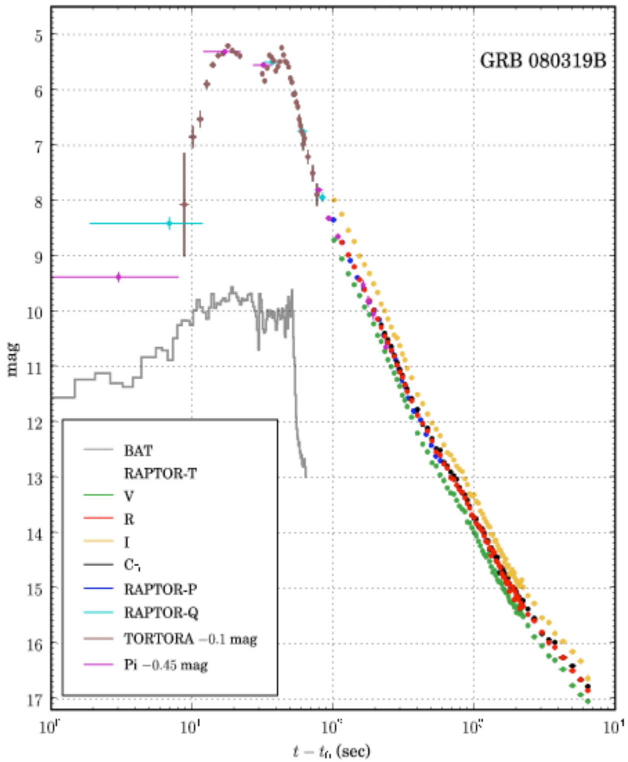
Data Challenges in Robotic Astronomy and Telescope Networking

Przemek Wozniak

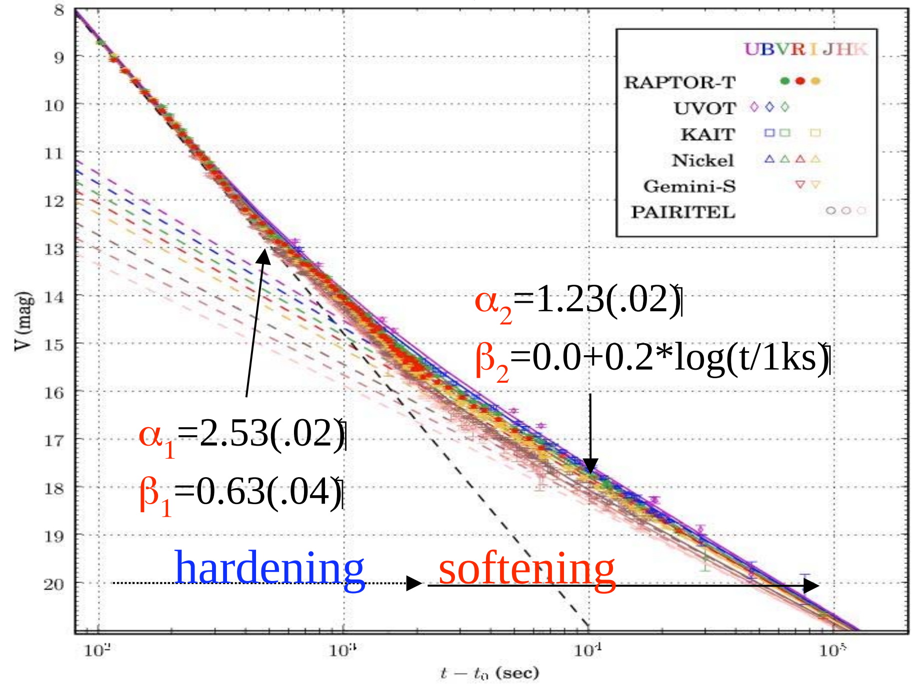
Los Alamos National Laboratory

Raptor/Thinking Telescopes Project

GRB 080319b: a naked eye burst from redshift ~0.9 !



Wozniak et al. 2009, Bloom et al. 2009



- Modelling required (nearly) simultaneous observations from a diverse set of instruments
- Need for optical searches without high energy triggers

Motivation for Autonomous Observatories and Telescope Networks

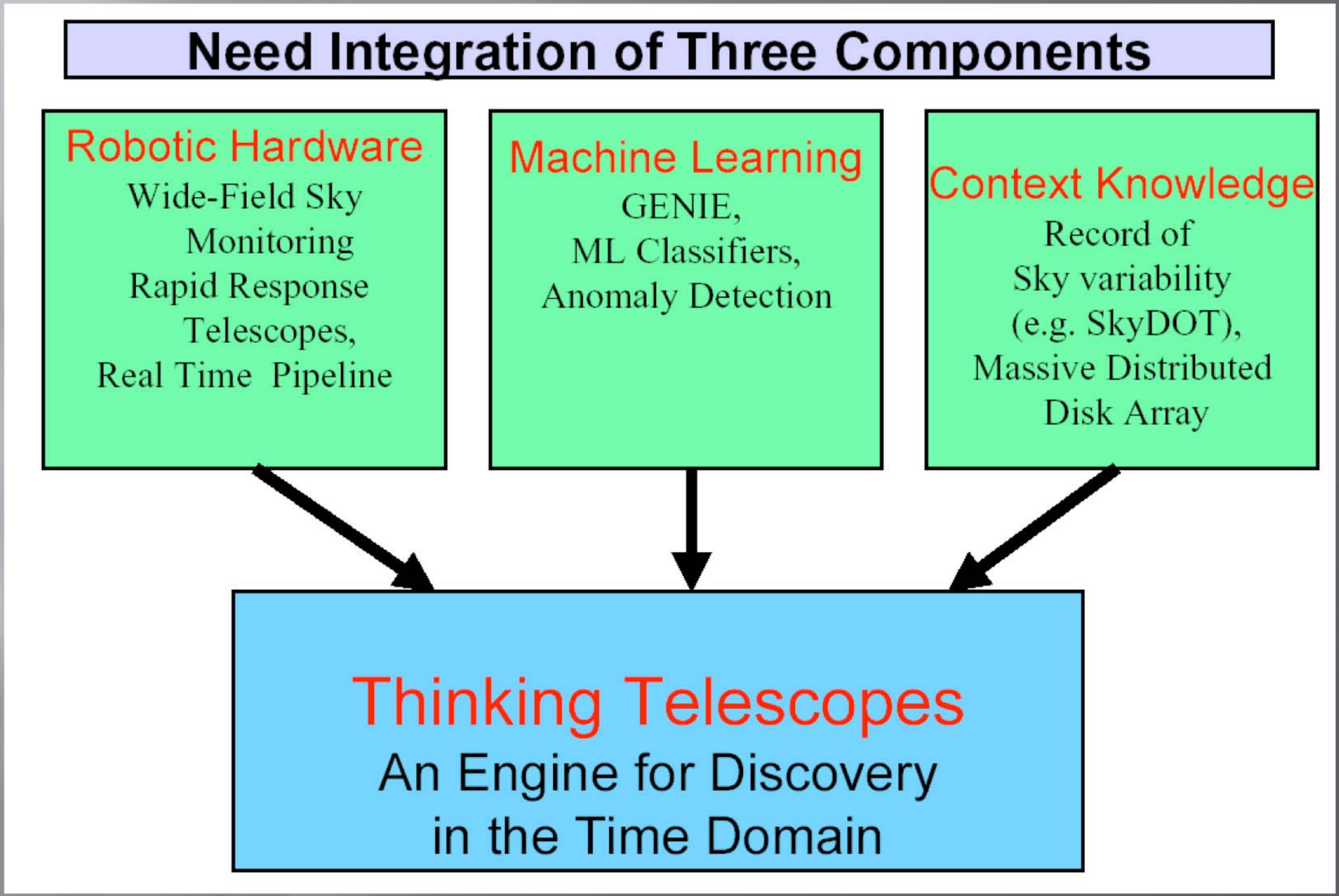
Streamline survey operations to minimize gaps in observing and maximize efficiency

Use coincidence observing to reject false positives in searches for explosive transients of cosmic origin

Optimize the interaction between surveys and follow-up in the regime of a large number of outstanding alerts

Take humans out of the loop to enable rapid response with very short reaction times

Deploy instruments to remote sites with minimal infrastructure support or in a hostile environment
(global sky coverage, avoid expensive travel)



Need Integration of Three Components
Robotic Hardware
Wide-Field Sky Monitoring
Rapid Response Telescopes,
Real Time Pipeline
Machine Learning
GENIE,
ML Classifiers,
Anomaly Detection
Context Knowledge
Record of Sky variability (e.g. SkyDOT),
Massive Distributed Disk Array
Thinking Telescopes
An Engine for Discovery in the Time Domain

Toward Continuous Sky Monitoring

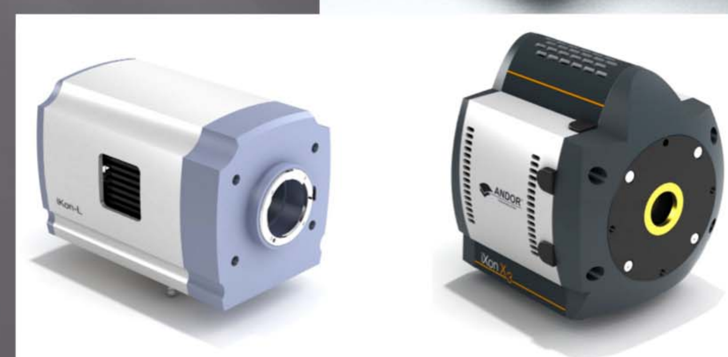
Robotic telescopes around the world

Aperture (m)	# telescopes	Fraction
< 0.25	93	41.5 %
0.25—0.50	66	29.5 %
0.50—0.75	15	6.7 %
0.75—1.00	25	11.2 %
1.00—1.25	7	3.1 %
> 1.25	18	8.0 %

COTS components:
cheaper and easy to use detectors
have a potential to drown us in data

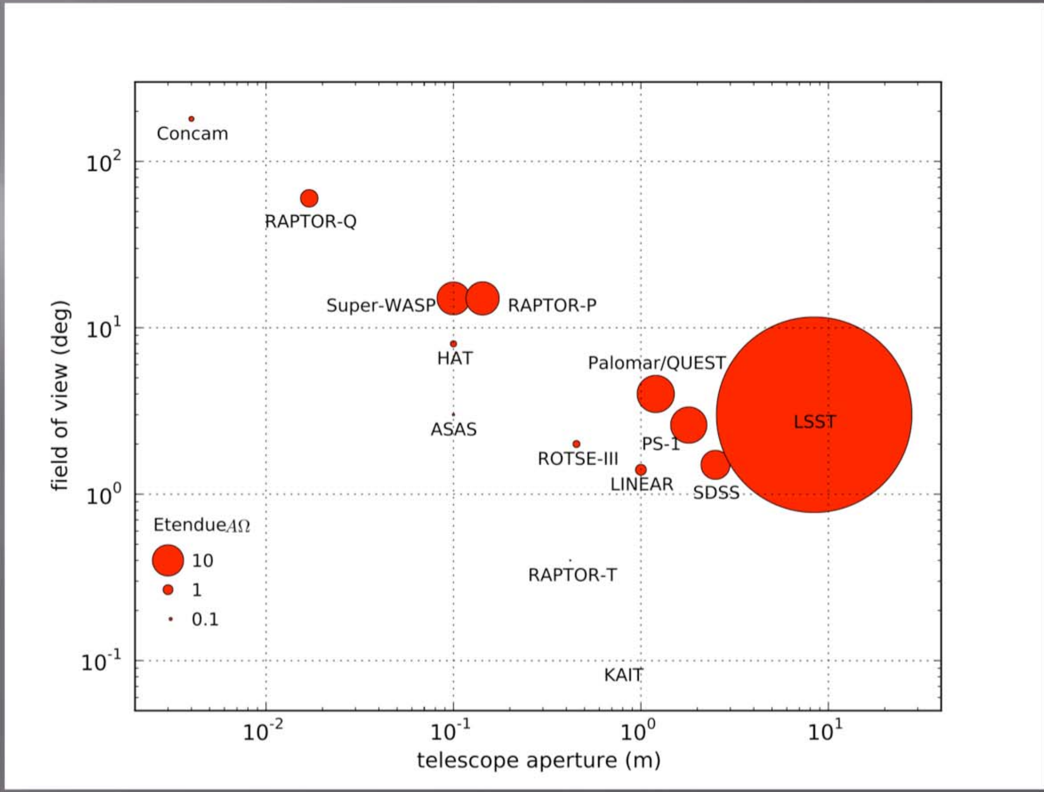
Science with robotic telescopes

Primary purpose	# telescopes	Fraction
Gamma-Ray Bursts	31	20.4 %
Service observations	26	17.1 %
Education	20	13.2 %
Exoplanet searches	18	11.8 %
Photometric monitoring	14	9.2 %
All-sky surveys	12	7.9 %
Supernovae search	10	6.6 %
Asteroids	8	5.3 %
Spectroscopy	4	2.6 %
Astrometry	4	2.6 %
AGN, Quasars	4	2.6 %
(Micro-)Lensing	1	0.7 %



Small Telescopes, Large Etendue

Instrument	Aperture	FOV	$A\Omega$	N_{tel}
	m	deg	m^2deg^2	
CONCAM	0.004	180.0	0.3	1
Super-WASP	0.100	15.0	11.1	8
ROTSE-III	0.453	2.0	0.5	1
RAPTOR-P	0.143	15.0	11.3	4
RAPTOR-T	0.420	0.4	0.02	1
RAPTOR-Q	0.017	60.0	3.1	5
ASAS	0.100	3.0	0.06	1
HAT	0.100	8.0	0.4	1
LINEAR	1.000	1.4	1.2	1
SDSS	2.500	1.5	8.7	1
Palomar/QUEST	1.200	4.0	14.2	1
PS-1	1.800	2.6	13.5	1
LSST	8.400	3.0	391.7	1
KAIT	0.800	0.1	0.004	1



Przemek Wozniak
Los Alamos National Laboratory

All-sky monitors

- Bottom of the food chain, but easy to replicate
- Complete observatory in a box

CONCAM

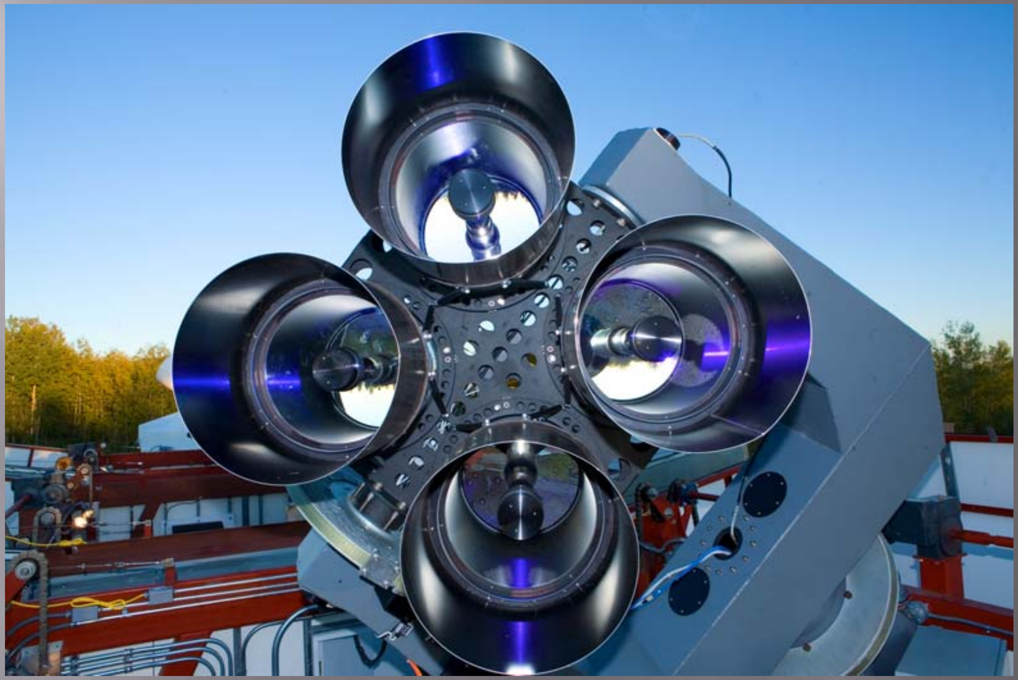
RQD2

- 5 2k x 2k frames every 20 s
- 0.4 TB of raw data per week

RAPTOR-K survey system

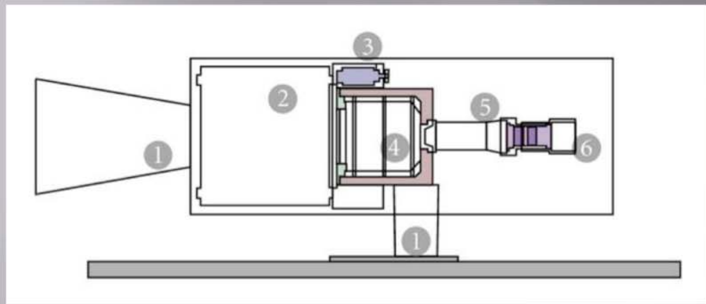
- 16-camera survey system covering a set of 1000 sq. deg to 16.5 mag
- Co-existing on a fast slewing mount
- A “circuit survey” looping over ~10 tiles covering most useful sky
- Generates 134 MB of data every ~30 seconds, or 0.5-1.0 TB per week
- Software automatically reduces data on-the-fly before the next exposure
- Anomaly detection constantly runs on ~500,000 light curves
- Events reported/response initiated in real time
- Prototype Raptor-P system is taking data

RAPTOR-T response array



- Four co-aligned 0.4 m telescopes
- Simultaneous multi-color imaging in VRI and Clear
- Fast-slewing mount can point anywhere in ~ 8 seconds
- Sensitive down to ~ 19 mag in 60 seconds
- f/5, 1k ×1k CCD, 24 arcmin FOV
- Back-illuminated Marconi CCD47-10 chip with 13μ pixels

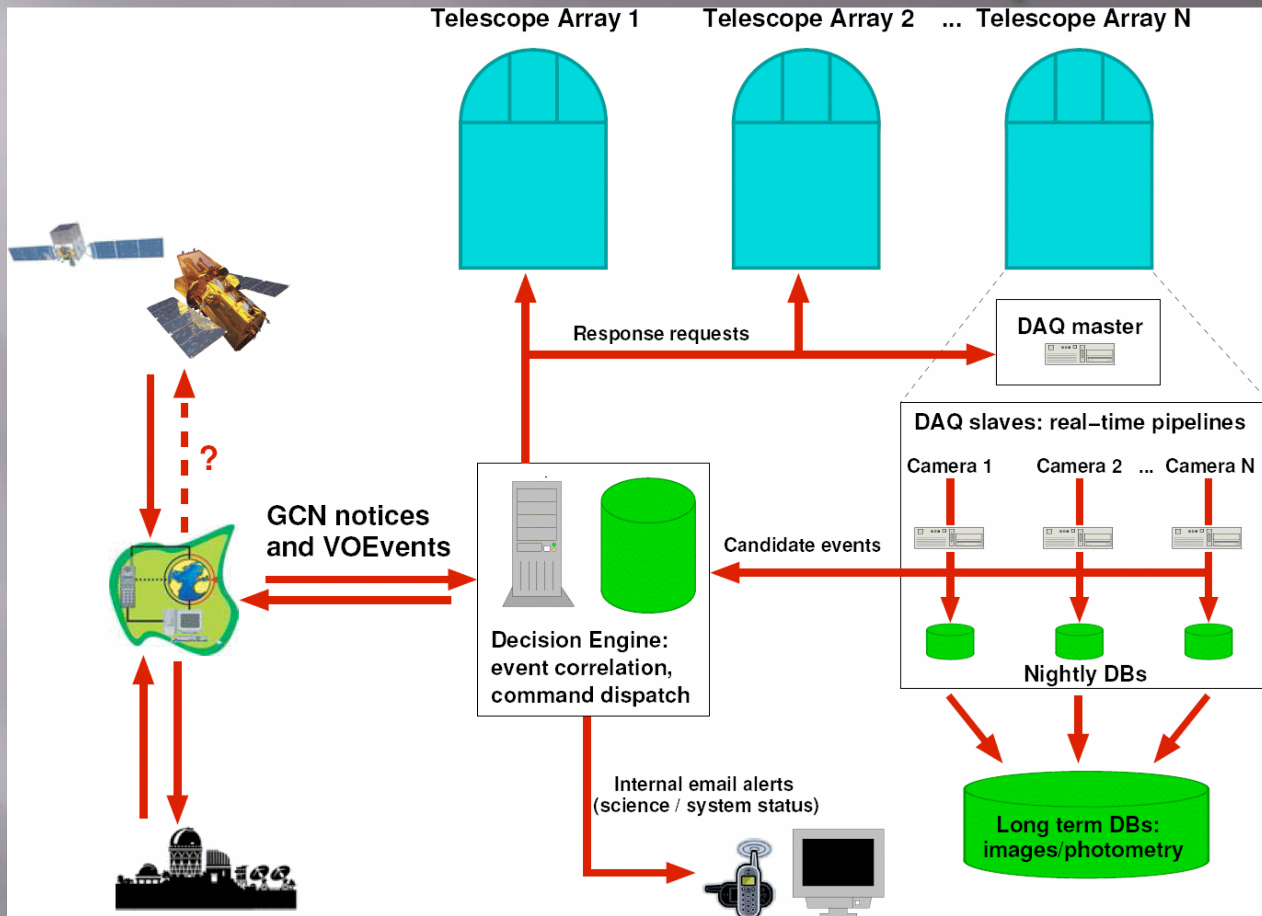
Ultra-fast Photometry



RAPTOR-Z
- 0.5-m telescope
- EMCCD can run at 10Hz or higher cadence with very little read noise

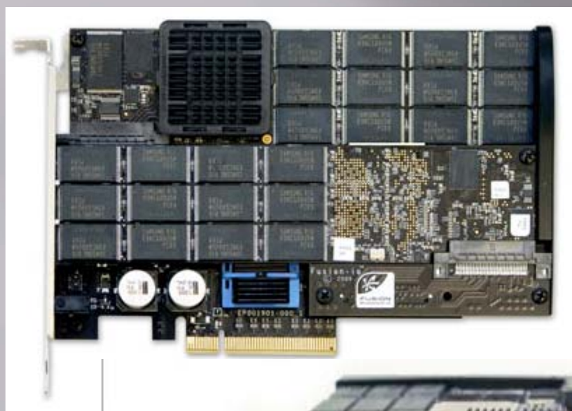
TORTORA
- 0.128 s exposures + 0.005 s gaps
- Terabytes of data per day

RAPTOR Autonomous Telescope Network



- Data transport using portable RAID arrays
- Local buffer storing a few months of data
- SQL databases with spatial HTM extension (Postgres)
- Web interface for remote data access

New Storage Technologies



- Limitations of network bandwidth drive computing and storage to the observing site
- Data intensive HPC systems at remote sites are subject to severe constraints on power, size, and robustness
- New generation Flash based SSD storage/compute boards such as Fusion-IO offer random data access at speeds 1000 times higher than HDD

But existing database systems are designed to hide the latency of disks and largely waste this performance

Communication

- Who
- What
- Where
- When
- Why
- How
- References
- Retractions

GCN (GRB Coordinates Network)

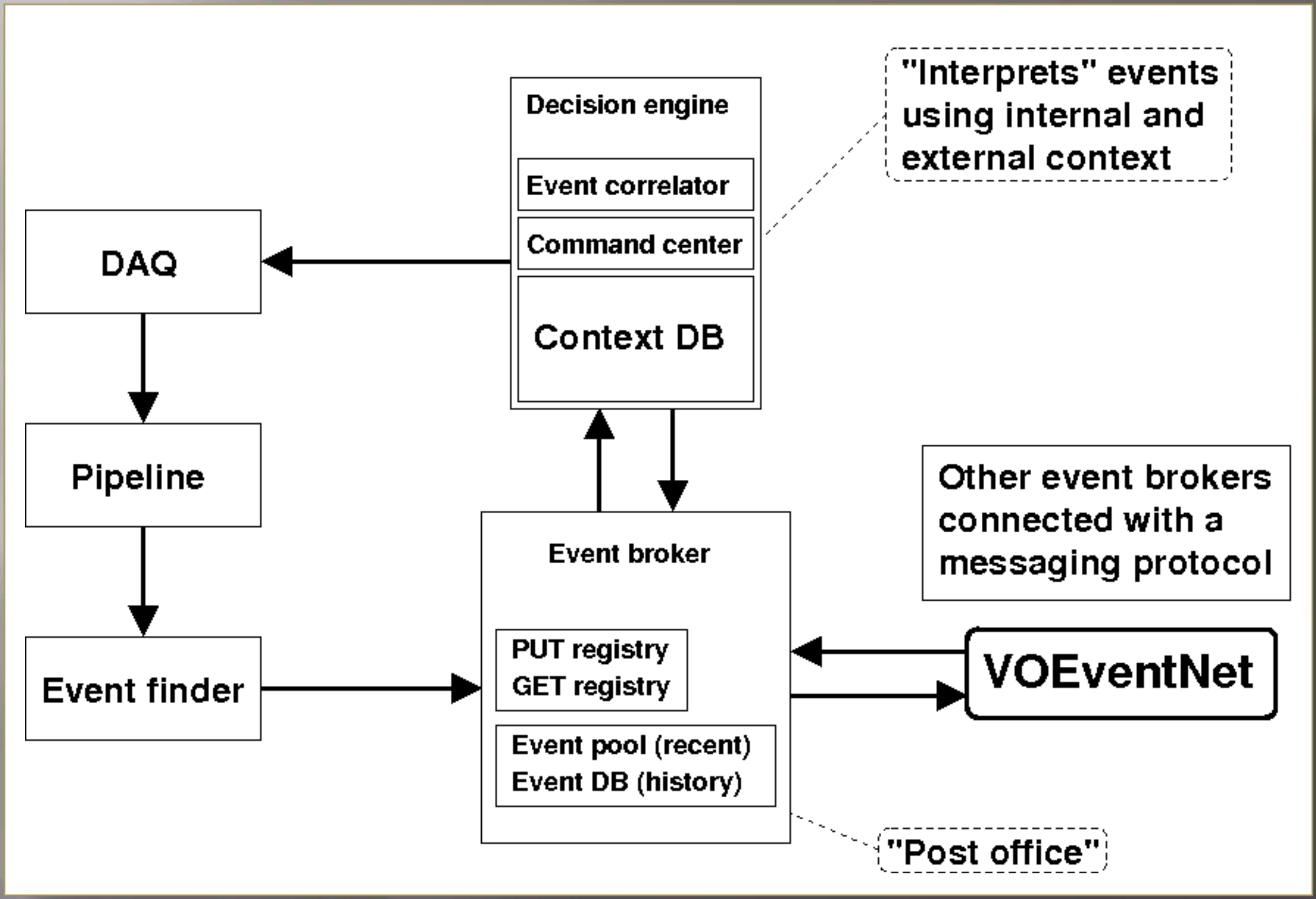
XML based IVOA standards:
http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaVOEvent
VOEvent
VOEventStream
VOEventServer
VOTimeSeries

Skyalert http://www.skyalert.org/ (VOEventNet)

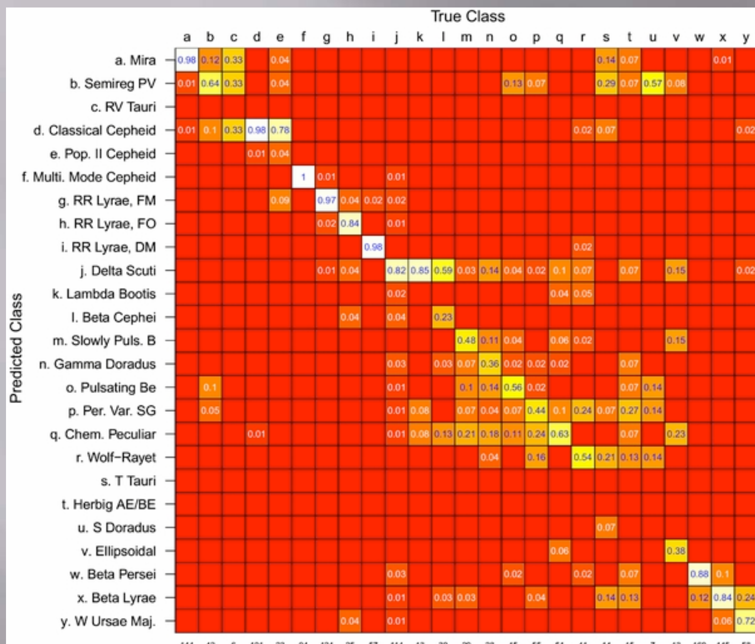
RTML (Remote Telescope Markup Language) is an XML dialect for controlling remote and/or robotic telescopes

Przemek Wozniak
Los Alamos National Laboratory

Aggregator Nodes and Event Brokers



LSST Classification Challenge



Richards et al. 2011
>20% error rate with
~25 types of variable stars

Transients must be selected and classified against the background of ~135 million variable stars (periodic and non-periodic)

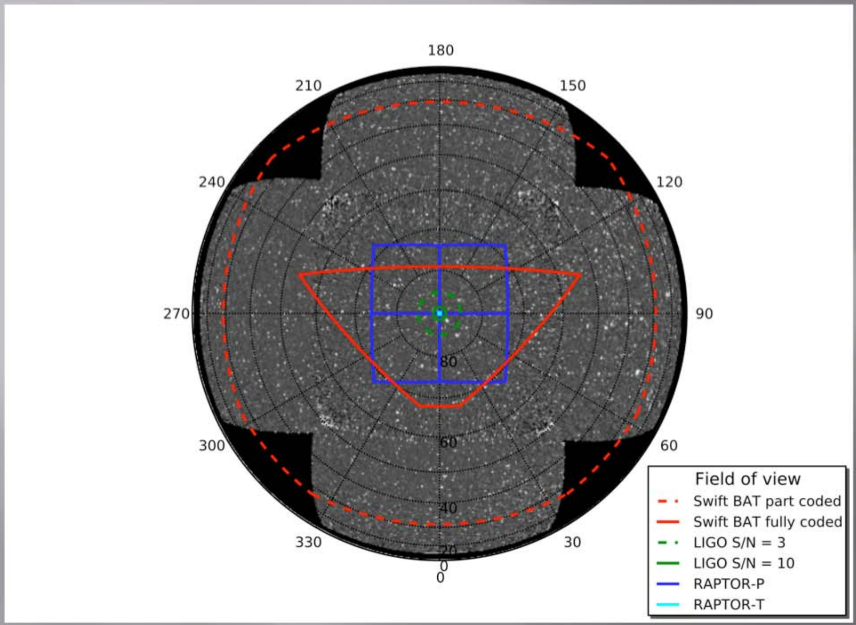
Existing algorithms cannot deliver the required classification accuracy

Transient classifiers mostly tuned for SN

~50,000 variability alerts per night will completely saturate the world’s supply of follow-up instruments

Need a comprehensive approach that can combine all available context

Real Time Cross-correlation and Follow-up Optimization



- Accurate real-time transient classification achieved by assimilating on the fly the required context information: multi-color time-resolved photometry, host galaxy and other environment data, broad-band spectral properties (gamma-rays, radio)
- High-level event classification using Bayesian belief networks
- Follow-up optimization using mutual entropy minimization (Mahabal et al.)

Need real-time data fusion on a global scale

The role of the VO ?

Summary

Autonomous instruments can deliver otherwise impossible observations

Time-domain surveys with relatively small telescopes are generating large amounts of data and face many problems of deep large scale surveys

Existing database systems are unable to utilize the random I/O performance of emerging Flash-based storage technologies

IVOA standards support machine and human readable representation of actionable event information and sharing of time sensitive data

Reliable real-time transient classification is an unsolved problem that must be addressed to make effective use of variability alerts from the next generation of massive photometric surveys

Need much better support for combining all available context from databases covering different energy ranges and spatial/temporal resolutions

Need much more sophisticated follow-up optimization algorithms