Table of contents

Summary:

In March 2020 Science Magazine published "Flux-induced topological superconductivity in full-shell nanowires" by Vaitiekėnas *et al.* (https://science.sciencemag.org/content/367/6485/eaav3392). The article reported on a contribution to the field of one-dimensional (1D) topological superconductivity, a topic that has maintained interest for over a decade because of its potential relevance to future quantum technology. The Vaitiekėnas *et al.* paper combined theory and experimental components in an effort to establish full-shell superconductor/semiconductor nanowires with axial magnetic flux as a promising platform for the experimental realization of topological superconductivity.

Computation with 1D topological superconductors is achieved by braiding Majorana zero modes (MZMs), protected zero-energy quasiparticles that are localized near the wire ends. Experimental identification of MZMs both confirms topological superconductivity and establishes the starting point of a path toward topological quantum computation. The theory part of the Science paper demonstrated that the nanowire platform could potentially succeed, but also pointed to possible challenges - namely that topological and trivial phases share the theoretical model phase space sporadically and that in the models the gaps in topological states were typically small compared to gaps in the absence of a magnetic field. The experimental part of the paper presented evidence that the desired quasiparticles had nevertheless been realized in a batch of full-shell nanowires grown at the University of Copenhagen.

Full-shell semiconductor nanowires are completely encapsulated by a superconducting shell so that the superconducting order parameter's phase winding around the nanowire is quantized. As the strength of an axially oriented magnetic field is varied, superconductivity occurs in lobes distinguished by integer phase winding numbers, which are referred to as Little-Parks lobes. Readers of the Vaitiekėnas *et al.* paper were given the impression that large gap topological superconductors appeared regularly in non-zero Little-Parks lobes. Two different types of experimental evidence were provided, both related to the presence of MZMs: i) zero-bias conductance peaks in NIS tunnel spectroscopy experiments and ii) measurements of the energy difference between even and odd Coulomb blockade gaps in a series of devices with different lengths. We are not aware of any publications in which these experimental results have been reproduced, either by the original authors or by other researchers. We mention in Section 5 similar experiments conducted by other researchers that have led to different conclusions.

After questions about the paper were raised by two quantum wire experts, Sergey Frolov and Vincent Mourik, a request was made by Science Editor Jelena Stajic that the authors of Vaitiekėnas *et al.* share additional data. In response the authors uploaded substantial additional data to the Zenodo repository in November 2020, https://zenodo.org/record/4263106. Frolov and Mourik constructed a critical analysis of the Vaitiekėnas *et al.* paper based partly on the additional data. They shared their analysis publicly on Zenodo (Post-publication analysis, version 2 posted on Zenodo, Sept 30, 2021, 10.5281/zenodo.6344446). In this analysis they argued that the additional Zenodo data, and even data published with the original paper, contradict many of the statements made in the paper and therefore invalidate the paper's conclusions. They also argued that the data presented in a key figure of the paper could not possibly be representative of the full set of experimental data obtained by the authors prior to the original submission. In particular they suggested that additional data must exist that covers broader parameter ranges than were originally presented. The complaints of Frolov and Mourik prompted Science Editor Jake Yeston to publish an Editorial Expression of Concern on July 30th 2021 and to file a "Form for reporting suspicion of research misconduct or questionable research practices" to the Practice Committee at the University of Copenhagen.

In his complaint, Yeston frames the questionable practice issue as follows: "(...) whether the data presented in the original paper accurately represented the outcome of the experiments undertaken. Journal editors and reviewers can only assess data to which they have access. If data that did not support the claims in the paper were withheld or suppressed, then the paper submitted to the journal implied greater statistical support for its conclusions than the experiments in fact bore out. (...)the source of greatest concern is the range of voltages and number of independently tested devices that were represented in the paper's second figure. The editors at Science believe that an independent, transparent investigation by experts in this subfield of Majorana physics is necessary to ascertain whether or not the authors unethically withheld data that undermined the conclusions of their paper."

The three authors of this report have performed the requested investigation to the best of our ability. Our task was to reach a conclusion on two points:

- whether the data presented in the *Science Magazine* article accurately represented the outcome of the experiments undertaken, and
- whether the authors deliberately or due to gross negligence withheld data that undermined the conclusions of their paper.

To perform our investigation, we were provided all material related to the publication process and the additional data request at *Science*. We interviewed many of the parties involved and many independent experts on Majorana physics. We also examined all the data gathered in connection with the paper in question. (The dedicated help of NBI IT specialists was invaluable in accomplishing this task.) We summarize our impressions below.

Vaitiekėnas *et al.* carried out many experiments, built and studied sophisticated theoretical models, and analyzed considerable data in an effort to confirm the theoretical hypothesis – original to their paper - that topological superconductivity should appear in the non-zero Little-Parks lobes of quantum wires with axial magnetic field. The *Science* paper describes the outcome of measurements made on those full-shell nanowires that satisfied acceptability criteria designed to filter out devices with excessive disorder, devices with contacts or tunnel barrier tuning gates that failed, and devices in which NIS and Coulomb blockade could not be reliably interpreted because of active degrees of freedom (quantum dots) in the tunnel barriers. We are confident that we have seen all the data and that no data was fabricated. The acceptability criteria applied by the authors have sound scientific motivations. However, they were not explicitly stated either in the paper or in the supplementary material. The authors' success rates in fabricating devices that met the acceptability criteria were also not provided. Many nanowires with various diameters and shell thickness were studied, but excluded from further analysis after an initial examination of their electrical transport characteristics. Upon close examination of the full set of data accumulated by the authors for their paper, we found that the dividing line between qualified and non-qualified devices had a substantial gray area. In a few cases, data that were clearly not supportive of the paper's conclusions were disqualified, even though they were not very different from data deemed to be qualified. Qualified devices were in most cases measured carefully in several different Little-Park lobes. The *Science* publication reported on four NIS devices, three of which displayed evidence for Majorana physics. After reexamining all available data and using our independent judgement to classify data sets, we found that out of sixty measured devices from multiple wire batches, fifteen had successful NIS tunneling spectroscopy, and of these, seven exhibited evidence of Majorana physics. The MZM success statistics are therefore not quite as compelling in our independent reexamination of the full data set as in the Science paper, and much less compelling than suggested by the research article summary published in the print edition of *Science.* Similarly, after close examination, we have concluded that the dependence of even-odd splitting in Coulomb blockade on wire length is not quite as systematic as suggested in the *Science*

publication. In our opinion, the authors should have been more forthright and explicit with readers and with referees in describing their success rate in fabricating devices that showed simple tunneling characteristics and had MZM behavior and, by flagging alternatives and uncertainties, more evenhanded in their discussion of interpretations.

We thus arrived at the following conclusions:

- The presented data do, for the most part, represent the outcome of the experiments: the authors have exercised scientific judgement in selecting which data to share using criteria whose application was partially subjective. Although data selection did, in our view, result in conclusions that did not adequately capture the variability of outcomes, the excluded data did not undermine the paper's main conclusions.
- The shortcomings we have noted in this manuscript do not constitute gross negligence.
- We do not view the authors' behavior in connection with this paper as an instance of scientific misconduct.

On the basis of our investigation, we recommend that:

1- A statement explaining the set of criteria used to select acceptable nanowire devices, and a statistical summary of the success rate for growth and fabrication of devices deemed acceptable by these criteria, should be appended to the Vaitiekėnas *et al.* paper as a note added.

2- Unpublished data from the Coulomb blockade experiments should be made public.

And more generally that:

3- The nanoelectronics and low-dimensional electron system community, both its authors and its referees, should maintain high standards for fulsome objective reporting on technical details of sample fabrication, and on success rates in fabricating devices that exhibit the behavior described in any publication. These standards should be enforced by referees.

4- Prescreening of data to identify relevant regimes in a large parameter space should be exercised in a fully documented, transparent fashion, and discarded data should be made available to the community via a long-term data repository.

5- Journal Editors should make it more clear to readers which parts of the published material have gone through peer review and which parts are editorial addenda.

Full Report

1 Context for this panel review

Uniform semiconductor quantum wires proximitized by conventional superconductors are theoretically expected to be topologically non-trivial over finite ranges of system parameters and, when topological, to support Majorana zero modes (MZM). In finite length wires the energy of a MZM lies close to the superconductor's Fermi level and its wavefunction is spatially localized near the ends of the wire. The MZM energy is expected to approach the Fermi energy of the superconductor exponentially in the limit of long wires. There is great interest in efforts to reliably realize Majorana zero modes experimentally because of their potential utility as a resource for topological quantum computation. Convincing identification of Majorana zero modes is challenging because of the lack of smoking gun experimental probes. In addition, growth and fabrication of nanowire devices with sufficiently weak disorder and sufficiently strong proximity coupling to the superconductor to achieve properties similar to those assumed in theoretical models is far from routine.

The paper by Vaitiekėnas *et al.* introduced a new idea to the quest for MZMs, namely the possibility of using non-zero quantized phase windings in the superconductor to promote topological superconductivity. The theory component of the paper shows by explicit calculation that Majorana zero modes can in principle be induced by this mechanism when the wire geometry and semiconductor/superconductor interface properties are favorable. Calculations were performed on models that describe the semiconductor/superconductor hybrid system reasonably realistically, except that they do not account for disorder. The topological states that support MZMs are present over a good fraction of the model's parameter space, but tend to have gaps that are smaller than in the absence of a magnetic field. The theoretical models studied in Vaitiekėnas *et al.* suggest that topological states are likely to be realized only sporadically as wire parameters are varied because of a sensitive dependence of the topology of the ground state to system details.

The experimental part of the Vaitiekėnas *et al.* paper presents two different types of evidence (measured on distinct devices) for MZMs. i) Zero-bias conductance peaks in NIS transport experiments and ii) Even-odd splitting in Coulomb blockade oscillations that gets smaller with increasing wire length. The authors argue that together these observations provide strong evidence that topologically non-trivial states were realized experimentally in the NBI-grown full-shell proximitized nanowires that they studied.

2 Review of Tunnel Spectroscopy and Coulomb Blockade Experiments presented in the Science manuscript

2A Core-shell nanowires in an axial magnetic field.

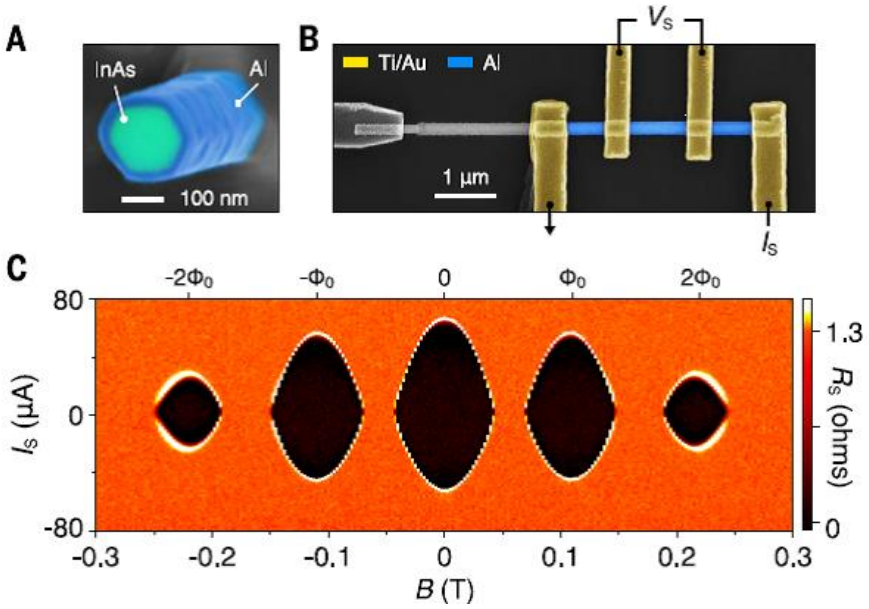


Figure 1: From Fig. 1 of Vaitiekėnas et al., presenting the core-shell nanowire and Little-Parks oscillations

The core-shell nanowires studied in the Vaitiekėnas *et al.* paper have an InAs core and an epitaxially grown Al shell. In full-shell wires, the superconductor is continuous around the core and the phase of its order parameter must therefore wind by an integer multiple of 2π. The superconducting gaps of full shell wires reach local maxima when the magnetic flux through the wire cross-section is close to an integer multiple of a flux quantum $\Phi = n\Phi_0$ = nh/2e, minimizing the frustration of states with the same integer winding number. The field regions with flux between (*n*-½) Φ_0 and (*n*+½) Φ_0 , *n* = 0, 1, 2, …, are referred to as the zeroth, first, second, *etc.* lobes. Superconductivity is weakened or destroyed when the flux through the core is close to half-integer flux quanta multiples. The periodic destruction and revival of superconductivity as flux quanta are threaded through the shell section constitute a Little-Parks effect and was reported in several previous publication on full-shell nanowires. The claim of the paper is that topological superconductivity can be induced in the non-zero lobes. This claim is supported by theory, in addition to the two experiments we now discuss.

2B Tunnel spectroscopy

The goal of the first type of experiment is to measure the density of states at one end of the core-shell nanowire via a NIS (normal/insulator/superconductor) tunneling experiment, in the configuration illustrated in the article in Fig 2A.

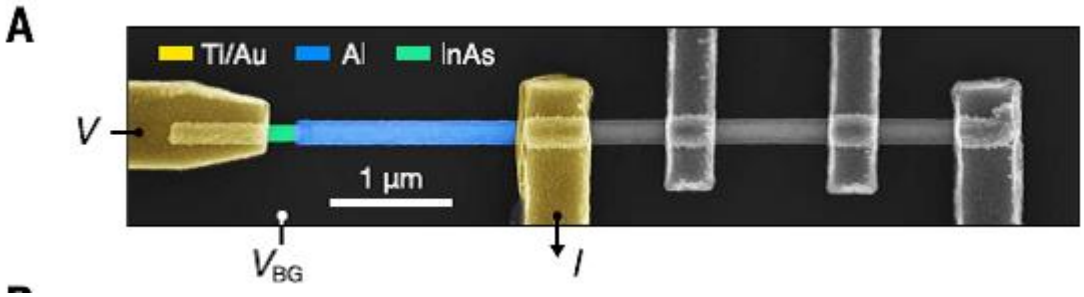


Figure 2: Figure 2A of Vaitiekėnas et al.

A non-superconducting (gold, colored yellow) electrode is connected to a short portion of the core-shell wire from which the superconducting shell has been etched (the green segment to the left end of the wire in the figure). A second non-superconducting electrode is connected to the wire shell further away. A back-gate voltage V_{BG} can be adjusted to change the conductivity of the exposed (green) core segment, thereby tuning the transparency of this barrier. Provided that the core segment acts like a simple tunnel barrier, the differential conductance dI/dV(V) of this configuration is proportional to the density of states at the junction between the core and the core-shell sections. A peak in differential conductance at zero voltage then indicates a peak in the tunneling density-of-states at zero energy at that end of the core-shell nanowire. (The tunneling density of states is related to energy changes associated with adding or removing an electron and to the imaginary part of the wire's one-particle thermal Green function.) The paper claims that in the zeroth lobe, the differential conductance is non-zero only at voltages above the superconducting gap, i.e. that the wire has a "hard gap in the zeroth lobe", whereas a zero-voltage conductance peak appears in the first lobe and in some higher lobes. The radically different behavior in the first lobe is interpreted as a signature of a Majorana zero mode, indicating that the system is in a topological phase. Other conductance peaks at higher subgap voltages are sometimes identified as analogs of the Caroli-Matricon-de Gennes states that occur inside superconducting vortices.

The figures supporting this claim are:

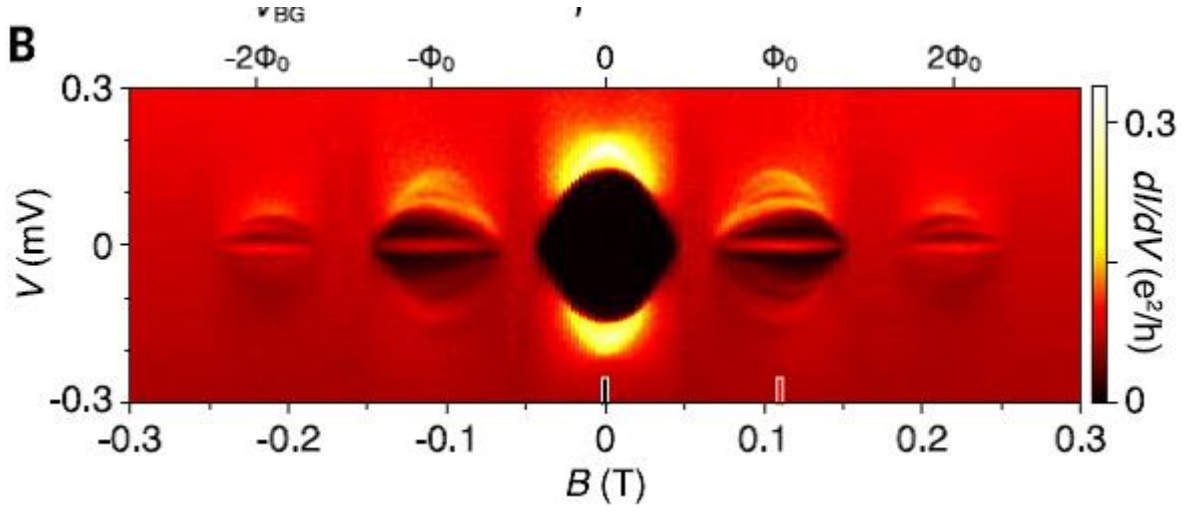


Figure 3: Figure 2B of Vaitiekėnas et al.

- Figure 2B, displays the color-coded conductance of device 1 as a function of bias voltage V and field (also labeled in units of flux quanta through the wire cross-section) measured at back gate voltage V_{BG}=-1.05 V. The zeroth lobe seems empty, whereas the first and second lobes display differential conductance peaks at zero voltage for both signs of magnetic field.

Figure 2C and 2D are meant to illustrate the absence of states in the gap, *i.e.* to demonstrate a so-called "hard gap", at zero field. Fig 2C is a color-coded plot of the differential conductance as a function of bias voltage V and gate voltage V_{BG}, in the range from -1.2 V to -0.8 V. Fig 2D plots differential conductance versus voltage at V_{BG}=-1.05 V. In this plot one can note a shoulder-like feature close to 0.15 mV, slightly below the maximum conductance corresponding, presumably, to the gap peak at 0.2 mV. Charge jumps in the effective back gate voltage are seen to occur around V_{BG}=-1.15, -0.95 and -0.9 V. This behavior is very common in nanodevices and is associated with uncontrolled changes in the distribution of background charges that are most likely located near the etched surfaces.

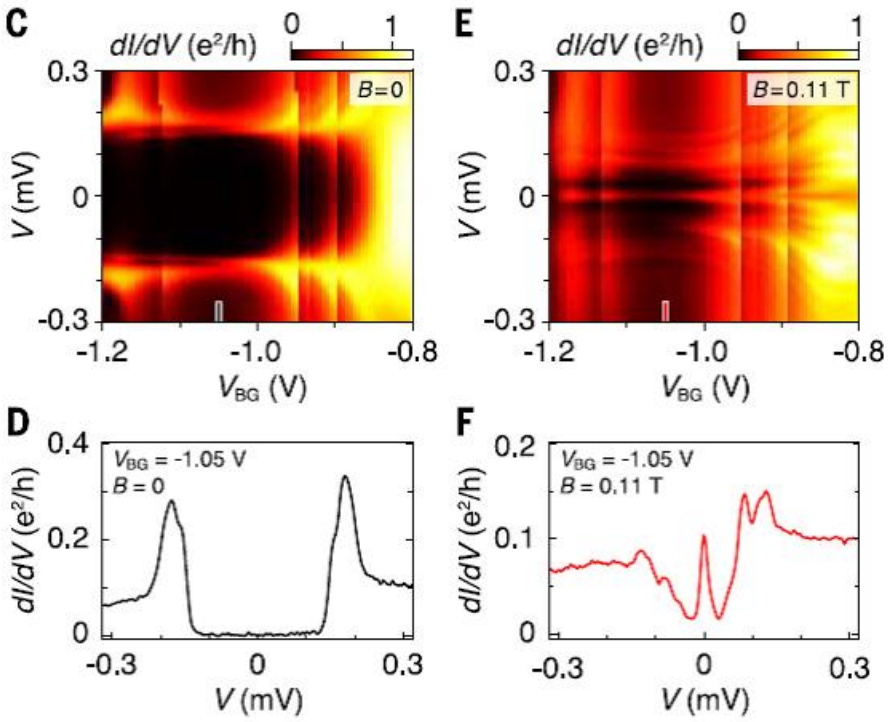


Figure 4 : Figs 2C,D,E,F of Vaitiekėnas et al.

- Figures 2E and 2F plot the same quantities as 2C and 2D, but for a field corresponding to approximately one flux quantum through the superconducting shell (winding number 1). Here the conductance peak at zero voltage is clearly visible, and seen to persist over a gate voltage range between -1.2 and -0.92 V. The zero bias peak seems to split between V_{BG}=-0.92 and -0.8 V. This behavior is described in the publication as a crossover from a zero bias peak to a zero bias dip. These figures are meant to illustrate the authors' claim that for non-zero, integer phase windings, topological superconductivity is present and manifested by a zero-energy density of states peak due to a Majorana fermion at the end of the full-shell hybrid nanowire.

2C Coulomb Blockade:

Coulomb blockade studies measure the dependence of the chemical potential of an isolated electronic system on electron number. When the tunnel barriers connecting the system to upstream and downstream contacts are tuned to approximately equal opacity, strong peaks in linear conductance appear when the energy needed to add an N^{th} electron to an island $\mu_N \equiv E_{oN} - E_{oN-1}$ is equal to the chemical potential of the leads. (E_{oN} is the ground state energy of an N electron island.) In small metallic islands $\mu_N - \mu_{N-1}$ is dominated by a Coulomb contribution that is expected to decrease

slowly with increasing island size L, approximately as L^{-1}. The term Coulomb blockade refers to suppression of transport over the chemical potentials ranges between the resonant peaks. The study of Vaitiekėnas *et al.* exploits a well-understood even-odd effect unique to superconductors, namely that odd N islands contain a single unpaired electron that must occupy a quasiparticle level. As a consequence, E_{oN} is increased by the smallest quasiparticle energy ε_0 when N is odd. μ_N is therefore increased by ε_0 for odd N and decreased by ε_0 for even N. This property of superconducting islands provides a second strategy for measuring small quasiparticle energies ε_0.

In Coulomb blockade studies, electrons are added to conducting islands by applying a plunger gate voltage V_G which lowers μ_N (relative to the source and drain chemical potentials: $\mu_N \rightarrow \mu_N - \lambda e V_G$, where λ is a dimensionless constant smaller than but of order 1. The lowest quasiparticle energy ε_0 can therefore be read off the pattern of gate voltage peaks: $2\varepsilon_0 = \lambda e|\Delta V_{GN} - \Delta V_{GN-1}|$ where V_{GN} is a gate voltage at which there is a conductance peak and $\Delta V_{GN} = V_{GN} - V_{GN-1}$. In the experiment by Vaitiekėnas *et al.* the dimensionless constant λ is read off the slopes of lines in Coulomb diamond color plots that capture the evolution of differential conductance peaks as source-drain bias voltage is varied. This method of measuring the minimum quasiparticle energy had been applied to hybrid quantum wires by both the Delft and the Copenhagen group in earlier papers. [J. Shen et al., Parity transitions in the superconducting ground state of hybrid InSb-Al Coulomb islands. Nat. Commun. 9, 4801 (2018).; S. M. Albrecht et al., Exponential protection of zero modes in Majorana islands. Nature 531, 206–209 (2016)].

Vaitiekėnas *et al.* applied the Coulomb blockade method to a series of full-shell islands fabricated on a single nanowire and summarized their results in Fig. 6 of the paper which is reproduced below in Figure 5. They studied islands with a range of device lengths from 210 to 970 nm, as illustrated in Fig. 6A of the manuscript. In all cases the Coulomb blockade spacings near zero magnetic field were doubled in the zeroth Little-Parks lobe, signaling that ε_0 in that lobe was larger than the Coulomb blockade energy and that only even N islands were stable as gate voltage V_G was varied, with the electron count jumping by two at each conductance peak. Results for the 210 nm device are summarized in Fig. 6B of the published paper.

Because the gap in the superconductor decreases with field when the wire radius is comparable to the coherence length, there is an interval of magnetic field (destructive regime) over which superconductivity is lost. The Coulomb blockade peaks were found to have 1e spacing when superconductivity reemerged in the first Little-Parks lobe, with the expected even-odd modulation from which the minimum quasiparticle energy could be extracted. The length dependence study reported on in the paper was performed at a particular magnetic field strength inside the first Little-Parks lobe. For each island, finite bias differential conductance measurements were also performed in order to determine λ for that island. Typical λ measurement data are illustrated in Fig. 6D (LP lobe 0) and Fig. 6E (LP lobe 1). Results for the magnetic-field dependence of the difference between the Coulomb blockade voltage spacings for even and odd N in the 210 nm device are illustrated in Fig. 6C of the paper. By comparing observations on devices with different lengths the authors conclude that $\varepsilon_0(L)$ is a monotonically decreasing function of wire length L, decreasing by approximately two orders of magnitude between the shortest and longest wires. They also conclude that the length dependence of this decline is incompatible with an L^{-1} or L^{-2} length dependence, but consistent with an exponential law, suggesting that the lowest energy quasiparticles are exponentially localized at the ends of the wires, implying that for each length L they are simple Majorana quasiparticles located at the wire ends, not MZMs or Andreev quasiparticles bound to some local potential fluctuations within the wires.

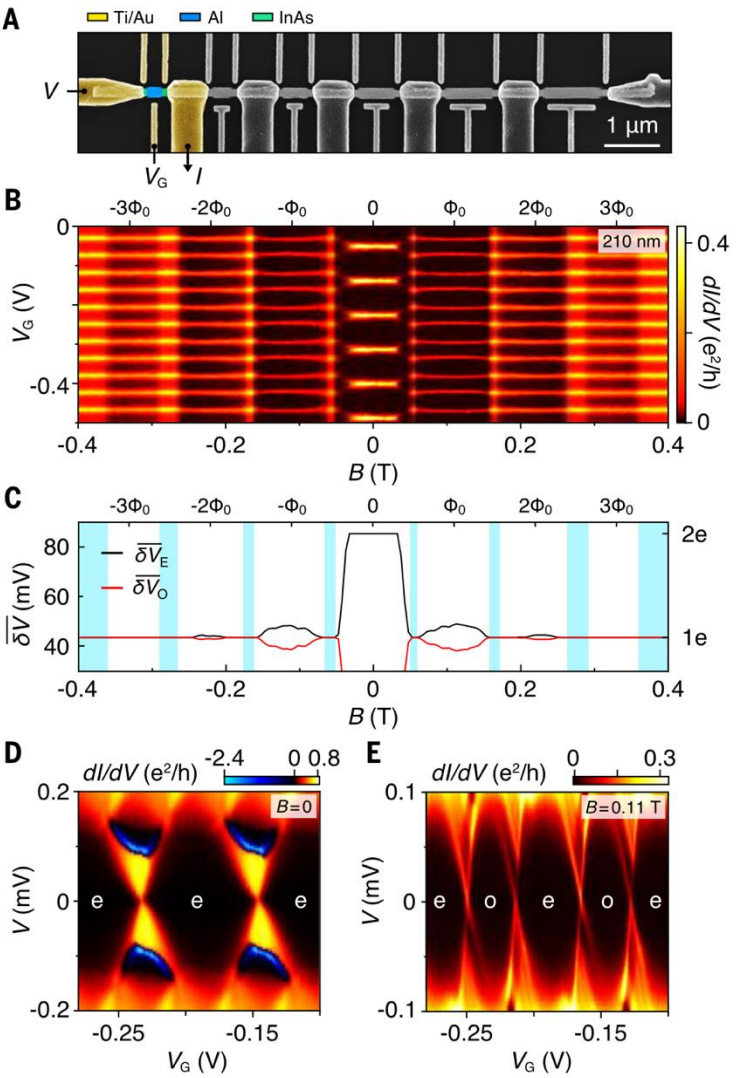


Fig. 6. Coulomb blockade: 2e peaks in the zeroth lobe, even-odd peaks in the first lobe. (**A**) Micrograph of device 2 comprising six islands with individual gates and leads, spanning a range of lengths from 210 to 970 nm. The measurement setup for the 210-nm segment is highlighted in color. (**B**) Zero-bias conductance for the 210-nm segment measured at 20 mK showing Coulomb blockade evolution as a function of plunger gate voltage, V_G, and axial magnetic field, B. (**C**) Average peak spacings for even (black) and odd (red) Coulomb valleys, $\overline{\delta V}$, from the data in (B) as a function of B, with destructive regimes shown in blue. Coulomb peaks spaced by $2e$ split in field and become $1e$-periodic around 55 mT. At higher field, odd Coulomb valleys shrink, reaching a minimum around 120 mT. In the second destructive regime, around 165 mT, peaks are $1e$-periodic again. (**D**) Zero-field conductance as a function of V and V_G, showing $2e$ Coulomb diamonds with even (e) valleys only. The negative differential conductance is associated with quasiparticle trapping on the island (see main text). (**E**) Similar to (D), but measured in the first lobe at B = 110 mT; the data reveal discrete, near-zero-energy state, even (e) and odd (o) valleys of different sizes, and alternating excited state structure.

Figure 5: Figure 6 of Vaitiekėnas et. al.

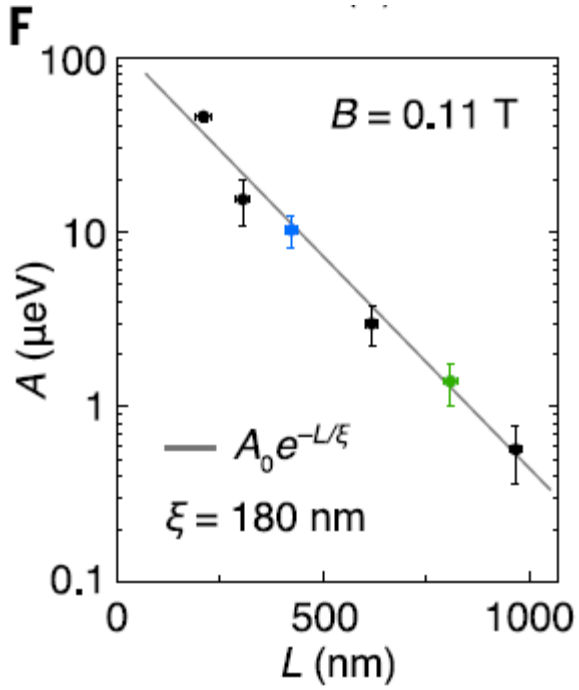


Figure 6: Figure 7F of Vaitiekėnas et al.

Figure 7F in the *Science paper* summarizes the dependence of the minimum quasiparticle energy on wire length extracted from Coulomb blockade measurements on six islands of lengths L between 200 and 1000 nm that were fabricated from a single nanowire (see Figure 6). The continuous line is the exponential fit obtained by the authors. The error bars express the variance of fits to measurements made on the same island with different magnetic fields and different source and drain contact tunings.

3 Review of additional data not provided with the manuscript

3A Additional data uploaded to Zenodo at the request of Science.

3A1 NIS devices: wider gate voltage range, bias voltage range, and additional devices

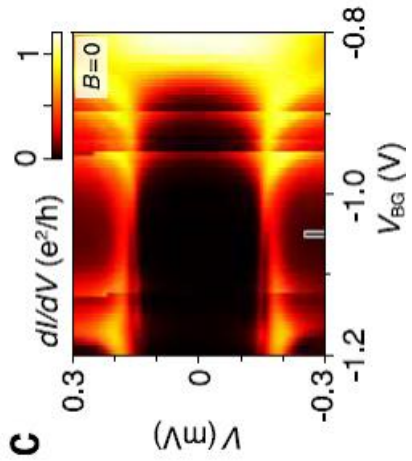


Figure 7: From Fig 2C of Vaitiekėnas et al, device 1, B=0, gate voltage range 1.2 V, bias voltage range +/-0.3 mV

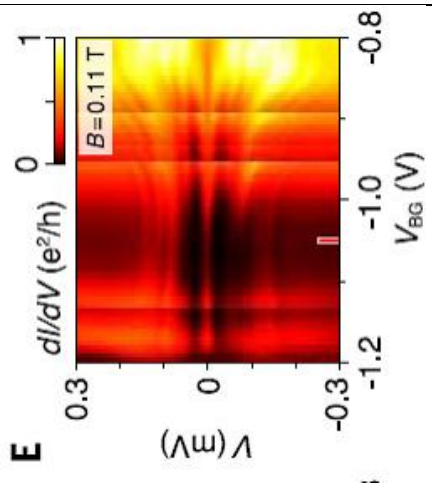


Figure 8: From Fig 2E of Vaitiekėnas et al, B=0.11 T, gate voltage range 1.2 V, bias voltage range +/-0.3 mV

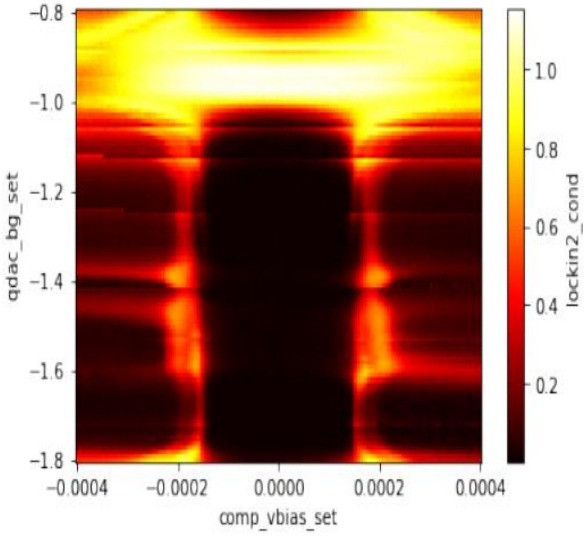


Figure 9: From Zenodo additional data, slide 53 of Summary.pdf, uploaded to the Zenodo repository at the request of the Science editor. Device 1, B = 0, gate voltage range 1 V, bias voltage range +/-0.4 mV

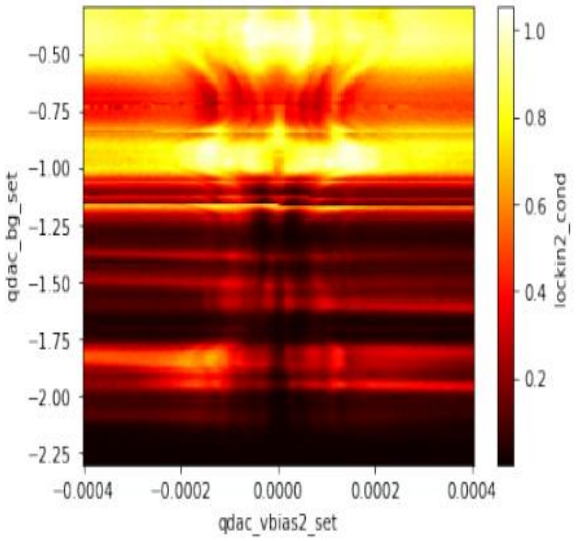


Figure 10: From Zenodo additional data, Summary.pdf, slide 55, Device 1 , B = 0.105T, gate voltage range 2 V, bias voltage range +/-0.4 mV

a Wider gate voltage range; definition of "tunneling regime"

The data released on Zenodo at the request of the responsible Science editor included results from measurements over a wider gate range than shown in the article. Figure 9 and Figure 10 above, displaying slides 53 and 55 from the authors' "summary.pdf" file uploaded on Zenodo, display differential conductance over 1 and 2 V gate voltage ranges respectively, much larger than the 0.4 V range of the article's Figs. 2C and 2F. The existence of this data taken prior to publication naturally raises questions about how the gate range that was shown in the paper was selected, and about the claim in the Science paper that the zero bias peak (ZBP) occurred "throughout the tunneling regime". Slides 51-54 in summary.pdf show that data was taken at zero field over a 3 V-wide back gate voltage

range, and slide 55 shows data taken at *B*=0.105 T (inside the first Little Parks lobe) that covers a 2V back gate voltage range. In the broader gate voltage region, there are extended regions (at least 1 V wide in slides 53 and 55) that display a hard gap at zero field but have conductance dips rather than peaks at zero bias in the first lobe. Since the high bias conductance over the wider voltage range is of the same order as in the narrower regime displayed in the main figure, one wonders why this region, which was not shown in the paper, was apparently deemed by the authors to be outside their "tunneling regime".

In discussions with the authors, the panel dug into their operational definition of "tunneling regime", and the protocol used to identify whether or not a specific device was in this regime at a specific back gate voltage. Indeed, as we learned, while the "tunneling regime" for quantum wires is often understood in a broad sense as meaning a regime in which the above-gap conductance of an interface is smaller than the conductance quantum, the authors' definition was narrower. The goal of the narrower definition was to exclude cases in which the NIS transport measurement cannot be interpreted as a measurement of the density of states on the superconducting side of the barrier. By the authors' definition, the "tunneling regime" is achieved only over a subset of the gate voltage range over which the conductance is sub quantum, and in some devices might not be achieved at all. For the authors, the tunneling regime is one in which the tunneling occurs through a featureless tunnel barrier - one that does not have internal degrees of freedom inside the barrier or other properties that might lead to tunneling I(V) features that are clearly not related to the density-of-states at the end of the wire. Only in this regime, the authors argued to the panel, should they trust the differential conductance measurements as a probe of Majorana physics.

This criterion excludes from consideration in particular those regions in which quantum dot–like features were visible in the tunneling spectroscopy. Such features include "resonances," i.e. conductance peaks at specific bias voltages, that are usually gate dependent. Andreev Bound states in a dot formed in the barrier are a likely source of subgap resonances. Above-gap resonances could be attributable to the Coulomb blockade behavior of such a dot. In the authors' judgment the "dotty" regions extend over wide low-conductance swaths corresponding to gate voltages more negative than those presented in the paper. In addition to using the narrow "tunneling regime" definition to identify gate voltage ranges in which the presence or absence of zero-bias conductance peaks can be safely identified with the presence or absence of MZMs, the authors employed the same criterion to quickly identify among the quantum wire devices they measured, which ones were likely to have a "tunneling regime" over some range of gate voltage and therefore be promising for detailed examination. Because of the multidimensional parameter space (field, back gate voltage, bias voltage), some protocol for the identification of promising devices was a practical necessity.

In the judgment of the committee the authors' criteria for a proper tunneling regime make physical sense. If the tunnel barrier is demonstrably complex, the additional NIS data over extended gate voltage ranges does not provide evidence against the author's conclusions even when it does not have a clear Majorana signature. However, we believe that the definition of tunneling regime intended by the authors should have been specified explicitly in the paper, and the range of gate voltage over which this tunneling regime is obtained compared to the total gate voltage range probed should have been spelled out in the supplementary material to allow readers (and referees) to make their own judgments. A plot with the broader range (in the Supplementary materials) would have given a clearer, more faithful, picture of the complex behavior of the family of devices studied for this publication.

The panel deliberated at length as to whether the omission of an explicit definition of "the tunneling regime" in the Science publication was a deliberate attempt to mislead readers as to the

certitude of the paper's conclusions, given the variability of NIS I(V) data. The key phrase "throughout the tunneling regime", used to describe where MZMs appear, is simplistic, and leaves less room for alternate interpretations than a more explicit qualification such as "over gate voltage ranges without tunneling resonances". The resonant states are mentioned in the supplementary material just in passing when discussing Fig. 7 "Tunneling spectroscopy without zero-bias peaks in a device with a thinner Al shell (device 5)": "For device 5, a discrete state crosses zero-energy around V_{BG} = 0.12 V and then again at 0.17 V, resembling a proximitized quantum dot state, similar to the one previously studied in Ref. [84], see Fig. S7. We usually associate such state with a resonant level in the barrier and, if possible, avoid it in the measurements." This phrasing might be understood by readers as suggesting that resonances are rare and easily avoided, and does not stand as a sufficiently strong warning on the importance of identifying and staying in the proper tunneling regime. Since the type of behavior seen in device 5 was seen in many quantum wire devices, the inclusion of this one example does not adequately convey the range of behaviors observed across quantum wire samples prepared using nominally identical growth and processing recipes.

b Larger bias voltage range.

The additional Zenodo material (for instance, slides 53 and 55 above) also shows that data was acquired over a voltage bias range of +/- 0.4 mV. The main text and supplementary materials display data over a bias voltage range restricted to +/- 0.3 mV. The authors of the paper were confident that their tunnel probe evaluation procedure properly selects regions free of resonances in the data. This confidence may have led them to overlook the importance of providing data up to higher bias voltages in the supplement of the original publication. The panel judges that, even though the omitted range of bias voltage is small, including the data over the entire voltage bias measured (at least in the Supplementary material) would have made resonant features more visible, thus providing the reader with better means to judge whether the selected regions correspond to a simple tunnel barrier.

Results for three other devices are given in the paper's Supplementary material. Two devices (device 3 and device 4) exhibit behavior similar to that of device 1 presented in the main text: a wide energy region in the zeroth lobe with no conductance (although a state close to the gap edge can sometimes be seen - see Fig. S6D for Device 4), and a zero bias peak in the first lobe (here however the width varies within the lobe). This data is displayed in Fig S5 and Fig S6. The third device, device 5, from a different wire growth batch, displays different behaviour. In particular there is no zero bias state in the first lobe.

Given the complexity of the studied devices, which unavoidably include disorder, the panel believes that data over the full range of parameters covered in the entire set of measurements (bias voltage, gate voltage) should have been summarized for readers, perhaps by including additional supplementary material.

3A2 Coulomb blockade devices

The authors checked several Coulomb blockade devices, but only three nanowires with altogether 16 islands were successfully investigated. The length dependence was determined for the one wire in which all six segments displayed tunable Coulomb blockade.

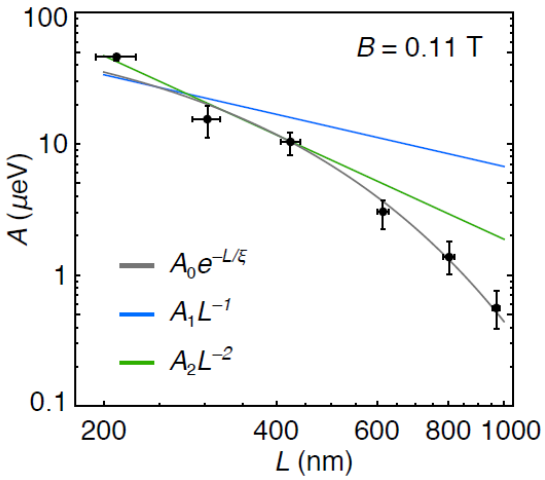


Fig. S18: **Exponential and power-law fits of peak spacing difference (device 2).** Same data as in main-text Fig. 7F: Average even-odd peak spacing difference, A, as a function of island length, L, measured at B = 110 mT. In case of topological bound states, the hybridization energy is expected to decay exponentially with L, whereas for trivial boundstates, the amplitude is expected to follow a power-law dependence. The best fit to the exponential form $A = A_0 e^{-L/\xi}$, gives $A_0 = (105 \pm 1)$ μeV and $\xi = (180 \pm 10)$ nm. For a particle in a box, energy level spacing scales as L^{-1}, and for a quantum harmonic oscillator—as L^{-2}. The best fit to $A = A_1 L^{-1}$ gives $A_1 = (7 \pm 1)$ meV nm, and to $A = A_2 L^{-2}$ gives $A_2 = (1.9 \pm 0.1)$ eV nm^2.

Figure 11: Fig. S18 from Supplementary Material

a Exponential versus power law for data presented in the paper

The Supplementary material (Fig. S18) compares the exponential fit to the length dependence of the even-odd splitting in Coulomb blockade to L^{-1} and L^{-2} power law fits (see Figure 11). Fig. S18 is intended to show that the exponential fit is superior to either power law fit, and argues that this constitutes evidence for MZMs. In order to assess the certitude of this conclusion, we have conducted a post-publication reexamination of the paper's Fig. 7F, and also examined additional data in the Zenodo post that was not analyzed for the original paper.

b Post-publication analysis of the data used for Fig. 7F

The paper's Fig. 7F (reproduced in this report's Figure 6) displays the variations with segment length of the energy difference between even and odd occupations of the segment, averaged over several data sets for each segment length.

The data sets differ in the sign of magnetic field, and in contact tuning. We display in Figure 12 below a plot in which each of the 25 data sets used to generate Fig. 7 is represented, illustrating the scatter in the data. This analysis, performed independently in summer 2021 by Bernard van Heck, agrees with the plot of Fig. 7F, which is also included in the figure. The points for the plot in the paper are obtained by calculating the mean and standard deviation of the points in the plot below and setting $1/N^{1/2}$ as the error. This analysis finds that exponential decay fits the data better than any power law. The best-fit power law has a power of -2.49 and a χ^2 twice as large as the exponential fit.

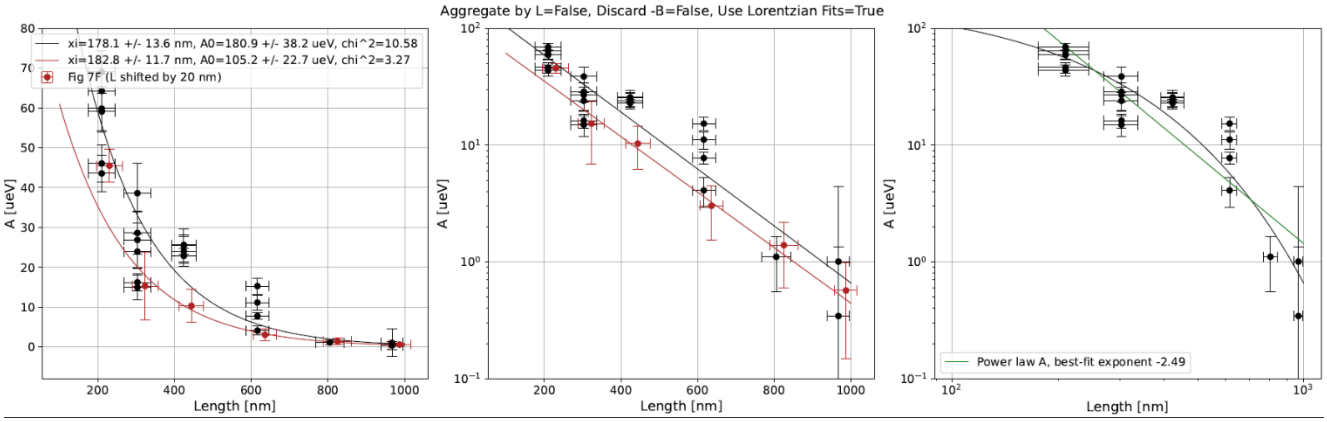


Figure 12 Post-publication analysis of the original Coulomb blockade data (25 files) performed by Bernard van Heck (black points). The published data points with error estimates are shown in red (length value increased by 20 nm for clarity).

On the basis of the full CB data sets of Vaitiekėnas et al. and the extended analysis, the panel concludes that there is indeed unambiguous evidence for a strong decrease in the even-odd effect with increasing length in the studied nanowires. The behavior is close to exponential, but in the light of scatter in the raw even/odd asymmetry values, one cannot exclude an inverse power law dependence with an exponent in the range of 2-3. The exact functional dependence is a matter of future debate. When discussing with experts, the panel found contrasting views on whether fluctuations in the even-odd spacings are to be expected. In theoretical models the MZM energy has oscillatory dependence on $k_F L$, where the k_F are Fermi wavevectors of occupied quantum wire channels. Some experts expect these dependences to be washed out by disorder. If present, they could explain the dependence of the (presumed) MZM energies on contact tuning that is evident in the data scatter.

c Additional data

Some Coulomb blockade data was excluded from the original analysis because the measurements were deemed to be unsuccessful. The panel was provided with the full set of data measured on the Coulomb islands of the same wire, and asked the authors to perform an analysis of all data that was not included in the original analysis. 56 data files were added to the 25 files originally analyzed. Appendix IX contains a list of the additional Coulomb blockade data sets that were compiled along with a brief explanation for their exclusion in the Science paper analysis. The additional data sets were assembled and analyzed with help from the authors; most of the additional data analysis was performed by Bernard van Heck and leads to the summary in Figure 13. We also performed an independent analysis of most of these data sets, with the assistance of Chao Lei (postdoctoral researcher UTexas) as a check.

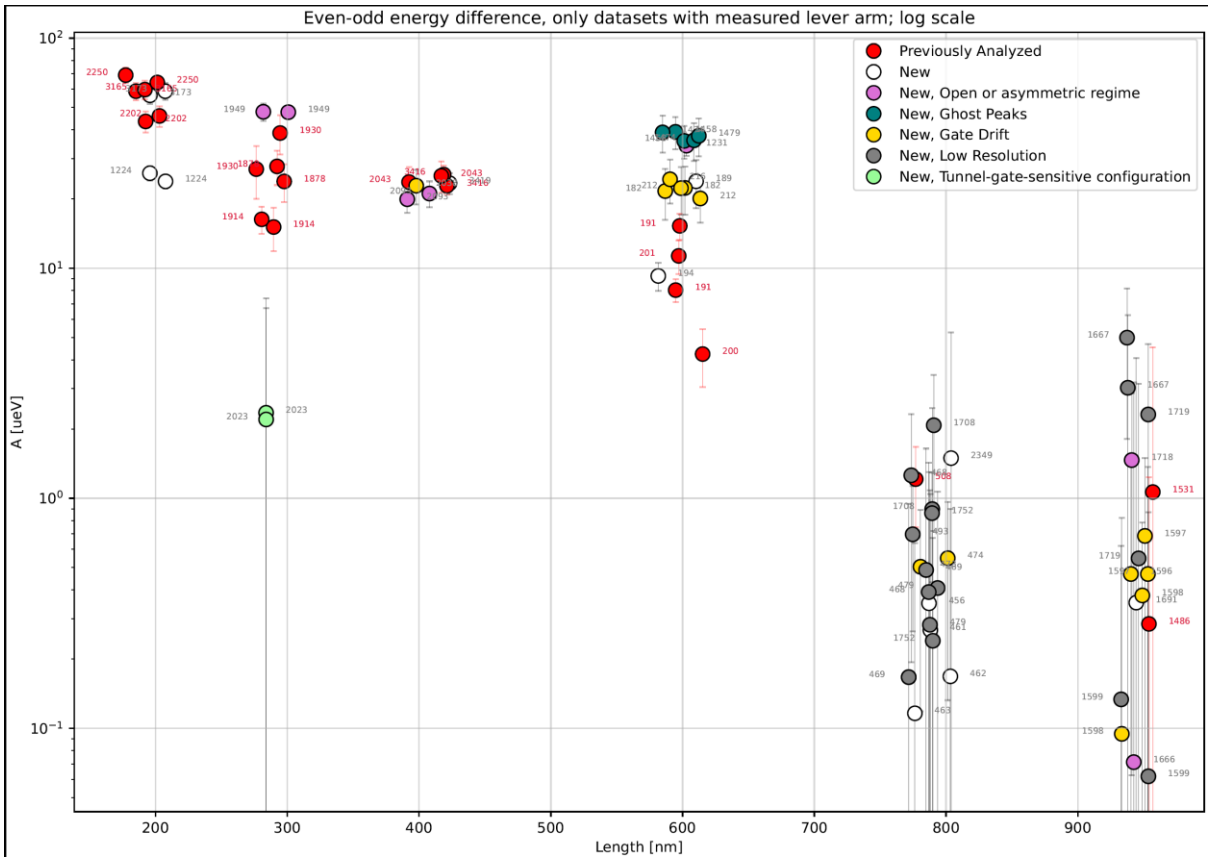


Figure 13: Difference between average even and odd Coulomb blockade energy spacings for the 25 data sets included in the original analysis (red) and 56 additional data sets. Most of these were excluded from the analysis in the Science paper because they did not satisfy selection criteria: purple – dependence on sign of field or large valley conductances; green – presence of ghost peaks that suggest fluctuations in particle number; yellow –drift in peak position in gate voltage during field sweep; blue – data taken with low gate voltage resolution that prevents accurate estimates of peak voltages; green – large, unidentified sensitivity to tunnel gate configuration, white – newly analyzed data that does satisfy selection criteria. This figure was compiled by Bernard van Heck, following discussions with the authors and the committee.

The 25 data files used for the original CB analysis in the Science paper were chosen based on a strict selection protocol. In the protocol, the authors insisted on the importance of tuning the devices to symmetric barriers and on low conductance in the blockade regions. The presence of ghost lines at zero magnetic field, even if minute, was also used to exclude data from the sets to be analyzed. In the panel's view such data, if taken with proper tunnel barriers that pass the protocol, should on general grounds display just a slightly reduced even/odd effect. It turns out that when analyzed most of the initially excluded data give larger even/odd effects, contrary to expectations. Of course, one may argue that these data are inferior to the pure protocol-based selection, but for the panel, the discarded data carries definite information value. This additional information does not however invalidate the finding that there is strong decrease of the even/odd effect with nanowire length. It does however, in the panel's view, undermine the claim that exponential decay of the splitting with L has been unambiguously established.

The CB data in the main article was for the most part collected in two stages: data on short samples were measured in the first cooldown, while longer samples were measured in a later cooldown. The gate voltage values in the two cooldowns were quite different, suggesting that some details of the doping profile may have changed between cooldowns. This raised the question of whether the rate of decay of the even-odd energy difference was influenced by parsing the two data subsets together. The panel asked the authors to identify and mark data points measured in different cooldowns. The new extended analysis suggests that the Coulomb blockade results are robust, and not sensitive to the change in doping profile between cooldowns.

In the panel's view, the conclusion that the even-odd splitting declines rapidly with wire length stands with the additional data included. There are two main changes. First, the scatter in the extended analysis is particularly large for the data for the 600 nm-long segment. If the new data is meaningful, its larger scatter emphasizes the fact that uncontrolled factors other than the segment length (for instance the Fermi wavelength) may influence the spacing.

Second, one contact tuning that was excluded for the 300 nm-long segment gives an even/odd effect that is one order of magnitude smaller than the average reported for that length in the Science manuscript. This discrepancy is apparent in the raw data, even before a careful analysis is conducted. The panel found no compelling reason for excluding that particular data set, which appears to be an unexplained outlier.

In discussions with the authors, it was often mentioned that they knew of no other explanation of the combined exponential dependence and NIS tunneling data than the MZM hypothesis. However, looking at the black data points in Figure 12 for example, it is hard to exclude inverse power law fits with exponents ~2.5. Hence, the claim of "Incompatibility with a power law" in the Science manuscript seems inaccurate. Statements such as "Incompatibility with an inverse power law with exponent smaller than 2" or "Incompatibility with a physically realistic power law" would have been more acceptable summary statements for the observations. Once the additional data is included, it appears that the Coulomb blockade spacing has a strong dependence on factors other than the length. Still, unlike the IST data on nominally similar quantum wires (described in part 5B), these observations show a clear trend toward smaller spacing with increasing length.

The panel suggests that the extended data analysis be deposited to Zenodo, along with the accompanying comments from the authors that identify datasets that were not entirely compliant with the protocol and therefore excluded in the original analysis. The original Coulomb blockade data files whose numbers are indicated in Figure 13 should also be deposited to Zenodo. This will allow researchers in the field to make their own judgments concerning the robustness of the conclusion.

3B Data from devices not shown in the paper or the supplementary material

Most of the data that was taken prior to publication but not represented in either the paper or its supplementary material is on NIS devices.

3B1 NIS devices uploaded on Zenodo after the request by the Science editor:

In addition to the devices 1, 3, 4 and 5 shown in the paper and the Supplementary Material, additional devices 8, 9, 10, 11, 12 were shared. Except for device 12, all display a zero bias peak in a large part of the first lobe. The ZBP features are less clear in these devices (sometimes broadened, sometimes split, sometimes not appearing throughout the lobe, or appearing in a region that displays resonances at zero field). When going over the measurements with the authors, the panel confirmed that device 12 displays no ZBP in the first lobe even though a featureless tunneling region is present at zero field. The panel and the authors also agreed that the ZBP displayed by device 9 was in a region that could be considered outside the tunneling regime, because a quantum dot may have formed in the barrier. The panel therefore considers that the statistics in the Zenodo file should be slightly modified, to six devices (from batch 439) in the tunneling regime with a ZBP (1,3,4,8,10,11), and two devices (device 12 from batch 439 and device 5 from batch 638) in the tunneling regime without a ZBP.

3B2 NIS devices not uploaded on Zenodo after the request by the Science editor, but made available to the panel

Except for one device (sample 5 from batch 638), all samples mentioned in the paper and the additional posted data came from a single nanowire growth batch, batch 439. The authors did not mention the existence of the many other devices they measured, which were fabricated using nanowires from a total of five different growth batches (numbered 439, 564, 637, 638, 829, grown over two years). The data taken on the other devices were organized with the help of NBI data scientist Rasmus Bjerregaard and made available in their entirety to the panel prior to a visit to the NBI by two of the panel members (SG and PH). These data were presented and discussed during the visit. The panel found that a majority of these additional devices did not "work", meaning that they did not satisfy the criterion of a good tunneling regime, and were therefore discarded. Different types of "bad behavior" could be observed. Some devices displayed resonances as soon as the gate voltage was tuned to a lower conductance, a behavior termed "dotty behavior" since it is most likely due to a quantum dot forming in the bare segment between the hybrid core-shell wire and the contact electrode. Some could not be "pinched off", meaning that the conductance could not be reduced sufficiently to access a tunneling regime. Others with a smaller diameter required a magnetic field to thread one flux quantum that was so large that the superconducting gap was suppressed considerably, which drove the subgap conductance features closer together, making them less distinguishable. In some samples, the contact from the Ti/Au lead to the nanowire was deemed bad, as characterized by a zero bias conductance dip ascribed to dynamical Coulomb blockade of the tunneling current (i.e. modes in the resistive environment which influence tunneling). Summarizing, out of the sixty-four NIS devices fabricated, twenty-eight devices were unsuccessful and were not studied in a systematic way (as described in the protocol); and twenty devices displayed no clear spectroscopic features or did not satisfy the protocol. The panel agrees with the authors that these failed and inconclusive experiments, which were not discussed explicitly, do not weaken conclusions based on the behavior of the samples with sound tunneling characteristics.

However, the panel also found that amongst devices that displayed good tunneling characteristics according to the protocol, i.e. displayed Little Parks lobes and had clear spectroscopy, some did not display a ZBP in the first LP lobe. In addition to device 5 (from growth batch 638) reported on in the Supplementary Materials of the manuscript, and device 12 (from growth batch 439) reported on in Zenodo, six devices from three different growth batches behaved in such a way: device 73 from growth batch 564, devices 20 and 23 from batch 638, and devices 59, 65, 68 from growth batch 829. Figure 14 displays data taken on such a device (device 65). Thus, amongst devices that are deemed acceptable because they follow the criteria of a good tunnel barrier (as described in the protocol, see appendix), seven displayed a ZBP in the first lobe (six from batch 439 and one from batch 829) and eight did not. A table summarizing the outcome of the NIS spectroscopy, prepared by the NBI team after jointly reviewing the data with two panel members, is shown in Appendix VI.

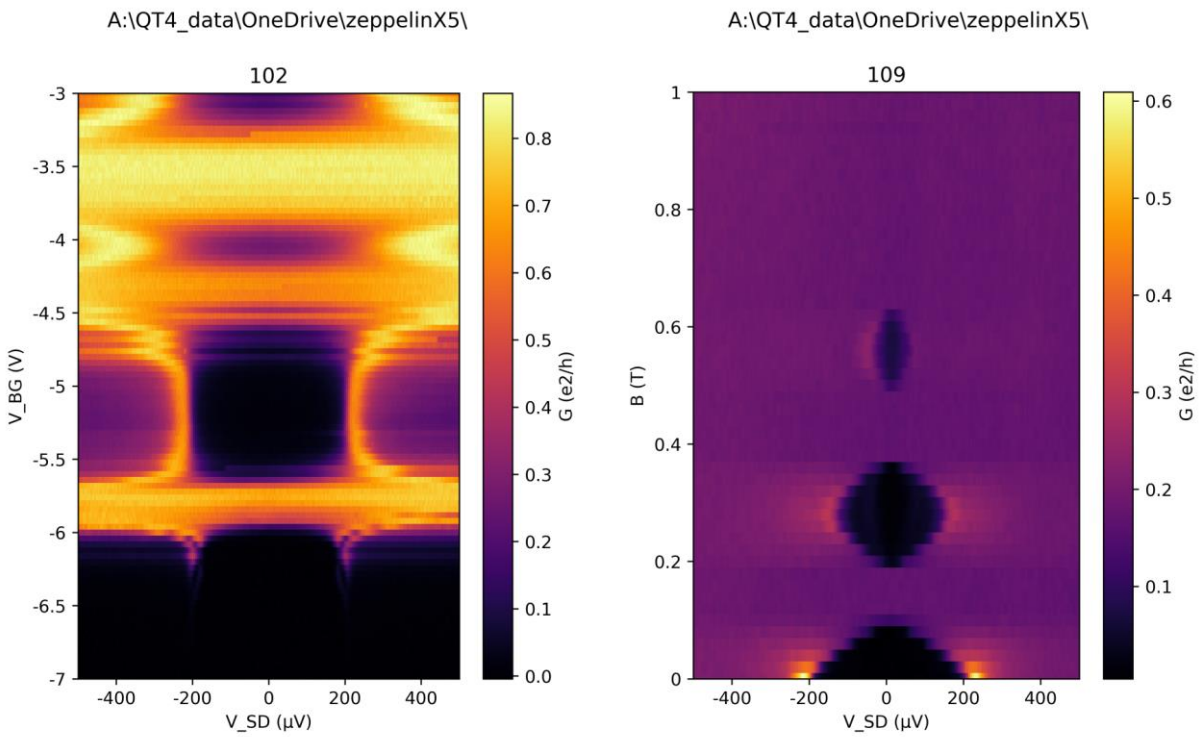


Figure 14 Example of a device with clear spectroscopy but no zero bias peak in the first LP lobe. (device 65, batch #829). Left, color-coded differential conductance at zero magnetic field as a function of back gate and source-drain voltages. Right, color-coded differential conductance as a function of voltage bias and magnetic field, measured at V_{BG}=-5.1 V. There are no low energy (subgap) features in the zeroth lobe, and no zero bias peak in the first lobe. The flux period corresponds to a nanowire radius of 48 nm.

The panel believes that the fact that devices with good tunneling characteristics from entire batches fail to display the zero bias peak is not *per se* an invalidation of the idea that topological superconductivity is possible in core-shell nanowires. Rather, the panel views this as confirmation of the theoretical result of the paper, which predicts a complex topological phase diagram in its wire-parameter model space, as summarized in Fig. 4 D of the paper, reproduced below. It is very possible for instance that the nanowires grown in an entire batch possess aluminum shells that are too thin to be in a region with topological superconductivity, as suggested by the authors in a comment about device 5. However, the panel believes that the *Science* paper should have provided a full accounting of devices that were valid according to the measurement and selection protocol, but did not display the behavior presented in the manuscript.

An additional realization that appeared as the panel was reviewing data with the authors was that it is often not so clear in practice whether a given device displays a good tunneling regime or not. In particular, a dot in the barrier with double occupancy can display a rather featureless behavior in gate voltage, and therefore be mistaken as a “good” region in which tunneling into the nanowire occurs directly.

Given that the article reports three devices (devices 1, 3 and 4) with a ZBP and one device (device 5) without a ZBP, and that the full NIS data points to seven devices with a ZBP and eight without, the statistical support suggested for the ZBP in the article was not grossly exaggerated. The number of unsuccessful devices (that did not satisfy the protocol), however, was clearly underreported.

different band offsets. (**C**) Topological phase diagram of the full-shell nanowire in the first lobe at B = 0.124 T as a function U_0 and spin-orbit coupling, α, close to the m_J = 0 topological phase. The gray lines indicate a change of the sign of the Pfaffian, Q. (**D**) Topological phase diagram for the same set of parameters as in (C), over a large range of band offsets. (**E**) Topological

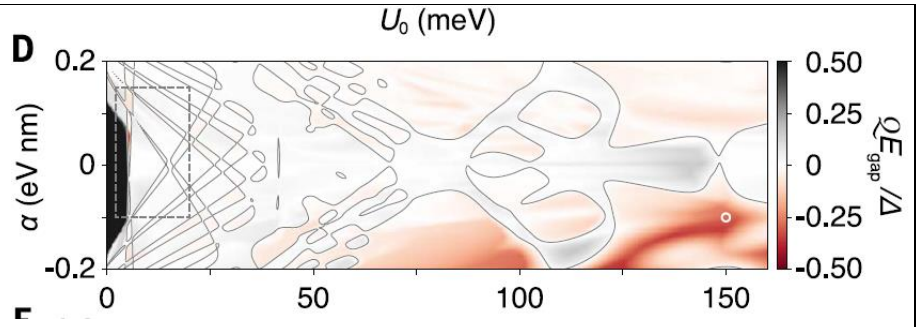


Figure 15 Figure 4 D of the manuscript, evidencing the fact that the regions of topological superconductivity are rather sporadic.

3B3 Coulomb blockade devices

Only one device had all seven segments working. The panel agrees with the decision of the authors to focus on that device.

4 Comments on some aspects of the manuscript

4A Tone of the paper

Theoretical models in physics are rarely perfect, and thus they are not normally expected to fit experimental data fully. There are often conflicts of detail in comparisons between experiment and theory, particularly in the complex systems studied in nanoelectronics, and one must allow for this. Consequently, the weight that should be given to small conflicting details is a matter of judgement: a discrepancy for some practitioners may be noise for some others. This variability in judgment applies also to the referees, and their choices of arguments ultimately reflect their preferences. Of course, the views of the referees are shaped by the state of understanding in a research field at the time that the paper is being considered. It is unfortunately impossible to reinstate that state of mind of the involved experts two to three years after the review process. Therefore, the judgement of the importance of conflicting details in the paper becomes a subjective matter. Nevertheless, the panel has considered some of the conflicting matters in the manuscript.

In the panel's view the article by Vaitiekėnas et al. would have benefited from a broader discussion of nuances in the full data set, which was requested by some referees and could have been supplied in the supplementary material. For example, an objective discussion of the relevance of Andreev bound states would have been useful. Every scientist has their own way to cope with conflicting details in comparisons between observations and theoretical expectations, and their own prejudices in weighing their significance. Differences in view are reflected in the way the manuscript is written, and in how conflicting details are brought up for the readers. Some authors, while presenting extensive data, concentrate nearly exclusively on supporting data, and present other data in very brief side remarks. Presentations in which discrepancies are minimized and not made explicitly visible to readers call for extreme care of the referees in their judgment of the paper. Could the selection of conflicting details and the manner in which they were presented in this paper have been influenced by the foreseen possibility that additional information would have impacted the reviewing process negatively? Unfortunately, this is impossible for us to judge. We also note that the statement in the research article summary "We show experimentally and theoretically that the winding of the superconducting phase around the shell induced by the applied flux gives rise to MZMs at the ends of the wire" pushes the conclusions beyond those made in the peer reviewed part of the article in which such claims are carefully qualified.

4B Questions about specific points in manuscript

In scientific publishing, including in high-level journals, scientific credibility lies heavily on the peer review process. It is expected that large blunders go through rarely, and that small ones can be sorted out as part of the ongoing scientific process. Issues that arise later can be considered as part of a healthy scientific debate without drastic implications on the scientific credibility of the paper. In the present paper, there are a few issues that fall in this category:

4B1 Hard gap, which sometimes displays a shoulder

Most of the experts the panel consulted accept that qualified devices in the Vaitiekėnas et al. study exhibited hard gaps in the zeroth Little-Parks lobe. Sometimes, however, a shoulder is observed in the experiments. The authors of the publication decided not to put much weight on the shoulders. However, these shoulders could be the signature of resonant barrier Andreev states that would shift down in magnetic field and could remain near zero energy over a large range of field, that could include the entire first lobe. This behavior has been seen repeatedly in past experiments. A proper supplementary discussion of the difference between the data in the Science paper and the older Andreev-based scenario would have been extremely valuable for full appreciation and understanding of the new results.

4B2 Aspects of the theory content:

a Checkered parameter range for topological superconductivity; smallness of topological gap

According to the theoretical models explained and simulated in the Vaitiekėnas et al. paper, the topological phase diagram is rather patchy (see Figs. 4 C and 4D). Because it is unclear where the measured nanowires should be positioned on the theoretical phase diagram, the theory is not able to reliably predict the presence or absence of MZMs in individual devices. The simulations suggest that the topological gap is often small compared to its value in the zeroth LP lobe when the ground state is topological (see for example Fig. S24). It seems to us that the experimental portion of the paper is more successful than theoretically expected in finding well isolated MZMs.

b Transition of the zero bias peak to a dip as tunnel barrier is opened:

The disappearance of the ZBP studied in the paper's main figure upon tuning the gate voltage toward weak barriers (increased conductance) does not proceed in full accordance with the theoretical simulations given in the paper (see Appendix X). The zero-bias peak transforms into a zero-bias dip, as expected, but the evolution of the bias voltage traces with gate voltage is not in full agreement with theoretical simulations, and not unambiguously distinct from what would be expected from crossing Andreev bound states that for unknown reasons have an anomalously slow gate voltage dependence. This discrepancy is noted in the paper, and briefly discussed by pointing to partial consistency and partial inconsistency with theory. The panel hopes that this discrepancy will trigger further theoretical and experimental work to see if its systematics can be understood.

c A zero bias peak can be present in the first lobe in trivial phases

This issue is apparent in Figs. S23 and S24, which shows a weak spin-orbit interaction example in which there is a ZBP (see for example S24 panel A). When combined with the discrepancy in the peak to dip transformation, the actual witnessing power of NIS data, according to present theoretical understanding, is weak for topological superconductivity. This situation may, we believe, change in the future if device control can be improved.

d Superconducting phase winding around the shell

The idea of phase winding as a driver of topological states is original to this paper and represents a significant scientific contribution. One important implication of the phase winding ideas is that it makes it possible to search for MZMs at smaller magnetic fields than in previous work and in materials with smaller g-factors. The experimental data are generally consistent with the performed theoretical modeling, but it was not possible to model individual devices mainly because of limited device control, at least at the time of this paper. This raises questions on whether the parity of the number of phase windings should play a role. In some parts of the paper (see e.g. Supplementary Fig. S6 and S7) the fact that the zero conductance peaks in the NIS devices appear in odd lobes and are split in even lobes is highlighted, as if confirming the model. In reality however the parity of the winding number should have an effect only in systems with cylindrical symmetry, which is not entirely preserved in the experiment.

5 Remarks and conclusions

5A Confirmation bias in data selection and analysis

The phenomenon searched for in the Vaitiekėnas *et al.* article occurs only if conditions that cannot be reliably controlled experimentally are satisfied. At present this is a fact of life for all experimental searches for Majorana particles, and data selection is therefore a natural element of the search process. It is expected that a large part of the measurement data in any study can be immediately dropped as uninteresting. The committee has no objection to disregarding data from samples in which the desired conditions have clearly not been achieved. However, the procedures by which the data presented in any publication are extracted from the full measurement parameter space should be fully documented and transparent.

In the Vaitiekėnas *et al.* paper in particular, both NIS spectroscopy and Coulomb blockade measurements are interpretable only if simple tunnel barriers with no internal degrees of freedom are produced. The authors have explained to us that, based on their experience, simple tunnel barriers tend to appear in the initial part of the pinch-off regime. In spectroscopy devices they searched for this sweet spot by scanning over a wide parameter space in initial survey experiments used to find promising devices, attempting to tune them to the gate voltage range judged to have a simple tunnel barrier – `*the tunneling regime'*. For the Coulomb blockade experiments they sought well-blockaded Coulomb islands that have symmetrical barriers and gate voltage dependences characterized by low valley and high peak conductance. In our judgement, this selection procedure makes scientific sense.

The committee had the opportunity to check the initial screening ("calibration" in the authors' terminology) of devices and found that it was performed carefully following a systematic protocol. The selection procedures employed to identify promising devices and voltage regions for further analysis were, however, not infallible. Resonances in the barriers can be missed and later misconstrued as field-induced Majorana features. The transition in gate voltage between the clean *tunneling* barrier regime, and a region with states in the barrier cannot always be unambiguously identified. In the Coulomb blockade experiments, the behavior of some wire segments does not follow expectations. In our view, it is important that such observations be included in the data analysis.

Given the occasional ambiguity of data selection protocols and its weight in choosing the data retained for analysis, it is important that authors guard themselves against confirmation bias. In the panel's view the Vaitiekėnas *et al.* data was prescreened using a narrower filtering than required by general considerations. Data that passed the protocol with some minor blemishes have been neglected from the analysis. Owing to the preselection procedure, the final amount of data analyzed for the publication was somewhat limited. In particular, the data for the Coulomb blockade analysis was limited to only a few even-odd splitting determinations, in a series of devices from a single

quantum wire. When we reexamined all the collected data, the picture of the device-length dependence of even-odd splitting that emerged was more nuanced than the one in the paper.

The data used in the Vaitiekėnas *et al.* paper came, with one exception, from wire batch 439. Sample 5 discussed in the paper was from batch 638, which had thinner 7 – 10 nm Al shell thicknesses. Although wires from batch 638 wires performed poorly in general, two other NIS devices from that batch exhibited good tunneling characteristics but had no first-lobe ZBPs. They were not included in the success rate statistics provided upon request to the referees of the manuscript. Although batch 638 devices were generally bad and viewed as not useful, the authors would have given a more realistic picture of the difficulty of finding MZM behavior by reporting briefly on devices from this batch, as well as other batches that they studied. The statistics on NIS junctions given in the Zenodo post would have changed slightly, but not so much as to undermine the conclusions of the paper substantially. The overall statistical summary would have stressed the importance for MZM behavior of having both high-quality wires with favorable geometrical properties and successful etching. When all measurements are included the statistics for the NIS devices are as follows: sixty-four devices were fabricated from wires originating from five growth batches. Fifty devices were measured, of which fifteen were acceptable according to the protocol. Out of those, seven displayed a ZBP in the first LP lobe and eight did not. The Vaitiekėnas *et al.* paper focuses mainly on wire batch 439. For this batch six acceptable devices displayed a ZBP and one did not. It appears that wire batch 439 was of high quality and accidentally had a core-shell geometry that was favorable for MZM physics. The manuscript presented three devices from batch 439 with a ZBP, and one device from a different batch, with no ZBP.

The committee notes that first lobe ZBPs can be present or absent even on the same wire probed by separate, but similarly prepared, NIS junctions. The data on samples 63J1 and 63J2 illustrate this (J1 and J2 refer to separate NIS junctions from device 63 in batch 829 (not uploaded on Zenodo): only one of these two junctions passes the *tunneling regime* protocol and displays a ZBP. Thus, uncontrolled variations in sample details can have an important effect either on the fragile topological phases, or on its visibility in tunneling experiments. In this sense, it is unfortunate that the authors did not extend their data collection and analysis, possibly in a follow-up paper, with the goal of pinpointing which factors are truly relevant for the observations of ZBPs.

In its Fig.2, the Vaitiekėnas *et al.* paper gives the impression that topological superconductors with gaps that are nearly as large as those in the bulk superconducting material can be achieved relatively routinely in full-shell nanowires. Given the theory component of the paper, which finds that topological states appear sporadically as gating conditions are varied and that gaps tend to be small, this experimental result is surprising. It has also not been reproduced publicly at this time to our knowledge. Could the authors have been misled by unintended confirmation bias in concluding from their study that full-shell wires with axial fields “provide a relatively easy route to creating and controlling MZMs in hybrid materials”? Perhaps. It could well be that the full-shell wires in batch 439 were accidentally in the topological part of the phase space and of a quality that is difficult to reproduce at present but might become routine eventually. We rely on future experiments and on the normal scientific process to determine if the axial magnetic field strategy is a competitive route to quantum wire MZMs.

The CB analysis is also susceptible to confirmation bias due to the application of data set preselection protocols that inevitably have gray areas. Out of the ~100 good-quality measured data sets, only ~25 were included in the initial analysis (cf. Figure 13). No further data was analyzed even though the principal authors must have known about at least one extreme outlier among the discarded data sets. This outlier, in the 300-nm-long segment, was immediately obvious to the committee upon glancing through the data for the first time. In the table compiled by the authors the

reason given for discarding the outlier is “Tunnel gate sensitive configuration”, but looking at the data plots, the panel did not find any distinct difference between the calibration data of the regular devices and the outlier device. We view this discrepancy in judgment between the committee and the authors as a normal scientific dispute that can be tolerated given the rest of the data which generally supports the authors’ interpretation. The apparent strong dependence of the even/odd CB results on the basic parameters (transmission and asymmetry of NIS junctions, Al shell thickness) is not discussed explicitly in the paper. Taking the enhanced scatter of the data points in the extended analysis into account, the conclusions in favor for MZMs become less strong. Nevertheless, the apparent strong decay of the even/odd effect with nanowire length is supported by the full set of data, and did not as far as the panel is aware have a convincing alternative to the MZM interpretation at the time of writing of the Vaitiekėnas et al Science paper.

5B Other experiments with Copenhagen core-shell nanowires

a NIS experiments

After the publication of Vaitiekėnas et al. in Science 2020, Valentini et al., from the Institute of Science and Technology in Vienna, Austria, reported in July 6 2021 in Science on “Nontopological zero-bias peaks in full-shell nanowires induced by flux-tunable Andreev states”. Their NIS spectroscopy measurements were performed on full-shell hybrid nanowires from the same growth batch as the Copenhagen group, albeit from a different area of the growth substrate, measured one year after, and were not etched using the recipe given in Vaitiekėnas et al. They reported that devices with short tunnel regions (the tunnel region is the region at one end of the core-shell nanowire where the superconducting shell has been etched away) do not display subgap states whereas long tunnel regions always do. The authors interpreted all subgap states as due to non topological Andreev states of the quantum dot formed in the tunnel junction between the normal electrode and the superconducting core-shell nanowire.

There are differences in the sample behaviors seen by the two groups. One difference is the value of the pinch-off voltage (the voltage needed to reduce the conductance of the NIS device toward one quantum), which was much greater in magnitude in the samples measured by the Vienna group. The pinch-off voltage could be related to the etching method they employed which may cause impurity doping of the nanowire. Gentle etching seems to be a must for good samples, and this may have been a problem also in Copenhagen in the thin-shell samples, which were etched using the same etching parameters as for the thicker layers.

b Coulomb blockade experiments combined with NIS experiments

A paper on the Coulomb blockade on the same systems was subsequently published by the Vienna group (Valentini et al, “Majorana-like Coulomb spectroscopy in the absence of zero bias peaks”) in Nature in December 2022. They report junction-dependent, even–odd modulated, single-electron CB peaks in InAs/Al hybrid nanowires without concomitant low-bias peaks in tunneling spectroscopy. They do not find a monotonic dependence of even-odd spacing with island length, but rather spacings that are extremely dependent on the barrier parameters. They interpret the data in terms of low-energy, longitudinally confined island states rather than overlapping Majorana modes.

c Need for reproduction of the experiment

The apparent discrepancy between the Vienna and Copenhagen observations, which has many potential explanations, reinforces the need for confirming experiments. The main experimental claims of the paper Vaitiekėnas et al. paper should not be viewed as established scientific fact until they are reproduced in at least one other laboratory, and reported in a refereed publication. The view of the panel is that the Copenhagen experiment would be completely vindicated if another lab combined wire growth, fabrication, and measurement and obtained similar results. Etching without

creating too much disorder in the barrier to achieve a small pinch-off voltage is likely the most challenging obstacle to confirming experiments. As far as we know, there is still no reproduction of the experiment in the public domain.

5C Charge jumps

There may be a tendency in the community to compare different Majorana works and make parallels between different experiments searching for MZM even though the involved author lists do not have any overlap. This has raised questions concerning the role of charge jumps in giving a misleading impression of the gate voltage stability of zero bias tunneling peaks. On the basis of our inspection of original data files, the panel concludes that the sizable displacement of the charge jump positions in gate voltage between Fig. 2 and its extended version are due to ordinary drift during the time interval (on the order of one week) which separates the data set given in Fig. 2 and the extended data set provided on Zenodo. Thus, there is no reason to suspect that the data traces presented in the Vaitiekėnas et al paper do not faithfully represent what has been measured. There is no specific cropping in the gate voltage range that was used to paint a picture favorable to the paper's thesis.

5D Summary and Conclusion

The questions we attempted to answer are:

- whether the data presented in the *Science Magazine* article accurately represented the outcome of the experiments undertaken, and
- whether the authors deliberately or due to gross negligence withheld data that undermined the conclusions of their paper.

We have concluded that:

- The presented data do, for the most part, represent the outcome of the experiments: the authors have exercised scientific judgement in selecting which data to share using criteria whose application was partially subjective. Although data selection did, in our view, result in conclusions that did not adequately capture the variability of outcomes, the excluded data did not undermine the paper's main conclusions.
- The shortcomings we have noted in this manuscript do not constitute gross negligence.
- We do not view the authors' behavior in connection with this paper as an instance of scientific misconduct

Remarks on Conclusions:

In our view the authors carried out many experiments, built and studied sophisticated theoretical models, and analyzed considerable data in an effort to confirm the theoretical hypothesis – original to their paper - that topological superconductivity should appear in the non-zero Little-Parks lobes of quantum wires with axial magnetic fields. We have examined all that data and its analysis, and have concluded that the experimental and theoretical findings were not grossly misrepresented in the Science publication. We are confident that no data was fabricated and that we have seen all the data. We acknowledge that data selection is problematical in this paper, as in many studies of nano-electronic devices, because of their exquisite sensitivity to atomic scale disorder. The authors presented data from a reasonably representative subset of those measurements that satisfied acceptability criteria designed to filter out devices that were too disordered or had active degrees of freedom embedded in their tunnel barriers. In most cases this amounted to restricting attention to simple tunneling regimes in gate voltage regions in which the tunnel barriers were weakly pinched off; in some devices no simple tunneling regime could be identified. The authors should have been more explicit with readers and with referees in explaining their success rate in fabricating devices that showed simple tunneling characteristics and had MZM behavior and, by flagging alternatives and uncertainties, more evenhanded in their discussion of interpretations. Upon reexamination of the

Coulomb blockade data, we find that the "Incompatibility with a power law", claim written both in the Research Article summary and in the article itself, is probably too strong. The MZM success statistics are not quite as compelling in our independent reexamination of the full data set as suggested in the Science paper, but still consistent with theory. The research article summary seemingly pushes the conclusions beyond those made in the peer reviewed part of the article by neglecting some essential qualifications. In high impact journals, papers are often written in a manner that provides an account of the experiments that is both polished and optimistic. A positive tone is natural in papers describing work that is regarded as groundbreaking by its authors. In this sense, this Science paper does not differ from many others that we encounter in the nanoscience field and in science more generally. The tendency of enthusiastic authors to have inflated expectations is perhaps as much a part of the normal scientific process as the tendency of readers and referees to be skeptical and ever alert to errors in logic or judgment. Within this familiar landscape that surrounds our daily work as scientists, we judge that Vaitiekėnas et al. have not crossed the line that separates scientific discourse and debate from scientific misconduct.

We recommend that:

Regarding the Vaitiekėnas paper

- A statement explaining the set of criteria used to select acceptable nanowire devices, and a statistical summary of the success rate for growth and fabrication of devices deemed acceptable by these criteria, should be appended to the Vaitiekėnas *et al.* paper as a note added. The number of NIS devices that did not have successful tunneling spectroscopy should be stated, along with the number of devices with successful tunneling spectroscopy exhibiting ZBPs in the LP1 lobe and the number not exhibiting the ZBPs.
- The full set (25+56) of Coulomb blockade data files be uploaded to Zenodo, along with the descriptive table explaining why some datasets were excluded and the 2023 analysis.

More generally

- We recommend that the nanoelectronics and low-dimensional electron system community maintain high standards for fulsome objective reporting on technical details of sample fabrication, and on success rates in fabricating devices that exhibit the behavior described in a publication. These standards should be enforced by referees.
- Prescreening of data to identify relevant regimes in a large parameter space should be exercised in a fully documented, transparent fashion, and discarded data should be made available to the community via a long-term data repository.
- Journal Editors should make it more clear to readers which parts of the published material have gone through peer review and which parts are editorial addenda.

Afterword:

In his famous 1959 talk *There's Plenty of Room at the Bottom,* Richard Feynman asked this question: `` What would happen if we could arrange the atoms one by one the way we want them …". It was a good question. Over the following 65 years we have made a great deal of progress toward Feynman's vision and plenty of good things have happened. Still, we are now only part way along Feynman's voyage of discovery and must live with many limitations on our ability to arrange atoms one by one to realize physical properties that are of interest or have technological value. That is the reality of nanoelectronics and of materials physics more generally. We cannot, for example, make a

proximitized nanowire just the way we want it. We cannot make a nanowire today that is identical to the one we made yesterday. We could not make a nanowire device in Vienna that is identical to one made in Copenhagen, even if the wires themselves were identical. Whether the long-term goal is quantum computation or greater energy efficiency or something in between, further progress toward Feynman's vision will be aided by the clearest possible communication between researchers on the device fabrication methods employed in a particular piece of work, and on the reliability with which those methods are able to achieve devices that exhibit the properties that are subject of a particular research subject. We encourage authors in our community to work diligently to communicate this important information, and referees to insist on high standards.

****

Appendix

I. Text of the complaint by Science Editor Jake Yeston:

"Essentially, the question is whether the data presented in the original paper accurately represented the outcome of the experiments undertaken. Journal editors and reviewers can only assess data to which they have access. If data that did not support the claims in the paper were withheld or suppressed, then the paper submitted to the journal implied greater statistical support for the conclusions than the experiments in fact bore out. (...)the source of greatest concern is the range of voltages and number of independently tested devices that were represented in the paper's second figure. The editors at Science believe that an independent, transparent investigation by experts in this subfield of Majorana physics is necessary to ascertain whether or not the authors unethically withheld data that undermined the conclusions of their paper."

A second "Form for reporting suspicion of research misconduct or questionable research practices to the Practice Committee at the University of Copenhagen" was filed by Frolov and Mourik on 15 dec 2021, requesting data from seven publications by the NBI group, published between 2016 and 2021. " (...) we made a complaint to Science about selective data presentation. Based on that experience, we think it's possible that Charles Marcus' group is being improperly selective in its treatment of its data to justify erroneous conclusions. Prof. Marcus has argued that this is a scientific debate that should happen in a collegial way. Such a debate cannot happen, however, while he persists in keeping so much of the data confidential. To understand if Prof. Marcus has made physics claims justified by his experimental data, we really need to see the full set of diagnostic and experimental data that was acquired on all similar devices and best practices in our field dictate that these be made available to other groups. Not doing this is helping to create a replicability crisis." The committee is not investigating this second claim.

II. Conclusions of the internal investigation conducted by the NBI (dated Feb. 21, 2021).

In short, our conclusions read as follows:

REF: JWT

(a) We find no problems with the paper, nor with the conclusions in the paper, nor with the data supporting the claims of the paper.

(b) We find the complains of Mr Frolov and Mr Mourik unjustified.

(c) All data connected with the present paper has - according to demand – been transmitted rightfully to third parties. No additional data is left out.

Should further information be needed I am most happy to assist.

Sincerely,

Jan W. Thomsen

Jan W. Thomsen
Professor
Head of Department
Niels Bohr Institute

III. Expert panel:

Was constituted of Allan MacDonald, Pertti Hakonen and Sophie Guéron. Alfredo Levy Yeyati took part in the initial discussions.

IV. Interviews were conducted, either in person or via Zoom, with the following people:

Sergey Frolov, Vincent Mourik, Saulus Vaitiekėnas, Roman Lutchyn, Jelena Stajic, Georgios Katsaros, Elsa Prada, Ramon Aguado, Pablo San-Jose, Charles Marcus, Karsten Flensberg, Joachim Mathiesen, acting head of department of the Niels Bohr Institute, Bernard van Heck.

The help of Bernard van Heck in conducting several analyses of the initial and additional Coulomb blockade data was invaluable.

V. Documents consulted by committee

-v4 of Post publication review by Frolov and Mourik, march 10, 2022 (https://doi.org/10.5281/zenodo.6344447)

-Niels Bohr Institute internal investigation report

-reviewing process exchanges with Science

- Authors' response to v3 or v2 of FM's post publication review (v4, March 2022 adds that not all data was communicated.)

- Data on the ~80 devices fabricated and measured between May 2018 and May 2019.

-dec 13-15 visit to Copenhagen by Pertti Hakonen and Sophie Guéron. People seen: Charles Marcus, Saulius Vaitiekėnas, Karsten Flensberg, Peter Krogstrup, Asbjorn Drachmann (technical research coordinator), Rasmus Bjerregaard (data scientist).

VI. Statistics of successful/ unsuccessful devices

The committee requested to be shown data relative to all nanowire devices fabricated between the first measurement of May 2018 and the resubmission date of the combined experimental and

theory paper. This important data sorting and presentation work was performed by two of the authors, Charles Marcus, Saulius Vaitiekėnas, with the help of Asbjorn Drachmann (technical research coordinator) and Rasmus Bjerregaard (data scientist). This resulted in 11.4 Gb of data (including device layout, bonding scheme, SEM images, pictures of a lab book, data curves and an extensive excel catalogue. CM and SV estimated that the committee went through 60% of the data curves and 100% of the devices measured during their visit to NBI. Additional zoom meetings to discuss the data occurred subsequently.

NIS tunneling devices from five different core-shell nanowire growth batches were measured. Of the tunable devices, for which the experimental protocol appended to this report was followed, which showed clear spectroscopy features, apart from the batch for the experiment (which showed 6 ZBP out of 11 devices), only one device out of 29 displayed a ZBP, whereas 9 devices displayed clear spectroscopy features and were considered in the tunneling regime. For statistics of success, our intent was to find samples with no low-energy structure in LP0 and a ZBP in LP1. The statistics established jointly by SG, PH, AMD and the NBI team (AMD remotely) for all batches are the following:

Batch-by-batch device yields related to" Flux-induced topological superconductivity in full-shell nanowires," by S. Vaitiekėnas et al., Science (2020). Dec. 29, 2022, amended by panel, May 2023

QDev 439	Unsuccessful fabrication	Protocol not followed/ Not studied systematically	No clear spectroscopy	Clear spectroscopy, protocol not satisfied, no ZBP*	Clear spectroscopy, protocol satisfied, no ZBP	Questionable ZBP observed	Clear spectroscopy ZBP observed	Total number of devices
NIS-junctions	24, 34, 38, 40-42, 81-82	13-15	36, 39, 4J2		12	9 (outside tunneling regime)	1, 3, 4J1, 8, ~~9~~, 10, 11	
Quantity	8	3	3	0	1	1	~~6~~ 7	22

4J1 and 4J2 refer to device 4 junctions 1 and 2. Devices 2, 6, 7, 35, 37, and 43 were CB-island devices. Devices 44 through 53 were SIS and had large resistances.*See "Table PNS".

QDev 637	Unsuccessful fabrication	Protocol not followed/ Not studied systematically	No clear spectroscopy	Clear spectroscopy, protocol not satisfied, no ZBP*	Clear spectroscopy, protocol satisfied, no ZBP	Questionable ZBP observed	Clear spectroscopy ZBP observed	Total number of devices
NIS-junctions	29	16-18, 26		19, 25, 27, 28				
Quantity	1	4	0	4	0	0	0	9

General comment: A different wire batch (QDev637) than the main batch (QDev439) studied in the paper.
This batch had smaller diameter and thinner shell visible on micrographs. Non-destructive or no Little-Parks oscillations in the 4-probe measurements. Various end-dot and ZBPs were observed inconsistently. No ZBPs in the tunneling regime. *See "Table PNS".

QDev 638	Unsuccessful fabrication	Protocol not followed/ Not studied systematically	No clear spectroscopy	Clear spectroscopy, protocol not satisfied, no ZBP*	Clear spectroscopy, protocol satisfied, no ZBP	Questionable ZBP observed	Clear spectroscopy ZBP observed	Total number of devices
NIS-junctions		30, 33	22	21, 31, 32	5, 20, 23			
Quantity	0	2	1	3	3	0	0	9

General comment: A different wire batch (QDev638) than the main batch (QDev439) studied in the paper. This is the same wire batch as Device 5 from the paper that did not show ZBPs in the tunneling regime. This batch had a smaller diameter and thinner shell visible on micrographs. Device 22 could not be pinched off. Device 30: Spectroscopy was not measured because 4-probe data showed a non-destructive Little Parks effect. Device 33: Very little data because of poor spectroscopy. *See "Table PNS".

QDev 564	Unsuccessful fabrication	Protocol not followed/ Not studied systematically	No clear spectroscopy	Clear spectroscopy, protocol not satisfied, no ZBP*	Clear spectroscopy, protocol satisfied, no ZBP	Questionable ZBP observed	Clear spectroscopy ZBP observed	Total number of devices
NIS-junctions	78, 79		75	71, 72, 74, 76, 77	73			
Quantity	2	0	1	5	1	0	0	9

General comment: A different wire batch (QDev564) than the main batch (QDev439) studied in the paper. Wires from this batch were used in Vaitiekenas, Krogstrup, Marcus, PRB 101, 060507(R) (2020). This batch had larger diameter and thicker shell visible on micrographs. Non-destructive Little-Parks oscillations in the 4-probe measurements. Subgap states were uncommon in these wires. Device 75 could not be pinched off. *See "Table PNS".

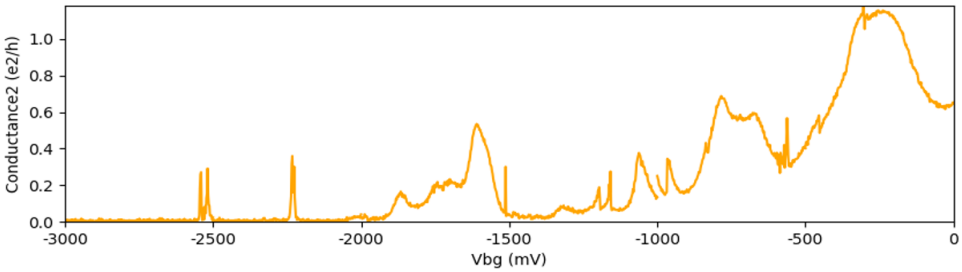
QDev 829	Unsuccessful fabrication	Protocol not followed/ Not studied systematically	No clear spectroscopy	Clear spectroscopy, protocol not satisfied, no ZBP*	Clear spectroscopy, protocol satisfied, no ZBP	Questionable ZBP observed	Clear spectroscopy ZBP observed	Total number of devices
NIS-junctions	60, 64, 66, 70	54-57	58	61, 63J2	59, ~~63J2~~, 65, 68	0	63J1	
Quantity	4	4	1	~~1~~ 2	4 3	0	1	15

63J1 and 63J2 refer to junctions 1 and 2 on Device 63. Devices 62, 67, and 69 are 4-probe devices.
General comment: A different wire batch (QDev829) than the main batch (QDev439) studied in the paper.
The wide distribution of diameters in this batch resulted in the first lobe position ranging from ~200 to 600 mT for different devices. No systematic behavior was observed. *See "Table PNS".

VII. Experimental protocol-Tunnel spectroscopy devices

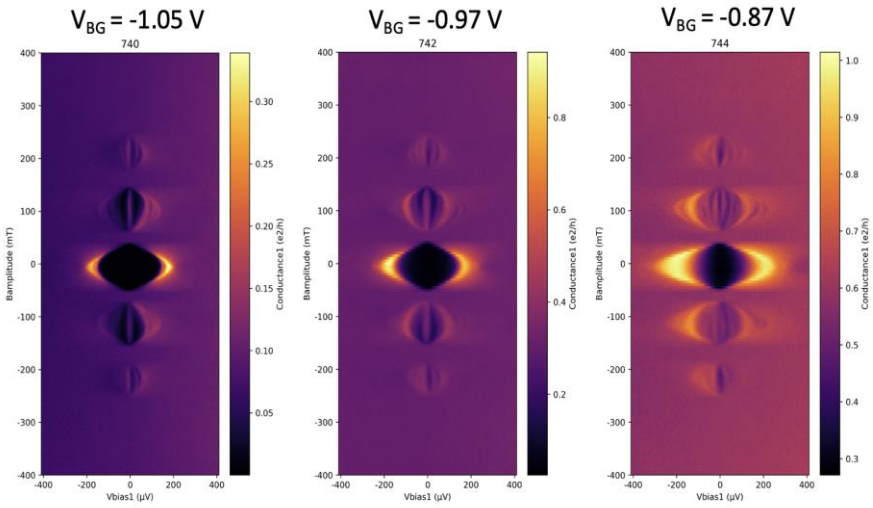
Based on their experience of nanowire-based devices, the experimentalists, after a few exploratory experiments on roughly 10 devices, developed a more systematic protocol to measure their samples. The protocol as written by the authors at the request of the committee is attached to this report. Summarizing, the experimentalists first looked for the pinch-off regime in zero magnetic field, as the back gate voltage was swept from positive to negative voltage. This back gate is common to all wires

on a chip, so the pinch-off curves were taken for all wires simultaneously. An example is shown in protocol- **Fig. 1** (provided by the authors). The pinch-off is defined as the condition dI/dV ~ 0 (unmeasurably small) independent of source-drain voltage, and the open regime as the regime where dI/dV > $2e^2/h$ at zero source-drain voltage. Next, conductance, dI/dV, as a function of dc source-drain voltage V and back gate, was measured, looking for the tunneling regime. Next, a wire to focus on was selected, based on the quality of the tunneling regime. "Tunneling regime " is defined in the author's protocol as a regime with conductance smaller than the open regime, in the range of dI/dV ~ 0.05 - 0.2 e^2/h, measured above the gap, with vanishing dI/dV below the gap. This regime is also restricted to a gate voltage that is less negative than where many resonances (extending above and below the gap) were typically found. The rationale for this definition of the tunneling regime being that it is the regime in which the junction is a featureless tunnel junction with a flat density of states, so that the differential conductance reflects the density of states at the end of the core-shell nanowire, and is not convoluted with features in the density of states of the barrier (constituted by the bared nanowire segment) itself. Indeed, at more negative gate voltages, reduced screening at strong depletion typically results in multiple potential maxima along the junction. This in turn creates quantum dots (QDs) whose signatures are resonances, both below the gap energy and above the gap energy. If the junction is too open, dI/dV is not proportional to the local density of states. If the junction is too closed, reduced screening typically leads to resonances and disordered transport.



Protocol-Fig. 1. Pinch-off curve. Conductance versus back gate voltage for Device 1.

Next, on devices with an accessible tunneling regime—meaning that the device had a relatively resonance-free region in a region with lower conductance than the open regime without strong, frequent resonance or dynamical Coulomb blockade—higher-resolution data were taken in the zeroth and first lobes, over the range from the open regime to the QD regime, spanning the tunneling regime. Next, high-resolution magnetic field sweeps were taken at a few representative back gate values within this range of back gate voltages found above. Typically, 1-3 values of back gate voltage were examined (protocol-**Fig. 5**). For devices that did not show a good tunneling regime (e.g., most devices from wire batches 637 and 638, for unknown reasons), measurements were either discontinued or a few field sweeps were taken at gate voltage values between resonances. Further details can be found in Attachment I.



Protocol-Fig. 5 Conductance as a function of source-drain voltage and axial magnetic field at three back gate voltages.

VIII. Experimental protocol-Coulomb blockade devices (taken from authors'text written for the panel)

The authors focused on the device with all Coulomb islands working. One at a time, each island was tuned into a symmetric weak-tunneling regime by tuning four gates per island: First, the back gate was swept until valley conductance vanished. Next, the source-drain offset was zeroed (numerically) by examining conductance versus source-drain and plunger-gate voltages. Next, the back gate was used to pinch off the island by sweeping until the CB peaks vanished. At that value of back gate, the device was next symmetrized by examining sweeps of left and right gates. Tuning left and right gates over large ranges revealed slowly varying nonmonotonic conductance, presumably due to resonances in the leads. The device was tuned for a rough maximum of peak height to find the rough diagonal where left and right gates had equal influence avoiding these nonmonotonic regions in left-right gate space. This procedure is typical of how quantum dots are tuned in a variety of contexts. After symmetrizing left and right conductances and avoiding nonmonotonic regions, a window of peaks was examined to find regimes in the plunger gate where the peak valleys went to zero conductance, peaks were moderately high (roughly 0.4 - 0.8 e2/h). Tuning-up was carried out at zero magnetic field. Next, plunger gate sweeps on the tuned island were carried out at zero field, in the destructive regime, and in the middle of the first lobe (Fig. 13). The magnetic field was aligned to the wire (...). Next, a high-resolution map of dI/dV as a function of magnetic field (outer loop) and plunger gate voltage (inner loop) was taken, spanning roughly 6-10 CB peaks and fields. These runs typically took several hours and swept continuously from positive to negative field. The symmetry of the data in magnetic field provided a further check on device stability. Immediately after the field sweep, without changing gate voltages, conductance was measured as a function of source-drain voltage (inner loop) and plunger gate (outer loop). These "CB-diamond" runs provided a measure of the lever arm of allowing gate sweeps to be converted to energy. The fact that island energy levels detected by CB measurements depend sensitively on the lever arm, though the mean CB spacing does not, motivated taking separate measurements of CB diamonds in the zeroth lobe, the destructive regime, and the first lobe be made. The CB diamonds in the first lobe were especially important for converting peak spacing to energy. This conversion allowed different islands to be compared. Diamonds were taken at the value of the magnetic field where the peak spacing cuts were extracted from the 2D maps. Separate lever arms for the source and drain were extracted from leading and trailing diamond edges to compensate for any remaining asymmetry in the coupling of the island to the two leads. To increase statistics for each island, the procedure was repeated at 1-2 additional back gate voltages (with corresponding required tune-up of left and right gates). For the 200 nm, 300 nm, 400 nm, and 600 nm islands, peak spacing fluctuations were extracted from the 2D maps of field and gate voltage. Then, CB diamonds measured in the first lobe were used to convert peak spacing to energy. For the 800 nm and 1000 nm islands, where peak spacing was nearly uniform in the first lobe, plunger sweeps at very high resolution were taken at several discrete values of the magnetic field in the first lobe. For the 1000 nm island, increased statistics were obtained by repeating the 110 mT high-resolution sweep at one other back gate voltage (with re-tuned left and right gates). Data at 110 mT and 140 mT were further analyzed using the first-lobe lever-arm data from 110 mT to test the dependence of island-length scaling on the field-dependent gap. Further details can be found in Attachment I.

IX. List of additional Coulomb blockade data with comments by the Authors

These comments refer to the curves included in the Coulomb blockade analysis, generating the figure shown in this report.

Files analyzed/Files to add in the second analysis

	First cooldown, analyzed	First cooldown, additional	Comments	Second cooldown, analyzed	Second cooldown, additional	Comments
200 nm	3165	3173	3173 appears to be a re-taken of 3165. Associated diamonds in 3166.		1224	We cannot recall why these data were not analyzed. We suspect this was a backup run that we never got around to analyzing. Associated diamonds in 1223.
200 nm	2202					
200 nm	2250					
300 nm	1878	1949	Run 1949 has large conductance, outside of our usual range. In fact, the reason for the peak-height criterion is visible by eye in 1949: the exaggerated peak motion in the first lobe results from transmission contribution to total capacitance (Flensberg, Aleiner-Glazman, Maurer *et al*). In this case, peak motion does not represent density of states. Note that valley conductance was not fully suppressed (run 1948), and Coulomb blockade in the destructive regime was nearly absent (run 1946). Data is reasonably quiet, but not in a regime to measure DoS. Associated diamonds in 1952.	1930	2023	Run 2024, which is the associated gate sweep to 2023, showed strongly non-monotonic features. Associated diamonds in 2020.
300 nm	1914					
300 nm		1863	Lower-resolution calibration sweeps. No diamonds.			

300 nm		1907	Calibration sweep with finite conductance background in the Coulomb valleys. No diamonds.			
400 nm	3416	3419	Concerned with possible drift in 3416, run 3419 was a re-take of 3416, taken after measuring the diamonds with no gate-voltage changes. Associated diamonds in 3425.		668	The subsequent run 669 showed strange line shapes, which raised our suspicion. In any case, we already had enough data for this island. No diamonds.
400 nm	2043	2038	A large switch occurred in the middle of run 2038, which also had a drift, so 2043 was taken analyzed instead. Associated diamonds in 2047.		763	We cannot recall why these data were not analyzed. We suspect this was a backup run that we never got around to analyzing. Associated diamonds in run 764.
400 nm					808	This was a backup run but showed jumps in the data. Associated diamonds in run 809.
400 nm					898	Only one side of magnetic field taken. Unusual switching behaviour at V_{PG} ~ -140 mV. No diamonds.
400 nm		2093	One switch visible. Zero-field Coulomb peaks have low conductance. Destructive regime displays high conductance indicating open regime or asymmetric tuning. Associated diamonds in run 2095.			
400 nm					654	1D sweep at 110 mT only. This lacks destructive regime data, which acts as calibration. No diamonds.
400 nm					929	1D sweep at 110 mT only. The sweep displays an irregular peak spacing at around -260 mV. We do not know if this is a jump in parity or in gating. No diamonds.

600 nm		1231	Because of the unusually high conductance in the destructive regime of run 1231, this did not follow the usual behavior and so was not pursued. Associated diamonds in 1235.	200	182 shifts but ok	Large drift and shifting. Some switches. Associated diamonds in run 183.
600 nm		1458	Ghost peaks [Albrecht PRL] in the zeroth lobe indicate parity switching amplified by asymmetric barriers. Associated diamonds in run 1461.	201	189	189 was repeat of 182 (only one side of magnetic field) taken after the device became more quiet. Associated diamonds in run 183.
600 nm		1878	1878 is from 300 nm island and was used for the analysis. The same is true for run 1914. Associated diamonds in 1883.	191		
600 nm					194	194 was repeat of 191 (only one side of magnetic field) taken after the device became more quiet. Associated diamonds in 192.
600 nm					216 large asym	Note the large drift in 212, which was taken at the same gate settings. The associated diamonds in run 214 display a large asymmetry.
600 nm		1428	Sweep with finite conductance background in the valleys, taken during calibration. The associated zero field data shows low conductance as seen in 1425. No diamonds.			
600 nm		1430	Sweep taken during calibration. The associated zero-field sweep shows ghost [Albrecht PRL] peaks as seen in run 1432. No diamonds.			
600 nm		1474	Ghost peaks [Albrecht PRL] indicate parity switches. Several jumps including one in the middle of the first lobe. Associated diamonds in run 1476.			

600 nm		1479	Similar run to 1474 displaying ghost peaks [Albrecht PRL], suggesting parity switches. Associated diamonds in run 1476.			
600 nm		1111	Sweep displays low zero-field Coulomb peak conductance and a very high conductance in the destructive regime, indicating either an open regime or asymmetrically tuned barriers. In addition, based on the subsequent run 1112, it appears to be taken close to an end resonance. No diamonds.			
600 nm		1125	There is a large switch at roughly 130 mT. Also, the sweep displays low zeroth lobe peak conductance and high destructive regime conductance. No diamonds.			
800 nm		118	118 and all runs from 48 to 238 are from zeppelinXi, wire B. This is a different wire from the one used in this study. Wire-to-wire comparison was not undertaken.	508	391	Large jump near the destructive regime. Only one side of the magnetic field. Low conductance in the zeroth lobe. No diamonds.
800 nm		1708	Low-resolution sweep inadequate for this length. In addition, the data display a drift and low peak conductance in the zeroth lobe. Associated diamonds in run 1712.		469	Drifting more than typical runs. Large jumps, including in the middle of both first lobes. Similar position indicates field driven jump. Associated diamonds in 465.
800 nm		1752	Low-resolution sweep inadequate for this length. In addition, the conductance is high in the destructive regime indicating an open regime. Associated diamonds in run 1754.		479	Low-resolution sweep inadequate for this length. Associated diamonds in run 477.

800 nm		2349 (random gate?)	Coulomb peaks invisible in the zeroth lobe. In notebook, this was described as "random gate config.", what this means is that the prior gate tuning did not follow the tune-up protocol. Associated diamonds in 2353.			
800 nm		1760	Several switch and drift. Run discontinued. No diamonds.			
800 nm					456, 461, 462, 463	These runs were used to develop the method of high-resolution one-dimensional sweeps. Associated diamonds in run 465.
800 nm					468, 469	Low-resolution sweep inadequate for this length. Also show significant switches. Associated diamonds in run 465.
800 nm					473, 474	Two back-to-back sweeps with different heights indicating drift not yet settled. Associated diamonds in run 470.
800 nm					493	493 was a low-resolution sweep, which was not analyzed further; higher resolution sweep 508, taken at the same settings, was used instead. Associated diamonds in run 470.
800 nm					539	Clear switch in the middle of the run. No diamonds.
1000 nm				1486		
1000 nm				1531		
1000 nm					1596-1599	After taking the high-resolution sweep 1596, the device was still drifting as seen in runs 1597-1599. Associated diamonds in run 1596.

1000 nm					1666, 1667	Not analyzed because of the very unusual behavior exemplified by the zero-field data in 1664. Associated diamonds in run 1665.
1000 nm					1691	At least two switches are visible bye between the first and the second, and the fifth and the sixths Coulomb peaks. We do not know if these are jumps in parity or in gating. The protocol was not followed for the tune up and there is no information on the zeroth lobe. Associated diamonds in run 1690.
1000 nm					1718, 1719	This data is unusual, but similar to data in 1664-1667. Associated diamonds in run 1717 (including a switch).

X. <u>"Peak to dip transformation"</u>

The published paper of Vaitiekenaas et al. does not elaborate on comparison of their own simulated tunneling conductance with the experimental NIS conductance data. Instead of comparing experimental results with their own simulation illustrated in Fig. 5 and in Supplement Fig. S23, qualitative comparison is made with the work of Vuik et al. (Ref. 44, SciPost Phys. 7, 061 (2019)) and the conclusion is "The increase of finite-bias conductance compared with zero-bias conductance as tunnel barrier decreases is in qualitative agreement with theory supporting MZMs (44), although the crossover from a peak to a dip occurs at lower conductance than expected." For comparison with the authors' own calculations, we have plotted the data of Fig. 5C that is provided on Zenodo (see Figure 16). One can notice that the low energy peak structure has a triplet form. The middle peak vanishes and two peaks remain. The remaining peaks are separated by an amount that depends on the geometrical parameters of the wire, and therefore this spacing is fixed by the sample dimensions. The simulated theoretical behavior does not fully agree with the measured data. In the data, the two appearing maxima seem to be at lower energy that expected from the theoretical simulations.

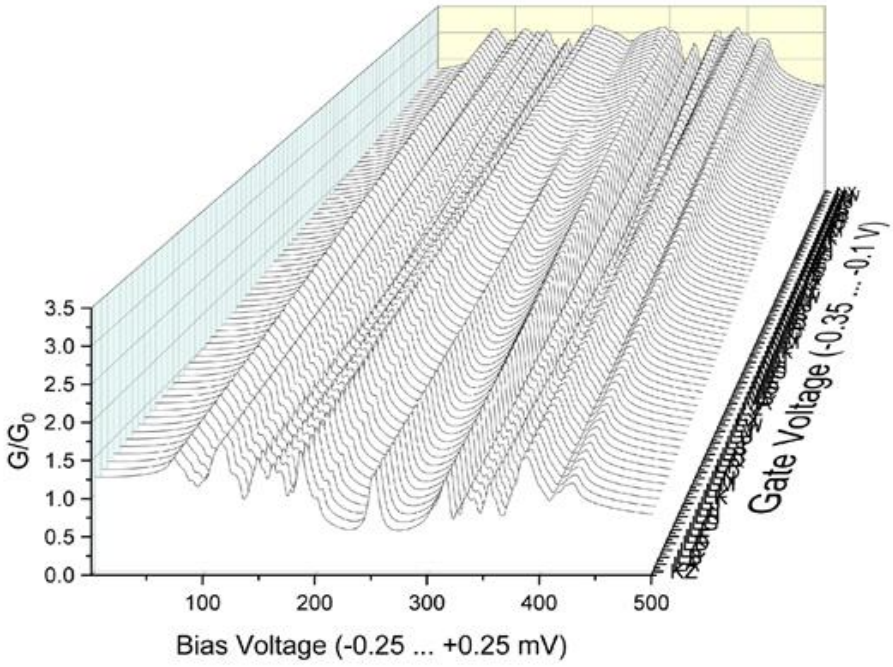


Figure 16 Waterfall plot of the simulated NIS conductance presented in Fig. 5C of the main paper.

Fabrication, tuning, and measurement protocols related to "Flux-induced topological superconductivity in full-shell nanowires," by S. Vaitiekėnas *et al.*, Science (2020).

Dec. 29, 2022

This document presents device fabrication, tuning, and measurement protocols, as presented during the visit by members of the Expert Panel on Dec. 15-16, 2022, and has been prepared at the request of the Expert Panel. The document has four sections: fabrication and measurement protocols for Normal-Insulator-Superconducting (NIS) devices, and fabrication and measurement protocols for Coulomb blockade (CB) devices. These protocols were developed over the course of the experiment and so were not followed strictly, particularly in early runs. The protocol also evolved over the course of the experiment as we sought to explore ranges of behavior, new wire batches, and alternative device designs.

We emphasize that a strategy designed to explore a new phenomenon will naturally differ from one aiming to gather yield statistics. In particular, gathering statistics requires a standard protocol, repeated as identically as possible, without exploring new parameter regimes or varying device design. Our study reports the discovery of a new physical effect, validated in several devices, consistent with community standards. We did not aim to report yield statistics beyond what we felt was needed for reasonable validation of our observations. When combined with theoretical and numerical support, our observations supported our interpretation.

1. Protocol for NIS fabrication:

   During the early experiments (May 2018-June 2018) fabrication began with wire selection. This was for wire batches 439, 637, and 638. The wire batch numbers reflect the growth sequence. Growth strategies and wire parameters differed from batch to batch. For instance, 439 wires were grown from unpatterned Au catalyst droplets, so there was a larger variance of diameters, while, for instance, 829 wires used a patterned catalyst. Some batches were grown vertically quickly and then fattened with lateral growth to avoid dislocations. Al thickness varied batch-to-batch by design. None of the batches used were attempted exact repeats.

   Wires were selected roughly from the middle of the growth substrate. Wires were transferred onto cleaved Si/SiOx chips with prefabricated Ti/Au bond pads, meanders, and alignment marks using a micromanipulator with a 100 nm tungsten needle. An adhesion promotor, AR300-80 was spun, baked, and rinsed. This was found earlier to prevent etch running, which was important to have the etch match the lithographic pattern. Parameters of the adhesion promoter step such as spin speed, bake time, and rinsing procedure, which may have important consequences, were not investigated.

   Next, a double layer of EL-6 MMA e-beam resist was spun and soft-baked. Next, the etching windows were exposed by electron beam lithography, developed using MIBK

and IPA, and cleaned with an oxygen plasma. Plasma cleaning is a critical step, removing residual polymer from the etch windows, but it was not optimized. The resist was then re-baked for another 60 s. Next, the chip was etched in MF321 for 75 s, followed by a two-step quench in deionized water. The resist was then stripped, and a new resist layer was spun and soft-baked. This was a double-layer resist system consisting of A-4 and A-6 PMMA e-beam resist stack. The leads and contacts to the shell were exposed and developed (as before, rinsing and oxygen-plasma cleaning), then milled with Ar plasma, then the Ti/Au leads were deposited, followed by a liftoff step. Another layer of A-6 was then spun and soft-baked. The leads were then written by electron beam lithography to expose the InAs core. Same development step. Again, Ar milling, but at lower power to prevent damage to the InAs. Argon milling was calibrated by QDev technical staff weekly to ensure uniformity and correct milling time which was crucial to not overly damage the InAs. Leads were metalized with Ti/Au and lift-off. After this, the wires were imaged for the first time.

Over the course of the experiment, this recipe evolved somewhat. For example, some devices have a top gate (Device 04), which required two additional fabrication steps. The first was a global layer of HfO deposited by atomic layer deposition, followed by e-beam patterning of the top gate and deposition of Ti/Au top gates with new meanders deposited.

These procedures were developed over several years mostly on half-shell wires. We cannot say if each step was necessary or how changes affected device behavior or fabrication yield. The change from Transene D, used in early experiments, to MF321 was made because the 7 s etch with Transene was more aggressive and too difficult to control and prevent over- or under-etching. One extra second made a significant difference. Also, Transene was slower to etch the tough oxide surface, and then when it broke through in a statistical process the pure Al underneath was etched almost immediately. This means that the 7 s etch was dominated by the statistical process of breaking through the oxide. The more aggressive Transene etch possibly caused more damage to the semiconductor. Overall, MF321 proved superior for most applications.

For other wire batches, this procedure was followed with slightly different MF321 etch times.

2. Protocol for CB fabrication:

The same fabrication procedure used for the NIS devices was followed for the CB devices, with a few differences: First, only 439 wires were used for the CB studies. Second, there was only one contacting step, which was the A-6 resist step with the lower-power Ar milling. Third, side gates were deposited in a separate A-6 step, this time without milling (because no contact was made). Also, devices were imaged after etching and before the first deposition step. This was necessary because precise alignment was needed for the CB devices. It is known that imaging devices change

the gate voltage pinch-off characteristics. Other possible effects of imaging are not known.

3. Protocol for NIS measurement:

Following wire bonding to an in-house designed circuit board, the sample was placed into a loading puck associated with a Bluefors XLD dilution refrigerator. RF and RC low-pass filters (QDevil) were installed in the fridge and additional RC low-pass filters were installed in the puck. The overall line resistance from the sample to room temperature was roughly 14 kΩ.

Typically, each chip contained 10-15 devices, sometimes from different wire batches, and all devices can be bonded using a custom multilayer circuit board with 100 bond pads. Circuit boards similar to our design are commercially available (QDevil). All devices were grounded during bonding, mounting, loading, and cooling. The chip was cooled in the dark, with LED illumination. After reaching base temperature (roughly 12 hours after loading) and ungrounding the sample, continuity checks were performed using a lock-in amplifier in manual operation rather than under computer control. Most devices passed this continuity test. In the case where the continuity test failed, further measurement of that device was discontinued.

In the initial device screening phase, multiple devices were connected, each to a separate lock-in. Wires were not oriented on the chip and so a parallel magnetic field could not be applied during the screening phase. Working at zero magnetic field, we looked for the complete pinch-off regime as a function of back-gate. Back-gate sweeps affect all wires on a chip. An example is shown in **Fig. 1**. Pinch-off curves were taken for all devices on a chip.

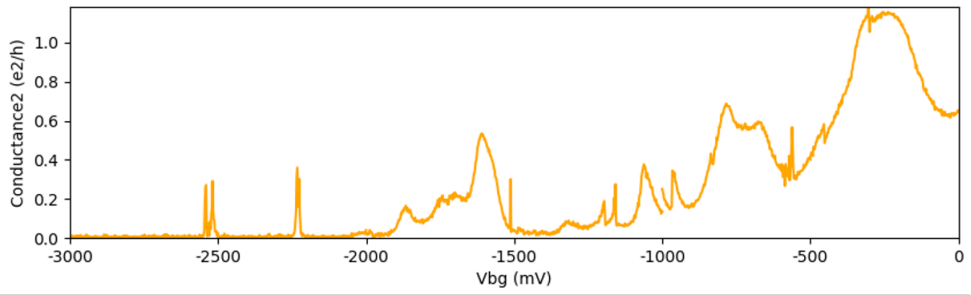


Fig. 1. Pinch-off curve. Conductance versus back gate voltage for Device 1.

Next, conductance, dI/dV, as a function of dc source-drain voltage and back gate, was measured, looking for the tunneling regime. An example is shown in **Fig. 2**. This was done for all working devices on the chip, typically the majority.

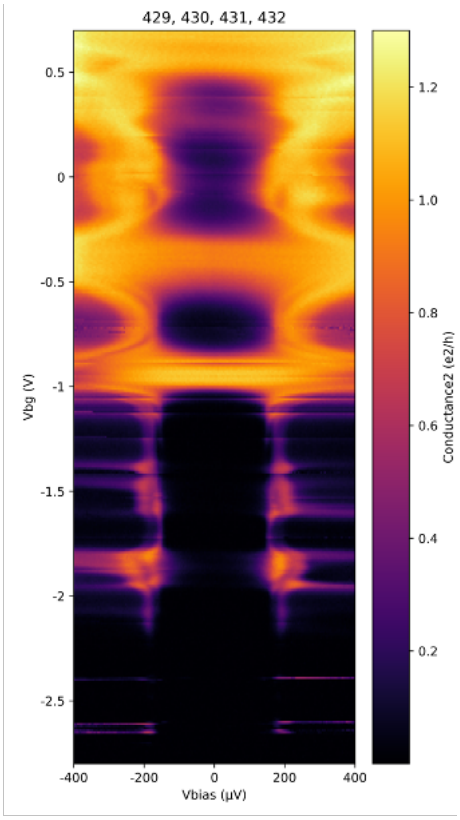


Fig. 2. Finding the tunneling regime. Conductance versus gate source-drain voltage and back gate voltage at zero magnetic field.

Next, a wire to focus on was selected, based on the quality of the tunneling regime. A protocol for field alignment was carried out for that wire. That involved measuring the angle of the wire on the chip from the micrograph and using the data-acquisition computer controlling the magnet to define field directions in the coordinates of the wire. Fine-tuning of wire angle was typically not needed. For a given wire batch, the lobe structure was mapped out by measuring dI/dV as a function of parallel magnetic field and source-drain voltage. For subsequent wires on that batch, a full-lobe sweep at this stage was typically not repeated. Instead, it was enough to move to the center of first lobe, knowing the rough field value for a given wire batch, and repeated the back-gate and source-drain measurement from pinch-off—defined as the condition dI/dV ~ 0 (unmeasurably small) independent of source-drain voltage—to the open regime, where dI/dV > $2e^2/h$ at zero source-drain voltage. An example is shown in **Fig. 3.** This procedure was repeated for all wires on each chip.

Next, zero-field data was examined to locate a good tunneling regime, that is, a regime with conductance smaller than the open regime, i.e., in the vicinity of dI/dV ~ 0.05 - 0.2 e^2/h, measured above the gap, with vanishing dI/dV below the gap, and at a less negative gate voltage than where many resonances (extending above and below the gap) were typically found.

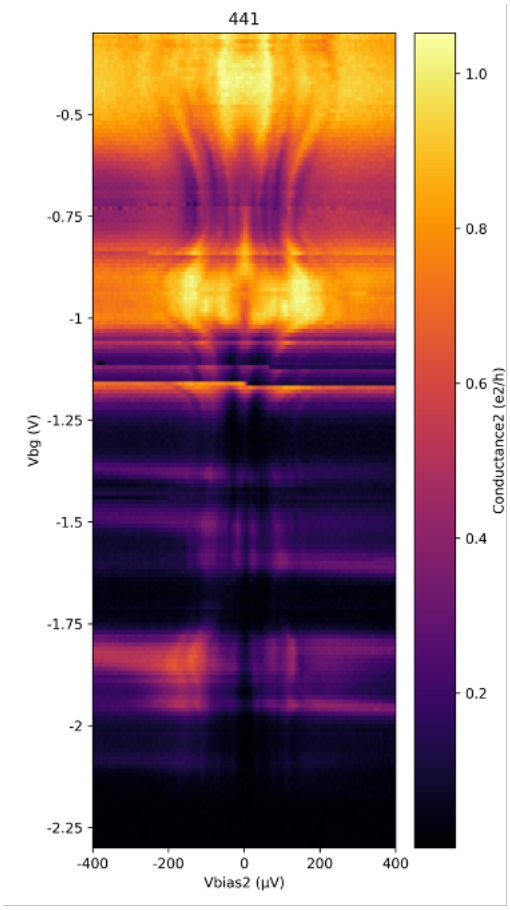


Fig. 3. Conductance versus source-drain voltage and back gate in the first lobe.

The rationale behind this part of the protocol was that reduced screening at strong depletion typically results in multiple potential maxima along the junction. In looking for a tunneling regime, the goal was to avoid multiple barriers along the junction. Only in that case, and for weak tunneling (< ~0.3 e^2/h), is tunneling current proportional to the local density of states, as discussed by Blonder, Tinkham, and Klapwijk (BTK). The local density of states is the quantity of interest in these studies.

Multiple barriers from weak screening *generically* create quantum dots (QDs) whose signatures are resonances. In fact, at low density, resonances may form even in the absence of disorder. This does not require fine-tuning, only an awareness that QDs mean that conductance is not proportional to the local density of states. The tunneling regime in modestly disordered junctions is therefore an intermediate condition: if the junction is too open, dI/dV is not proportional to the local density of states *a la* BTK; if the junction is too closed, reduced screening typically leads to resonances and disordered transport. Tuning tunnel barriers for the tunneling regime and avoiding quantum dots is a familiar part of mesoscopic physics, used whenever gate-controlled tunable barriers are needed for spectroscopy. In highly disordered junctions, many barriers in series give rise to a different effect, dynamical Coulomb blockade, a reduction of conductance around zero bias lacking periodic gate response. While essentially all devices showed a QD regime near pinch-off (i.e., gate voltages more negative than the tunneling regime), for some wire batches or

individual devices, dynamical Coulomb blockade was observed on top of any lobe structure. This feature, along with other signatures of strong disorder such as gate hysteresis, was taken to be a sign that something had gone badly with the junction during fabrication, and typically further measurement on that device was discontinued.

Next, on devices with an accessible tunneling regime—meaning that the device had a relatively resonance-free region below the open regime without strong, frequent resonance or dynamical Coulomb blockade—higher-resolution data were taken in the zeroth and first lobes, over the range from the open regime to the QD regime, spanning the tunneling regime (**Fig. 4**).

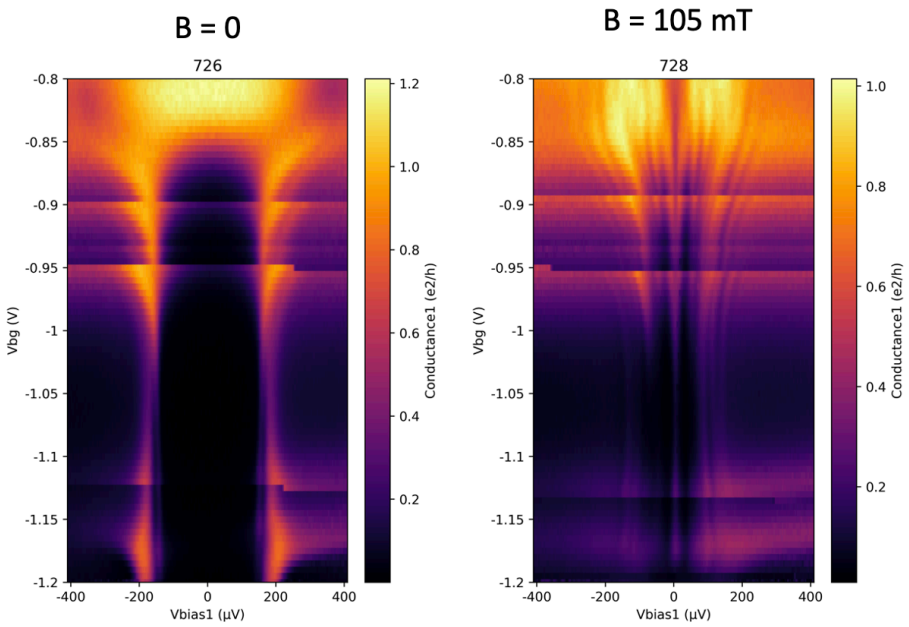


Fig. 4. High resolution scans of conductance versus source-drain voltage and back gate at zero field and in the middle of the first lobe.

Next, high-resolution magnetic field sweeps were taken at a few representative back-gate values within this range of back-gate voltages found above. Typically, 1-3 values of back-gate voltage were examined (**Fig. 5**).

For devices that did not show a good tunneling regime (e.g., most devices from wire batches 637 and 638, for unknown reasons), measurements were either discontinued or a few field sweeps were taken at gate-voltage values between resonances.

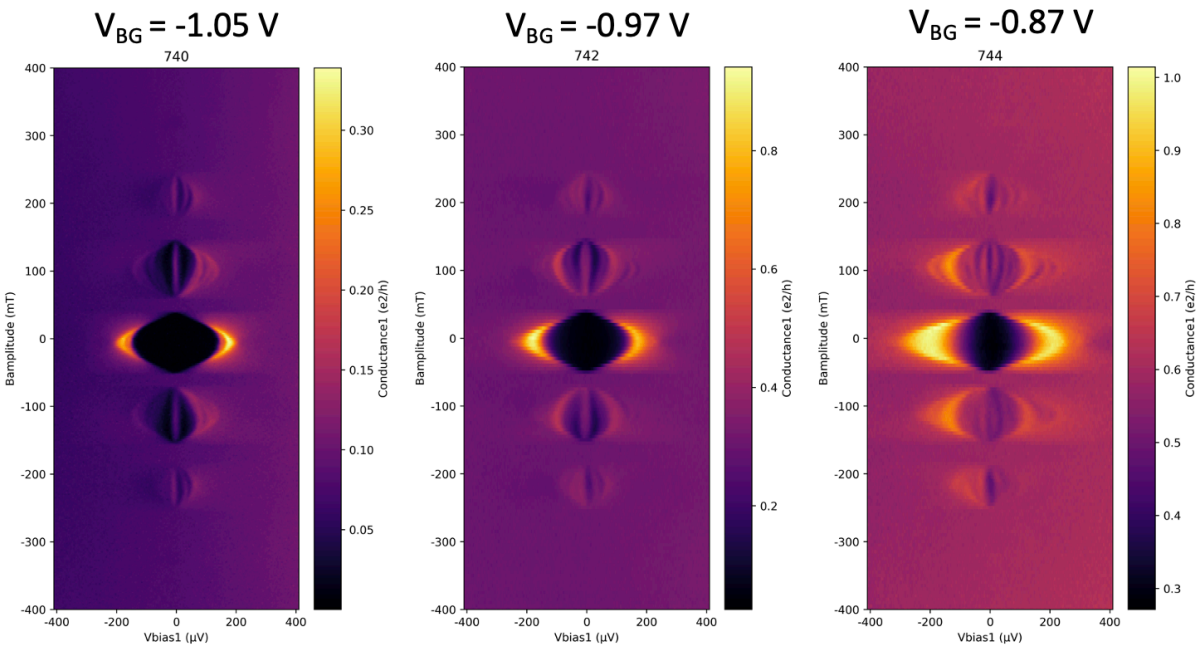


Fig. 5 Conductance as a function of source-drain voltage and axial magnetic field at three back gate voltages.

4. Protocol for CB measurement:

The availability of full-shell wires of length 10 μm allowed multiple devices to be fabricated on the same wire. This was important given the known quantitative variability among devices that were qualitatively the same, including devices from the same wire batch (for instance, Device 1 and Device 3). Multi-island devices were designed with two contacts and three side gates per island in addition to a global back gate, allowing the two tunnel barriers and island occupancy to be independently controlled. The four gates acting on each island are denoted *left*, *right*, *plunger*, and *back*. The protocol that was followed aimed to compare multiple islands measured on the same wire. We began by looking for devices with all island segments working.

Each chip contained up to four bonded devices. Within a cooldown, devices were measured island-by-island using manual lock-in measurements to check for continuity and gateability. The first chip with CB devices that was cooled had no CB devices with all segments working. Non-working means failing the continuity test and cannot be turned on with the side or back gates. These devices were not measured further. On the second chip with CB devices, one of the three devices checked on that chip during the first cooldown had all islands working. The other two had non-working islands. On a subsequent cooldown, a fourth device was examined. It also had a non-working island and was not measured further.

Each island on the all-working device on the second chip immediately showed gross features of CB, i.e., oscillations of conductance as a function of any of the gate voltages.

One at a time, each island was tuned into a symmetric weak-tunneling regime by tuning four gates per island. The procedure is illustrated with data for the shortest island (200 nm length) the same process was carried out serially for all segments. First, the back gate was swept until valley conductance vanished, as shown in **Fig. 6**. Next, the source-drain offset was zeroed (numerically) by examining conductance versus source-drain and plunger-gate voltages (**Fig. 7**). Next, the back gate was used to pinch off the island by sweeping until the CB peaks vanished (**Fig. 8**). At that value of back gate, the device was next symmetrized by examining sweeps of left and right gates (**Figs. 9, 10**), noting that this generally resulted in unequal voltages on left and right gates. Tuning left and right gates over large ranges revealed slowly varying nonmonotonic conductance, presumably due to resonances in the leads. The device was tuned for a rough maximum of peak height to find the rough diagonal where left and right gates had equal influence avoiding these nonmonotonic regions in left-right gate space. A typical large-range sweep of left gates is shown in **Fig. 11**. This procedure is typical of how quantum dots are tuned in a variety of contexts.

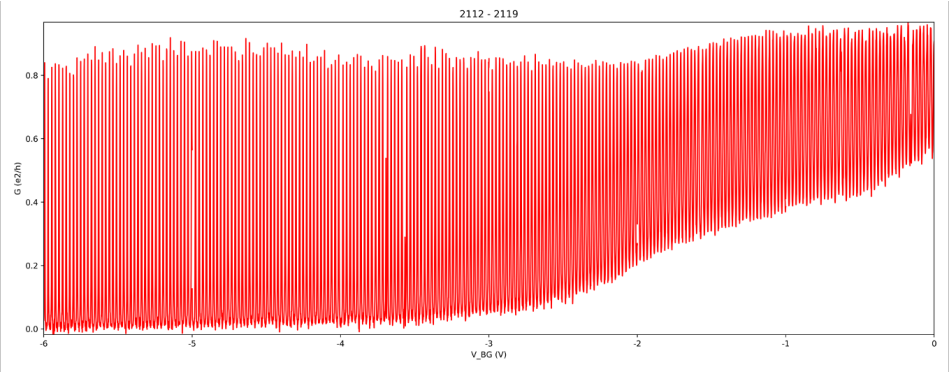


Fig. 6. Initial sweep of back gate to tune an island.

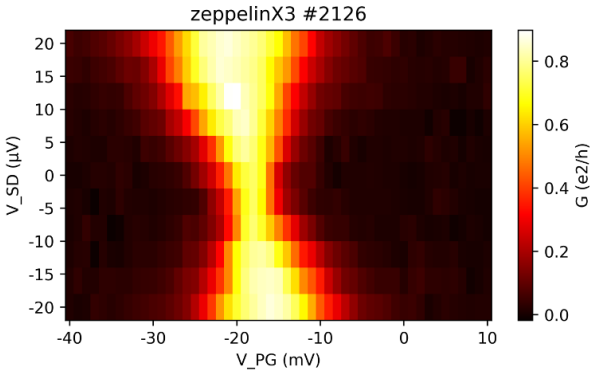


Fig. 7. Correcting source-drain bias offset.

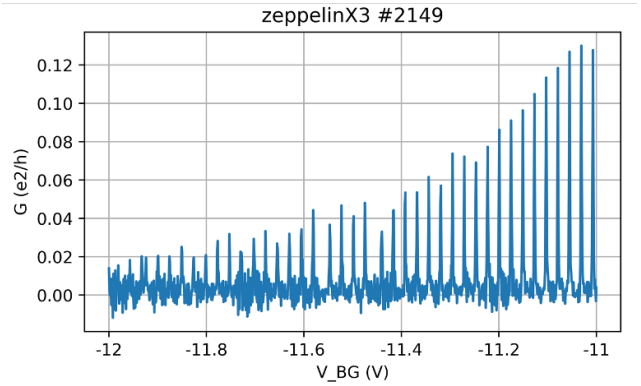


Fig. 8. Back gate pinch off. Conductance versus back gate.

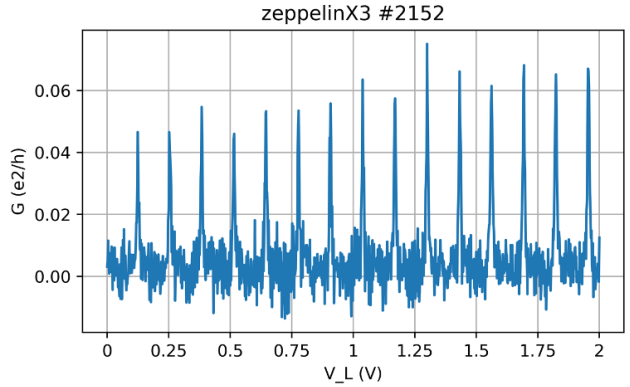


Fig. 9. Turning conductance back on with left gate.

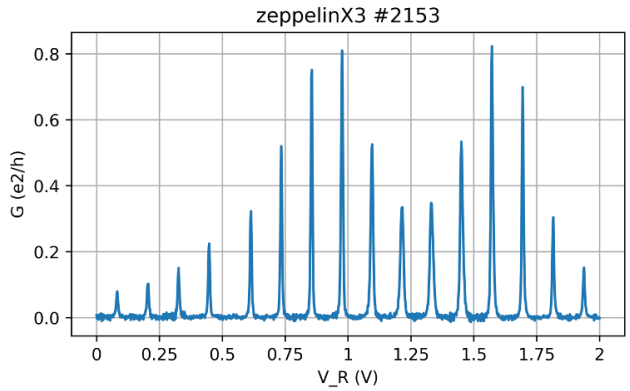


Fig. 10. Turning conductance back on with right gate.

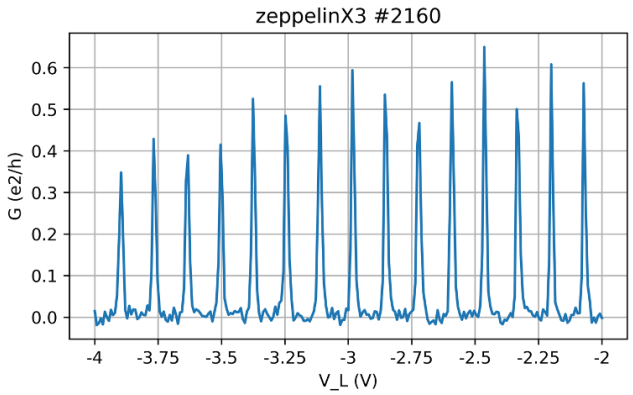


Fig. 11. Long-range left gate sweep.

After symmetrizing left and right conductances and avoiding nonmonotonic regions, a window of peaks was examined to find regimes in the plunger gate where the peak valleys went to zero conductance, peaks were moderately high (roughly 0.4 - 0.8 e^2/h) **Fig. 12**. This completes the island tune-up for one island.

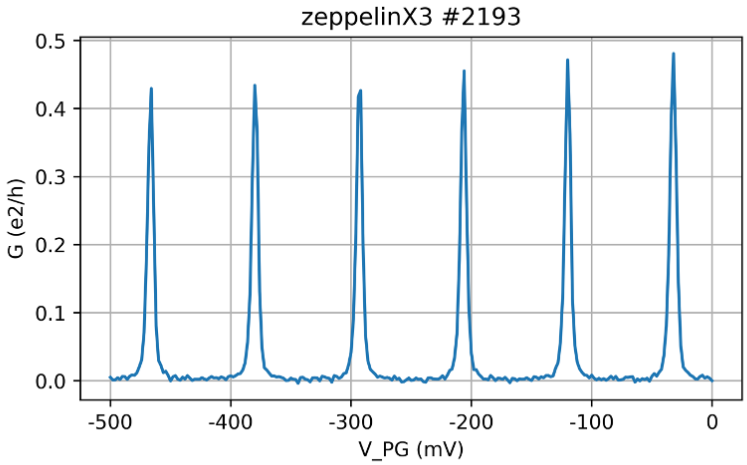


Fig. 12. Plunger gate sweep. Conductance versus plunger gate.

Tuning-up was carried out at zero magnetic field. Next, plunger gate sweeps on the tuned island were carried out at zero field, in the destructive regime, and in the middle of the first lobe (**Fig. 13**). The magnetic field was aligned to the wire axis following the same procedure as in the NIS devices, based on micrographs of the device. (Minor curvature of wires was not accounted for). The field magnitudes for the destructive regime and first lobe were known from previous measurements on the same wire batch. This was carried out as a check, which all segments passed.

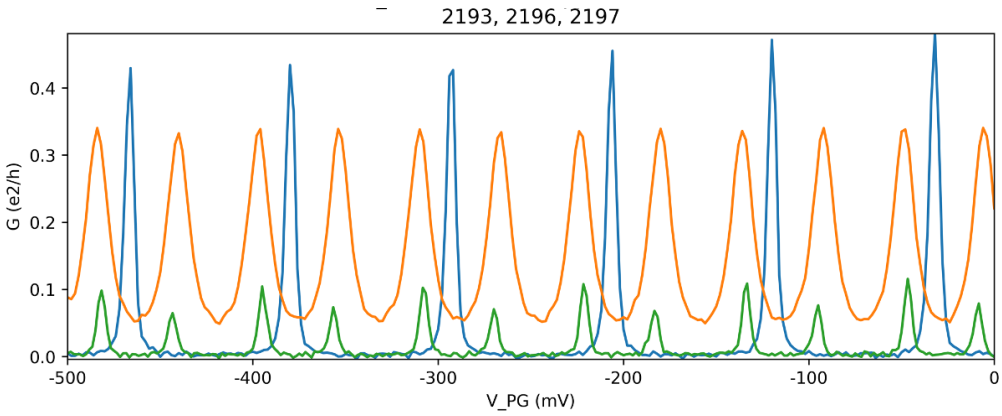


Fig. 13. Plunger gate sweep on the 200 nm island at zero field (blue) showing 2e peak spacing, destructive regime (orange) showing precise 1e peak spacing (orange), and in the middle of the first lobe, showing even-odd peak spacing (green).

Next, a high-resolution map of dI/dV as a function of magnetic field (outer loop) and plunger gate voltage (inner loop) was taken, spanning roughly 6-10 CB peaks and fields (**Fig. 14**). These runs typically took several hours and swept continuously from positive to negative field. The symmetry of the data in magnetic field provided a further check on device stability.

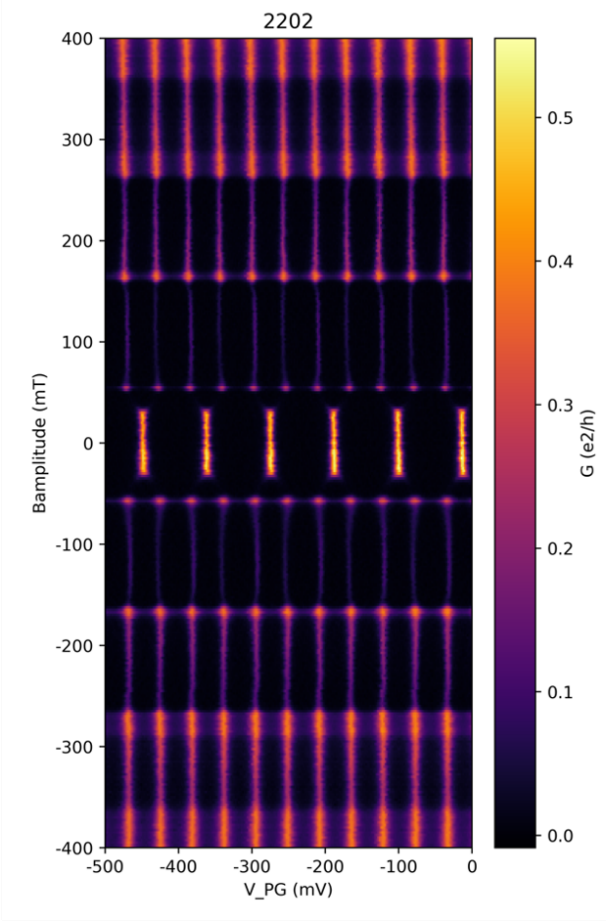


Fig. 14. Conductance as a function of plunger gate and axial magnetic field for 200 nm island.

Immediately after the field sweep, without changing gate voltages, conductance was measured as a function of source-drain voltage (inner loop) and plunger gate (outer loop). These "CB-diamond" runs (**Fig. 15**) provided a measure of the lever arm of allowing gate sweeps to be converted to energy. The fact that island energy levels detected by CB measurements depend sensitively on the lever arm, *though the mean CB spacing does not,* motivated taking separate measurements of CB diamonds in the zeroth lobe, the destructive regime, and the first lobe be made. The CB diamonds in the first lobe were especially important for converting peak spacing to energy. This conversion allowed different islands to be compared. Diamonds were taken at the value of the magnetic field where the peak spacing cuts were extracted from the 2D maps. Separate lever arms for the source and drain were extracted from leading and

trailing diamond edges to compensate for any remaining asymmetry in the coupling of the island to the two leads. **Figure 15** also shows the discrete subgap spectrum in the first lobe. This important feature is not visible in linear response (zero dc source-drain voltage) CB blockade data. **Figure 15** also shows that CB diamonds in the destructive regime were blurred, as expected for metallic islands, making it difficult to extract a lever arm there.

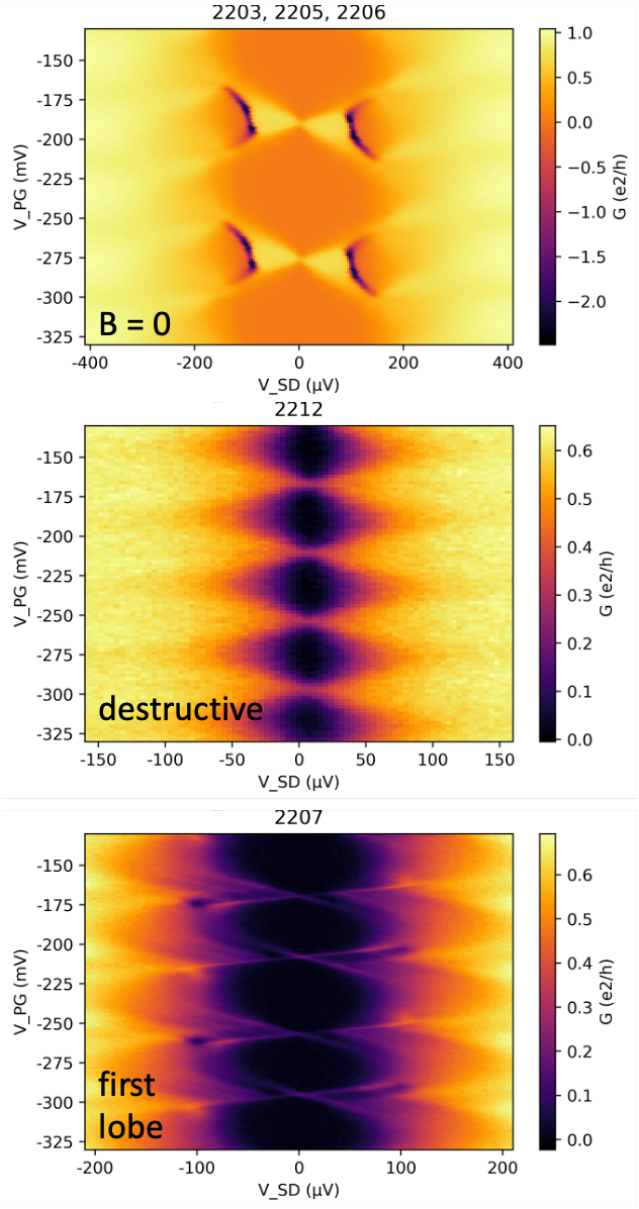


Fig. 15. Conductance as a function of source drain voltage and plunger gate at zero field (top), destructive regime (middle), and middle of the first lobe (bottom), used to measure lever arm. Note discrete subgap spectrum in the first lobe.

To increase statistics for each island, the procedure was repeated at 1-2 additional back gate voltages (with corresponding required tune-up of left and right gates).

For the 200 nm, 300 nm, 400 nm, and 600 nm islands, peak spacing fluctuations were extracted from the 2D maps of field and gate voltage. Then, CB diamonds measured in the first lobe were used to convert peak spacing to energy.

For the 800 nm and 1000 nm islands, where peak spacing was nearly uniform in the first lobe, plunger sweeps at very high resolution were taken at several discrete values of the magnetic field in the first lobe. For the 1000 nm island, increased statistics were obtained by repeating the 110 mT high-resolution sweep at one other back gate voltage (with re-tuned left and right gates). Data at 110 mT and 140 mT were further analyzed using the first-lobe lever-arm data from 110 mT to test the dependence of island-length scaling on the field-dependent gap.