



# Prompt Processing

Kian-Tat Lim



U.S. DEPARTMENT OF  
**ENERGY**

**SLAC**

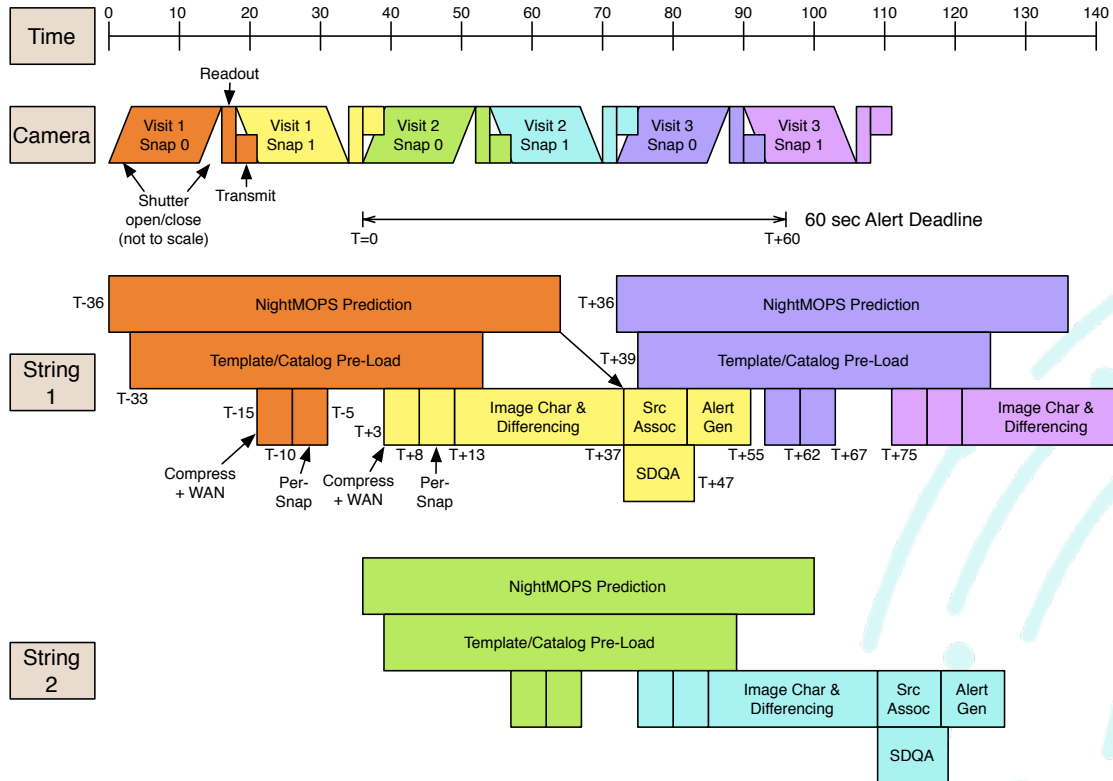


# Original Concept and Key Features

---

- Executes Alert Production in parallel across hundreds of cores
- Can execute calibration reductions
- Data-driven
- Visit-driven (not arbitrary collections of input data)
- Can preload data (or, equivalently, block waiting for input)
- Low overhead to minimize added latency
- Events (or even log messages) with metrics conveyed to Chile via DM-internal messaging; converted to SAL messages there

# Timeline



# Gen3 MW Issues

---

- Group id vs. exposure id
  - Only group known at `next_visit` event time
  - Exposure id could change if previous script is aborted
- In current Gen3 MW, cannot block for data within pipeline
  - All dataids need to be known at QG generation
- Is pre-staging to local (RAM?) disk sufficient?
  - Memory-based pipeline dataset transfer unproven
- High overhead for QG generation
  - Only a few pipelines; precompute?
- Need more than one core per CCD? One or more gather/scatter steps?

# Event-Triggered “OCPS”

---

- OCPS is nominally commanded to execute a Gen3 pipeline
- But it could be triggered by an event as well (e.g. data available or `next_visit`)
- Back-end is a UWS service accessed via REST API; can be anywhere, including USDF
- Can extend to execute a custom Prompt Processing shell script
  - Gather/scatter could be difficult
  
- Aside: in the cloud, might use “serverless computing”: Amazon Lambda, Google Cloud Run

# Event Generation

---

- How do we emit events over LHN?
- ActiveMQ no longer in use; use Kafka? (Irony: Kafka to SAL to Kafka again)
- Datasets, logs, custom code?
- Is `lsst.verify + faro` a suitable framework for metric computation and output?

# Move Prompt Processing Again?

---

- Originally at Base because Summit was too cramped
- Moved to NCSA for ease of deployment, maintenance, and elasticity during daytime or even at night (e.g. for crowded fields)
- Requires reliable, high-bandwidth LHN
- Requires transfer of data products back to Chilean DAC (less critical if delayed 24 hours)
- Could move back to Base or Summit
- Share resources with Commissioning and/or Camera Diagnostic Cluster?