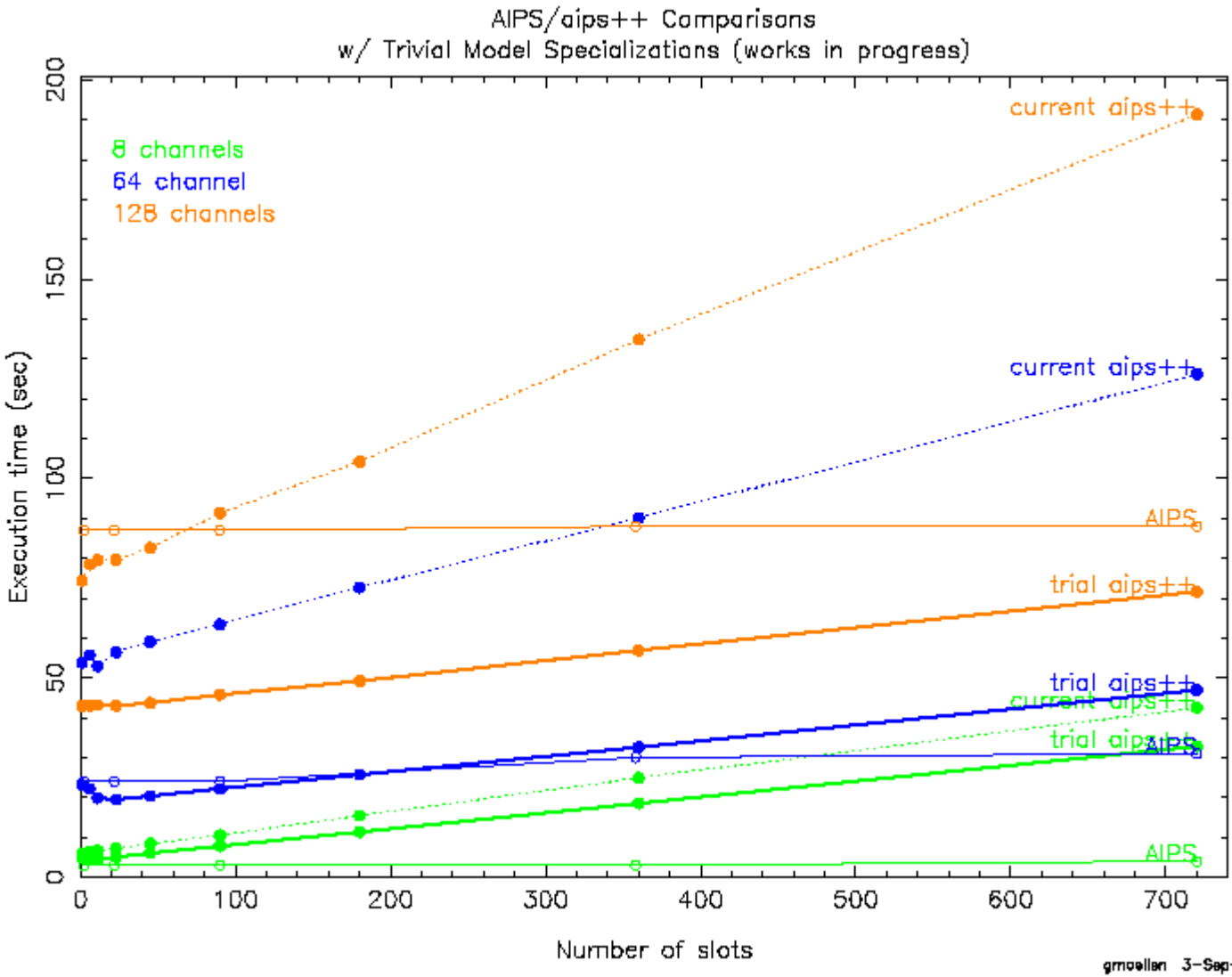
Factor of ~10 improvement for 720 solution intervals.



Performance: Calibrater

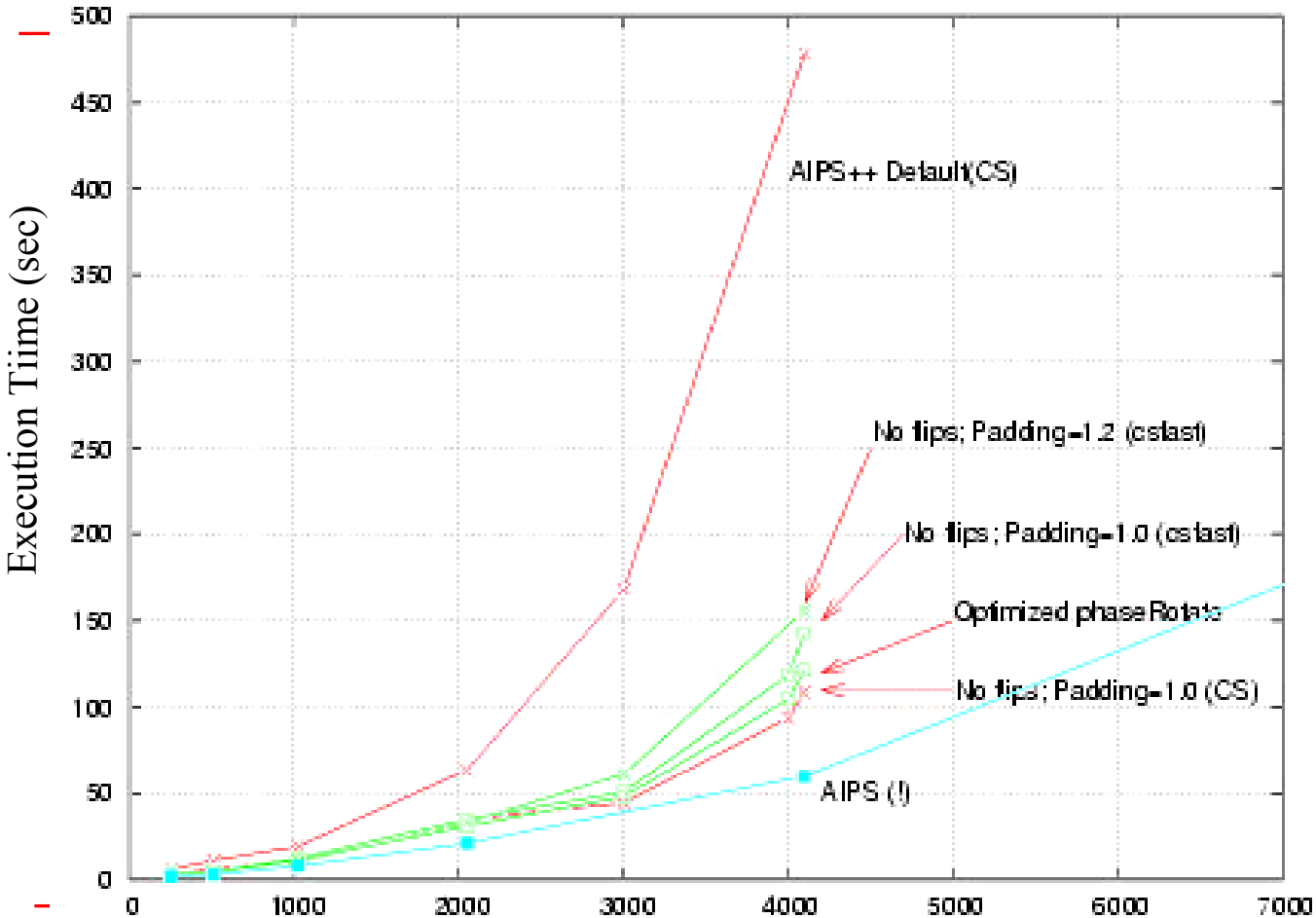
Trial aips++ fixes identified but not checked in:

- Don't read model from disk if pt src.
- Unnecessary copy of channel data removed.
- Slope now dominated by sol'n write-to-disk.



McMullin

Performance: Imager



Imaging performance improved through convolution optimization & tuning of minor/major cycles.

Improved by factor of 1.8 for 2048x2048 pixels.

Improved by factor of 4.4 for 4096x4096 pixels.

AIPS++/AIPS ratio now 1.6 for 2048x2048 pix & 1.8 for 4096x4096 pix.

Now dominated by more general polarization processing in AIPS++?

ALMA Benchmark

- **SSR has identified 2 initial benchmark datasets:**

- ***Pseudo GG Tau*** – PdBI data of 25 March. Original observation expanded to 64 antennas with GILDAS simulator & source structure converted to point source. 3 & 1 mm continuum & spectral line emission. Data in ALMA-TI FITS format (same data used during AIPS++ re-use Phase III test).
  - Ensure continuous comparisons in time with AIPS++ Ph III 're-use' test
  - Compare core functions (fill, calibrate, image) on ALMA-size dataset
  - Exercise mm-specific processing steps

- ***Polarized continuum data*** – VLA polarized continuum emission in grav lens 0957+561, 6cm continuum, 1 spectral window. Snapshot observation, 27 antennas, extended in time with AIPS++ simulator to increase run-time. Data in Standard FITS format.
  - Exercise full polarization calibration, self-calibration, non-point source imaging (polarization processing can only be compared with MIRIAD/AIPS).

Performance: Status: Initial results (benchmark 1 and benchmark 2)

	GILDAS/CLIC	AIPS++	A/G	Comments
• Filler	743	3931	5.3	
• Init (write header info)	319		n/a	
• Fill model/corr data cols.		1814	n/a	
• PhCor (Check Ph-corr data)	440	1129	2.6	(Glish)
• RF (Bandpass cal)	2943	1365	0.5	
• Phase (Phase cal)	1294	822	0.6	
• Flux (Absolute flux cal)	942	928	1.0	(Glish)
• Amp (Amplitude cal)	932	289	0.3	
• Table (Split out calib src data)	607	2135	3.5	(FITS write)
• Image	124	281	2.3	
Total	8344s	12694s	1.5	

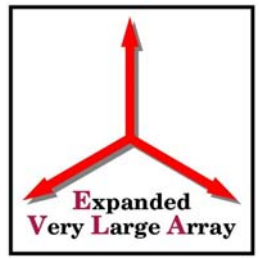
Timing on a 1.4 GHz machine with 512 MB RAM (very preliminary):
Miriad: 123s, AIPS: 783s, AIPS++: 465s ➔ A++/M = 3.7 A++/A = 0.6

What: Robustness

Testing “in-depth”

- Project Testing
  - Unit tests, assay
  - PM doc, function tests -> data set, script
- NAUG Testing
- ALMA Testing

ALMA Software User Test Plan: Status



- *Use Cases* – descriptions of operational modes and what external dependencies exist. Designed to exercise subsystem interfaces, functionality, & user interfaces.
- *Test Cases* – Use Case subset designed to test specific functions.
- *Testing timeline* (when tests run in relation to Releases, CDRs).
- *Test Definitions* - specifies which test case will be run, what the test focus is, and whether the test is automated or involves users.
- *Test Reports* (e.g., user reports, audit updates, summary).

What: Framework

- Open architecture which leverages user software development
- Modular architecture permits ongoing technological evolution
- Integration of C/C++, Java, and Python within the same framework
- New technology such as for VO is readily incorporated
- Legacy software can live on within a modern open framework

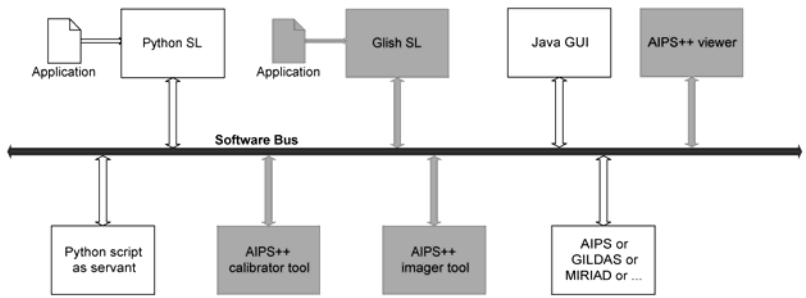
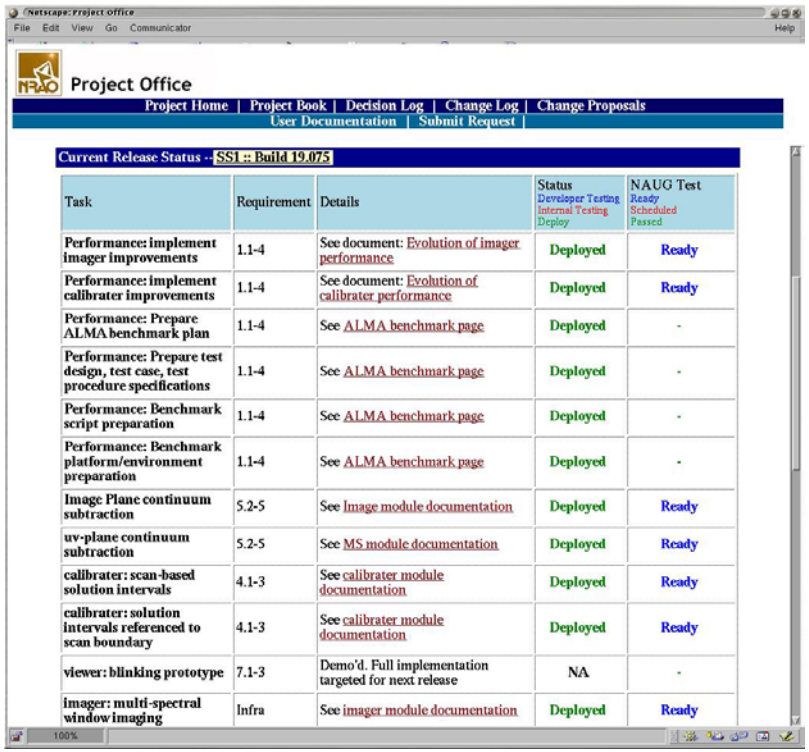


Figure 2. AIPS++ Architecture

Project Office



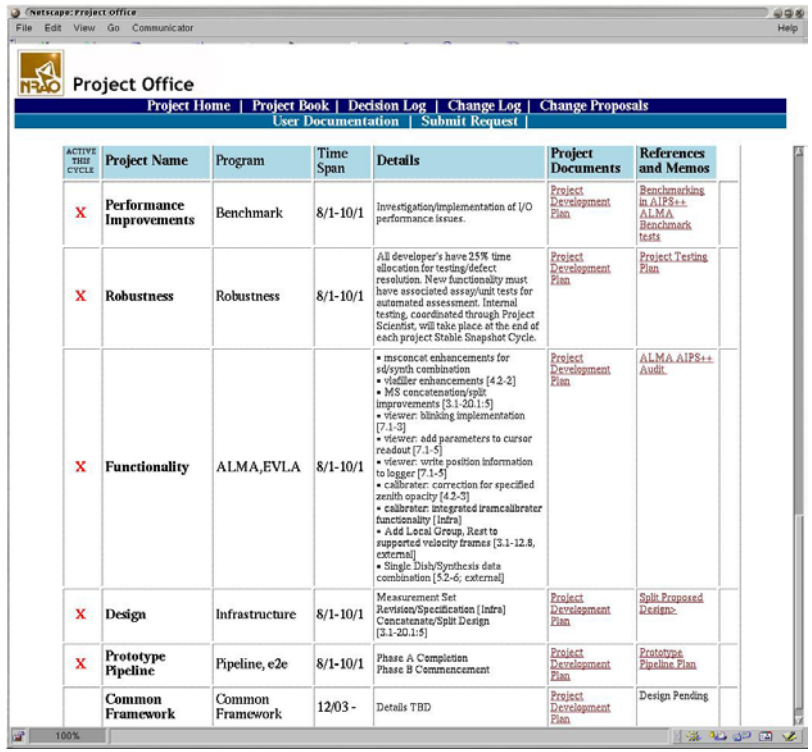
Project Office

Project Home | Project Book | Decision Log | Change Log | Change Proposals

User Documentation | Submit Request

Current Release Status -- SS1 :: Build 19.075

Task	Requirement	Details	Status (Developer Testing, Internal Testing, Deploy)	NAUG Test (Ready, Scheduled, Passed)
Performance: implement imager improvements	1.1-4	See document: Evolution of imager performance	Deployed	Ready
Performance: implement calibrater improvements	1.1-4	See document: Evolution of calibrater performance	Deployed	Ready
Performance: Prepare ALMA benchmark plan	1.1-4	See ALMA benchmark page	Deployed	-
Performance: Prepare test design, test case, test procedure specifications	1.1-4	See ALMA benchmark page	Deployed	-
Performance: Benchmark script preparation	1.1-4	See ALMA benchmark page	Deployed	-
Performance: Benchmark platform/environment preparation	1.1-4	See ALMA benchmark page	Deployed	-
Image Plane continuum subtraction	5.2-5	See Image module documentation	Deployed	Ready
uv-plane continuum subtraction	5.2-5	See MS module documentation	Deployed	Ready
calibrater: scan-based solution intervals	4.1-3	See calibrater module documentation	Deployed	Ready
calibrater: solution intervals referenced to scan boundary	4.1-3	See calibrater module documentation	Deployed	Ready
viewer: blinking prototype	7.1-3	Demo'd. Full implementation targeted for next release	NA	-
imager: multi-spectral window imaging	Infra	See imager module documentation	Deployed	Ready



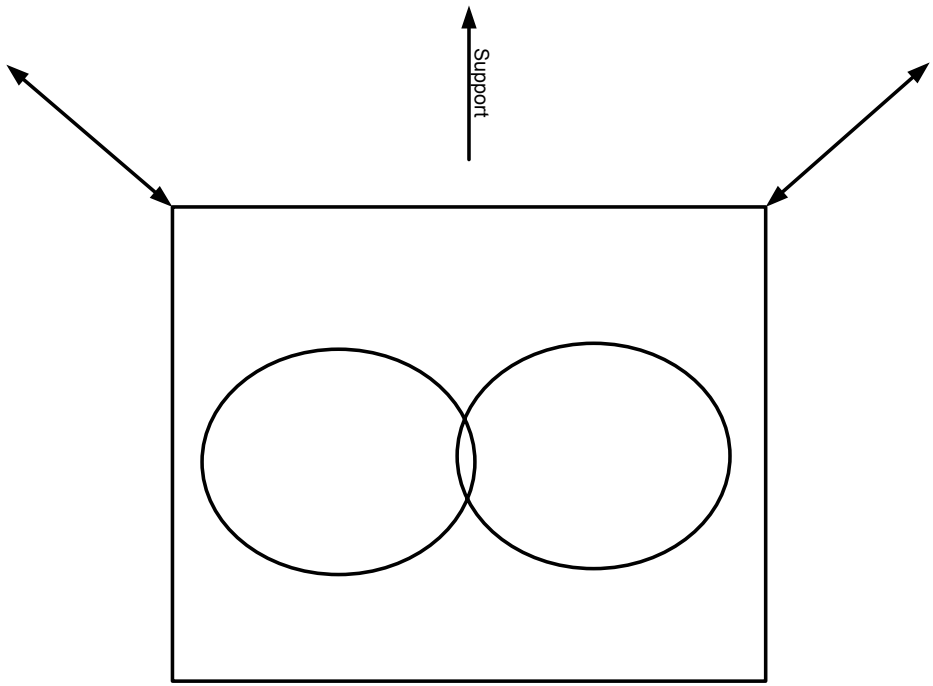
Project Office

Project Home | Project Book | Decision Log | Change Log | Change Proposals

User Documentation | Submit Request

ACTIVE THIS CYCLE	Project Name	Program	Time Span	Details	Project Documents	References and Memos
X	Performance Improvements	Benchmark	8/1-10/1	Investigation/implementation of I/O performance issues.	Project Development Plan	Benchmarking in AIPS++; ALMA Benchmark tests
X	Robustness	Robustness	8/1-10/1	All developer's have 25% time allocation for testing/defect resolution. New functionality must have associated assay/unit tests for automated assessment. Internal testing, coordinated through Project Scientist, will take place at the end of each project Stable Snapshot Cycle.	Project Development Plan	Project Testing Plan
X	Functionality	ALMA,EVLA	8/1-10/1	• msconcat enhancements for sd/synth combination • viewfiller enhancements [4.2-2] • MS concatenation/split improvements [3.1-20.1:5] • viewer: blinking implementation [7.1-3] • viewer: add parameters to cursor readout [7.1-5] • viewer: write position information to logger [7.1-5] • calibrater: correction for specified zenith opacity [4.2-3] • calibrater: integrated tramcalibrater functionality [infra] • Add Local Group, Rest to supported velocity frames [3.1-12.8, external] • Single Dish/Synthesis data combination [5.2-6; external]	Project Development Plan	ALMA AIPS++ Audit
X	Design	Infrastructure	8/1-10/1	Measurement Set Revision/Specification [infra] Concatenate/Split Design [3.1-20.1:5]	Project Development Plan	Split Proposed Design-
X	Prototype Pipeline	Pipeline, e2e	8/1-10/1	Phase A Completion Phase B Commencement	Project Development Plan	Prototype Pipeline Plan
	Common Framework	Common Framework	12/03 -	Details TBD	Project Development Plan	Design Pending.

Who: Off-line Organization



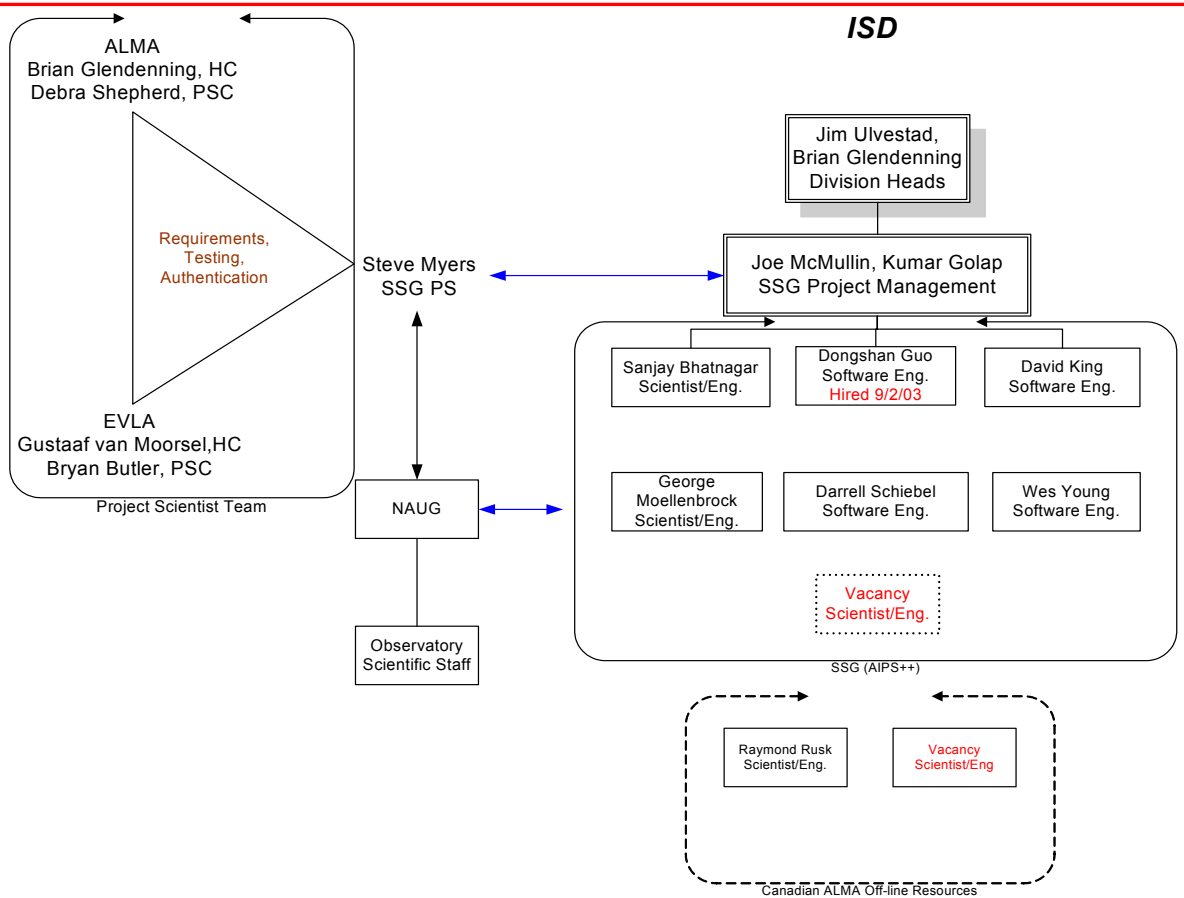
Project-Driven!

Tighter Structure

Better coordination, communication, collaboration.

More observatory-wide involvement.

Who: Offline Resources



SSG:

Development

5 FTEs (NRAO)

1 FTE (grant)

1 (+1) ALMA

Project Scientist

0.25 FTE

Testing

~1 FTE

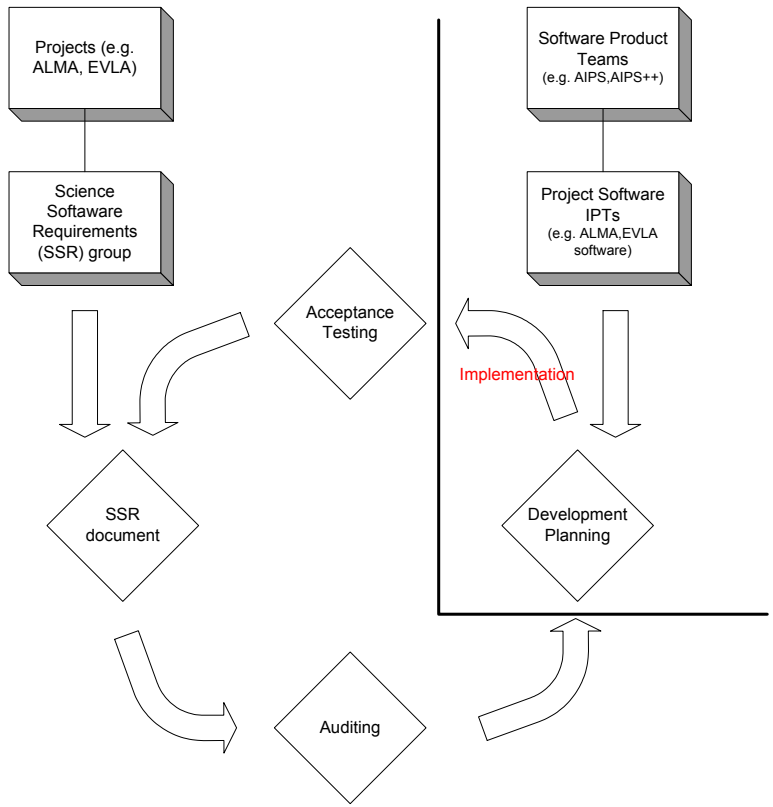
How: Process

Short Development Cycle
- 2 month development cycle
- Robustness: 25% time allocated for defects/testing per developer.

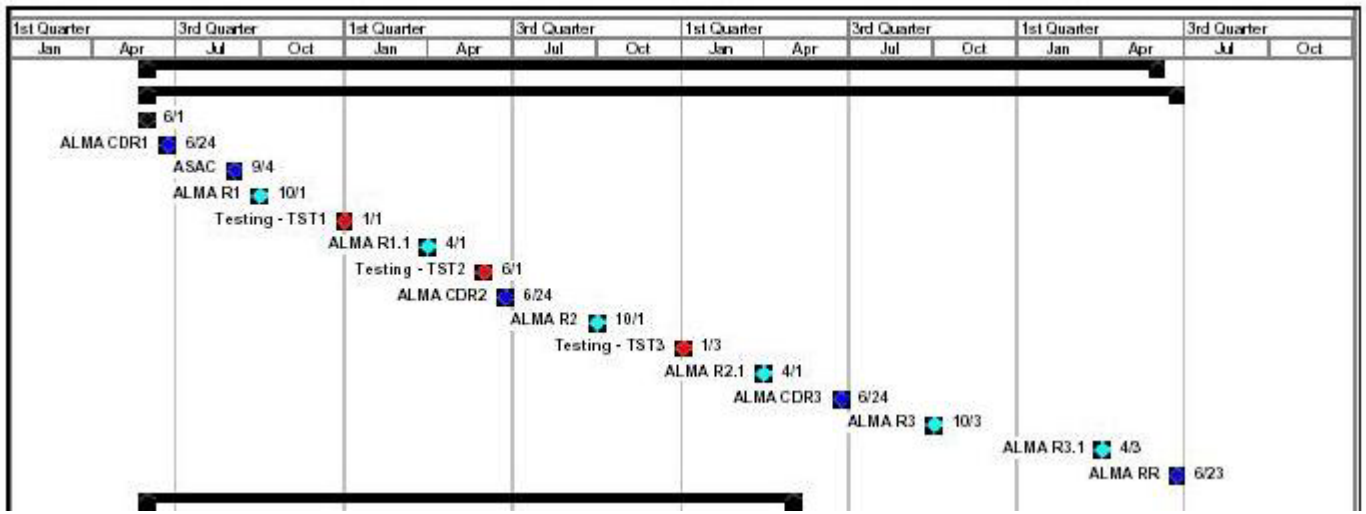
Coordination
- Subsystem scientist=in-house customer
- Feedback at all stages of development process from design to implementation

Testing
- Project level
  - developer (function tests)
  - project management (confirms)

  Stable declared
- NAUG integrated testing
- SSR integrated testing



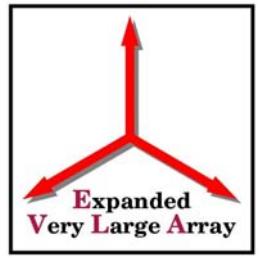
When: Milestones



- SS1 -- Aug 03 (Complete)
- CDR1 -- Jun 03 (Passed)
- ANASAC -- Aug 03 (Complete)
- ESAC, ASAC -- Sep 03 (Complete)

- R1 -- Oct 03
- TST1 -- Jan 04
- R1.1 -- Apr 04
- CDR 2 -- Jun 04

Issues/Questions



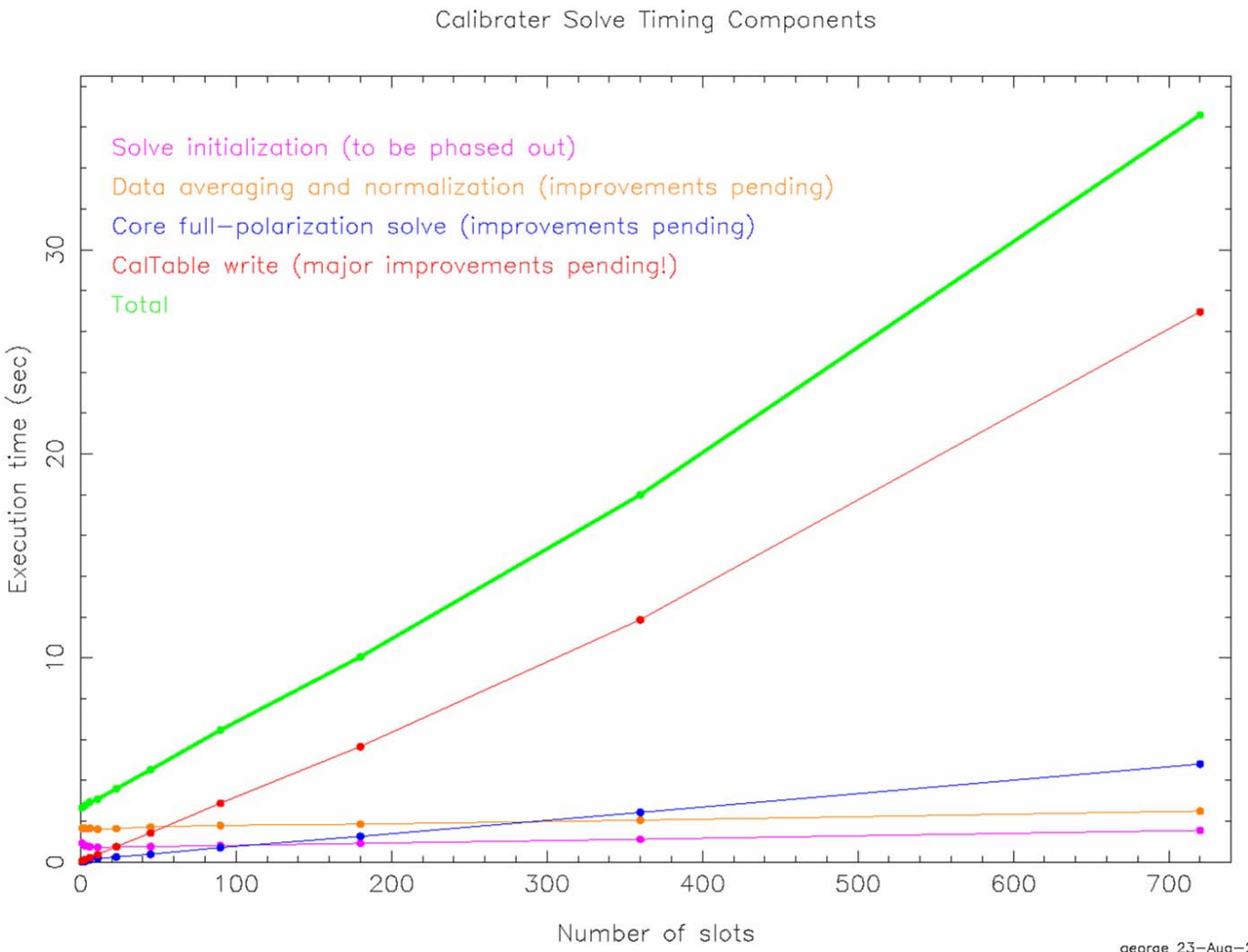
- Resource Allocations
  - algorithm development
- Code base framework change
- Organizational changes

Summary

- NRAO has altered its computing organization to better serve the projects – projects drive development, direct management structure.
- New structure facilitates involvement of the whole observatory (scientific staff, computing groups (AIPS, AIPS++, others), etc).
- EVLA is poised to take advantage of ALMA's lead in software development.
- Substantial improvements in performance, functionality; predictability in deliverables.

- SSG Project Office Page (Plans, Designs, Status, Development)
- http://projectoffice.aips2.nrao.edu.
- AIPS++ Technical Review
- http://projectoffice.aips2.nrao.edu/aipstr/aipstr.html
- ALMA CDR 1
  http://projectoffice.aips2.nrao.edu/almacdr.html
- ALMA Benchmark Page
  http://shiraz.drao.nrc.ca:8080:AlmaDRPBenchmarks

Performance: Calibrater



AIPS++/IRAM Comparisons

- Phase 1: Can AIPS++ reduce real mm wave data? Is AIPS++ code base flexible enough to implement new algorithms to match IRAM processing?
  - Yes, but schedule was very extended
    - Partly underestimated effort, mostly priority setting
    - ALMA/NRAO and EVLA now directly manages AIPS++
      - And ALMA has complete control of priorities until CDR2
- Phase 2: Can new users process similar but new data on an accelerated schedule (3 weeks)?
  - *Fill, Calibration, Editing:* generally, yes
  - *Imaging:* only experienced aips++ users could produce images - too difficult, some defects encountered & there was not adequate time to resolve problems.
- Phase 3: Performance