



Universiteit
Leiden
The Netherlands

Dancing with the stars

Albert, J.G.

Citation

Albert, J. G. (2020, October 28). *Dancing with the stars*. Retrieved from <https://hdl.handle.net/1887/137988>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/137988>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/137988> holds various files of this Leiden University dissertation.

Author: Albert, J.G.

Title: Dancing with the stars

Issue Date: 2020-10-28



INNOVATION

Astronomy has been around for millennia in various forms, from the prehistoric nomadic peoples scratching the phases of the Moon into animal bones in 36,000 BC to present-day globally envisioned experiments being carried out in Space. The outer cosmos undoubtedly remains the greatest fascination of all Humankind, so much so that we spend entire lives and immeasurable resources trying to know its past, present, and future. It is the search for understanding of ourselves that we truly seek; a pursuit of understanding of existence and reason. It is a worldwide effort, perpetrated by various forces throughout history from the shamanic traditions to the religious and now by science, for no other reason than Human curiosity. Indeed, we find many reasons why astronomy is of great indirect value, for instance, due to spin-off of technology like the internet, but no astronomer ever expects their discoveries will be bottled up and sold for anything other than human pleasure. And, in that sense it is of ultimate value, and explains why it has a special role in the world.

By coincidence we were born with consciousness in an era where the Universe is still observable. By observable, we mean the electromagnetic waves our eyes perceive – the visible spectrum – happens to correspond to a night sky that has striking noticeable features. To anyone who has seen it, we cannot imagine the night sky without this constellation of twinkling bright spots. However, it is conceivable that had we been born much later the expansion of the Universe would have redshifted all star light out of human perception, and we would never have had the impulse to be curious about the cosmos.

Radio emission mechanisms

It was only several decades ago, after looking up with our eyes for 2.5 million years that homo sapiens first thought to look in radio frequency, defined here as having a wavelength longer than a centimetre or frequency less than 30 GHz. The radio sky is completely different from the sky in any other frequency range, as the physical processes that emit radio frequency radiation trace different astrophysical structures. In radio astronomy the physical processes that produce the emission can be classified based on the spectral nature of the emission. Discrete emission mechanisms give rise to spectral lines and are associated with atomic and molecular energy

transitions, while continuum emission mechanisms give rise to a broad band spectra.

In the realm of discrete spectra emission mechanisms we enumerate two main types, and their birth in radio astronomy. The first is the 21 cm emission from the hyperfine transition in neutral hydrogen (HI), which was discovered as a ‘hiss’ coming from the interstellar medium by Ewen and Purcell [1951] and corroborated by Muller and Oort [1951]. Neutral hydrogen consists of a proton and electron, both of which have a quantum angular momentum spin. Correspondingly, each has a quantised magnetic dipole moment which is proportional to the charge times spin. When the angular momentum spins are aligned then the magnetic dipole moments are anti-align – due to the opposite charges – and the energy of the HI molecule is higher than if the magnetic dipoles were aligned. Occasionally, with a decay rate of 2.9×10^{-15} seconds, an HI molecule with anti-aligned magnetic dipoles will spontaneously undergo a spin-flip to the lower energy level, and emit a photon at approximately 1.4 GHz. Due to the vanishingly small decay rate, in order to see the occurrence of 21 cm emission there must be an extremely large volume of HI. This makes it an ideal candidate to study the cosmic dark ages during the few hundred million years that followed the Big Bang, in which there is abundant HI and little star formation [e.g. Patil et al., 2017].

Another important branch of discrete emission mechanism radio astronomy is radio recombination lines (RRLs). A RRL is the emission associated with the transition of an ion from higher to lower principle quantum numbers, $n \rightarrow n - N$, where N is the number of energy levels transitioned. By far the highest probability transitions are of the form $n \rightarrow n - 1$. Furthermore these transitions of principle quantum numbers are orders of magnitudes more likely than fine and hyperfine transitions such as the one involved in 21 cm HI emission. They were first hypothesised to be observable in interstellar ionised hydrogen (HII) by Kardashev [1959], and have since been observed in galactic HII regions over a range of frequencies 36.5 GHz ($56 \rightarrow 55$) to 404 MHz ($254 \rightarrow 253$). In Ball et al. [1970] the interpretation of RRLs suggested that molecules other than HII were needed to fully explain their data, and since has lead to a rich exploration of the galaxy’s composition in the heavier nuclei [Brocklehurst and Seaton, 1972]. Recently, Emig et al. [2019] have observed the first RRLs at cosmological distances, opening a new field for RRLs in studying distant galactic nuclei.

Continuum spectra radio emission mechanisms form a broad class of non-discrete physical process which generally are to do with moving charges. Here we enumerate two classes, and the physical processes that generate them. The first is emission related to deceleration of charged particles during a scattering event. The typical scenario is an electron scattering off of a proton ion, where kinetic energy is converted into emission. This free-free scattering – so called because the electron begins and ends unbound – is a thermal process when driven by the internal energy of the medium. It can be shown [e.g. Rybicki and Lightman, 1986] that the thermal free-free emission flux density from optically thin systems has very little frequency dependence ($S \propto \nu^{-0.1}$), whereas for optically thick systems there is a steep dependence ($S \propto \nu^2$). Thus, thermal free-free emission can shed light on the density of plasma environments.

The second class of continuum emission mechanisms is called synchrotron emission named after the first observation in 1947 by Robert Langmuir and Herb Pollock at the General Electric synchrotron accelerator. Synchrotron emission is any emission related to acceleration of charged particles perpendicular to their velocity, i.e. $|\mathbf{a} \times \mathbf{v}| \neq 0$, and has a continuous spectrum spanning all frequencies. In laboratory experiments, synchrotron radiation is produced by forcing electrons to move in circular orbits using a magnetic field, and is seen as

a nuisance source of energy loss.

The history of astrophysical synchrotron emission is more complicated as noted by Breus [2001],

In particular, the Russian physicist V.L. Ginzburg broke his relationships with I.S. Shklovsky and did not speak with him for 18 years. In the West, Thomas Gold and Sir Fred Hoyle were in dispute with H. Alfvén and N. Herlofson, while K.O. Kiepenheuer and G. Hutchinson were ignored by them.

While the prediction of astrophysical synchrotron *may* be attributed to Shklovsky [1953], the first observation seems to be by Burbidge [1956] of Messier 87, whose jet is shown in Figure 1.

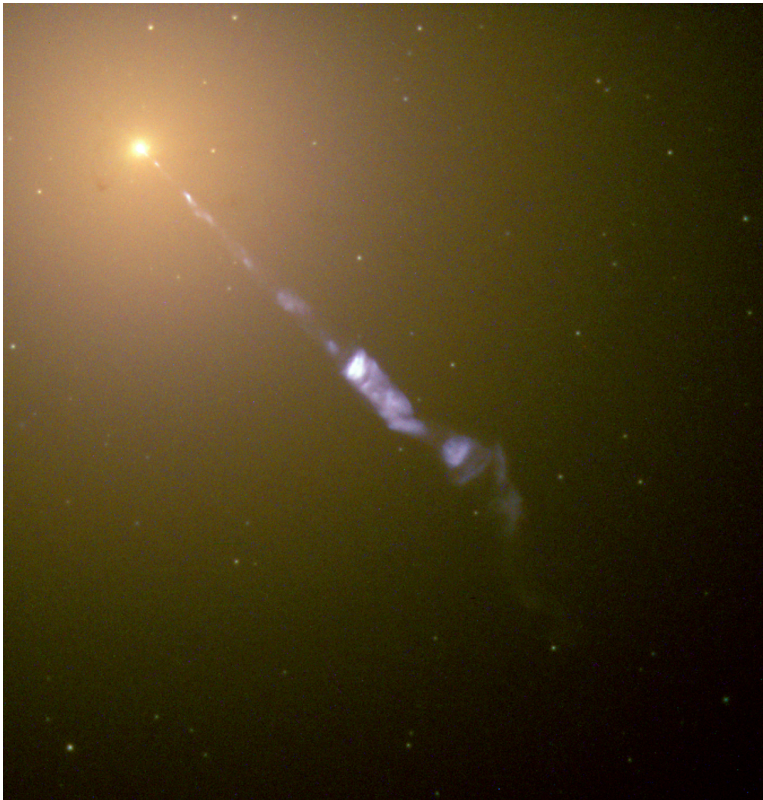


Figure 1: Messier 87 and the radio jet produced by synchrotron. The jet was the first observed astrophysical synchrotron emission.

Astrophysical radio-frequency synchrotron emission comes from charged particles – typically electrons – moving through a magnetic field, where Lorentz force causes acceleration perpendicular to the velocity. This causes particles to trace out circular spirals as they move forward with a gyro-frequency $\nu_e \propto B_{\perp}$, where B_{\perp} is the strength of the magnetic field perpendicular to the velocity of the particle. For $B_{\perp} = 1 \mu\text{Gauss}$ the gyro-frequency of the

electron is approximately 3 Hz. For particles at relativistic speeds, the frequency of emitted photons is boosted by two factors of the Lorentz factor γ due to the relativistic Doppler effect and length contraction. Thus, in the relativistic regime¹, the frequency of maximal emission is then related to the gyro-frequency by $\gamma^2 \nu_e$. For an ultra-relativistic electron with energy of 10 GeV, the Lorentz factor is about 1.96×10^4 and the frequency of maximal radiation is 1.14 GHz.

There are several types of physical processes that can boost an electron to nearly the speed of light. One is a jet stemming from rapidly spinning neutron stars and super massive black holes forming the basis of pulsars [Hewish et al., 1968] and active galactic nuclei [Ambartsumian, 1956, Osterbrock, 1989, Begelman et al., 1984], respectively. Another is diffusive shock acceleration [DSA; Drury, 1983, Blandford and Eichler, 1987], a mechanism by which a charged particle is repeatedly reflected between magnetic ‘walls’ with increasing velocity, which plays an important role in understanding merging galaxy clusters, super nova shocks, and solar flares.

Diffuse emission in galaxy clusters

The cosmic web is the structure of the matter, magnetic field, and kinetic energy on the largest scales of the observable Universe [e.g. van de Weygaert and Schaap, 2009]. The web-like structure evolved from very homogeneous Gaussian primordial conditions and is the perfect example of how complexity can emerge from simple rules and non-complex initial conditions [Vazza, 2020]. Supported by cosmological simulations, the current modern perspective on assembly of the cosmic web is hierarchical in nature [Aragón-Calvo et al., 2010]. Large clusters of galaxies formed at slight over-densities in the matter distribution, from which tenuous filamentary structures extended for many Mpc. Smaller structures formed in the filaments and were drawn toward the clusters, with velocities exceeding 1000 km s^{-1} . Due to hierarchical merging of clusters with accretion from the filaments, galaxy clusters grow to be very massive with masses on the order of $10^{14-15} M_\odot$, and the mass of nearby filamentary structures can be equally massive [Metzler et al., 2001].

Occasionally, galaxy clusters merge converting 10^{56} J of gravitational energy into kinetic energy through a variety of processes including DSA of electrons to ultra relativistic energies in the intra cluster medium (ICM) as well as turbulence. The resulting ultra relativistic cosmic ray electrons (CRE) give off diffuse synchrotron emission in the presence of an ICM magnetic field [for a review see van Weeren et al., 2019]. Diffusive shock acceleration can accelerate non-relativistic thermal populations of CRE, or reaccelerate older aged populations of CRE. The synchrotron emission from the first type is referred to as a radio relic [e.g. Ensslin et al., 1998, van Weeren et al., 2010], while that of the latter is referred to as radio fossil emission [e.g. de Gasperin et al., 2017, Mandal et al., 2020]. There are a number of physical parameters that can be inferred from observations of DSA driven synchrotron emission, including information on the CRE population and ICM magnetic field [Brüggen and Vazza, 2020]. One of the important factors is the speed of the shock which is ultimately coupled to the dynamics of the merger, and premerger initial conditions. Another is the acceleration efficiency which is a phenomenological number describing how much energy is transferred

¹There are non-relativistic astrophysical sources of synchrotron, which, instead of depending on Lorentz boosting, rely on extremely large magnetic fields. This is known as cyclotron emission and we pay it no further attention here.

to an electron diffusing through the shock front once.

In order to perform inference with non-thermal synchrotron emission – whether from radio halos, relics, or fossils – an electron ageing model must be chosen, which describes how the spectral index of synchrotron emission from a population of ultra-relativistic CRe evolves over time. Two standard models exist, colloquially known as the Jaffe and Perola [JP; Jaffe and Perola, 1973] model and the Kardashev and Pacholczyk [KP; Kardashev, 1962, Pacholczyk, 1970] model. The key difference between the models is that the JP model assumes an anisotropic pitch-angle distribution of the CRe, and the KP model assumes an isotropic pitch-angle distribution of the CRe. The initial formulation of both the JP and KP models assumed that energy was injected into the CRe suddenly at single moment in time after which the electrons age in isolation. Then in Komissarov and Gubanov [1994] both models were extended to incorporate a finite period of continuous energy injection and the effect of energy loss due to inverse Compton scattering.

Another important aspect is the magnetic field of the ICM. Galaxy clusters have magnetic fields on the order of $1 \mu\text{G}$ [e.g. Roettiger et al., 1999], however the magnetic field is highly dependent on the dynamical state of the cluster [Bonafede et al., 2010]. Galaxy clusters can have many dynamical states ranging from disturbed to relaxed [Cassano et al., 2010b], which is characterised by the X-ray morphology of the intra cluster medium (ICM) [Poole et al., 2006, Maughan et al., 2008, Santos et al., 2008]. Difusive shock acceleration driven emission is highly polarised, reflecting the fact that synchrotron emission emissivity is proportional to $|\mathbf{a} \times \mathbf{v}|$ and that the Lorentz force produces acceleration perpendicular to the direction of velocity. Radio relics are primarily observed near the periphery of the galaxy cluster, therefore they probe magnetic fields far from the densest part of the ICM [e.g. Wittor et al., 2019].

Turbulence is another mechanism for kinetic energy dissipation into the ICM following a merger, accounting for approximately 20% – 30% of the thermal energy in merging clusters. Thermal populations of electrons also produce emission due to thermal free-free scattering. In contrast with DSA driven synchrotron emission, turbulence driven emission should have zero average polarisation. Such emission is referred to as a radio halo. Coupled with simulations they can still provide constraints on the magnetic field structure, however due to the dispersive nature of turbulence only statistical studies are possible [Vazza et al., 2011]. **Figure 2** shows an example slice of temperature from a galaxy cluster merger simulation with turbulence.

Radio interferometry

History

One thing that becomes clear is that the radio sky is rich in information on a wide variety of physical processes, so how was radio astronomy born? This account can be found in a number of sources [e.g. Southworth, 1956, Sullivan, 2009]. The answer goes back to 1839 when Carl Gauss postulated that a conducting layer at atmospheric heights should account for the variations in a compass reading of magnetic north – an effect that navigators of the time well knew. This mysterious electric current had supporters throughout the 1800’s, such as Lord Kelvin. Around sixty years later, in 1901, Guglielmo Marconi transmitted a radio signal across the Atlantic ocean, which should only have been possible if some reflecting layer bounced the radio signal around Earth’s curvature. An electric current surrounding

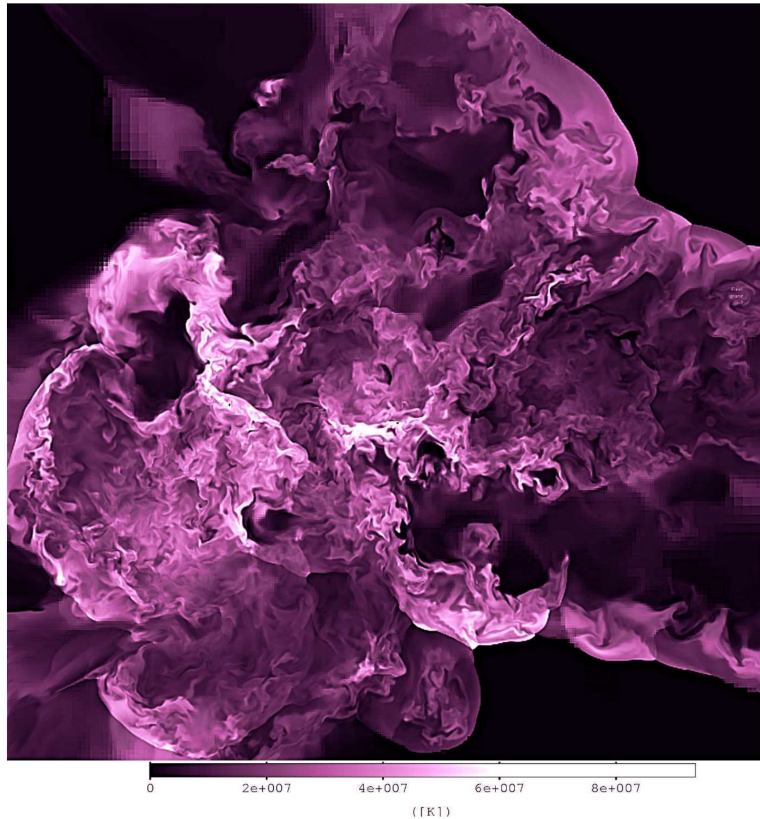


Figure 2: 2-dimensional slice showing the gas temperature for the innermost region of a simulated galaxy cluster, during its main merger event ($z = 0.6$). The side of the slice is $8.8 \text{ Mpc } h^{-1}$ and the depth along the line of sight is $25 \text{ kpc } h^{-1}$. Taken from [Vazza et al., 2011].

Earth seemed like a plausible explanation. For the next three decades astronomers assumed that, while there should be radio emission from Space, it would be bounced back by the same electric layer around the Earth, just as Marconi's signal had been bounced back down. There were few attempts to look for extraterrestrial radio emission, and the development of trans-Atlantic radio communications flourished.

In 1928 a young 23 year old Karl Jansky, newly having attained his Bachelors in physics, was hired by Bell Labs. He was stationed at their Holmdel, New Jersey site and tasked with identifying sources of atmospheric radio interference in telecommunications. Karl built a 100-ft in diameter steerable antenna, which he called the 'merry-go-round' for its construction included four sets of Ford model-T tires on which it could be rotated azimuthally. Over the course of months Karl measured the various components of radio noise with his rig as a function of time of day and azimuth. He identified two types of static-like noise related to thunderstorms, which were well-known at the time to cause radio interference: one due to local thunderstorms, and the second due to distant thunderstorms reflecting off of the ionosphere. In August of 1931, such thunderstorms and associated interference were

common occurrences of night-time, however as Karl continued irregular monitoring into the autumn and winter he began to notice a third hiss-like component to the interference, which he referred to as a 'very steady continuous interference... that changes direction continuously throughout the day, going completely around the compass in 24 hours'. The direction of this component seemed entirely aligned with the Sun causing him to refer to the hiss as 'Sun static'. Despite equipment damages due to ferocious winter weather, he was able to monitor into the spring at which point he noticed the hiss direction appeared to drifted in time 'in accordance with the lengthening day'. This further convinced him that the hiss was somehow related to the vernal Northerly swings of the Sun.

The idea that the source of interference might be sidereal, fixed to the stars, came some time in December of 1932. By this time the reality of the Great Depression had hit Bell Labs, and working hours had been severely cut back limiting his time available to study the hiss. Karl had instead spent most of 1932 studying general methods of measurement of noisy signals, still taking daily measurements but not deeply analysing them. He had noticed that the signal continued to advance in time relative to the Sun, and during a partial solar eclipse in August Karl noted that the hiss intensity had little dependence on the solar occlusion. Despite these observations, Karl must have still held the notion that the hiss was still somehow related to the Sun. For it wasn't until December 1932 that George C. Southworth, a correspondent working for AT&T in New York, asked Karl to plot his entire years' data to see if it correlated with his study of 'diurnal changes in Earth currents'. Upon performing this task, Karl, not an astronomer, was shocked and puzzled to find that the signal had slipped by exactly one day. To an astronomer such a pattern would be obvious, for the stars rise four minutes earlier each day due to the Earth's rotation about the Sun, adding up to one day over a year. Then Karl's good friend and bridge partner, Albert Melvin Skellett, helped make the realisation that the period of signal was remarkably the same as the sidereal day. (Albert lead a dual life, being both a radio engineer for Bell Labs, and a graduate student of astronomy at the near-by Princeton University.) This lead Karl to learn about astronomical coordinates, and by Christmas 1932 he had determined the signal was fixed in space with a direction of 18 hours right ascension, and -4 degrees declination,

Since I was home [early November] I have taken more data which indicates definitely that the stuff, whatever it is, comes from something not only extraterrestrial, but from outside the solar system. It comes from a direction that is fixed in space and the surprising thing is that... [it] is the direction towards which the solar system is moving in space. According to Skellett, there are clouds of 'cosmic dust' in that direction through which the Earth travels.

Karl had fixated on a region of sky in the Sagittarius constellation [Jansky, 1933], which we now know to be the galactic centre. Karl Jansky was subsequently ordered by Bell Labs back to more pressing tasks. The fundamental unit of flux in radio astronomy is named after Karl, the jansky (Jy), $1 \text{ Jy} = 10^{-26} \text{ W Hz}^{-1} \text{ m}^{-2}$. The electric layer that surrounds Earth is the ionosphere, which is the central topic of this thesis.

Following World War II, and the debut of the atomic bomb, there was a great desire to make progress in radio technology. Radio astronomy gained much momentum in the tail winds of radar technology, and indeed there was a great deal of shared knowledge between the two fields. While we can only surmise, radio technology flourished in large part for better methods of surveillance and communication, and perhaps ultimately for the prevention

of another man-made global catastrophe. The renewed interest in radio technologies was a great boon to radio astronomy which led to radio interferometry being born [Ryle and Vonberg, 1946]. In the ‘West’, radio research institutes sprang up in Europe, especially in Britain and the Netherlands, the USA at the National Radio Astronomy Observatory (NRAO), and Australia at Commonwealth Scientific and Industrial Research Organisation (CSIRO). The Iron Curtain, separating ‘East and West’, provided an – arguably *ideal* – opportunity for independent progress on many topics of mathematics, physics, and engineering [Pesch and Plail, 2009, Gillespie et al., 1977], however in radio astronomy the result was quite the opposite. Just like in the ‘West’, radar and radio astronomy flourished in post-war USSR. However, unlike their counterparts in Europe, USA, and Australia, who primarily worked out of universities, all USSR radio astronomy was conducted at military institutes and was heavily censored. Most discoveries were never published in academic journals, and those that did get published were often missing key data [Braude et al., 2012]. This policy of secrecy ultimately led to a lack of trust by their contemporaries, and hence today there are very few references to early-day radio astronomy out of the ‘East’.

Eventually, radio astronomy spread around the world to places like India, South America, Canada, South Africa, and China. Today, there are global collaborations undertaking some of the most massive, in terms of data, human experiments, such as the Low-Frequency Array (LOFAR), with even bigger experiments in the near future such as the Square Kilo-Metre Array (SKA). Aperture synthesis radio interferometry, which will shortly be described, naturally breeds unification among countries; for the need to cover long baselines – in order to resolve the most distant objects in the Universe – necessitates an international community. The irony is profound as radio interferometry, which sprang from war, promotes peace.

We’ll now proceed with explanation of interferometry from a modern perspective in a way that is foundational to Chapters 2–4.

Imagine yourself lying on the bottom of a swimming pool, and looking up through the water. You are the radio telescope and the water is the Earth’s atmosphere and ionosphere. Now imagine that there’s a ceiling up overhead, with faint frescoes that you’re trying to make out through the water. The frescoes are the faint objects of the universe. Then, somebody goes and puts a thousand bright lights on the ceiling (the radio-bright objects of the universe) – you’re now trying to make out the frescoes behind the lights’ glare. And then the wind picks up so the water gets choppy.

OLEG SMIRNOV

Technical introduction to scattering and radio interferometry

Most of the rich information in astrophysical radio emission requires good angular resolution, which a single antenna is unable to provide. Interferometry is a way of inferring the shape and location of a source by measuring the interference pattern of light which has simultaneously travelled two different optical pathways. Interference is a property of waves and is completely described by Maxwell’s equations of electromagnetism, thus we start there. We present the

theory of scattering as it relates to the ionosphere and the observables of radio interferometry, which sets the stage for Chapters 2 to 4. We treat the topic with slightly more detail than a typical introduction to the ionosphere and radio interferometry, though we skip rigour where appropriate for flow. The aim is that this might also serve as a summary reference for future work on the topic.

The ionosphere is a low-density multispecies ion plasma situated between 50km and 1000km from Earth's surface [for a detailed description see Kivelson and Russell, 1995]. The plasma precipitation is fed by extreme ultraviolet radiation from the Sun, and the bulk motions are driven by a complex network of forces. The ionosphere is likely a turbulent medium [Mevisius et al., 2016], and displays many discrete behaviours [Jordan et al., 2017]. The spatially and temporally varying free electron density (FED) in the plasma alters the local dielectric properties of ionosphere, and this leads to a refractive index less than unity that decreases linearly with FED and wavelength of the radiation to second order. The refractive index is given by the Lassen-Appleton-Hartree equation [Hutchinson, 2005]. The ionosphere thus acts as a random scattering medium, and its main effects on the image quality are a reduction of dynamic range, distortion of the shapes of sources, and decoherence of the visibilities. We present a description of the effect of the ionosphere from first principles.

To begin, consider monochromatic electric and magnetic fields of the form $\mathbf{E}(\mathbf{r}, t) = e^{-i\omega t}\mathbf{E}(\mathbf{r})$ and $\mathbf{B}(\mathbf{r}, t) = e^{-i\omega t}\mathbf{B}(\mathbf{r})$ respectively. Going forward we'll drop the spatial functional argument for clarity. Using Fourier representation we can build any electric or magnetic field from various monochromatic components and the following derivation thus applies to generic fields. Recall that in an isotropic linear medium that the electric field is related to the electric displacement field by $\mathbf{E} = \epsilon^{-1}\mathbf{D}$ and the magnetic field is related to the magnetic flux density by $\mathbf{B} = \mu\mathbf{H}$, where ϵ and μ are the permittivity and permeability respectively. We assume that both ϵ and μ do not change over the timescale of propagation, and that μ is also constant over space. The FED in the ionosphere induces a spatially² varying permittivity ϵ but does not alter permeability μ . Furthermore, we assume there are no sources of charge or current, i.e. that macroscopically the ionosphere is charge neutral on scales comparable to the wavelength. Maxwell's equations of electromagnetism are then,

$$\nabla \cdot (\epsilon\mathbf{E}) = 0 \tag{1}$$

$$\nabla \cdot (\mu\mathbf{H}) = 0 \tag{2}$$

$$\nabla \times \mathbf{E} = i\omega\mu\mathbf{H} \tag{3}$$

$$\nabla \times \mathbf{H} = -i\omega\epsilon\mathbf{E}, \tag{4}$$

where we've used the fact that all temporal dependence is contained in the unitary factor $e^{-i\omega t}$. We now define the refractive index as $n^2 \triangleq c^2\mu\epsilon$, and the wavenumber as $k = (2\pi)n/\lambda$ where λ is the wavelength in vacuum. In general, the wavenumber changes spatially due to the refractive index. Taking the curl of Eq. 3, inserting Eq. 4, and using a vector calculus identity we get,

$$\nabla^2\mathbf{E} + k^2\mathbf{E} = \nabla(\nabla \cdot \mathbf{E}). \tag{5}$$

By making use of Eq. 1, the right-hand side is found to be equivalent to $\nabla(\nabla \cdot \mathbf{E}) = -\nabla(\mathbf{E} \cdot \nabla \log \epsilon_r)$, and is zero if $\mathbf{E} \perp \nabla \log \epsilon_r$ or $\log \epsilon_r = 0$ everywhere, where $\epsilon_r = \epsilon/\epsilon_0$ is the relative

²The FED is assumed to be constant in time over the timescale of propagation of light through the medium.

permittivity. These conditions are true if the wave is normal incident on a scattering surface, or if there is no scatterer. More generally, we can neglect $\nabla(\nabla \cdot \mathbf{E})$ if the length scale over which ϵ_r changes is much larger than the wavelength. In the ionosphere, the irregularities vary over kilometres which is thousands of times larger than the wavelength. Since we assume $\nabla(\nabla \cdot \mathbf{E}) = 0$ we can apply the linear operator to each electric field component and treat the electric field as a scalar. We therefore drop the vector notation for the electric field. In this case Eq. 5 turns into,

$$(\nabla^2 + k_m^2)E = -k_m^2 \Phi E, \quad (6)$$

where $k_m = \frac{2\pi}{\lambda}$ is the constant wavenumber of the background medium which we take to be vacuum, and $\Phi = (n^2 - 1)$ is the scattering potential.

Green's function. The general solution to Eq. 6 can be written in terms of the Green's function for the Helmholtz equation, which is a function satisfying $(\nabla^2 + k_m^2)g(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}')$. Since the coefficients on the left side are coordinate-independent the Green's function is translation-invariant, thus Green's function can be found by taking the Fourier transform³ and solving,

$$\tilde{g}(\mathbf{s}) = \frac{1}{4\pi^2 s^2 - k_m^2} \quad (7)$$

$$\implies g(\mathbf{r}, \mathbf{r}') = \frac{e^{ik_m|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|}. \quad (8)$$

We have skipped a large amount of rigour in deriving Eqs. 7 and 8, notably we have not even mentioned generalised functions. For a very clear account of all assumptions including the Sommerfeld radiation condition see Schmalz et al. [2010]. Note, it is possible to derive the Green's function including the $\nabla(\nabla \cdot \mathbf{E})$ term [Krüger et al., 2017]. The additional terms in this complete Green's function are negligible when $k_m^2 r^2 \gg O(25)$, which is where the condition to drop the $\nabla(\nabla \cdot \mathbf{E})$ term comes from. Both the Green's function and its Fourier transform will be useful in the following.

The general solution to Eq. 6 is then found by [Evans et al., 1999],

$$E = k_m^2 \int_V g \Phi E \, dV + \oint_S (g \nabla E - E \nabla g) \cdot dA \quad (9)$$

Equation 9 is an implicit solution since the solution appears also in the right-hand side, therefore further assumptions are required to explicitly solve. Importantly, in order for the solution to be defined we must stipulate boundary conditions on a surface S enclosing our volume V .

Interestingly, George Green (ca. 1793 – 1841) was at least several decades ahead of his time in formulating his theory on the solution to differential equations, despite having no formal education in – or indeed any real exposure to – mathematics [Challis and Sheard, 2003]. He was not even really known about until years after his death, such that this day no

³Note that in this work we use the unitary definition of the Fourier transform with frequency variable in units of cycles.

one knows what he looked like. He grew up as the son of a miller near Nottingham, which was rural farmland in the early 19th Century, and had little in the way of education. It would have been hard to find access to the then cutting-edge mathematics that were happening in places like Cambridge, and Paris. At the time Nottingham was mostly farmland, and the only resources he had access to were a local library where there certainly was nothing in the way of advanced mathematics to be found. The only possible connections to non-banal mathematics would have been through a local teacher who had studied mathematics before becoming a teacher. Nonetheless, George Green's works are the foundation upon which all equations of interest in physics are solved, such as Maxwell's equations, the elastic wave equation, the Helmholtz equation, the Schrödinger equation, the Dirac equation, the viscous fluid equation, the Klein-Gordon equation, and Einstein's equations to name but a few.

There are a number of ways to solve Eq. 9 which differ in the functional basis taken and in the assumed conditions on the scattering potential Φ . In this work we discuss the Born and Rytov approximations whose first-order approximations are equivalent [e.g. Born and Wolf, 1999, Blackledge, 2005]. As will become apparent, the two approximations differ in the space where solutions are additive – in the sense of linear differential operators – and therefore each has its own advantages in interpretation.

Born approximation. In the Born approximation we assume additivity in the complex wave function, i.e. we solve the linear equation Eq. 9 given boundary conditions for E . We choose the boundary conditions to be a super position of temporally incoherent plane waves on the celestial sphere,

$$E_i(\mathbf{r}) = \int_{\mathbb{S}^2} a(\hat{\mathbf{s}}) e^{ik_m \hat{\mathbf{s}} \cdot \mathbf{r}} d\hat{\mathbf{s}} \quad (10)$$

where $|a(\hat{\mathbf{s}})|^2$ is the specific intensity, or surface brightness, from the direction $\hat{\mathbf{s}}$ on the celestial sphere, in units Jysr^{-1} . It gives the power per area per frequency per solid angle. Clearly, E_i satisfies the homogeneous Helmholtz equation $(\nabla^2 + k_m^2)E_i = 0$. By temporally incoherent we mean that each $a(\hat{\mathbf{s}})$ has a unitary factor with random phase, η , of the form $e^{i\eta t}$ such that $\langle e^{i2\pi(\eta_1 - \eta_2)t} \rangle = \delta(\eta_1 - \eta_2)$ where averaging is over a time much longer than λ/c [Mandel and Wolf, 1965]. Using Green's theorem the second integral in Eq. 9 simplifies to E_i [Evans et al., 1999] and the Born solution becomes,

$$E^{\text{Born}} = k_m^2 \int_V g \Phi E dV + E_i \quad (11)$$

$$\triangleq E_s + E_i. \quad (12)$$

Equation 11 is still an implicit solution, however a series representation known as the Born series is available,

$$E_{n+1}^{\text{Born}} = E_i + k_m^2 \int_V g \Phi E_n^{\text{Born}} dV, \quad (13)$$

with $E_0^{\text{Born}} = E_i$. The series defined by Eq. 13 converges if the maximal absolute eigenvalue of the matrix representation – the spectral radius – of the convolution $g \star \Phi$ is less than unity [Osnabrugge et al., 2016]. Equivalently, $k_m^2 \left| \int_V g \Phi dV \right| < 1$ everywhere in V [Blackledge, 2005].

In the ionosphere this is almost always true, though periods of intense solar activity can decrease the refractive index which increases the absolute value of the scattering potential. The assumptions of radio interferometry are violated during these periods and the data should be thrown away.

Going forward we will assume the first-order Born approximation, $E^{\text{Born}} \approx E_1^{\text{Born}}$, which has a convenient representation via Fourier transforms. Since Green's function, Eq. 8, is translation-invariant the first-order Born approximation can be written as a convolution, $E^{\text{Born}} = E_i + k_m^2 (g \star \Phi E_i)$. We apply the convolution theorem and write the first-order Born approximation as,

$$E^{\text{Born}}(\mathbf{r}) = E_i(\mathbf{r}) + \underbrace{k_m^2 \int_{\mathbb{S}^2} e^{i2\pi\mathbf{s}\cdot\mathbf{r}} \tilde{g}(\mathbf{s}) \int_{\mathbb{S}^2} a(\hat{\mathbf{s}}) \tilde{\Phi}(\mathbf{s} - \lambda^{-1}\hat{\mathbf{s}}) d\hat{\mathbf{s}} d\mathbf{s}}_{E_s(\mathbf{r})}, \quad (14)$$

where $\tilde{\Phi}$ is the 3D Fourier transform of the scattering potential. Equation 14 is the scattering equation first written down by Max Born in the 1920s. It implies that if we can measure the scattered wave, $E_s = E^{\text{Born}} - E_i$, then this allows us to reconstruct the scattering potential by solving an inverse problem. Such problems are called inverse scattering problems and are some of the hardest problems in mathematics and physics. Indeed, Richard Feynman's creation in the 1960s of Feynman diagrams is all about solving scattering problems, and is also based on Green's functions.

The Born approximation has a nice connection to the plane wave spectrum of Fourier optics [Born and Wolf, 1999], which will be useful for us. Let the coordinate frame be Cartesian $\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}$ with $\hat{\mathbf{z}}$ pointing towards the observer, and the origin fixed to the centroid of the scattering potential. In this coordinate system we parametrise the plane wave directions with $\hat{\mathbf{s}} = (l, m, n \triangleq \sqrt{1-l^2-m^2})$ with $l^2 + m^2 < 1$. We will use notation that subscript 2D denotes the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ components of the vector, e.g. $\hat{\mathbf{s}}_{2D} = l\hat{\mathbf{x}} + m\hat{\mathbf{y}}$. Consider the 2D Fourier transform of Eq. 14 in a plane spanned by $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$. Since the incoming wave, E_i , is a superposition of plane waves, its 2D Fourier transform can easily be derived using the definition of the Fourier transform of the Dirac delta, and is,

$$\tilde{E}_i(\mathbf{s}_{2D}; z) = \int_{\mathbb{S}^2} a(\hat{\mathbf{s}}) \delta(\mathbf{s}_{2D} - \lambda^{-1}\hat{\mathbf{s}}_{2D}) e^{ik_m n z} d\hat{\mathbf{s}}. \quad (15)$$

The 2D Fourier transform of the scattered wave term, E_s , requires a bit more work. From Eq. 14 we can read off the 3D Fourier transform of E_s ,

$$\tilde{E}_s(\mathbf{s}) = \frac{k_m^2}{4\pi^2 s^2 - k_m^2} \int_{\mathbb{S}^2} a(\hat{\mathbf{s}}) \tilde{\Phi}(\mathbf{s} - \lambda^{-1}\hat{\mathbf{s}}) d\hat{\mathbf{s}}, \quad (16)$$

where we substituted the Green's function Fourier transform from Eq. 7. The 2D Fourier transform is then found by inverse transforming s_z back to the spatial domain. In order to do this we use the residue theorem to integrate around the singularity at $s_z^0 = +\sqrt{\frac{k_m^2}{4\pi^2} - s_x^2 - s_y^2}$ using the partial fraction expansion $\frac{1}{4\pi^2 s^2 - k_m^2} = \frac{1}{2s_z^0} \left(\frac{1}{s_z - s_z^0} - \frac{1}{s_z + s_z^0} \right)$,

$$\tilde{E}_s(\mathbf{s}_{2D}; z) = ik_m^2 \frac{e^{i2\pi s_z^0 z}}{4\pi s_z^0} \int_{\mathbb{S}^2} a(\hat{\mathbf{s}}) \tilde{\Phi}(\mathbf{s}^0 - \lambda^{-1}\hat{\mathbf{s}}) d\hat{\mathbf{s}}, \quad (17)$$

where $\mathbf{s}^0 = (s_x, s_y, s_z^0)$. Note this requires that $\tilde{\Phi}$ contains no singularities, which is true of the ionosphere.

Equations 15 and 17 are exact for the first-order Born approximation. Let us examine them qualitatively. Both equations contain exponential terms of a similar form which encode the propagation of the plane wave spectrum along $\hat{\mathbf{z}}$. Notice that in the plane wave spectrum of the scattered wave, Eq. 17, for directions of propagation where $s_x^2 + s_y^2 > \frac{k_m^2}{4\pi^2}$ the wave decays with z . Such waves are called evanescent waves and are not detectable by radio interferometers.

Equation 17 is a weighted sum of the incoming wave amplitudes, $a(\hat{\mathbf{s}})$, with the scattering potential spectrum, $\tilde{\Phi}$, evaluated on a sphere of radius λ^{-1} with origin offset by \mathbf{s}^0 . For the purpose of illustration, imagine the scattering potential is an isotropic Gaussian blob with a characteristic length scale l . This imposes a characteristic length scale in the Fourier domain of $\sigma = (2\pi l)^{-1}$ beyond which the spectrum amplitude decays to zero, i.e. $\tilde{\Phi}(\mathbf{s}) \approx 0$ when $|\mathbf{s}| \gg \sigma$. This implies that observers looking in direction $\hat{\mathbf{s}}^0$ will only see scattered emission from celestial sources with directions, $\hat{\mathbf{s}}$, satisfying $\arccos(\hat{\mathbf{s}}^0 \cdot \hat{\mathbf{s}}) \lesssim \frac{\lambda}{2\pi l}$. The smallest ionospheric structures often range between 1–5 km with sub-km structures occasionally occurring [Yeh and Swenson, 1959, Yeh, 1962, Mevius et al., 2016]. At a radio frequency of 150 MHz this implies light from a single source is scattered into a cone of approximately $1.1[l/\text{km}]^{-1}$. This implies that diffraction effects are weak.

A useful property of an optical system is whether or not it is isoplanatic [Fellgett and Linfoot, 1955]. In general, when the electric field plane wave spectrum (Eqs. 15 and 17) at the detector surface can be sufficiently approximated by a planar convolution of the plane wave spectrum on the object surface (\mathbb{S}^2 in our case), then the optical system is called isoplanatic. We note that both Eqs. 15 and 17 are not planar convolutions but rather convolutions over the 2-sphere. Therefore, in general, an astronomical radio interferometer is not an isoplanatic optical system. In radio interferometry there are two sources of non-isoplanaticity: one from the curvature of the celestial sphere, and one from scattering. This can be seen by interpreting the convolution kernels in Eqs. 15 and 17 as the sources of non-isoplanaticity. Various methods exist in radio interferometry for handling non-isoplanaticity due to curvature of the celestial sphere in the case of non-coplanar arrays [e.g. Cornwell and Perley, 1992, Cotton et al., 2004, Cornwell et al., 2008, van Weeren et al., 2016, Tasse et al., 2018]. The methods introduced in Chapters 2 and 3, and applied in Chapter 4, are about handling non-isoplanaticity due to scattering. In general, non-isoplanaticity can be handled by dividing the object surface into isoplanatic patches [Fellgett and Linfoot, 1955].

In radio interferometry it is not the electric field that is measured but rather the electric field coherence, $V(\mathbf{r}, \mathbf{r}') = \langle E(\mathbf{r}, t)E^*(\mathbf{r}', t) \rangle$, where $\langle \cdot \rangle$ is averaging over a time period much longer than λ/c . For simplicity, we assume a perfect receptive sensitivity, i.e. the radio antennas measure an unbiased electric field⁴. Using the inverse 2D Fourier transform of the plane wave spectrum the electric field coherence is given by,

$$V(\mathbf{r}, \mathbf{r}') = \int \langle (\tilde{E}_i(\mathbf{s}_{2D}; z) + \tilde{E}_s(\mathbf{s}_{2D}; z))(\tilde{E}_i^*(\mathbf{s}'_{2D}; z') + \tilde{E}_s^*(\mathbf{s}'_{2D}; z')) \rangle e^{i2\pi(\mathbf{s}_{2D} \cdot \mathbf{r}_{2D} - \mathbf{s}'_{2D} \cdot \mathbf{r}'_{2D})} d\mathbf{s}_{2D} d\mathbf{s}'_{2D} \quad (18)$$

⁴Incorporating a receptive sensitivity is simply a matter of multiplying the plane wave spectrum by the directional sensitivity gain of the antenna.

The standard equation for electric field coherence seen in radio interferometry is stationary [e.g. Perley et al., 1989], i.e. depends only on the separation between two positions, however the assumptions arriving there are often absent. In general, the electric field coherence is non-stationary [e.g. Mandel and Wolf, 1965, Garcia-Sucerquia et al., 2002], i.e. depends on two positions, and is only truly stationary for incoherent emission satisfying the homogeneous Helmholtz equation. This is true of the celestial radiation, E_i , but not true of the scattered wave, E_s . The scattered wave gains spatial coherence similar to how spatially incoherent monochromatic radiation passing through two closely separated pin holes gains spatial coherence and produces an interference pattern that is not translation-invariant. However, the electric field coherence can be approximated as stationary using the same observation noted earlier; that the sparse incoherent celestial emission is scattered into a narrow cone of diameter approximately l' , so that two scattered waves interfere only if their wave vectors are nearly perfectly aligned. Under this assumption, using Eqs. 15 and 17, Eq. 18 can be worked out,

$$V(\mathbf{r}-\mathbf{r}')=k_m^2 \int_{\mathbb{S}^2} \frac{e^{i2\pi\lambda^{-1}\hat{\mathbf{s}}\cdot(\mathbf{r}-\mathbf{r}')}}{4s_z^2} \left(\int_{\mathbb{S}^2} |\tilde{\Phi}(\lambda^{-1}\hat{\mathbf{s}}-\lambda^{-1}\hat{\mathbf{s}}')|^2 \langle |a(\hat{\mathbf{s}}')|^2 \rangle d\hat{\mathbf{s}}' \right) d\hat{\mathbf{s}} + \int_{\mathbb{S}^2} e^{i2\pi\lambda^{-1}\hat{\mathbf{s}}\cdot(\mathbf{r}-\mathbf{r}')} \langle |a(\hat{\mathbf{s}})|^2 \rangle d\hat{\mathbf{s}}. \quad (19)$$

Note that due to the stationary assumption there are no cross-coherence terms, i.e. $\langle \tilde{E}_i(\mathbf{s}_{2D}; z) \tilde{E}_s^*(\mathbf{s}_{2D}; z') \rangle + \langle \tilde{E}_s(\mathbf{s}_{2D}; z) \tilde{E}_i^*(\mathbf{s}_{2D}; z') \rangle = 0$, which can be verified directly. Equation 19 comprises inverse Fourier transforms on the 2-sphere, which are in the class of restricted Fourier transforms [e.g. Barcelo et al., 1997, Tao, 2003]. Efficient methods of computing (inverse) Fourier transforms and convolutions on the 2-sphere exist [e.g. Driscoll and Healy, 1994, Jarosz, 2008]. The use of such methods result in an exact treatment of non-isoplanatic systems [Eastwood et al., 2018], yet still remain relatively unexplored in radio interferometry. Equation 19 is an oscillatory integral, so another way to exactly treat non-isoplanaticity is via oscillatory integral methods such as the method of stationary phase. Such methods are completely unexplored in radio interferometry.

For incoherent radiation the isoplanatic assumption is equivalent to the condition that the specific intensity in the detector surface can be written as a planar convolution of the specific intensity on the object surface. The kernel of the convolution in this case is called the point spread function (PSF). Again, there are two sources of non-isoplanaticity due to celestial sphere curvature, and scattering. For a coplanar array (detector surface of constant z) an optical system with incoherent radiation becomes isoplanatic. However, recall that scattering introduces coherence, and therefore a system with scattering is not isoplanatic even for coplanar arrays.

For illustrative purposes, assume the special case of a coplanar array $z = z'$, and let the electric field coherence be measured everywhere in the plane of the interferometer. We observe that the second term in Eq. 19 reduces to a 2D Fourier transform relation between the electric field coherence and the 2D power spectrum of the incoming electric field⁵, which we call the specific intensity, sky-brightness, or image, in units of Jy sr^{-1} . This is known as the Wiener-Khinchin theorem. Observe that the first term does not simplify for a coplanar

⁵This technically requires changing the domain of integration to \mathbb{R}^2 and defining the integrands as zero outside the unit disk.

array due to the s_z^2 in the denominator. However, if we additionally add the assumption that the celestial emission is restricted to a narrow field of view, such that $s_z^2 \approx 1$, then we can also apply the Wiener-Khinchin theorem. In this case, we see that there are two images; one corresponding to the celestial emission and one to scattered emission respectively,

$$\mathcal{I}_i(\hat{\mathbf{s}}) = \langle |a(\hat{\mathbf{s}})|^2 \rangle \quad (20)$$

$$\mathcal{I}_s(\hat{\mathbf{s}}) = \frac{k_m^2}{4s_z^2} \int_{\mathbb{S}^2} |\tilde{\Phi}(\lambda^{-1}\hat{\mathbf{s}} - \lambda^{-1}\hat{\mathbf{s}}')|^2 \langle |a(\hat{\mathbf{s}}')|^2 \rangle d\hat{\mathbf{s}}'. \quad (21)$$

The interpretation of Eq. 21 can be thought of with a toy model. Imagine the ionosphere is a flat layer perpendicular to zenith with a sinusoidal wave of refractive index with a spatial frequency, \mathbf{s}_{ion} , and let there be one source at zenith. Then, clearly $\tilde{\Phi}(\mathbf{s})$ will have a delta peak at $\pm\mathbf{s}_{\text{ion}}$. Now, Eq. 21, says that a fraction of the source flux will be shifted symmetrically by $\pm\mathbf{s}_{\text{ion}}$. This would look like two speckles offset from the source at zenith. Therefore, the intuition behind Eq. 21 is that speckles are symmetrically displaced radially from sources according to the magnitude and direction of strong spatial modes in the ionosphere. Over the course of an observation the integral effect of scattering would be perceived as a halo of speckles, which grows brighter proportional to the specific intensity of the source. Since the aperture of an Earth rotation synthesis radio interferometer [Högbom, 1974] is sparsely covered, a displaced speckle would appear as a displaced instantaneous PSF [Born and Wolf, 1999]. Another characteristic comes from the property of temporal continuity in the ionosphere. Since wind speed in the ionosphere can be several hundreds of metres per second [Phillips, 1952] we expect temporal continuity of the FED and the speckles should move continuously. Therefore, during periods of high ionosphere activity the speckles should ‘draw’ radial spokes emanating from bright sources, e.g. as seen in Figure 3.3 in Chapter 3.

Finally, observe that Eq. 21 is a relation between the 3D scattering potential spectrum and the 2D scattered intensity. The relation is invertible given knowledge of the celestial sky intensity, $\langle |a(\hat{\mathbf{s}})|^2 \rangle$, a measured scattered intensity $\mathcal{I}_s(\hat{\mathbf{s}})$, and a prior over the scattering power spectrum. Given the above restriction on diffraction to a cone of approximately $1'$, without priors the relation is only invertible if $\langle |a(\hat{\mathbf{s}})|^2 \rangle$ is dense with respect to this separation. The fact that the complete information of a 3D object can be stored in a 2D object is the principle of holography, first introduced by Dennis Gabor for microscopic wavefront reconstruction that bypasses the diffraction limit of a lens [Gabor, 1949].

Incoherent light is radiation of which we know only certain data of a statistical nature: the energy or intensity, the polarization, the spectral composition. If we interrupt an optical tract at an arbitrary plane, we see only an almost-uniform blur. It must strike one as something of a miracle that a lens can sort out the intermingling wavelets and restore a picture of an object – at least in one plane.

DENNIS GABOR, 1969

Rytov approximation. The Rytov approximation gives a different interpretation of the effects of scattering in terms of what are called *gains*. Gains form the basis of most radio interferometry calibration procedures, and are closely related to Chapters 2 to 4.

The Rytov approximation makes the eikonal transformation, $E^{\text{Rytov}}(\mathbf{r}) = E_0(\mathbf{r})e^{\phi(\mathbf{r})}$, where E_0 satisfies the homogeneous Helmholtz equation $(\nabla^2 + k_m^2)E_0 = 0$ and $(\nabla E_0)/E_0 = \nabla\phi_0$, for some function ϕ_0 [Born and Wolf, 1999]. We choose $E_0(\mathbf{r}, \hat{\mathbf{s}}) = a(\hat{\mathbf{s}})e^{ik_m\hat{\mathbf{s}}\cdot\mathbf{r}}$ which clearly satisfies the homogeneous Helmholtz equation and $(\nabla E_0)/E_0 = \nabla(ik_m\hat{\mathbf{s}}\cdot\mathbf{r})$. If we substitute the eikonal transformation into Eq. 9 then we get a non-linear Riccatian equation for ϕ . If $|\nabla\phi\cdot\nabla\phi| \ll |k_m^2\Phi|$ everywhere in V , then the non-linear Riccatian equation can also be solved by making the substitution $w = \phi/E_0$ yielding [Blackledge, 2005],

$$\phi = \frac{k_m^2}{E_0} \int_V g\Phi E_0 dV. \quad (22)$$

In the Rytov approximation we choose additivity – in the sense of linear operators – in phase rather than electric field. In other words, by making the eikonal transformation, we get a linear differential equation for phase, and not electric field as in the Born approximation. Nonetheless, the principle of super position of electric fields applies and the scattered electric field is given by the difference between the solution and the incoming electric field, $E_s^{\text{Rytov}} = E_0(e^\phi - 1)$. We note that Eq. 22 looks similar to the scattered wave in the first-order Born approximation,

$$E_s^{\text{Born}} = E_0\phi, \quad (23)$$

when we make the choice $E_i = E_0$. Equation 23 in combination with Eq. 14 implies that $E^{\text{Rytov}} = E_0[\exp(E_1^{\text{Born}}/E_0) - 1]$, and when the spectral radius of $g \star \Phi$ is less than unity then, $E^{\text{Rytov}} \approx E_1^{\text{Born}}$. This shows that the two approaches are equivalent in the regime of the typical ionosphere.

Let us now make the connection between the Rytov approximation and classical non-diffractive tomography as introduced by Johann Radon [Radon, 1917]. From the discussion on the Born approximation we already know that diffraction is approximately limited to a cone on the angular scale of $1'$. Given that the average separation between bright celestial sources is greater than this, it makes sense to consider the non-diffractive limit. In this setting we choose to align the coordinate axes with $\hat{\mathbf{z}}$ pointing in the direction of the single plane wave $\hat{\mathbf{s}}$. If we take the 2D Fourier transform of Eq. 23, use the definition of the plane wave spectrum of the Born scattered wave, Eq. 17, and rearrange we get,

$$\tilde{\Phi}(\mathbf{s}^0 - \lambda^{-1}\hat{\mathbf{s}}) = -i\frac{4\pi}{k_m^2} s_z^0 e^{-i2\pi s_z^0 z} \tilde{\phi}(\mathbf{s}_{2D} - \lambda^{-1}\hat{\mathbf{s}}_{2D}; z). \quad (24)$$

where $\mathbf{s}^0 = (s_x, s_y, s_z^0 = \sqrt{\lambda^{-2} - s_x^2 - s_y^2})$ and $\hat{\mathbf{s}} = (0, 0, 1)$. Equation 24 is an exact relation between the Rytov and first-order Born approximations. Let us make the substitution $\mathbf{s}' = \mathbf{s}^0 - \lambda^{-1}\hat{\mathbf{s}}$. Now, from our discussion in the Born approximation we have that the ionosphere is only weakly diffracting and therefore $|\mathbf{s}'|^2 = 2\lambda^{-2}(1 - \hat{\mathbf{s}}^0 \cdot \hat{\mathbf{s}}) \ll \lambda^{-2}$ from which we have that $s_z^0 \approx \lambda^{-1}$ and $s_z' \approx 0$. Equation 24 becomes,

$$\tilde{\Phi}(\mathbf{s}') \approx -i\frac{4\pi}{\lambda k_m^2} \tilde{\phi}(\mathbf{s}'_{2D}; z). \quad (25)$$

If we take the inverse Fourier transform of both sides, using that $s_z' \approx 0$, and rearrange we get

the classic non-diffractive tomography relation,

$$\int e^{-i2\pi s'_z z} \Phi(\mathbf{r}) dz = -i \frac{4\pi}{\lambda k_m^2} \phi(\mathbf{r}) \quad (26)$$

$$\implies \phi(\mathbf{r}) \approx i \frac{\lambda k_m^2}{4\pi} \int \Phi(\mathbf{r}) dz. \quad (27)$$

This is precisely the equation that is foundational to Chapter 2. Another way to arrive at this result is via the Jeffreys-Wentzel-Kramers-Brillouin (JWKB) approximation [e.g. Jeffreys, 1925]. In the JWKB approximation one makes the extra assumption on the eikonal phase, that $\nabla^2 \phi$ can also be neglected, which results in the geometric optics regime. The eikonal representation in the JWKB approximation gives rise to Jones and Müller algebra treatment of ray optics [Jones, 1941, Parke, 1949].

Given Eq. 27 the Rytov solution for a super position of incoming incoherent plane waves, Eq. 10, is given by,

$$E^{\text{Rytov}}(\mathbf{r}) = \int G(\hat{\mathbf{s}}, \mathbf{r}) a(\hat{\mathbf{s}}) e^{ik_m \hat{\mathbf{s}} \cdot \mathbf{r}} d\hat{\mathbf{s}}, \quad (28)$$

where we define the *gains*, $G(\hat{\mathbf{s}}, \mathbf{r}) = \exp\left[i \frac{\lambda k_m^2}{4\pi} \int_{\hat{\mathbf{r}}=\hat{\mathbf{s}}-1} \Phi(\mathbf{r}) dV(\mathbf{r})\right]$. This implies that the electric field coherence is given by,

$$V(\mathbf{r}, \mathbf{r}') = \int G(\hat{\mathbf{s}}, \mathbf{r}) G^*(\hat{\mathbf{s}}, \mathbf{r}') (|a(\hat{\mathbf{s}})|^2) e^{ik_m \hat{\mathbf{s}} \cdot (\mathbf{r} - \mathbf{r}')} d\hat{\mathbf{s}}. \quad (29)$$

Equation 29 is the radio interferometry measurement equation for scalar electric fields and perfect antenna sensitivity [Hamaker et al., 1996]. Importantly, we note that in the non-diffractive limit the gains contain no amplitude information. When the full Eq. 22 is used then the resulting gains would contain amplitude information, and this is possibly an avenue of future research.

In comparing Eq. 29 with Eq. 19 we can clearly see the different approximations lead to different interpretations. For the Born approximation, the effect of the scattering on the image is intuitive in terms of the plane wave spectrum. For the Rytov approximation, the effect on the electric field coherence is intuitive in terms of gains. It should be made clear that both are valid – indeed equivalent – approaches. It is the author’s opinion that there is still much to gain in radio interferometry by better understanding the foundational equations. Often improvements and understanding come from returning to the basics of a theory, e.g. understanding ‘ghosts’ in images due to over-calibration [Grobler et al., 2014].

Next-generation radio interferometry

The next generation of radio interferometry is pushing technology to its limit. In particular, science with the near-future Square Kilometre Array [SKA; Dewdney et al., 2009] at full capacity is not technologically feasible with current techniques. The key parameters of a radio interferometer are its sensitivity, angular and spectral resolution, bandwidth, and field of view. In terms of instrumental attributes these parameters correspond to surface area, baseline coverage, number of channels, and the type of antenna.

In an aperture synthesis radio interferometer [Högbom, 1974], as collecting surface area increases the volume of data grows quadratically as this typically corresponds to adding more antennas, which adds one extra source of data for every pre-existing antenna. The SKA is a near-future radio interferometer that has been envisioned for over two decades [Braun, 1996, van de Weygaert and van Albada, 1996], and aims to provide 1 km^2 of effective collecting area [Hollitt et al., 2017]. The SKA is the "pie in the sky" of radio interferometry, since with current technology it would only be possible to utilise a mere fraction of its capacity for science. The expected data rate of the SKA is tens of petabytes per day, or about 1 Internet – the same volume as the entire internet. This challenge has spawned a huge interest in algorithms [e.g. Lindner et al., 2015, Offringa, 2016, Wijnholds et al., 2018, Cai et al., 2018, 2019, Lyon et al., 2019] and technology [e.g. Hassan et al., 2013, Isaev and Tarasov, 2016] for handling huge data volumes.

The demand for higher angular resolution means longer baselines are required, and antennas become further separated. This raises the challenge of keeping antennas synchronised and transferring massive amounts of data to a common correlator. This ultimately would result in improved internet infrastructure around the world. Another challenge with long baseline arrays is dealing with the curvature of the Earth, which breaks the coplanarity assumption. This challenge was already beginning to be addressed several decades ago [e.g. Cornwell and Perley, 1992], and was largely solved within the last decade [Cornwell et al., 2008]. The divergence of the radio interferometer from the classical optical detector plane description, wherein fast Fourier transforms do everything, also sparked much work in more general treatments of interferometric imaging [e.g. Tasse et al., 2013].

In this thesis we work with two world-class radio interferometers: Low-Frequency Array [LOFAR; van Haarlem et al., 2013], and the Giant Metre-wave Radio Telescope [GMRT; Gupta et al., 2017].

LOFAR is a high-sensitivity, high-resolution, wide-band, wide-field, low-frequency, fiber-optic connected, phased-array aperture synthesis radio interferometer that spans Europe, from Ireland to Latvia, and France to northern Sweden. It has two operating bands, the high-band antennae (HBA; 115–240 MHz) and the low-band antennae (LBA; 10–80 MHz). In this thesis, we make use of the Dutch LOFAR array, with baselines out to 80 km, and an effective collecting area of 0.04 km^2 at 150 MHz. The expected sensitivity of a standard 8.5 hour observation with the full Dutch LOFAR-HBA and 24 subbands is $72.8 \mu\text{Jy beam}^{-1}$. The original primary aim of LOFAR was to survey the entire Northern sky with unprecedented resolution, depth, and fidelity. The science that comes out of LOFAR now spans all radio fields and domains of astronomy including: cosmic dawn [e.g. Patil et al., 2017, Mertens et al., 2020], early star-formation [e.g. Saxena et al., 2018], active galactic nuclei [e.g. Williams et al., 2018], radio transients [e.g. Kuiack et al., 2019], galaxy clusters [e.g. van Weeren et al., 2019, Mandal et al., 2019], radio exo-planets [e.g. Callingham et al., 2020], and more. LOFAR is a pathfinder for SKA-low, and comes with a great number of technological challenges [Noordam, 2004]. In some ways the Murchison Widefield Array [MWA; Tingay et al., 2013] in Australia is a Southern-hemisphere counterpart to LOFAR, acting as a precursor to the SKA-low.

The GMRT is an aperture synthesis radio interferometer composed of 30 massive steerable 45 m parabolic dishes situated just 80 km north of Pune, India. It had three narrow bands (32 MHz wide) situated at 150 MHz, 320 MHz, and 610 MHz. Given the large collecting area of the dishes, approximately 0.03 km^2 , the GMRT was one of the most sensitive radio

interferometers prior to LOFAR, being approximately three times larger than the Very Large Array [VLA; Condon et al., 1998] and equally large as the single dish giant Arecibo Radio Telescope [Heiles et al., 2001]. It has enabled a host of science ranging from cosmic dawn [e.g. Paciga et al., 2013], radio transients [e.g. Bhattacharyya et al., 2016], radio exo-planets [e.g. Hess and Zarka, 2011], galaxy clusters [e.g. Venturi et al., 2008], to stellar radio emission [Quiroga-Nuñez et al., 2020]. The GMRT was recently upgraded to a wideband version [uGMRT; Gupta et al., 2017], however in this thesis we only made use of the GMRT.

About this thesis

In some sense the SKA, being a far-off intangible goal, has provided the motivation to come together as a community and solve the technological challenges of today's telescopes. Radio interferometry has seen a huge boost in its algorithms in the last decade as we prepare to meet the challenges of SKA. One of the many challenges is direction dependent ionospheric calibration. There are several common methods for ionospheric calibration, e.g. field-based calibration [Cotton et al., 2004] which models image domain apparent source shifts, and facet-based calibration [Intema et al., 2009, van Weeren et al., 2016, Tasse et al., 2018] which performs calibration on disjoint patches of the sky called facets. In this thesis we focus entirely on the facet-based approach.

Facet-based calibration has several orders of complexity. The simplest method is to form an independent data set per facet, by subtracting a model of the rest of the sky, and then perform calibration independently per facet. The result solutions can then be spatially filtered to remove noise and to try to make them spatially consistent with some model. This is the basis of Source Peeling and Atmospheric Modelling [SPAM; Intema et al., 2009]. An extension on this is to also enforce a linear phase model during solve, e.g. that the ionospheric phase distortions are $\propto \nu^{-1}$. This makes forms the basis of Facet Calibration for LOFAR [FACTOR; van Weeren et al., 2016]. The more complicated method is to simultaneously and jointly calibrate all facets on all the visibilities [Tasse, 2014b, Smirnov and Tasse, 2015]. Following this post-processing can enforce a linear phase model, e.g. $\propto \nu^{-1}$ dependence for ionospheric terms. This forms the basis of killMS and DDFacet [Tasse et al., 2018]. All facet-based methods make use of self-calibration [Pearson and Readhead, 1984].

Common to all of these methods is that 1) they are limited by the density of infield-calibrators to form facets, and 2) they do not use any spatial information of the ionosphere. One of the big open questions in radio interferometry is: Is it possible to use the physics of a weakly scattering random medium to improve on these methods of ionospheric calibration? In this thesis we answer this question in theory and in practice.

The thesis is split into two parts. In the Chapter 1 we investigate diffuse synchrotron emission in a high redshift galaxy cluster. We use this investigation to gain a working experience of radio interferometry, as well as to explore an interesting field in astronomy, i.e. the study of the largest scale structure in the Universe. In Chapters 2 to 4 we then tackle the problem of probabilistic calibration of the ionosphere using a physical model of weakly scattering random media. The scope of the project is aimed at investigating ways to solve the problem of the ionosphere for LOFAR, as a path finder for SKA.

In Chapter 2 we apply well-known physics of non-diffractive, or non-dispersive, phase distortions to derive a probabilistic tomographic prior for the ionospheric distortions. The

result is a mean and covariance function for a Gaussian process [Rasmussen and Williams, 2006b], with hyper parameters that encode information about the geometry of the ionosphere including height, thickness, and spectral density. We show, using simulated ionospheres, that the resulting Gaussian process surpasses all other common Gaussian processes in modelling ionospheric distortions and is computationally feasible.

Following this, in Chapter 3, we devise a method that can apply this to real interferometric data. Using DDFacet we acquire calibration data which we use as data in our model. There are a number of challenges we face. Firstly, the calibrator data is low signal-to-noise. This complicates extracting ionospheric information from them. We devise a variational hidden Markov model [HMM; e.g. Rabiner and Juang, 1986] which provides a Bayesian estimate of the distribution of the ionospheric distortions, which forms the input to our model. In applying the method to real data we are forced to make several approximations to the model for computational feasibility. Ultimately, we apply it to a single LOFAR observation and find we are able to significantly reduce ionospheric artefacts.

Finally, in Chapter 4 we assess the feasibility of the method on the deep Lockman Hole data set comprising 12 8.5 hour LOFAR observations. We find that the method significantly improves 10/12 data set, and that 2/12 data sets only show subtle improvements. In no cases it the image quality degraded over the current state-of-the-art. We also jointly image all 100 hours of the deep Lockman Hole data set and find that the improvements extend to multi-epoch observations. In the process we uncover a modelling systematic, which we are able to isolate and characterise. We find that there is a direction-dependent non-ionospheric component that limits the resulting noise level around sources to a few times thermal noise. We propose a correction to the method which shows, in principle, that we can account for it, although we leave the extension of the method to a following paper.

Future directions

Given the success of the model at LOFAR-HBA frequencies, a natural extension is to apply it to LOFAR-LBA. This likely comes with a number of challenges. For example, at lower frequencies the number of suitable calibrators will be far fewer because sensitivity over appropriate time scales will be much lower. In Chapter 2 we showed that the full tomographic version of the model can handle LOFAR-LBA conditions. Therefore, a possible future avenue will be making the full tomographic version computationally feasible to handle this sparse and noisy regime.

In this thesis we focus purely on the ionospheric modelling aspect, and do not attempt to alter the way in which calibration and imaging occurs. However, the author recognises the potential for new calibration and imaging techniques that implicitly account for the ionosphere. Currently, there are no methods that take into account the spatial information and physics of the ionosphere when inverting the visibilities. Furthermore, currently the standard description of the ionosphere in radio astronomy excludes the amplitude information of the ionosphere from the model. The model could be extended to incorporate a model for amplitudes, however this would likely require modelling the ionosphere directly on visibilities.

Another possible future direction has to do with the definition of the radio interferometry measurement equation. The current definition [Hamaker et al., 1996] was formed in a time prior to big-data, tomographic ionosphere corrections, and all development of the past two

decades. Currently, the definition stems from the use of Jones algebra which means that we choose additivity in phase space and that wave-fronts never distort. We therefore see the visibilities as a convex sum of complex exponentials weighted by the celestial specific intensity. There are a several side-effects of this including non-linearity of the visibility model, and phase-wrapping of all propagation effects. However, if we were to take the Born approximation, we would naturally gain linearity of the visibility model, which would make least-squares calibration much more well-founded. This could have extensive implications for calibrating the ionosphere phase and amplitude effects directly from visibilities.

