Nature paper “Quantized Majorana conductance”

Report from independent experts

Expert members
Prof. Piet Brouwer
Prof. Klaus Ensslin
Prof. David Goldhaber-Gordon
Prof. Patrick Lee
(below, “the experts” and “we” are used interchangeably)

TU Delft support
Prof. Pieter Kruit
Dr. René Haverslag

Index

Preamble

In 2020 serious concerns were raised by members of the scientific community regarding some of the conclusions reached by this publication in Nature. The four of us were contacted by the TU Delft to provide advice on this issue. The authors provided us access to unpublished data available to them prior to publication.

This public version of our report focuses on the scientific methodology and interpretation, in light of the data and the theoretical background available to the authors at the time of submission. In particular, we aim to provide a snapshot of that timeframe, and pay minimal attention to more recent theoretical developments that may influence the authors' current thinking. We also do not consider subsequent experimental findings, except inasmuch as they shed light on the thinking prior to publication.

1. Introduction

Since 2010 the physics community has been excited by the prospect of observing Majorana fermions -- an analogue to an elegant concept in elementary particle physics -- in carefully designed and tuned nanometer-scale solid state structures. Two theoretical papers that year suggested assembling known building blocks, semiconductor nanowires and superconductors, into a hybrid structure with new properties, and then tuning that hybrid with magnetic field and nanopatterned gate electrodes to manifest the elusive Majorana. This looked experimentally daunting, since superconductivity and magnetic field generally do not coexist easily, and the high electron density in a superconductor might oppose using electric fields from gate electrodes to effectively tune conduction in the sample. Nonetheless, the group of Leo Kouwenhoven at Delft boldly set out to make and measure the relevant structures, with support from not only traditional scientific funding agencies but also Microsoft, which hoped Majoranas might act as "topological qubits", forming the basis for a robust quantum computer. Quite soon, in 2012, Kouwenhoven and his team reported low-temperature electrical transport measurements showing some of the key expected signatures of Majoranas. This spurred great interest: theorists published papers proposing how to manipulate the state of topological qubits by moving one Majorana around another, Microsoft ramped up their efforts in topological quantum computing, several experimental groups reproduced Kouwenhoven's basic findings, and theorists offered alternative scenarios to explain the data (along with more experiments to test their ideas.) From the start, there has been controversy in the community about how to interpret the transport experiments, with the most prominent alternative explanations centering around less exotic Andreev bound states.

For the next five years, Kouwenhoven and others observed and studied a rich set of physics in structures similar to that in which Majoranas were first reported. Crucially, improvements in materials and interfaces (notably in the Copenhagen group also supported by Microsoft) made some electronic properties far cleaner. One prediction about electron flow through such structures had been known from the first theoretical proposals, and gained added weight over time. Under appropriate conditions, an electron could split into two Majoranas, one at each end of a superconducting nanowire. If it did, conduction through the nanowire should be enhanced at low applied voltages, indicating the presence of a state at zero energy (one of the Majoranas.) This enhanced conductance at zero bias was a central finding of the original 2012 experimental paper. But because an electron was equally shared between the two ends of the wire, **the conductance should be enhanced to a very special value derived from fundamental constants of nature, $2e^2/h$ or ~(13 kohm)$^{-1}$, the same conductance as a perfect 1D wire**[1]. Counterintuitively, this conductance should not depend on how strongly or weakly electrons are allowed to tunnel into the Majorana state from an external lead. In the 2012 paper and all follow-ons from Delft and elsewhere for 5 years, the zero-bias peak conductance remained far lower than this value. This discrepancy could be explained by finite-temperature effects. Nevertheless, by 2017 seeing the full $2e^2/h$ conductance would have been considered a holy grail or a new "smoking gun" demonstration that Majoranas had indeed been found (moving Majoranas around each other and seeing resulting interference phenomena would be at least as impressive, and is today considered a more conclusive target, but it too had proved challenging.)

Just seeing a much larger conductance would not be enough: its value should be precisely $2e^2/h$, and it should be stable against modest variations in magnetic field, electron density in the nanowire (tuned by a so-called "Super Gate", SG), and tunneling rate into the end of the wire (tuned by a

[1] The specific value, 2, of the prefactor is theoretically unambiguous, but explaining it is beyond the scope of this survey.

“Tunnel Gate”, TG.) Larger variations in any of these parameters should remove the conditions for formation of a well-isolated Majorana and thus depart from the “quantized” $2e^2/h$ conductance, either upward or downward. Nor would there be any guarantee that the right conditions could be achieved in a given “device”: each nanowire would have different dimensions, different electron density, different gate geometry. So failing to see a $2e^2/h$ zero bias conductance peak, stable with tuning parameters, should not indicate anything wrong with the simple and elegant Majorana theory: when one eventually made the right wire, and tuned it to the right regime, the Majorana should emerge. The paper in question reported success in this quest. According to the paper’s extended data section, by the time of publication over 60 nanowire devices had been fabricated, 11 measured extensively at ultralow temperatures, and 2 showed behavior resembling what was hoped for. Hence the paper title “Quantized Majorana Conductance.”

In this report we will examine the claims made about how the data resemble theoretical expectations, and how the selection and processing of the data influenced those resemblances.

Before moving to the detailed discussion below, the experts wish to note that they all feel that at the time of publication the samples and data reported represented a big and important step from earlier reports on zero-bias peaks (ZBPs) in superconductor/semiconductor hybrid nanowires, from the Delft group and others. Such earlier studies had shown ZBPs at only a fraction of ($2e^2/h$). By forming a more perfectly transparent junction between superconductor (Al) and semiconductor (InSb), the present authors raised this to 1x($2e^2/h$) and above. Roughly contemporaneously the Copenhagen group, which had in 2015 introduced this approach to junctions (epitaxial growth of Al on semiconductors, InAs in their case), also achieved ZBPs close to 1.0x($2e^2/h$), in narrow channels etched from 2D heterostructures with epitaxial Al atop InAs.[2] Their paper focused on temperature dependence of the ZBP whereas the Delft group’s paper focused on stability of ZBP conductance as a function of tuning parameters at low temperature.

The expert group received a large set of measurement data from the authors, from *devices A* and *B*, as well as other documents the authors had prepared (a) in response to the questions and concerns brought to them by members of the scientific community, (b) as part of the process of withdrawing the paper from *Nature*, or (c) in response to the experts’ questions. The authors were responsive to additional requests for data plotting. With the assistance of TU Delft staff, the experts verified (by spot checks) that the plotted data in both the paper and other documents shared with the experts were indeed drawn from the raw data files provided. The experts did not attempt to ascertain that these raw data files were those originally recorded at Delft (e.g. by checking time stamps, which were not preserved in the files provided), but we see no reason to question that assumption. The provided plots were studied carefully, discussed among the experts, and then discussed over video conference with the first author and separately with the corresponding author. The experts focused mainly on the data from *device A*, which formed the primary basis for the paper’s central claims.

We take the opportunity here to introduce two concepts we will refer to extensively below:

1. Plateaus: As noted above, tunneling into a Majorana mode at the end of a nanowire is supposed to yield a ZBP with a height 1x($2e^2/h$). Tracking this height as a function of B, TG, and SG should give a plateau in each case -- a constant section of finite extent -- as contrasted with a peak, a monotonic variation, or a more complex dependence. The plateau might occupy

---

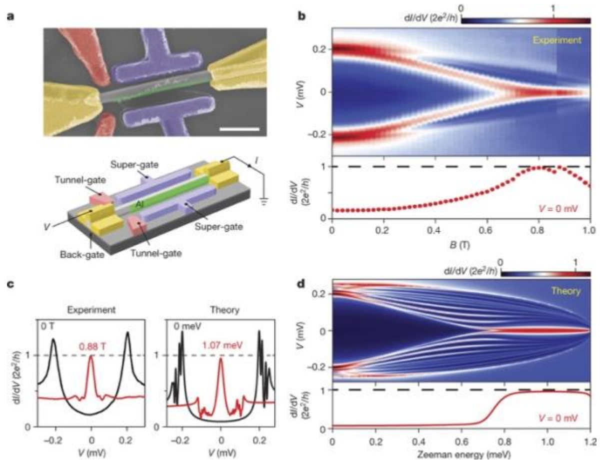
[2] Scaling of Majorana Zero-Bias Conductance Peaks, F. Nichele et al., Phys. Rev. Lett. 119, 136803 (2017)

a small region of this three-dimensional parameter space, but its presence and its quantitative value $1x(2e^2/h)$ play central roles in the paper's claims.

2. Charge jumps: These are changes in the state of localized charges near the conducting channel, which can move from one charge configuration to another over time or driven by tuned parameters such as voltages on gate electrodes or magnetic field. These configuration changes effectively reset the parameters of the conducting channel (tunnel barrier, chemical potential, detailed potential landscape.) The experts realize that charge jumps complicate measurements of most mesoscopic electronic devices. In conversation, the corresponding author stated, and the experts agree, that this issue is especially prevalent -- charge jumps occur more frequently -- in devices in which a high-quality dielectric is not available. This is the case in the work in question because of materials and processing constraints. Devices can change with time, or even "go bad" after some time, which can legitimately lead researchers to stop measurement on one device in favor of another device which might prove to have cleaner behavior. As noted above, the researchers reported making over 60 devices and extensively measuring 11 of them. In response to experts' questions, they explained that they discarded ones that displayed lithography flaws, failed to conduct, or carried supercurrent (indicating that the superconducting Al layer was not properly etched.) According to discussion with the authors, device measurement could only be done in series: only one dilution refrigerator was available on which only 1-2 working devices could typically be mounted at a time, and after unmounting a device and moving on to a new device they did not attempt to return to measuring a previous device, in part because they saw evidence of degradation of device properties over time at room temperature.

2. Fact finding about data presented in the paper, including context of full data set

Figure 1, in particular figure 1b



Many measurements of conductance versus control parameters were made on *device A* (even more were made on *device B*, but as noted above both the paper and this report focus on *device A*.) Each measurement, which lasts on the order of an hour, gives a 2D map of differential conductance as a function of bias voltage (vertical) and some other control parameter such as magnetic field (horizontal), as in the top half of figure 1b. 1D plots as in the bottom half of figure 1b are extracted from these 2D data sets. Most measurements as a function of B show that the linear conductance rises with B, reaches a maximum near B=0.8T, and falls off at higher B. The maximum values vary from about 0.8x($2e^2/h$) to about 1.3x($2e^2/h$). In their paper, the authors chose to present the data where, according to their conductance scale at that time, the conductance reached the value 1.0x($2e^2/h$). Among five data sets taken at the same time frame with this property, they further chose the only one which returns to 1.0x($2e^2/h$) at a slightly higher B after dipping down. (The other 4 sets are reproduced below in fig A) Referring to Fig. 1b, page 1 of the paper mentions “a robust ZBP” forms with increasing field, “reaches the quantized value of $2e^2/h$” and “the ZBP height remains close to $2e^2/h$ over a sizable range in B field (0.75–0.92 T).” The authors do not mention the other data in which the conductance as a function of field consistently rises to a single maximum and falls smoothly. Neither do they explicitly mention the many curves that do not go to exactly 1.0x($2e^2/h$) but instead reach substantially lower or higher conductance.

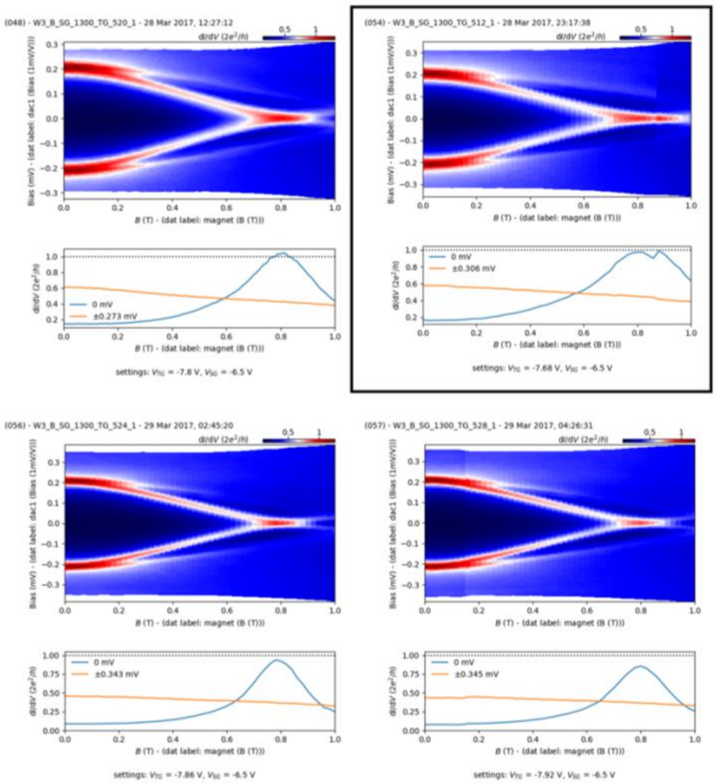
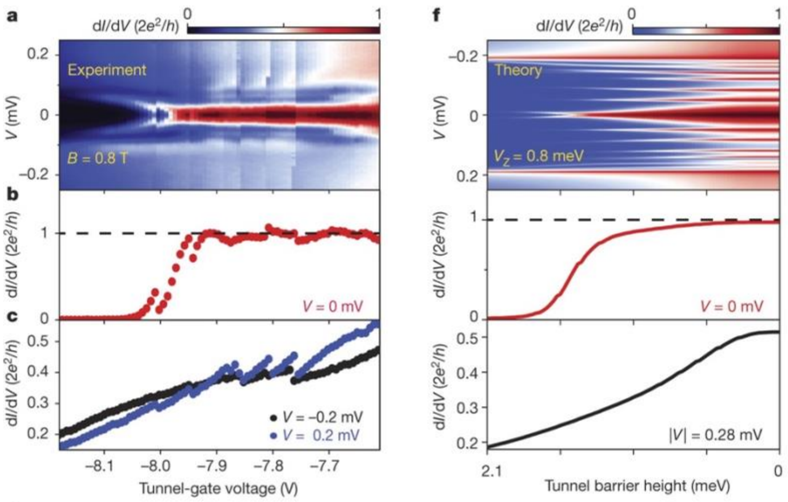


Fig A. A series of measurements of conductance vs magnetic field as the tunnel gate voltage was varied. Only the scan with the charge jump (in black box) and with a value closest to $2e^2/h$ was selected for publication as Fig 1b.

The experts interpret the abrupt return of conductance to 1.0x($2e^2$/h) as the result of a charge jump which has reset the parameters of the device, so that the new conductance vs B has a higher maximum value, from which it is now falling. In conversation with the experts, the corresponding author agreed that this is a charge jump. This is not remarked upon in the paper, though “switches” (charge jumps) are explicitly noted there in Fig. 2. A simulation (Fig. 1d) shows a broad plateau, which on cursory inspection bears more resemblance to the data of 1b than to otherwise similar data sets without a charge jump.

Figure 2b

This figure shows conductance as a function of Tunnel Gate voltage. There were several other measurements of conductance versus Tunnel Gate voltage in the full data set provided to the experts. In each of these measurements, super gate voltage and magnetic field were fixed. In many cases, the values of super gate voltage and magnetic field were different from those for the data of Figure 2. None of those other measurements gave a plateau as wide or flat as the one in Figure 2. The general trend was a fast rise in conductance as Tunnel Gate voltage is increased from very negative values, then a maximum conductance, then a slow decrease in conductance (plateau-like, with conductance varying by no more than 15%) over a range of 0.2 V, and finally a faster conductance decrease as the Tunnel Gate voltage becomes less negative. The experts interpret the relatively large (0.35V) range of the “plateau” in Fig 2b as a result of charge jumps repeatedly resetting the effective Tunnel Gate voltage, as can be seen in Fig 2c.



Below in Fig B (right panel) we show an example of another measurement of conductance vs Tunnel Gate at similar B, SG and TG voltage but measured at a different time than that used in the published Figure 2, to illustrate that the behavior described above is quite general. In particular the broad peak in the window of 0.12 V delimited by red vertical lines marking charge jumps at -7.8 and -7.68V looks similar to that in the window of 0.15V (between charge jumps at -7.75 and -7.6V) in the published Fig 2b shown above. Furthermore, we show in the left panel a magnetic field scan at the same SG and TG voltage which shows the same kind of broad peak described earlier in Fig A. There is no mention of these data in the paper. In the opinion of the experts, this omission is problematic because these data suggest that a conductance at 1.25x(2e2/h) has the same level of stability as a function of B and tunnel gate voltage (plateau-like, in the words of the authors) as the published data at 1.0x(2e2/h) once the effect of charge jumps in Figure 2b is removed. If the authors had chosen to publish these and similar data, the case for a quantized value at 1.0x(2e2/h) would have been seriously weakened.

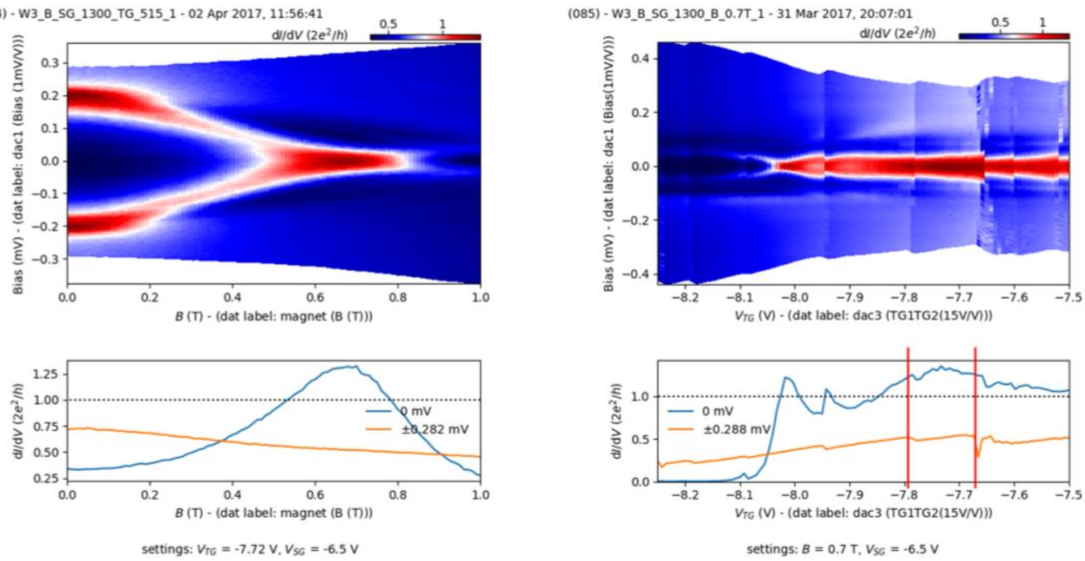


Fig B. Unpublished conductance measurements as a function of magnetic field at supergate voltage of -6.5 V, and as a function of supergate voltage at B=0.7T. In these parameter ranges, peak conductances considerably exceed the quantized value. These are to be compared with the published data Fig 2a,b and Fig 1b.

Comparing the original data for Figure 2b to the published data, members of the scientific community pointed out that the authors removed a part of the data and shifted the left part of the data to the right, contracting a range of Tunnel Gate voltage of about 0.03V. The authors acknowledged this in their response to members of the scientific community, and stated that the failure to describe this data processing in the paper (and the removal of this specific part of the data and not others) was an error in judgment. In discussion with us, the first author further explained that the removal of these data without any indication in the text was facilitated by a sloppy chain of data stewardship as figures were drafted. The authors do not think the removal significantly influences the interpretation of the data. The experts agree with that judgment, but do note that the other charge jumps left in the data give the misleading impression that the plateau is much wider than it really is as a function of effective Tunnel Gate voltage, something the corresponding author also acknowledged in his response to members of the scientific community. Also of note, the measured data for Tunnel Gate voltage less negative than -7.6V, where the conductivity starts decreasing, are not included in the published figure, subtly suggesting to readers that the plateau could be even wider.

In response to members of the scientific community, and already in dialogue with Referee 2, the corresponding author emphasized that the width of the plateau ”does not indicate much”; rather, he emphasized the relative flatness of zero-bias conductance even while high-bias differential conductance (a proxy for strength of tunneling into the wire) changed.

Figure 2d and 2e

The data for these figures are all represented in Figure 2a and/or 2b. The purpose indicated by the authors for panels d and e is to show that “although the ZBP width does change with GN [the ‘normal state’ conductance], the quantized height remains unaffected”. This is the same goal alluded to at the end of the discussion of 2b above, only here the proxy for tunneling strength is the width of the ZBP rather than the high-bias differential conductance. To support their contention, the authors selected data which showed zero bias peak conductance of exactly $2e^2/h$ (within a few percent) and plotted the width of the peaks. This is explained in the paper in the following way: “Figure 2d (red curves) shows several line-cuts of the quantized ZBP. The extracted height and width are plotted in Fig. 2e”. The experts consider this explanation unclear and the resulting Figure 2e misleading, especially since the discussion in the main text refers to Fig. 2d and the upper panel of Fig. 2e to make the point that the quantized peak height is found over a large range of normal state conductance G_N. With all data included, Figures 2d and 2e would have looked as in the versions (provided by the authors to members of the scientific community in February 2020 and then modified at our request) that we place directly below the originals.

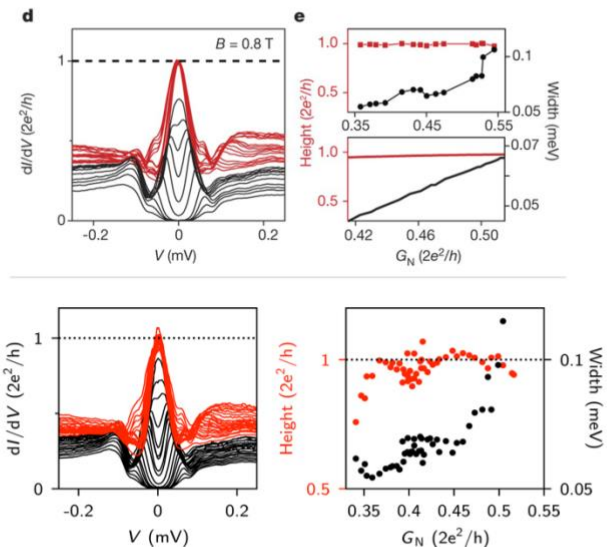
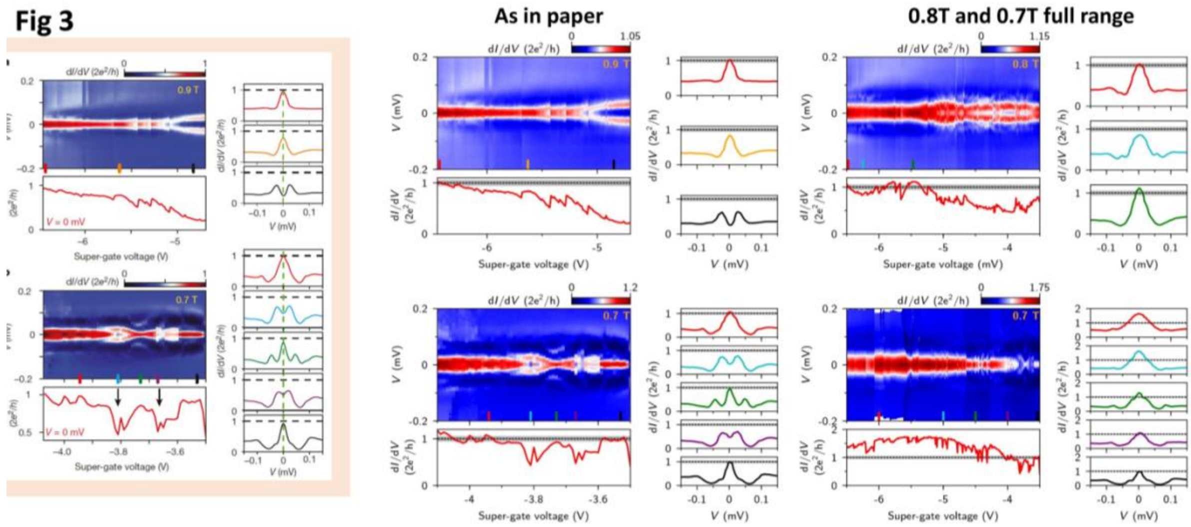


Figure 3

This figure shows the dependence of conductance on Super Gate voltage. Changing Super Gate is intended to change the chemical potential in the main body of the nanowire, but would normally have an undesired secondary effect of changing the transmission of the tunnel barrier. The authors note in the Figure 3 caption that Tunnel Gate voltage is adjusted to keep tunneling transmission constant (as reflected in roughly constant above-gap conductance.) Such simultaneous tuning of multiple gates to try to isolate adjusting just one parameter is a standard practice in the field. The values of Tunnel Gate at each edge of the Super Gate sweep are not stated, making it difficult to draw correspondence with other data sets which have stated, fixed Tunnel Gate settings (The compensated tunnel gate voltages were added in the plots provided to the experts.)

If one envisioned the left-hand side of the 0.9T data of Figure 3a as representing the same conditions as applied in Figures 1 and 2, the data show a gradual downward drift in conductance starting at -6.5 V with an average slope of about $-0.3x(2e^2/h)$ per V. The trend may indicate that the conductance simply crosses $(2e^2/h)$ at -6.5 eV, but no data are available for more negative gate voltage to support or refute this possibility. The authors view these data as supportive of a plateau in linear conductance as a function of Super Gate. Concerning the data at 0.7T, it was pointed out by members of the scientific community that the plot shows only a small section of the available data (compare lower middle to lower right panel below.) The full data show that for voltage more negative than the narrow range shown in the published figure, the conductance increases above $(2e^2/h)$, reaching a value as large as $1.7x(2e^2/h)$. Looking at the entire data set, the experts have the impression that the conductance just happened to cross $(2e^2/h)$ at -4.05 V, the point where the published data were truncated. The first author explained that the original purpose of the data cropping was to present a data set over which the tunneling rate remained constant: around -4.05 V a charge jump occurred together with a significant change in normal-state conductance. The experts consider it highly unfortunate that the information contained in the broader voltage range showing conductance considerably above $2e^2/h$ was not made available to the readers to form their own judgement. The reason is that for those who can now see the omitted data, these data show that the conductance is smoothly crossing the quantized value $2e^2/h$, as mentioned above, rather than stopping their rise right at $2e^2/h$ as suggested by the data provided in the paper.

The published data at 0.9 T and 0.7T are not the same magnetic field condition (B = 0.8T) as in the earlier figures. The available data at 0.8 T, and data over the full range of Super Gate at 0.7 T, are shown below. Middle column: plots for 0.9T (top) and 0.7T (bottom) almost the same as those in the paper (up to a recalibration of the conductance by ca. 10%). Right hand column: 0.9T data replaced with 0.8T, 0.7T extended to whole gate range.



The experts notice that the conductance at 0.8 T fluctuates around $2e^2/h$ with fluctuations of up to 25% in the Super Gate range -5 to -6.5 V. The data in Figures 1 and 2 were taken at Super Gate voltage -6.5V, where the conductance crosses $2e^2/h$, changing linearly with the Super Gate voltage with a rather large slope of $1x(2e^2/h)$ per V. The experts think that these data argue against a local plateau as a function of Super Gate voltage under conditions comparable to those in Figures 1 and 2, though the last sentence of the paper effectively indicates this as a necessary criterion for supporting the paper's central claim: "Only a stable quantized tunnel-conductance plateau, robust against variations in all gate voltages and magnetic field strength, can uniquely identify a topological Majorana zero-mode in tunnelling spectroscopy." It should be noted that in his conversation with the experts, the corresponding author disagreed with the experts' view, and continued to interpret the relative flatness of the conductance for SG between -6.5 and -5.5 V at B=0.8T as evidence for a plateau.

Figure 4

The data of conductance versus B are similar to those in figure 1, but for *device B* rather than *device A*. There are no data on conductance versus Tunnel Gate voltage: in the paper the horizontal axis is mislabeled as tunnel gate, whereas the measurement was actually versus Super gate voltage as recognized by the authors in earlier correspondence.

The conductance versus Super Gate voltage may show a plateau over a range of 0.06 V. It should be noted that this is a very small range on the scale of the Super Gate voltages shown in Fig 2, even though the devices are different and a precise comparison is not possible.

No further comments by the experts. As noted above, we focused our close analysis on data from *device A*.

Figure 5

Unlike the other figures, this figure is intended to show data that do not represent Majorana physics, to illustrate that data in non-Majorana regimes look different (these data are from *device C*.)

As in Figures 1 and 4 (for *devices A* and *B* respectively), a gap closes as in-plane magnetic field is applied, and the zero-bias conductance peak that emerges reaches around $1x2e^2/h$ The authors explain that this example shows an apparently quantized peak with similar stability with respect to B field and to gate voltage (back gate in device C and tunnel gate in device A) as sample A, but that it is not a Majorana zero-mode because it has no stability with respect to tunnel gate voltage. While the experts agree that the example shown in Fig. 5 clearly does not reflect a Majorana state, there is no additional information given in the paper that would allow the readers to decide whether device A and device C are examples of a continuum of behavior that can be found if a large parameter and sample space is searched or representatives of classes of behaviour.

3. The authors' methodology

In a 1974 commencement address to graduating students at Caltech, Richard Feynman shared his philosophy on how to approach scientific exploration: "*The first principle is that you must not fool yourself—and you are the easiest person to fool. So you have to be very careful about that. After you've not fooled yourself, it's easy not to fool other scientists. You just have to be honest in a conventional way after that.*" The experts wish to clearly state that we found no evidence of fabrication: all data in the publication seem to be genuine results of measurements[3]. However, the research program the authors set out on is particularly vulnerable to self-deception, and the authors did not guard against this as Feynman warned. We find that the authors have indeed fallen into the trap described by Feynman and have fooled themselves.

We paraphrase the authors' narrative, shared in discussion with the experts: *In their experiments they were motivated by theoretical predictions to search for quantized conductance at a level of $2e^2/h$. When they found that, they were exuberant and wanted to share their success with the community. Of course, they selected their best data to present, notably in Figures 1, 2b and 3, where other data sets could have been chosen instead. Nothing was intentionally hidden. In a full-length paper, they would have shown all the subtleties, but journals like Nature don't give space for that. The further subselection of data points in Figures 2d and 2e was meant to avoid the confusion and clutter that would have resulted if all data were plotted (as seen in alternate versions recently prepared by the authors and juxtaposed with those figure panels in the present report), and to highlight that zero bias conductance can have the same value even as peak width varies (where peak width is an alternative to normal state conductance, as a proxy for rate of tunneling into the wire.) They stated that, given the fact that the data in Figure 2e were extracted from those in Figure 2b which show broader variations, they did not anticipate that the horizontal red line would give a wrong impression to the readers, namely that the conductance is extremely stable at exactly $2e^2/h$ while changing the rate of tunneling into the wire.*

The experts wish to respond to this narrative.

The search for "best data" is central here. Every scientific team faces the challenge of curating their data, both what to measure and record and what to present and highlight. If an audience is given a comprehensive data set with no commentary, there is no story, nothing to learn, unless the audience does all the interpretive work themselves. To appreciate the search for "best data" in the context of the present article, recall from the introduction what the authors were seeking: a zero-bias conductance peak of precisely $2e^2/h$, stable against modest variations in magnetic field, electron density in the nanowire (tuned by a so-called "super gate", SG), and tunneling rate into the end of the wire (tuned by a "tunnel gate", TG.) There is no guarantee that the right conditions can be achieved in a given "device". Each nanowire would have different dimensions, different electron density, different gate geometry. So, failing to see a $2e^2/h$ zero bias conductance peak, stable with tuning parameters, should not indicate anything wrong with the simple and elegant Majorana theory: when one eventually made the right wire, and tuned it to the right regime, the Majorana should emerge.

Given this situation, the authors adopted the following strategy. First, make and measure many devices with nominally appropriate ingredients for correspondence to the idealized model:

[3] In one case (Figure 4), the parameter tuned was mislabeled. Throughout, the conductance values were slightly miscalibrated. Though each of these errors has significant implications for the narrative, each seems to have been inadvertent.

semiconductor nanowire with spin-orbit coupling, superconductor, tunnel gate, super gate, in-plane magnetic field. If a given device does not show properties coarsely resembling theoretical expectations (e.g. zero-bias conductance of order $2e^2/h$ at nonzero in-plane magnetic field), move on to the next device, as this one must not be easily tunable into a Majorana regime. Indeed, over sixty devices were made and many of them at least briefly measured. Once one has a promising device, quickly survey parameter space (magnetic field and two gate voltages) to find a parameter regime that yields a closed gap and zero-bias peak with conductance $1.0x(2e^2/h)$.

For the discussion of the data presented in the article, we focus on the final stage, once a device has been identified as promising for the potential observation of a conductance plateau heralding Majorana bound states. For this device, the experimental team has found a conductance close to $1.0x(2e^2/h)$ and seeks a maximum as a function of bias at zero bias, and local stability as a function of the three other parameters B, TG, SG. Here, our first comment is, that if one allows the local stability demand to devolve from requiring a finite-width plateau to merely finding zero or small partial derivatives with respect to some of the parameters , the target of the search is generic enough that it is likely to find some parameter setting where at least two partial derivatives are zero. Zero derivative (here: a maximum) with respect to magnetic field at some finite field seems a generic feature of *devices A, B,* and *C,* apparently explainable in terms of either Andreev bound states or Majoranas. Since the peak value varies from smaller to larger than $1.0x(2e^2/h)$, it is possible to adjust SG and possibly TG to get to a peak height of precisely $1.0x(2e^2/h)$, which the experimenters did. As a next step, the experimental team apparently "zoomed in" on such a "sweet spot" and took scans as a function of each parameter: B, TG, SG – as seen both in the actual sequence of data taken on *device A* and in the corresponding author's description in conversation with the experts.

The experimental team knew what they were seeking, the corresponding author told us, and of course would zero in on the promising regime rather than pursuing a long, exhaustive search. The corresponding author gave the telling analogy of a moon shot: when you set out for the moon and get to the moon once (here he meant: demonstrate a gap closed with magnetic field, a zero bias peak, $2e^2/h$ conductance, and insensitivity to B, TG, and SG) there's no question you've gotten there. A complication in these experiments is that the devices evolve with time and measurement history. Since only one partial derivative can be measured at a time, it was necessary to consider different parameter settings when measuring different partial derivatives. (Indeed, the authors looked at different values of B and different ranges of SG in going from Fig. 2 – sweeping TG – to Fig. 3 – sweeping SG.) As a result of the instability of a device, a consistent measurement of all three derivatives in the vicinity of the same point in parameter space turned out to be impossible. Given the difficulty of reproducibly verifying that partial derivatives are zero over a finite range in all three directions about a single point in parameter space, with this methodology it is easy to fall into the trap of focusing only on data that support one's goal, while ignoring warning signs presented by the rest of the data. In particular, we note that in our narrative (as opposed to that advanced by the authors in the paper) there was nothing special about $1.0x(2e^2/h)$, except that that is the conductance value the tuning procedure focused on.

Summarizing and expanding, the experts understand how the authors came to adopt the methodology outlined above given the experimental circumstances – notably, the instability of the devices, which precluded a careful high-precision mapping of the full parameter space – and given the specific scientific goal – establishing that a device can be fine-tuned to a "sweet spot" at which it shows a quantized conductance plateau. It would have been almost hopeless to do the search any other way, absent further improvements in fabrication control beyond the researchers' world-class state of the art at the time of the research, which already reflected years of sustained effort, driven by the complex combination of materials requirements. Nevertheless, the experts find the methodology dangerous, because it can easily lead to self-deception. Indeed, the methodology makes it likely that the self-deception will support the conclusions the researchers set out to reach: If most of the data as a function of fine gate sweeps are taken at a magnetic field at which the linear conductance peaks, and gates have already been used coarsely to tune conductance to around 1.0x($2e^2/h$), evidently plateaus can *only* be found at conductance 1.0x($2e^2/h$). Indeed, in the opinion of the experts, this self-deception is precisely what occurred. In addition to the methodology used, the self-deception and thereby deception of readers was likely facilitated by the way the data were presented and selected. This includes, but is not limited to, the authors' choice to select magnetic-field dependent data and tunnel-gate dependent data each with a maximum that is widened by one or more charge jumps; their choice to remove data showing another charge jump that might have alerted more readers to such jumps' ubiquity; their choice to crop data sets right where conductance crosses 1.0x($2e^2/h$) and their selection for further analysis points on the "plateau" that were tightly clustered at 1.0x($2e^2/h$).

If the observed conductance plateau were more dramatic, the methodology could in principle become sensible. For example, a conductance that sticks to $2e^2/h$ within a few percent over a sizeable range of magnetic field and both gate voltages (which is not what is seen in the present article) is unlikely to be a statistical artefact, but a conclusion that conductance is meaningfully stable at $2e^2/h$ may be fundamentally impossible to draw if the fluctuations around the target value are much larger and/or if (as in the present article) robustness is not established with respect to all relevant parameters at the same point in parameter space. Under such conditions an approach in which you only look for what you want to see may well be *inherently* flawed.

That it is possible to "zoom in" on the "wrong" plateau was demonstrated inadvertently in the present case: The authors recently reinvestigated and discovered that their conductance measurement had not been properly calibrated. After correcting for this calibration error, the remarkably well "quantized" conductance plateau of Fig. 2b and 2e turned out to be at 1.1 x $2e^2/h$ instead of 1.0 x $2e^2/h$, a difference that must be considered significant when seen in comparison to the size of the fluctuations around the plateau value (see two modified versions of top panel of original Figure 2e, reproduced below in Fig. C). The left panel, already presented during our first discussion of Figure 2e, uses the original calibration (but unlike the published version it includes data with conductance not precisely 1.0 x $2e^2/h$), whereas the right panel shows the same data with the corrected calibration applied. It is unlikely that the authors would ever have zoomed in on this particular plateau-like feature had they used the correct calibration when performing the experiment.

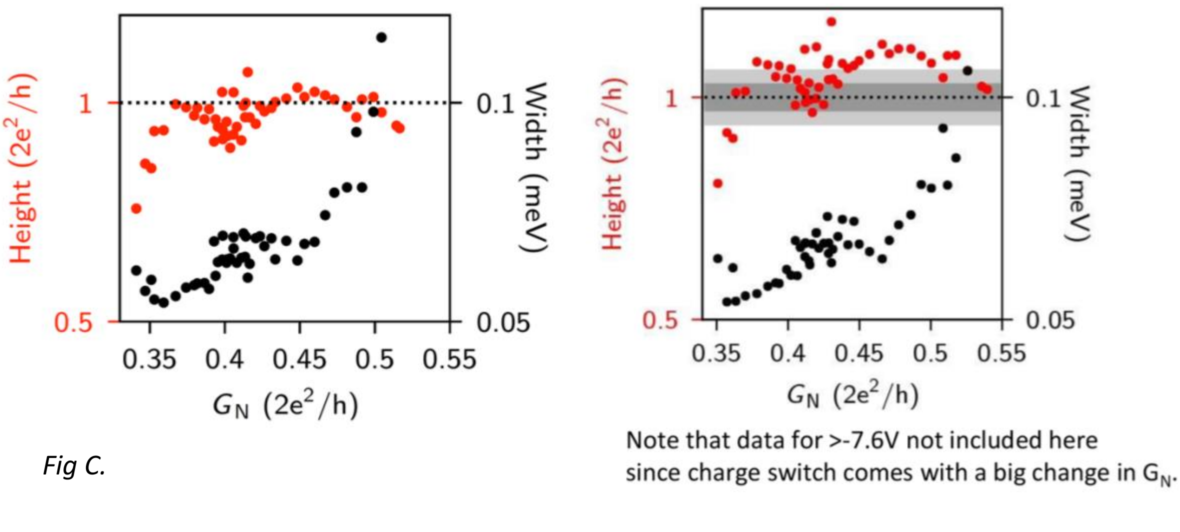


Fig C.

4. Conclusions from the data

As discussed above, one can speak about a true conductance plateau if the conductance does not change more than a small amount (1%? 5%?) over a finite though possibly small range of all relevant parameters. In the experiment the relevant parameters are the magnetic field B, the tunnel gate voltage TG, and the super gate voltage SG. The authors' approach was to establish the existence of a conductance plateau at zero bias around the "sweet spot" at B = 0.8 T, SG = -6.5 V, TG ≈ -7.7 V. We now discuss stability with respect to B, TG, and SG separately for *device A*.

Stability with respect to magnetic field B

The published magnetic-field dependent data at TG = -7.68 V and SG = -6.5 V merely shows a conductance maximum at $2e^2/h$, stretched by an artefact (probably a charge jump). The experts conclude that the *published* data do not support the claim of a conductance plateau vs. magnetic field B. Unpublished conductance vs. B traces at different but nearby values of SG also show a single maximum, with a height between ~ 0.8 and 1.3 (in units of $2e^2/h$). The unpublished data provide additional evidence that there is no conductance plateau as a function of B and, in addition, that there is nothing special about the value $2e^2/h$. At the least, these data show an *absence of evidence* of anything special about $2e^2/h$: the gate sweeps that would be needed to establish whether or not plateaus formed at other conductance levels were never called for by the authors' methodology.

Stability with respect to tunnel gate voltage TG

This claim may be considered supported by the published data in Fig. 2b, which shows the zero bias conductance vs. TG at B = 0.8 T, SG = -6.5 V, although the size of the conductance plateau (which may also be a broad maximum) is increased by a factor ~2 by artefacts (several charge jumps). The very high degree of quantization that seems to exist in Fig. 2e is caused by the authors' unfortunate selection of data points from Figure 2b and does not reflect the quality of the "plateau" in the unprocessed data of Fig. 2b. Unpublished data at nearby values of SG, notably a data set at SG = -6.8 V, show no plateau at $2e^2/h$, but instead a similar broad maximum at a slightly larger value of the conductance.

Stability with respect to super gate voltage SG

The article presents no data that indicate that the zero-bias conductance at the sweet spot does not vary with respect to SG. An unpublished conductance vs. SG trace at B = 0.8 T covers the range SG > -6.5 V only and has finite slope at SG = -6.5 V, which is not consistent with the existence of a local plateau near SG = -6.5 V.

Summarizing

Of the three necessary criteria for a conductance plateau, only one (robustness vs. TG) may be backed up by the data. The experts conclude that the data do not support the conclusion of a robust conductance plateau at $G = 2e^2/h$. With the data sets nominally available to the experimental team well before publication, the team could have arrived at the same conclusion. In fact, even from the selected data that were in the publication, a critical and dedicated observer could have noticed the widening of the peak in Figure 1b into an apparent plateau by a charge jump, the extension of the plateau in Figure 2b by several such jumps, the misleading representation of the data in Figure 2e, and the absence of any data showing stability with respect to SG. (Some observers did, which led to the present report.)

5. Concluding remarks

The experts conclude that the way the results in the manuscript were presented showed that the authors have selected data (conductance peaks reaching $2e^2/h$) which support the phenomenon that they were seeking, while omitting data that would have raised doubts in the reader's mind about their proclaimed success story. If this were done intentionally, it clearly would have been a serious offense. However, based on the material made available to them and after discussions with the authors, the experts did not find evidence for intent. Instead, they consider the most plausible explanation that the authors were caught up in the excitement of the moment, and were themselves blind to the data that did not fit the goal they were striving for. They have "fooled themselves" in the way forewarned by Feynman in the speech we quoted at the beginning of section 3.

The experts appreciated the openness of the authors concerning requests for additional data and for clarifications concerning the analysis. In discussion, it appeared that they understood the experts' concerns about how data were analyzed and presented in connection with the 2018 Nature paper in question.