PROCEEDINGS OF SCIENCE

Performance Modeling and Improvements on the GRB Source Localization Streaming Pipeline Aboard the Antarctic Demonstrator for the Advanced Particle-Astrophysics Telescope (ADAPT)

Ye Htet,[a] **Marion Sudvarg,**[b] **Honghao Yang,**[a] **Jeremy Buhler,**[a,*] **Roger Chamberlain,**[a] **Wenlei Chen**[c] **and James Buckley**[b]

[a] *Dept. of Computer Science and Engineering, Washington University*
One Brookings Drive, St. Louis, MO, USA
[b] *Dept. of Physics, Washington University*
One Brookings Drive, St. Louis, MO, USA
[c] *Dept. of Physics, Oklahoma State University*
501 Athletic Avenue, Stillwater, OK, USA
E-mail: {htet.ye, msudvarg, honghao, jbuhler, roger, buckley}@wustl.edu, wenlei.chen@okstate.edu

The Advanced Particle-astrophysics Telescope (APT) is a mission concept for a space-based gamma-ray telescope whose capabilities include prompt localization of gamma-ray bursts (GRBs) to support multi-wavelength and multi-messenger astrophysics. ADAPT — APT's balloon-borne prototype — can localize GRBs in well under a second using on-board computing hardware. ADAPT will partner with ground-based, fast-slewing optical telescopes, rapidly providing alerts that enable the partner to observe a short-duration burst within a few seconds of detection.
In this work, we investigate the utility of having ADAPT issue progressively more accurate location estimates for a GRB as detected Compton events from the burst accumulate over time. We develop a computational model to estimate how frequently ADAPT can compute these estimates, finding that it can do so at least every 150 ms for a 1 MeV/cm^2 burst on a low-power quad-core Intel Atom processor. We then assess how quickly ADAPT's localization improves as it observes more events and show that a partner instrument can slew to a burst's location faster if it exploits progressive location estimates than if it waits for one final estimate. Real-time, on-board source localization thus has a role to play in cooperative observation of gamma-ray transients even when data collection time, rather than computing time, dominates the cost of detection.

39th International Cosmic Ray Conference (ICRC2025)
15–24 July 2025
Geneva, Switzerland

ICRC 2025
The Astroparticle Physics Conference
Geneva July 15-24, 2025

*Speaker

1. Introduction

The Advanced Particle-astrophysics Telescope (APT) is a mission concept for a space-based observatory designed to observe high-energy gamma rays and cosmic rays in support of multi-wavelength and multi-messenger astrophysics [1–5]. APT will promptly detect energetic transient events in the distant universe, especially gamma-ray bursts (GRBs) in the MeV energy range, and will rapidly communicate these events to narrow-field-of-view partner instruments for follow-up observation at other wavelengths. APT will be deployed in a Sun-Earth Lagrange L_2 orbit, achieving nearly full-sky field of view and order-of-magnitude improvement in GRB detection sensitivity compared with current instruments such as Fermi [2, 6, 7]. The Antarctic Demonstrator for APT (ADAPT), a small-scale technology demonstration mission for APT's hardware design and computational capabilities, will launch using a high-altitude balloon in late 2026 [4, 8–11].

Prompt optical counterparts of dim, short-duration transients might be visible for only seconds, so they must be accurately localized, and partner instruments notified to observe them, within a short time. In prior work on ADAPT, we developed a localization pipeline [4, 10] based on Compton reconstruction that runs aboard the telescope and is typically capable of localizing a short 1 MeV/cm^2 GRB in the sky to within 5–6 degrees in only a few hundred milliseconds after a one-second exposure. (APT's larger detector should deliver sub-degree localization accuracy at similar speed). This on-board capability avoids communication bandwidth constraints and latency associated with sending raw detector signals to earth for localization and so decreases the time to produce an informative alert for observing partners.

Because localization runs on a shorter timescale than some GRBs, its latency is limited by the time needed to observe enough Compton events to obtain an accurate location. The necessary exposure time could stretch to multiple seconds for dimmer bursts. A question, then, is whether observing partners can benefit from ADAPT's delivering an earlier, less accurate localization – particularly if it is followed by later, more accurate updates. For optical telescopes that slew fast enough to promptly observe a GRB, we argue that the answer is "yes." Early localization could enable the partner to begin slewing toward the source immediately, while later location updates allow it to course-correct. In the end, the partner could reach its observing target sooner than if it had waited for the final result.

In this work, we investigate ADAPT's ability to provide progressively more accurate localizations in real time during a GRB, along with this capability's utility to observing partners. In Section 3, we first devise a computational performance model of ADAPT's pipeline to assess how frequently localizations can be computed, given that the pipeline is also responsible for reconstructing a stream of individual Compton events as they arrive. We then obtain empirical estimates of the model's parameters to show that, on a low-power Intel quad-core processor, localization can produce a new result at least every 150 ms. Section 4 assesses the utility of progressive localization by modeling a partner telescope that uses early source location estimates to begin slewing as soon as possible and corrects its path as localization improves. ADAPT's progressive localization for a one-second, 1MeV/cm^2 burst could reduce a fast-slewing (20 deg/sec) partner instrument's time to target by up to 19% compared to waiting for one final result.

2. Background

When ADAPT's detector triggers due to a GRB in the MeV energy range, it begins to observe gamma-ray interactions, primarily via Compton scattering. The detector's attached ASIC and FPGA logic [9, 12] converts each such interaction to its position $\mathbf{r} = (x, y, z)$ and its deposited energy E. For a single gamma ray, the flight software gathers its list of interactions (r_i, E_i) and communicates the list to the analysis pipeline via an in-memory *event queue*.

As shown in Figure 1, the analysis pipeline consists of two stages: *reconstruction* and *localization*. Reconstruction converts the interaction list for each photon to a constraint on the source's location, in the form of a *Compton ring* – a circle on the unit sphere on which the source must lie. A photon's interactions are converted to (one, or two possible) Compton rings by inferring their temporal ordering from energetic and kinematic constraints via the Compton law [13]. In prior work, we developed an accelerated algorithm for reconstruction [3, 4] that can process $> 10^5$ photons per second on ADAPT's flight computer [14].

Figure 1: ADAPT GRB localization pipeline.

Reconstruction continuously accumulates Compton rings in a *Compton ring queue*. At any time, the pipeline can perform *localization* by combining the constraints from all stored rings to produce an estimated source direction vector $\mathbf{s}$. Localization must contend both with uncertainty in the parameters of each inferred Compton ring and with spurious rings arising from atmospheric background radiation unrelated to the GRB; such rings can account for $\geq 50\%$ of all rings seen by ADAPT even during a 1 MeV/cm^2 burst. As described in [4], ADAPT's localization algorithm first constructs a coarse estimate of the source direction from a random sample of rings in the queue, then performs iterative least-squares refinement of this estimate using all available rings to obtain its final result. We recently added a machine-learning component [15] to localization to help filter out rings that are likely caused by background particles. With this addition, localization now iterates steps of approximation, refinement, and filtering to maximize the fraction of non-background Compton rings used to estimate the source's location.

An estimated source location is communicated to partner instruments in an *alert* message. A telescope with a narrow field of view that receives an alert must slew its detector toward the source to observe it. While traditional optical instruments can move at most a few degrees per second, more recent mounts [16] permit direct geodesic movement at tens of degrees per second. Although this work focuses on simplified alerts with a single point estimate of the source, our ongoing work on real-time production of more informative source *likelihood maps* [17, 18] shows that these maps can be computed on a timescale similar to that assumed for point estimates and so could also be used in the future to guide a fast-slewing optical telescope.

3. Performance of the GRB Localization Pipeline

To assess the feasibility of progressive localization during a GRB, we must model its added computational cost. The same processor that performs localization must also reconstruct a continuous stream of events from ADAPT's detector into Compton rings. Hence, time spent in extra localization decreases the computational resources available for reconstruction. Because the goal is

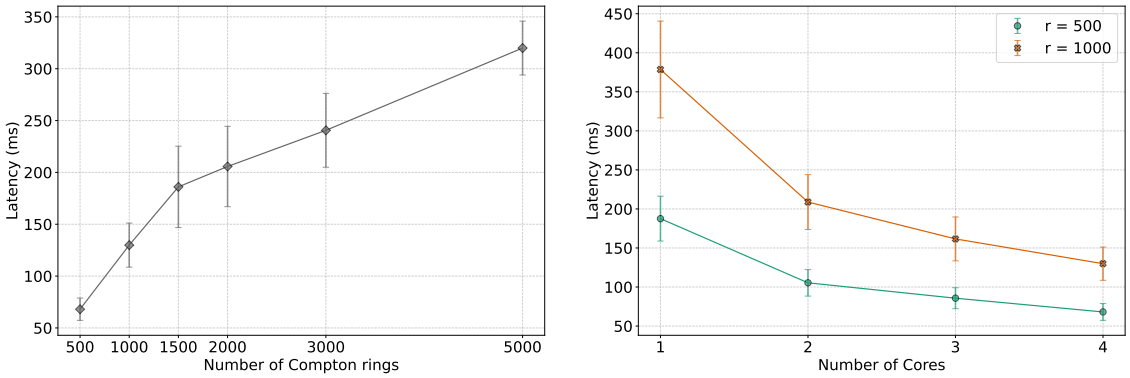
to analyze a GRB in real time, it would be undesirable to delay reconstruction indefinitely; hence, our model should help us judge the extent to which progressive localization can be added without undue impact on reconstruction.

Performance Modeling. In this work, the pipeline runs on ADAPT's flight instrument computer, a WINSYSTEMS EBC-C413 with a quad-core, 1.92 GHz Intel Atom E3845 CPU. We may utilize its multiple CPU cores in two ways: first, by partitioning the stream of events across parallel reconstruction threads on different cores (all of which write to the shared ring queue), and second, by using PyTorch and OpenMP multithreading to exploit the lower-level parallelism in localization's machine-learning and linear-algebraic operations, respectively. Because reconstruction is empirically much faster than localization, we assume here that a single core is responsible for reconstruction, while all four cores participate in localization whenever it runs. Reconstruction therefore pauses for the duration of each localization run. Future work will explore a broader set of strategies to map analysis tasks to processor cores.

We can empirically determine the time $t_{\text{loc}}(m, r)$ to perform localization on r rings using m cores. We also measure the time $t_{\text{recon}}(e)$ to perform reconstruction on e events using a single core. As noted, we assume that localization prevents reconstruction from running whenever it occurs. If a burst generates E events, it requires time $t_{\text{recon}}(E)$ for reconstruction; if all data should be analyzed within time T (e.g., the end of the burst, or a fixed deadline), that leaves time $T - t_{\text{recon}}(E)$ for localization. If each run of localization processes a maximum of R Compton rings from the ring queue, then we have time for at least $n_{\text{loc}} = \left\lfloor \frac{T - t_{\text{recon}}(E)}{t_{\text{loc}}(m,R)} \right\rfloor$ intermediate localizations before time T.

Measurements. We measured the performance of reconstruction and localization on our target platform for a 1 MeV/cm^2 burst occurring over one second, normally incident to ADAPT's detector. The burst's spectrum followed a Band function with parameters $\alpha = -0.5$, $E_{\text{peak}} = 490$ keV, and $\beta = -2.35$. The GRB light curve followed a Gaussian distribution, which was added to an atmospheric background as described in [8] that contributed a steady flux of 4.48×10^5 particles/sec. We generated incident particles from the burst and background and simulated their interactions with ADAPT's detector using GEANT4, then applied the detector simulation described in [9] to produce events for the analysis pipeline. We ran 300 trials with randomly generated sets of incident particles to assess the induced variability in reconstruction and localization costs.

Figure 2a shows the cost t_{loc} of localization with $m = 4$ cores as a function of the number of input Compton rings r. t_{loc} is not a simple linear function of r in part because its approximation sub-stage is limited to use a sample of at most 1000 Compton rings, and because the number of iterations needed to localize a source decreases as more of its Compton rings are observed. In this figure, the data point for r rings includes 1 s worth of background plus as many source rings as are needed to make up r in total; hence, larger ring counts reflect localization's expected latency at 1 s for increasingly bright GRBs. (The test burst falls at r between 500 and 600.) Overall, localization's average cost was under 200 ms even for more than 1500 Compton rings. In practice, this cost would vary depending on the rate at which putative background rings can be identified and discarded, which depends on GRB brightness as well as its angle of incidence to the detector. Figure 2b shows how t_{loc} varies with the number of cores. For 500–1000 rings, localization exhibited a speedup of



(a) t_{loc} vs. r for $m = 4$

(b) t_{loc} vs. m for $r = 500$ and $r = 1000$

Figure 2: Characterization of localization time t_{loc} on Compton rings from a test burst as a function of number of cores m and number of rings r. Error bars reflect one standard deviation over 300 trials.

2.9–2.7× with four cores versus a single core, justifying the utility of using all available cores in subsequent experiments.

Reconstruction received, over 300 trials, an average of $E = 31,746$ events from the burst with minimal variation (1 s.d. = 154). After filtering unreconstructable, low-quality, and suspected background events, the average yield of Compton rings was only $r = 581$ (1 s.d. = 24), corresponding to a localization time of < 100 ms with $m = 4$ cores. The average time $t_{\text{recon}}(E)$ was 134 ms, with a maximum observed time of 140 ms over all trials. Based on these observations, we estimate that localization for this burst can run at least $n_{\text{loc}} = \left\lfloor \frac{1-0.14}{0.1} \right\rfloor = 8$ times per second together with reconstruction without exceeding the capacity of the processor or delaying event reconstructions past one second after triggering.

4. Utility of Progressive Localization

The performance predictions of Section 3 show that running localization multiple times per second is feasible on our computing platform. We now turn to the question of how to exploit this capability in the context of cooperative observation between ADAPT and a fast-slewing optical telescope. The goal is to reduce the time for the partner instrument to reach the source location found by ADAPT, so that it can begin to capture light from any optical counterpart to the gamma-ray transient as soon as possible.

We use the following schedule for progressive localization. We first wait 200 ms after triggering for an initial set of events to be processed by reconstruction. Starting at 200 ms, we schedule a localization run to start every 150 ms using all cores (as guided by our performance model), with reconstruction running on one core between each localization run. Each run uses all available Compton rings in the queue at the time. This process continues until the burst ends. The 200 ms time used here is somewhat ad hoc, though guided by our estimate of the Compton event rate. Future work will systematically evaluate how ADAPT's localization error depends on the number of Compton rings and will use that knowledge, along with the rate at which rings accumulate, to set lower bounds on the time to first localization and the interval between runs.

Accuracy of Progressive Localization. Figure 3 illustrates the timing and accuracy of intermediate localizations computed under the above schedule as ADAPT observes the test burst described

in Section 3. As before, we ran 300 trials with randomly generated incident particles. Vertical error bars show the variability in the angular error between ADAPT's inferred source direction **s** and the true source direction, while horizontal error bars show the variability in when the localization run actually finishes. The "final" run occurs only after all Compton events from the burst have been processed (new data collection ceases at 1 s, after the last intermediate localization starts). The first localization is delivered in slightly under 400 ms and typically has a large error. However, this error rapidly decreases with the second and third localizations as more Compton rings accumulate, ultimately matching the final result. Variation in when each localization result is delivered is small, indicating that inter-trial variation in Compton ring accumulation rates was limited.

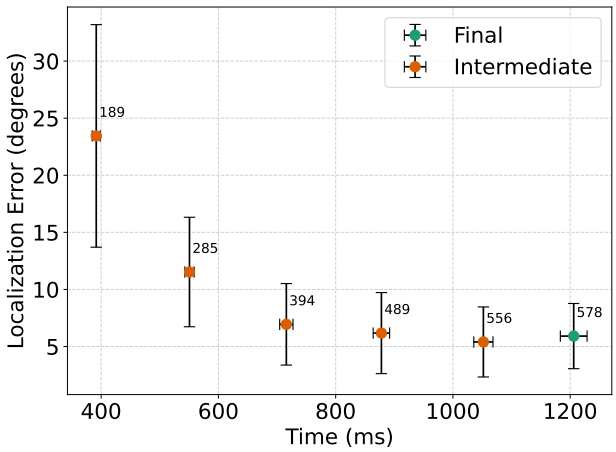


Figure 3: Accuracy of intermediate localization runs (orange) vs. final run with all data (green). Error bars reflect one standard deviation over 300 trials; counts on each point are average number of rings available for that localization run.

Impact on Cooperative Pointing. We now consider the impact on a partner telescope of using ADAPT's progressive localization to locate a transient. We assume that the partner can slew at a constant rate of 20°/sec in any direction, consistent with the Planewave direct-drive mounts used by, e.g., the 7DT telescope system [16]. When ADAPT generates its first source estimate **s**, the partner receives it with negligible delay and begins to move toward **s**. Each subsequent estimate is communicated to the partner (again with negligible delay), which changes its movement target to the updated source location. The partner stops moving once it reaches the most recently received **s**.

We compare the time needed for the partner to reach ADAPT's final estimated source direction under two strategies: the progressive strategy outlined above, and a baseline in which the partner waits to receive the final source direction before moving. The baseline requires no trajectory corrections but delays any movement for at least the full duration of the burst.

Figure 4 illustrates the difference in time-to-target for the two strategies for the test burst. For these experiments, we varied the actual source's polar angle to be either 0° (normally incident), 30°, or 60°; for the latter two, the azimuthal angle was 45°. The partner was assumed to begin pointing at the zenith for the 30- and 60-degree bursts, or at 60° for the normally incident burst, so that substantial movement was always necessary. Each experiment was again run for 300 trials with different sets of input particles to assess the impact of variable accuracy in ADAPT's source estimates. In all cases, mean time to target was roughly 4 s under the baseline strategy, while progressive localiza-

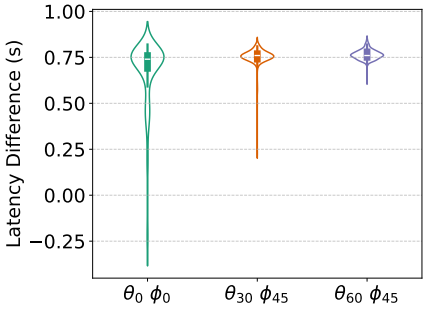


Figure 4: Latency advantage of progressive over baseline strategy for bursts at polar angle 0°, 30°, and 60°.

tion reduced this time by around 0.75 s (19%). Variability in the latency reduction between trials was due primarily to errors in earlier source directions, which were greatest for normally-incident bursts. We note that faster slew speeds improve the observed latency reduction: if the partner instrument is capable of slewing at 50°/sec (achievable with some optical telescopes today), exploiting progressive localization reduces the time to target by 40% vs. the baseline.

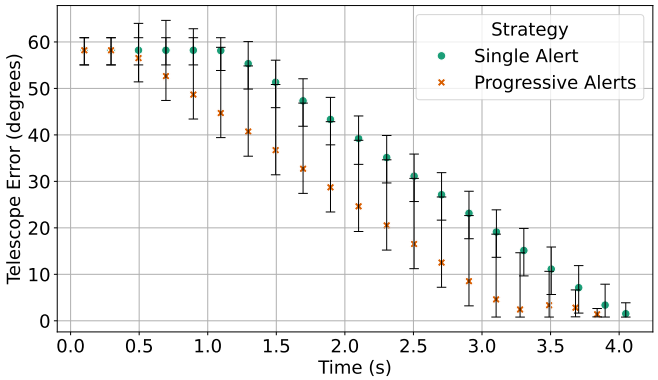


Figure 5: Trajectory of a telescope tracking the final alert from the two different strategies. The telescope initially points at 60° polar and 45° azimuthal and must move to a target within 5° of the zenith. Points show median errors over 230 trials, while error bars show range from min to max error.

Figure 5 illustrates for the normally-incident case how the *telescope error*, that is, the difference between the partner instrument's current pointing direction and the source direction given by the final localization run, evolves over time. Because telescope error is sensitive to the placement of ADAPT's final source estimate (which impacts distance to travel), the figure reflects only those trials (230/300) where the final estimate lies within 5° of the true source location. In the baseline ("single alert") strategy, the partner does not start to move until 1.2-1.3 s after the burst triggers the detector. In contrast, exploiting intermediate localizations begins to reduce the median telescope error (with the min and max error bars) vs. baseline at around 500 ms after triggering, which agrees with the observed improvement in time to target.

5. Conclusion and Future Work

The ADAPT and APT instruments will perform source localization and alert generation in real time via onboard computation. The low latency of these computations opens up new possibilities for how to use them in cooperative observation of high-energy transient phenomena at multiple wavelengths. We have demonstrated that real-time, progressive GRB localization is computationally feasible for ADAPT's analysis pipeline and have provided evidence that using progressive estimates computed every 150 ms can allow a fast-slewing optical telescope to reach the predicted location of a burst more quickly.

Future work will further develop and assess the utility of progressive localization. We will extend our computational cost model to accommodate different burst fluences and different processor resource allocations. We will improve the scheduling of progressive localization based on more systematic quantification of accuracy versus number of available Compton rings. We will consider more realistic models of partner telescopes that account for communication and control delays and settling time. Finally, we will study how to exploit progressive localization when the reported alert data is not a single source direction but rather an entire likelihood map. Ongoing work [17, 18] suggests that these maps can be produced using onboard computation on the same timescale as our current point localizations.

References

[1] W. Chen et al., *The Advanced Particle-astrophysics Telescope: simulation of the instrument performance for gamma-ray detection*, in *Proc. 37th Int'l Cosmic Ray Conf.*, vol. 395, pp. 590:1–590:9, 2021, DOI.

[2] J. Buckley et al., *The Advanced Particle-astrophysics Telescope (APT) project status*, in *Proc. 37th Int'l Cosmic Ray Conf.*, vol. 395, pp. 655:1–655:9, July, 2021, DOI.

[3] M. Sudvarg et al., *A fast GRB source localization pipeline for the Advanced Particle-astrophysics Telescope*, in *Proc. of 37th Int'l Cosmic Ray Conf.*, vol. 395, pp. 588:1–588:9, July, 2021, DOI.

[4] Y. Htet et al., *Prompt and accurate GRB source localization aboard the Advanced Particle Astrophysics Telescope (APT) and its Antarctic Demonstrator (ADAPT)*, in *Proc. 38th Int'l Cosmic Ray Conf.*, vol. 444, pp. 956:1–956:9, July, 2023, DOI.

[5] J.H. Buckley, J.D. Buhler and R.D. Chamberlain, *The Advanced Particle-astrophysics Telescope (APT): computation in space*, in *Proc. 21st ACM Int'l Conf. Computing Frontiers Workshops and Special Sessions*, pp. 122–127, May, 2024, DOI.

[6] C. Meegan, G. Lichti, P.N. Bhat et al., *The Fermi gamma-ray burst monitor*, *Astrophysical J.* **702** (2009) 791.

[7] J. D. Meyers, Curator, "Overview of the Fermi GBM." *https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Cicerone/Cicerone_Introduction/GBM_overview.html*, Jan., 2020.

[8] W. Chen et al., *Simulation of the instrument performance of the Antarctic Demonstrator for the Advanced Particle-astrophysics Telescope in the presence of the MeV background*, in *Proc. 38th Int'l Cosmic Ray Conf.*, vol. 444, pp. 841:1–841:9, July, 2023, DOI.

[9] M. Sudvarg et al., *Front-end computational modeling and design for the Antarctic Demonstrator for the Advanced Particle-astrophysics Telescope*, in *Proc. 38th Int'l Cosmic Ray Conf.*, vol. 444, pp. 764:1–764:9, July, 2023, DOI.

[10] M. Sudvarg, C. Zhao, Y. Htet, M. Konst et al., *Hls taking flight: Toward using high-level synthesis techniques in a space-borne instrument*, in *Proc. 21st ACM Int'l Conf. Computing Frontiers*, pp. 115–125, 2024.

[11] Y. Htet, M. Sudvarg, J.D. Buhler, R.D. Chamberlain and J.H. Buckley, *Localization of gamma-ray bursts in a balloon-borne telescope*, in *Proc. Wkshps. Int'l Conf. High Performance Computing, Network, Storage, and Analysis (SC-W)*, pp. 395–398, Nov., 2023, DOI.

[12] M. Sudvarg et al., *FPGA-Based Data Processing using High-Level Synthesis on the Antarctic Demonstrator for the Advanced Particle-astrophysics Telescope (ADAPT)*, in *Proc. 39th Int'l Cosmic Ray Conf.*, vol. 501, pp. 852:1–852:9, July, 2025.

[13] S.E. Boggs and P. Jean, *Event reconstruction in high resolution Compton telescopes*, *Astronomy and Astrophysics Suppl. Series* **145** (2000) 311.

[14] M. Sudvarg et al., *Parameterized workload adaptation for fork-join tasks with dynamic workloads and deadlines*, in *Proc. of 29th Int'l Conf. on Embedded and Real-Time Computing Systems and Applications (RTCSA)*, 2023, DOI.

[15] Y. Htet, M. Sudvarg, D. Butzel, J.D. Buhler, R.D. Chamberlain and J.H. Buckley, *Machine learning aboard the ADAPT gamma-ray telescope*, in *Proc. of Workshops of the International Conference on High Performance Computing, Network, Storage, and Analysis (SC-W)*, pp. 4–10, Nov., 2024, DOI.

[16] J.H. Kim, M. Im, H. Lee, S.-W. Chang, H. Choi and G.S.H. Paek, *Introduction to the 7-Dimensional Telescope: commissioning procedures and data characteristics*, in *Ground-based and Airborne Telescopes X*, H.K. Marshall, J. Spyromilio and T. Usuda, eds., vol. 13094, p. 130940X, International Society for Optics and Photonics, SPIE, 2024, DOI.

[17] D. Wang, Y. Htet, M. Sudvarg, R. Chamberlain, J. Buhler and J. Buckley, *Coordinating instruments for multi-messenger astrophysics*, in *Proc. of 22nd ACM International Conference on Computing Frontiers Workshops and Special Sessions*, pp. 213–218, May, 2025, DOI.

[18] J. Buhler and M. Sudvarg, *Real-time Likelihood Map Generation to Localize Short-duration Gamma-ray Transients*, in *Proc. 39th Int'l Cosmic Ray Conf.*, vol. 501, pp. 587:1–587:9, July, 2025.

Acknowledgments

We wish to acknowledge our colleagues in the APT collaboration (https://adapt.physics.wustl.edu/). This work was supported by NASA award 80NSSC21K1741.