Needle-like structures discovered on positively charged lightning branches

B. M Hare, O. Scholten, J. Dwyer, T. Trinh, S. Buitink, S. ter Veen, A. Bonardi, A. Corstanje, H. Falcke, J. Hörandel, et al.

▶ To cite this version:

B. M Hare, O. Scholten, J. Dwyer, T. Trinh, S. Buitink, et al.. Needle-like structures discovered on positively charged lightning branches. Nature, Nature Publishing Group, 2019, 568 (7752), pp.360-363. 10.1038/s41586-019-1086-6 . insu-02149526

https://doi.org/10.1038/s41586-019-1086-6

Needle-like structures discovered on positively charged lightning branches

B. M. Hare[1]*, O. Scholten[1,2]*, J. Dwyer[3], T. N. G. Trinh[1], S. Buitink[4,5], S. ter Veen[6], A. Bonardi[5], A. Corstanje[5], H. Falcke[2,6,7], J. R. Hörandel[5,7], T. Huege[4,8], P. Mitra[4], K. Mulrey[4], A. Nelles[9,10], J. P. Rachen[5], L. Rossetto[5], P. Schellart[5,11], T. Winchen[4], J. Anderson[12,13], I. M. Avruch[6,14], M. J. Bentum[6,15], R. Blaauw[6], J. W. Broderick[6], W. N. Brouw[6,16], M. Brüggen[17], H. R. Butcher[18], B. Ciardi[19], R. A. Fallows[6], E. de Geus[6,20], S. Duscha[6], J. Eislöffel[21], M. A. Garrett[22,23], J. M. Grießmeier[24,25], A. W. Gunst[6], M. P. van Haarlem[6], J. W. T. Hessels[6,26], M. Hoeft[21], A. J. van der Horst[27], M. Iacobelli[6], L. V. E. Koopmans[16], A. Krankowski[28], P. Maat[6], M. J. Norden[6], H. Paas[29], M. Pandey-Pommier[25,30], V. N. Pandey[6,16], R. Pekal[31], R. Pizzo[6], W. Reich[32], H. Rothkaehl[33], H. J. A. Röttgering[23], A. Rowlinson[6,26], D. J. Schwarz[34], A. Shulevski[26], J. Sluman[6], O. Smirnov[35,36], M. Soida[37], M. Tagger[24], M. C. Toribio[23], A. van Ardenne[6], R. A. M. J. Wijers[26], R. J. van Weeren[23], O. Wucknitz[32], P. Zarka[38] & P. Zucca[6]

---

[1]KVI-Center for Advanced Radiation Technology, University of Groningen, Groningen, The Netherlands. [2]Inter University Institute for High Energies, Vrije Universiteit Brussels, Brussels, Belgium. [3]Department of Physics and Space Science Center (EOS), University of New Hampshire, Durham, NH, USA. [4]Astrophysical Institute, Vrije Universiteit Brussel, Brussels, Belgium. [5]Department of Astrophysics/IMAPP, Radboud University Nijmegen, Nijmegen, The Netherlands. [6]ASTRON, Netherlands Institute for Radio Astronomy, Dwingeloo, The Netherlands. [7]NIKHEF, Science Park Amsterdam, Amsterdam, The Netherlands. [8]Karlsruhe Institute of Technology (KIT), Institute for Nuclear Physics, Karlsruhe, Germany. [9]Institut für Physik, Humboldt-Universität zu Berlin, Berlin, Germany. [10]DESY, Zeuthen, Germany. [11]Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA. [12]Institute of Geodesy and Geoinformation Science, Technical University of Berlin, Berlin, Germany. [13]Department 1, Geodesy GFZ German Research Centre for Geosciences, Potsdam, Germany. [14]Science and Technology, Delft, The Netherlands. [15]Eindhoven University of Technology, Eindhoven, The Netherlands. [16]Kapteyn Astronomical Institute, University of Groningen, Groningen, The Netherlands. [17]University of Hamburg, Hamburg, Germany. [18]Research School of Astronomy and Astrophysics, Australian National University, Canberra, Australian Capital Territory, Australia. [19]Max Planck Institute for Astrophysics, Garching, Germany. [20]SmarterVision BV, Assen, The Netherlands. [21]Thüringer Landessternwarte, Tautenburg, Germany. [22]Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, The University of Manchester, Manchester, UK. [23]Leiden Observatory, Leiden University, Leiden, The Netherlands. [24]LPC2E—Université d'Orleans/CNRS, Orléans, France. [25]Station de Radioastronomie de Nancay, Observatoire de Paris, CNRS/INSU, Université d'Orleans, OSUC, Nancay, France. [26]Anton Pannekoek Institute for Astronomy, University of Amsterdam, Amsterdam, The Netherlands. [27]Department of Physics, The George Washington University, Washington, DC, USA. [28]University of Warmia and Mazury in Olsztyn, Space Radio-Diagnostics Research Centre, Olsztyn, Poland. [29]Center for Information Technology (CIT), University of Groningen, Groningen, The Netherlands. [30]CRAL, Observatoire de Lyon, Université Lyon, UMR5574, Saint Genis Laval, France. [31]Poznan Supercomputing and Networking Center (PCSS), Poznan, Poland. [32]Max-Planck-Institut für Radioastronomie, Bonn, Germany. [33]Space Research Center PAS, Warsaw, Poland. [34]Fakultät für Physik, Universität Bielefeld, Bielefeld, Germany. [35]Department of Physics and Electronics, Rhodes University, Grahamstown, South Africa. [36]SKA South Africa, Pinelands, South Africa. [37]Jagiellonian University, Astronomical Observatory, Krakow, Poland. [38]LESIA & USN, Observatoire de Paris, CNRS, PSL/SU/UPMC/UPD/SPC, Meudon, France. *e-mail: b.h.hare@rug.nl; scholten@kvi.nl

Needle-like structures discovered on positively charged lightning branches: Supplementary Information

1 Method

The lightning flashes used in this work were mapped using data from the LOFAR (LOw Frequency ARray) radio telescope. Due to its effective lightning protection system, LOFAR is able to continue to operate during thunderstorm activity[1]. The Dutch LOFAR stations consist of 38 (24 core + 14 remote) stations spread over 3200 km^2 in the northern Netherlands. The largest baseline between core stations is about 3 km, the largest baseline between remote stations is about 100 km. From each station we use 6 dual-polarized low band dipole antennas (LBA), sampled at 200 MHz, to observe the $30 - 80$ MHz band. The raw time series data were saved to the transient buffer boards, which continuously buffer the last 5 s of data from a maximum of 48 dual-polarized antennas per station. The resulting relative timing accuracy is better than 1 ns. See [2] for more details on LOFAR. When a lightning flash occurs within the area enclosed by the Dutch LOFAR stations, as observed by www.lightningmaps.org, the transient buffer boards are stopped and the data is read to disk.

The method we used to map each lightning flash has three major steps. In the first step we fitted plane-waves to the time of pulses received by individual LOFAR stations. Note that the LOFAR stations are less than 100 m in diameter and the lighting is many kilometers from the closest LOFAR station, so that a plane-wave approximation is very good for individual LOFAR stations. These plane-waves were used to identify non-functional antennas, and the intersection of their arrival directions gave a rough first estimate of the flash location, accurate to a few kilometers. Since each station has its own clock and cable delays, in the second step we found the clock offsets between the different LOFAR stations by simultaneously fitting the locations of multiple events and station clock offsets to the measured times of radio pulses, with a Levenberg-Marquardt minimizer. In order to achieve the highest precision, we chose to fit locations of 5 events that created pulses that were strong but not saturating, had a simple structure, and did not change shape significantly across different stations. After fitting, the root-mean-square difference between the modeled and measured arrival times of the radio pulses was around 1 ns. The resulting station clock offsets are consistent with LOFAR station clock calibrations, which are known for the core stations but not the remote stations.

After finding the station clock offsets, in the third step we mapped the lightning flash using a new interferometric imaging algorithm inspired by those used in astronomy. The goal of this algorithm is to find for each source the 3D location, $\vec{X}$, that corresponds to the maximum value of the image intensity, $I(\vec{X})$, defined by

$$I(\vec{X}) = \frac{1}{N} \sum_{\substack{i,j \\ i\neq j}} \frac{\left|C_{i,j}(\Delta T_{i,j}(\vec{X}))\right|}{\max(|C_{i,j}|)} \,, \tag{S1}$$

where $C_{i,j}(t)$ is the complex-valued cross-correlation between the signals recorded on antennas i and j, $\Delta T_{i,j}(\vec{X})$ is the time difference between a signal on antenna i and a signal on antenna j if those signals were both emitted by a point source at $\vec{X}$, and $\max(|C_{i,j}|)$ is the maximum value of the absolute value of the cross-correlation between antenna i and j. In different mapping stages the sum is performed differently as described in more detail below. N is the length of the sum. The cross correlation is built from the complex valued voltages as discussed later, it is also up-sampled by a factor of 8 and interpolated with a periodic cubic spline in order to achieve sub-sample timing accuracy. The cross-correlation is normalized so that if $\vec{X}$ is exactly equal to the location of the source region, and the source region is a point source, then $\Delta T_{i,j}(\vec{X})$ will correspond to the maximum of the cross-correlation, and the value of the image intensity will be exactly 1. Therefore, if we find the correct source location, then the intensity of the image should be close to 1 with lower values indicating poorer fits. Because of noise and the fact that our sources are not perfect point sources, we never expect the peak of the image intensity to actually be precisely 1. Note that if the signal in an antenna is dominated by noise, then the cross-correlation of that signal with any other signal will be fairly flat, and so will not affect the location of the peak of the image.

The two difficulties are finding the image intensity maximum without forming the entire image, and in choosing the correct signal length to use in the cross-correlation. Our algorithm was specifically designed to solve these problems for imaging lightning. Before imaging, a 3D rectangular bounding box is defined around the lightning flash using the approximate location from the first step and assuming the lightning flash is 5 km wide and 6 km tall [3]. If, after imaging, the flash is less than 500 m from the edge of the bounding box, the box is expanded by 500 m and the flash re-imaged. The imaging is run on data blocks that are 2^{16} samples long (327 μs). The data block for each antenna has a timing offset so that a point source in the middle of our bounding box will produce a pulse in the middle of the data block for every antenna. Since the pulses near the edge of the data block cannot be imaged, as discussed further below, each block has a slight overlap so that every received pulse can be imaged without repetition.

Radio signals from the lightning will sometimes saturate the LOFAR digitizers, therefore, first, saturated data are removed by setting the saturated areas and the following 50 samples to zero. We also include a half-Hann window (length of 50 samples) around these windows that are set to zero. Furthermore, since man-made artificial radio frequency interference (RFI) corrupts parts of the measured spectrum, the data blocks are band-pass filtered between 30 – 80 MHz and RFI lines are removed according the method described in [4]. This filtering also removes the negative frequency components so that the resulting signal is complex valued and analytic. After filtering the raw data, the lightning is imaged in two stages. The goal of the first stage is to find an approximate location for the source inside our bounding box. However, since this stage is designed to search the entire bounding box it is not as precise as possible. Therefore, we follow with a second stage that assumes that the result of the first stage is close to the real source location and attempts to refine the location.

For each block of data we image each pulse in order of strength, with stronger pulses imaged first. Once we know the time of the pulse that we want to image on a central antenna, t_p, we calculate a window, $L_i + t_p$ and $H_i + t_p$, for every other antenna, i, such that we know that the corresponding pulse will arrive between times $L_i + t_p$ and $H_i + t_p$ on antenna i, under the assumption that the source location is within the boundary box. This central antenna is referred to as the reference antenna, and is chosen so that the difference between the minimum L_i and maximum H_i is as small as possible. This leads us to the algorithm of the first stage:

1) Form a window centered on the largest source on the reference antenna. This window is 50 data points long (250 ns) in order to accommodate the typical width of the radio pulse. Note that we ignore pulses that are within the maximum L_i of the beginning of the block and within the maximum H_i of the end of the block in order to avoid imaging sources that may not be seen on all antennas within this data block.

2) On every other antenna i form a window between times L_i+t_p and H_i+t_p.

3) Calculate the cross-correlations, $C_{i,p}(t)$, between the windowed data on the reference antenna and the windowed data on every other antenna. Since the windows are different lengths on every antenna, every window is zero-padded to a length that is the next power of two larger than the largest $H_i - L_i$.

4) Find the location $\vec{X}$ that maximizes equation S1, where j is the reference antenna, and the sum over i is over all other antennas. This maximization is done with the Nelder-Mead algorithm [5] with the starting point chosen to be a random point within the bounding box. This procedure is repeated until the same maximum (within 1 m) is found 100 times, with a maximum of 1500 iterations.

The location of the maximum of the image intensity in stage 1 is now $\vec{X}_{\text{guess}}$, an approximate location for stage 2. It is assumed that this location is very close to the actual source location, and stage 2 is used to refine the source location. The steps in stage 2 are:

1) For every antenna create a window centered at t_i, where t_i is the time that this antenna would receive the radio pulse if that pulse was emitted from the approximate location, $\vec{X}_{\text{guess}}$ from the previous stage, and received at the reference antenna at time t_p. The windows are chosen to be 50 samples long for all antennas.

2) Calculate the cross-correlations, $C_{i,j}(t)$, between the windowed data on every pair of antennas.

3) Find the maximum of equation S1, where the sum is now over every pair of antennas, excluding the antennas in the LOFAR core (except station CS002, so that we still have one station in the area of the core). The LOFAR core is excluded because it is small (about 3 km diameter) and so only provides information on large scales. This large scale information is important in stage 1, but is not only not useful in stage 2 but also significantly increases the time needed in stage 2. The maximization in this step is similar to that in stage 1, except the Nelder-Mead guesses are all within a box of 100 m width centered on $\vec{X}_{\text{guess}}$, only 5 iterations are required for convergence, and there are a maximum of 50 iterations.

4) If the new maximum is more than 1 m away from $\vec{X}_{\text{guess}}$, then our new location becomes $\vec{X}_{\text{guess}}$ and we repeat from step 1 of stage 2. Most sources converge within 2 or 3 iterations, however if we have already had 8 repetitions of stage 2, then we quit looping with the status that stage 2 has not converged.

If stage 2 converges, we then "remove" this source from the data. This is done on every antenna by first calculating t_i, which is the time that radio pulse would arrive if the location from stage 2 is correct and the reference antenna received a pulse from the source at time t_p. Then a window of data centered on t_p is set to zero. As in stage 2, this window is 50 samples long. This is done so that the stronger sources do not interfere with the imaging of the weaker sources.

There are three primary differences between stage 1 and stage 2. First, stage 1 has much longer signal lengths than stage 2. This is because stage 2 assumes that the source location is roughly known and stage 1 only assumes that the source is inside the bounding box. Secondly, the sum (in equation S1) in stage 2 is over all pairs of antennas, while the sum in stage 1 is only over pairs of antennas where one antenna is the reference antenna. This is done because almost all antennas in stage 1, except the reference antenna, have windows that are very large. If the cross-correlation were formed between these large windows it would be dominated by pulses that do not come from the source we wish to image. This is different from the reference antenna, which has a very small window that only includes the pulse we want to image, and stage 2 which has small windows on all antennas. Finally, as explained above, stage 2 excludes all antennas from the LOFAR core (except station CS002) and stage 1 includes all antennas.

We repeat stage 1 and stage 2 on the next strongest source in the data block until 100 sources have been found or the amplitude of the pulse on the reference antenna is the same as our noise level. We stop at 100 sources to minimize the possibility of imaging a weaker source whose reconstructed location is artificially influenced by a stronger source. Most of the time there are more then 100 radio sources in a block of data. This algorithm relies heavily on finding the correct global maximum in stage 1, and this maximum is easily missed. We perform three cuts in order to separate the miss-located sources from well-located ones.

We require that stage 2 converges, that the result of stage 2 is within 50 m of the result of stage 1, and that the image intensity in stage two at the location of the source is greater than 0.85. The requirement that the distance between the stage 1 and stage 2 results are within 50 m is to help guarantee that the assumption in stage 2 is met. Out of 17,290 sources from the 2017 flash with an image intensity larger than 0.85, only 473 have a stage 1-stage 2 distance larger than 50 m. The intensity cut at 0.85 was chosen to best balance image quality and number of sources; a histogram of intensity of the radio sources for the 2017 flash is shown in Fig. S36. As this cut is increased or decreased, the image quality smoothly improves or worsens; no image features completely appear or disappear at different cut levels. About 30% of the pulses we attempted to image pass all three cuts. The time of the source is easily found from the location, $\vec{X}$, and the time the pulse was received on the reference antenna, t_p.

1.1 Artifact Checks

We have performed a number of tests of our imaging technique in an attempt to search for potential artifacts. First, we checked if the needles could be due to some dependence between the reconstructed locations of different sources. I.E. it is possible that finding the location of one source (perhaps the wrong location) could affect the imaged location of a different source. We have found that this is unlikely to be the case because different twinkles of the needles are always on different 327 μs long (2^{16} samples) blocks of data, and each data block is processed independently. This implies that the spatial and temporal structure between needle twinkles cannot be due to any imaging dependence between the needle twinkles. Furthermore, there are many individual needle twinkles which are spread across two blocks of data, strongly implying that the structure of individual needle twinkles cannot be due to imaged source location affecting each other. Finally, we have also imaged the 2017 flash without removing the pulses of imaged sources, and found that while our source location efficiency decreases to 20%, the properties of the needles, and the lightning flash in general, remains the same.

As discussed above, we remove pulses from the data that saturate the LOFAR antennas. We have checked if this process introduces artifacts by imaging the 2017 flash without removing saturated pulses. The resulting image contains all the same features as our image from the standard approach.

In our imaging algorithm we pick a reference antenna in order to determine which pulse to image. We have checked that the final image does not depend upon the choice of reference antenna by re-imaging the 2017 flash using a different reference antenna that is 20 km from the typically reference antenna. The resulting image is indistinguishable from the normal image.

Finally, it is conceivable that if the overlap between blocks is not handled correctly, then pulses could be double counted, leading to artifacts. In order to check this, we have created a histogram of times, modulus block size, of imaged sources that belong to needles. This histogram is flat, within statistical deviation, indicating that imaged sources on needles do not come form any preferred time in a block.

2 Additional data on 2017 flash

Fig. S1 shows the radar reflectivity around the 2017 lightning flash, provided by KNMI. The radar shows that the 2017 flash occurred just to the east of the main core of a storm. figref2017-temp shows the altitude versus temperature for the time and location of the 2017, extracted from the Global Data Assimilation System (see https://ready.arl.noaa.gov/gdas1.php).

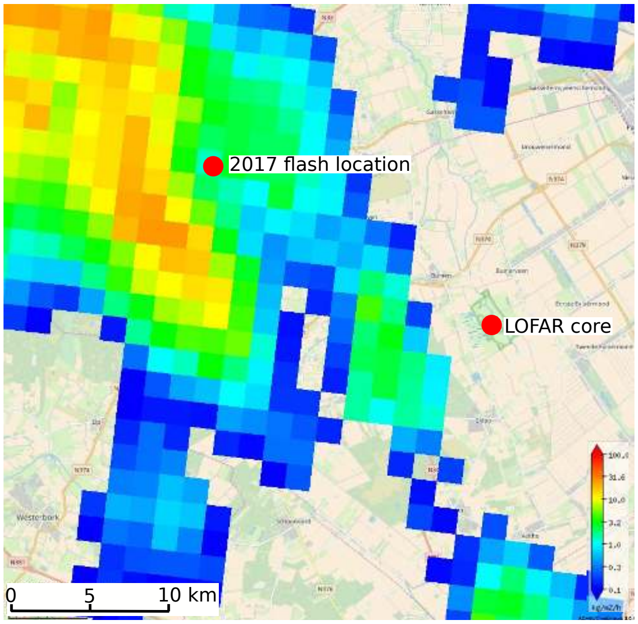

Figure S1: Radar reflectivity during the time of the 2017 flash, provided by KNMI (available from `http://geoservices.knmi.nl/viewer2.0/`).

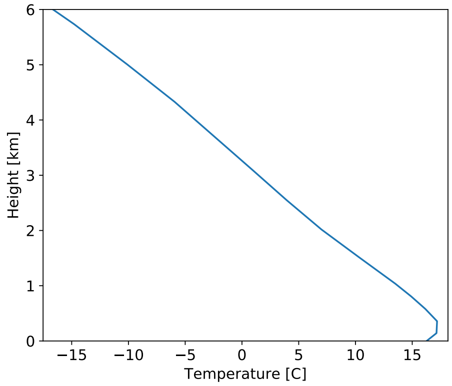

Figure S2: Altitude versus temprature profile during the 2017 lightning flash, derived from the Global Data Assimilation System.

Three animations are also included in the supplementary information.

`Lightning_complete_rotating.mov` is an animation of the whole 2017 flash.

`Lightning_detail_needle.mov` details a section of positive leader, with sources from N4 shown in red.

`Lightning_detail_negative_leader.mov` details a section of negative leader, in order to illustrated the difference between negative and positive leaders.

3 2017 Additional Imaged Needles

Below are a number of figures detailing more needles along the length of the positive channel shown in Fig. 2 left. We have chosen this particular section of the positive leader at random, other sections show similar features. Each figure shows a grey line that illustrates the approximate location of the positive leader channel, the same location as shown in Fig. 2 left. The location of the positive leader channel is found by drawing lines by eye through the source locations shown in Fig. 2 left, keeping in mind that the needles tend to extend out from the positive leader channel.

Fig. S3 shows a zoom-in on a region around N1 in Fig. 2 left. The “N1” label shows the location of the needle that is about 45 degrees off the positive leader channel. The sources along the positive channel could be due to the needle N1, smaller needles that we could not image, or some other phenomenon that we could not image. This needle clearly twinkles at least four times: at t=44 ms, 58 ms, 80 ms, and 95 ms. The two sources at T=48 ms and the one source at T=88 ms could also be twinkles, but it is difficult to be certain. With so few sources in each twinkle, it is entirely reasonable that some twinkles were not imaged. Fig. S4 details one twinkle at 80 ms of N1. This twinkle has few located points, but it is clear that the VHF sources farther from the positive channel occurred later.

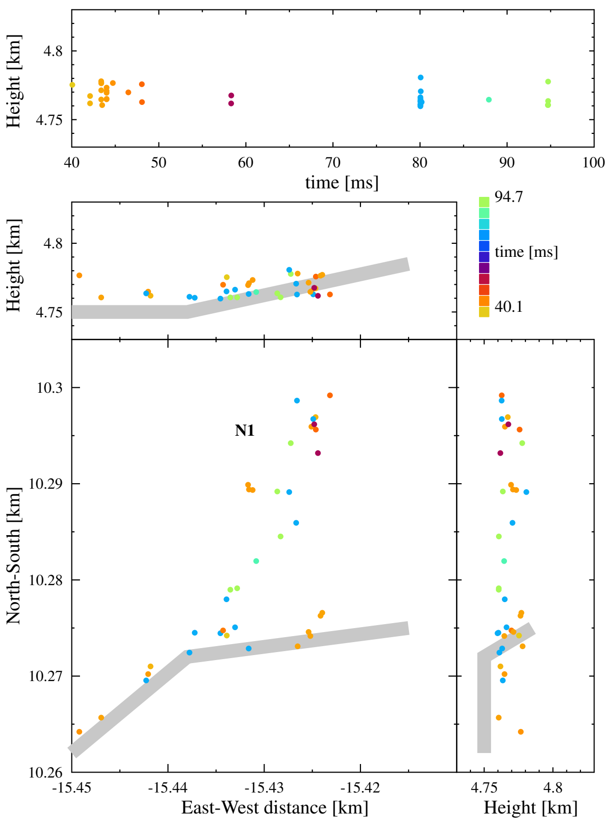

Figure S3: Region around N1 of the 2017 flash (see Fig. 2 left). The grey line shows the approximate location of the positive leader channel. The label "N1" indicates the needle N1.

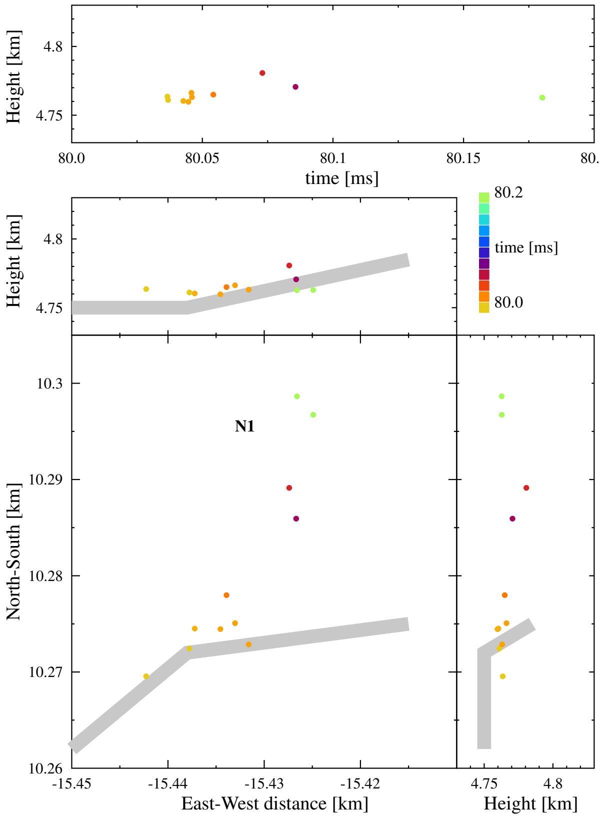

Figure S4: Region around one twinkle at T=80 ms of needle N1.

Fig. S5 shows a zoom-in on a region around N2 in Fig. 2 left. There do not appear to be any sources along the leader channel in this region, all the sources seem to come from needle N2. This needle has at least 5 twinkles at T= 81 ms, 85 ms, 89 ms, 83 ms, and 98 ms. Each of which are about 3–5 ms apart. There is one source at T=76 ms that could be another needle twinkle, but it is difficult to tell with only one located source. Fig. S6 shows one twinkle of N2, at T=81 ms.

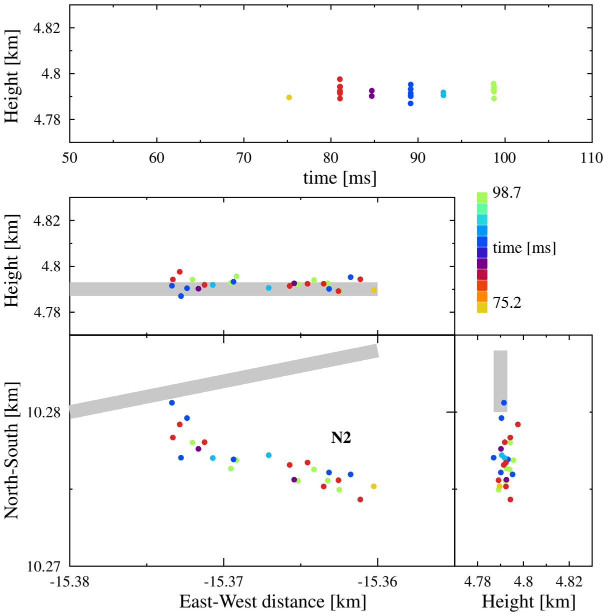

Figure S5: Region around N2 of the 2017 flash (see Fig. 2 left). The grey line shows the approximate location of the positive leader channel.

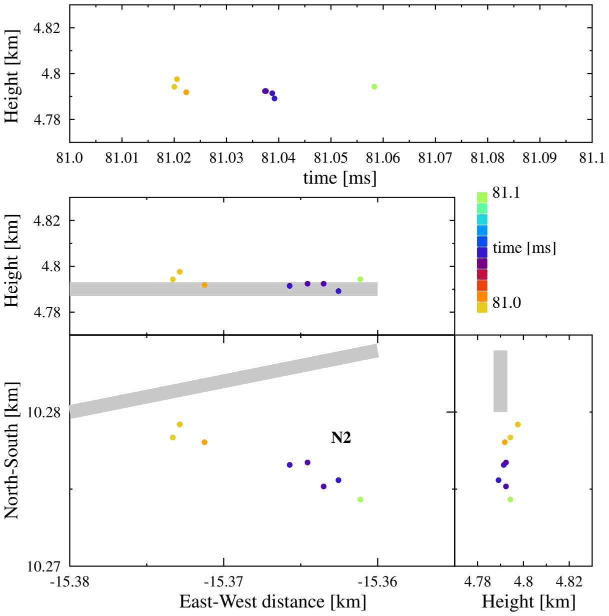

Figure S6: One twinkle, at T=81 ms, of needle N2.

Fig. S7 shows a zoom-in on a region around N3 in Fig. 2 left. This region has three separate needles in close proximity, labeled "A", "B", and "N3", where needle N3 is the one most clearly seen in Fig. 2. Height vs Altitude shows a number of twinkles from these three needles, and it is difficult to distinguish which twinkle goes with which needle from this figure. It is clear that the last two twinkles at T=96 ms and 103 ms both come from needle N3. These two twinkles are about 7 ms apart. Fig. S8 shows one relatively well imaged twinkle of needle N3. While there is some scatter, in general the later VHF sources tend to occur farther from the positive leader channel.

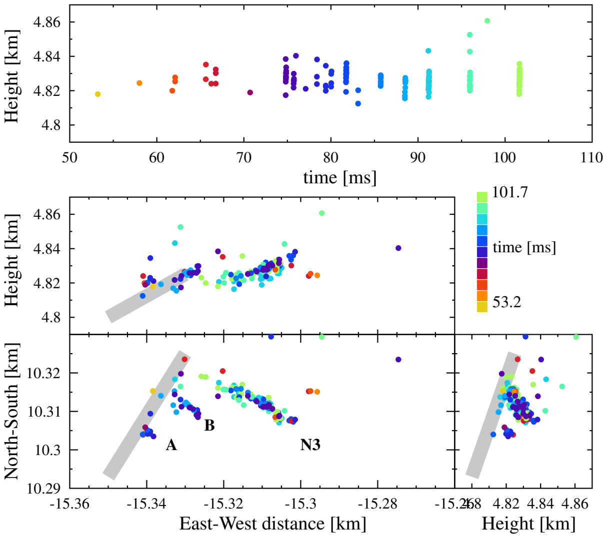

Figure S7: Region around N3 of the 2017 flash (see Fig. 2 left). The grey line shows the approximate location of the positive leader channel. The three labels, "A", "B", and "N3" distinguish between three separate needles.

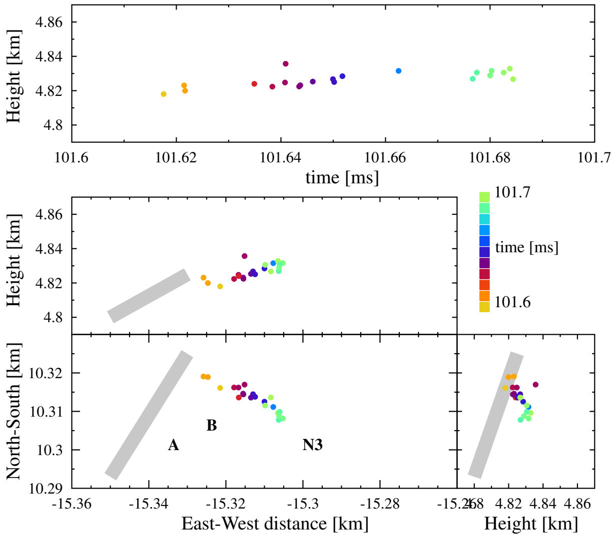

Figure S8: One twinkle, at T=101 ms, of needle N3.

Fig. S9 shows a zoom-in on a region around N4 in Fig. 2 left, the same region as shown in Fig. 3. As discussed in the main body this needle has 5 well-imaged twinkles. Fig. S9 also shows a few sources from a separate structure, labeled "C", which could be another needle.

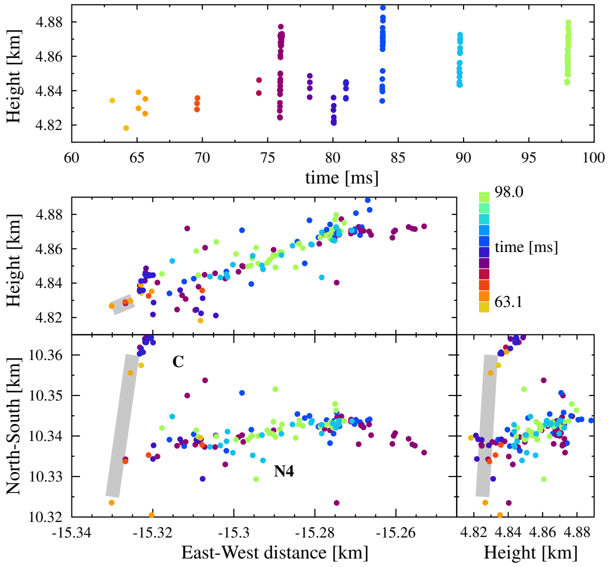

Figure S9: Region around N4 of the 2017 flash (see Fig. 2 left). The grey line shows the approximate location of the positive leader channel.

Fig. S10 shows a zoom-in on a region around N5 in Fig. 2 left. This region has three separate groups , labeled “D”, “E”, and “N5”. Despite being imaged with very few sources, Groups E and N5 are clearly needles, as they clearly have multiple twinkles, with each twinkle having multiple sources. Group D only has three sources from different times. It is possible that Group D is a small needle with at least three twinkles, but there are too few sources to be certain.

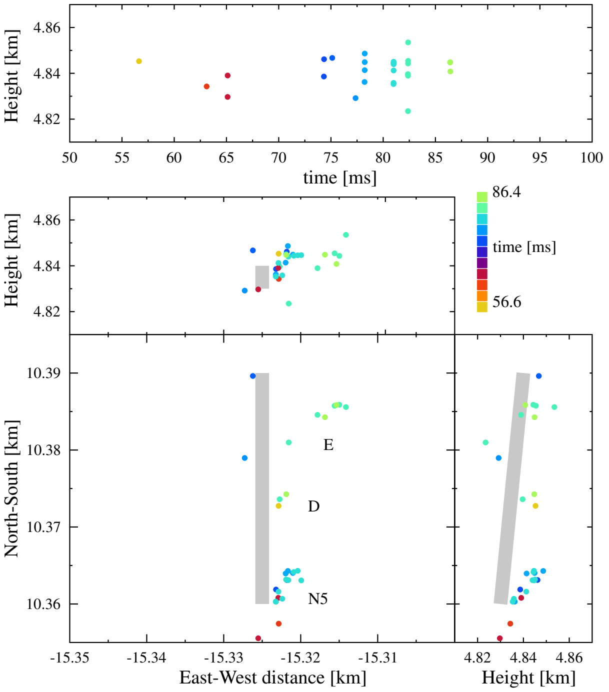

Figure S10: Region around N5 of the 2017 flash (see Fig. 2 left). The grey line shows the approximate location of the positive leader channel. The three labels, “D”, “E”, and “N3” distinguish between three separate groups of sources. Groups E and N5 are clearly needles.

Fig. S11 shows a zoom-in on a region around N6 in Fig. 2 left. This needle has at least four activations at T=69 ms, 78 ms, 85 ms, and 88 ms. There are single sources at T=63 ms and 94 ms that could also be twinkles. There are also two sources at 4.92 km altitude, T=65 ms, that seem to be from a higher altitude than the rest of the needle. It is not clear how these two sources are related to needle N6. This needle only has a few sources in each twinkle, as so is not a well-imaged needle. It appears to be parallel to the positive leader channel; however, the fact that this needle is poorly imaged and that the path of the positive leader is difficult to discern makes it difficult to tell if N6 is truly parallel to the positive leader channel or not.

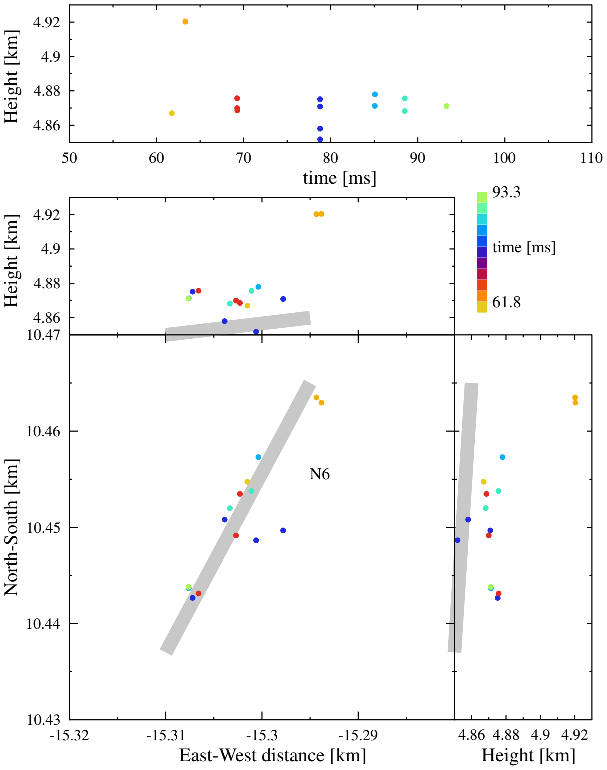

Figure S11: Region around N6 of the 2017 flash (see Fig. 2 left). The grey line shows the approximate location of the positive leader channel.

Fig. S12 shows a zoom-in on a region around N7 in Fig. 2 left. This needle has at least 5 activations at T = 68 ms, 80 ms, 88 ms, 92 ms, and 98 ms. There are also two sources at T=63 ms and 77 ms that could be poorly imaged twinkles. The shortest time between imaged activations is about 4 ms and the longest is about 10 ms. However, given that there are multiple twinkles with few, or perhaps only one, sources, it is entirely possible that some twinkles were not imaged at all.

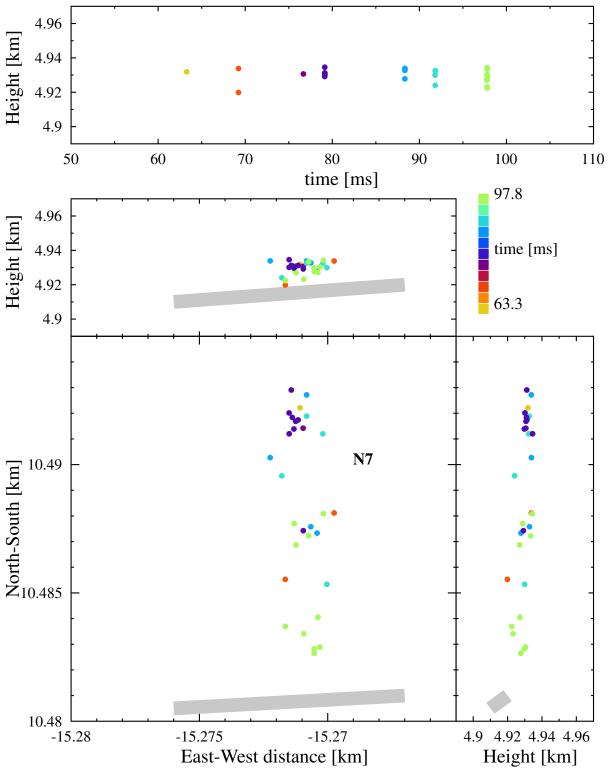

Figure S12: Region around N7 of the 2017 flash (see Fig. 2 left). The grey line shows the approximate location of the positive leader channel.

Fig. S13 shows a zoom-in on a region around N8 in Fig. 2 left. This needle has at least 6 activations at T=70 ms, 76 ms, 85 ms, 87 ms, and 90 ms, as well as a single source at T=81 ms. It is interesting to note that the activation at T=87 ms, altitude of 4.94 km, is only about 2 ms after the previous activation. Furthermore, this twinkle has three sources that appear on the opposite side of the leader channel (between Y=10.475 km and 10.48 km). It is not clear why this is the case, especially given that the location of the leader channel in Fig. S13 is highly approximate.

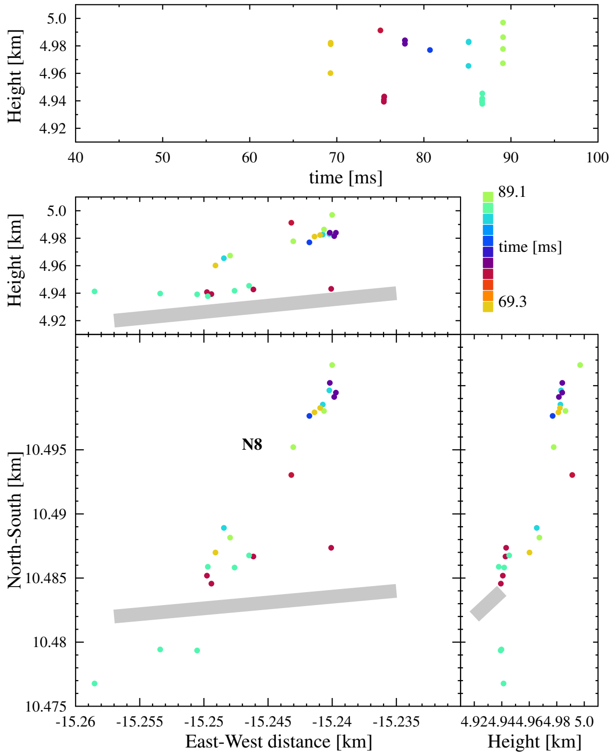

Figure S13: Region around N8 of the 2017 flash (see Fig. 2 left). The grey line shows the approximate location of the positive leader channel.

Fig. S14 shows a zoom-in on a region around N9 in Fig. 2 left. This region has two needles labeled "F" and "N9". Needle N9 has four clear twinkles at T=77 ms, 83 ms, 87 ms, and 93 ms. There is about 4–6 ms between each twinkle. Needle E clearly twinkles twice at T=78 ms and 85 ms. These two needles can be clearly distinguished, as the 6 clear twinkles shown in Height vs Time fall into two groups. The first group, containing the twinkles from N9, are above an altitude of 5 km. The other group, the activations corresponding to Needle F, clearly come from a lower altitude with a slight time offset relative to needle N9. Furthermore, needle N9 and needle F come from two distinct separate regions in the plan projection. The minimal distance between needle N9 and needle F is about 5 m. The fact that we can distinguish two needles that have a minimal distance around 5 m is one piece of evidence that our location error is better than 5 m. Fig. S14 shows one twinkle of needle N9, at T=87 ms.

335

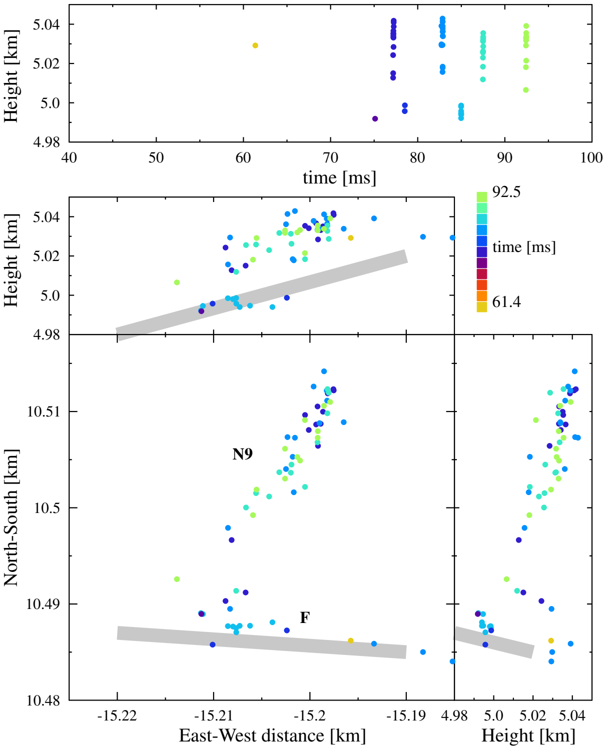

Figure S14: Region around N9 of the 2017 flash (see Fig. 2 left). The grey line shows the approximate location of the positive leader channel. Labels “F” and “N9” distinguish between the two needles.

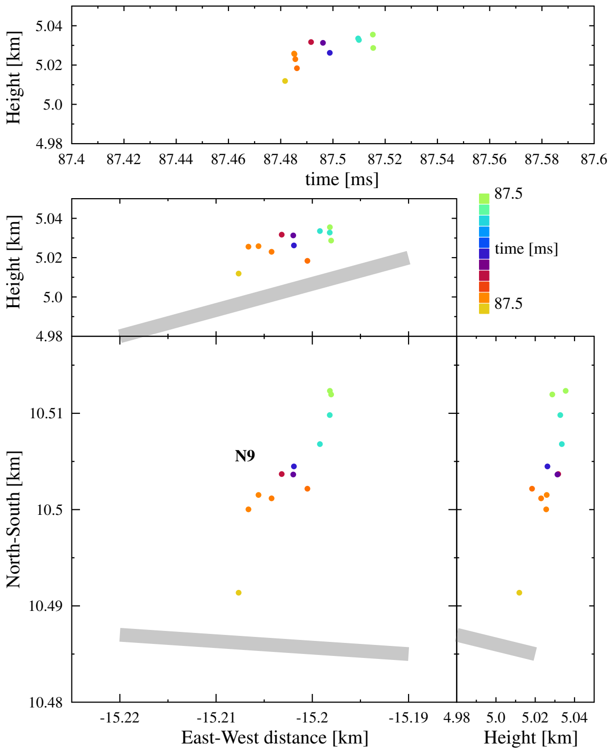

Figure S15: One twinkle of needle N9, at T=87 ms.

Fig. S16 shows a zoom-in on a region around N10 in Fig. 2 left. Apart from needle N10, this region also contains multiple sources along the positive leader channel. These sources are grouped together in time at T=83 ms, similar to a needle twinkle. These sources could be due to a small needle that is poorly imaged and close to parallel with the positive leader channel. Needle N10 has 6 clear activations at T=76 ms, 81 ms, 85 ms, 90 ms, 94 ms, and 99 ms. Needle N10 appears to have sources on both sides of the positive leader channel. However, the sources south of the channel are most likely a smaller, poorly imaged needle. Fig. S17 details the twinkle at T=75 ms.

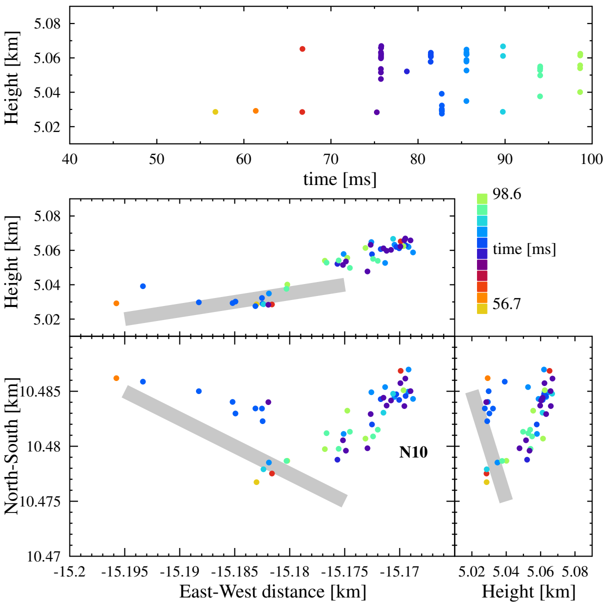

Figure S16: Region around N10 of the 2017 flash (see Fig. 2 left). The grey line shows the approximate location of the positive leader channel. The "N10" shows the sources related to needle N10.

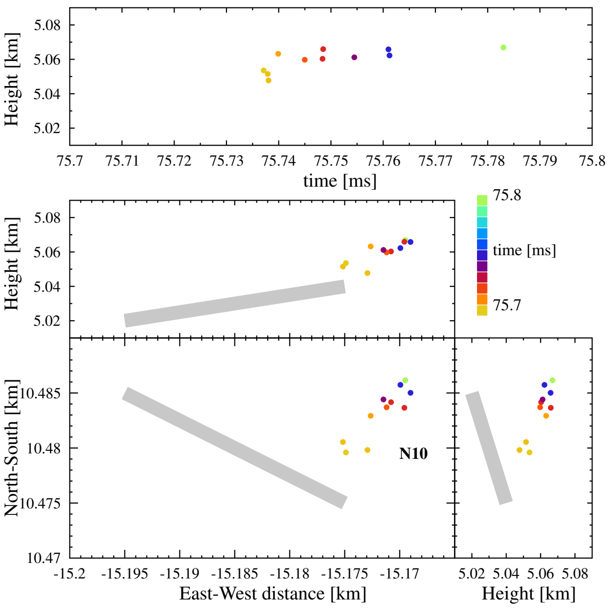

Figure S17: One twinkle of needle N10, at T=75 ms.

Fig. S18 shows a zoom-in on a region around N11 in Fig. 2 left. This needle is very short in length but still contains a larger number of sources. N11 has 7 clear twinkles at T=70 ms, 75 ms, 77 ms, 81 ms, 85 ms, 88 ms, and 96 ms. There are also two smaller groups of sources, labeled "G" and "H" separate from needle N11. These smaller groups could be smaller poorly imaged needles, but they do not have enough sources to say for certain. Fig. S19 shows a twinkle of N11 at T=95 ms.

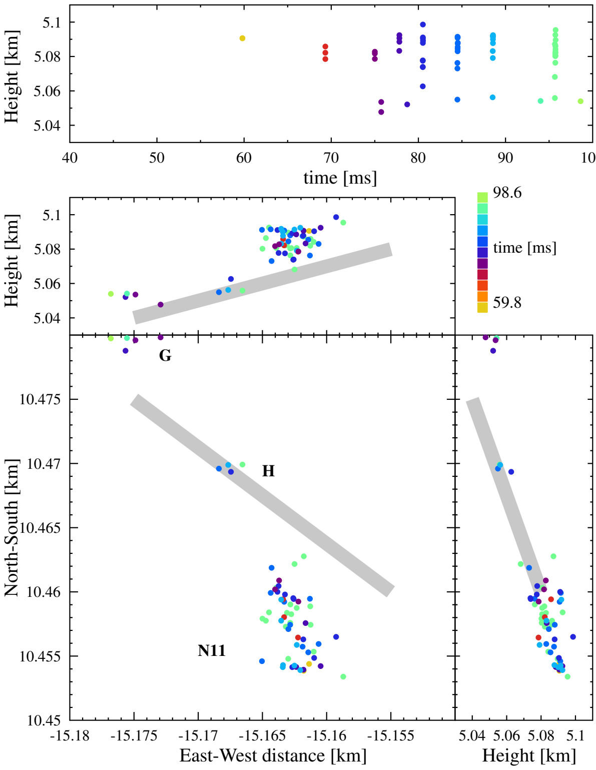

Figure S18: Region around N11 of the 2017 flash (see Fig. 2 left). The grey line shows the approximate location of the positive leader channel. The labels "G", "H", and "N11" distinguish between the three groups of sources, where N11 is clearly a needle.

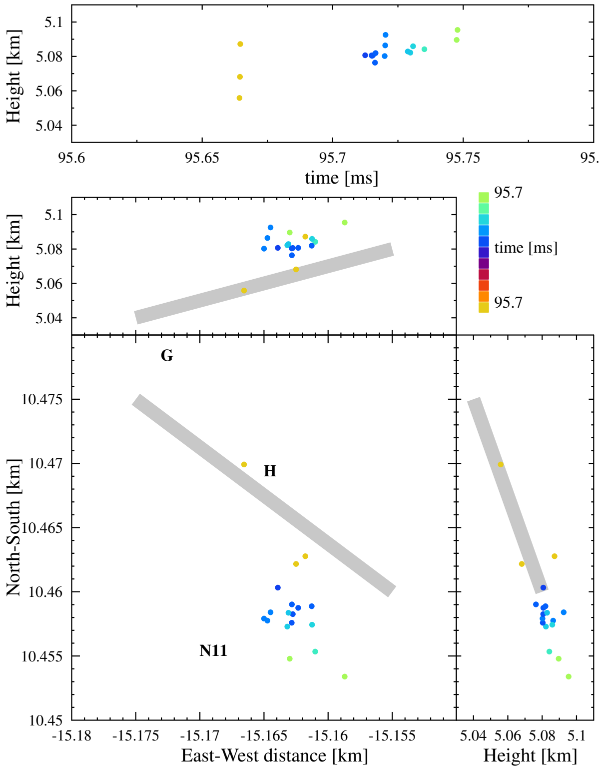

Figure S19: One twinkle of N11 at T=95 ms.

4 2016 Additional Imaged Needles

Fig. S20 shows an overview of the flash from 12, July, 2016. This was an inverted intra-cloud flash, 40 km from the core of LOFAR and outside the area enclosed by LOFAR, and so was not imaged with as high of a quality as the 2017 flash. The negative leader started around 5.5 km altitude, propagated down to 2.5 km altitude by t=25 ms, and then propagated horizontally into a positive charge region. The positive leader was not seen until 25 ms after the start of the negative leader, and its first located sources were about 300 m above the start of the negative leader. The positive leader then propagated into a negative charge region at 6 km altitude. Later, at t=75 ms, a branch of the negative leader propagated into a small positive charge region at about 4.5 km altitude. Fig. S21 shows the altitude versus temperature for the time and location of the 2016, extracted from the Global Data Assimilation System (see https://ready.arl.noaa.gov/gdas1.php). Fig. S22 shows a section of positive leader. Most of the sources are scattered along the leader channel, not at the tip of the leader, and so are consistent with needles, however the leader is too poorly imaged for many of the needles to be visible.

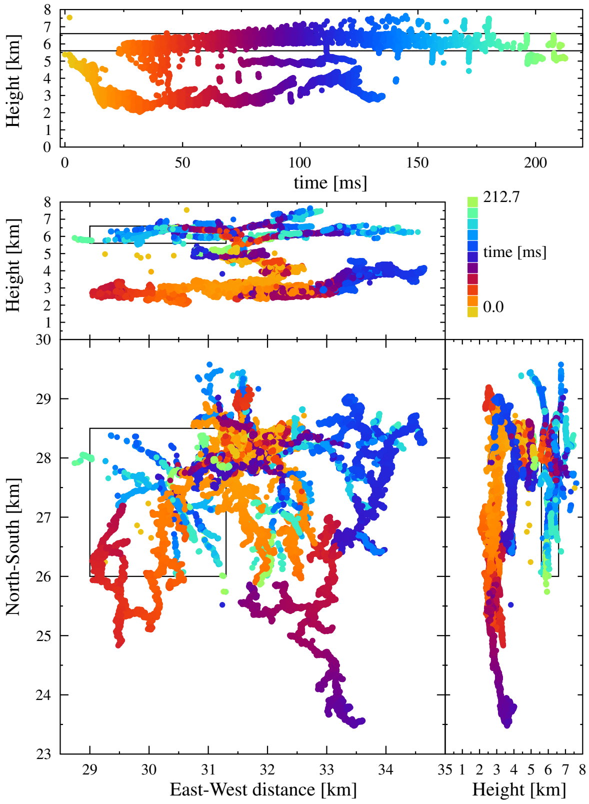

Figure S20: Image of a lightning flash that occurred on 12, July, 2016. Each dot is the location of a VHF source. The sources below 5.5 km altitude are from the negative leader, and the sources above 5.5 km altitude are from the positive leader. This was an inverted intra-cloud flash, and did not connect to ground. The box shows the region expanded in Fig. S22.

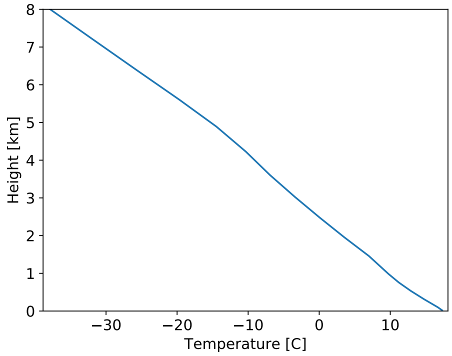

Figure S21: Altitude versus temprature profile during the 2016 lightning flash, derived from the Global Data Assimilation System.

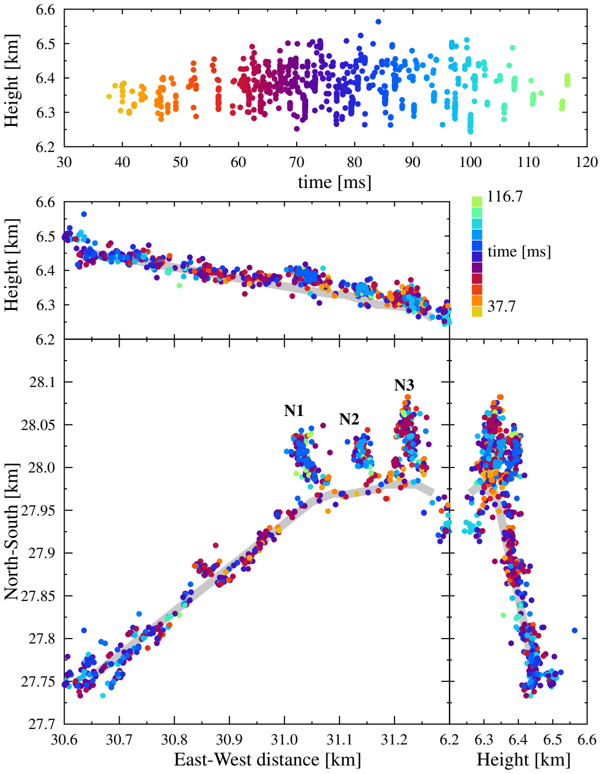

Figure S22: Image of a section of positive leader during the 2016 flash shown in Fig. S20. This section shows three clear needle-like structures, in addition to many sources that are scattered along the body of the leader that are consistent with being due to needles, but the image is not clear enough to be certain. The grey lines show the approximate location of the positive leader channel.

Fig. S23 details needle 2016-N1 shown in Fig. S22. This needle is about 60 m long, and twinkles over 18 times. Most of the twinkles occur 5–10 ms after the previous one, but the last well-imaged twinkle seems to occur over 50 ms from its previous twinkle. It is not obvious if this is physical, or due to twinkles not being imaged. Fig. S24 shows one twinkle of needle 2016-N1 at T=175 ms.

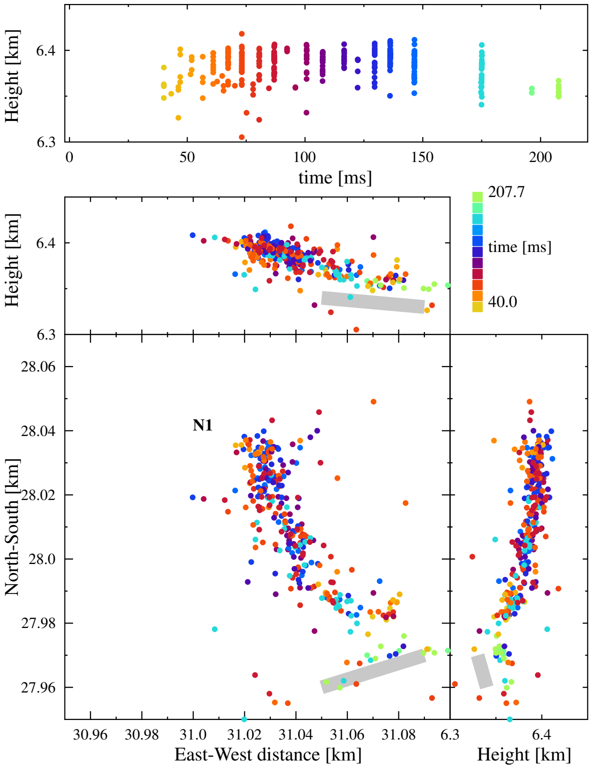

Figure S23: Region around 2016-N1 shown in Fig. S22.

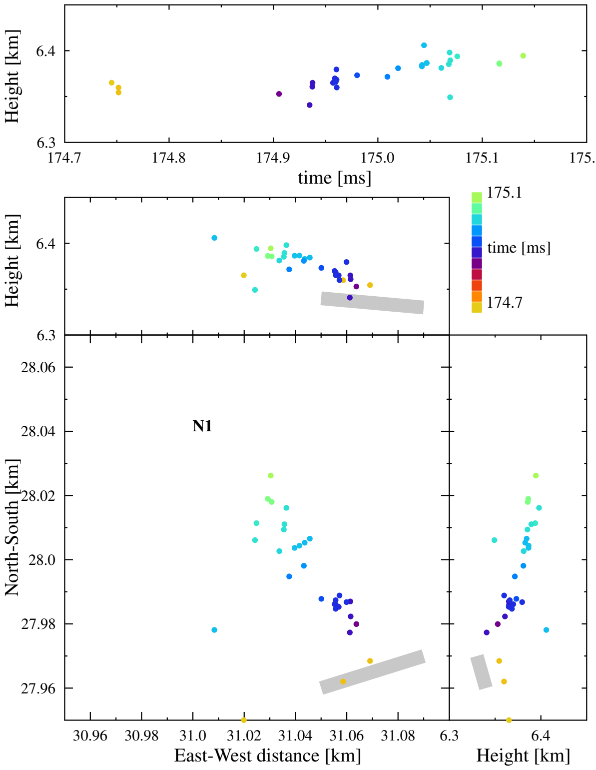

Figure S24: One twinkle of 2016-N1 at T=175 ms

Fig. S25 details needle 2016-N2 shown in Fig. S22. This needle is about 40 m long, and twinkles about 9 times. There are also three groups of sources that occur t=150 ms, but these seem to occur along the positive leader channel and not 2016-N3, and so may be part of a poorly imaged needle. Fig. S26 shows one twinkle at T=85 ms.

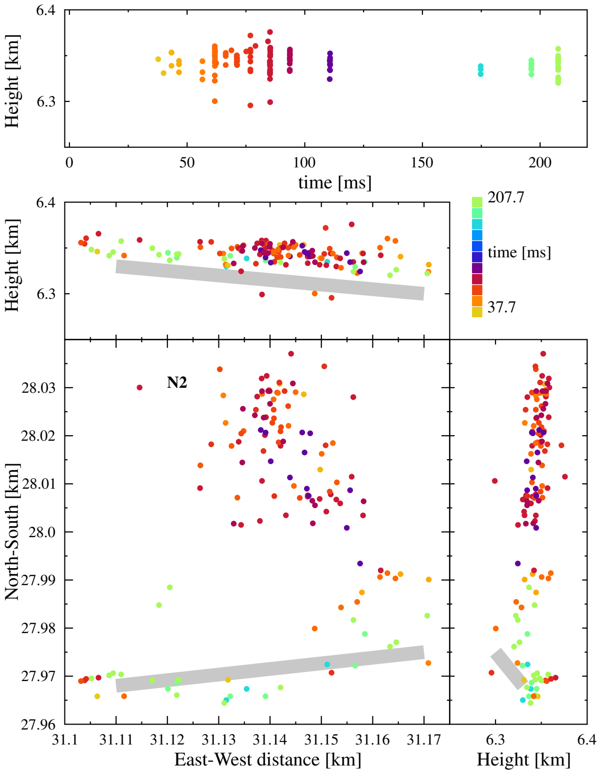

Figure S25: Region around 2016-N2 shown in Fig. S22.

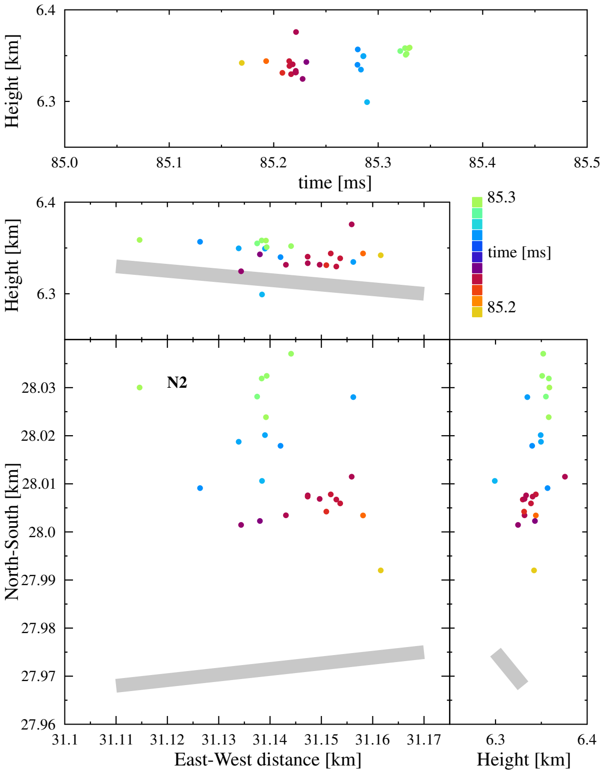

Figure S26: One twinkle of needle 2016-N2 at T=85 ms.

Fig. S27 details needle 2016-N3 shown in Fig. S22. This needle is about 100 m long, one of the longest needles we have found. This needle twinkles over 14 times. Most of the twinkles are 5 – 10 ms apart, but like 2016-N1, there is a later twinkle that occurs over 50 ms since the previous twinkle. There is also a small needle, labeled “I”, that is about 10 m long, and twinkles 4 times. Fig. S28 shows one twinkle of needle 2016-N3 at T=175 ms. This twinkle seems to be clustered in time. It is not clear if this clustering is an imaging artifact or is physical.

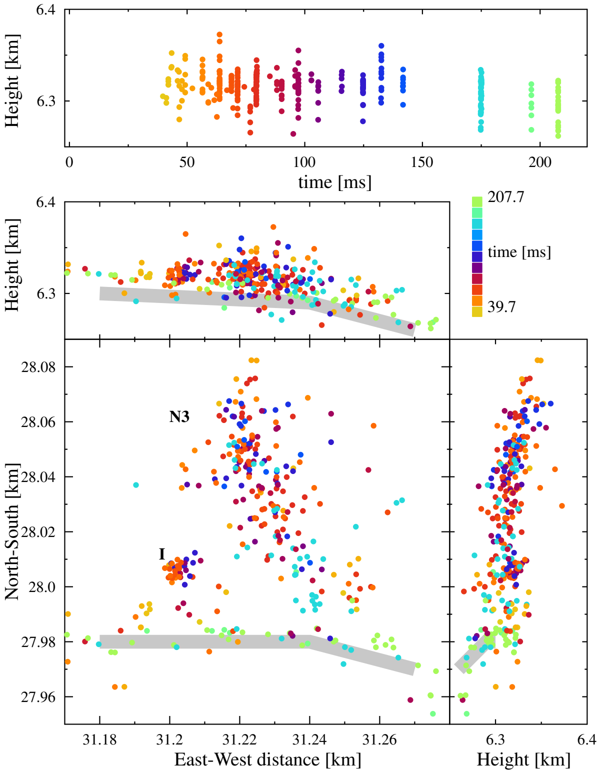

Figure S27: Region around 2016-N3 shown in Fig. S22. The labels "2016-N3" and "I" show two different needles.

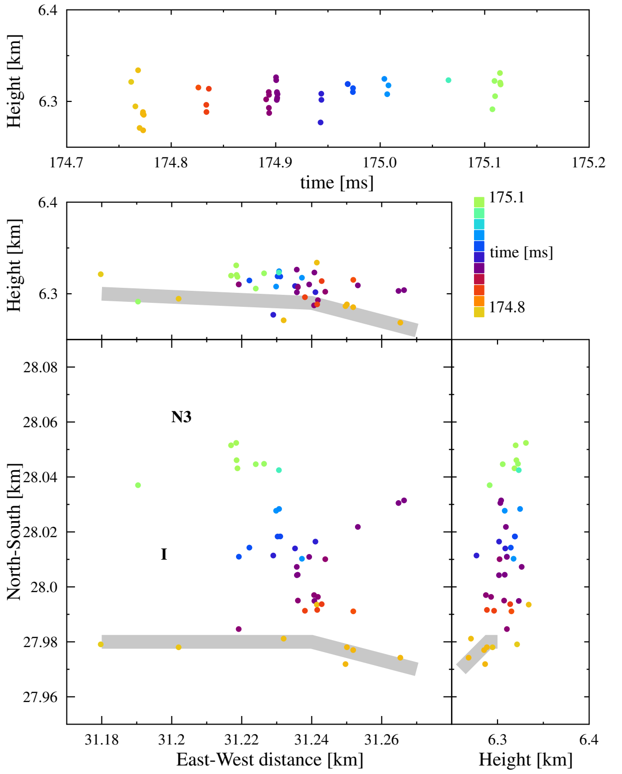

Figure S28: One twinkle of needle 2016-N3 at T=175 ms

5 Additional Hypothesis

Apart from the hypothesis listed in the main body, there are two other ways that the electric field perpendicular to the leader channel could flip direction. The first effect is due to the fact that the highest charge densities around a lightning leader are found at the tip of the leader. This is illustrated by Fig. S29, which shows the line charge density along a short and long cylinder in a uniform electric field, calculated using a method-of-moments simulation [6]. This figure show a large spike in the line-charge density at the tips of the cylinders. This implies that as a leader grows, the charge density at one point on the leader starts high but then must decrease as the leader grows. For an infinitely conducting cylinder this is no problem, but lightning is a different story. The corona around the lightning channel that holds the positive charge must, by definition, be very poorly conducting. If it was not, then any significant charge density would flow away quickly. As a result, the positive charges cannot be removed easily, therefore negative charge must be added to the inside of the corona as the leader gets longer. Fig. S30 shows a diagram of the charge corona around the lightning leader according to this hypothesis. The corona of a section of positive leader that is immediately behind the leader tip is mostly dominated by positive charges, but the corona on a section of leader far behind the leader tip will have a layer of negative charge in order to lower the total line charge density. Eventually the negative charge will cause dielectric breakdown, which we map as a needle twinkle. However, the charge-density difference between a short and long channel is only large near the tip of the leader, so it is not clear if this effect is strong enough to produce needles over hundreds of meters, this is compounded by the fact that the leader channel is generally curved, and not straight as used in Fig. S29.

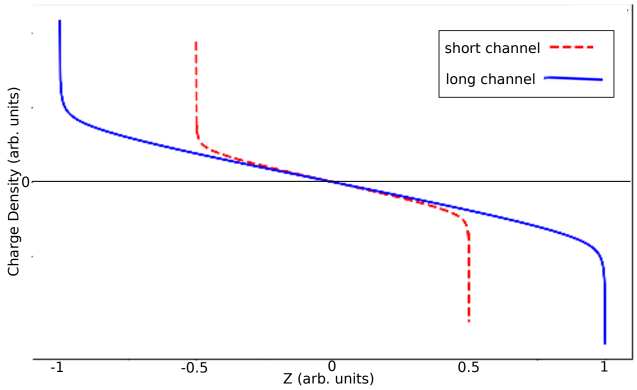

Figure S29: Line charge density along a short and long perfectly conducting cylinder in a uniform electric field.

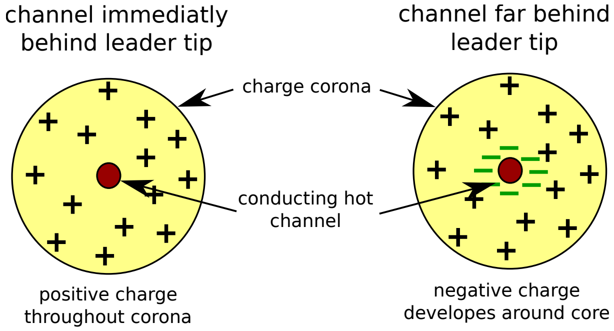

Figure S30: Diagram showing the corona immediately behind the positive leader tip which is mostly positive charge, and the corona far behind the leader tip which contains a layer of negative charge.

The third mechanism is comparatively simple. It is known that there are current pulses that start on the positive leader and propagate along the leader channel towards the negative leader. They are typically called retrograde negative breakdown, also known K-changes or recoil leaders [7, 8, 9, 10, 11, 12]. Since these current pulses are a form of negative breakdown, they move negative charge along the channel as they propagate. So it is possible that as they propagate by a needle they could deposit enough negative charge to initiate breakdown that we then see as a needle twinkle. In our data see many instances of negative retrograde breakdown, however only a single one is associated with needle activity.

Fig. S31 shows this retrograde breakdown seen in the 2016 flash, and three needles that twinkled shortly afterwards. The retrograde breakdown is poorly imaged, and only has an imaged length of about 400 m. The three needles include N1 and N3 shown in Fig. S22, and a third needle that is labeled "J". It is interesting to note that needle N2, which is between N1 and N3, does not twinkle. Needle J is after the imaged length of the the retrograde breakdown. However, given how soon needle J happens after the retrograde breakdown, it is most likely that the retrograde breakdown continued down the channel, past needle J, and simply was not imaged.

Most of the retrograde negative breakdown imaged by LOFAR is similar to that reported in previous literature. Future work will be needed to see if LOFAR sees any new properties that have not been previously reported.

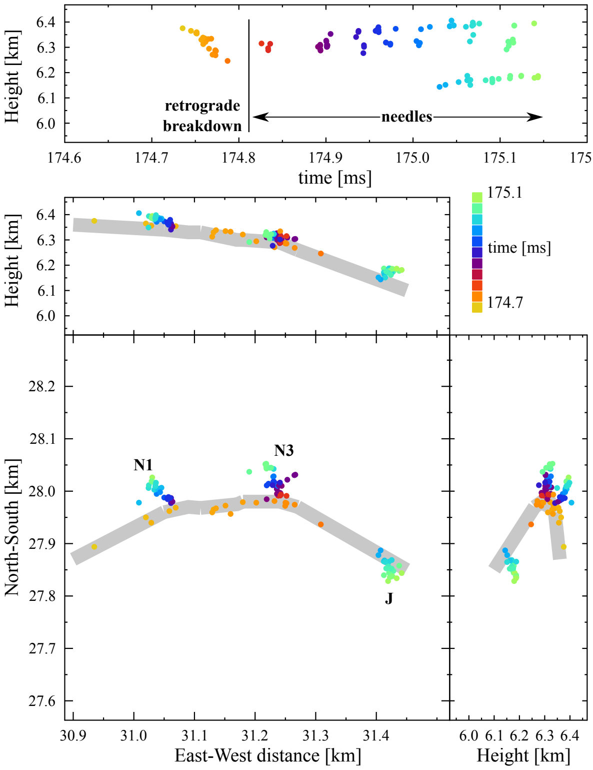

Figure S31: An instance of negative retrograde breakdown that is shortly followed by the twinkles of three needles.

Unfortunately, for all three hypothesis given, too little is known about lightning leaders, such as their capacitance, how much charge they collect as they propagate, the distribution of that charge along the lightning channel, and the structure of the leader corona, to produce a numerical model to compare against our observations. Similarly, we are presently unable to estimate the current on the leader channel or how needles affect that current more precisely then the hypothesis given main text.

6 Potential for Optical Observations

One possibility to gain more insight into the physics behind needles is to image them with a high-speed camera. However, it is not clear if needles can be detected through optical observations. The primary difficulties being that needles are small and positive leaders are often hidden from view inside thunderclouds. This uncertainty is compounded by the fact that we don't know if needles are highly conductive, similar to negative leaders, and thus very luminous or needles only consist of clusters of streamers and do not form a conducting channel, and so are not very luminous.

Upward positive leaders near ground have been studied extensively (see [13] and citations therein) without detection of needles, but it is possible that upward positive leaders behave differently then positive leaders in intra-cloud or negative cloud-to-ground lightning. One obvious possibility is that upward positive leaders may have a higher current, and therefore do not form disconnections that can lead to needles.

It should be noted that the structure of needles have three primary characteristics: 1) They are relatively small, around 100 m long or less. 2) They are roughly perpendicular to the positive leader channel. 3) They flicker multiple times, with about 2-10 ms between flickers.

7 Location Error Analysis

The dominant source of location error in our analysis comes from our assumption that our sources are point sources. In reality, lightning VHF source regions are extended sources. The visible effect is that our measured waveforms have a different shape and a different polarization depending on the antenna. Antennas that are relatively close, as are all antennas in the LOFAR core (diameter 2 km) for example, have very similar waveforms, however antennas that are far apart can have very different waveforms, as they view the radio source region from completely different angles. The resulting effect in our analysis is that the waveforms do not interfere with each other in the same way that one would expect for short-baseline interferometry. This is the primary reason that we take the absolute value of the cross-correlation before summing over pairs of antennas in equation S1, as it makes our analysis less sensitive to this problem.

We have estimated the location error of our sources by looking at the size of the smallest structures that we can image, which must be larger than twice our location error. Such structures would be the width of the needles and the negative leader and the distance between needles. The width of the needles and the negative leaders all tend to be 5 m, and we have not found any well-imaged structure with a size smaller than 5 m. Furthermore, Fig. S14 shows two needles that we have imaged that are about 5 m apart. Therefore, since our location error cannot be much larger than the diameter of our image size (5 m diameter), and we can image structures as small as 5 m in size, then we conclude that our location accuracy must be better than 2 m in X and Y. Our location accuracy in Z is much worse, but better than 15 m.

Fig. S32, Fig. S33, Fig. S34, and Fig. S35 show the image intensity around a typical source, defined by equation S1, sliced at constant height and constant North-South distance at the location of the source. These four figures show the main beam around the found source location. These figures show that the image is smooth, with minor side beams

Not all sources are equally well-imaged. As discussed previously, about 30% of the sources we attempt to image pass our final cuts. Our primary cut is that the image intensity must be larger than 0.85 (maximum of 1). Fig. S36 shows the distribution of image intensities. The source used to produces the images in Fig. S32, Fig. S33, Fig. S34, and Fig. S35 had an intensity of 0.91.

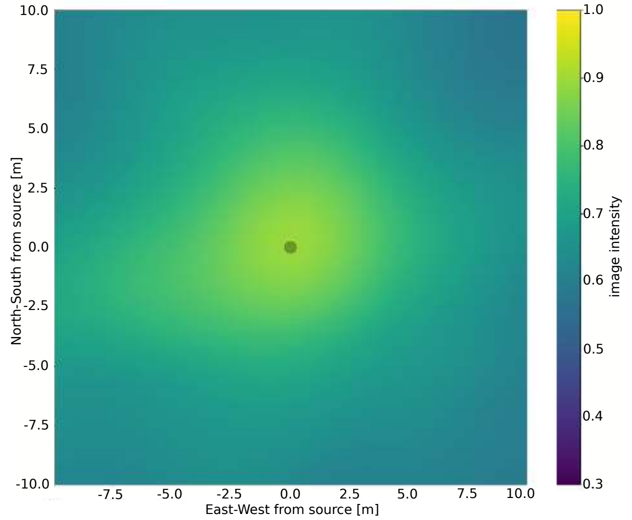

Figure S32: Image of a source sliced at a constant height. The found location of the source is at the grey circle.

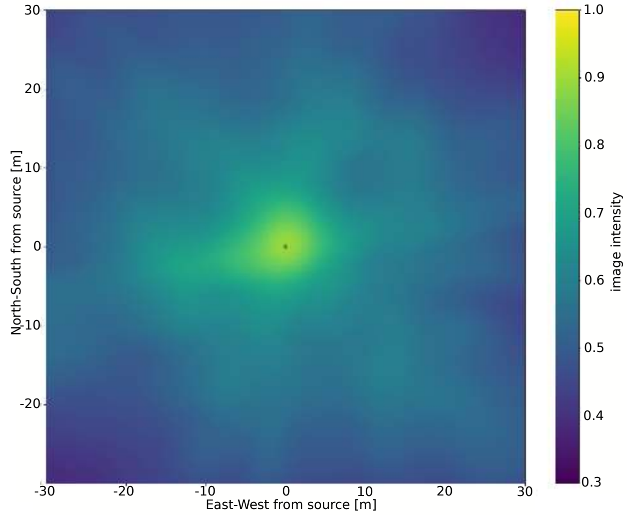

Figure S33: Image of a source sliced at a constant height. The found location of the source is at the grey circle.

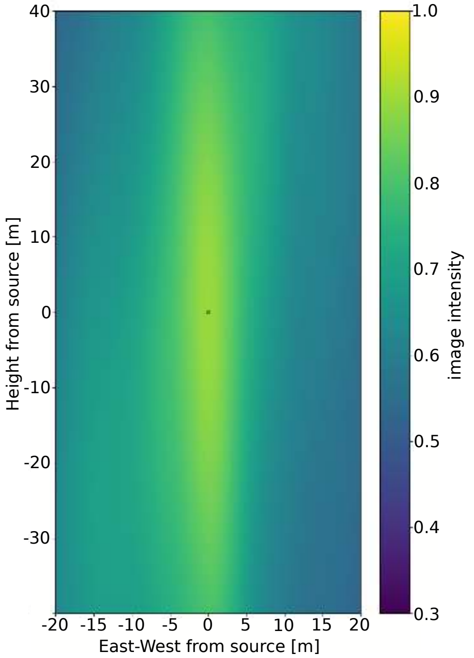

Figure S34: Image of a source sliced at a North-South distance. The found location of the source is at the grey circle.

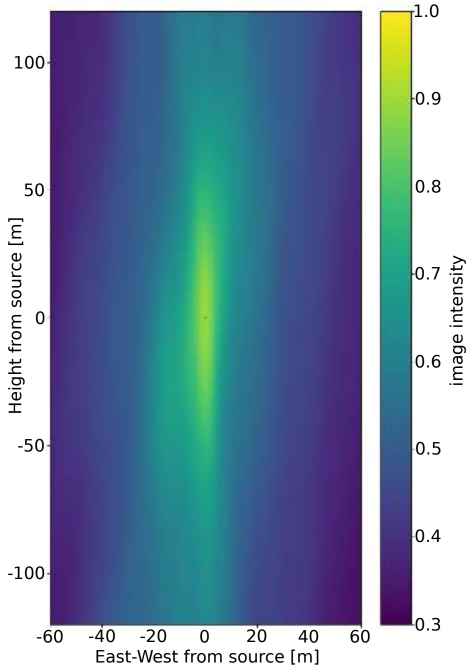

Figure S35: Image of a source sliced at a North-South distance. The found location of the source is at the grey circle.

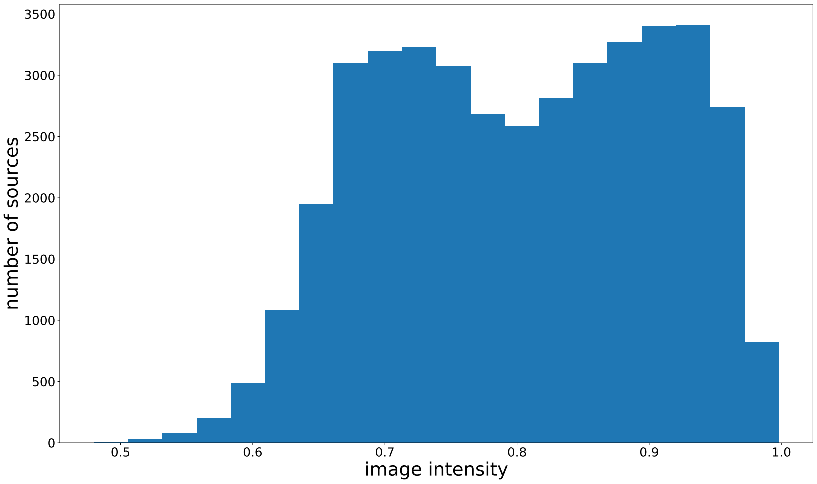

Figure S36: Histogram of image intensities for the 2017 flash.

References

[1] Norden, M. & D Bregman, J. Lightning protection strategy used in lofar radio telescope. In *9th International Symposium on Electromagnetic Compatibility Joint with the 20th International Wroclaw Symposium on Electromagnetic Compatibility (EMC EUROPE 2010)*, 569–575 (2010).

[2] van Haarlem, M. P. *et al.* LOFAR: The LOw-Frequency ARray. *A&A* **556**, A2 (2013).

[3] Dwyer, J. R. & Uman, M. A. The physics of lightning. *Physics Reports* **534**, 147 – 241 (2014). The Physics of Lightning.

[4] Corstanje, A. *et al.* Timing calibration and spectral cleaning of LOFAR time series data. *Astronomy and Astrophysics* **590** (2016).

[5] Gao, F. & Han, L. Implementing the nelder-mead simplex algorithm withadaptive parameters. *Computational Optimization and Applications* **51**, 259–277 (2012).

[6] Gibson, W. C. *The Method of Moments in Electromagnetics* (Chapman and Hall/CRC, 2008).

[7] Mazur, V. Triggered lightning strikes to aircraft and natural intracloud discharges. *Journal of Geophysical Research: Atmospheres* **94**, 3311–3325 (1989).

[8] Shao, X. M., Krehbiel, P. R., Thomas, R. J. & Rison, W. Radio interferometric observations of cloud-to-ground lightning phenomena in florida. *Journal of Geophysical Research: Atmospheres* **100**, 2749–2783 (1995).

[9] Mazur, V. Physical processes during development of lightning flashes. *Comptes Rendus Physique* **3**, 1393 – 1409 (2002).

[10] Edens, H. E. *et al.* VHF lightning mapping observations of a triggered lightning flash. *Geophysical Research Letters* **39**, L19807 (2012).

[11] Stock, M. G. *et al.* Continuous broadband digital interferometry of lightning using a generalized cross-correlation algorithm. *Journal of Geophysical Research: Atmospheres* **119**, 3134–3165 (2014).

[12] Mazur, V. The physical concept of recoil leader formation. *Journal of Electrostatics* **82**, 79 – 87 (2016).

[13] Rakov, V. A. & Uman, M. A. *Lightning: Physics and Effects* (Cambridge University Press, 2007).