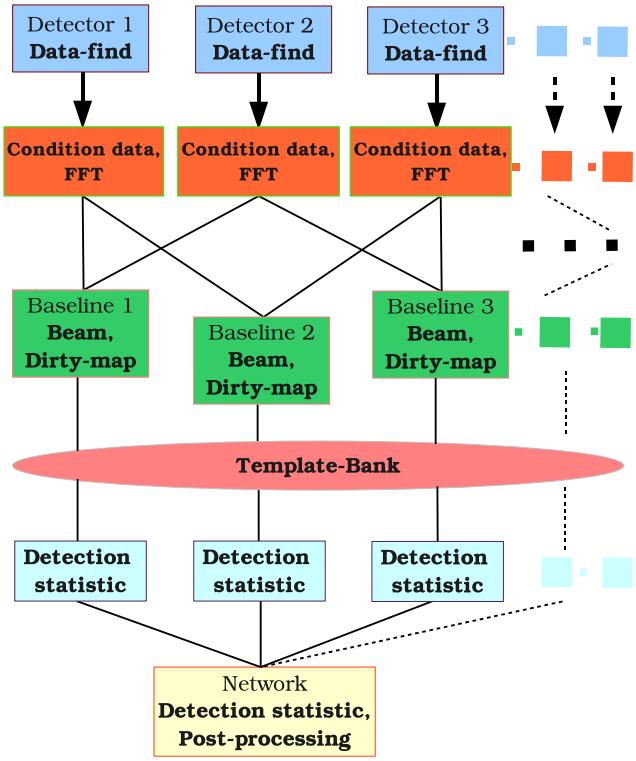
OSG Summer School 2011
DIPONGKAR TALUKDER
Department of Physics and Astronomy,
Washington State University.

Supporting Multi-baseline Gravitational Wave Radiometry using High-Throughput Computing

A gravitational wave (GW) is a fluctuation in the curvature of space-time that propagates with the speed of light that, as predicted by Einstein's theory of general relativity. When two massive objects, like neutron stars, orbit each other, space-time is stirred by their motion, and gravitational radiation ripples outward into the universe. The Laser Interferometer Gravitational-Wave Observatory (LIGO), is a large-scale physics experiment aiming to directly detect gravitational waves. In the past few years, the first generation LIGO detectors have successfully operated at or near design sensitivity [1]. The LIGO detectors are currently in the process of being upgraded to advanced LIGO (aLIGO) and are due to come on line in 2015. The new detectors will improve the strain sensitivity of current instruments by a factor of ten, with a thousandfold increase in the observable volume of space. With advanced detectors and from year-long observations, we expect to see stochastic GW background [2-4]. This gives rise to the expectation that the detection of gravitational waves will reveal a new view of the universe. A stochastic GW background can arise from an unresolved superposition of multiple astrophysical sources, such as: (a) core collapse supernovae, (b) rotating neutron stars with hydrodynamical instabilities, (c) rapidly spinning asymmetric neutron stars, and (d) coalescing compact binaries [4]. Multi-baseline radiometry is a tool for searching GW from (anisotropic) stochastic sources with a network of baselines [5].



Since the targeted source is stochastic, we search for its GW signal by looking for correlated patterns in the data of two or more detectors after accounting for time delays and detector responses consistent with a given sky location. This is done by cross correlating the data from the detectors, taken in pairs, with a

sky-position-dependent time-frequency filter. This cross-correlated output is termed as dirty map. In the presence of a signal in the detector data, the dirty map can be express as sum of two terms: beam-pattern function convolved with pixel-strength vector and pixel-noise [5]. The dirty map is filtered against a bank of templates modeled by distribution of sources across the sky. The normalized filter output is termed as maximum-likelihood-ratio (MLR) statistic. The MLRs from the baselines can be combined to make a Network MLR statistic. The two major parts in this analysis that demand computation are computing dirty maps and filtering.

For our research, we used computer simulations to develop and test our search method. So far we have managed to prototype our method on a few computers running Linux with four processors each. But the applications of this method to (real) LIGO data have necessitated the use of a network of computers or grid computing. The analysis of a month of LIGO data requires many thousand of jobs, running for days on hundreds of CPUs. The above figure is a schematic diagram of the search pipeline. The analysis described here will be carried out on the LIGO Data Grid (LDG) clusters. The nodes are managed by a Condor system for large scale computing, using its directed acyclic graph (DAG) option to execute the whole pipeline. A part of the pipeline (up to computing the Dirty-map) has already been developed for the LDG clusters. We are planning to build the rest and make it ready for aLIGO. This extended pipeline will also be tested on existing LIGO data in a few months.

The pipeline is designed as follows: The DAG is used to parcel the data into time periods (eg. a month of data). The data is segmented into 60 second intervals, high-passed at 32Hz, decimated to 4096 Hz, Hann windowed, and a cross-correlation analysis is performed on each segment. The segments are then combined into the final result (baseline dirty-maps). The Condor DAGMan is used to manage these complex work-flows. The complex work-flows ensure data from multiple detectors are analyzed in sequence. The analysis is not CPU intensive–computer constraint is the disk I/O in loading the recorded strain data into memory. For a month of data, a dag file contains approximately 43200 jobs. For the search, computing the cross-correction statistic for each frequency bin for a baseline requires on the order of 5TB of data storage. We do not save all this data; instead we run the full analysis and only save the dirty maps (few GB). Once the dirty maps are computed, the filtering jobs start. There are tens of thousands of templates that have to be filtered (fitted or matrix multiplied) against the dirty maps. The filtering process will also be handled using Condor DAGMan (I will be working on it soon). The total CPU time for this process is a few hundred hours. The whole run for a month of data will take about 3 days to complete on the cluster of 300+ nodes with 4GB RAM in each. The results are collected at the end and post-processed if needed.

Acknowledgments

First of all, I would like to express my gratitude and warm-hearted thanks to OSG summer school program committee chair Tim Cartwright and my advisor Sukanta Bose for their inspiration and kind co-operation. I am delighted to interact with the people involved in the Condor team. Special thanks to Alain Roy, Igor Sfiligoi, Brian Bockelman, and Zach Miller for managing the hands-on activities. I take this opportunity to thank my friends in OSG summer school and TeraGrid'11 conference for their friendship and entertainment. I cannot end without thanking my wife Shampa Biswas for providing a loving environment for me.

References

[1] http://www.ligo.caltech.edu/ .

[2] T. Regimbau, Res. Astron. Astrophys. **11** 369 (2011) .
[3] S. Dhurandhar et al., arXiv:1105.5842v1 [gr-qc] .
[4] X.-J. Zhu et al., arXiv:1104.3565v3 [gr-qc] .
[5] D. Talukder et al., Phys. Rev. D **83** 063002 (2011).