SCARIe FABRIC: Building a distributed software correlator for e-VLBI

Nico Kruithof, Yurii Pydoprihora,
Mark Kettenis, JIVE
Marcin Okon, Dominik Stoklosa, PSNC
Damien Marchal, University of Amsterdam

Starting Point

C++ implementation of the SFXC correlator used for the Huygens probe tracking.

Algorithm developed by Sergei Pogrebenko.

- FX correlator
- Narrow-band, high spectral resolution
- Fractional bit-shift before fringe rotation

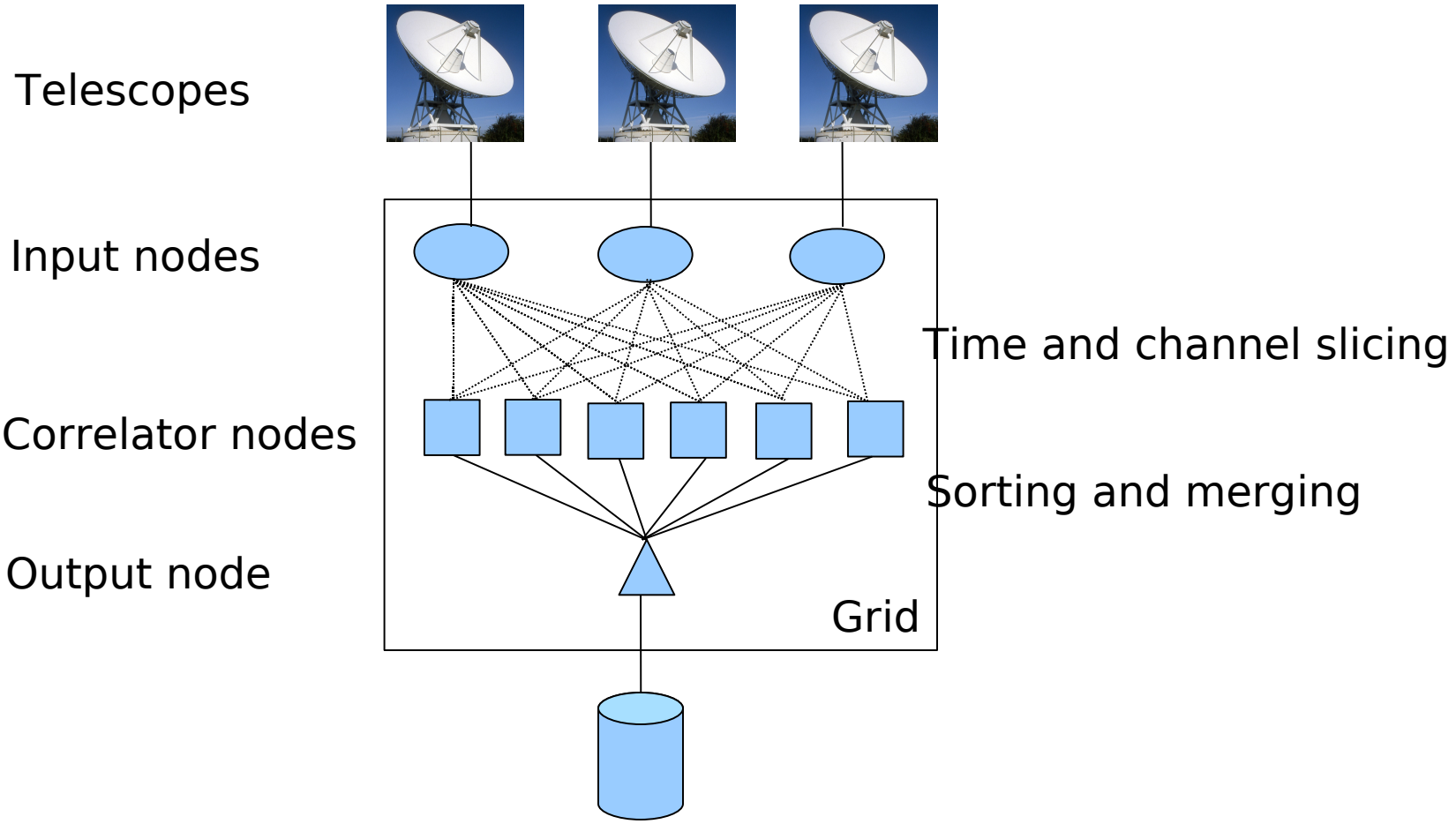
- Largely unoptimized
- Correlator core only

Design Decisions

- Parallelisation in both subbands and time.
- All baselines for a single subband processed on a single cluser node.
- Further course grained parallelisation (in time only) to distribute over multiple clusters.

- Input node (one per station)
- Correlator node (as many as feasable)
- Output node
- Manager node

Design Decisions

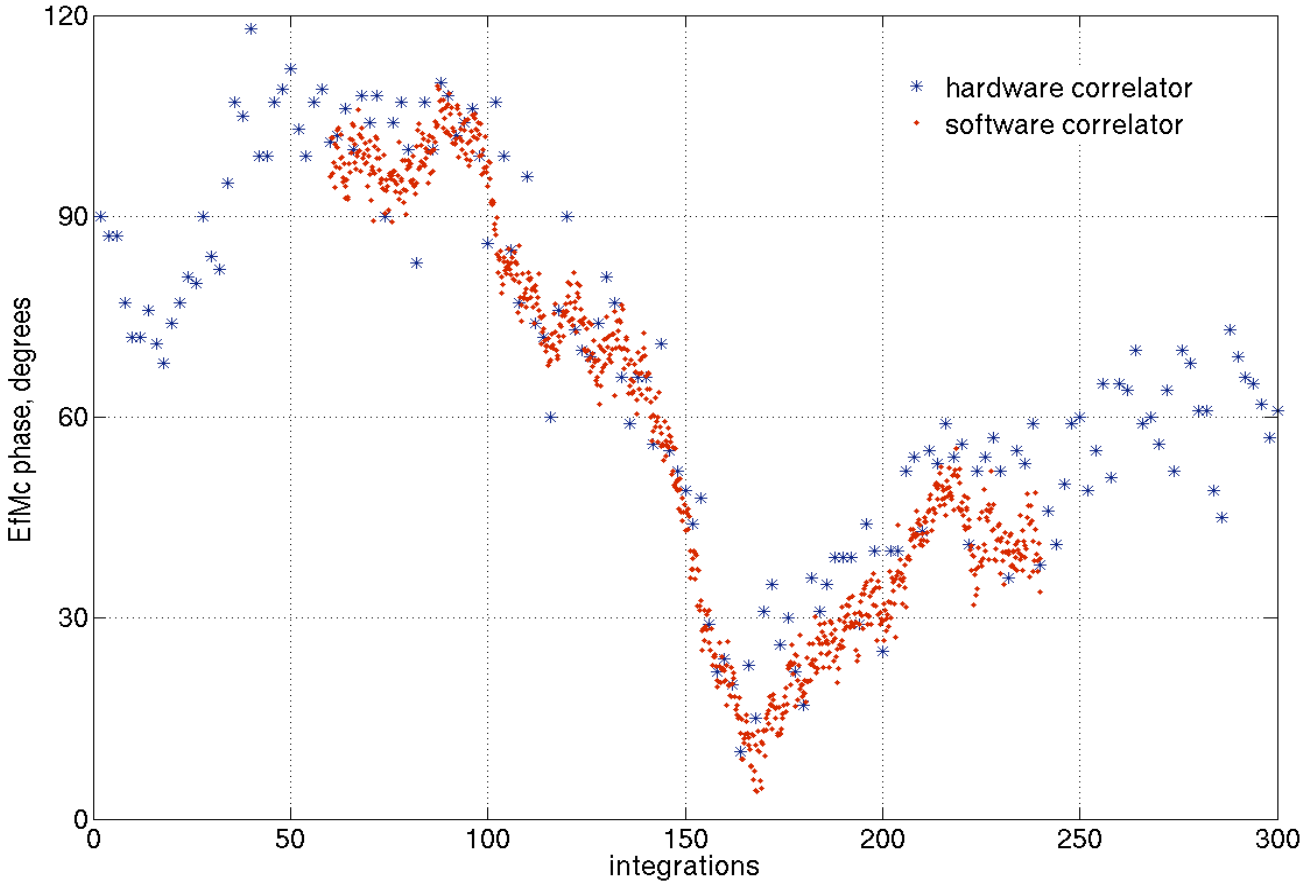


Current State

- VEX-driven, with small JSON-based control file specifying correlation parameters.
- Modular.
- Parallelized using MPI.
- Scales from SMP machines to largish clusters.
- Integrated delay model based on CALC 10.
- All subbands processed, both LSB and USB, cross-correlations.
- Takes Mark4 (Mark5A) and Mark5B input data.
- Output data format that includes (some) metadata.

Validation

- Comparison with the Mk4 Hardware Correlator

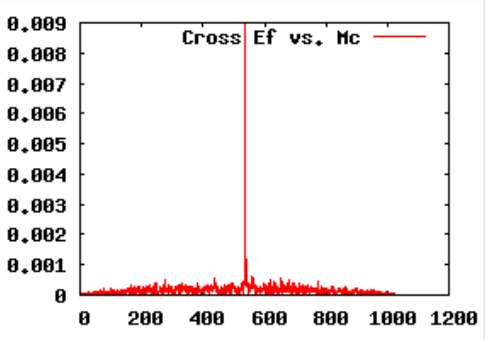


FTP Fringe Tests

- Cluster running NICT correlator broke down in May 2007
- SFXC was used ever since

Show plots

	Auto correlation				Cross correlation					
	Ef	Nt	Mc	On	Ef-Nt	Ef-Mc	Ef-On	Nt-Mc	Nt-On	Mc-On
CH01, 2203.99 MHz, USB, Rcp	A	A	A	A	34.02	45.62	17.38	20.74	10.95	12.62
CH02, 2219.99 MHz, USB, Rcp	A	A	A	A	30.6	53.39	13.23	26.08	13.67	12.88
CH03, 2243.99 MHz, USB, Rcp	A	A	A	A	29.94	19.99	16.44	13.75	10.98	4.723
CH04, 2283.99 MHz, USB, Rcp	A	A	A	A	25.51	53.06	14.29	23.27	12.13	9.903
CH05, 2299.99 MHz, USB, Rcp	A	A	A	A	35.51	4.39	13.7	2.677	13.71	3.064
CH06, 2331.99 MHz, USB, Rcp	A	A	A	A	2.304	41.2	11.97	2.685	2.646	12.37
CH07, 8211.99 MHz, LSB, Rcp	A	A	A	A	86.87	144.1	37.35	26.65	5.897	10.25
CH08, 8211.99 MHz, USB, Rcp	A	A	A	A	112.5	208.6	46.89	33.62	6.784	12.25
CH09, 8227.99 MHz, USB, Rcp	A	A	A	A	105.7	209.4	41.66	33.51	5.963	12.1
CH10, 8251.99 MHz, USB, Rcp	A	A	A	A	2.694	4.214	2.609	31.76	5.27	12.23
CH11, 8315.99 MHz, USB, Rcp	A	A	A	A	85.97	201	38.35	32.83	8.808	14.01
CH12, 8427.99 MHz, USB, Rcp	A	A	A	A	103.9	192	42.19	38.02	7.186	10.66
CH13, 8499.99 MHz, USB, Rcp	A	A	A	A	92.58	199.3	42.48	37.24	6.661	13.43
CH14, 8539.99 MHz, USB, Rcp	A	A	A	A	80.98	201.6	44.96	28.93	8.184	14.69
CH15, 8579.99 MHz, LSB, Rcp	A	A	A	A	33.49	141.8	35.61	8.895	3.097	12.05
CH16, 8579.99 MHz, USB, Rcp	A	A	A	A	11.14	182.4	42.42	3.811	3.607	14.92



Post Processing

- JIVE uses j2ms2 to convert hardware correlator output into MeasurementSet.
- j2ms2 has been adapted to accept the SFXC output format.

Benefits:

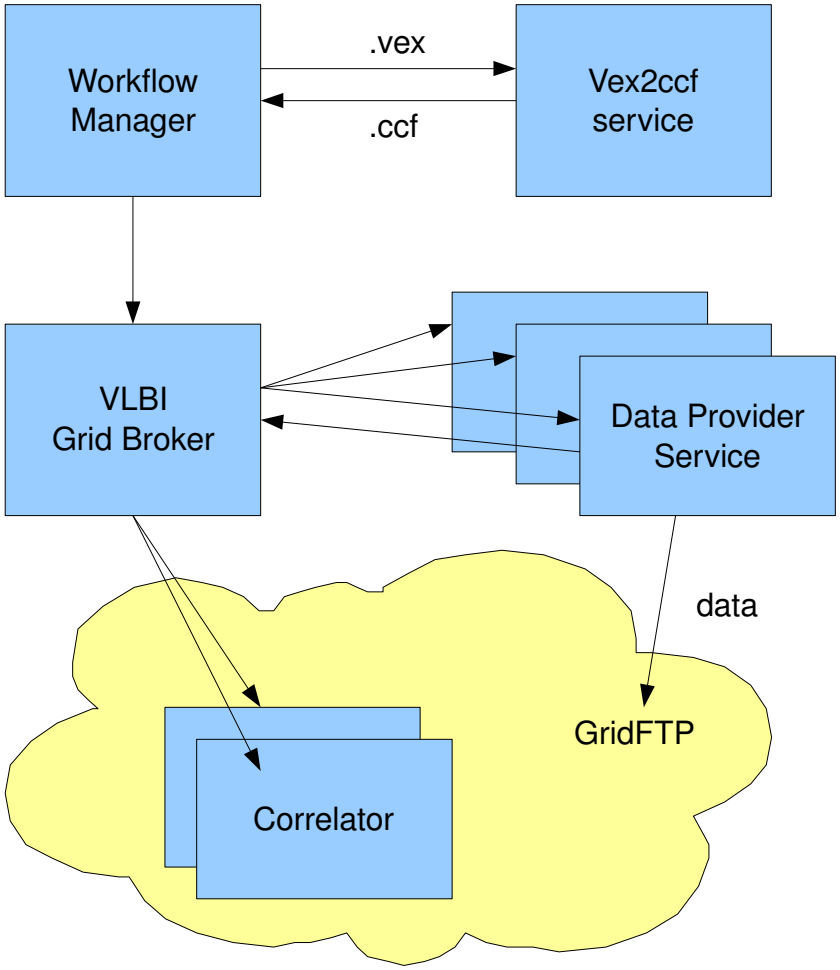
- Standard analysis tools can be used.
- Translation into FITS through standard pipeline (skipping some steps).

The Bigger Picture: FABRIC

How to get things on the Grid?

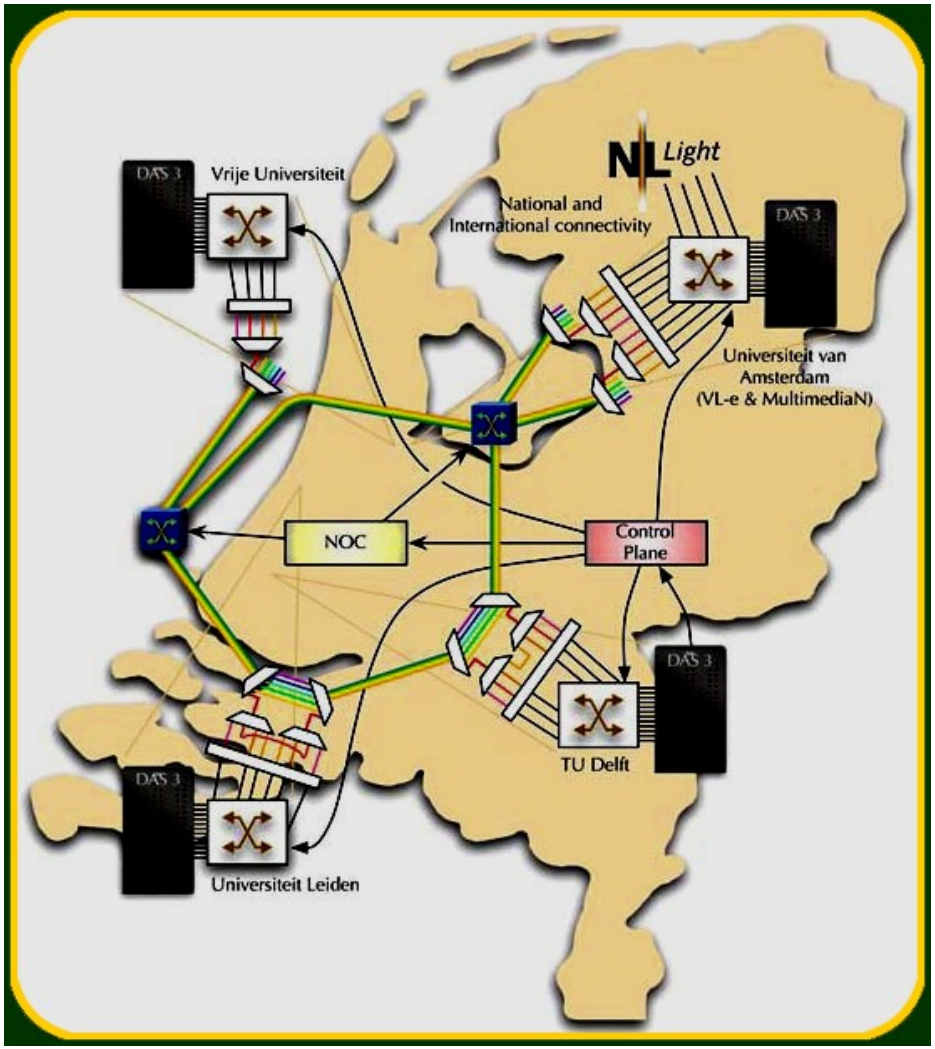
- Workflow Manager (PSNC)
- VLBI Grid broker (PSNC)
- Web Services (SOAP)
  - Vex2ccf: Creates control file template from VEX
  - Data providers: Provides chunks of data and transfers them to GridFTP servers.
- Staged correlation (uncorrelated data hits disk)

The Picture


Workflow Manager
.vex
.ccf
Vex2ccf service
VLBI Grid Broker
Data Provider Service
data
GridFTP
Correlator

Advanced Networking: SCARIe

- DAS-3 cluster
- Distributed over 5 locations
- Multiple 10 Gb/s access links
- 1 Gb/s ethernet interconnect
- 10 Gb/s Myrinet interconnect
- Starplane: User-controllable optically switched network (DRAC) between locations.



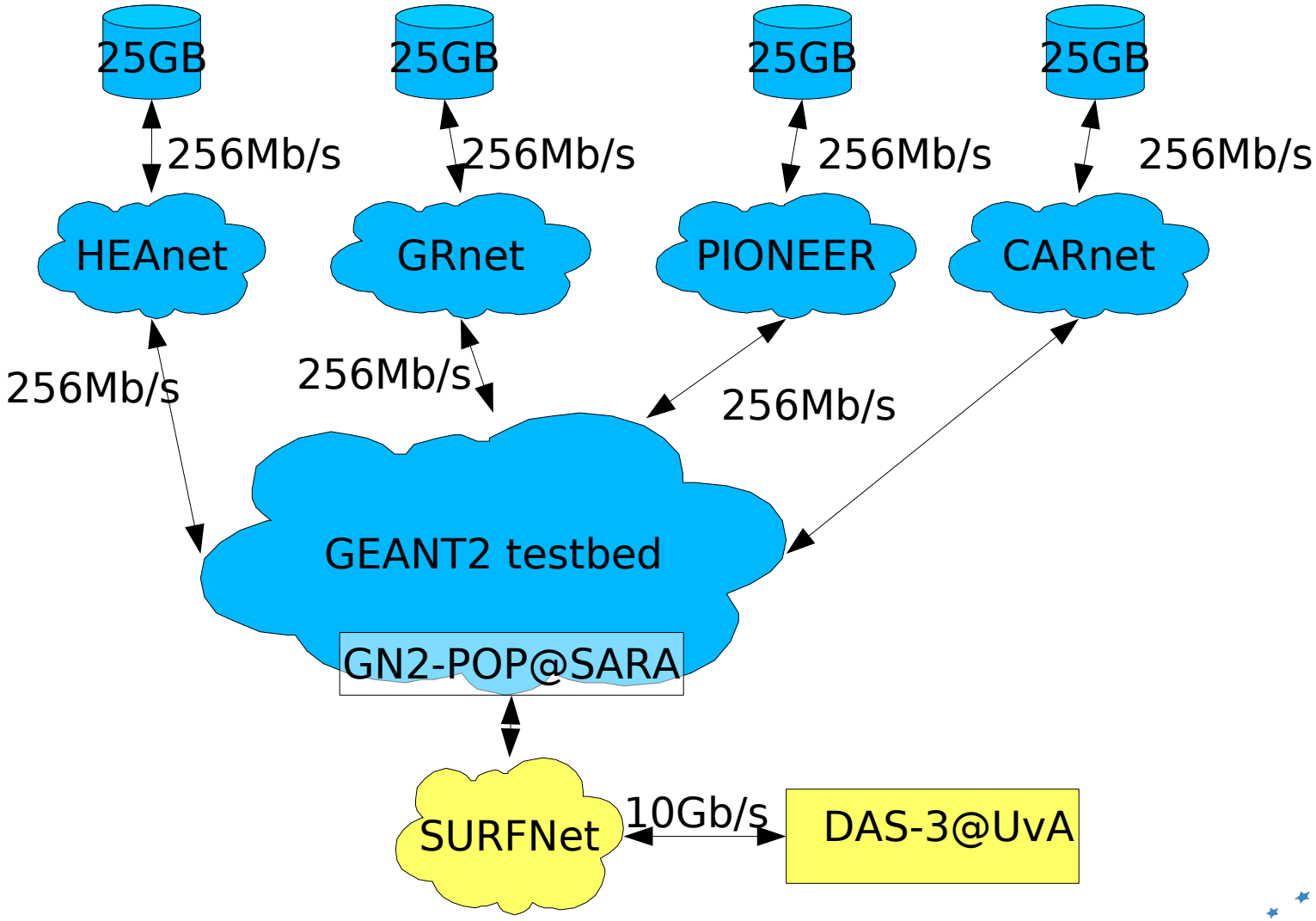
Benchmark

- 60 dual-core, dual-CPU machines
- 4 stations, 256 Mb/s, 1024 spectral channels per subband at 54% of real time
- Bottleneck: Conversion of input data
- Dedicated input nodes effectively use only 1 core

AutoBAHN

- JRA within GEANT2 devloping a bandwidth-on-demand facility across domains
- Being deployed by NRENs of Greece, Ireland, Poland and Croatia, using the GEANT2 testbed
- Interfaces to other BoD-systems being developed (Internet2, DRAC).
- Two ways of making reservations:
  - Interactive through web interface
  - Using a webservice

AutoBAHN demo



Conclusions

- Argument reduction for (co)sine functions is computationally expensive.
- Interoperability of web services is a major issue.
- Bandwidth-on-demand looks promising.