University of Groningen

# Vast and Fast Data in the era of the large astrophysics and particle physics experiments

Gazagnes, Simon

*DOI:*
10.33612/diss.179743481

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*
Publisher's PDF, also known as Version of record

*Publication date:*
2021

Link to publication in University of Groningen/UMCG research database

*Citation for published version (APA):*
Gazagnes, S. (2021). *Vast and Fast Data in the era of the large astrophysics and particle physics experiments*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. https://doi.org/10.33612/diss.179743481

**Copyright**
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

**Take-down policy**
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

Download date: 20-03-2023

# SUMMARY FOR NON-EXPERTS

Here, I provide a summary of the work completed in the context of this thesis. I guarantee that this summary does not require an extensive scientific background to understand the main outcomes presented in the different chapters of this book.

Since the beginning of the third millennium, we have entered the Information Age, an era dominated by vast volumes of data and information generated by our modern societies. This new era has critical implications for research, the progress of which now mostly relies on large cutting-edge research experiments. The data collected by these experiments will be complex to handle. Data complexity can be the result of several aspects. In this thesis, we principally dealt with vast data volumes and fast data acquisition systems. Extracting and analyzing the information content from these systems is one of the major endeavors that we need to overcome to open new pathways towards data-centered discoveries.

To provide adapted solutions to tackle the hurdles that these upcoming cutting-edge experiments will face, the Data Science community has been growing at an incredible pace. Inter-disciplinary projects, grouping scientists from diverse fields such as mathematics, computer science, physics, and biology, are modern approaches to dealing with the growing data-centric challenges. This thesis is one example of such effort: the project [VF]ast data, funded by a grant at the Centre for Data Science and Systems Complexity (DSSC) from the University of Groningen, has been meant to benefit from the latest developments in Mathematical morphology, a growing branch in the field of Image and Signal processing, to provide new tools allowing us to deal optimally with current and upcoming research experiments in astrophysics and particle physics.

This inter-disciplinary Ph.D. project led to five main scientific projects focusing on different aspects. In **Chapter 2** and **Chapter 3**, I presented two published studies that detail a novel computational tool, DISCCOFAN (DIStributed Connected COmponent Filtering and ANalysis), based on the recent developments in the field of Mathematical Morphology. This tool provides an efficient method to analyze the structures and patterns observed in two- and three-dimensional data sets. Several domains such as biomedical imaging and astronomy have applications working with the notion of objects (e.g. a blood vessel or a galaxy), such that we need to efficiently analyze the properties of these structures. To optimally deal with these applications, a specific class of mathematical morphology techniques, referred to as component trees, has been developed. Component trees are hierarchical structures that represent the nested relations of connected regions (i.e. structures) in the image. Hence, they provide a different representation of the original data set that helps us to optimally analyze the information contained in the data. When I started this Ph.D., these techniques were limited to relatively modest image sizes, up to a few gigapixels (a few billion points in the data set). However, they could not handle gigantic images with hundreds or thousands of billions of pixels because manipulating these data sets requires a gigantic amount of computational power on a single machine and very few machines in the world dispose of such capabilities. Benefiting from the progress of parallel computing, we developed DISCCOFAN, a novel computational tool that enables the use of component tree techniques to deal with enormous two and three-dimensional images. DISCCOFAN distributes the computational footprint over several independent processes that can complete series of tasks concurrently. Hence, it is a promising tool to deal with the large and high-resolution images and volumes that the next generation of research experiments will collect.

Chapter 4 and Chapter 5 present two astrophysical projects focusing on exploring the physical processes that governed the Epoch of Reionization, a phase transition in the Universe that happened in the first billion years. This cosmic epoch has major implications for the Universe as we know it today, especially because the first generation of astronomical objects formed during this period. Because it happened billions of years ago, collecting observations of these objects is extremely challenging. Hence, despite being a critical milestone in the Universe's history, we still have little information about the underlying physical processes that governed this epoch. However, several upcoming telescopes will enable us to observe the first stars and galaxies that existed during this period. These observations will help us answering fundamental questions related to the formation of the first astronomical objects and their impact on the evolution of our Universe.

In Chapter 4, I analyzed the properties of nearby galaxies that have similar characteristics to the first astronomical objects in the Universe. By doing so, we aim at establishing a consistent theoretical framework that can be used, in turn, to interpret the upcoming observations of the first population of galaxies that existed during the Epoch of Reionization. This analysis provided valuable insights to understand how these objects might have impacted the Universe during this epoch. Then, the upcoming James Webb Space Telescope and ground-based extremely large telescopes will provide us revolutionary observations to assess the outcomes of this work.

Chapter 5 builds upon a different observation probe, referred to as "21-cm observations", to explore the physical processes that governed the Epoch of Reionization. Rather than directly observing astronomical objects such as stars or galaxies, 21-cm observations provide a unique way to explore the Universe during the first billion years based on the evolution of the gas in between stars and galaxies. The formation of the first astronomical objects directly impacted the morphology and topology of the gas regions in the Universe such that we can use 21-cm observations to explore their properties. Using the method presented in the first two chapters, I studied the morphology of the gas at these epochs using simulations and highlighted that this approach should provide valuable insights into the astrophysics governing the evolution of the Universe at these epochs. This work is particularly relevant in the context of the Square Kilometer Array, an upcoming radio-telescope that will collect revolutionary 21-cm observations.

Finally, Chapter 6 details the last project completed in the context of this thesis. This project focused on designing an efficient track reconstruction algorithm for upcoming particle accelerator experiments. The latter builds upon high-energy collisions of fundamental particles (e.g. protons, neutrons, electrons) to probe the properties of the fundamental building blocks of matter and the physics governing the very small scales. In order to detect and analyze very rare particles, accelerator experiments need to operate at extremely high interaction rates, with billions of billions of particle collisions per second. Because the amount of data produced by these experiments is too large to be stored, it requires real-time data processing systems, referred to as "trigger systems", that enable one to select the most relevant interactions. These systems often embed a track reconstruction algorithm that identifies in real-time the particle trajectories using the data collected from the detectors. This is because reconstructing the trajectories of the particles in the detectors allows us to extract their key properties. In this last chapter, we designed a fast track-reconstruction algorithm for a specific particle collision experiment to be performed in the near future that will operate at very high interaction rates (hence, collecting fast data). We showed that this method is promising for reconstructing particle trajectories on the fly.

Overall, this inter-disciplinary thesis provided novel computational methods to deal with upcoming data challenges related to *vast* and *fast* research experiments in astrophysics and particle physics. More importantly, it paves the way for future works, extending this research in diverse directions. Finally, this thesis highlights that the current and upcoming progress in data analysis techniques will enable key data-centered discoveries, providing crucial insights to answer fundamental questions in diverse scientific fields. Hence, while this era will have critical challenges to overcome, there is a bright future for the science that will emerge from these upcoming *vast* and *fast* research experiments.