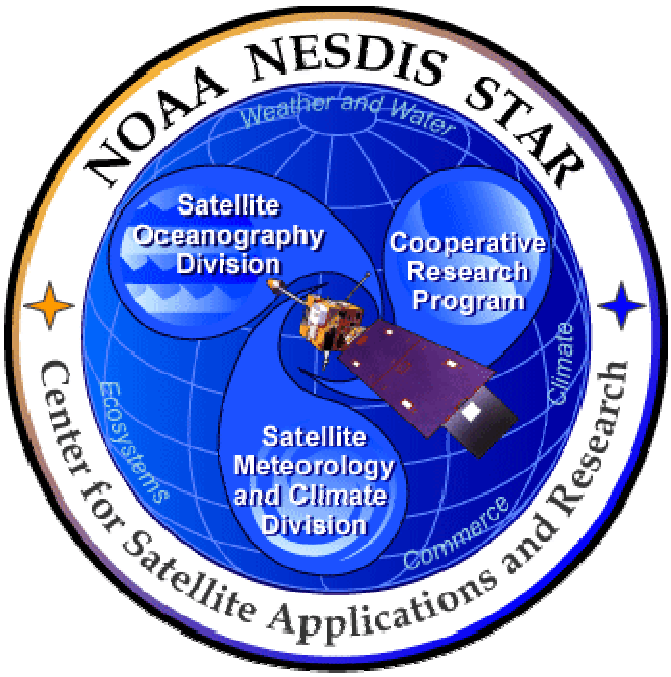

NOAA NESDIS
CENTER for SATELLITE APPLICATIONS and RESEARCH

ALGORITHM THEORETICAL BASIS DOCUMENT

GLM Lightning Cluster-Filter Algorithm

Steve Goodman, NOAA/NESDIS

Douglas Mach, University of Alabama in Huntsville
William Koshak, NASA-MSFC
Richard Blakeslee, NASA-MSFC

Version: **2.0**

September 24, 2010

Table of Contents

LIST OF FIGURES 3
LIST OF TABLES 3
LIST OF ACRONYMS 4
ABSTRACT 5
1. INTRODUCTION 6
1.1 Purpose Of This Document 6
1.2 Who Should Use This Document 6
1.3 Inside Each Section 6
1.4 Related Documents 7
1.5 Revision History 7
2. OBSERVING SYSTEM OVERVIEW 7
2.1 Products Generated 7
2.2 Instrument Characteristics 9
2.2.1 GLM Measurement Approach 10
2.2.2 On-Board Signal Processing 11
3. ALGORITHM DESCRIPTION 12
3.1 Algorithm Overview 12
3.1.1 Background 13
3.1.2 Event 13
3.1.3 Group 13
3.1.4 Flash 14
3.2 Processing Outline 14
3.3 Algorithm Input 17
3.3.1 Primary Sensor Data 17
3.3.2 Ancillary Data 18
3.3.3 Derived Data 18
3.4 Theoretical Description 18
3.4.1 Physics of the Problem 18
3.4.2 Mathematical Description 20
3.4.3 Algorithm Output 26
4. Test Data Sets and Outputs 30
4.1 Simulated Input Data Sets 30
4.1.1 NLDN Data 30
4.1.2 VHF Data 31
4.1.3 Re-Sampled LIS Data 32
4.2 Output From Simulated Input Data Sets 32
4.2.1 Precisions and Accuracy Estimates 33
4.2.2 Error Budget 33
5. Practical Considerations 33
5.1 Numerical Computation Considerations 33
5.2 Programming and Procedural Considerations 34
5.3 Quality Assessment and Diagnostics 34
5.3.1 Code Processing Speed 34
5.3.2 Clustering Results 35
5.3.3 Marking Non-Lightning Events, Groups, and Flashes 35

5.3.4 Diagnostics Imported from L0-L1b Code 35
5.4 Exception Handling 35
5.4.1 Flash QA Bits (13 Bits Total) 35
5.4.2 Group QA Bits (9 Bits Total) 36
5.4.3 Event QA Bits (6 Bits Total) 37
5.5 Algorithm Validation 37
6. ASSUMPTIONS AND LIMITATIONS 38
6.1 Performance 38
6.2 Assumed Sensor Performance 38
6.3 Pre-Planned Product Improvements 38
References 38
Appendix 1: Algorithm Implementation and Test Plan Document GLM Lightning Cluster-Filter Algorithm 40
1. OBJECTIVE 40
1.1 Purpose of this Document 40
1.2 Specific Objectives 40
1.3 Scope of Document 40
1.4 Related Documents 40
2. BRIEF DESCRIPTION OF ALGORITHM 41
2.1 Overview 41
2.2 Product Summary 41
3. DATASETS 41
3.1 National Lightning Detection Network Data 42
3.2 VHF Lightning Mapping Data 42
3.3 Lightning Imaging Sensor Data 42
4. METHODOLOGY 42
4.1 Algorithm Definition Phase 43
4.1.1 Speed Tests 43
4.1.2 Accuracy Tests 44
4.1.3 Resiliency Tests. 45
4.1.4 Specific List of Tests 45
4.2 Algorithm Refinement Phase 47
5. RESULTS 48
5.1 CPU Load Test (S) 48
5.2 CPU Load Test II (S) 49
5.3 CPU Load Test III (S) 51
5.4 Longitude Test I (International Dateline) (R) 54
5.5 Longitude Test II (Prime Meridian) (R) 56
5.6 Simulated Regrouping Test (GLM_test_MergingFlash.txt) (A) 57
Appendix 2: A Physical Understanding of the Event, Group, and Flash “Radiance” Data Products 59
Appendix 3: The Flash Radiance Data Product 63
A3.1 Preliminary Definitions and Derived Quantities 63
A3.2 An Expression for the Data Product 65
Appendix 4: Additional Considerations on Centroiding 67

LIST OF FIGURES

Figure 1. Optical groups attempt to track bright transient emissions from lightning. Inter-stroke processes also produce optical groups. 9
Figure 2. GLM field-of-view coverage (thin line is for GOES-West configuration, and thick line is for GOES-East configuration). 10
Figure 3. High level flow chart of the GLM Lightning Cluster Filter Algorithm (LCFA). 16
Figure 4. Details of the GLM Lightning Cluster Filter Algorithm LCFA). 17
Figure 5. The first group (in red). 20
Figure 6. The next group (in green). 21
Figure 7. The next two groups (in yellow and blue). 21
Figure 8. Two additional single event groups (in cyan and magenta). 22
Figure 9. Two additional single event groups (in grey and black). 22
Figure 10. The 21 July 2003 NLDN proxy dataset. 30
Figure 11. Plot comparing LIS (squares), LMA (dots), and NLDN (Xs) datasets. 31
Figure 12. The fractional coincidence of LIS and LMA flashes as a function of range. . 32
Figure 13. Overall validation strategy involving the LCFA. This figure extends beyond the definition/performance testing of the LCFA (see the Validation Plan for details). 43
Figure 14. Event Rates 49
Figure 15. Processing speed of the GLM code 49
Figure 16. Event Rates. 51
Figure 17. Processing speed of the GLM code 51
Figure 18. Data rate as a function of data time. 54
Figure 19. Processing speed of the GLM code. 54
Figure 20. Location of all groups in the flash (centered around the International Dateline). 56
Figure 21. Location of all groups in flash (centered around Prime Meridian). 57
Figure 22. Location of all the groups in the one flash. 58
Figure 23. Geometry for energy density calculations. 59

LIST OF TABLES

Table 1. Revision History 7
Table 2. Lightning Product Requirements 8
Table 3. Resultant Flash Data 23
Table 4. Resultant Group Data 23
Table 5. Event Data 24
Table 6. Group Data 24
Table 7. Flash Data 24
Table 8. Output Flash/Group/Event Order from the Example Processing 25
Table 9. Alternate Output Flash/Group/Event Order from the Example Processing 25
Table 10. QA Bits for GLM LCFA 26
Table 11. Estimated Bandwidth Needed to Transmit the GLM LCFA Data 27
Table 12. Metadata for Lightning Product 29
Table 13. Error Budget Summary 33

LIST OF ACRONYMS

AIT – Algorithm Integration Team
AITP – Algorithm Implementation and Test Plan
ATBD – Algorithm Theoretical Basis Document
AWG – Algorithm Working Group
CCD – Charge Coupled Device
FAR – False Alarm Rate
F&PS – Functional & Performance Specification
GLM – Geostationary Lightning Mapper
GOES – Geostationary Operational Environmental Satellite
GRB – GOES Rebroadcast
GSP – Ground Segment Project
IFOV – Instantaneous Field Of View
LAN – Local Area Network
LCFA – Lightning Cluster Filter Algorithm
LEO – Low Earth Orbiting
LIS – Lightning Imaging Sensor
LMA – Lightning Mapping Array
MRD – Mission Requirements Document
NASA – National Aeronautics and Space Administration
NESDIS - National Environmental Satellite, Data and Information Service
NLDN – National Lightning Detection Network
NOAA – National Oceanic and Atmospheric Administration
OTD – Optical Transient Detector
PG – Product Generation
RTEP – Real-Time Event Processor
STAR – (Center for) Satellite Applications and Research
VAGL – Vendor Allocated Ground Latency
VHF – Very High Frequency

ABSTRACT

This document provides a detailed description of the Lightning Cluster Filter Algorithm (LCFA) that generates Level 2 lightning products (flashes, groups, events) from Level 1b Geostationary Lightning Mapper (GLM) geo-located, time-tagged lightning event data. The LCFA builds a parent-child tree-structure that identifies the clustering of optical events into groups, and groups into flashes. An overview of the products generated is provided with some basic background on the GLM instrument characteristics. A description of simulated (proxy GLM) event data is also provided. The GLM proxy event data is used as input to the LCFA, and the LCFA output is evaluated. Practical considerations (e.g., LCFA exception handling) are examined, and all assumptions and limitations are provided along with plans to mitigate potential issues.

1. INTRODUCTION

This section summarizes the overall purpose and intended users of this document, and provides an overview of each main document section. It also provides a concise description of related documents and the revision history of this document.

1.1 Purpose Of This Document

The Algorithm Theoretical Basis Document (ATBD) for the Geostationary Lightning Mapper (GLM) Lightning Cluster-Filter Algorithm (LCFA) provides a high level description of, and a physical basis for, the creation of clustered lightning "events", "groups" and "flashes" from geolocated Level 1b lightning "events". The central purpose of this ATBD is to facilitate development of operational Product Generation (PG) software; i.e. the ATBD is not intended to be an information-only document.

1.2 Who Should Use This Document

The intended users of this document are those interested in understanding the theoretical and physical basis of the LCFA. The primary users of this document are those who want to develop operational lightning clustering/filtering code for the GLM. This document also provides information useful to anyone maintaining or modifying the original algorithm, and those who wish to optimally and appropriately apply the LCFA output.

1.3 Inside Each Section

This document is broken down into the following main sections.

- **Introduction**: Basic background and context (purpose, use, overview, related documents, revision history).
- **Observing System Overview**: Provides a brief description of the products generated by the algorithm, and some pertinent GLM instrument characteristics.
- **Algorithm Description**: Provides a detailed description of the algorithm including its physical/mathematical basis, its input, and its output.
- **Test Data Sets and Outputs**: Describes the test data sets used to characterize the performance of the algorithm and quality of the data products, including the breadth of the domain (typical versus stressing states) used in the analysis and assessment. This section also describes results (precision, accuracy, error budgets) from algorithm processing on simulated input data.
- **Practical Considerations**: Describes issues involving numerical computation, programming and procedures, quality assessment and diagnostics, and exception handling. It also references the algorithm validation document.
- **Assumptions and Limitations**: Provides an overview of the current assumptions and limitations of the approach and gives the plan for overcoming these limitations with further algorithm development.
- **References**: Provides pertinent references, primarily from the scientific literature on lightning.

1.4 Related Documents

This ATBD document draws heavily upon the heritage *Algorithm Theoretical Basis Document (ATBD) for the Lightning Imaging Sensor (LIS)*. Other related documents include the specifications in the *Mission Requirements Document (MRD)*, the references provided in section 5, the *Algorithm Implementation and Test Plan (AITP)* document (Appendix 1) for the LCFA, and the *Product Validation Plan* for validating GLM Level 2 lightning products (flashes, groups, events).

1.5 Revision History

Table 1. Revision History

Revision	Date	History
Draft	5-30-08	Submitted to the Algorithm Integration Team (AIT).
Revised Draft	9-25-08	Revised draft submitted to the AIT.
Another Revised Draft	1-6-09	A revised format was requested on January 6, 2009 by the AIT, so the ATBD was upgraded and re-submitted to the AIT shortly thereafter.
Pre 80% Version	5-27-09	To meet format changes made by NOAA on May 21 & 27, 2009
80% Version	5-29-09	This version is the official 80% level ATBD.
Post 80% Version	9-30-09	Various format changes, inputs to subsections, and corrections were made, and submitted to AIT upon AIT's request.
100% Version	6-22-10	This version is the first 100% level ATBD.
100% Version, Revised	9-24-10	Subsequent upgrades in response to Suscynsky and AIT review comments produced this revised 100% version .

2. OBSERVING SYSTEM OVERVIEW

This section provides an overview of the products generated and instrument characteristics (including instrument mission objectives).

2.1 Products Generated

The product requirements are shown in Table 2. At the time of this writing, some of the entries are planned and possibly have not been formally implemented within the similar requirements tables of the F&PS. For data rates < 20,000 events s^{-1}, the LCFA uses 1 s for input buffering, less than 1 s for processing, and 1 s for output buffering. The LCFA has been allocated 4 s of the 20 s VAGL shown in Table 2.

Table 2. Lightning Product Requirements

Name	User & Priority	Product Geographic Coverage	Vertical Resolution	Horizontal Resolution	Mapping Accuracy	Measurement Range	Measurement Accuracy	Product Refresh Rate /Coverage Time	VAGL	Measurement Precision	Temporal Coverage Qualifiers	Product Extent Qualifier	Cloud Cover Conditions Qualifier	Product Statistics Qualifier
Lightning Detection -Events -Groups -Flashes	GOES-R	Full Disk	Sfc to Cloud Top.	10 km	5 km	Real Time	70% minimum Flash Detection Efficiency (FDE)	Continuous	20 sec	5% (Std. Dev. of FDE)	Day and night	Quantitative out to at least 65 degrees LZA and qualitative beyond	Cloud cover conditions permitting obs. of lightning associated with threshold accuracy	Over lightning cases and mesoscale-sized surrounding regions

There are 2 format types (netCDF4, McIDAS), with 3 base products (events, groups, flashes) for a total of 6 products. The GLM products are expected to meet the target F&PS requirements; however, whether or not the "5 km (or less) Mapping Accuracy" requirement is actually met depends on how well parallax error will be mitigated. GOES Rebroadcast (GRB) will distribute all base products. The base products will be geo-located, time-stamped, and sized (i.e., characterized by a footprint). Geo-location of the groups and flashes are based on computing optical amplitude-weighted centroids (the mathematical/pseudocode is provided in Section 3.4.3). The base products will be provided as close to real time as the software can produce. These base products are linked by a tree structure that shows what events construct what groups, and what groups construct what flashes. Note that the definition of a group is intended as a proxy to individual return strokes in a ground flash (as depicted in Figure 1) or to individual high current discharges (K-changes) in a cloud flash. The correspondence is not normally one-to-one however; there tends to be more groups per flash than there are strokes per flash since inter-stroke processes can also produce groups. In addition, note that the GOES-R Ground Segment Project (GSP) considers the Instrument Vendor (Lockheed Martin) data as Level 1b data. The events, groups, and flashes calculated with the LCFA are considered as Level 2+ end products. More information on events, groups, and flashes is provided in the next section. Finally, statistics shall be calculated to provide an indication of how likely a flash is actually lightning. The statistics will be based on the estimated noise event rate provided by the L0-L1b algorithm output. The statistics will be the likelihood that the flash could have been created from random noise alone, based on the input estimate of the noise rate (and Monte Carlo simulations of the GLM FOV).

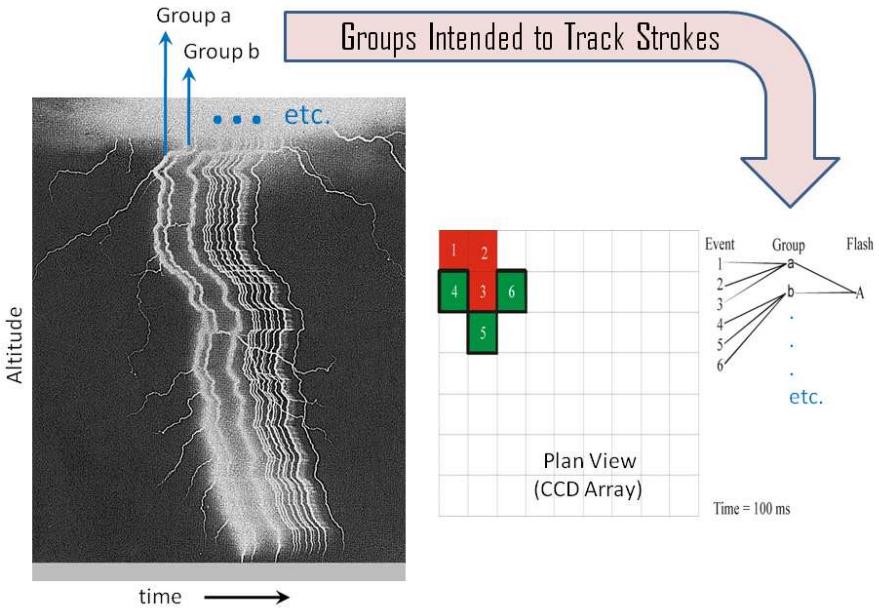

Figure 1. Optical groups attempt to track bright transient emissions from lightning. Inter-stroke processes also produce optical groups.

2.2 Instrument Characteristics

The mission objectives of the GLM are to: (1) provide continuous, full disk lightning measurements for storm warning and nowcasting, (2) provide an early warning of tornadic activity, and (3) accumulate a long-term database to track decadal changes of lightning. The GLM will be deployed as an instrument on the GOES-R spacecraft (36,000 km altitude, 0^o inclination). The GLM shall have continuous monitoring capability across a near hemispheric coverage (18,000 km diagonal field-of-view) with a near-uniform round sample 8-14 km pixel footprint resolution. Figure 2 shows the field-of-view coverage that depends on the satellite positioning (thin outline is for a GOES-W configuration at -137^o longitude, and the thick outline is for a GOES-East configuration at -75^o longitude); the annualized lightning flash density (in units of flashes/km^2/yr) is also provided from the combined Optical Transient Detector (OTD) and LIS climatology. The GLM will map lightning location to within 5 km and will have a minimum lightning flash detection efficiency of 70%.

Sections 2.2.1 and 2.2.2 describe the processing done on-board the GLM instrument or on-ground by the L0-L1b algorithms. They provide a background for the processing of the LCFA which is described in Section 3.

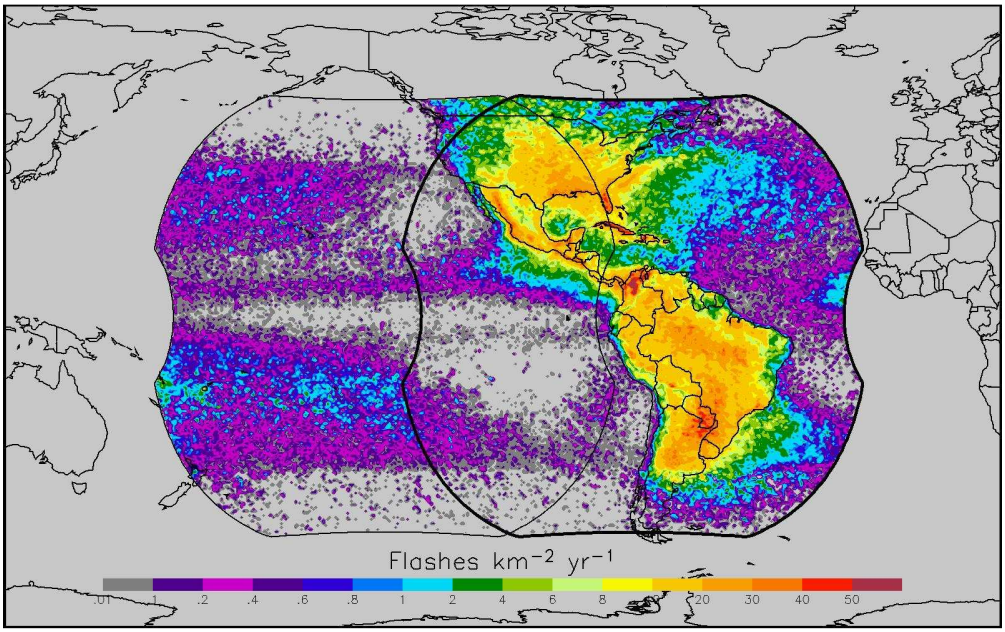

Figure 2. GLM field-of-view coverage (thin line is for GOES-West configuration, and thick line is for GOES-East configuration).

2.2.1 GLM Measurement Approach

The GLM is a conceptually simple device. It images the scene much like a television camera; however, because of the transient nature of lightning, its spectral characteristics, and the difficulty of daytime detection of lightning against the brightly lit cloud background, actual data handling and processing is much different from that required by a simple imager. In order to achieve the performance goals required to meet the scientific objectives, the GLM combines components in a unique configuration. Conceptually, a wide field-of-view lens combined with a narrow-band interference filter is focused on a high speed 1372 x 1300 Charge Coupled Device (CCD) focal plane. Signals are read out in parallel from the focal plane into real-time event processors for event detection and data compression. The resulting event detections are formatted, queued, and sent to the satellite Local Area Network (LAN).

The particular characteristics of the GLM sensor design, well proven by the heritage OTD and LIS instruments, result from the requirement to detect weak lightning signals during the day. During the day, the background illumination, produced by sunlight reflecting from the tops of clouds, can be much brighter than the illumination produced by lightning. Consequently, the daytime lightning signals tend to be buried in the background noise, and the only way to detect lightning during daytime is to implement filtering techniques that increase or maximize the lightning signal relative to this bright background. These filtering techniques take advantage of the significant differences in the spatial, spectral, and temporal characteristics between the lightning signal and the background noise. A combination of four filtering methods is employed by the GLM for this purpose.

A form of spatial filtering is used first by matching the instantaneous field-of-view (IFOV) of each detector element in the GLM focal plane array to the typical cloud-top

area illuminated by a lightning stroke (i.e., 5-10 km on a side). The matching of IFOV with the typical cloud size results in an optimal sampling of the lightning scene relative to the background illumination. For example, an IFOV having dimension greater than 10 km would typically be under-filled with lightning signal relative to the background because the diffuse cloud-top lightning emission associated with an embedded vertical lightning source decays radially at cloud-top whereas the sunlit cloud maintains cloud-top brightness in the radial direction. If the IFOV has dimension smaller than 10 km, both the lightning signal and the background signal decrease, but electronic noise does not, so again the signal to (total) noise ratio drops.

Second, spectral filtering is realized by using a narrow-band interference filter centered on a strong optical emission multiplet (e.g., the first ionized Oxygen OI(1) line at 777.4 nm) in the lightning spectrum. The narrow-band filtering is needed to further enhance the lightning signal relative to the reflected daylight background, which otherwise would overwhelm the much smaller lightning signal.

Third, through time integration of the signal on the focal plane, the GLM achieves temporal filtering that takes advantage of the difference in lightning pulse duration of the order of 400 μs versus the background illumination which tends to be constant on the time scale of the order of seconds. In an integrating sensor, such as the GLM, the integration time specifies how long a particular pixel accumulates charge between readouts. The lightning signal-to-noise ratio improves as the integration period approaches the pulse duration. If, however, the integration period becomes too short, the lightning signal tends to be split between successive frames; this decreases the signal-to-noise ratio. A 2 ms integration time has been adopted in the GLM instrument design to minimize pulse splitting and maximize lightning detectability.

Even with the three "filtering" approaches discussed above, the ratio of the background illumination to the lightning signal may still exceed 150 to 1 at the focal plane. Therefore, a fourth technique – a modified frame-to-frame background subtraction – is implemented to remove the slowly varying background signal from the raw data coming off the GLM focal plane. A more detailed discussion on the measurement approach adopted for the GLM is given in Christian et al. (1989).

The real time event processor generates an estimate of the background scene imaged at each pixel of the focal plane array. This background scene is updated during each frame readout sequence and, at the same time, the background signal is compared with the off-the-focal-plane signal on a pixel-by-pixel basis. When the difference between these signals exceeds a selected threshold, the signal is identified as a potential lightning event and an event processing sequence is enabled. The implementation of this real-time data processor results in a 10^5 reduction in data rate requirements while maintaining high detection efficiency for lightning events.

2.2.2 On-Board Signal Processing

The GLM is a staring imager optimized to detect and locate lightning. An imaging system, a focal plane assembly, a real-time signal processor and background remover, an event processor and formatter, power supply, and interface electronics are the six major subsystems of the sensor.

The GLM Real-Time Event Processors (RTEPs) can extract weak lightning flashes from an intense but slowly evolving background. The daytime background varies primarily with sun angle, cloud albedo, and ground albedo. When a cloud-top lightning optical emission is detected within a single frame, the lightning signal is superimposed on top of the essentially constant background signal. Each RTEP continuously averages the output from the focal plane over several frames on a pixel-by-pixel basis in order to generate a background estimate. It then subtracts the average background estimate from the current signal to obtain a difference signal.

The difference signal essentially consists of shot noise fluctuating about a zero with occasional peaks due to lightning events. When a peak exceeds the level of the variable threshold, it is considered to be a lightning event and is processed by the rest of the circuit. Clearly, the threshold must be higher during daytime when shot noise is dominated by the solar background. Since the original signal contains background plus lightning, or just background, the difference signal will be either a lightning signal, near zero, or a false alarm.

Once the event is identified it is time-tagged. The event amplitude, (x,y) pixel address, and time tag are passed to platform telemetry via the interface electronics to the satellite LAN. Event geolocation, and determination if the event is a noise source (via a coherency filter analysis) are performed on the ground by the Ground Segment.

3. ALGORITHM DESCRIPTION

This section provides a complete description of the algorithm at the current level of maturity.

3.1 Algorithm Overview

The concept of the LCFA is closely based on the (only predecessor) heritage OTD/LIS data processing algorithm. Consequently, there are no competing pre-existing algorithms to compare or contrast with, or to detail in a long overview. However, this ATBD does invoke modifications to the heritage algorithm that are needed when one moves from the Low Earth Orbiting (LEO) OTD/LIS application to the geostationary application associated with GLM.

There is one major product produced by the GLM software: a lightning dataset. To obtain this dataset, the satellite data stream needs to be decoded, filtered, clustered, and output to the appropriate file. The LCFA only generates the lightning dataset. Specifically, the LCFA receives as input the Level 1b pixel-level optical “event” data and processes this data into more convenient lightning data products that are easily utilized by the scientific research and broader operational user communities. Therefore, the LCFA must take the event data and assemble the higher level clustered lightning data products (events, groups and flashes), and in so doing, it will generate derived lightning characteristics associated with these higher level products. It will also interrogate individual flashes, groups, and events on a statistical basis to see if they are associated with lightning or noise [i.e., the Lightning AWG Team does not assume that the noise filtering performed by the Instrument Vendor is perfect]. Definitions of the basic data storage classes (events, groups, flashes) that drive the LCFA are provided below.

3.1.1 Background

Note that the lightning AWG group is not responsible for the background image and will not use the background information in the LCFA, but it has been included here for perspective. A background image is a "snap shot" of the background estimate made possible by the GLM RTEPs; because of the large field-of-view, the GLM instrument employs several RTEPs. The background image is transmitted in the data stream along with event data. When the transmission of one background is begun, the next background image is captured. New images are sent to the ground as frequently as the transmission rate allows. Though the background image is not used by the LCFA, it has scientific uses; e.g., it provides the geographical distribution of clouds in the near infrared over which the lightning occurs.

3.1.2 Event

An event is defined as the occurrence of a single pixel exceeding the background threshold during a single frame. In other words, each pixel output from the RTEP produces a separate event. The Level 1b GLM instrument data consists of time, (x, y) pixel address, latitude and longitude locations, and calibrated amplitude of the event. An event is the basic unit of data from the GLM.

Although an event can be thought of as a single optical pulse due to lightning, it is possible that multiple pulses occurring within the 2 ms integration window may contribute to an event. Therefore, we purposely did not use ‘pulse’ or ‘stroke’ (or other similar name) to describe the basic unit of data from the GLM. Note that an event may sometimes not be due to lightning at all. It may be produced by noise in the analog data stream exceeding the background threshold. In that case, the event is a false alarm. When the LCFA determines that an event has a non-zero probability of being from a non-lightning source, it will be marked as such in the data, but it will still be clustered along with the lightning data.

3.1.3 Group

A lightning discharge will often illuminate more than one pixel during a single integration time. The result is two or more adjacent events within the same time frame. When these multiple events are adjacent to each other (a side or corner of the events touching), they will be placed in a single group. The formal definition of a group is one or more simultaneous events (i.e., events that occur in the same time integration frame) that register in adjacent (neighboring or diagonal) pixels in the focal plane array. A group may consist of only one event or include many events. The location data for a group will be calculated in earth-based (latitude/longitude) coordinates. This is done to provide consistent representation in the group/flash processing and because the ultimate goal of the analysis is to locate lightning with respect to the earth’s surface.

Although a group may often correspond to a single lightning optical pulse, it is also possible that multiple lightning pulses occurring within the 2 ms integration window may contribute to a group. A false event due to noise at a pixel exceeding the background threshold can also contribute to a group (although noise groups often contain only one event). Note that if an event can be assigned to more than one group, all of the groups it can be assigned to will be combined into one group (and then the event added to it). If a

group is determined to be from a non-lightning source, it will be marked as such in the QA bits (see Section 5.4) but it will still be clustered by the LCFA.

3.1.4 Flash

A lightning flash consists of one to multiple optical pulses within a specified time and distance. For the GLM algorithm, we define a flash as a set of groups sequentially separated in time by no more than 330 ms and in space by no more than 16.5 km in a weighted Euclidean distance format (details in Section 5). Note that for two (or more) groups to be considered part of the same flash, any two events in the two groups can meet the 330 ms and 16.5 km spacing. In other words, for the GLM algorithm, we do not use the group centroids to determine if two (or more) groups are part of the same flash. The temporal and spatial rules can be easily adjusted in the GLM algorithm processing software. We will continue to examine the rules closely during the analysis of OTD, LIS, and GLM data to "fine tune" the rules defining a flash. A flash may include as few as one group with a single event or it may consist of many groups, each containing many events. Spatial characteristics for a flash (and all higher level parameters) are calculated in ground coordinates (i.e., latitude and longitude).

We have used the above definition of a "flash" because we believe it will usually produce results that correspond to the customary definition of a conventional lightning flash. Note that with GLM data alone, we cannot determine if a flash is a ground or cloud flash. It is possible that future versions of the GLM algorithm may incorporate data from ground flash locating systems along with GLM optical characteristics to help interpret the GLM flashes as either cloud-to-ground or intra-cloud. We do acknowledge that, on occasion, distinct conventional lightning flashes may result in a single flash being produced by the GLM algorithm (e.g., possibly in high flashing rate mesoscale convection systems). Other mismatches between algorithm flashes and actual conventional flashes will undoubtedly also occur. Note that there is no absolute time limit to a flash. That is, as long as subsequent groups are produced in an area within the 330 ms time windows, all groups will be assigned to a single flash. However, practical considerations do limit the total size and time span of a flash (see Section 5). Also note that if a group can be assigned to more than one flash, all flashes it can be assigned to will be combined into one flash (and then have the group added to it).

3.2 Processing Outline

The processing outline for the LCFA is provided in Figures 3 and 4. Many of the subroutines are candidates for parallel processing, so this allows for simple parallelization of the GLM code. Such parallelization would take advantage of modern computer architecture (e.g., multi-core processors and/or multi-threaded operating systems). The tasks do not *require* parallelization, but are designed to accommodate parallelization. If the code runs faster in parallel mode, then the GLM code can take advantage of this strategy. We will weigh the advantages of parallel processing against the extra code/memory overhead needed by tuning and optimizing the algorithm using proxy datasets. Details of the clustering algorithm itself are shown in Figure 4.

From our past experience with on-orbit filtering of data (OTD and LIS) we assume that some noise events will make it past the on-board/ground Level 0 to Level 1b GLM filters.

Because of this, we have placed "hooks" in our LCFA for filtering of events/groups/flashes based on the clustering results (the steps labeled “Mark Non-Lightning Events” and “Mark Non-Lightning Groups & Flashes” in Figure 3). However, from our past experience, we also know we can’t predict what these noise events that get past the L0-L1b filters will be like. If we did, they would already be in the L0 to L1b filtering set. We will not know how much filtering will be needed until the actual GLM instrument is being tested. Therefore, at this point, the event/group/flash filtering by the LCFA is TBD. The method will be to mark the events/groups/flashes by setting the appropriate QA bits (described later). Removal of events/groups/flashes is not the designated task of the LCFA, so the “noise” events/groups/flashes will (at least initially) be clustered.

A unique aspect of the GLM processing (in contrast with instruments like the Advanced Baseline Imager, ABI) is that the input data when on orbit will be continuous. Although in testing, the program will work in a batch mode (processing the proxy files), in real-time on-orbit processing, the L1b data will not be staged in chunks larger than about a second or the latency requirements will be violated. The flowcharts presented here represent the program operation for testing within the AIT Framework and not real-time. In real-time, it should not be computationally efficient to start/stop the code (with all of the buffer and memory preparations) after each unit of L1b data. The code will operate in a more continuous mode.

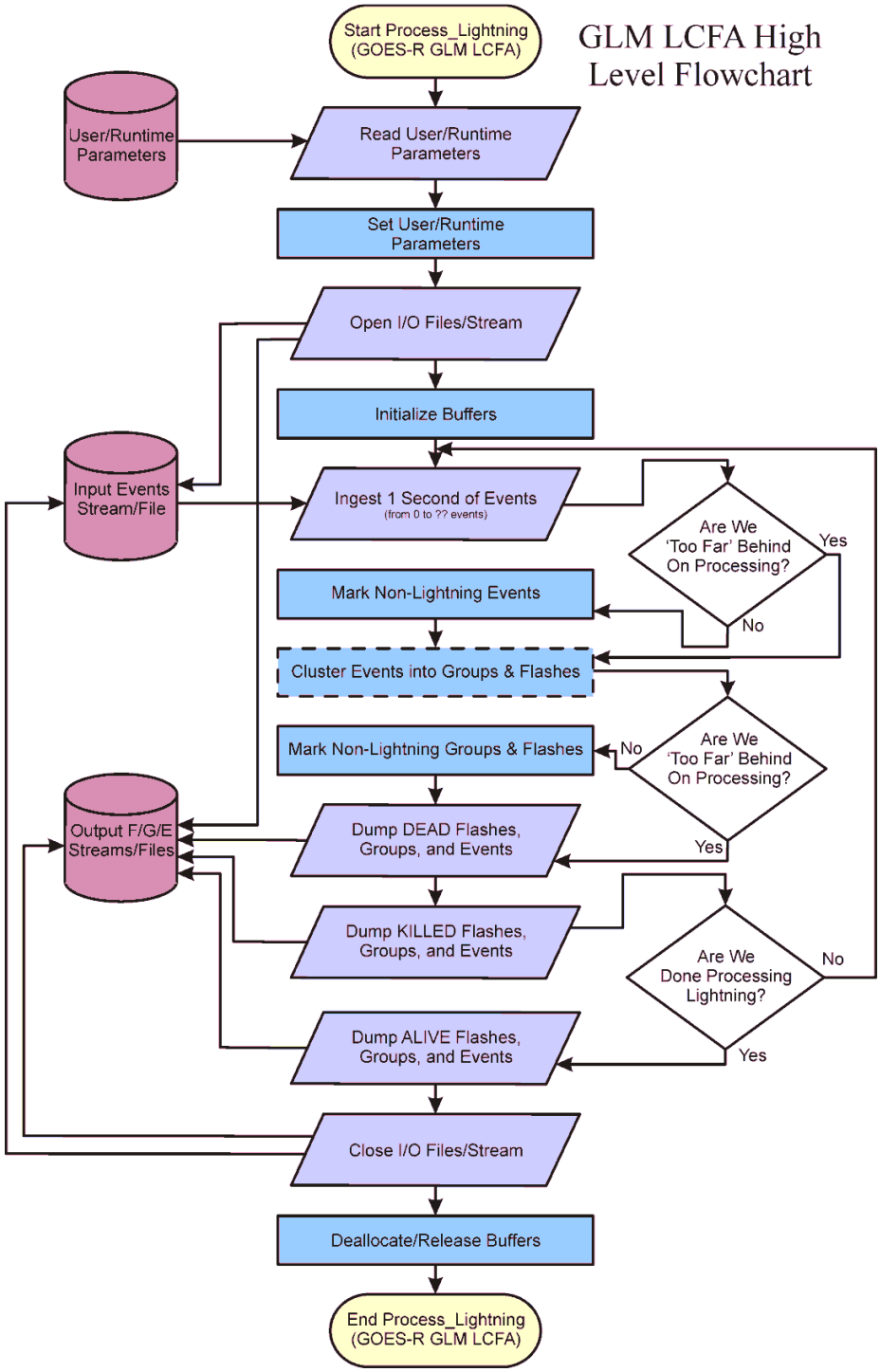

Figure 3. High level flow chart of the GLM Lightning Cluster Filter Algorithm (LCFA).

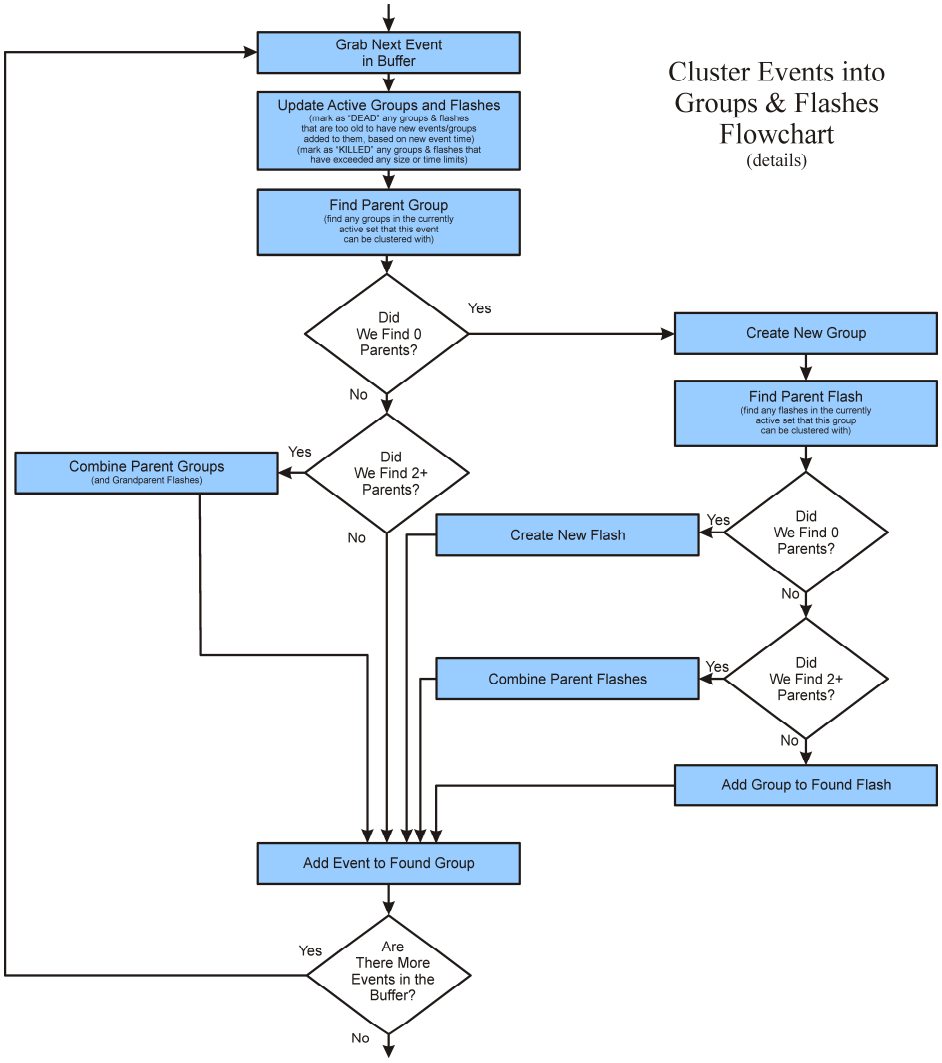

Figure 4. Details of the GLM Lightning Cluster Filter Algorithm LCFA).

3.3 Algorithm Input

This section describes the input needed by the LCFA.

3.3.1 Primary Sensor Data

The LCFA only requires the Level 1b pixel-level event data as input. This includes the event pixel time-stamp, the (x,y) pixel address within the focal plane array, the associated geolocation of the center of the event in latitude/longitude coordinates, the raw event amplitude in counts, and the calibrated event optical energy density amplitude (in microJoules per steradian per square meter per micrometer).

The LCFA assumes the data will be in time order. If the input stream is not time-ordered, the LCFA will still be able to process the L1b data into events, groups, and flashes, but the accuracy will be suspect. The main issue is that some of the groups and flashes may be split due to non-sequential input data. Widely spaced data (far enough apart that they

would not be put into the same flash) that is slightly out of time order (on the order of propagation time differences) will not cause flashes and groups to be artificially split.

3.3.2 Ancillary Data

No ancillary data is required for input into the LCFA.

3.3.3 Derived Data

There are no derived data (such as from radiative transfer models) that is used as input to the LCFA. The time-stamp, pixel address, latitude and longitude, and calibrated amplitude are all that is needed to process the data with the LCFA.

3.4 Theoretical Description

As discussed above, lightning detection is the process of discriminating transient optical emissions against either an un-lit (nighttime) or solar-illuminated (daytime) cloud. Optical transients due to noise must be considered under both day and night conditions and must be clearly distinguished from lightning optical transients. Even though noise will be removed from the Level 1b event data by the Instrument Vendor, this removal process is not assumed to be perfect. So the overarching theoretical problem for the LCFA is to assemble *legitimate* (noise-free) lightning groups and flashes from lightning event data that may or may not contain noise. This means that the LCFA, as the acronym implies, will both cluster and filter the event data. The clustering and filtering "feed" off of one another in the sense that accurate filtering produces accurate clustering and accurate clustering allows one to better filter the data. Indeed, because of the clustering, we anticipate being able to identify and filter noise that might not be identified and filtered by the Instrument Vendor. The following subsection details some important lightning physics that, together with the GLM instrument characteristics discussed above, are fundamental to the development of the LCFA.

3.4.1 Physics of the Problem

The occurrence of lightning is accompanied by the sudden release of electrical energy which is converted into rapid heating in the vicinity of the lightning channel, the generation of a shock wave (which rapidly decays into an acoustic wave, i.e., thunder), and electromagnetic radiation ranging from extremely low frequency (ELF) radio waves to X-rays. One of the strongest radiation regions is in the optical wavelengths with peak power typically between 100 to 1000 MW. These optical emissions result from the dissociation, excitation, and subsequent recombination of atmospheric constituents as they respond to the sudden heating in the lightning channel. The heating is so intense (electron temperatures > 20,000º K) that the optical emissions occur primarily at discrete atomic lines with some continuum at shorter wavelengths. Measurements from a NASA U-2 airplane have shown that the strongest emission features in the cloud top optical spectra are produced by the neutral oxygen and neutral nitrogen lines in the near infrared (e.g., the OI(1) line at 777.4 nm and the NI(1) multiplet at 868.3 nm are consistently strong features).

Temporally, the optical lightning signal is comprised of a series of fast rise time, short-duration pulses associated with the energetic discharge processes occurring within the

cloud. The individual pulses of cloud-to-ground lightning are generally associated with return strokes and K-changes (high current in-cloud discharges). The optical pulse widths and rise times are highly variable and are similar for intracloud and cloud-to-ground lightning discharges.

The thundercloud is an optically thick medium and therefore strongly affects the temporal and spatial characteristics of the optical signals produced by lightning which would be observed by a satellite sensor. Although the thundercloud is optically thick, there is very little absorption at optical wavelengths. Hence, the major effect of the cloud on the optical signal is to blur the source geometry and to delay and time-broaden the pulses due to multiple scattering. Extensive measurements from an instrumented NASA U-2 aircraft have shown that the rise time of an optical pulse is typically lengthened 215 μs and the pulses' widths tend to be 210 μs wider as a result of this multiple scattering. The median pulse rise time and the full-width-at-half-maximum obtained from the analysis of nearly 1300 pulses produced by 79 lightning flashes are 240 μs and 370 μs, respectively (Christian and Goodman, 1987).

It is important to stress that, while the cloud significantly alters the temporal characteristics of the cloud top optical signals, the cloud does not block these emissions. When viewed from above, the optical lightning signals appear as a diffuse light source radiating from the cloud top. Measurements of the total optical energy radiated from the cloud top are in good agreement with ground-based measurements of cloud-to-ground flashes and support the theory that the cloud acts like a conservative scatterer, i.e., that most of the optical energy escapes the cloud (Christian and Goodman, 1987). Of the 79 discharges referred to above, 90% produced peak radiant energy densities of 4.7 $\mu J\ m^{-2}\ sr^{-1}$ or greater. The region of cloud top that is illuminated by a lightning flash depends on where the flash occurred within the cloud, the geometry and physical extent of the flash, and the characteristics of the cloud through which the lightning channel propagated and the radiation scattered. Monte Carlo simulations of the radiation transfer of the optical lightning signals, and the NASA U-2 aircraft studies indicate that the diameter of the cloud top illumination associated with a single storm cell will typically be on the order of 10 km. Observations of large storm systems from the Space Shuttle have shown that illuminated regions can exceed 60 km in length (Goodman and Christian, 1993).

Finally, it should be noted that both intracloud and cloud-to-ground lightning flashes are readily observed from above. Extensive observations with the NASA U-2 aircraft flying over the tops of thunderstorms in coordination with ground-based measurements made under the same storms (Goodman et al., 1988) have clearly established the viability of optical detection of all major types of lightning. Because the majority of the channel of a cloud-to-ground flash occurs within the cloud, the light emerging from the top of the cloud has undergone a similar scattering process as an intracloud flash (the portion of the channel below cloud base is essentially undetectable from above as most light produced outside of the cloud is reflected away from the cloud interior). Further, since the scattering process dominates the characteristics of the optical signature, the optical pulses from both intracloud and cloud-to-ground flashes are very similar (Guo and Krider, 1982; Thomason and Krider, 1982; Goodman et al., 1988). We are unable at this time to distinguish between intracloud and cloud-to-ground lightning from the optical signatures

alone. While this is a limitation, it is more important to determine the total lightning rate than either just the cloud-to-ground or intracloud rate.

3.4.2 **Mathematical Description**

The purpose of this section is to graphically portray the mathematical algorithm employed by the LCFA to assemble individual events into groups and flashes by "walking through" a typical GLM data scenario. In this illustrative exercise, all times indicated are times after the first event time. Also, all of the data is considered to be within the same region. Numbers indicate event numbers while lowercase letters represent the groups. The flashes are designated by capital letters. Each subsequent section describes how the algorithm processes the events that occurred at that integration time. For the purpose of this demonstration, it is assumed that there were no events prior to the events at time 0 and that the pixel grid is 0.02° wide in latitude and longitude. In general, the latitude/longitude grid in earth-based coordinates and the pixel grid will not be the same size or co-registered. In addition, the times will be time from the start of the epoch.

Time = 0 ms: The first time integration is shown in Figure 5. Three (**1, 2, 3**) events occur at this time integration. Since the events are simultaneous and register in adjacent (i.e., neighboring or diagonal) pixels, they are collected into a single group (**a**). The group is assigned a new parent flash (**A**). Note that the flash information is not moved to the output stream/files until the flash is completed (DEAD).

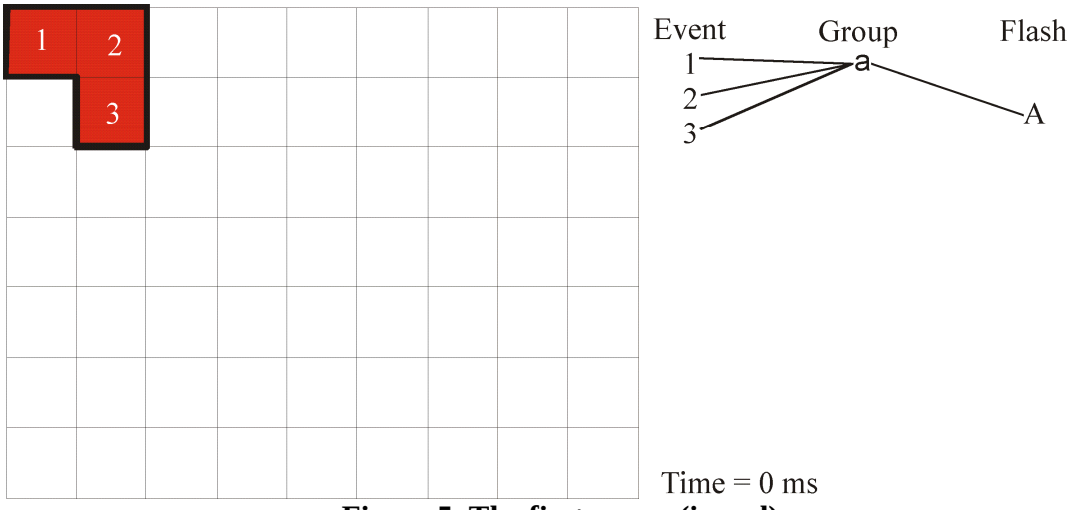

Figure 5. The first group (in red).

Time = 100 ms: The next time integration with data is shown in Figure 6. At this time (100 ms after the first one), there are three more events (**4, 5, 6**). As in the previous case, these three new events are all assigned to a new group (**b**). These events are not assigned to group **a** since they occur at a different time. Since group **b** is within 16.5 km of group **a** (actually, they touch), and the groups occur within 330 ms of each other, group **b** is assigned to flash **A**.

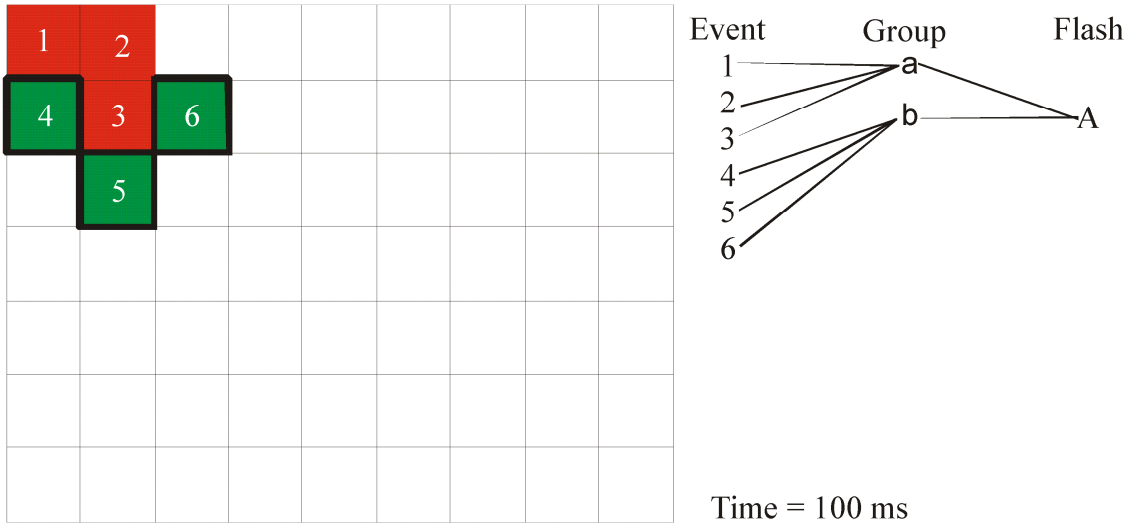

Figure 6. The next group (in green).

Time = 350 ms: The next integration time with data is shown in Figure 7. The time is 350 ms after the time of the first events, but only 250 ms after the time of the last events. At this time there are four (**7, 8, 9, 10**) more events. Events **7** and **8** are adjacent to each other and are assigned to a new group (**c**). Events **9** and **10** are not adjacent to events **7** and **8**, but are adjacent to each other. They are assigned to another new group (**d**). Since group **c** is within 330 ms of the last group of flash **A** (250 ms) and is also within 16.5 km of the parts of flash **A**, group **c** is assigned to flash **A**. Although group **d** also occurred within 330 ms of the last group of flash **A**, it is greater than 16.5 km away from any part of flash A so it is assigned to a new flash (**B**).

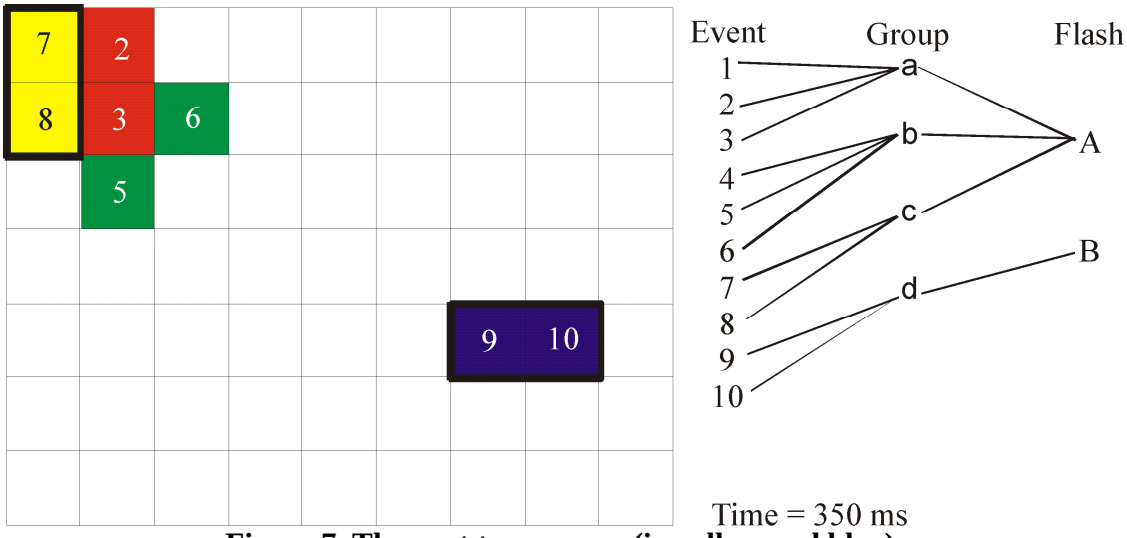

Figure 7. The next two groups (in yellow and blue).

Time = 400 ms: Figure 8 shows the next integration time with data. The time is 400 ms after the first events and 50 ms after the latest events. Two more events occur (**11, 12**) at this time. These two events are at the same time, but they are not adjacent to each other. They are assigned to two new groups (**e** for **11** and **f** for **12**). The two new groups are less

than 330 ms (50 ms) from the time of the last group of flash **B** and are within 16.5 km (adjacent) of the parts of flash **B** so the two groups are assigned to flash **B**.

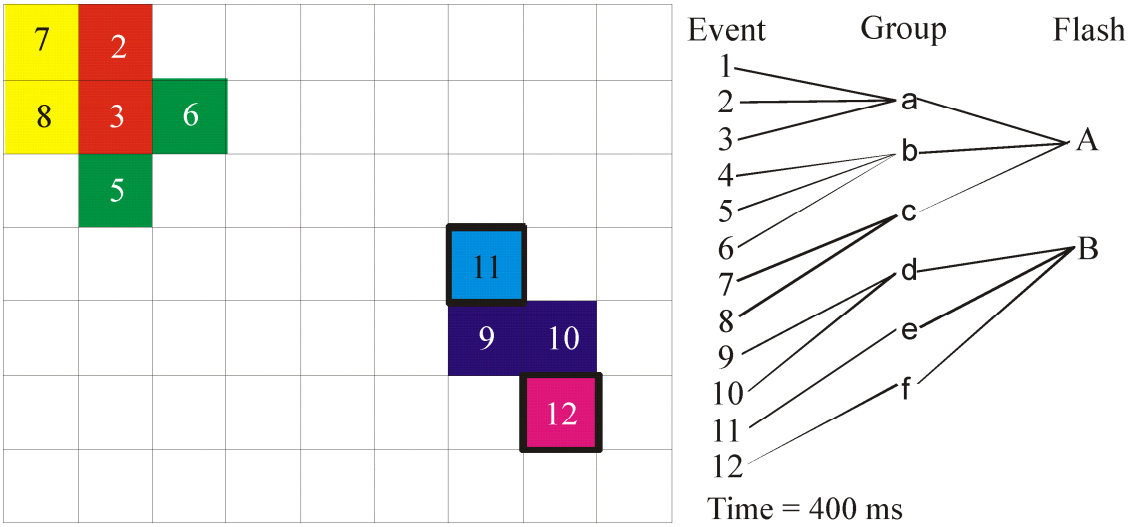

Figure 8. Two additional single event groups (in cyan and magenta).

Time = 750 ms: The last time with events (for this example) is shown in Figure 9. At this time integration, 750 ms after the first events and 350 ms after the last events, there are two new events (**13, 14**). The events are not adjacent, so they are assigned to two new groups (**g** for **13** and **h** for **14**). Group **g** overlaps the parts of flash **A**, however, it has now been 400 ms (greater than 330 ms) since the last group associated with flash **A**. Therefore, group **g** is assigned to a new flash (**C**). Group **h** is not within 16.5 km of any current flash, so it is assigned another new flash (**D**). Since flashes **A** and **B** have now 'expired', their summary statistics are sent to the output streams and files as bandwidth and processing power permits. Later, 330 ms after the last events for flashes **C** and **D**, they are also sent to the output streams/files.

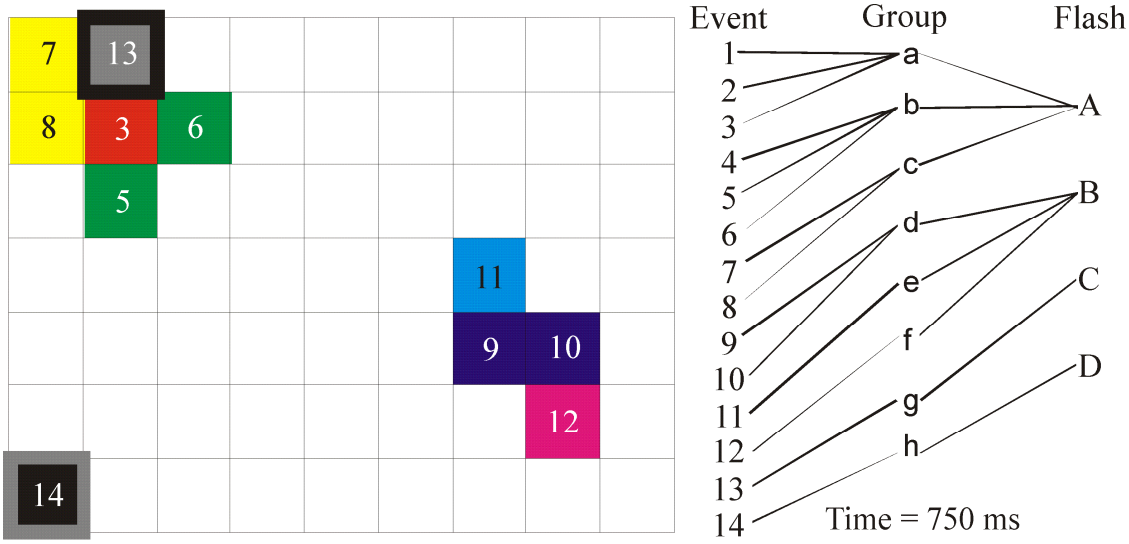

Figure 9. Two additional single event groups (in grey and black).

Summary Data: In the example data processing sequence just described, there were fourteen events, eight groups, and four flashes (see Tables 3 and 4). This example shows how the GLM algorithm will convert events into groups and flashes. The full list of what will be output is listed in Tables 5, 6, and 7. During the GLM mission, the start-time of the flashes (the time of the first event in the flash) will be a relative time (seconds from mission epoch, currently 1/1/2000). The group (first event time of the group) and event times will be in GLM frames (2 ms) from the start of the parent flash to minimize data transmission bandwidth (see Tables 5, 6, and 7). Two alternate output orders for the data from this example are shown in Tables 8 and 9.

Table 3. Resultant Flash Data

Flash ID	Start Time	Delta Time	Event Count	Lat/Lon Count	Child Count	Child IDs
A	0	350	8	6	3	a, b, c
B	350	50	4	4	3	d, e, f
C	750	0	1	1	1	g
D	750	0	1	1	1	h

Table 4. Resultant Group Data

Group ID	Parent ID	Group Time	Event Count	Lat/Lon Count	Child Count	Child IDs
a	A	0	3	3	3	1, 2, 3
b	A	100	3	3	3	4, 5, 6
c	A	350	2	2	2	7, 8
d	B	350	2	2	2	9, 10
e	B	400	1	1	1	11
f	B	400	1	1	1	12
g	C	750	1	1	1	13
h	D	750	1	1	1	14

Table 5. Event Data

bits	Description
13	uint - delta GLM frames (~2 ms) from the start of the grandparent flash to this event
31	int - lat/lon x 100 per ea 16 bit long (signed), 15 bit lat (signed)
12	uint - event optical energy density ($\mu Jm^{-2}sr^{-1}\mu m^{-1}$)
10	uint - parent group index
6	QA bits
72	Total (9 bytes)

Table 6. Group Data

bits	Description
13	uint - delta GLM frame (~2 ms) count from the start of parent flash to this group
31	int - lat/lon x 100 per ea 16 bit long (signed), 15 bit lat (signed)
11	uint - # of GLM frames (~2 ms) in group (max 4.1 s)
19	uint - footprint, scaled log value
14	uint - group optical energy density ($\mu Jm^{-2}sr^{-1}\mu m^{-1}$)
9	uint – parent flash index
11	uint - event count-1 (max 2049)
11	uint - first event index
9	QA bits
128	Total (16 bytes)

Table 7. Flash Data

bits	Description
56	uint - time, 24 bit Day, 32 bit µs/day, epoch 1/1/2000 (time of first event in flash)
31	int - lat/lon x 100 per ea 16 bit long (signed), 15 bit lat (signed)
13	uint - # of GLM frames (~2 ms) in flash (max 16.4 s)
20	uint – footprint, scaled log value
15	uint - flash optical energy density ($\mu Jm^{-2}sr^{-1}\mu m^{-1}$)
10	uint - group count-1 (max 1025)
10	uint - first group index
13	QA bits
168	Total (21 bytes)

Table 8. Output Flash/Group/Event Order from the Example Processing

Sequence Number	Item	Item Level	Sequence Number	Item	Item Level
0	A	Flash	13	d	Group
1	a	Group	14	9	Event
2	1	Event	15	10	Event
3	2	Event	16	e	Group
4	3	Event	17	11	Event
5	b	Group	18	f	Group
6	4	Event	19	12	Event
7	5	Event	20	C	Flash
8	6	Event	21	g	Group
9	c	Group	22	13	Event
10	7	Event	23	D	Flash
11	8	Event	24	h	Group
12	B	Flash	25	14	Event

Table 9. Alternate Output Flash/Group/Event Order from the Example Processing

Sequence Number	Item	Item Level	Sequence Number	Item	Item Level
0	A	Flash	13	2	Event
1	B	Flash	14	3	Event
2	C	Flash	15	4	Event
3	D	Flash	16	5	Event
4	a	Group	17	6	Event
5	b	Group	18	7	Event
6	c	Group	19	8	Event
7	d	Group	20	9	Event
8	e	Group	21	10	Event
9	f	Group	22	1	Event
10	g	Group	23	12	Event
11	h	Group	24	13	Event
12	1	Event	25	14	Event

3.4.3 Algorithm Output

The primary output from the LCFA is summarized in Tables 5 to 11. Tables 5, 6, and 7 present the proposed bit sizes and total byte widths for the event, group, and flash data. Tables 8 and 9 present possible data ordering and Table 10 presents the designations for the QA bits while Table 11 estimated bandwidth needed to transmit the GLM data given the estimated rates listed in the table and the byte sizes listed in Table 5. Note that the parameter labels in Tables 3 and 4 are not the official parameters, but are used just to help with understanding the flash clustering exercise in the section. The official parameters and their sizes are listed in Table 5 and described in detail in the next paragraph.

Table 10. QA Bits for GLM LCFA

Level	Bit	Designation	Source
Flash	0	Flash Artificially Terminated (reasons in next bits)	Clustering
	1	Too Many Groups	Clustering
	2	Too Long (too much time)	Clustering
	3	Flash Buffer Overflow	Clustering
	4	Group Buffer Overflow	Clustering
	5	Event Buffer Overflow	Clustering
	6	Processing Capacity Exceeded	Clustering
	7	Parallax Warning	Input
	8	Ephemeris Warning	Input
	9	Event Time Warning	Input
	10	Flash Bypassed “Mark Non -Lightning Groups & Flashes” Section	Filtering
	11	Non-Lightning Source Probability (bit 0)	Filtering
	12	Non-Lightning Source Probability (bit 1)	Filtering
Group	0	Group Artificially Terminated (reasons in next bits)	Clustering
	1	Too Many Events	Clustering
	2	Too Long (too much time)	Clustering
	3	Group Buffer Overflow	Clustering
	4	Event Buffer Overflow	Clustering
	5	Group Bypassed “Mark Non-Lightning Groups &Flashes”	Clustering
	6	Non-Lightning (Noise) Warning bit 0	Filtering
	7	Non-Lightning (Noise) Warning, bit 1	Filtering
	8	Reserved for Future Use	
Event	0	Event Bypassed “Mark Non-Lightning Events” section	Filtering
	1	Non-Lightning (Noise) Warning,bit 1	Filtering
	2	Non-Lightning (Noise) Warning, bit 0	Filtering
	3	Reserved for Future Use	
	4	Reserved for Future Use	
	5	Reserved for Future Use	

Time: Each flash and child/grandchild group/event sets will need a precision time tag. The time of each item (flash/group/event) will be the time of the first event in that item. Flashes will be provided with the full GOES-R time tag which is a 56 bit structure representing the number of seconds from the epoch time (currently 1/1/2000) and the number of microseconds from the start of the day. Note that if the base time representation of the mission changes from this, the time will be stored in whatever

format designated for the mission (as long as the time has sub-millisecond resolution). The times for the groups and events linked to the parent/grandparent flash will be a delta-time between the start of the flash (first event in the flash) represented by the number of GLM frames (2 ms) between the start of the flash and the start of the individual group or event. For example, in Table 4, the time associated with group "a" will be 0 (started at the same frame as the parent flash) and the time associated with group "c" will be 175 (started 350 ms after the start of the parent flash, which is 175 GLM frames later). Determining the precision time of a particular group or event will require the user to take the time of the parent/grandparent flash and adding the delta-frame count (converted back to milliseconds) of the group/flash. This extra step is justified to maintain the minimum bandwidth for the GLM data.

Table 11. Estimated Bandwidth Needed to Transmit the GLM LCFA Data

Value	Description/units
10,000	peak event rate per sec (528 0 events/sec is vendor estimate)
5000	sustained event rate per sec (3120 events/sec is vendor estimate)
40	mean events per flash (based on Mach cluster paper)
4	mean events per group (based on Mach cluster paper)
10	mean groups per flash (based on Mach cluster paper)
Peak Event Rate (with parent and child indices; compressed event/group time)	
720,000	bits/sec – events
320,000	bits/sec – groups
42,000	bits/sec – flashes
1,082,000	total bits/sec
Sustained Event Rate (with parent and child indices; compressed event/group time)	
360,000	bits/sec – events
160,000	bits/sec – groups
21,000	bits/sec – flashes
541,000	total bits/sec

Location: Each flash and child/grandchild group/event will need a precision geolocation on the Earth. For each level (flash/group/event) that will be provided by a 16-bit scaled Longitude and a 15-bit scaled Latitude (total of 31 bits). The scaling is nominally 100 (Lat/Long multiplied by 100 and then rounded to the nearest integer and masked to 15 bits for the Latitude and 16 bits for the Longitude). This scaling will provide a signed range for Longitude from -327.68° to +327.68°. Note that if the GLM instrument is stationed such that the Longitude may exceed these limits, a different scaling may be needed. The Latitude signed range will be from -163.84° to +163.84°. This range should cover any possible station and orientation for the GLM instrument. The location associated with the events will be that provided by the GLM L0-L1b algorithm. The locations associated with the groups and flashes will be the amplitude weighted mean location based on the events in the particular group or flash. A specific pseudocode example is shown here (additional details/considerations are provided in Appendix 4):

```
Amplitude_Weighted_Group_Latitude = SUM(over i)(event_latitude
      * event_amplitude)/SUM(over i)(event_amplitude);
```

Delta Time: Each flash and child group will need a precision delta time for the flash/group. It will be the time difference between the first event and the last event in a flash/group. It will be stored in number of GLM frames (2 ms) to conserve bandwidth. Note that with the current definition of a group, this value will always be 0. Future enhancements of the GLM LCFA may include groups with events crossing GLM frame boundaries which may result in groups having delta-time values greater than zero.

Footprint: Each flash and child group will have a footprint. The footprint is defined as the locations on the ground covered by all events in the specific group/flash. This footprint can be as small as a single event or as large as 100's of square kilometers for large flashes. Again, to conserve bandwidth, the actual footprint (in square kilometers) will be scaled by a log factor and stored in 20 bits for flashes and 19 bits for groups. Values that are too big to fit within the 19 or 20 bit masks will be truncated to the largest value (19 or 20 bits). The actual scaling factor is listed in the conversion pseudocode example shown here:

```
Scaled_Footprint =
LOG10(actual_footprint_km2/min_pixel_footprint_km2)*1.66e5;
```

Radiance: We loosely use the word "radiance" to refer to what is actually a solid angle averaged spectral energy density (SAASED). The SAASED for a group or a flash has, by heritage, been taken as the sum of the SAASEDs of the events composing the group or flash. However, actual formulas for the group and flash SAASEDs are more complicated [see Appendix 2: *A Physical Understanding of the Event, Group, and Flash "Radiance" Data Products* for details]. The present AWG GLM LCFA simply sums energy densities (i.e., the present LCFA computes *ER*, *GR*, and *FR* defined in Appendix 2). However, the optimal "energy products" would be (e_E, e_G, e_F) defined in Appendix 2; these products are most accurately obtained when a look-up table containing the solid angle of each pixel is provided. However, Appendix 2 also provides approximative expressions for (e_E, e_G, e_F) when the look-up table is not provided. In the present AWG GLM LCFA, the "event radiance" *ER* is the 12-bit number from the GLM L0-L1b algorithm. The "group radiance" *GR* is defined by the sum of the *ER* in that group scaled by a factor of 8 and truncated to fit in 14 bits. If the *GR* will not fit in 14 bits, the maximum value that will fit in 14 bits will be sent. The "flash radiance" is estimated by the sum of the *ER* in that flash scaled by a factor of 16 and truncated to fit in 15 bits. If the *FR* will not fit in 15 bits, the maximum value that will fit in 15-bits will be sent.

Parent/Child Referencing: The flash/group/event sets will be internally linked with references sufficient to reconstruct the parent/child relationships between the flash and child groups and between groups and child events. The links are needed because not all information needed to fully describe a flash are contained in the flash information (Table 7). The flash data includes the number of groups in the flash and an index that links the flash to the first group. Implied in the flash/group clustering is that groups associated with a flash are stored/transmitted near the flash information. The "index to the first group" for a flash is equivalent to a pointer in languages such as C. The index allows for quick referencing of the groups associated with the flash. The group data includes the number of events in the group, an index that links the group to the first event in the group, and an index that links the group to its parent flash. Implied in the group/event clustering is that events associated with a group are stored/transmitted sequentially after

the group information. The events have an index that links the event to its parent group. Using the example in Section 3.4.2, the data for this set of flashes/groups/events would be stored in the order shown in Tables 8 or 9.

For the storage sequence in Table 8, the "first group index" for "A" would be 1 (the first group of flash "A" is at position 1) and the "group count-1" would be 2 (3-1 = 2). For group "a", the "parent flash index" would be 0 (flash "A" is at position 0 in the sequence) and the "event count -1" would be 2 (3-1 = 2). For event "1", the "parent group index" would be 1 (group "a" is at position 1). The "first group index" for "C" would be 21 (the first group in flash "C" is at position 21). The "group count – 1" would be 0 (1 – 1 = 0). For group "g", the "parent flash index" is 20 (flash "C" is at position 20 in the sequence). The "event count – 1" for group "g" would be 0 (1-1=0). The "first event index" for group "g" would be 22. For event "13", the "parent group index" would be 21. The indices do not need to be universally unique over the lifetime of the GLM data. They only need to be unique over the transmission packets and archival storage files so that 1) given a group, the parent flash can be found and the child events can be found, 2) given a flash, the child groups can be found, and 3) given an event, the parent group can be found.

For the storage sequence in Table 9, the "first group index" for "A" would be 4 (the first group of flash "A" is at position 4) and the "group count-1" would be 2 (3-1 = 2). For group "a", the "parent flash index" would be 0 (flash "A" is at position 0 in the sequence) and the "event count -1" would be 2 (3-1 = 2). The first event index would be 12 (the first event for group "a" is at index 12). For event "1", the "parent group index" would be 4 (group "a" is at position 4). The "first group index" for "C" would be 10 (the first group in flash "C" is at position 10). The "group count – 1" would be 0 (1 – 1 = 0). For group "g", the "parent flash index" is 2 (flash "C" is at position 2 in the sequence). The "event count – 1" for group "g" would be 0 (1-1=0). The "first event index" for group "g" would be 24. For event "13", the "parent group index" would be 10. The indices do not need to be universally unique over the lifetime of the GLM data. They only need to be unique over the transmission packets and archival storage files so that 1) given a group, the parent flash can be found and the child events can be found, 2) given a flash, the child groups can be found, and 3) given an event, the parent group can be found.

QA Bits: Each level (event/group/flash) will have its own set of QA bits. Some will be inherited from the L0-L1b processing, some will be calculated during clustering, and others will be designated during output. The QA bits and their creation are listed in Table 10 and explained in Sections 5.4.1, 5.4.2, and 5.4.3.

Metadata: Table 12 shows the metadata is needed for Product Monitoring.

Table 12. Metadata for Lightning Product

Flash	***Group***	***Event***
Flash count	**Group count**	**Event count**
Number of QA flag values	**Number of QA flag values**	**Number of QA flag values**
Definition of each QA flag value	**Definition of each QA flag value**	**Definition of each QA flag value**
Number of flashes with each QA value	**Number of groups with each QA value**	**Number of events with each QA value**

4. Test Data Sets and Outputs

The datasets used to test the performance of the LCFA, and the types of LCFA outputs, are discussed below. Detailed test procedures and LCFA output analyses are beyond the scope of this document; the reader is referred to the *Algorithm Implementation and Test Plan (AITP)* document (Appendix 1) for these details.

4.1 Simulated Input Data Sets

A key enabling activity for development and testing of the LCFA is the development of GLM proxy datasets derived from other ground-based and satellite sensors. There are three main types of GLM proxy datasets that are presently being generated: a simple proxy based on National Lightning Detection Network (NLDN) data, a proxy based on empirically mapping ground-based VHF lightning data into optical data, and a proxy based on simply resampling heritage LIS data. The three proxies are discussed below.

4.1.1 NLDN Data

NLDN data was used to identify a very active storm day (July 21, 2003; see Figure 6) that was used to try to "break" the LCFA. That is, the realistic high flash rates of this storm day allow us to test the LCFA computation speed and determine if the algorithm can comply with data latency requirements. The total flash rates are estimated from the NLDN ground flash rates by making reasonable assumptions about the cloud per ground flash ratio. The ratio averages about 2.94 (Boccippio et al., 2001).

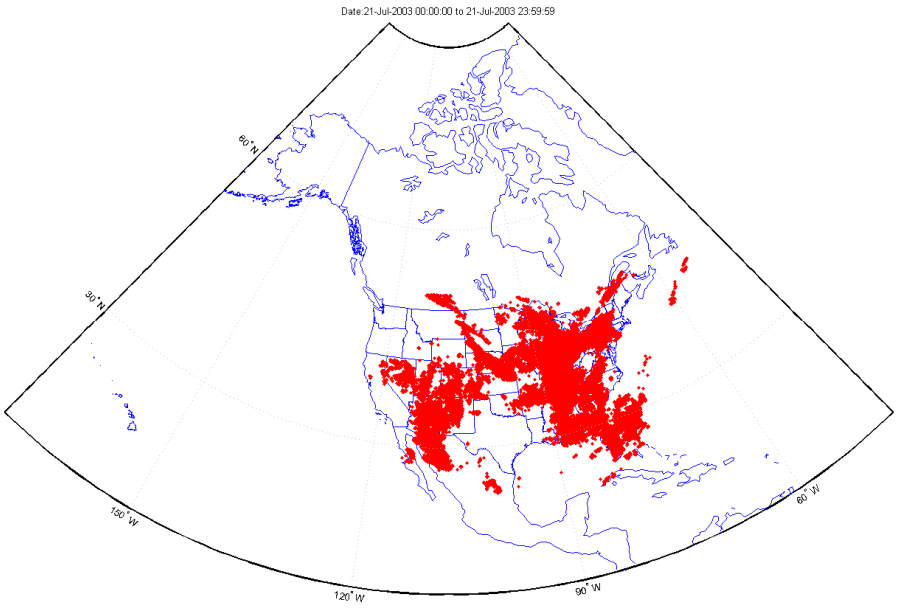

Figure 10. The 21 July 2003 NLDN proxy dataset.

4.1.2 VHF Data

To create a proxy that contains realistic spatio-temporal thunderstorm evolution, concurrent LIS and ground-based lightning VHF observations are being compared to construct an empirical model that is capable of mapping VHF lightning observations to optical emissions. This approach, for example, allows one to simulate the spatio-temporal characteristics of event-based (pixel-level) data detected by a GLM by applying the empirical model to a database of VHF lightning observations from several thunderstorms. Figure 7 shows the display output of a tool that animates the coincident data files.

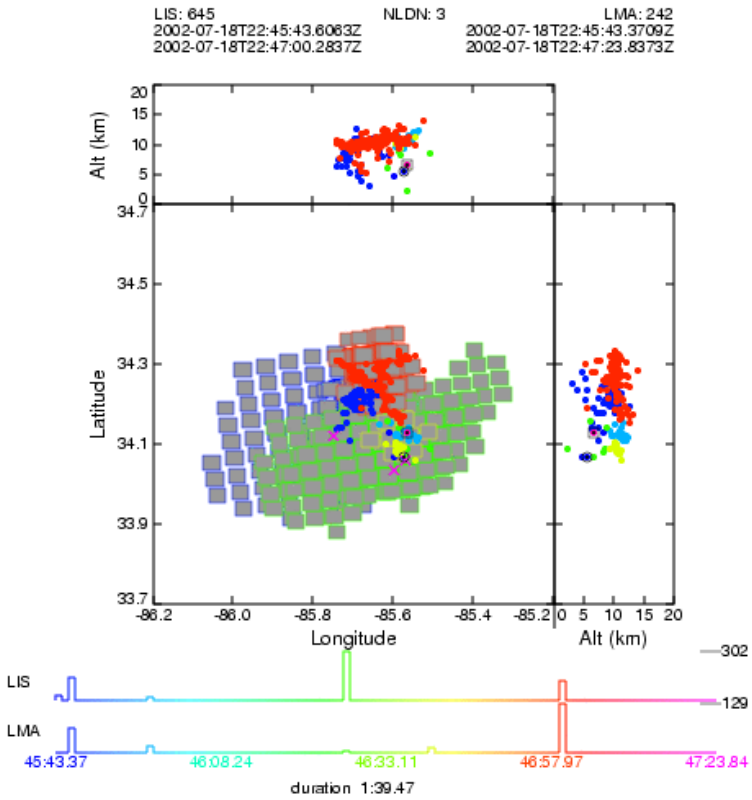

Figure 11. Plot comparing LIS (squares), LMA (dots), and NLDN (Xs) datasets.

LIS & VHF Lightning Mapping Array (LMA) data compare very favorably with regard to several variables (time, geolocation, flash algorithms, detection efficiency). Database statistics and fractional coincidences are provided in the list below and in Figure 12:

- May 2002 — July 2007
- 2525 total overpasses
- 1066 where both saw lightning
- LIS: pixels —> v4.1 algorithm —> flashes
- LMA: sources —> v1 algorithm —> flashes
- Tolerance: ± 0.20 s, 35km range
- LIS flashes = 56570, LMA(>49 srcs) = 44339
- Mean LMA sources per coincident LIS flash = 66
- Mean multiplicity of LMA flashes: 1.19 per coincident LIS flash

- LMA flash splitting or LIS flash smearing
- Geolocation comparison:
- mean error: (-0.1, 0.4) km
  - from LIS —> LMA
  - slightly NNW

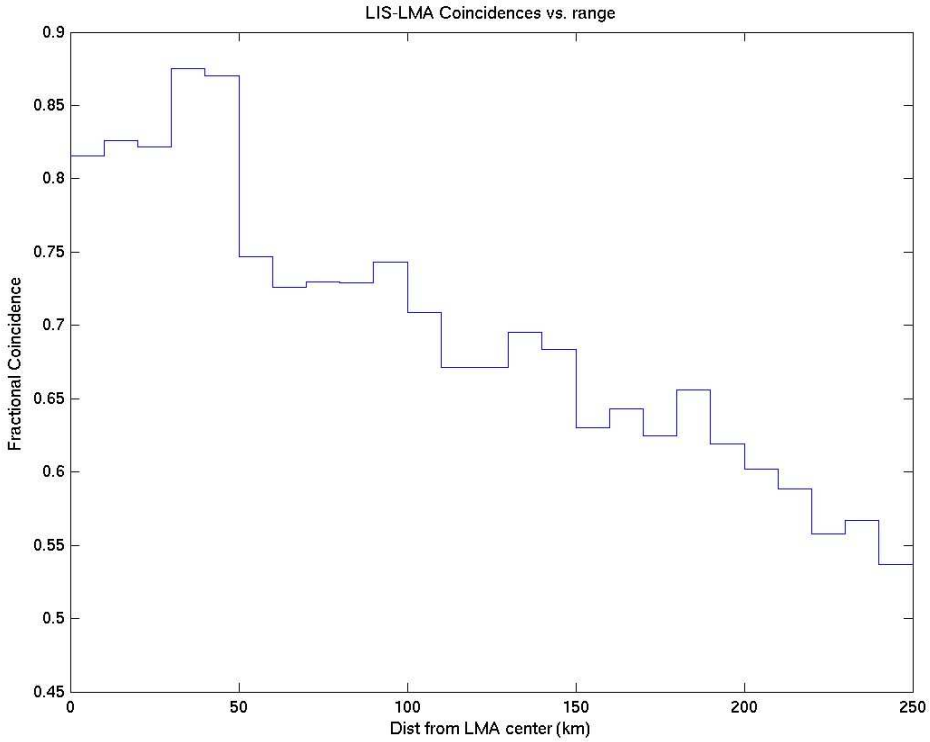

Figure 12. The fractional coincidence of LIS and LMA flashes as a function of range.

4.1.3 Re-Sampled LIS Data

The LIS observes lightning at a higher pixel resolution (4 km at nadir) than the GLM, so a "re-sampled" LIS dataset will allow performance modeling of GLM characteristics over its entire FOV and diversity of background scene viewing conditions. It is relatively easy to resample LIS data at a lower spatial resolution, and the resulting proxy is adequate for completing tests that only require "snapshots" of lightning. In addition, we will also use heritage Optical Transient Detector (OTD) data near boresight, since this is already ~8 km resolution.

4.2 Output From Simulated Input Data Sets

Details of LCFA output using simulated input data is provided in the *Algorithm Implementation and Test Plan (AITP)* document (Appendix 1).

4.2.1 Precisions and Accuracy Estimates

The LCFA simply "passes on" the accuracy and precision of the GLM instrument and Level 0 to Level 1b processing. The LCFA will not modify the incoming event data accuracy (it shall simply "pass on" the GLM values). Similarly, the instrument *flash* detection efficiency of 70% or better with a False Alarm Rate (FAR) of < 5% will not be significantly altered by the LCFA; very minimal changes in flash detection efficiency will occur due to LCFA flash merging or flash splitting (see *Algorithm Implementation and Test Plan (AITP)* document (Appendix 1) for specific statistics). All input lightning events from the GLM will be clustered into groups & flashes. The LCFA will compute flash location using a weighted optical centroiding approach. The flash location accuracy will be < 5km (~ ½ pixel resolution), and the lat/lon location data will be stored to 0.01° (1.1 km & 1% precision). As discussed in section 2.1, for data rates < 20,000 events s^{-1}, the LCFA uses 1 s for input buffering, less than 1 s for processing, and 1 s for output buffering. The LCFA has been allocated 4 s of the 20 s VAGL

4.2.2 Error Budget

A comparison of LCFA output data (under heading "actual") against requirements is provided in the error budget summary below (Table 13). The values have been discussed in the previous section. We do not anticipate much change in the error budget limitations as the LCFA is already highly mature.

Table 13 shows that the GLM products meets the F&PS product measurement accuracy and product measurement precision requirements; however, whether or not the "5 km (or less) Mapping Accuracy" requirement is actually met depends on how well parallax error will be mitigated.

Table 13. Error Budget Summary

Name	Mapping Accuracy[A]		Measurement Accuracy[A]		Vendor Allocated Ground Latency (VAGL)		Product Measurement Precision (flash FAR)	
	Actual	F&PS Specification	Actual	F&PS Specification	Actual	F&PS Specification	Actual	F&PS Specification
Flashes	<5 km	5 km	>70%	70% flash detection	<3 s	20 s (4 s)	<5%	5%

5. Practical Considerations

5.1 Numerical Computation Considerations

The data latency requirement is 20 sec, which includes 10 sec for Instrument Vendor processing of Level 0 to Level 1b, and ~ 4 sec for the LCFA processing of Level 1b to Level 2. Therefore, for the LCFA processing chain, if the data rate becomes too high, lower priority code sections (i.e., "Mark Non-Lightning Events" and "Mark Non-Lightning Groups & Flashes") will be suspended to allow for the production of the real time lightning product (under the assumption that a lower quality on-time product is

better than a higher quality delayed product). The exact procedure for dropping the filtering depends on how far behind the LCFA can get before it is considered out of latency.

5.2 Programming and Procedural Considerations

One of the assumptions made is that the coding environment will be similar to that of products such as VxWorks or Linux, which allow for task spawning, blocking, and task terminations.

5.3 Quality Assessment and Diagnostics

5.3.1 Code Processing Speed

The primary quality assessment of the LCFA is that it process a seconds worth of GLM L1b data into events, groups, and flashes in less than one second. If the code cannot 'keep up with" the input data stream, the code will "fall behind" and fail latency requirements. Therefore, the single most critical measurement of data quality and diagnosis is the amount of time it takes to process one second of GLM data. The current Framework version of the GLM LCFA measures this quantity, but does not output it as part of the data stream (due to bandwidth restrictions). However, since different computers have different processing speeds (based on code efficiency, processor count, memory latencies and speed, and disk I/O limitations), the values for this parameter may be different depending on which machine and the maximum L1b data rates encountered. The only requirement is that the code takes less than one second to process one second of GLM L1b data.

One expected aspect of the GLM L1b data is that it will be very "bursty". Event rates can range from 0 events per second to upwards of 20,000 events per second, or higher. Therefore, a second-by-second measurement of the code processing rate is not the best measure of the code efficiency. Rules on how long a time period that can be used to "average" the bursty data rates is unknown (set by standards outside of our control). However, a reasonable time window would be 5 seconds. As long as the code is able to process 5 seconds of data in less than 5 seconds of computer time, the code should pass the latency requirements. However, these values may need to be changed based on outside documentation.

If the code does not meet the latency requirements, it can be scaled back to quicken event processing. The first area where the code can scale back processing is to bypass the detection of non-lightning source events, groups, and flashes. If the processing gets far enough behind that it triggers the bypass, the incident will be marked in the QA bits (see Section 5.4). One of the main drains on processing speed is the maintenance of "ALIVE" flashes and groups (flashes and groups that can still accept new groups and events). To speed processing, ALIVE flashes and groups can be terminated early ("KILLED") and then dumped to the output streams/files. This will also result in setting a QA bit.

5.3.2 Clustering Results

The next critical quality assessment of the LCFA is that it clusters the events into the proper groups and flashes. The definition of the clustering scheme leaves little room for assessment (it is either wrong or right). However, if flashes and groups are KILLED because they exceed some limit, those flashes and groups have QA bits set to indicate why the flash or group was terminated early.

5.3.3 Marking Non-Lightning Events, Groups, and Flashes

There are two currently TBD routines that will use information gleamed from the testing of the GLM instrument to mark events, groups, and flashes that are likely not due to lightning. The makeup of the routines is currently unknown as we have no idea of what type of "artifacts" will be created by the GLM instrument. To allow for "fuzzy" determination of non-lightning event, groups, and flashes, the QA bits associated with the non-lightning items have 4 levels. These will correspond approximately to 0% chance of the item being non-lightning, 33% chance, 67% chance, and 100% chance of being from a non-lightning source.

5.3.4 Diagnostics Imported from L0-L1b Code

The LCFA will import and pass on QA warnings from the L0-L1b code. These include parallax warning, ephemeris warning, and event time warning. In future versions of the code, the LCFA may also check the data and set the parallax, ephemeris, and event time warning QA bits if they need to be set based on calculations made in the LCFA.

5.4 Exception Handling

5.4.1 Flash QA Bits (13 Bits Total)

- ***Flash Termination (1 bit).*** Was the flash terminated by an exception, or did the flash expire naturally? If there are no groups that can be added to a flash within the time/space parameters set by the GLM team (currently 330 ms and 16.5 km) a flash will terminate and be written to the output. If a flash is terminated early due to some problem (later QA bits), the flash is flagged as being "killed". There are both practical and theoretical reasons to limit the size of flashes. Flashes have a physical limit to how long they can be. However, the exact cutoff for the time limit for a flash is dependent on the computer hardware used to process the GLM data. The code should be able to handle any reasonable time limit for a flash (from a few seconds to no more than 10 seconds). Flash sizes are limited to the size of the parent storm plus any propagation into the surrounding area. Again, flashes cannot be infinitely sized, but the exact limit on a flash size is not yet set and depends on both physical and computational limits. The code should be able to handle sizes up to thousands of square kilometers. Practical limitation also may come into play. Computers only have the ability to handle a limited number of items (and their pointer links) in memory at one time. Depending on the computer system used to process GLM data, it is not clear if the practical or the theoretical limits will come into play first.

- ***Reason for Early Flash Termination (7 bits).*** Why was the flash terminated early? The reasons are (1 bit each as the flash could be terminated for more than one reason):
  - *Too many children.* We may put limits on the # of children (groups) that a flash can have. The # of children can also be limited due to time or buffer restrictions.
  - *Flash reaches time limit.* We may put limitations on the total time length a flash may have. The time limit may also be restricted due to buffer limitations.
  - *Flash buffer overflow.* The total number of active and dead but not written to the output flashes may have exceeded the capacity of the flash buffer. If so, a flash will need to be terminated (and written to the output) to allow for more flashes.
  - *Group buffer overflow.* The total number of active or dead but not written to the output groups may have exceeded the capacity of the group buffer. If so, a flash will need to be terminated (and written to the output) to allow for more groups.
  - *Event buffer overflow.* The total number of active or dead but not written to the output events may have exceeded the capacity of the event buffer. If so, a flash will need to be terminated (and written to the output) to allow for more events.
  - *Processing capacity exceeded.* A flash (or a series of flashes) may need to be terminated if the GLM data processing code gets too far behind. The quickest way for the code to "catch up" is to terminate the current set of active flashes and write them to the output.
- ***Parallax Warning (1 bit).*** This bit will be set if there is reason to believe the location of the flash may be in error due to parallax issues.
- ***Ephemeris Warning (1 bit).*** This bit will be set if there is reason to believe the location of the flash may be in error due to ephemeris problems.
- ***Event Time Warning (1 bit).*** This bit is set if there is reason to believe the time of the flash may be in error due to problems with the event times.
- ***Flash Bypassed Non-Lightning Probability Filter (1 bit).*** This bit is set if the flash did not go through the non-lightning flash detection filter.
- ***Non-Lightning Probability Bits (2 bits).*** Probability (0, 33%, 67%, 100%) the flash is due to non-lightning sources.

5.4.2 Group QA Bits (9 Bits Total)

- ***Group Termination (1 bit).*** Was the group terminated by an exception, or did the group expire naturally? If there are no events that can be added to a group within the time/space parameters set by the GLM team (currently 0 ms and adjacency) a group will terminate and be written to the output. If a group is terminated early due to some problem (later QA bits), the group is flagged as being "killed". There are both practical and theoretical reasons to limit the size of groups. Groups have a physical limit to how large they can be. Currently, groups have no time span, that is, they happen in a single time frame. Future versions of the code may employ groups that extend in time. The code should be able to handle this possibility. However, the exact cutoff for the time limit for a future group is yet

to be determined. The code should be able to handle any reasonable time limit for a group (up to a few seconds). Group sizes are limited to the size of the parent storm plus any propagation into the surrounding area. Again, groups cannot be infinitely sized, but the exact limit on a group size is yet to be determined. The code should be able to handle sizes up to thousands of square kilometers. Practical limitation also may come into play. Computers only have the ability to handle a limited number of items (and their pointer links) in memory at one time. Depending on the computer system used to process GLM data, it is not clear if the practical or the theoretical limits will come into play first.

- ***Reason for Early Group Termination (4 bits).*** Why was the group terminated early? The reasons are (1 bit each as the group could be terminated for more than one reason):
  - *Too many children.* We may put limits on the # of children (events) that a group can have. The # of children can also be limited due to time or buffer restrictions.
  - *Group reaches time limit.* We may put limitations on the total time length a group may have. The time limit may also be restricted due to buffer limitations.
  - *Group buffer overflow.* The total number of active and dead but not written to the output groups may have exceeded the capacity of the group buffer. If so, a group will need to be terminated (and its parent flash written to the output) to allow for more groups.
  - *Event buffer overflow.* The total number of active or dead but not written to the output events may have exceeded the capacity of the event buffer. If so, a group will need to be terminated (and its parent flash written to the output) to allow for more events.
- ***Group Bypassed Non-Lightning Probability Filters Warning Bit (1 bit).*** Group was not filtered by the non-lightning probability filter.
- ***Group Non-Lightning Probability Filter Warning Bits (2 bits).*** Chance group is not due to lightning (0%, 33%, 67%, 100%).
- ***Spare Bit (1 bit).*** Reserved for future used.

5.4.3 Event QA Bits (6 Bits Total)

- ***Event Bypassed Non-Lightning Probability Filters Warning (1 bit).*** Chance group is not due to lightning (0%, 33%, 67%, 100%).
- ***Event Non-Lightning Probability Filter Warning Bits (2 bits).*** Chance event is not due to lightning (0%, 33%, 67%, 100%).
- ***Spare Bits (3 bits).*** Reserved for future use.

5.5 Algorithm Validation

The *Product Validation Plan* discusses the plans for validating simulated GLM Level 2 lightning products (flashes, groups, events) during the pre-launch phase (current time up to time of launch) and actual GLM Level 2 lightning products during the post-launch phase (time of launch out to 2 years). The validation of simulated and actual Level 2 products serves as a fundamental means for validating the performance of the GLM Lightning Cluster Filter Algorithm. The plan describes the validation approach and

identifies correlative datasets, including those from known/planned community field campaigns or dedicated GOES-R field campaigns, which will or could be used to validate the products. The *Algorithm Implementation and Test Plan* (Appendix 1) provides additional details of pre-launch algorithm testing with focus on algorithm speed, accuracy, and resiliency (i.e., input error handling).

6. ASSUMPTIONS AND LIMITATIONS

6.1 Performance

Details of LCFA performance tests are provided in sections 4-5 of the *Algorithm Implementation and Test Plan* (AITP) document (Appendix 1). The algorithm tests fall into three possible categories: (1) Speed Tests, (2) Accuracy Tests, and (3) Resiliency Tests. There are several possible types of tests under each category, and the test themselves have facilitated algorithm definition as well as algorithm performance.

6.2 Assumed Sensor Performance

The LCFA assumes that the input data from the GLM instrument meets the performance specifications in the appropriate documents. The LCFA will also assume that the data is correct and is within the designated FOV of the GLM instrument. Data that is outside the designated FOV will be ignored. The LCFA will also assume that the GLM data is in near time order. The only exception to the time order will be corrections needed for propagation time differences between the sub satellite point and the limb. Times out of this range will cause flashes and groups to be terminated and flagged.

6.3 Pre-Planned Product Improvements

None.

References

Boccippio, D. J., K. L. Cummins, H. J. Christian, and S. J. Goodman, Combined satellite- and surface-based estimation of the intracloud-cloud-to-ground lightning ratio over the continental United States, *Monthly Weather Review*, **129**, 108-122, 2001.

Christian, H. J. and S. J. Goodman (1987), Optical observations of lightning from a high altitude airplane, *J. Atmos. Oceanic Technol.*, **4**, 701-711.

Christian, H. J., R. J. Blakeslee, and S. J. Goodman (1989), The detection of lightning from geostationary orbit, *J. Geophys. Res*., **94**, 13329-13337.

Christian, H. J., R. J. Blakeslee, S. J. Goodman, D. M. Mach, Algorithm Theoretical Basis Document (ATBD) for the Lightning Imaging Sensor (LIS), *NASA-MSFC, Earth Science Department*, February 1, 2000.

Goodman, S. J., and D. R. MacGorman, Cloud-toground lightning activity in mesoscale convective complexes, *Mon. Wea. Rev*., **114**, 2320-2328, 1986.

Goodman, S. J., H. J. Christian, and W. D. Rust, Optical Pulse Characteristics of intra-cloud and cloud-to-ground lightning observed from above clouds, *J. Appl. Met*., 1369-1381, 1988.

Goodman, S. J., and H. J. Christian. Global observations of lightning, Atlas of Satellite Observations related to Global Change, R. Gurney, J. Foster, and C. Parkinson, eds., *Cambridge University Press*, New York, 1993, 191-219.

Guo, C., E. P. Krider, The optical and radiation field signatures produced by lightning return strokes, *J. Geophys. Res*., **87**, 8913-8922, 1982.

Mach, D. M., D. R. MacGorman, W. D. Rust, and R. T. Arnold, Site errors and detection efficiency in a magnetic direction-finder network for locating lightning strikes to ground. *J. Atmos. Oc. Tech*., **3**, 67-74, 1986.

Mach D. M., H. J. Christian, R. J. Blakeslee, D. J. Boccipio, S. J. Goodman, W. L. Boeck, Performance assessment of the Optical Transient Detector and Lightning Imaging Sensor, *J. Geophys. Res.*, **112**, D09210, DOI:10.1029/2006JD007787, 2007.

Thomason, L. W., and E. P. Krider, The effects of clouds on the light produced by lightning, *J. Atmos. Sci.*, **39**, 2051–2065, 1982.

Appendix 1: Algorithm Implementation and Test Plan Document GLM Lightning Cluster-Filter Algorithm

1. OBJECTIVE

1.1 Purpose of this Document

This Algorithm Implementation and Test Plan (AITP) document for the Geostationary Lightning Mapper (GLM) Lightning Cluster-Filter Algorithm (LCFA) provides a detailed description of the various tests to be conducted to define the form of the LCFA and to verify its overall performance.

1.2 Specific Objectives

This document provides methodologies for conducting algorithm definition phase tests that examine LCFA *speed*, *accuracy*, and *resiliency*. Results from these tests are largely based on proxy dataset analyses. This document also provides a schedule for the tests, and identifies those responsible for conducting and documenting the test results. Subsequent algorithm refinement phase tests are mentioned but will not be detailed until after the LCFA critical design review (CDR).

1.3 Scope of Document

The primary focus of this document is on pre-launch evaluation of the LCFA consisting of algorithm definition and performance testing using proxy datasets derived from a variety of actual lightning measurements. This document does not comment on possible definition/performance testing of the LCFA using proxy data derived from calibration lab optical sources; this type of testing would only occur after a calibration system has been designed and assembled and the GLM instrument is made ready for calibration. GLM calibration activities are beyond the scope of this document. In addition, an assessment of the true on-orbit GLM properties (data latency, flash detection efficiency, flash location accuracy, …) is beyond the scope of this document since such an assessment depends, at a minimum, on the final GLM instrument characteristics, the final version of the level 0 to level 1b processing algorithm, and the final LCFA. Post launch validation strategies are discussed in the *Validation Plan for AWG Lightning Algorithms & GLM Level 2 Dat*a.

1.4 Related Documents

This AITP is related to the Validation Plan for AWG Lightning Algorithms & GLM Level 2 Data, the GLM LFCA Algorithm Theoretical Basis Document (ATBD), and GLM specifications in the Mission Requirements Document (MRD).

2. BRIEF DESCRIPTION OF ALGORITHM

2.1 Overview

A detailed description of the LCFA is provided in the ATBD. As an overview, the LCFA receives as input the level 1b geolocated pixel-level optical "event" data and processes this data into more convenient lightning data products that are easily utilized by the scientific research and broader operational user communities. Therefore, the LCFA must take the event data and assemble the higher level lightning data products (groups and flashes), and in so doing, it will generate derived lightning characteristics associated with these higher level products. It will also interrogate individual events on a statistical basis to see if they are associated with lightning or noise [i.e., the Lightning AWG Team does not assume that the noise filtering performed by the Instrument Vendor is perfect]. So the overarching theoretical problem for the LCFA is to assemble *legitimate* (noise-free) lightning groups and flashes from lightning event data that may or may not contain noise. This means that the LCFA, as the acronym implies, will both cluster and filter the event data. The clustering and filtering "feed" off of one another in the sense that accurate filtering produces accurate clustering and accurate clustering allows one to better filter the data. Indeed, because of the clustering, we anticipate being able to identify and filter noise that might not be identified and filtered by the Instrument Vendor. The LCFA will provide statistics that indicate how likely a flash is actually lightning.

2.2 Product Summary

There are two major products produced by the GLM software: a lightning dataset and a corresponding background dataset. The LCFA only generates the lightning dataset. To obtain these datasets, the satellite data stream needs to be decoded, filtered, clustered, and output to the appropriate file. Detailed definitions of the basic data storage classes (background, events, groups, flashes) that drive the LCFA are provided in the ATBD. Note that there will be 2 format types (netCDF4, McIDAS), with 3 base products (events, groups, flashes) for a total of 6 products. GOES Rebroadcast (GRB) will distribute all base products. The base products will be geo-located, time-stamped, and sized (i.e., characterized by a footprint). Geo-location of the groups and flashes are based on the LCFA computing an optical amplitude-weighted centroid. The base products will be provided as close to real time as the software can produce. These base products are linked by a tree structure that shows what events construct what groups, and what groups construct what flashes. In addition, note that the Ground Segment Project (GSP) considers the Instrument Vendor (Lockheed Martin) level 1b as level 2+ end products since it has been geolocated and background removed.

3. DATASETS

A key enabling activity for development and testing of the LCFA is the development of GLM proxy datasets derived from other ground-based and satellite sensors. There are three main types of GLM proxy datasets that are presently being generated: a simple proxy based on National Lightning Detection Network™ (NLDN) data, a proxy based on

empirically mapping ground-based VHF lightning data into optical data, and a proxy based on resampling heritage LIS data. The three proxies are discussed below.

3.1 National Lightning Detection Network Data

National Lightning Detection Network^{TM} (NLDN) data was used to identify a very active storm day (July 21, 2003; see Figure 10) that will be used to try to "break" the LCFA. That is, the realistic high flash rates of this storm day allow us to test if the LCFA can keep up with its processing chain given the 4 sec latency requirement (the total latency requirement is 20 sec, with 10 sec allotted to the Instrument Team in the creation of level 1b data). The total flash rates are estimated from the NLDN ground flash rates by making reasonable assumptions about the cloud per ground flash ratio. The ratio averages about 2.94.

3.2 VHF Lightning Mapping Data

To create a proxy that contains realistic spatio-temporal thunderstorm evolution, concurrent LIS and ground-based lightning VHF observations are being compared to construct an empirical model that is capable of mapping VHF lightning observations to optical emissions. This approach, for example, allows one to simulate the spatio-temporal characteristics of event-based (pixel-level) data detected by a GLM by applying the empirical model to a database of VHF lightning observations from several thunderstorms. Figure 11 shows the display output of a tool that animates the coincident data files.

3.3 Lightning Imaging Sensor Data

The LIS observes lightning at a higher pixel resolution (4 km at nadir) than the GLM, so a "re-sampled" LIS dataset will allow performance modeling of GLM characteristics over its entire FOV and diversity of background scene viewing conditions. It is relatively easy to resample LIS data at a lower spatial resolution, and the resulting proxy is adequate for completing tests that only require "snapshots" of lightning. In addition, we will also use heritage Optical Transient Detector (OTD) data near boresight, since this is already ~8 km resolution.

4. METHODOLOGY

For context, an overview of the validation strategy for all future GLM algorithms and GLM Level 2 products is provided in Figure 13. These phases are described in detail in the *Validation Plan for* AWG *Lightning Algorithms & GLM Level 2 Data* document. Note that the figure encompasses all lightning-related validation processes and interconnections and therefore extends beyond just the validation of the LCFA.

The focus of this document is on the pre-launch evaluation of the LCFA. The pre-launch evaluation phase of the LCFA can be further divided into two phases: (1) an Algorithm Definition Phase (ADP), and (2) an Algorithm Refinement Phase (ARP).

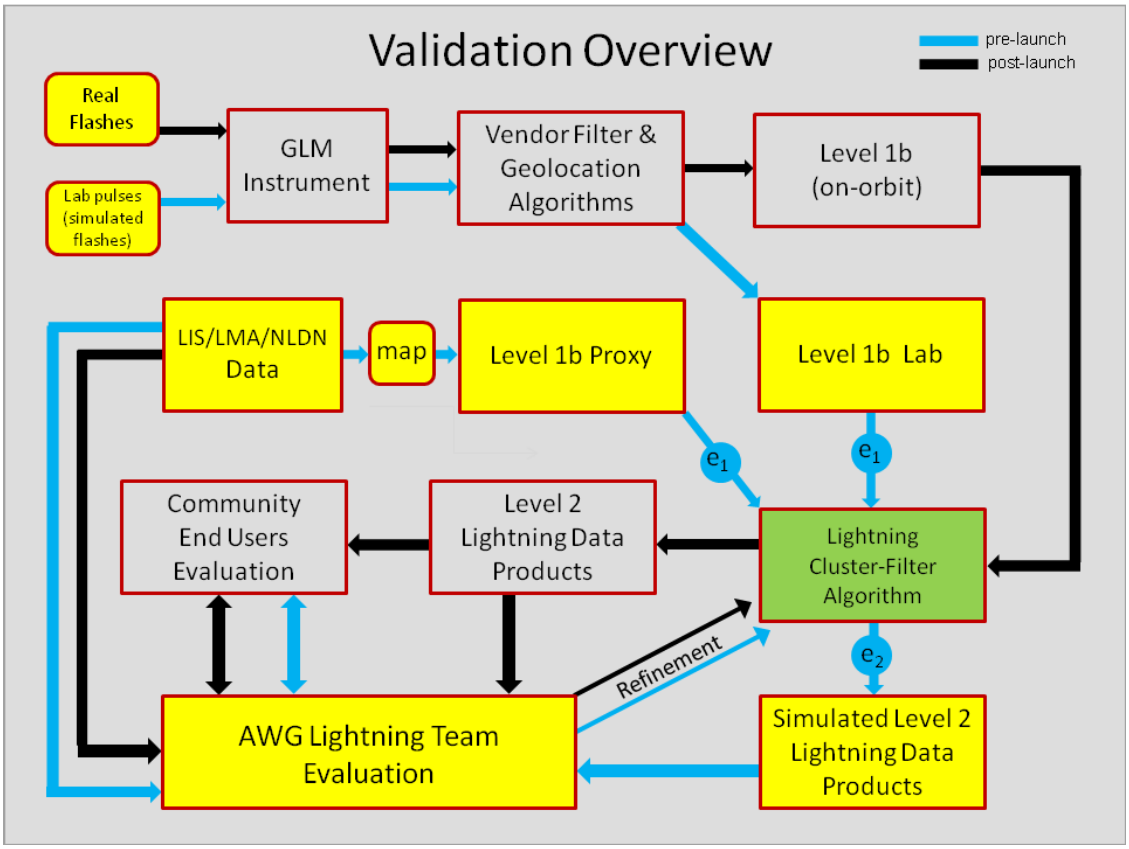

Figure 13. Overall validation strategy involving the LCFA. This figure extends beyond the definition/performance testing of the LCFA (see the Validation Plan for details).

4.1 Algorithm Definition Phase

The intent of the ADP is to construct tests and perform calculations that help clearly define the basic form and methodology to be employed by the LCFA. Tests and calculations are specifically devised to illustrate, substantiate, or justify the use of specific algorithm practices or approaches. The tests/calculations are geared toward answering specific questions, and are associated with clear metrics that are used to interpret the results. The official ADP tests are termed *Perform Testing & Validation of GLM Algorithm (80%*). However, much ADP pretesting is done before this time (see Test Schedule and Responsibilities section below).

The ADP tests involve three basic types: *speed*, *accuracy*, and *resiliency*. These basic test types are described below:

4.1.1 Speed Tests

The definition of a Speed Test is a test that tests how many events per second the LCFA can process, and determines if the LCFA can keep up with latency requirements. The data latency requirement is 209 sec, which includes 10 sec for Instrument Vendor processing of level 0 to level 1b, and 4 sec for the LCFA processing of level 1b to level 2. Therefore, for the LCFA processing chain, if the data rate becomes too high, lower priority code

sections (such as real-time filtering, or full fitting) will be suspended to allow for the production of the real time lightning product (under the assumption that a lower quality on-time product is better than a higher quality delayed product).

Calculations have been made to estimate the peak and sustained flash rates expected from the total GLM FOV. The calculations have determined that the Lockheed Martin estimates of 5280 events/sec (peak) and 3120 events/sec (sustained) are reasonable, but smaller than our initial peak estimates. Moreover, there is an instrument processing limit of 20,000 events/sec, so it is our understanding that this will be the absolute max input event rate to the LCFA.

4.1.2 Accuracy Tests

An accuracy test involves first constructing (by simulation or by using a data source) an event-level dataset. Next, one clusters the event-level dataset to construct a known flash-level dataset. Finally, the event-level dataset is input to the LCFA and the LCFA output is compared to the known flash-level dataset to assess LCFA clustering accuracy. Note that the input event-level dataset has no errors (effect of input errors is examined in resiliency tests). Hence, the Accuracy Test sees how well the LCFA retrieves known flash properties. Specifically, the flash properties of interest include:

- **Percent Flashes Detected**. The number of flashes computed by the LCFA will be compared with the number of flashes in the known flash-level dataset. The ratio of the LCFA-derived number of flashes to the number of known flashes, times one hundred, is the LCFA *percent flashes detected* (PFD), but only if the LCFA has not split known flashes. In general, the nuances associated with the possibility of splitting/merging flashes must be handled appropriately so as not to skew overall results. Note that the PFD is NOT the same as, nor an estimate of, the GLM on-orbit flash detection efficiency. Changes in the PFD will be noted as a function of increasing flash rate; multiple prototype storms will be used to increase flash rate across the GLM FOV.
- **Flash Location Error**. The flash optical centroid of each known flash will be directly computed and compared with the associated flash optical centroid computed by the LCFA. The distance between the two centroids defines the *flash location error* (FLE). The FLE distribution will be obtained, with the mean, standard deviation, and median errors cited. Changes in the FLE distribution will be noted as a function of increasing flash rate; multiple prototype storms will be used to increase flash rate across the GLM FOV.
- **Flash Radiance Error**. The radiance of each known flash will be directly computed and compared with the associated flash radiance computed by the LCFA. The difference between the two radiances defines the *flash radiance error* (FRE). The FRE distribution will be obtained, with the mean, standard deviation, and median errors cited. Changes in the FRE distribution will be noted as a function of increasing flash rate; multiple prototype storms will be used to increase flash rate across the GLM FOV.
- **Flash Area Error.** The area of each known flash will be directly computed and compared with the associated flash area computed by the LCFA. The difference between the two areas defines the *flash area error* (FAE). The FAE distribution

will be obtained, with the mean, standard deviation, and median errors cited. Changes in the FAE distribution will be noted as a function of increasing flash rate; multiple prototype storms will be used to increase flash rate across the GLM FOV.

- **Flash Duration Error**. The duration of each known flash will be directly computed and compared with the associated flash duration computed by the LCFA. The difference between the two durations defines the *flash duration error* (FDE). The FDE distribution will be obtained, with the mean, standard deviation, and median errors cited. Changes in the FDE distribution will be noted as a function of increasing flash rate; multiple prototype storms will be used to increase flash rate across the GLM FOV.

4.1.3 Resiliency Tests.

A resiliency test examines how well the LCFA can control/minimize the deleterious effects of input errors or other peculiar/unusual input conditions (e.g., event radiance noise or location errors, corrupted data packets, extremely long-duration flashes, events near the prime-meridian or international date-line, negative event times, event times that run backwards, latitudes/longitudes that are out-of-range). Peculiar/unusual conditions do not include high event rate input, since this is handled separately in the Speed Test. The Resiliency Tests should include statistical measures (PFD, FLE, FRE, FAE, FDE) mentioned above to help quantify the deleterious effects.

A Quality Assurance Tool (*QA Tool*, written by Mach and Bateman, and inspected by the remaining R3/AWG lightning team) will be used to help assess and minimize the effect of input errors or other peculiar/unusual input conditions on LCFA flash retrieval results.

4.1.4 Specific List of Tests

Below is a list of specific tests. Each test falls into one or more categories (S = Speed, A = Accuracy, R = Resiliency); the test category is indicated after the name of each test.

- **CPU Load Test (S).** This test uses 300 flashes per second and 5000 events per second. The NASA-LDAR legacy event-to-flash associator processed this data set at a rate of 2X real-time. The event file contains synthetic GLM events representing a two hour period of time created from the most active two hour period of time in the most active day observed by the NLDN network between the years 1995 and 2007 (13 years). The events were created from NLDN flashes using the following method: [Step 1] For each flash reported in the NLDN network (mostly ground flashes), other flashes were created, and offset in position and time. The flashes included (a) a Southern Hemisphere flash and (b) 10% of the time a "random flash" somewhere within the FOV of GLM. [Step 2] For each of the flashes produced in step 1 (i.e. a NLDN flash, a Southern Hemisphere flash and possibly a random flash), a variable number of groups were created around each flash. The number of groups per flash were chosen at random from the observed Cumulative Probability Distribution function for the expected number of GLM groups per GLM flash. [Step 3] For each of the groups produced in step 2, a variable number of events were created around each group. The number of events

per group is chosen at random from the observed Cumulative Probability Distribution function for the expected number of GLM events per GLM group.

- **CPU Load Test II (S).** This test starts at 0 events per second and peaks at about 12,000 events per second (sustained rate) at T=1hour and then goes into a "variable rate" mode where it simulates "bursty" data, varying between 1000 and 10,000 events per second on a 10 second time-scale. The NASA-LDAR legacy event-to-flash associator falls behind at a 6,000 events per second (sustained rate) but kept up at 12,000 events per second (variable rate).
- **CPU Load Test III (S).** This test starts at 0 events per second and peaks at about 20,000 events per second (sustained rate) at T=1 hour and then goes into a "variable rate" mode where it simulates "bursty" data, varying between 1000 and 20,000 events per second on a 10 second time-scale. The NASA-LDAR legacy event-to-flash associator falls behind at a 6,000 events per second (sustained rate).
- **Break Code Test using NLDN Data (S).** This test will feed total estimated flash rates from the 21 July 2003 NLDN storms to the LCFA. To boost overall flash rate, the storms should be approximately repeated over other regions of the GLM FOV. The times of the repeated storms should be shifted relative to each other, but contain some temporal overlap.
- **Break Code Test using Re-sampled LIS Data (S).** This test will be based on re-sampling LIS data to generate a GLM proxy. Due to the limited LIS view time, several storms observed by LIS will be required to generate a GLM proxy. Nonetheless, the re-sampled LIS data can be combined and repeated over several different regions within the GLM FOV. The times of the repeated data should be shifted relative to each other, but contain some temporal overlap. We will see how much of the re-sampled and repeated LIS data within the GLM FOV the LCFA can handle.
- **Break Code Test using VHF Data (S).** This test will be based on a very active storm of GLM proxy data derived from one VHF network such as the NALMA or the OKLMA. This active storm “prototype” will be placed in several different regions within the GLM FOV. The times of the repeated prototype storms should be shifted relative to each other, but contain some temporal overlap. We will see how many prototype storms within the GLM FOV the LCFA can handle.
- **Lag Mitigation Tests (S).** If one or more of the “break code tests” above indicate that the latency requirement is likely to be violated under expected sustained and/or extreme lightning flash rate conditions, then several mitigation approaches (regionalization, full-to-first fit conversion, …) must be run to assess the quality of each mitigation approach. The highest quality mitigation approach must be identified and substantiated. The lag mitigation approaches should be evaluated based on the following metrics: accuracy, speed, ease of implementation, and robustness.
- **Simulated Stress Test (S).** This is a test where we cover a large percentage of the disk with long-lived continuously illuminated flash.
- **Simulated Distance/Time Thresholding Test (S).** This test involves normal distributions with mean group separations of 16.5 km and 330 ms.
- **Moving Lines of Storms (S, A).** Data source is NLDN data. These are long lines of storms that persist for many hours and move across the field of view.

- **Simulated High Flash Rate Storms (S, A).** These are simulated storms that make lightning at (and beyond) the peak rates we are able to handle.
- **Supercell Storm with Much Lightning (S, A).** Data source is VHF data.
- **Medium Storm with Average Lightning (A).** Data source is VHF and LIS data.
- **Small Storm with Little Lightning (A).** Data source is VHF and LIS data.
- **Storm Merging/Splitting (A).** Data source is VHF data.
- **Storm Splitting (A).** Data source is VHF data.
- **Storms that Cycle (A).** Data source is VHF data.
- **Simulated Oddly-Moving Storms (A).**
- **Simulated Regrouping Test (A).** This test involves proxy data where flashes start as separate entities but grow together. In the GLM algorithm, all flashes that overlap will be combined into a single flash.
- **OK Storm with Very Large/Long Flashes (R).** Data source is VHF data.
- **Simulated Noise-Only at Various Densities (R).**
- **Simulated Checkerboard Test (R).** This test involves various checkerboard patterns in space and time to see if LCFA distance criteria perform as expected.
- **Corrupted Data Packets & Noise Test (R).** The QA Tool will generate simulated corrupted data packet and noise errors indicated by "e_1" in Figure 13. These random errors will be added to the same known event-level dataset used for Accuracy Tests. The input errors, e_1, will corrupt to a certain extent the simulated Level 2 products produced by the LCFA (and shown in the lower right corner of Figure 13). By comparing LCFA output for $e_1 = 0$ and for $e_1 \neq 0$, we will be able to assess the overall effects of corrupted packets and noise on the five flash variables (detection, location, radiance, area, duration) discussed above. In other words, we will determine how PFD, FLE, FRE, FAE, and FDE change due to nonzero e_1.
- **QA-Procedure (R).** Since the errors, e_1, can produce errors in the flash retrievals, we will want to define an LCFA QA-procedure that flags and possibly removes the effects of e_1. Hence, to test/assess the LCFA's QA-procedure for handling corrupted data packets and noise, we will run the QA-procedure on two virtual Level 1b data streams, one stream free of corrupted data packets and noise, and one with (randomly added) corrupted data packets and noise. A comparison of the QA-procedure output from these two streams will allow us to identify and remove deficiencies in the QA-procedure. The error results will be checked manually to assure that the QA-procedure is accurate.
- **Longitude Test (R).** This test will verify that the LCFA can handle a ±180 degree or 0/360 degree longitude numbering system, as well as events locations that cross either the International Dateline or the Prime Meridian.

4.2 Algorithm Refinement Phase

The results of CDR can possibly motivate additional tests/calculations that should be performed to help refine the LCFA. Moreover, the AWG lightning team will naturally want to continuously refine the LCFA. Some of the ARP tests will look similar to the ADP tests but would be more robust and precise in order to fully optimize the LCFA.

5. RESULTS

5.1 CPU Load Test (S)

This test uses 300 flashes per second and 5000 events per second. The NASA-LDAR legacy event-to-flash associator processed this data set at a rate of 2X real-time.

- Notice File Output:

```
2009/04/07 10:00:54: Open_NOTICE_FILE: GLM_Process_Lightning program version: 3.0,
summary: Takes geo-located lightning events and clusters them
2009/04/07 10:00:54: Open_NOTICE_FILE: Started: 2009-04-07
2009/04/07 10:00:54: Read_User: Will use 16.500000 as the FLASH_DELTA_KM and 0.330000 as
the FLASH_DELTA_TIME
2009/04/07 10:00:54: Read_User: Will use 14.000000 as the GROUP_DELTA_KM and 0.000000 as
the Group_Delta_time
2009/04/07 10:00:54: Read_User: WILL limit the size of items
2009/04/07 10:00:54: Read_User: Will use 0.010000 as the PIXEL_SPACING, 0.050000 as the
NOISE_REJECT_LEVEL, and WILL do Noise_Reject
2009/04/07 10:00:54: Read_User: Events input:
I:/GLM/Proxy_data/CPU_Load_Test_FakeEvents.txt
2009/04/07 10:00:54: Read_User: Output binary clustered data:
I:/GLM/output_R3/glm_fge_CLTFE.txt
2009/04/07 10:00:54: Read_User: Output text flashes:
I:/GLM/output_R3/flashes_out_CLTFE.txt
2009/04/07 10:00:54: Read_User: Output text groups: I:/GLM/output_R3/groups_out_CLTFE.txt
2009/04/07 10:00:54: Read_User: Output text events: I:/GLM/output_R3/events_out_CLTFE.txt
2009/04/07 10:00:54: Read_User: Event stats: I:/GLM/output_R3/events_i_out_CLTFE.txt
2009/04/07 10:00:54: Read_User: Group stats: I:/GLM/output_R3/groups_i_out_CLTFE.txt
2009/04/07 10:00:54: Read_User: Flash stats: I:/GLM/output_R3/flashes_i_out_CLTFE.txt
2009/04/07 10:00:54: Read_User: Processing rate stats:
I:/GLM/output_R3/processing_rate_CLTFE.txt
2009/04/07 10:00:54: main: Will not output ASCII E/G/F data (for speed test)
2009/04/07 10:00:54: Create_New_Group: Increasing group buffer to 500
2009/04/07 10:00:54: Create_New_Group: Increasing group index to 500
2009/04/07 10:00:54: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 10:00:55: Create_New_Group: Increasing group buffer to 1000
2009/04/07 10:00:55: Create_New_Group: Increasing group index to 1000
2009/04/07 10:00:55: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 10:00:59: Create_New_Group: Increasing group buffer to 2000
2009/04/07 10:00:59: Create_New_Group: Increasing group index to 2000
2009/04/07 10:00:59: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 10:01:08: Create_New_Flash: Increasing Item buffer to 400
2009/04/07 10:01:08: Create_New_Flash: Increasing Item index to 400
2009/04/07 10:01:48: Create_New_Group: Increasing group buffer to 4000
2009/04/07 10:01:48: Create_New_Group: Increasing group index to 4000
2009/04/07 10:01:48: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 10:04:07: main: WF:770206, WG:5996164, WE:17977528
2009/04/07 10:04:07: main: Delta time:193
2009/04/07 10:04:07: Terminate: Done with processing GLM data
```

The code produced 770206 flashes and 5996164 groups out of the 17977528 events in the proxy file. It took 193 seconds to process the data, resulting in a 93 kevents/sec processing rate. The default group buffer/index size of 250 was too small. The code automatically increased the buffer/index size to 4000. The Item buffer/index default size of 200 was too small. The code automatically increased the size to 400. The event rate as a function of data time is shown in Figure 14 and the processing speed of the GLM LCFC is shown in Figure 15.

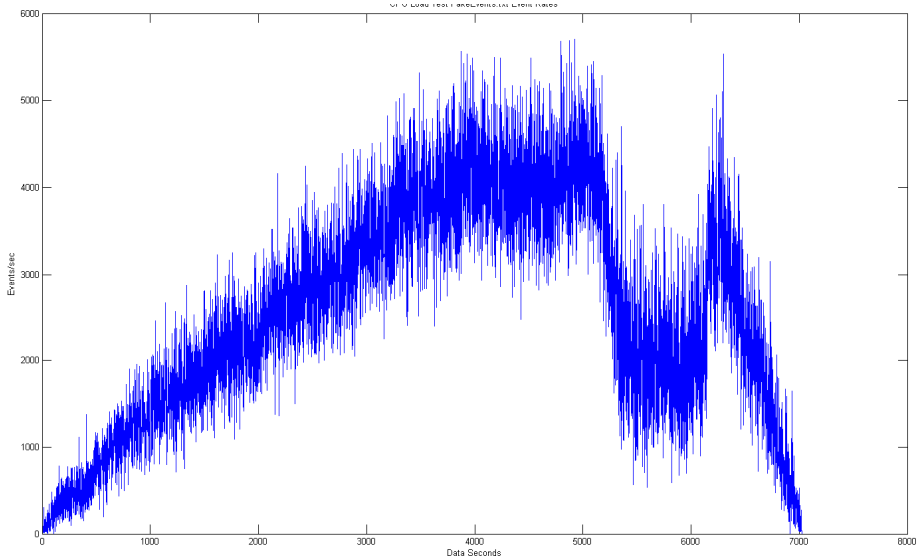

Figure 14. Event Rates

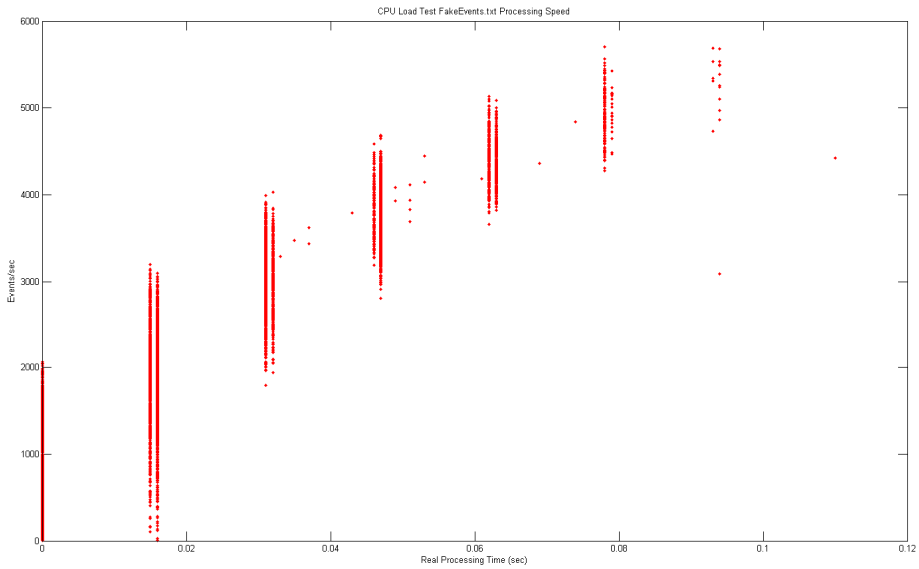

Figure 15. Processing speed of the GLM code

The GLM code is able to process up to 6000 events in less than 0.12 seconds of real-time.

5.2 CPU Load Test II (S)

This test starts at 0 events per second and peaks at about 12,000 events per second (sustained rate) at T=1 hour and then goes into a "variable rate" mode where it simulates "bursty" data, varying between 1000 and 10,000 events per second on a 10 second time-scale. The NASA-LDAR legacy event-to-flash associator falls behind at a 6,000 events per second (sustained rate) but kept up at 12,000 events per second (variable rate).

- Notice File Output:

2009/04/07 10:45:20: Open_NOTICE_FILE: GLM_Process_Lightning program version: 3.0, summary: Takes geo-located lightning events and clusters them
2009/04/07 10:45:20: Open_NOTICE_FILE: Started: 2009-04-07
2009/04/07 10:45:20: Read_User: Will use 16.500000 as the FLASH_DELTA_KM and 0.330000 as the FLASH_DELTA_TIME

```
2009/04/07 10:45:20: Read_User: Will use 14.000000 as the GROUP_DELTA_KM and 0.000000 as the Group_Delta_time
2009/04/07 10:45:20: Read_User: WILL limit the size of items
2009/04/07 10:45:20: Read_User: Will use 0.010000 as the PIXEL_SPACING, 0.050000 as the NOISE_REJECT_LEVEL, and
WILL do Noise_Reject
2009/04/07 10:45:20: Read_User: Events input: I:/GLM/Proxy_data/CPU_Load_2.FakeEvents.txt
2009/04/07 10:45:20: Read_User: Output binary clustered data: I:/GLM/output_R3/glm_fge_CLTFE2.txt
2009/04/07 10:45:20: Read_User: Output text flashes: I:/GLM/output_R3/flashes_out_CLTFE2.txt
2009/04/07 10:45:20: Read_User: Output text groups: I:/GLM/output_R3/groups_out_CLTFE2.txt
2009/04/07 10:45:20: Read_User: Output text events: I:/GLM/output_R3/events_out_CLTFE2.txt
2009/04/07 10:45:20: Read_User: Event stats: I:/GLM/output_R3/events_i_out_CLTFE2.txt
2009/04/07 10:45:20: Read_User: Group stats: I:/GLM/output_R3/groups_i_out_CLTFE2.txt
2009/04/07 10:45:20: Read_User: Flash stats: I:/GLM/output_R3/flashes_i_out_CLTFE2.txt
2009/04/07 10:45:20: Read_User: Processing rate stats: I:/GLM/output_R3/processing_rate_CLTFE2.txt
2009/04/07 10:45:20: main: Will not output ASCII E/G/F data (for speed test)
2009/04/07 10:45:20: Create_New_Group: Increasing group buffer to 500
2009/04/07 10:45:20: Create_New_Group: Increasing group index to 500
2009/04/07 10:45:20: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 10:45:20: Create_New_Group: Increasing group buffer to 1000
2009/04/07 10:45:20: Create_New_Group: Increasing group index to 1000
2009/04/07 10:45:20: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 10:45:23: Create_New_Group: Increasing group buffer to 2000
2009/04/07 10:45:23: Create_New_Group: Increasing group index to 2000
2009/04/07 10:45:23: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 10:45:26: Create_New_Flash: Increasing Item buffer to 400
2009/04/07 10:45:26: Create_New_Flash: Increasing Item index to 400
2009/04/07 10:45:38: Create_New_Group: Increasing group buffer to 4000
2009/04/07 10:45:38: Create_New_Group: Increasing group index to 4000
2009/04/07 10:45:38: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 10:46:15: Create_New_Flash: Increasing Item buffer to 800
2009/04/07 10:46:15: Create_New_Flash: Increasing Item index to 800
2009/04/07 10:47:23: Create_New_Group: Increasing group buffer to 8000
2009/04/07 10:47:23: Create_New_Group: Increasing group index to 8000
2009/04/07 10:47:23: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 11:01:40: main: WF:1817234, WG:14542838, WE:43648565
2009/04/07 11:01:40: main: Delta time:980
2009/04/07 11:01:40: Terminate: Done with processing GLM data
```

The code produced 1,817,234 flashes and 14,542,838 groups out of the 43,648,565 events in the proxy file. It took 980 seconds to process the data, resulting in a 44 kevents/sec processing rate. The default group buffer/index size of 250 was too small. The code automatically increased the buffer/index size to 8000. The Item buffer/index default size of 200 was too small. The code automatically increased the size to 800. The event rate as a function of data time is shown in Figure 16 while the response of the GLM LCFA is shown in Figure 17.

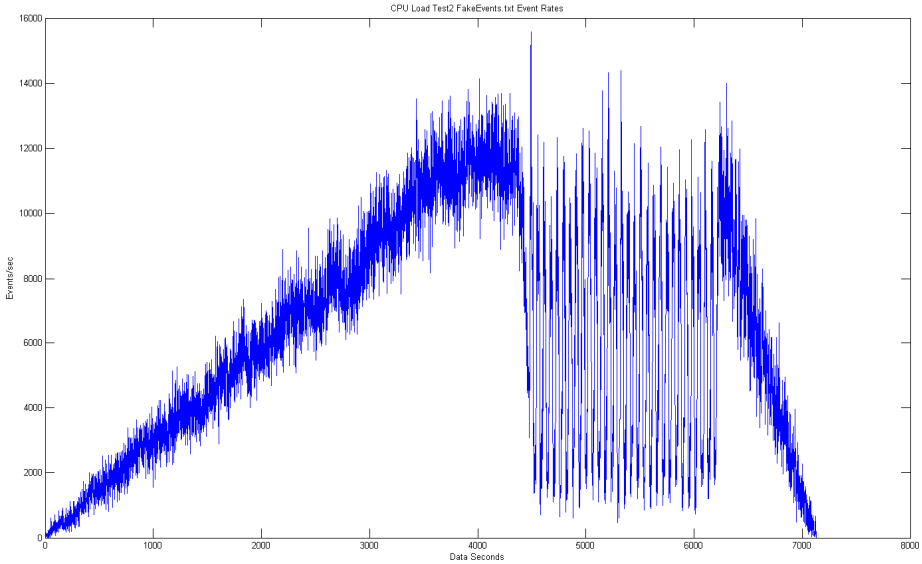

Figure 16. Event Rates.

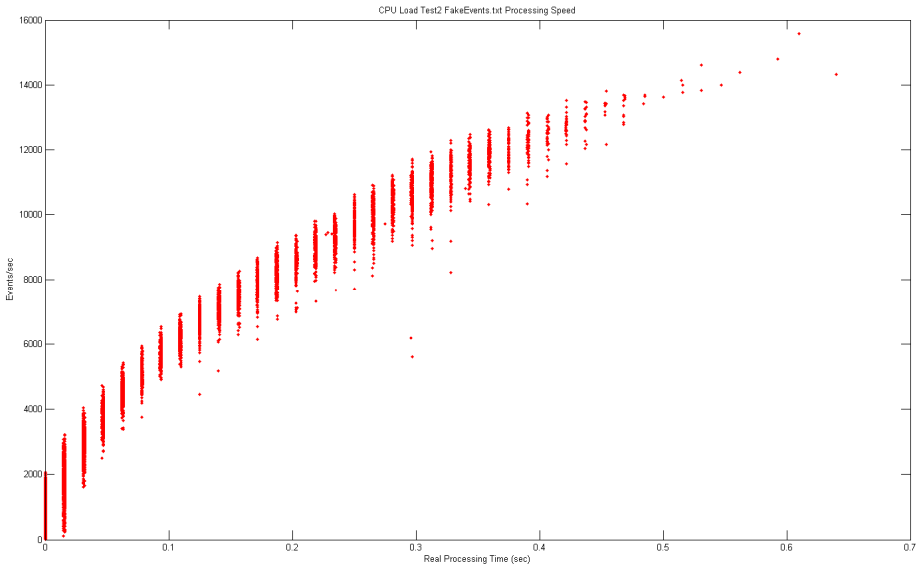

Figure 17. Processing speed of the GLM code

The GLM code is able to process up to 16,000 events in less than 0.7 seconds of real-time.

5.3 CPU Load Test III (S)

This test starts at 0 events per second and peaks at about 40,000 events per second (sustained rate) at T=1hour and then goes into a "variable rate" mode where it simulates "bursty" data, varying between 1000 and 20,000 events per second on a 10 second time-scale. The NASA-LDAR legacy event-to-flash associator falls behind at a 6,000 events per second (sustained rate).

- Notice File Output:

2009/04/07 14:13:38: Open_NOTICE_FILE: GLM_Process_Lightning program version: 3.0, summary: Takes geo-located lightning events and clusters them

2009/04/07 14:13:38: Open_NOTICE_FILE: Started: 2009-04-07
2009/04/07 14:13:38: Read_User: Will use 16.500000 as the FLASH_DELTA_KM and 0.330000 as the FLASH_DELTA_TIME
2009/04/07 14:13:38: Read_User: Will use 14.000000 as the GROUP_DELTA_KM and 0.000000 as the Group_Delta_time
2009/04/07 14:13:38: Read_User: WILL limit the size of items
2009/04/07 14:13:38: Read_User: Will use 0.010000 as the PIXEL_SPACING, 0.050000 as the NOISE_REJECT_LEVEL, and WILL do Noise_Reject
2009/04/07 14:13:38: Read_User: Events input: I:/GLM/Proxy_data/FakeEvents_20K.txt
2009/04/07 14:13:38: Read_User: Output binary clustered data: I:/GLM/output_R3/glm_fge_20k.txt
2009/04/07 14:13:38: Read_User: Output text flashes: I:/GLM/output_R3/flashes_out_20k.txt
2009/04/07 14:13:38: Read_User: Output text groups: I:/GLM/output_R3/groups_out_20k.txt
2009/04/07 14:13:38: Read_User: Output text events: I:/GLM/output_R3/events_out_20k.txt
2009/04/07 14:13:38: Read_User: Event stats: I:/GLM/output_R3/events_i_out_20k.txt
2009/04/07 14:13:38: Read_User: Group stats: I:/GLM/output_R3/groups_i_out_20k.txt
2009/04/07 14:13:38: Read_User: Flash stats: I:/GLM/output_R3/flashes_i_out_20k.txt
2009/04/07 14:13:38: Read_User: Processing rate stats: I:/GLM/output_R3/processing_rate_20k.txt
2009/04/07 14:13:38: main: Will not output ASCII E/G/F data (for speed test)
2009/04/07 14:13:38: Create_New_Group: Increasing group buffer to 500
2009/04/07 14:13:38: Create_New_Group: Increasing group index to 500
2009/04/07 14:13:38: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 14:13:38: Create_New_Group: Increasing group buffer to 1000
2009/04/07 14:13:38: Create_New_Group: Increasing group index to 1000
2009/04/07 14:13:38: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 14:13:40: Create_New_Group: Increasing group buffer to 2000
2009/04/07 14:13:40: Create_New_Group: Increasing group index to 2000
2009/04/07 14:13:40: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 14:13:49: Create_New_Group: Increasing group buffer to 4000
2009/04/07 14:13:49: Create_New_Group: Increasing group index to 4000
2009/04/07 14:13:49: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 14:13:52: Create_New_Flash: Increasing Item buffer to 400
2009/04/07 14:13:52: Create_New_Flash: Increasing Item index to 400
2009/04/07 14:15:14: Create_New_Group: Increasing group buffer to 8000
2009/04/07 14:15:14: Create_New_Group: Increasing group index to 8000
2009/04/07 14:15:14: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 14:15:34: Create_New_Flash: Increasing Item buffer to 800
2009/04/07 14:15:34: Create_New_Flash: Increasing Item index to 800
2009/04/07 14:27:17: Create_New_Group: Increasing group buffer to 16000
2009/04/07 14:27:17: Create_New_Group: Increasing group index to 16000
2009/04/07 14:27:17: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 14:42:33: Create_New_Flash: Increasing Item buffer to 1600
2009/04/07 14:42:33: Create_New_Flash: Increasing Item index to 1600
2009/04/07 15:09:10: Copy_Event: Increasing Nominal Events to 180000 (and clearing it)
2009/04/07 16:09:54: Create_New_Group: Increasing group buffer to 32000
2009/04/07 16:09:54: Create_New_Group: Increasing group index to 32000
2009/04/07 16:09:54: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/07 20:48:59: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 20:54:38: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 20:54:38: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 20:54:57: Double_Child_Buffer: Increasing item child count buffer to 3200
2009/04/07 21:05:53: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 21:07:30: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 21:21:22: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 21:22:50: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 21:28:39: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 21:29:50: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 21:44:28: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 21:44:32: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 21:44:32: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 22:05:26: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 22:07:55: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 22:10:55: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 22:15:46: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 22:20:33: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 22:20:41: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 22:20:49: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 22:21:25: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 22:24:27: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 22:26:31: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 22:29:19: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 22:50:53: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 22:51:32: Double_Child_Buffer: Increasing item child count buffer to 1600

```
2009/04/07 22:55:44: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 23:01:49: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 23:03:16: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 23:18:33: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 23:18:40: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 23:18:40: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 23:19:13: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 23:20:41: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 23:29:08: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 23:53:51: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/07 23:54:26: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:08:30: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:15:05: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:17:26: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:19:13: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:19:55: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:24:09: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:25:21: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:28:00: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:31:08: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:35:01: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:38:42: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:39:17: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:39:19: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:39:19: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:40:53: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:42:26: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:46:56: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:47:22: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:47:22: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:47:38: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:51:02: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 00:57:48: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 01:01:23: main: WF:3648249, WG:48333625, WE:321270562
2009/04/08 01:01:23: main: Delta time:38865
2009/04/08 01:01:23: Terminate: Done with processing GLM data
```

The code produced 3,648,249 flashes and 48,333,625 groups out of the 321,270,562 events in the proxy file. It took 38865 seconds to process the data, resulting in an average 8.2 kevents/sec processing rate. The default group buffer/index size of 250 was too small. The code automatically increased the buffer/index size to 1600. The Item buffer/index default size of 200 was too small. The code automatically increased the size to 1600. Many flashes/groups needed their child buffer increased from 800 to 1600. The code needed to increase the nominal event count from 90,000 to 180,000. The event rate as a function of data time and GLM LCFA response are shown in Figures 18 and 19.

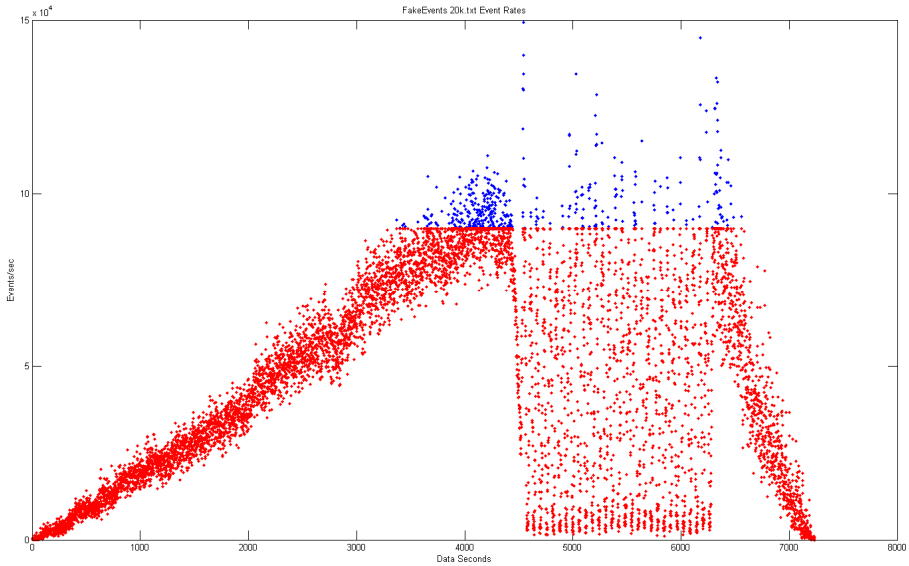

Figure 18. Data rate as a function of data time.

Note that the current limit for the event statistics to produce this plot is 90,000 events. The blue data above that number is an estimate of the true number of events/sec in the proxy dataset.

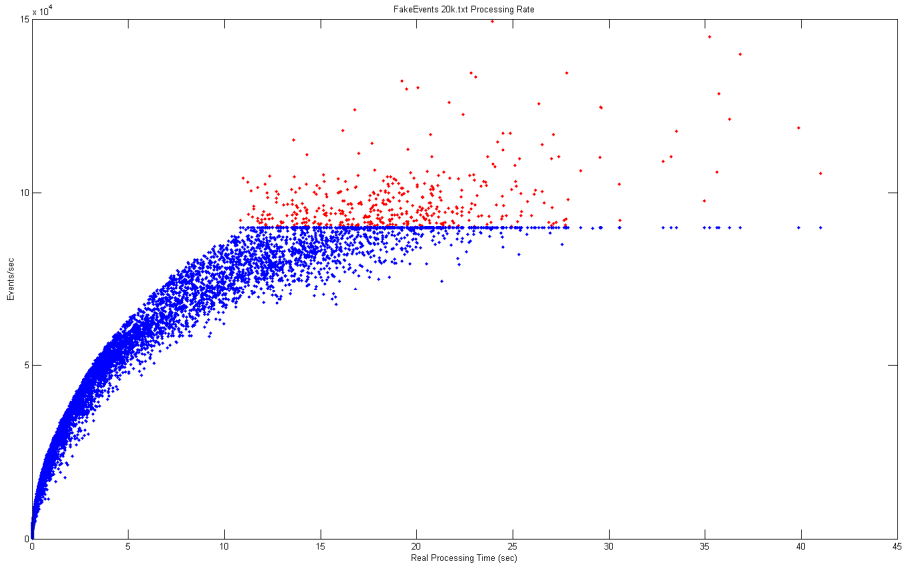

Figure 19. Processing speed of the GLM code.

Note that the current limit for the event statistics to produce this plot is 90,000 events. The red data above that number is an estimate of the true number of events/sec in the proxy dataset.

5.4 Longitude Test I (International Dateline) (R)

This test will verify that the LCFA can handle a ±180 degree or 0/360 degree longitude numbering system, as well as events locations that cross either the International Dateline or the Prime Meridian.

- Notice File Output:

```
2008/12/12 14:10:47: open_notice_file: GLM_Process_Lightning program version: 2.0, summary: takes geo-located lightning
2008/12/12 14:10:47: open_notice_file: Started: 2008-12-12
2008/12/12 14:10:47: read_user: Will use 16.500000 as the Flash_Delta_Km and 0.330000 as the Flash_Delta_Time
2008/12/12 14:10:47: read_user: Will use 14.000000 as the Group_Delta_Km and 0.000000 as the Group_Delta_time
2008/12/12 14:10:47: read_user: WILL limit the size of items
2008/12/12 14:10:47: read_user: Will use 0.010000 as the Pixel_Spacing, 0.050000 as the Noise_Reject_Level, and WILL do
Noise_Reject
2008/12/12 14:10:47: read_user: Events input: I:/GLM/Proxy_data/GLM_test_InternationalDateline.FakeEvents.txt
2008/12/12 14:10:47: read_user: Output binary clustered data: I:/GLM/output_R2/glm_fge_IDFE.txt
2008/12/12 14:10:47: read_user: Output text flashes: I:/GLM/output_R2/flashes_out_IDFE.txt
2008/12/12 14:10:47: read_user: Output text groups: I:/GLM/output_R2/groups_out_IDFE.txt
2008/12/12 14:10:47: read_user: Output text events: I:/GLM/output_R2/events_out_IDFE.txt
2008/12/12 14:10:47: read_user: Event stats: I:/GLM/output_R2/events_i_out_IDFE.txt
2008/12/12 14:10:47: read_user: Group stats: I:/GLM/output_R2/groups_i_out_IDFE.txt
2008/12/12 14:10:47: read_user: Flash stats: I:/GLM/output_R2/flashes_i_out_IDFE.txt
2008/12/12 14:10:47: read_user: Processing rate stats: I:/GLM/output_R2/processing_rate_IDFE.txt
2008/12/12 14:10:47: main: Will output ASCII E/G/F data
2008/12/12 14:10:47: create_new_group: Increasing group buffer to 500
2008/12/12 14:10:47: create_new_group: Increasing group index to 500
2008/12/12 14:10:47: create_new_group: Initializing (clearing) the new group buffer
2008/12/12 14:10:47: create_new_group: Increasing group buffer to 1000
2008/12/12 14:10:47: create_new_group: Increasing group index to 1000
2008/12/12 14:10:47: create_new_group: Initializing (clearing) the new group buffer
2008/12/12 14:10:47: double_child_buffer: Increasing item child count buffer to 1600
2008/12/12 14:10:47: create_new_group: Increasing group buffer to 2000
2008/12/12 14:10:47: create_new_group: Increasing group index to 2000
2008/12/12 14:10:47: create_new_group: Initializing (clearing) the new group buffer
2008/12/12 14:10:47: main: WF:2, WG:1001, WE:1001
2008/12/12 14:10:47: main: Delta time:0
2008/12/12 14:10:47: terminate: Done with processing GLM data
```

The test dataset was to produce 2 flashes. The first would have 1000 groups and events clustered around the International Date Line. The GLM code reproduced that flash (D Flash #0 Start Time:0.001 End Time:1.000 Centroid Lat:0.000 Centroid Lon:-180.000 Child_Count:1000). Figure 20 below shows the location of all the groups in the one flash (centered around the International Dateline).

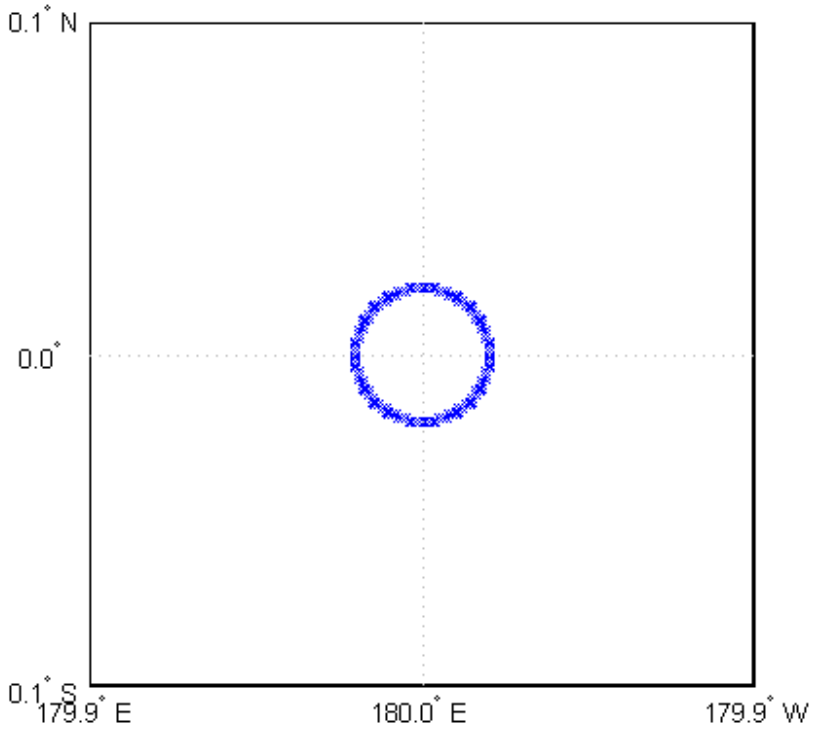

Figure 20. Location of all groups in the flash (centered around the International Dateline).

5.5 Longitude Test II (Prime Meridian) (R)

This test will verify that the LCFA can handle a ±180 degree or 0/360 degree longitude numbering system, as well as events locations that cross either the International Dateline or the Prime Meridian.

- Notice File Output:

2008/12/12 18:18:09: open_notice_file: GLM_Process_Lightning program version: 2.0, summary: takes geo-located lightning
2008/12/12 18:18:09: open_notice_file: Started: 2008-12-12
2008/12/12 18:18:09: read_user: Will use 16.500000 as the Flash_Delta_Km and 0.330000 as the Flash_Delta_Time
2008/12/12 18:18:09: read_user: Will use 14.000000 as the Group_Delta_Km and 0.000000 as the Group_Delta_time
2008/12/12 18:18:09: read_user: WILL limit the size of items
2008/12/12 18:18:09: read_user: Will use 0.010000 as the Pixel_Spacing, 0.050000 as the Noise_Reject_Level, and WILL do Noise_Reject
2008/12/12 18:18:09: read_user: Events input: I:/GLM/Proxy_data/GLM_test_PrimeMeridian.FakeEvents.txt
2008/12/12 18:18:09: read_user: Output binary clustered data: I:/GLM/output_R2/glm_fge_PMFE.txt
2008/12/12 18:18:09: read_user: Output text flashes: I:/GLM/output_R2/flashes_out_PMFE.txt
2008/12/12 18:18:09: read_user: Output text groups: I:/GLM/output_R2/groups_out_PMFE.txt
2008/12/12 18:18:09: read_user: Output text events: I:/GLM/output_R2/events_out_PMFE.txt
2008/12/12 18:18:09: read_user: Event stats: I:/GLM/output_R2/events_i_out_PMFE.txt
2008/12/12 18:18:09: read_user: Group stats: I:/GLM/output_R2/groups_i_out_PMFE.txt
2008/12/12 18:18:09: read_user: Flash stats: I:/GLM/output_R2/flashes_i_out_PMFE.txt
2008/12/12 18:18:09: read_user: Processing rate stats: I:/GLM/output_R2/processing_rate_PMFE.txt
2008/12/12 18:18:10: main: Will output ASCII E/G/F data
2008/12/12 18:18:10: create_new_group: Increasing group buffer to 500
2008/12/12 18:18:10: create_new_group: Increasing group index to 500
2008/12/12 18:18:10: create_new_group: Initializing (clearing) the new group buffer
2008/12/12 18:18:10: create_new_group: Increasing group buffer to 1000
2008/12/12 18:18:10: create_new_group: Increasing group index to 1000
2008/12/12 18:18:10: create_new_group: Initializing (clearing) the new group buffer
2008/12/12 18:18:10: double_child_buffer: Increasing item child count buffer to 1600
2008/12/12 18:18:10: create_new_group: Increasing group buffer to 2000
2008/12/12 18:18:10: create_new_group: Increasing group index to 2000

2008/12/12 18:18:10: create_new_group: Initializing (clearing) the new group buffer
2008/12/12 18:18:10: main: WF:2, WG:1001, WE:1001
2008/12/12 18:18:10: main: Delta time:0
2008/12/12 18:18:10: terminate: Done with processing GLM data

The test dataset was to produce 2 flashes. The first would have 1000 groups and events clustered around the Prime Meridian. The GLM code reproduced that flash (D Flash #0 Start Time:0.001 End Time:1.000 Centroid Lat:0.000 Centroid Lon:-0.000 Child_Count:1000). Figure 21 shows the location of all the groups in the one flash (centered around the Prime Meridian).

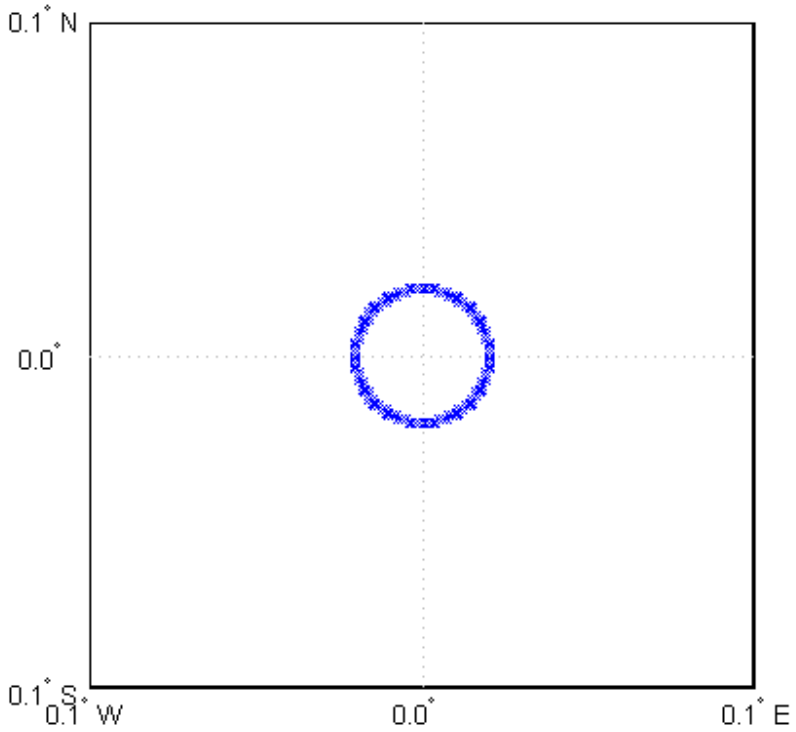

Figure 21. Location of all groups in flash (centered around Prime Meridian).

5.6 Simulated Regrouping Test (GLM_test_MergingFlash.txt) (A)

This test involves proxy data where flashes start as separate entities but grow together. In the GLM algorithm, all flashes that overlap will be combined into a single flash.

- Notice File:

2009/04/08 10:53:59: Open_NOTICE_FILE: GLM_Process_Lightning program version: 3.0, summary: Takes geo-located lightning events and clusters them
2009/04/08 10:53:59: Open_NOTICE_FILE: Started: 2009-04-08
2009/04/08 10:53:59: Read_User: Will use 16.500000 as the FLASH_DELTA_KM and 0.330000 as the FLASH_DELTA_TIME
2009/04/08 10:53:59: Read_User: Will use 14.000000 as the GROUP_DELTA_KM and 0.000000 as the Group_Delta_time
2009/04/08 10:53:59: Read_User: WILL limit the size of items
2009/04/08 10:53:59: Read_User: Will use 0.010000 as the PIXEL_SPACING, 0.050000 as the NOISE_REJECT_LEVEL, and WILL do Noise_Reject
2009/04/08 10:53:59: Read_User: Events input: I:/GLM/Proxy_data/GLM_test_MergingFlash_mod.txt
2009/04/08 10:53:59: Read_User: Output binary clustered data: I:/GLM/output_R3/glm_fge_GtMFM.txt
2009/04/08 10:53:59: Read_User: Output text flashes: I:/GLM/output_R3/flashes_out_GtMFM.txt
2009/04/08 10:53:59: Read_User: Output text groups: I:/GLM/output_R3/groups_out_GtMFM.txt
2009/04/08 10:53:59: Read_User: Output text events: I:/GLM/output_R3/events_out_GtMFM.txt
2009/04/08 10:53:59: Read_User: Event stats: I:/GLM/output_R3/events_i_out_GtMFM.txt
2009/04/08 10:53:59: Read_User: Group stats: I:/GLM/output_R3/groups_i_out_GtMFM.txt
2009/04/08 10:53:59: Read_User: Flash stats: I:/GLM/output_R3/flashes_i_out_GtMFM.txt

```
2009/04/08 10:53:59: Read_User: Processing rate stats: I:/GLM/output_R3/processing_rate_GtMFM.txt
2009/04/08 10:53:59: main: Will output ASCII E/G/F data
2009/04/08 10:53:59: Create_New_Group: Increasing group buffer to 500
2009/04/08 10:53:59: Create_New_Group: Increasing group index to 500
2009/04/08 10:53:59: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/08 10:53:59: Create_New_Group: Increasing group buffer to 1000
2009/04/08 10:53:59: Create_New_Group: Increasing group index to 1000
2009/04/08 10:53:59: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/08 10:53:59: Double_Child_Buffer: Increasing item child count buffer to 1600
2009/04/08 10:53:59: Create_New_Group: Increasing group buffer to 2000
2009/04/08 10:53:59: Create_New_Group: Increasing group index to 2000
2009/04/08 10:53:59: Create_New_Group: Initializing (clearing) the new group buffer
2009/04/08 10:53:59: Double_Child_Buffer: Increasing item child count buffer to 3200
2009/04/08 10:54:01: main: WF:2, WG:1932, WE:2001
2009/04/08 10:54:01: main: Delta time:2
2009/04/08 10:54:01: Terminate: Done with processing GLM data
```

The test dataset was to produce 2 flashes. The first would have 2000 events clustered around 45° Latitude and 45°Longitude. The flash started in 4 different locations (a cross around the centroid) and the GLM code should have combined the 4 initial flashes when they overlapped in the middle. The GLM code reproduced that flash (D Flash #0 Start Time: 0.001 End Time: 1.000 Centroid Lat: 45.001 Centroid Lon: 45.001 Child_Count:1931). Figure 22 shows the location of all the groups in the one flash (note that some of the groups consist of more than one event).

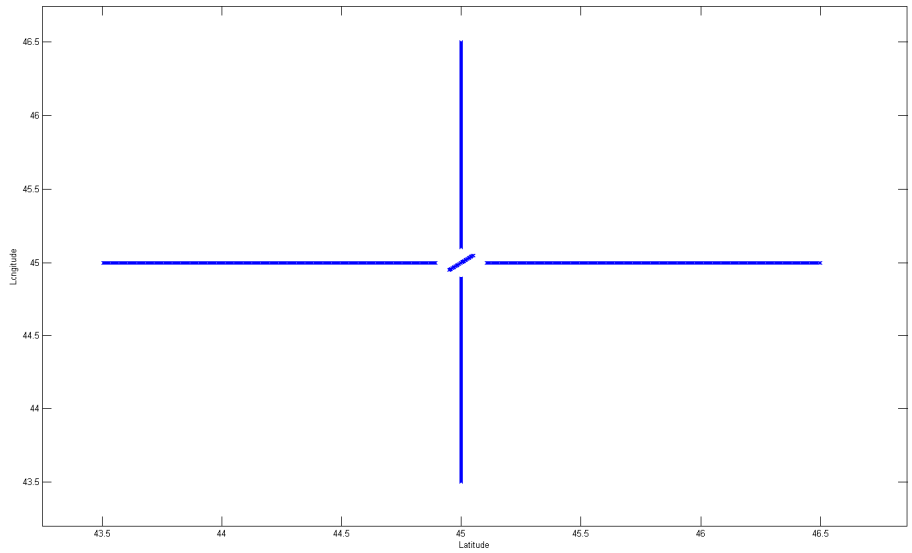

Figure 22. Location of all the groups in the one flash.

Appendix 2: A Physical Understanding of the Event, Group, and Flash "Radiance" Data Products

Overview: The GLM ATBD reviews asked how we would calculate group "radiance" and flash "radiance". Presently, the Lightning Cluster Filter Algorithm (LCFA) follows what was done for OTD/LIS, which was only approximative. This writing attempts to quantify what those old OTD/LIS "radiance" data products are, in physical terms. Consider a lightning imager (e.g., OTD, LIS, or future GLM) detecting a flash that is fully within its field of view, and neglect any instrument measurement errors/noise, threshold settings, and false alarms. Here, we are just interested in the illumination of the pixel array by a legitimate flash. Under these circumstances, one can define several solid-angle-averaged spectral energy densities, and then compare these (physically understandable) variables with the previous OTD/LIS "radiance" data products. This sheds light on the physical meaning of these previous data products. So this begins with the energy density descriptions. The figure below helps define basic quantities; the flash subtends a solid angle $\Delta\Omega$ and occupies *m* (two millisecond) frames.

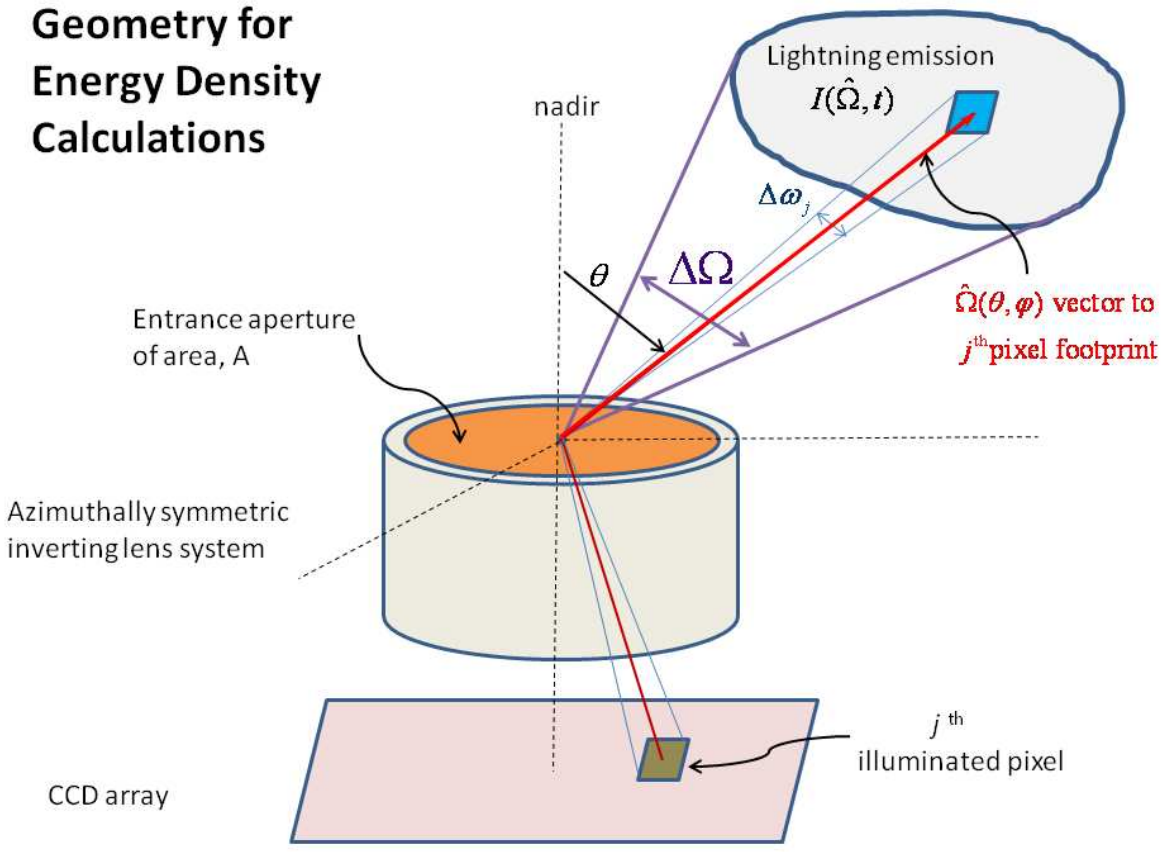

Figure 23. Geometry for energy density calculations.

Event Solid-Angle-Average Spectral Energy Density: The solid-angle-average spectral energy density (units of $\mu J/m^2/sr/\mu m$) collected by the j^{th} illuminated pixel in the i^{th} frame (which defines the j^{th} event in the i^{th} frame) is

$$\bar{\xi}_{\lambda j}(i) \equiv \frac{1}{\Delta\omega_j}\int_{\Delta\omega_j}\xi_\lambda(\hat{\Omega},i)d\Omega = \frac{1}{\Delta\omega_j}\int_{\Delta\omega_j}\int_{t_{i-1}}^{t_i} I_\lambda(\hat{\Omega},t)dtd\Omega \ . \qquad \textbf{(A1)}$$

Group Solid-Angle-Average Spectral Energy Density: The solid-angle-average energy density (units of $\mu J/m^2/sr/\mu m$) collected by a set of n_{pi} illuminated pixels in the i^{th} frame (which defines the p^{th} group in the i^{th} frame) is

$$\bar{\zeta}_{\lambda p}(i) \equiv \frac{1}{\Delta\Omega_{pi}}\int_{\Delta\Omega_{pi}}\xi_\lambda(\hat{\Omega},i)d\Omega = \frac{1}{\Delta\Omega_{pi}}\int_{\Delta\Omega_{pi}}\int_{t_{i-1}}^{t_i} I_\lambda(\hat{\Omega},t)dtd\Omega = \frac{1}{\Delta\Omega_{pi}}\sum_{s=1}^{n_{pi}}\Delta\omega_{pis}\bar{\xi}_{\lambda s}(i) \ . \qquad \textbf{(A2)}$$

Flash Solid-Angle-Average Spectral Energy Density: The solid-angle-average energy density (units of $\mu J/m^2/sr/\mu m$) collected by a set of n illuminated pixels (which defines the flash) is

$$\left\langle \xi_\lambda(\hat{\Omega}) \right\rangle \equiv \frac{1}{\Delta\Omega}\int_{\Delta\Omega}\xi_\lambda(\hat{\Omega})d\Omega = \frac{1}{\Delta\Omega}\int_{\Delta\Omega}\int_{t_o}^{t_m} I_\lambda(\hat{\Omega},t)dtd\Omega = \frac{1}{\Delta\Omega}\sum_{i=1}^{m}\sum_{j=1}^{n}\Delta\omega_j\bar{\xi}_{\lambda j}(i) \ , \qquad \textbf{(A3)}$$

Previous OTD/LIS "Radiance" Data Products: The "event radiance" (ER), "group radiance" (GR), and "flash radiance" (FR) data products derived for OTD and LIS, are not really radiances, but were defined as follows:

$$ER \equiv \bar{\xi}_{\lambda j}(i) \ , \quad GR \equiv \sum_{s=1}^{n_{pi}}\bar{\xi}_{\lambda s}(i) \ , \quad FR \equiv \sum_{i=1}^{m}\sum_{k=1}^{n_i}\bar{\xi}_{\lambda k}(i) \ . \qquad \textbf{(A4)}$$

The definitions of the various solid angles appearing above are:

$$\begin{aligned}
&\Delta\omega_j = \text{ solid angle of } j^{th} \text{ event subtended at entrance aperture = solid angle of } j^{th} \text{ pixel,}\\
&\Delta\omega_{pis} = \text{ solid angle of } s^{th} \text{ event of } p^{th} \text{ group in } i^{th} \text{ frame subtended at entrance aperture,}\\
&\Delta\Omega_{pi} = \sum_{s=1}^{n_{pi}}\Delta\omega_{pis} = \text{ solid angle of } p^{th} \text{ group in } i^{th} \text{ frame subtended at entrance aperture,}\\
&\Delta\Omega = \sum_{j=1}^{n}\Delta\omega_j = \text{ solid angle of flash subtended at entrance aperture .}
\end{aligned} \qquad \textbf{(A5)}$$

The definition of the various sample sizes are defined as follows:

$$\begin{aligned}
&n_{pi} = \#\text{ events in } p^{th} \text{ group of } i^{th} \text{ frame,}\\
&n_i = \#\text{ events in } i^{th} \text{ frame,}\\
&n = \#\text{ pixels illuminated by the flash.}
\end{aligned} \qquad \textbf{(A6)}$$

So the issue is to determine how the ER, GR, and FR data products are related to the (physically understandable) solid-angle-averaged spectral energy densities given in (A1), (A2), and (A3). The first one is trivial; i.e., ER is identically equal to the expression in (A1). To find relationships for the other two, the assumption that the pixel field-of-view is constant is used; i.e.,

$$\Delta\omega_j = \Delta\omega_{pis} \cong \Delta\omega = \text{const} \ ,$$

$$\Delta\Omega_{pi} = \sum_{s=1}^{n_{pi}} \Delta\omega_{pis} \ \cong \ \sum_{s=1}^{n_{pi}} \Delta\omega = n_{pi}\Delta\omega \ , \qquad \textbf{(A7)}$$

$$\Delta\Omega = \sum_{j=1}^{n} \Delta\omega_j \cong \ \sum_{s=1}^{n} \Delta\omega = n\Delta\omega \ .$$

Using the approximations in (A7), the expressions in (A2) and (A3) reduce to

$$\bar{\zeta}_{\lambda p}(i) = \frac{1}{\Delta\Omega_{pi}} \sum_{s=1}^{n_{pi}} \Delta\omega_{pis} \bar{\xi}_{\lambda s}(i) \ \cong \frac{1}{n_{pi}\Delta\omega} \sum_{s=1}^{n_{pi}} \Delta\omega \bar{\xi}_{\lambda s}(i) = \frac{1}{n_{pi}} \sum_{s=1}^{n_{pi}} \bar{\xi}_{\lambda s}(i) = \frac{1}{n_{pi}} GR \ ,$$

$$\left\langle \xi_\lambda(\hat{\Omega}) \right\rangle = \frac{1}{\Delta\Omega} \sum_{i=1}^{m} \sum_{j=1}^{n} \Delta\omega_j \bar{\xi}_{\lambda j}(i) \ \cong \frac{1}{n\Delta\omega} \sum_{i=1}^{m} \sum_{j=1}^{n} \Delta\omega \bar{\xi}_{\lambda j}(i) = \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n} \bar{\xi}_{\lambda j}(i) = \frac{1}{n} FR \ . \qquad \textbf{(A8)}$$

So the "radiance" data products are related to the solid-angle-average energy densities as follows:

$$ER = \bar{\xi}_{\lambda j}(i) \ ,$$

$$GR \cong n_{pi} \bar{\zeta}_{\lambda p}(i) \ , \qquad \textbf{(A9)}$$

$$FR \cong n \left\langle \xi_\lambda(\hat{\Omega}) \right\rangle \ .$$

The expressions in (A9) are the desired results. Again, these results are provided with the understanding that instrument measurement error/noise, threshold settings, and false alarms were all ignored. Accounting for these effects is more complicated (i.e., see Appendix 3 which was extracted from the recent journal article: Koshak, W. J., Optical Characteristics of OTD Flashes and the Implications for Flash-Type Discrimination, accepted in *J. Atmos. Oceanic Technol.*, May 2010).

It would be best if a look-up table was available that provided the solid angle of each pixel. Then one could define the following 3 energy products (in units of $\mu J/m^2/\mu m$) for events, groups, and flashes, respectively:

$$e_E \equiv \bar{\xi}_{\lambda j}(i) \Delta\omega_j \cong \Delta\omega \cdot ER \ ,$$

$$e_G \equiv \bar{\zeta}_{\lambda p}(i) \Delta\Omega_{pi} = \sum_{s=1}^{n_{pi}} \Delta\omega_{pis} \bar{\xi}_{\lambda s}(i) \ \cong \Delta\omega \sum_{s=1}^{n_{pi}} \bar{\xi}_{\lambda s}(i) = \Delta\omega \cdot GR \ , \qquad \textbf{(A10)}$$

$$e_F \equiv \left\langle \xi_\lambda(\hat{\Omega}) \right\rangle \Delta\Omega = \sum_{i=1}^{m} \sum_{j=1}^{n} \Delta\omega_j \bar{\xi}_{\lambda j}(i) \ \cong \Delta\omega \sum_{i=1}^{m} \sum_{j=1}^{n} \bar{\xi}_{\lambda j}(i) = \Delta\omega \cdot FR \ .$$

If the look-up table is available, then the exact sums in (A10) could be carried out. If the look-up table is not available, then a constant estimate $\Delta\omega$ would have to be used [along with the approximative expressions shown in (A10)]. Note that the definitions, (e_E, e_G, e_F) , track the full energy of the event, group, or flash intercepted by the GLM sensor (when the values are integrated over the entrance aperture and filter bandwidth). Whereas, the alternate values $(\bar{\xi}_{\lambda j}(i), \bar{\zeta}_{\lambda p}(i), \left\langle \xi_\lambda(\hat{\Omega}) \right\rangle)$ track only the solid-angle-averaged energy

densities; i.e., loosely speaking, the “average brightness” of the events, groups, and flashes.

Appendix 3: The Flash Radiance Data Product

In the prior appendix, the flash "radiance" data product, based on heritage OTD/LIS definitions, was described in terms of familiar physical quantities. In this appendix, additional rigor is added in describing this product.

A3.1 Preliminary Definitions and Derived Quantities

Given an optical emission like a lightning flash, the spectral radiance $R_\lambda = R_\lambda(\mathbf{r}, \hat{\Omega}, t)$ detected by an observer depends on the observer's location $\mathbf{r}$, the direction the observer is looking $\hat{\Omega} = \hat{\Omega}(\theta, \varphi)$ where (θ, φ) are polar and azimuth angles respectively, and on the time *t* of observation. The variable λ is the wavelength, and the units of R_λ are in $W/m^2/sr/\mu m$. Since one is interested in the spectral radiance observed by GLM, $\mathbf{r} = \mathbf{r}_o(t)$ where $\mathbf{r}_o(t)$ is the location of the center of the GLM entrance aperture and is a known vector function of time related to the satellite emphemeris and attitude. Eliminating $\mathbf{r}$ in favor of *t* gives the spectral radiance $I_\lambda = I_\lambda(\hat{\Omega}, t)$ impinging on the GLM entrance aperture.

The GLM instrument frame time τ is 2 *ms*. Suppose that only one flash occurs within the GLM field-of-view, is completely within the field-of-view, and subtends a solid angle $\Delta\Omega$ as viewed from GLM, i.e., as viewed at location $\mathbf{r}_o(t)$. Let us also assume that the flash occupies $i = 1, ..., m$ frames, where the i^{th} frame begins at time t_{i-1} and ends at time t_i, and $\tau = t_i - t_{i-1} = 2\ ms$. The spectral energy density $\xi_\lambda(\hat{\Omega}, i)$ (in units of $J/m^2/sr/\mu m$) impinging on the GLM entrance aperture from direction $\hat{\Omega}$ that is collected during the i^{th} frame is defined as

$$\xi_\lambda(\hat{\Omega}, i) \equiv \int_{t_{i-1}}^{t_i} I_\lambda(\hat{\Omega}, t)dt \ . \qquad \textbf{(A11)}$$

In general, the flash of solid angle $\Delta\Omega$ illuminates $j = 1, ..., n$ pixels and the solid angle subtended at the sensor by the j^{th} pixel footprint is given by $\Delta\omega_j$. Therefore, the flash solid angle can be discretized as

$$\Delta\Omega = \sum_{j=1}^{n} \Delta\omega_j \ . \qquad \textbf{(A12)}$$

So the spectral radiance and (single frame) spectral energy density averaged over the flash solid angle $\Delta\Omega$ are, repectively

$$\left\langle I_\lambda(\hat{\Omega},t)\right\rangle \equiv \frac{1}{\Delta\Omega}\int_{\Delta\Omega} I_\lambda(\hat{\Omega},t)d\Omega \ , \quad \left\langle \xi_\lambda(\hat{\Omega},i)\right\rangle \equiv \frac{1}{\Delta\Omega}\int_{\Delta\Omega} \xi_\lambda(\hat{\Omega},i)d\Omega \ , \tag{A13}$$

and the same variables averaged over the j^{th} pixel footprint of solid angle $\Delta\omega_j$ are

$$\bar{I}_{\lambda j}(t) \equiv \frac{1}{\Delta\omega_j}\int_{\Delta\omega_j} I_\lambda(\hat{\Omega},t)d\Omega \ , \quad \bar{\xi}_{\lambda j}(i) \equiv \frac{1}{\Delta\omega_j}\int_{\Delta\omega_j} \xi_\lambda(\hat{\Omega},i)d\Omega \ . \tag{A14}$$

Because of (A11) one also has the relations

$$\left\langle \xi_\lambda(\hat{\Omega},i)\right\rangle = \int_{t_{i-1}}^{t_i} \left\langle I_\lambda(\hat{\Omega},t)\right\rangle dt \ , \quad \bar{\xi}_{\lambda j}(i) = \int_{t_{i-1}}^{t_i} \bar{I}_{\lambda j}(t)dt \ . \tag{A15}$$

Since the integrals in (A13) can be rewritten as a sum of integrals, i.e.,

$$\int_{\Delta\Omega}(\ )d\Omega \ \rightarrow \ \sum_{j=1}^{n}\int_{\Delta\omega_j}(\ )d\Omega \ , \tag{A16}$$

(A14) can be used to rewrite the expressions in (A13) as

$$\left\langle I_\lambda(\hat{\Omega},t)\right\rangle = \frac{1}{\Delta\Omega}\sum_{j=1}^{n}\Delta\omega_j\bar{I}_{\lambda j}(t) \ , \quad \left\langle \xi_\lambda(\hat{\Omega},i)\right\rangle = \frac{1}{\Delta\Omega}\sum_{j=1}^{n}\Delta\omega_j\bar{\xi}_{\lambda j}(i) \ . \tag{A17}$$

Since the flash duration is $\Delta t = t_m - t_o$ and occupies $i = 1,...,m$ frames, the total spectral energy density received from direction $\hat{\Omega}$ during flash time Δt is

$$\xi_\lambda(\hat{\Omega}) \equiv \int_{t_o}^{t_m} I_\lambda(\hat{\Omega},t)dt = \sum_{i=1}^{m}\int_{t_{i-1}}^{t_i} I_\lambda(\hat{\Omega},t)dt = \sum_{i=1}^{m}\xi_\lambda(\hat{\Omega},i) \ . \tag{A18}$$

Averaging (A18) over the flash solid angle $\Delta\Omega$ and using the second equation in (A17) gives

$$\left\langle \xi_\lambda(\hat{\Omega})\right\rangle = \left\langle \sum_{i=1}^{m}\xi_\lambda(\hat{\Omega},i)\right\rangle = \sum_{i=1}^{m}\left\langle \xi_\lambda(\hat{\Omega},i)\right\rangle = \frac{1}{\Delta\Omega}\sum_{i=1}^{m}\sum_{j=1}^{n}\Delta\omega_j\bar{\xi}_{\lambda j}(i) \ . \tag{A19}$$

The quantity $\left\langle \xi_\lambda(\hat{\Omega}) \right\rangle$ is important since it represents the solid-angle-averaged spectral energy density intercepted by GLM due to the flash of duration Δt. Another important quantity is the total spectral energy per area (in units of $J/m^2/\mu m$) intercepted by GLM, call it u_λ, given by

$$u_\lambda \equiv \int_{t_o}^{t_m} \int_{\Delta\Omega} I_\lambda(\hat{\Omega}, t) d\Omega dt = \sum_{i=1}^{m} \sum_{j=1}^{n} \Delta\omega_j \overline{\xi}_{\lambda j}(i) . \quad \textbf{(A20)}$$

The last result is obtained by using the first equation in (A14) and the second equation in (A15), and by rewriting the integrals as sums over i^{th} frames and j^{th} pixels. Comparing (A19) and (A20) allows one to link the two fundamental quantities as follows

$$u_\lambda = \left\langle \xi_\lambda(\hat{\Omega}) \right\rangle \Delta\Omega \ . \quad \textbf{(A21)}$$

A3.2 An Expression for the Data Product

Consider the following proxy variable P_λ (the simple sum of the mean spectral energy densities across each illuminated pixel from each frame, in units of *J/m^2/sr/μm*)

$$P_\lambda \equiv \sum_{i=1}^{m} \sum_{j=1}^{n} \overline{\xi}_{\lambda j}(i) = \sum_{i=1}^{m} \sum_{k=1}^{n_i} \overline{\xi}_{\lambda k}(i) \ . \quad \textbf{(A22)}$$

Here, note that n_i is the number of pixels in the i^{th} frame that are illuminated by the flash, and n is the total number of pixels that are illuminated by the flash. The two sums are necessarily equivalent because $\overline{\xi}_{\lambda j}(i) = 0$ when the j^{th} pixel is *not* illuminated by the flash during the i^{th} frame. The heritage flash "radiance" data product can be written

$$P_\lambda^* = \sum_{i=1}^{m} \sum_{k=1}^{n_i} \overline{\xi}_{\lambda k}^*(i) + \sum_{i=1}^{m} \sum_{l=1}^{N_i} \varepsilon_{\lambda l}(i) \ . \quad \textbf{(A23)}$$

Here, N_i is the number of pixels in the i^{th} frame that are *not* illuminated by the flash. The (threshold-modified) average spectral energy density for the j^{th} pixel illuminated in the i^{th} frame, $\overline{\xi}_{\lambda j}^*(i)$, and the (threshold-modified) false alarm spectral energy density, $\varepsilon_{\lambda l}(i)$, are given by

$$\begin{aligned} \overline{\xi}_{\lambda k}^*(i) &= \overline{\xi}_{\lambda k}(i) + e_{\lambda k}(i) + T_{\lambda k}(i) , \\ \varepsilon_{\lambda l}(i) &= e_{\lambda l}(i) + T_{\lambda l}(i) \ . \end{aligned} \quad \textbf{(A24)}$$

Note that the instrument measurement errors are described by $e_{\lambda k}(i)$, and the effect of employing a threshold setting is described by $T_{\lambda k}(i)$. For example, if the instrument measurement $\overline{\xi}_{\lambda k}(i)+e_{\lambda k}(i)$ is below the instrument threshold setting, then $T_{\lambda k}(i)=-\overline{\xi}_{\lambda k}(i)-e_{\lambda k}(i)$ such that $\overline{\xi}^{*}_{\lambda k}(i)=0$, otherwise, $T_{\lambda k}(i)=0$. Similarly, given that the l^{th} pixel in the i^{th} frame is not illuminated, if the instrument measurement error $e_{\lambda l}(i)$ is below the instrument threshold setting, then $T_{\lambda l}(i)=-e_{\lambda l}(i)$ such that $\varepsilon_{\lambda l}(i)=0$, otherwise, $T_{\lambda l}(i)=0$ and a false alarm occurs. Note also that the flash "radiance" data product is not technically a radiance value but rather a spectral energy density; however this was the vernacular used in describing the data product within the OTD/LIS software distribution. Plugging (A24) into (A23) gives

$$P^{*}_{\lambda}=\sum_{i=1}^{m}\sum_{k=1}^{n_i}\overline{\xi}_{\lambda k}(i)+\left[\sum_{i=1}^{m}\sum_{k=1}^{n_i}\left\{e_{\lambda k}(i)+T_{\lambda k}(i)\right\}+\sum_{i=1}^{m}\sum_{l=1}^{N_i}\left\{e_{\lambda l}(i)+T_{\lambda l}(i)\right\}\right]\equiv P_{\lambda}+E_{\lambda}\ , \qquad \textbf{(A25)}$$

where (A22) was used, and the term in the square brackets is defined as the total spectral error, E_{λ}. Hence, if one assumes that $\Delta\omega_j\cong\Delta\omega=const$, then: $\Delta\Omega\cong n\Delta\omega$, $u_{\lambda}\cong\Delta\omega\sum_{i=1}^{m}\sum_{j=1}^{n}\overline{\xi}_{\lambda j}(i)$ and the data product can be readily written in terms of the well known radiometric quantities derived in the previous section; i.e.,

$$P^{*}_{\lambda}\cong\frac{1}{\Delta\omega}u_{\lambda}+E_{\lambda}=n\left\langle\xi_{\lambda}(\hat{\Omega})\right\rangle+E_{\lambda}\ . \qquad \textbf{(A26)}$$

This is the desired result. It clearly indicates how the flash "radiance" data product is related to familiar quantities [the total intercepted spectral energy per area, u_{λ}, and the average spectral flash energy density $\left\langle\xi_{\lambda}(\hat{\Omega})\right\rangle$].

Appendix 4: Additional Considerations on Centroiding

Consider basic centroiding for say a group composed of *N* events. It would look like:

$$\varphi_{Centroid} = \frac{\sum_{i=1}^{N} A_i \varphi_i}{\sum_{i=1}^{N} A_i} \quad . \qquad \textbf{(A27)}$$

Here: $\varphi_i = $ latitude of ith pixel, and $A_i = $ optical amplitude of ith pixel. A similar expression holds for the longitude.

First, note that one would like to do the centroiding using real lat/lon for each pixel, so that implies a desire to have a lat/lon for each pixel (i.e., a look-up table; call it table #1). The AWG Lightning Team has been told that such a table already exists as of this writing, but the team has not obtained or analyzed the table.

Second, the instrument lightning signal count output for a particular ith pixel that is governed by a particular RTEP is presumed to have associated physical units given by $\mu J / m^2 / \mu m / ster$ and is an average radiance across that pixel's field-of-view; i.e., the counts denote a solid-angle-averaged radiance across the pixel. [How this connection is presumably obtained: In the calibration lab, one illuminates the pixel across its entire field-of-view with an optical source of known radiance, and then simply reads the instrument count output, thereby relating solid-angle-averaged radiance to instrument count output.] All of this assumes that GLM will be calibrated like OTD and LIS were. Note that in such a calibration procedure, the particular pixel field-of-view is not considered because instrument counts are just being related to a "per steradian" quantity. Nonetheless, these counts are then shipped over to the LCFA for centroiding. So one has instrument counts, C_i, and the associated calibrated instrument counts, I_i, in units of $\mu J / m^2 / \mu m / ster$ to work with. Clearly, one would opt to use $A_i = I_i$, not $A_i = C_i$ because $I_i(C_i)$ is in general presumed to be a nonlinear function.

But, one can be even more careful and ask what one physically wants from a centroiding. Each ith pixel corresponds to some footprint on cloud-top that is emitting some (diffuse, multiple-scattered) lightning optical radiance. Some footprints are very bright, and some are dimmer. The radiance across a single footprint varies in general, and also has some average value. What one wants, ideally, is to get the best optical centroiding of (in this example) a group. The idea is to weight each pixel-footprint of the group by how much optical energy it is associated with. To do this, one would want to weight the center location of each footprint with the appropriate optical energy intercepted by the pixel associated with that footprint.

This intercepted energy is $B_i = I_i \Delta\omega_i$, which is in units of $\mu J / m^2 / \mu m$, where $\Delta\omega_i$ is the solid angle of the ith pixel in units of steradians. [Technically, one could also multiply by filter bandwidth which would also slightly vary with lens boresight (and hence pixel location in the CCD), but this is presumably a smaller effect due to near normal incidence of lightning optical rays into filter. Also, the aperture area (which gives the per square meter units) is presumed fixed.]

In other words, if the amplitude in (A27) is to be correct, it appears that it must include the solid angle of the pixel.

So, in a nutshell, if one is not given a look-up table #2 of each pixel's field of view (in units of steradians) then one is forced to just use the following formula for the centroiding:

$$\varphi_{I} \equiv \frac{\sum_{i=1}^{N} I_i \varphi_i}{\sum_{i=1}^{N} I_i} \quad . \qquad \textbf{(A28)}$$

Without look-up table #2, the ATBD just implies using equation (A28) above with the units, $\mu J / m^2 / \mu m / ster$ for I_i.

Otherwise, with look-up table #2, one could do the following more accurate computation:

$$\varphi_B \equiv \frac{\sum_{i=1}^{N} B_i \varphi_i}{\sum_{i=1}^{N} B_i} = \frac{\sum_{i=1}^{N} I_i \Delta\omega_i \varphi_i}{\sum_{i=1}^{N} I_i \Delta\omega_i} . \qquad \textbf{(A29)}$$

So, in summary, note that the computations in (A28) and (A29) are NOT equivalent, and (A29) is more accurate than (A28) because it actually captures the true weight to be assigned to the ith pixel; i.e., it includes specific information about the solid angle of the pixel.

Finally, note that (A28) incorrectly assumes that the solid angle of each pixel is equivalent; i.e., if the pixel solid angles are all equivalent then one obtains

$$\frac{\sum_{i=1}^{N} I_i \Delta\omega_i \varphi_i}{\sum_{i=1}^{N} I_i \Delta\omega_i} = \frac{\sum_{i=1}^{N} I_i \Delta\omega \varphi_i}{\sum_{i=1}^{N} I_i \Delta\omega} = \frac{\Delta\omega \sum_{i=1}^{N} I_i \varphi_i}{\Delta\omega \sum_{i=1}^{N} I_i} = \frac{\sum_{i=1}^{N} I_i \varphi_i}{\sum_{i=1}^{N} I_i} \equiv \varphi_I . \qquad \textbf{(A30)}$$