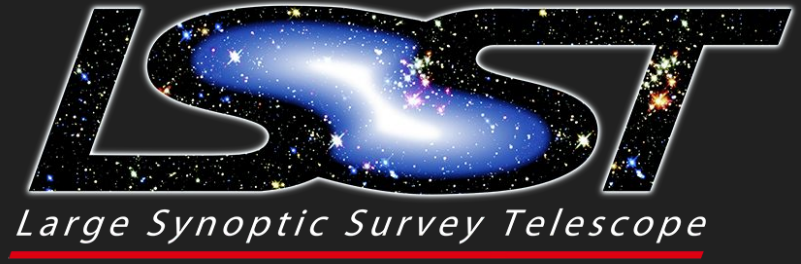




# From DES to LSST

## Transient Processing Goes from Hours to Seconds

Eric Morganson, NCSA
LSST Time Domain Meeting
Tucson, AZ
May 22, 2017

# Hi, I’m Eric

- Dr. Eric Morganson, Research Scientist, Nation Center for Supercomputing Applications (NCSA)
- Astronomer from Pan-STARRS1, SDSS and Dark Energy Survey
- Run DES transient pipeline at NCSA (with other research interests)
- Will be production scientist for LSST transient processing
- My first LSST meeting
- Inexplicably resemble Tycho Brahe

# NCSA: LSST Data Facility

- NCSA Astronomy Services Group
  - Runs DES and LSST Data Processing
- “Data Processing” includes
  - File transfer
  - Database administration
  - High performance computing coordination
  - Software maintenance
  - Archiving and serving the data
- LSST Data Processing requires
  - Daily interaction (transients, L1)
  - Major development for yearly (L2) releases
  - Coordination between astronomers, programmers and system administrators

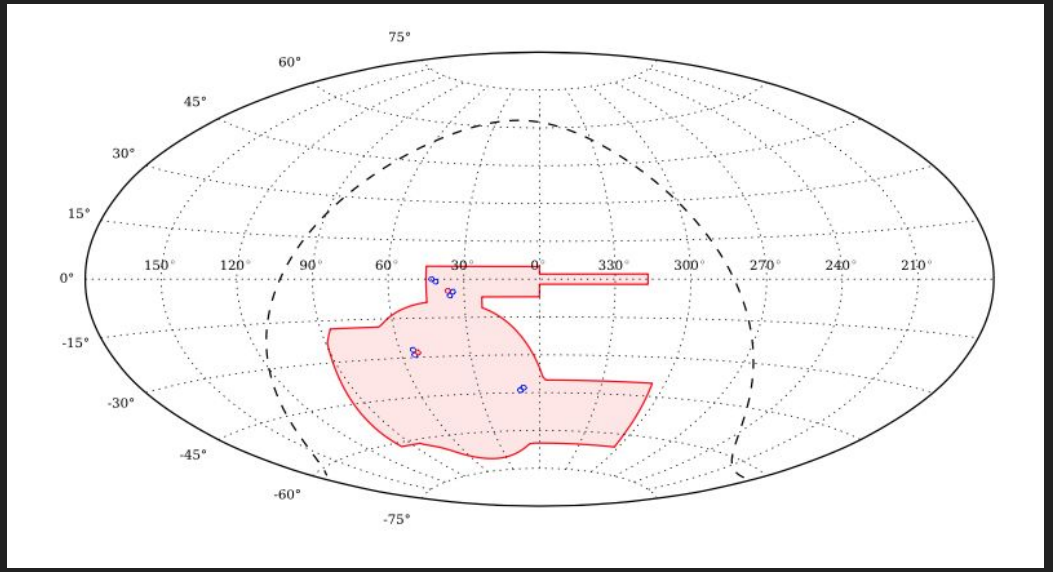
The National Center for Supercomputing Applications (NCSA) in Urbana, IL

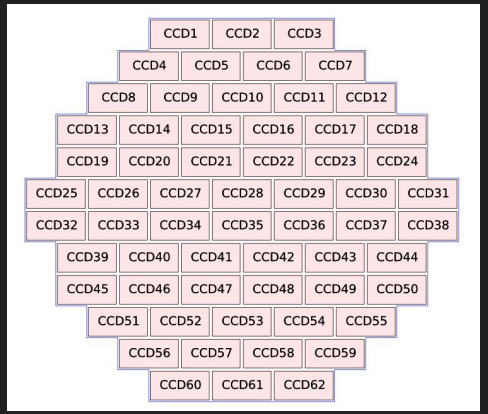
# Incidents Response: When Things go Wrong

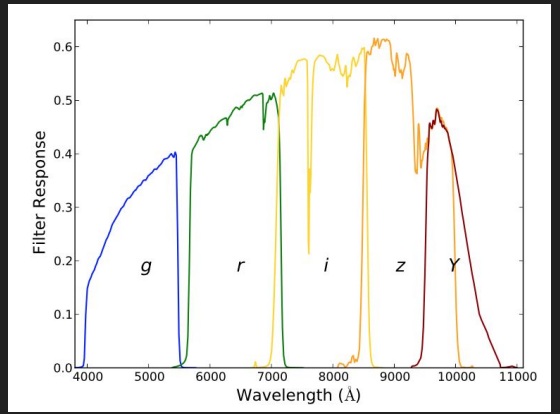
- Pan-STARRS and DES required 2-3 years to run smoothly
- Random things go wrong
- Pan-STARRS:
  - Shutter broke
  - Storm wiped out powerline
- DES:
  - Power surge caused massive data loss (no raw data lost)
  - Lightning caused DB outage
- LSST:
  - Network outages
  - Hardware/software problems
  - Operations team recover and mitigate

# Dark Energy Survey

- “Mini-LSST”
  - Like SDSS, PS1, ZTF, HSC
- 4m telescope, 3 Deg$^2$ FoV
- 5,000 Deg$^2$ Survey
- Includes 10 SN fields
  - Transient detection
  - Taken every 5 days
  - Pre-generated templates
  - 3 hour processing time
- Widefield transients
  - LIGO events
  - Not processed regularly
  - Templates “hand made”



# DES Software: Not Reinventing the Wheel

- 

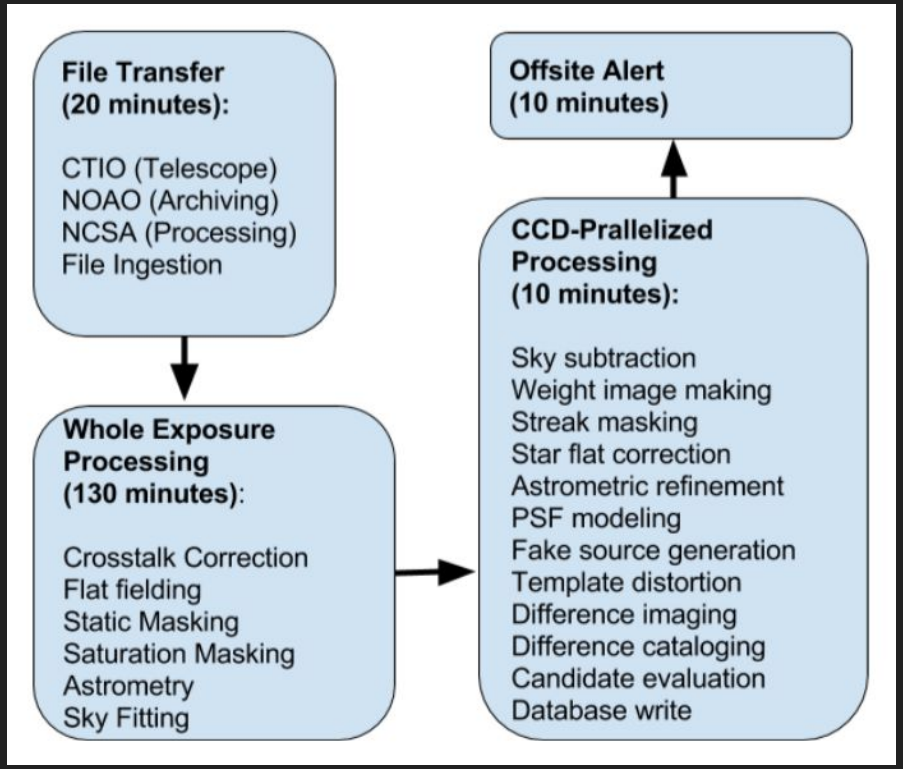
    - SExtractor: source extraction, catalog making
    - PSFEx: PSF extraction and modeling
    - SCAMP: Astrometric calibration
    - SWarp: Image regridding  for coaddition
    - Emmanuel Bertin (Astromatic author) is collaboration member
- HOTPANTS (Andrew Becker): difference imaging
- DES-developed algorithms and software
    - Flat fielding
    - Sky background modeling
    - Calibration
    - Multi-Object Fitting
    - Weak lensing analysis

# How is LSST different?

- Wide field transients
- Generating templates “on the fly”
- Devoted software stack
- Transient response time: 1 minute (not 3 hours)
- Where does 200x come from?

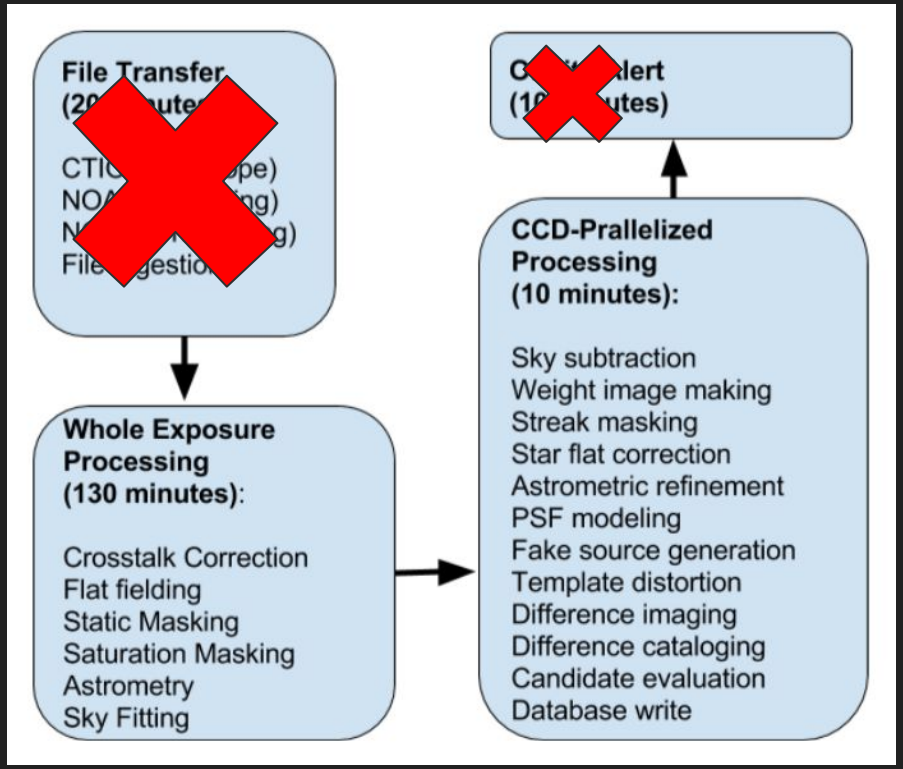
# DES Transient Pipeline:  3 Hours

- DES wants Type 1a Supernovae
  - Follow-up within 1 week desirable
  - Can obtain redshift from galaxy
- Need image evaluation in 10 hours
- Transient pipeline takes 3 hours, because it can



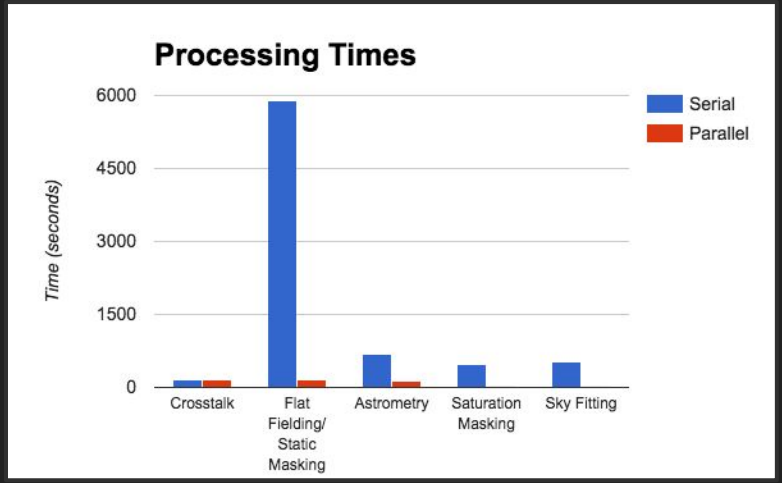
# Transfer/Alert Speedups

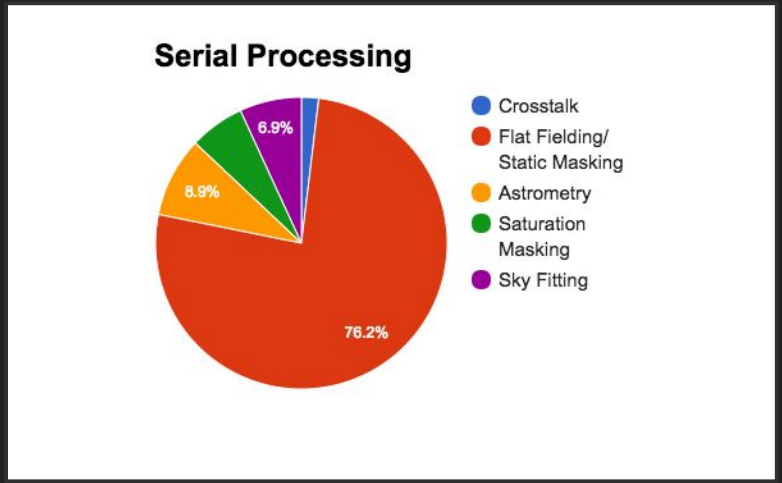
- DES file transfer: 20 minutes
  - Transfer time: slow
  - Data archived before processing
  - Waiting for 10 minutes cron jobs
  - Auxiliary files loaded as needed
- LSST transfer will be seconds
  - Devoted 100 GBPS line
  - Parallel archiving and processing
  - Working “live” no cron
  - Auxiliary files prestaged
- DES alert system: 10 minutes
  - Literally waiting for a cron job
- LSST alerts instantaneous



# Whole Exposure Processing

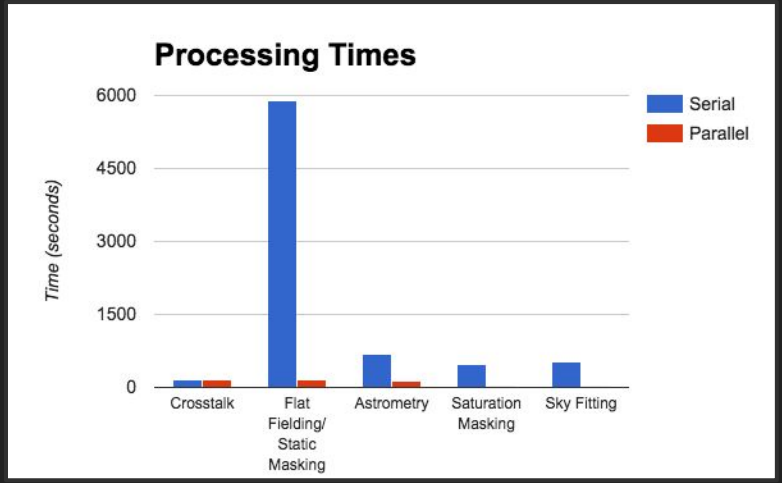
- Crosstalk correction, astrometry, sky fitting best done on whole exposure
- Some tasks run 60x serially
  - Once per CCD
- Dominated by flat fielding
- Parallelization possible
  - Serial 130 minutes
  - Parallel 8 minutes
- LSST Crosstalk Correction performed at telescope
  - Without crosstalk, DES is 5.5 minutes

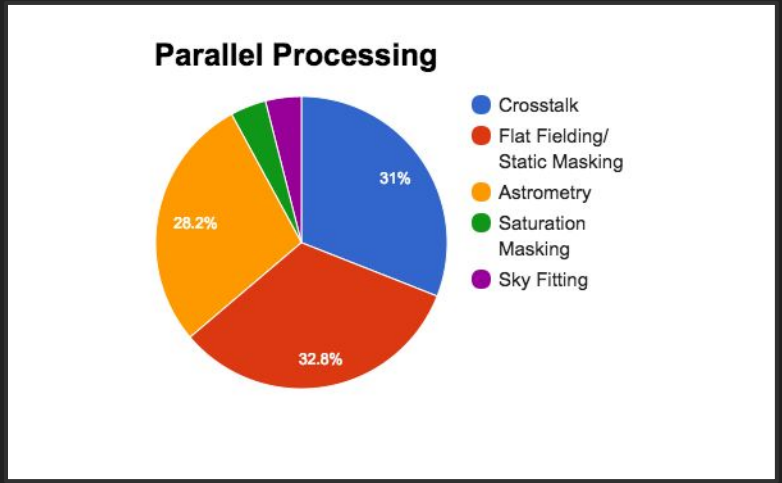


# Whole Exposure Processing

- Some tasks run 60x serially
  - Once per CCD
- Dominated by flat fielding
- Sky fitting/astrometry best done on whole exposure
- Parallelization possible
  - Serial 130 minutes
  - Parallel 8 minutes
- LSST Crosstalk Correction performed at telescope
  - Without crosstalk, DES is 5.5 minutes



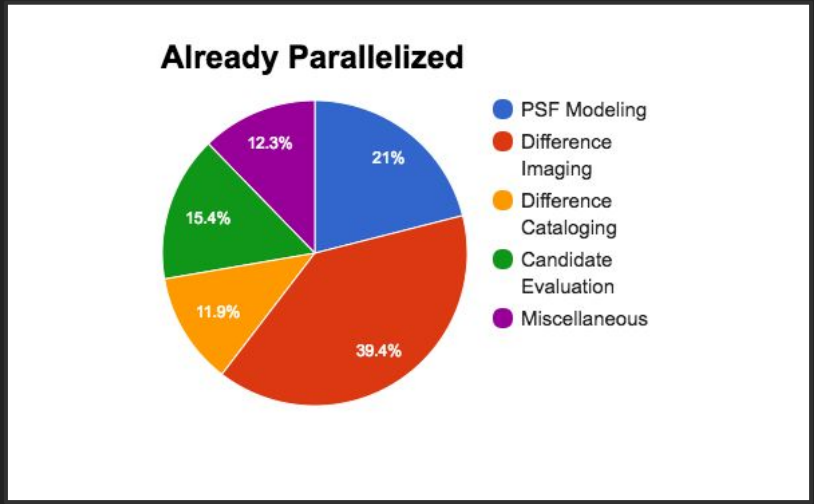



# Real Reasons for Not Parallelizing

- Crosstalk Correction: whole exposure necessary
  - LSST does it on the mountain
  - LSST only corrects neighboring CCDs
  - DES has crosstalk of up to $10^{-5}$ for non-neighboring CCDs
- Astrometry: best done on whole exposure
  - Better constrained with more stars
  - DES difference imaging fails ~15% of the time
    - But SN taking with worst seeing
  - Single CCD astrometry increases failure rate in marginal weather
- Sky (Background) Fitting: best done on whole exposure
  - Limit L1 calibration
- L2 processing will perform full exposure analysis

# DES Transient Pipeline: CCD Parallelized

- Last 8 minutes: parallelized by CCD
- Dominated by difference imaging
  - HOTPANTS
- PSF Modeling next
  - SExtractor, PSFEx
- Fully parallelized, post-crosstalk: 13.5 minutes
- One more way to speed things up...



# Keep data in memory

- DES reads and writes files each of 46 processes
  - Usually a large fits file
  - Most not returned
- Rough estimate: 25% of “processing time” file i/o
  - Keeping all data in memory takes us down to 10 minutes
- Increased operations difficulties
  - Debugging easier with distinct processes and output
  - Even moreso with batch operations
    - Processes killed externally

# From 10 Minutes to 1 Minute: Magic

- Processors still getting faster
- Optimizing code for LSST
- Algorithmic magic
- I am excited to learn about the magic

# Conclusions

- Previous surveys show: operations team essential
- LSST detect transients 200X faster than DES
- 20X from parallelization, algorithms, working in memory
  - Real data sacrifices and operations challenges
- 10X from faster code
  - Interested in understanding how that works