Cosmic depth and detail: advancing LOFAR imaging workflows to unveil the deep high-resolution universe

Jong, J.M.G.H.J. de

Citation

Jong, J. M. G. H. J. de. (2025, May 9). *Cosmic depth and detail: advancing LOFAR imaging workflows to unveil the deep high-resolution universe*. Retrieved from https://hdl.handle.net/1887/4245860

Popular scientific summary

Radio from the heavens

While investigating sources of noise that could interfere with telecommunications, physicist and engineer Karl Jansky unexpectedly detected in 1933 radio emission at a frequency of 20.5 MHz coming from the centre of our galaxy. With this discovery, he unknowingly opened a new window to explore the Universe at radio frequencies. Jansky's discovery was a few years later confirmed and expanded upon by engineer and amateur radio astronomer Grote Reber, who detected signals from the same direction as Jansky at 162 MHz, along with radio emissions from the Andromeda galaxy and other areas in the sky. The following post-Second World War era marked the true emergence of radio astronomy, fueled by advancements in radar technology developed during the war and a subsequent shift toward peaceful scientific endeavours.

Key discoveries

This new window to study our Universe at radio frequencies led to many remarkable discoveries. The largest and brightest objects were identified first, including galaxies and remnants of powerful star explosions within our own galaxy. Soon after, other phenomena were uncovered, such as the hydrogen line, which allows astronomers to explore the structure and dynamics of galaxies and the early universe. Another ground-breaking discovery was the cosmic microwave background, providing a snapshot of the faint afterglow of the Big Bang. Pulsars – rapidly rotating neutron stars that emit beams of radiation – were also detected, offering a precise cosmic clock and deepening our understanding of extreme physics. More recently, fast radio bursts, mysterious and powerful flashes of radio waves, have become a new exciting topic in radio astronomy, while even brown dwarfs have been detected with radio telescopes as well. The following two sections will highlight discoveries at radio wavelengths most relevant to this thesis.

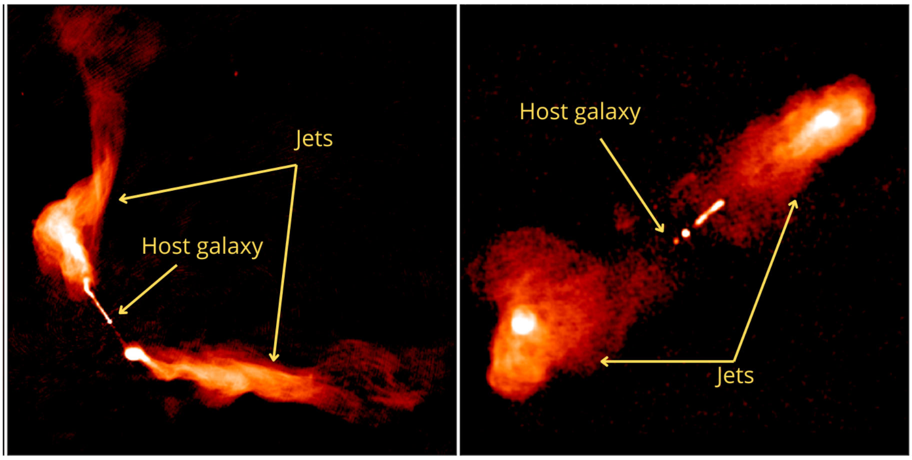

Figure I: Examples of a core-brightened jet (*left*) and edge-brightened jet (*right*) corresponding to their host galaxies. Source: Hardcastle & Croston (2020).

Radio-emitting galaxies

A major early discovery in radio astronomy was the identification of bright radio-emitting sources initially called 'quasi-stellar'. This name was given to these objects because their counterparts, detected with optical telescopes, appeared star-like. These were later commonly known as 'quasars'. Measurements of quasar distances revealed that their radio emissions originated from distant galaxies. It is now understood that their powerful radio emissions result from large amounts of material being drawn toward a supermassive black hole at the galaxy's centre. This material forms a hot, rotating disk, which radiates energy in particular at wavelengths associated with hot ultra-violet (UV) and X-ray radiation.

The radio-emitting galaxies exhibit in many cases jet-like structures emerging from their cores, as indicated in Figure I. These jets can extend far beyond their host galaxies, injecting vast amounts of energy into the surrounding mediums. This energy can disrupt the formation of new stars and so prevent the growth of galaxies. Jetted radio-emitting galaxies are classified by their morphological appearances into Fanaroff-Riley type I (core-brightened) and type II (edge-brightened). Core-brightened jets are characterised by having the brightest parts of their jet near the host galaxy while diffusing out into more plume-like structures towards at the outer edges. The edge-brightened jets exhibit the brightest parts (also called hotspots) of their jet closer to the outer edges. Figure I presents examples of both types. This difference in appearance is attributed to a combination of the power of their jets and the environments in which they reside: Core-brightened jets lack the power to penetrate through their (dense) surrounding environment, causing their jets to diffuse outward, while the more powerful edge-brightened jets remain strong enough to push through their environment, which keeps their hotspots at the outer parts of their jets. Consequently, galaxies with core-brightened jets are typically found in dense environments such as galaxy clusters, while galaxies with edge-brightened jets are more often located in isolated regions.

Although our understanding of these radio-emitting galaxies has improved over time, the complete picture to explain the formation and evolution of their jets is still not fully understood. This is partly due to observational limitations in resolution and sensitivity, which introduced biases that have led to several revisions on the origins of their jets. This highlights the need for more comprehensive data, including larger samples of radio-emitting galaxies observed with more sensitive and high-resolution telescopes.

Radio bridges

Groups of galaxies are called galaxy clusters. These large-scale cosmic structures can consist of hundreds to thousands of galaxies bound together by gravity, making them the largest gravitationally bound objects in the Universe. They often contain not just galaxies, but also vast amounts of hot gas, dark matter, and other intergalactic material. An important process in the evolution of our Universe is the merging of galaxy clusters to form larger structures. The merging processes generate large-scale faint radio emission that can be detected by our telescopes. The source of this emission can be attributed to several physical processes, one of which involves shock-like behaviour, while the other involves randomised turbulence in the presence of magnetic fields.

More recently, faint radio emission between galaxy clusters that are about to merge (in a pre-merging phase) has been detected as well at very low radio frequencies. An example of such radio emission bridging a pair of pre-merging clusters is presented in Figure II. While some have proposed that shock-like behaviour is responsible for their origin, more recent work suggests that the emission may arise from turbulence. However, since only a few direct observations from two confirmed radio bridges have been analyzed, further studies are required to fully understand their physical origin and prevalence.

Interferometers and the Low Frequency Array

Multiple telescopes observing across the radio spectrum have been developed over the last 100 years. Among the most famous are the Very Large Array (VLA),

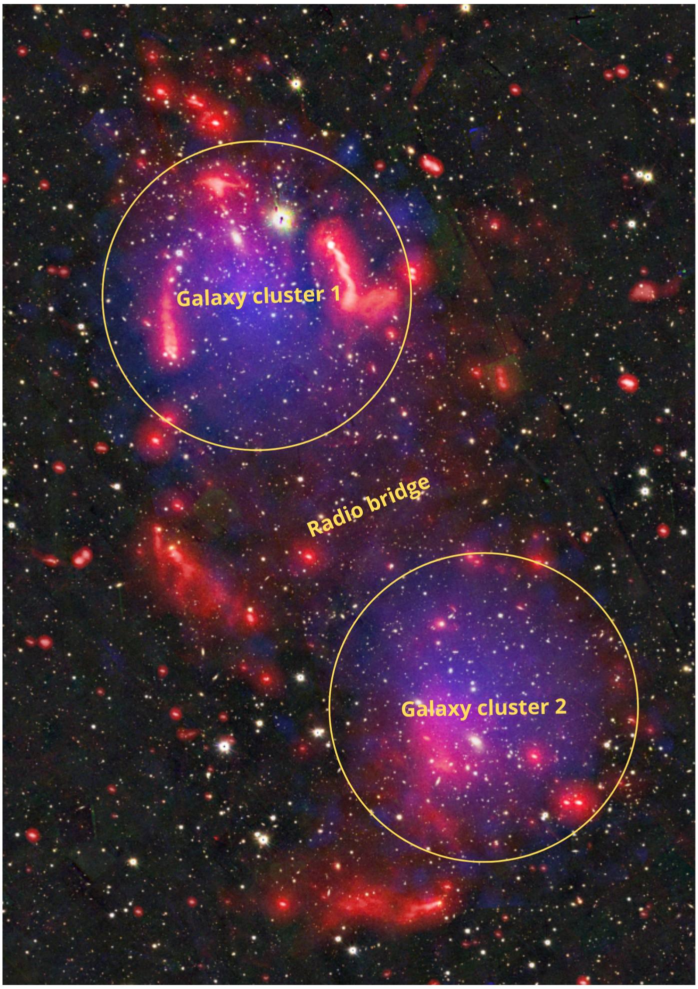

Figure II: Radio bridge in between a pair of pre-merging galaxy clusters. The red colour corresponds to radio emission, the blue colour corresponds to X-ray emission, and the background is optical data. This image is designed by F. Sweijen, using data from Chapter 2.

the Westerbork Synthesis Radio Telescope (WSRT), and the Atacama Cosmology Telescope (ALMA). These are large instruments that make use of interferometry to improve resolution. This is a technique where multiple antennas or dishes are working together to observe the same object. Combining the observed signals from different locations on Earth, allows astronomers to simulate a much larger telescope to study objects at much finer resolutions. While an instrument such as the VLA gave us high-quality radio data of our Universe since the 1980s, there was no data available at comparable quality at wavelengths larger than one meter (below about

Figure III: The LOFAR Superterp is located in Exloo, Drenthe, in the Netherlands. The word ‘terp’ comes from the Dutch language and refers to an artificially elevated area created by early settlers in the Netherlands, particularly in the northern regions, to protect their homes and livestock from flooding in low-lying areas. In the context of LOFAR, the Superterp is a larger, more concentrated LOFAR station located within the core area of the network, built on elevated land and surrounded by a canal to protect it from wildlife discovering this tool to study the radio universe as well. The large grey tiles cover the high-band antennas, while the smaller, more randomly distributed squares represent the low-band antennas. Source: ASTRON.

300 MHz). This frequency regime conceals valuable information about astrophysical processes related to diffuse radio sources from the previously mentioned jets and radio bridges. Fortunately, due to technological advancements and growing interest (and the availability of financial resources), this has since the 2000s changed and several interferometers observing at low radio frequencies have been developed. One of the main instruments observing below 250 MHz is the Low-Frequency Array (LOFAR).

LOFAR is an instrument equipped with two types of antennas: low-band antennas that detect frequencies between 10 and 80 MHz, and high-band antennas that detect frequencies from 110 to 240 MHz. A LOFAR station and the two antenna types are depicted in the images in Figures III and IV. None of LOFAR’s components, including its antennas, have moving parts, as it uses technology to digitally steer its focus across the sky. This means that all the data processing happens within computers, making the construction of its antennas relatively inexpensive compared to other radio telescopes. The low construction costs and international

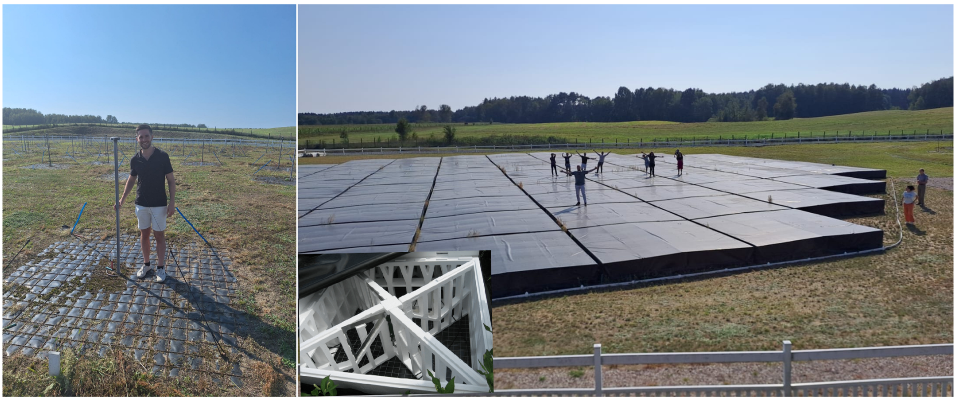

Figure IV: Different LOFAR antennas. On the left, a low-band antenna is shown with a human for scale. In the middle below, an image of a high-band antenna is displayed, which is concealed beneath the tiles on the right, with humans standing on top for scale. These pictures were taken at the PL612 station in Bałdy (Poland) by the author and his colleagues.

collaborations have contributed to the rapid expansion of LOFAR with stations across Europe. LOFAR's core is located in the Netherlands, with international stations in Germany, Poland, England, Ireland, Sweden, Latvia, France, and Finland.[16] Soon, LOFAR will also be expanded to Italy and Bulgaria. By utilizing all of these stations, LOFAR effectively simulates a radio telescope the size of the European continent, offering exceptionally high resolution in addition to its impressive sensitivity due to its dense array of antennas.

Powerful computers and advanced software

LOFAR can capture images of large areas of the sky, up to tens of square degrees in a single observation,[17] allowing it to detect many radio-emitting sources simultaneously. However, creating these large images poses significant computational challenges, particularly when combining data from all of LOFAR's international stations. The computational complexity arises from the need to correct various effects that distort low-frequency observations, one of the most significant being the ionosphere – a layer of Earth's atmosphere filled with charged particles that interfere with propagating radio waves. Correcting for these distortions requires a carefully designed strategy, and once calibration is complete, the data must be transformed into images – a process known as 'imaging' – to produce data ready for scientific

---

[16] The Finnish station in Kilpisjärvi operates independently and is not connected to the LOFAR network.

[17] For comparison: the diameter of the moon covers roughly 0.5 square degrees

analysis. All these steps need specialised software.

Over the past decade, most of the software development efforts have focused solely on processing data from the Dutch LOFAR stations, as this is computationally less demanding while still yielding valuable insights into the Universe. The next major step is to extend this capability to automatically process data from all LOFAR stations. This will enable us to conduct a full survey of the northern sky at a much higher resolution (see Figure V) and to target specific regions with very long exposure times by combining multiple observations of the same area, without the need for manual intervention to produce high-quality science-ready images. Each of these observations produces about 16 terabytes (TB) of data.[18] Processing these amounts of data cannot be done on standard laptops or desktops. Instead, it requires the use of large high-performance computing clusters designed to handle the immense computational load efficiently (see Figure VI).

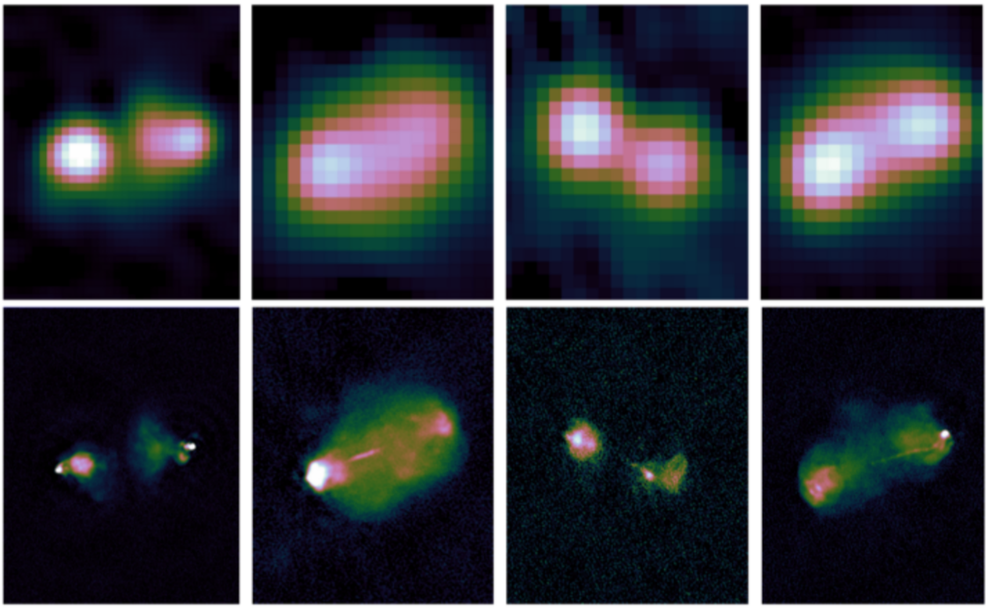

Figure V: Resolution improvement when going from Dutch-only LOFAR station imaging (first row) to imaging with the full international LOFAR (second row). Columns correspond to the same radio source. Source: de Jong et al. (2024).

---

[18] 16 TB is enough storage to hold around 4 million photos from a modern smartphone, 8,000 hours of HD video, or over 100 million pages of text documents. An average laptop has on average about 1 to 2 TB.

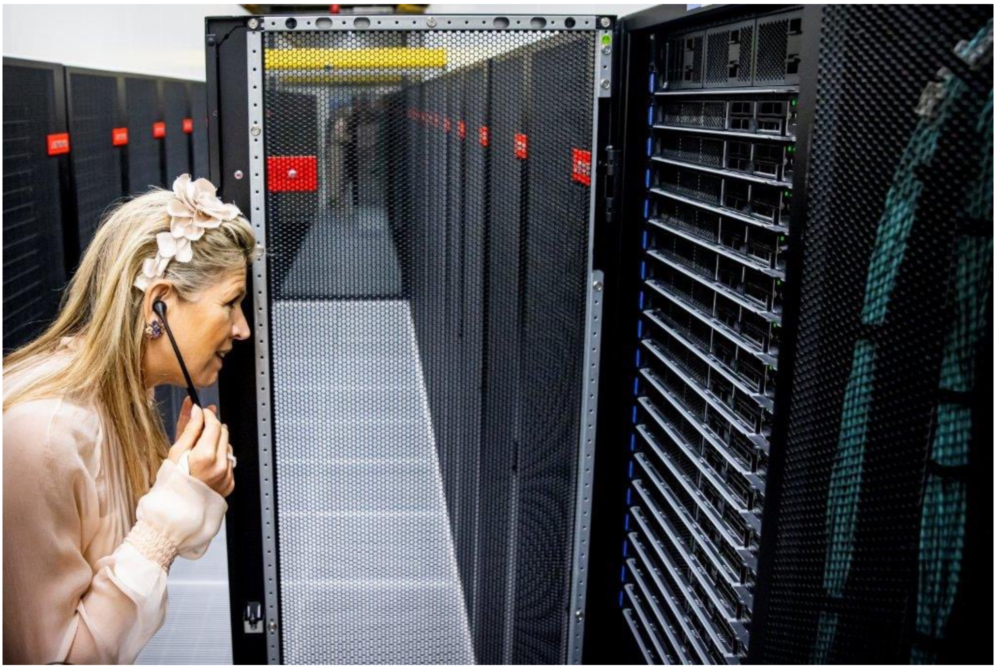

Figure VI: Queen Máxima of the Netherlands inaugurating the Snellius supercomputer in Amsterdam. This supercomputer is part of the SURF computing infrastructure. Using the GPUs of this system, we trained a neural network to detect when the calibration of our radio data has converged. Source: Het Koninklijk Huis.

This thesis

We begin this thesis in **Chapter 2** at the larger scales of our Universe, where we combine six observations from the Dutch LOFAR stations to create the deepest image to date of a radio bridge between a pair of galaxy clusters. These clusters are in the process of merging into a single galaxy cluster. At their current pre-merging stage, they exhibit radio emission that forms a bridge between them. This unique phenomenon was only recently discovered, and our work aims to follow up on it to better understand its origin, using our new deep image. Our findings suggest a turbulent re-acceleration process, where particles gain energy through interactions with turbulence in the presence of magnetic fields, causing them to emit radiation at radio frequencies.

We then step to smaller scales by gathering known core-brightened (FRI) and edge-brightened (FRII) radio-emitting galaxies with powerful jets detected by LOFAR's Dutch stations, along with their cosmic distances measured by other telescopes. This allows us to investigate in **Chapter 3** how these two types of jets have evolved over cosmic time. We find that the earlier Universe was more densely populated by the brightest sources of both morphologies, likely due to the greater availability of gas in the past when the Universe was also smaller. However, we also observe a decline in the number density of fainter sources, a result that is not easily explained. We suggest that this could be due to the resolution limits of our data, motivating the need for higher-resolution images of our sky to better understand this trend.

Driven by the need for higher-resolution images, we focus in **Chapter 4** on processing data from four observations of a single sky area, taken with all Dutch and international LOFAR stations. Combining multiple observations enables the creation of a deeper image which is needed to enable the detection of more distant and faint radio sources.[19] However, this also presents challenges, as it provides us with four times more data to process. To streamline the data reduction strategy, we develop techniques to automatically select areas of the sky with sufficient signal to optimise our data processing. This automatic selection is crucial, as stronger signals provide more information about how the sky changes over time and frequency due to distortions, allowing us to correct better for these with specialised software. After completing all data processing, we achieved the deepest image over a 2.5 by 2.5 degrees sky area with a 20-fold improvement in resolution compared to making images with only the Dutch LOFAR stations (see Figure V), meaning that our image pixels cover 400 times smaller area than those in the data used in Chapter 3.[20]

Although we achieved in the previous chapter the production of the deepest and high-resolution image with LOFAR to date, our work revealed that the computational costs are too high to scale this approach effectively (and responsibly, given the large energy consumption). To tackle this issue, we revisit in **Chapter 5** a technique to reduce data sizes when combining multiple observations of the same sky area. This method makes use of the sidereal day – the period of approximately 23 hours and 56 minutes during which the Earth completes one full rotation around its axis. This allows us to average data samples from our observations, corresponding to a similar sidereal timestamp, such that we reduce the data volume without significant loss in information. In this way, our method enables much faster processing when converting the calibrated data into a final sky image.

Finally, in **Chapter 6**, we present and discuss additional advancements to improve the processing of LOFAR's vast data volumes. This includes quality improvements, implementing a framework for automatic processing on high-performance

---

[19] Similar to setting a longer exposure time on your camera to catch more light during darker evenings.

[20] This also means that the final image, with its size of 32 GB, is too large to be opened on a standard laptop.

computing systems, and utilizing artificial intelligence (AI) to automatically determine when calibration is sufficiently refined, eliminating the need for human intervention. Moreover, we optimise the most computationally intensive aspects of our workflow, making the entire process up to approximately four times faster. These improvements will ultimately allow us to process much larger and many more datasets in a more efficient and cost-effective manner.