A Novel FRB Detection Pipeline For NenuFAR

Julien Plante, Damien Gratadour, Louis Bondonneau, Cédric Viou

A Novel FRB Detection Pipeline For NenuFAR

Julien Plante ,[1] Damien Gratadour,[1] Louis Bondonneau,[1] and Cédric Viou[2]

[1]*LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, Meudon, Hauts-de-Seine, France;* `julien.plante@obspm.fr`

[2]*Station de Radioastronomie de Nançay, Observatoire de Paris, PSL Research University, CNRS, Université d'Orléans, Nançay, Cher, France*

Abstract. NenuFAR is a very large low-frequency radiotelescope, designed to observe the sky at frequencies ranging from 10 to 85 MHz. One of the main science case is the study of pulsars and Fast Radio Bursts (FRBs), and this program already has detected hundreds of these objects and events. However, the current detection pipeline is not able to perform a 24/7 survey of the sky. This is not an issue for the observation of recurring events such as pulsars emissions, but limits the number of FRB detections, and discovery opportunities. Here, we propose a novel real-time FRB detection system for NenuFAR, including a custom hardware platform and a software solution, designed to detect transient events in real-time and trigger signal storage on event detection to reduce memory footprint. This experiment on NenuFAR has also been designed as a pathfinder for more efficient transient events detection on SKA, with scalability as one of the core specifications. In this paper, we present the design of the experiment, detail the underlying hardware and software technologies and discuss initial results from a benchmarking campaign.

1. Introduction

Multiple Fast Radio Burst (FRB) detection pipelines are available nowadays (Petroff et al. 2019). Because of the low observation frequency of the NenuFAR telescope (10-85MHz) (Zarka et al. 2020), the dispersive delays of FRBs are up to a thousand times greater than for usual radiotelescopes (with typical observations centered around 1GHz). This makes existing pipelines perform poorly on NenuFAR data (increased memory footprint and latency, reduced Dispersion Measure (DM) search capabilities).

To address this limitation, we propose a pipeline tailored for low-frequency, real-time FRB detection.

2. Experiment Description

The pipeline currently under development is summarized on Figure 1, while more in-depth explanations are given in this section.

NenuFAR data is multicasted to experiments using UDP packets. To maximize the bandwidth and minimize memory footprint, we use DPDK + GPUDirect. This cutting-edge technology introduced by Nvidia enables direct transfer from the Network Interface Controller (NIC) to the GPU, without the need for a bounce buffer on CPU.

Incoming data is distributed on GPUs by frames of a few seconds in a round-robin fashion. Using multiple GPUs increases the time available for one iteration, making more complex operations possible.

Radio Frequency Interference (RFI) mitigation is performed after a simple pre-processing (time integration + power law), to identify and remove signal parts contaminated by undesired radio signals such as human emissions. This stage is not yet implemented in our pipeline.

Dedispersion is then applied to compensate the dispersion effect caused by the Interstellar Medium (ISM). We based our work on the Dispersion Measure Transform (DMT), with a custom strategy to address the detection of highly dispersed (> 10 s) bursts. Using piecewise integration, we are able to compute the same DMT by iterating on shorter time frames, while being independent of the underlying DMT algorithm. Current implementation uses the Bruteforce DMT algorithm (Barsdell et al. 2012).

The final stage detects energy peaks in the DMT. We propose an image processing approach to take advantage of highly optimized GPU libraries for image processing, such as Nvidia Performance Primitives (NPP). The current implementation is very limited, simply based on a global maximum and threshold, but we are working on more robust algorithms.

FRB candidates will trigger an external ring buffer of several seconds-long based on memory-only storage. That will allow raw waveform data retrieval and their recording on HDD to bring only signals of interest to the final user.

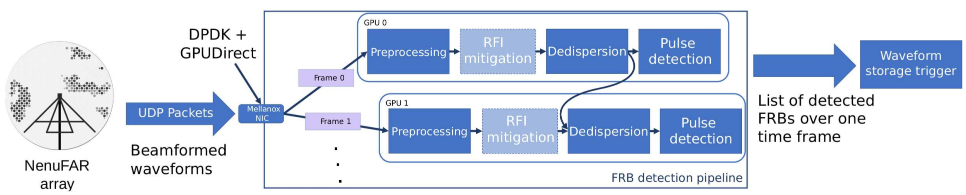

Figure 1. Pipeline architecture

This pipeline will be powered by the COSMIC framework (Ferreira et al. 2020) for real-time GPU computing, to provide stable performance, flexibility and a comprehensive user interface.

3. Initial Results

We compared the current performance of our pipeline to that of Heimdall, another GPU-accelerated FRB detection pipeline, and studied the new data acquisition approach.

3.1. DPDK + GPUDirect Benchmark

A benchmark has been realized to assess the performance of the DPDK + GPUDirect data acquisition method. Figure 2 show the results of a 1-minute benchmark of the latency, at 1.5 Gbit/s (the maximum bandwidth of NenuFAR for this experiment). These results were obtained with two Mellanox Connectx-5 interconnected with a 100 GbE fibre. The latency is measured as the delay from the first send to the full reception in GPU memory of 100kB chunks of data.

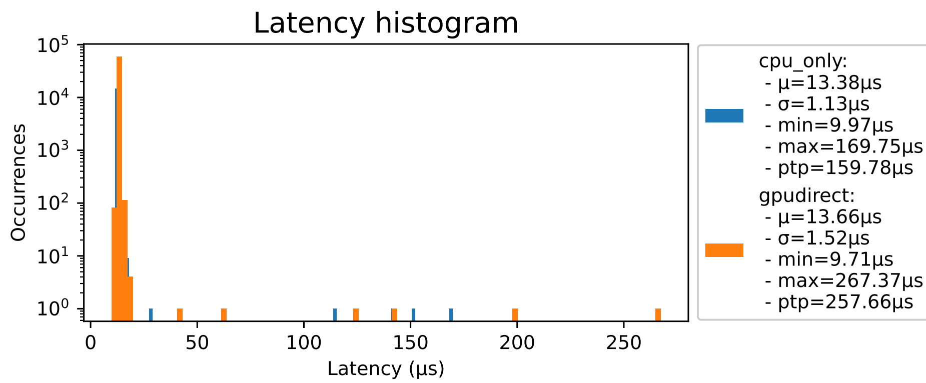

Figure 2. Latency over 1 minute at 1.5 Gbit/s

The mean latency in both NIC to CPU and NIC to GPU modes is about 13 µs, which will allow for a quick response of the detection pipeline. Some outliers are present, and will have to be dealt with to use this technology in a strong real-time environment. We believe this can be solved by tuning the system configuration (CPU shielding, real-time kernel, ...).

Moreover, we found that this result is strongly dependent on the PCIe topology of the experiment. With the NIC on the same PCIe bridge than the GPU, bandwidths up to 80 Gbit/s were measured. However, with a different topology, bandwidths down to 2 Gbit/s were noticed, with much higher latencies (hundreds of µs).

3.2. Computations Benchmark

To compare with Heimdall, we used NenuFAR-like data: 200 kHz sampling frequency (with time integration over 16 samples), 2^{21} samples per frame, 256 DM. The benchmark was run on one Nvidia Quadro GV100 (no multi-GPU for fair comparison of the implementations).

	Heimdall	This pipeline
RFI mitigation	25 s	N/A
Dedispersion	3.3 ms	4.6 ms
Pulse detection	800 ms	6.8 ms

Table 1. Computation time comparison for each stage

Both Heimdall and our pipeline are already able to detect FRBs in real-time (2^{21} samples at 200 kHz is about 10.5s) using high-end CPU and GPU, and still have a great margin for their processing time compared to a single frame time. This gives us room for thorough RFI mitigation, or opens the door to using lower-end hardware.

Our current implementation of dedispersion is 40% slower than Heimdall's, but enables the detection of FRBs with a dispersive delay longer than the time frame duration. It also allows easy pipelining across multiple GPUs, thus increasing the available processing time with each added GPU. Still, our implementation is very naive, and speedups are expected with further optimizations.

Our implementation of pulse detection is more than 100x faster than Heimdall's, likely thanks to our 2D approach to pulse detection in the DMT space. Though this approach is promising, the current implementation is not robust to RFIs, and needs fine-tuning of a threshold.

4. Conclusion

We presented initial results concerning a new FRB detection pipeline tailored for use with NenuFAR. This system will make possible to scale down the memory requirement for FRB study, by detecting FRBs on-the-fly, and saving only the FRB-related signal on disk.

Using DPDK + GPUDirect, we are able to provide a low-latency, high-bandwidth, and rather standard data acquisition method. While this technology is currently implemented using Nvidia hardware, we believe this could be ported to other manufacturer's hardware. This data acquisition method is more than enough to fill the requirements in this application, but enables high scalability.

A new approach to dedispersion emerged from this work, especially useful for low-radio observation, by removing the need for overlap and making possible to detect pulses longer than the time frame studied. This approach easily enables multiple-GPUs parallelism, increasing the available time for processing one time frame.

This progress is promising for the deployment of an online robust FRB detection pipeline for the low-frequency radiotelescope NenuFAR, with a first prototype integration planned for the first half of 2022.

5. Perspectives

The current priority is to complete the pipeline by linking the DPDK + GPUDirect to the rest of the pipeline. Further important steps are to implement a state-of-the-art RFI mitigation tuned for NenuFAR, and more robust pulse detection algorithms to provide an end-to-end prototype ready for integration at Nançay.

Next iterations will also focus on optimizing the DMT implementation, either by optimizing memory accesses in the current Bruteforce DMT, or by using the Fast DMT (Zackay & Ofek 2017), an algorithm with reduced complexity for the DMT computation. The outliers in latency for the data acquisition method are also to be looked into.

References

Barsdell, B. R., Bailes, M., Barnes, D. G., & Fluke, C. J. 2012, MNRAS, 422, 379. 1201.5380

Ferreira, F., Sevin, A., Bernard, J., Guyon, O., Bertrou-Cantou, A., Raffard, J., Vidal, F., Gendron, E., & Gratadour, D. 2020, in Adaptive Optics Systems VII, edited by L. Schreiber, D. Schmidt, & E. Vernet, International Society for Optics and Photonics (SPIE), vol. 11448, 239

Petroff, E., Hessels, J. W. T., & Lorimer, D. R. 2019, The Astronomy and Astrophysics Review, 27, 4. URL https://doi.org/10.1007/s00159-019-0116-6

Zackay, B., & Ofek, E. O. 2017, The Astrophysical Journal, 835, 11

Zarka, P., Denis, L., Tagger, M., Girard, J., Coffre, A., Dumez-Viou, C., Taffoureau, C., Charrier, D., & other members of the NenuFAR-France team 2020, in GASS (URSI)