

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](https://doi.org/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.



rijksuniversiteit  
groningen

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

**Proefschrift**

ter verkrijging van de graad van doctor aan de  
Rijksuniversiteit Groningen  
op gezag van de  
rector magnificus prof. dr. C. Wijmenga  
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

vrijdag 1 oktober 2021 om 12:45 uur

door

**Simon Gazagnes**

geboren op 18 augustus 1993  
te Martigues, Frankrijk

**Promotores**

Prof. dr. N. Kalantar-Nayestanaki

Prof. dr. L.V.E. Koopmans

**Copromotores**

Dr. M.H.F. Wilkinson

Dr. J.G. Messchendorp

**Beoordelingscommissie**

Prof. dr. G. Mellema

Prof. dr. J. Ritman

Prof. dr. J.B.T.M. Roerdink

*Everything has its wonders, even darkness and silence...*

Helen Keller

COVER DESIGN: Remco Wetzels & Simon Gazagnes, image from Sergey Nivens.  
PRINTED BY: ProefschriftMaken  
ISBN: 978-94-6423-410-7

# CONTENTS

<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
1.1	MATHEMATICAL MORPHOLOGY	4
1.1.1	First developments	5
1.1.2	Connected operators	6
1.1.3	Tree construction algorithms	10
1.2	THE EPOCH OF REIONIZATION	13
1.2.1	Exploring the properties of the first galaxies with UV spectroscopic observations	15
1.2.2	Investigating cosmic reionization with 21-cm observations	18
1.3	HIGH-ENERGY PARTICLE COLLISIONS	24
1.4	THESIS OVERVIEW	28
<b>2</b>	<b>DISTRIBUTED COMPONENT FORESTS IN 2-D: HIERARCHICAL REPRESENTATIONS SUITABLE FOR TERA-SCALE IMAGES</b>	<b>35</b>
2.1	INTRODUCTION	36
2.2	CONNECTED FILTERS AND MULTI-SCALE TOOLS	37
2.3	COMPONENT TREES AND FORESTS	38
2.3.1	Parallel Computation	39
2.3.2	Distributed Component Forests	40
2.3.3	Granulometries and Pattern Spectra	44
2.3.4	Differential Area Profiles	45
2.4	RESULTS	47
2.4.1	Timings and speed-up	49
2.4.2	Memory efficiency	53
2.5	CONCLUSIONS	56
<b>3</b>	<b>DISTRIBUTED CONNECTED COMPONENT FILTERING AND ANALYSIS IN 2-D AND 3-D TERA-SCALE DATA SETS</b>	<b>61</b>
3.1	INTRODUCTION	62

3.2	CONNECTED COMPONENT ANALYSIS	63
3.2.1	Connected Components	63
3.2.2	Attribute Filters and Multi-Scale Analysis	64
3.3	THE COMPONENT TREE	65
3.3.1	Tree construction	66
3.3.2	Parallelization strategies, literature and limitations	67
3.4	DISTRIBUTED CONNECTED COMPONENT FILTERING AND ANALYSIS	71
3.4.1	Building the Local Component Trees	71
3.4.2	Optimal selection of border nodes	72
3.4.3	Merging Boundary Trees	74
3.4.4	Updating the trees	77
3.5	PERFORMANCE TESTING	80
3.5.1	2-D random images	81
3.5.2	2-D realistic data sets	86
3.5.3	3-D data sets	89
3.6	CONCLUSION	92
<b>4</b>	<b>THE ORIGIN OF THE ESCAPE OF LYMAN-<math>\alpha</math> AND IONIZING PHOTONS IN LYMAN CONTINUUM EMITTERS</b>	<b>97</b>
4.1	INTRODUCTION	98
4.2	DATA	102
4.3	METHOD	104
4.3.1	Neutral gas properties	104
4.3.2	Constraining the dust extinction	112
4.3.3	Ly $\alpha$ properties	116
4.4	THE ISM POROSITY ENABLES THE ESCAPE OF LY $\alpha$ AND LYC PHOTONS	120
4.4.1	The scattering of Ly $\alpha$ photons in a porous ISM	120
4.4.2	Narrower H I absorption lines lead to larger observed Ly $\alpha$ emission	124
4.4.3	The Ly $\alpha$ equivalent width scales with low $C_f(\text{H I})$	130
4.4.4	The escape of Ly $\alpha$ and LyC photons is larger in galaxies with lower $C_f(\text{H I})$ and $E_{\text{B-V}}$	132
4.4.5	Predicting the LyC escape fraction from absorption line measurements	136
4.4.6	The flux at the Ly $\alpha$ profile minimum correlates with the LyC escape fraction	139
4.5	DISCUSSION	142
4.5.1	The ISM porosity enables the escape of Ly $\alpha$ and LyC photons	142

4.5.2	Indirectly constraining $f_{\text{esc}}(\text{LyC})$ at high- $z$	145
4.6	SUMMARY	148
4.A	FITS	150
4.B	TABLES	150
4.B.1	Lyman series covering fraction measurements	150
4.B.2	Lyman series velocity width measurements	150
4.B.3	Significance levels of the trends in this work	150
4.C	PLOTS OF $\text{Ly}\alpha$ AND H I ABSORPTION LINE PROFILES	156
<b>5</b>	<b>INFERRING THE PROPERTIES OF THE SOURCES OF REIONIZATION USING THE MORPHOLOGICAL SPECTRA OF THE IONIZED REGIONS</b>	<b>163</b>
5.1	INTRODUCTION	164
5.2	SIMULATIONS AND INSTRUMENT SENSITIVITY	167
5.2.1	21CMFAST	168
5.2.2	Interferometer characteristics	171
5.3	THE MORPHOLOGY OF THE IONIZED REGIONS	172
5.3.1	Morphological attributes	172
5.3.2	Extracting morphological attributes with DISCCOFAN	175
5.4	MORPHOLOGY OF THE IONIZED REGIONS DURING THE EOR	176
5.4.1	Evolution during reionization	176
5.4.2	Morphological spectra for different reionization scenarios	179
5.4.3	Impact of the point spread function and noise	182
5.5	21CMMC SETUP	186
5.5.1	21cmmc using the power spectrum	187
5.5.2	21cmmc using tomographic statistics	188
5.5.3	Mock observations	190
5.6	RESULTS	192
5.6.1	Inference using the ionized regions number-count and morphology	192
5.6.2	Comparison with the power spectrum	198
5.6.3	Impact of different initial conditions	201
5.6.4	Impact of foreground residuals	204
5.7	DISCUSSION	208
5.7.1	The ionized regions morphology probes the average ionizing budget per baryon	208
5.7.2	Combining power spectrum and tomography	212
5.7.3	Limitations of using tomography	213
5.8	CONCLUSIONS	215



5.A	COMPARISON OF THRESHOLDING METHODS FROM MOCK TOMOGRAPHIC OBSERVATIONS	217
5.B	POWER SPECTRUM OF THE MOCK OBSERVATIONS	219
5.C	LIKELIHOOD FOR N-DIMENSIONAL DISTRIBUTIONS	219
5.D	SAMPLE VARIANCE TESTS	222
<b>6</b>	<b>FAST TRACK RECONSTRUCTION USING THE PANDA STRAW TUBE TRACKER</b>	<b>229</b>
6.1	INTRODUCTION	230
6.2	THE PANDA STRAW TUBE TRACKER	232
6.2.1	x-y projection	233
6.2.2	z-y projection	237
6.3	THE EVENT RECONSTRUCTION ALGORITHM	240
6.3.1	x-y reconstruction	240
6.3.2	z-reconstruction	248
6.4	PERFORMANCE ASSESSMENT	253
6.4.1	Comparison with BABAI20	254
6.4.2	Performance as a function of track complexity	260
6.5	CONCLUSIONS	269
6.A	ALGORITHMS	270
6.A.1	The <i>connect</i> phase	271
6.A.2	The <i>fitting</i> phase	271
6.A.3	z reconstruction	274
<b>7</b>	<b>CONCLUSIONS AND FUTURE PROSPECTS</b>	<b>279</b>
7.1	OVERVIEW PER CHAPTER	280
7.2	GENERAL CONCLUSIONS AND FUTURE PROSPECTS	284
	<b>SUMMARY</b>	<b>293</b>
	<b>SAMENVATTING</b>	<b>297</b>
	<b>RÉSUMÉ</b>	<b>301</b>
	<b>ACKNOWLEDGEMENTS</b>	<b>305</b>

# LIST OF ACRONYMS

<b>AGN</b>	Active Galactic Nuclei
<b>CLASSY</b>	COS Legacy Archive Spectroscopic Survey
<b>CMB</b>	Cosmic Microwave Background
<b>COS</b>	Cosmic Origins Spectrograph
<b>DCF</b>	Distributed Component Forest
<b>DSSC</b>	Data Science and Systems Complexity
<b>E-ELT</b>	European Extremely Large Telescope
<b>EMC</b>	Electromagnetic Calorimeter
<b>EoR</b>	Epoch of Reionization
<b>FAIR</b>	Facility of Antiproton and Ion Research
<b>FUV</b>	Far-Ultraviolet
<b>GMRT</b>	Giant Metrewave Radio Telescope
<b>GMT</b>	Giant Magellan Telescope
<b>HERA</b>	Hydrogen Epoch of Reionization Array
<b>HESR</b>	High Energy Storage Ring
<b>HI</b>	Neutral hydrogen
<b>HST</b>	Hubble Space Telescope
<b>IGM</b>	Intergalactic medium
<b>ISM</b>	Interstellar medium
<b>JWST</b>	James Webb Space Telescope
<b>LHC</b>	Large Hadron Collider
<b>LOFAR</b>	Low Frequency Array
<b>Ly<math>\alpha</math></b>	Lyman- $\alpha$
<b>LyC</b>	Lyman continuum
<b>MPI</b>	Message Passing Interface
<b>MVD</b>	Micro Vertex Detector
<b>MWA</b>	Murchison Widefield Array
<b>PANDA</b>	antiProton ANnihilation at DArmstadt
<b>PAPER</b>	Precision Array to Probe Epoch of Reionization
<b>SKA</b>	Square Kilometer Array
<b>STT</b>	Straw Tube Tracker
<b>TMT</b>	Thirty Meter Telescope
<b>UV</b>	Ultraviolet

