Topological insulators and superconductors are very peculiar materials with a gap in the bulk electron spectrum and a low-lying branch of subgap excitations on their surface (see [1] for a review). This surface metallic state appears due to topological reasons and is robust with respect to any (sufficiently small) perturbations. In particular, topological properties prevent these surface states from Anderson localization. One common example of a topological insulator is a two-dimensional system in the integer quantum Hall effect regime. The bulk of such a system has a spectral gap between successive Landau levels and is hence an insulator. At the same time quantized Hall conductance appears due to a fixed integer number of chiral propagating edge modes on the background of the bulk gap. This type of materials are referred to as \( Z \) topological insulators.

Another type of topological insulator is realized in three-dimensional (3D) semiconductor crystals with sufficiently strong spin-orbit interaction (BiSb, BiTe, BiSe, strained HgTe etc). The spin-orbit interaction leads to inversion of the spectral gap. As a result subgap surface excitations appear with a dispersion of the massless Dirac type. The topological invariant in these materials has a \( Z_2 \) nature. When the number of surface states is odd, one of them always remains gapless due to topological protection.

A general classification of topological insulators was developed in Refs. [2, 3] based on the symmetry of the underlying Hamiltonian. The quantum Hall effect is an example of a topologically nontrivial state of the unitary symmetry (class A of the Altland-Zirnbauer classification [4]) in 2D. Strong spin-orbit interaction leads to a topological state in the symplectic class (AII) in 3D. Another important example is the 1D topological superconductor of the class BD symmetry (superconductor with both time-reversal and spin rotation symmetries broken). The topologically protected mode in this case is zero-dimensional and is known as the Majorana bound state (MBS). It appears, in particular, in the core of an Abrikosov vortex in the spinless \( p \)-wave superconductor [5]. A similar MBS appears [6] in a vortex in an ordinary s-wave superconductor brought in contact with the surface of a 3D \( Z_2 \) topological insulator, which corresponds to the symmetry class AII. The MBS appears as a descendant of the topologically protected massless Dirac state on the free surface of the topological insulator.

There are two general methods to observe the Majorana level. One of them relates to the anomalous Josephson effect [7, 8]. Another, and a more direct, way involves tunneling into the region where the MBS is supposed to be localized [9]. The differential conductance in such a tunneling experiment yields the local density of states with spatial and energy resolution. The Majorana state in the vortex core manifests itself by resonant Andreev reflection at zero energy [10].

In this Letter we study local tunneling conductance for a setup depicted in Fig. 1. A superconducting (with a spectral gap \( \Delta \)) film with a circular hole of the radius \( R \) is deposited on the surface of a 3D topological insulator, e.g., Bi\(_2\)Te\(_3\) crystal. Perpendicular magnetic field applied to the system produces a vortex pinned to the hole. The radius \( R \) is supposed to be relatively large,
is a conjugate eigenstate $s$ of the Hamiltonian $H$ that poses boundary conditions for the Hamiltonian $H$. This creates a gap in the electronic spectrum outside the hole. To analyze this, we assume the dirty limit with a disorder induced mean free path $l$ from the Dirac point by the random disorder potential. The Fermi level is shifted by $\mu$. At the same time the total BdG Hamiltonian $H_{\text{BdG}}$ counts every physical excitation twice due to the doubling in NG space. An unpaired level is thus "half" of a true excitation — a Majorana state.

The particle-hole symmetry $Q$ defines the class BD. On the level of random matrices there are two distinct versions of this class with even and odd number of eigenstates referred to as D and B class, respectively. In class B, one unpaired eigenstate has exactly zero energy and is self-conjugate: $\psi = s_y \tau_y \psi^*$. The BdG Hamiltonian counts every physical excitation twice due to the doubling in NG space. An unpaired level is thus "half" of a true excitation — a Majorana state.

We will calculate the density of states inside the hole with the help of the supersymmetric non-linear sigma model. Two-dimensional Dirac fermions with potential disorder are described by a very peculiar model of the class AII with $\mathbb{Z}_2$ topological term $\cal{Z}_2$. This topological term appears as a consequence of the chiral anomaly of Dirac fermions. We will consider the minimal model operating with the $8 \times 8$ supermatrix $Q$. Apart from Fermi-Bose superspace, this matrix operates in the space of retarded-advanced (RA) fields and in a specific time-reversal (TR) space introduced to take into account Cooperons and diffusons on equal footing. RA space is completely analogous to the NG space (see below); we denote Pauli matrices in this space by $\tau$. Notation $\sigma$ is used for Pauli matrices in TR space. A detailed derivation of the sigma model can be found in [12].

In class AII, the matrix $Q$ obeys the non-linear constraint $Q^2 = 1$ and the linear constraint

$$Q = Q = C Q^T C^T, \quad C = \tau_x \left( \begin{array}{cc} \sigma_x & 0 \\ 0 & i \sigma_y \end{array} \right).$$

As a result, $Q$ contains 8 commuting and 8 anticommuting (Grassmann) variables in total. Commuting degrees of freedom parameterize diagonal fermion-fermion (F) and boson-boson (B) blocks of $Q$. The general Q-matrix can be decomposed as

$$Q = U^{-1} \left( \begin{array}{cc} Q_F & 0 \\ 0 & Q_B \end{array} \right) U,$$

where the central part contains only commuting variables while Grassmann parameters define the unitary supermatrix $U$. We will use the following explicit form of the central part of $Q$ in terms of eight angle parameters:

$$Q_F = \left[ \begin{array}{cc} \tau_z \cos \theta_f + \sigma_x \sin \theta_f (\tau_x \cos \phi_f + \tau_y \sin \phi_f) \\ \sigma_x \cos k_f + \tau_z \sin k_f (\sigma_x \cos \chi_f + \sigma_y \sin \chi_f) \end{array} \right] \times \left[ \begin{array}{cc} \sigma_x \cos k_f + \tau_z \sin k_f (\sigma_x \cos \chi_f + \sigma_y \sin \chi_f) \\ \tau_z \cos \phi_f + \tau_y \sin \phi_f \end{array} \right].$$

This representation fulfills all the constraints imposed by the symmetry class AII of the Hamiltonian $H_0$. The F and B sectors of the sigma model are compact and non-compact, respectively. This is achieved by demanding that the angles $\theta_b, k_b$ are imaginary while all other angles in Eqs. (6), (7) are real. Below we will find that the saddle point describing the density of states in the hole occurs
with real $\theta_b$. This implies a proper shift of the integration contour for this angle.

The sigma model of class AII is designed to study transport properties of a disordered system. This implies that the $Q$ matrix operates, in particular, in the RA space allowing for averaging the product of retarded and advanced Green functions. We are interested in the density of states and hence it suffices to average just the single retarded function. At the same time, superconducting boundary conditions implemented in the BdG Hamiltonian require to introduce an additional doubling of fields in the Nambu-Gor’kov space. This can be achieved within the standard AII class sigma model with the $8 \times 8$ supermatrix $Q$ while the role of NG space is played by the Usadel equation with this boundary condition:

$$\theta_1 = 2 \arctan(r/R),$$  \hspace{1cm} (11)

$$\theta_2 = \pi - 2 \arctan(r/R).$$  \hspace{1cm} (12)

Saddle point equations also require $k_f = 0$ and hence the angle $\chi_f$ drops from the matrix $Q$ and from the action. The two remaining angles $k_b$ and $\chi_b$ are free and can take any constant values.

The spatial profile of $Q$ is thus fixed by the Usadel equation. The solutions $\theta_{1,2}$ represent two disconnected saddle points in the F sector while in the B sector only the saddle point $\theta_b = \theta_1$ is reachable. If the integration contour for $\theta_b$, which runs along the imaginary axis, is shifted to the point $\theta_2$ a divergence occurs in the $k_b$ integral. Thus the B sector is reduced to a hyperboloid parameterized by $\chi_b > 0$ and $0 < \chi_b < 2\pi$, see Ref. [13]. This is exactly the structure of the sigma model of class BD as we anticipated from symmetry analysis. The distinction between even (D) and odd (B) versions of this class is related to the disjoint character of the manifold due to the discrete degree of freedom in the F sector. Namely, in class D (B) the two parts of the manifold contribute to the partition function with the same (opposite) sign. In our problem the odd symmetry class B occurs. To demonstrate it, we compare the value of the action $S$ at the two minima in the F sector. These two solutions indeed contribute with the opposite sign since the corresponding values of the topological term $\Theta$ differ by exactly $i\pi$.

The density of states is given by the integral

$$\rho(E, r) = \frac{\nu}{8} \Re \int DQ \text{Str}[k\Lambda Q(r)] e^{-S[Q]}.$$  \hspace{1cm} (13)

At low energies, this integral is to be calculated over the saddle manifold. Apart from the two variables $k_b$ and $\chi_b$ and two disconnected points in the F sector this manifold involves two Grassmann variables in the matrix $U$, Eq. [13]. In order not to spoil the saddle point, these variables must be constant in space and fulfill the condition $[U, \Delta] = 0$. We introduce them according to

$$U = \exp \left[ \frac{1}{2} \begin{pmatrix} 0 & \eta \sigma_x + \zeta \sigma_y \\ -i \eta \sigma_y & 0 \end{pmatrix}_{FB} \right].$$  \hspace{1cm} (14)

Within this parameterization, we can rewrite the integral [13] in terms of $k_b, \chi_b, \eta, \zeta$. Thus the problem is reduced to the 0D sigma model of class B (the two disjoint points in the F sector contribute with opposite signs). Explicit calculation of the integral [13, 12] yields the local DOS in the factorized form

$$\rho(r, E) = \nu n(r) f(E/\omega_0),$$  \hspace{1cm} (15)
energies much less than the Thouless energy $E_T^\gamma$. The Majorana state appears as a half of a fermionic level. Result [15] up to a factor 2 due to BdG double counting.

In the extreme tunneling limit $\nu \rightarrow 0$, the solid blue oscillating curve represents the low-energy result (19), the dashed curve shows the high-energy result (20). Oscillating coefficient for the B class contains a narrow lorentzian characteristic for the B class and contains a narrow lorentzian.

The results (22) and (23) apply at low temperature and voltage. When temperature and/or voltage are higher than the level spacing $\omega_0$, all the resonances get smeared and the normal average DOS $\nu$ is recovered.

The noise produced by the Majorana level is T- and V-independent as long as $\gamma \omega_0 \ll T$. The results (22) and (23) apply at low temperature and voltage. When temperature and/or voltage are higher than the level spacing $\omega_0$, all the resonances get smeared and the normal average DOS $\nu$ is recovered.

To conclude, we have studied the local density of states in the superconducting vortex on the surface of a topological insulator in the superconducting antidot setup depicted in Fig. 1. The spatial and energy dependence of the density of states factorize at low energies and the latter is given by the 0D sigma model of symmetry class B. We have identified the zero-energy Majorana state occurring due to the topological properties of the system. This Majorana state exhibits itself via the resonant Andreev reflection at zero energy yielding the peak in differential conductance with the universal amplitude $e^2/\pi h$ and width proportional to the normal conductance $G$. We are grateful to B. Sacepe for stimulating discussions. This work was supported by the RFBR grant 10-02-00554-a and by the German Ministry of Education and Research (BMBF).
Derivation of the sigma-model

Sigma model for hybrid structure

In this section we derive the sigma model describing the dynamics of the topological insulator surface excitations in the presence of a random potential. We assume the hole geometry depicted in Fig. 1 of the main text. The derivation starts with the microscopic Bogolyubov – de Gennes Hamiltonian \((1)\). The sigma model is aimed at averaging the single retarded Green function determining the density of states. We will also demonstrate the equivalence of this sigma model to the model of the symplectic symmetry class AII describing the transport properties of massless Dirac fermions. It is known that the sigma model for Dirac fermions possesses the specific \(Z_2\) topological term. The equivalence of the two models will prove the appearance of the topological term in the superconducting case considered in the present paper.

Let us start with the retarded Green function in the system governed by the Hamiltonian \((1)\). We will use the supersymmetric integral representation

\[
G^R_{E}(r, r') = -i \int D\Phi^* D\Phi \, \text{str} \, [k\Phi(r)\Phi^\dagger(r')] \, e^{-S}, \quad S = -i \int d^2r \, \Phi^\dagger(E + i0 - H)\Phi. \tag{24}
\]

Here the vector fields \(\Phi\) and \(\Phi^\dagger\) contain 4 commuting and 4 Grassmann parameters each. Apart from the spin and Bogolyubov – de Gennes structure of the Hamiltonian, we also introduce the superstructure in order to get rid of the normalizing denominator and facilitate further disorder averaging. The pre-exponential factor contains the supermatrix \(k = \{(1, -1)\}_\text{FB}\) and supertrace is defined as in Ref. \([17]\): \(\text{str} A = A_{\text{FF}} - A_{\text{BB}}\).

The superconducting symmetry of the Hamiltonian \(H\), Eq. \((26)\), gives rise to specific soft modes – Cooperons – of the type \(\langle G^R_{E}G^{-}_E\rangle\). These modes are relevant for the average density of states since they are built out of retarded functions only. In order to include them into our effective theory, we transform the action by writing half of it in the form \((24)\) and another half in the time-reversed (transposed) form:

\[
S = -\frac{i}{2} \int d^2r \left[ \Phi^\dagger(E + i0 - H)\Phi - \Phi^T k(E + i0 + s_y\tau_y H s_y\tau_y)\Phi^* \right]
= -\frac{i}{2} \int d^2r \left( \Phi^\dagger \tau_z, \, i\Phi^T k s_y\tau_x \right) \left[ (E + i0)\sigma_z\tau_z - H_0 - \tau_z \hat{\Delta} \right] \left( s_y\tau_y \Phi^* \right). \tag{25}
\]

Here we have introduced the matrix \(\sigma_z\) operating in the space of time-reversed blocks – TR space. The matrix \(k\) appeared due to anticommutation of Grassmann variables. We will denote the doubled vectors as

\[
\Psi = \frac{1}{\sqrt{2}} \begin{pmatrix} \Phi \\ s_y\tau_y \Phi^* \end{pmatrix}, \quad \bar{\Psi} = \frac{1}{\sqrt{2}} \begin{pmatrix} \Phi^\dagger \tau_z, \, i\Phi^T k s_y\tau_x \end{pmatrix}. \tag{26}
\]

They are no longer independent, as \(\Phi\) and \(\Phi^*\) were, but rather obey the linear constraint

\[
\bar{\Psi} = (C\Psi)^T, \quad C = -i s_y\tau_x \begin{pmatrix} \sigma_x & 0 \\ 0 & i\sigma_y \end{pmatrix}_{\text{FB}}. \tag{27}
\]
At this stage, we are ready to perform the disorder averaging. We adopt the standard Gaussian white noise disorder model characterized by the correlator
\[
\langle V(\mathbf{r})V(\mathbf{r}') \rangle = \frac{\delta(\mathbf{r} - \mathbf{r}')}{\pi \nu \tau}. \tag{28}
\]
The single-particle density of states is defined as \( \nu = |\mu|/2\pi v_F^2 \). Note that the right-hand side is twice bigger compared to the standard definition applied to normal metals. The reason for this extra factor 2 is the Dirac nature of electron spectrum and hence doubling of the phase space due to the two types of excitations (electrons and holes).

Disorder averaging produces the quartic term in the action,
\[
S = \int d^2r \left\{ \frac{(\bar{\Psi}\Psi)^2}{2\pi \nu \tau} - i\bar{\Psi} \left[ (E + i0)\Lambda + \mu - v_0 s \cdot p - \tau_z \hat{\Delta} \right] \Psi \right\}. \tag{29}
\]
Here we use the notation \( \Lambda = \sigma_z \tau_z \). The interaction term is further decoupled with the help of the Hubbard-Stratonovich transformation by introducing an auxiliary matrix field \( Q \). The construction of the nonlinear sigma model implies that the field \( Q \) contains all relevant slow modes of the disordered system – diffusons and Cooperons. There are in total three ways to decouple the four-fermion interaction \([17]\). One of them involves a scalar field \( Q \). This leads only to an irrelevant renormalization of the chemical potential. Two other decoupling schemes introduce a matrix field with the structure \( \langle \bar{\Psi} \Psi \rangle \) and \( \langle \bar{\Psi} \Psi^T \rangle \) corresponding to diffusons and Cooperons, respectively. In order to include all relevant slow modes in the theory, we have to perform decoupling in both ways, which can be achieved with a single matrix \( Q \). Details of this calculation can be found in Ref. [17]. The resulting action has the form:
\[
S = \frac{\pi \nu}{16\tau} \text{str} Q^2 - i \int d^2r \bar{\Psi} \left[ E\Lambda + \mu - v_0 s \cdot p - \tau_z \hat{\Delta} + \frac{iQ}{2\tau} \right] \Psi. \tag{30}
\]
The vectors \( \Psi \) and \( \bar{\Psi} \) are related by Eq. (27). This allows us to limit the matrix \( Q \) by the linear constraint \( Q = \bar{Q} \equiv CQ^T C^T \). This relation keeps only those parameters in \( Q \) that couple to the product \( \bar{\Psi} \bar{\Psi} \). Integrating out the field \( \bar{\Psi} \) yields the action for the matrix \( Q \):
\[
S = \frac{\pi \nu}{16\tau} \text{str} Q^2 - \frac{1}{2} \text{str} \ln \left[ E\Lambda + \mu - v_0 s \cdot p - \tau_z \hat{\Delta} + \frac{iQ}{2\tau} \right]. \tag{31}
\]
Derivation of the sigma model proceeds with the saddle-point analysis of the above action. In the dirty limit, the energy and \( \hat{\Delta} \) terms in the argument of the logarithm are relatively small compared to \( Q/\tau \). With these small terms neglected, the action possesses the uniform saddle point \( Q = \Lambda \). This saddle point corresponds to the self energy of the average electron Green function in the self-consistent Born approximation. Other saddle points can be achieved by rotations \( Q = T^{-1} \Lambda T \), if the matrix \( T \) commutes with the spin operator \( s \) (\( E \) and \( \hat{\Delta} \) terms are still neglected). Rotations \( T \) define the target manifold of the non-linear sigma model. The effective action of the sigma model within this manifold is derived with the help of the gradient expansion of Eq. (31) allowing for slow spatial variation of \( Q \) and perturbative expansion to the linear order in \( E \) and \( \hat{\Delta} \). Since the \( Q \) matrix is trivial in the spin space, we can safely reduce its size to \( 8 \times 8 \) keeping only Nambu, TR, and FB structure. Then the self-conjugacy relation acquires the form of Eq. (4).

The gradient expansion of Eq. (31) is a highly nontrivial procedure in view of the chiral anomaly of the Dirac fermions. The momentum integrals arising after the expansion of the logarithm are divergent in the ultraviolet limit and require a proper regularization. The result of the expansion is independent of a particular regularization scheme provided the gauge invariance is preserved. The anomaly affects only the imaginary part of the action and leads to the appearance of the topological term \([12]\). At the same time, the real part can be obtained in a straightforward way since all the arising momentum integrals are convergent. The result reads
\[
\text{Re} S = \frac{\pi \nu}{8} \int d^2r \text{str} \left[ D(\nabla Q)^2 + 4i(E\Lambda - \tau_z \hat{\Delta})Q \right]. \tag{32}
\]
Here the diffusion coefficient is \( D = v_F^2/2 \) and the matrix \( Q \) is reduced to the \( 8 \times 8 \) size. The action (32) used in the main text, differs from Eq. (32) only by a \( \pi/2 \) rotation of the superconducting phase and by an imaginary contribution to the energy term. The latter corresponds to a finite dwell time of the electron due to the coupling to the tunneling microscope tip as we elaborate below.

In order to explain the emergence of the topological term in the imaginary part of the action, we will prove the equivalence of the sigma model derived here for the hybrid structure of Fig. 1 to the symplectic sigma model for Dirac fermions derived in Ref. [12].
The sigma model obtained by the gradient expansion of Eq. (31) belongs to the symplectic symmetry class AII. This is quite natural since the Hamiltonian possesses only time-reversal symmetry in the absence of the $\Delta$ term. Let us demonstrate the equivalence between our sigma model (8) describing the density of states for the Hamiltonian $H_0$ derived in the previous section and used in the main text. This is quite natural since the Hamiltonian possesses only time-reversal symmetry in the absence of the $\hat{\Delta}$ term. Let the Wess-Zumino-Witten term, cf. Ref. [18].

These relations bring the charge conjugation matrix $\Lambda_0$ operating in the RA space. Subscript 0 is used to distinguish this model from the sigma model derived in the previous section and used in the main text. The time-reversal symmetry $H_0 = s_y H_0^T s_y$ is taken into account by further doubling of variables in the TR space,

$$S_0 = -\frac{i}{2} \int d^2r \left[ \Phi^T (\omega/2 + i0 - \tau_z H_0) \Phi - \Phi^T k (\omega/2 + i0 - \tau_z s_y H_0 s_y) \Phi^* \right]$$

Note that the matrix $\Lambda_0 = \tau_z$ multiplying the frequency term is different from its counterpart $\Lambda$ from the previous section. The vectors $\Psi$ and $\bar{\Psi}$ as well as the charge conjugation matrix $C_0$ are also defined differently,

$$\Psi = \frac{1}{\sqrt{2}} \begin{pmatrix} \Phi \\ i s_y \Phi^* \end{pmatrix}, \quad \bar{\Psi} = (C_0 \Psi)^T = \frac{1}{\sqrt{2}} \begin{pmatrix} \Phi^T \tau_z, & i \Phi^T k s_y \tau_z \end{pmatrix}, \quad C_0 = -i s_y \tau_z \begin{pmatrix} 0 & \sigma_y \\ \sigma_x & 0 \end{pmatrix}_{FB}. \quad (35)$$

Already at this stage we can prove the equivalence of the two models by selecting a proper unitary rotation of the field vector $\Psi$. This unitary rotation should obey the following properties:

$$U^T C_0 U = C, \quad U^{-1} \Lambda_0 U = \Lambda. \quad (36)$$

These relations bring the charge conjugation matrix $C_0$ and the matrix $\Lambda_0$ to the representation used in the previous section for the sigma model of the hybrid system. There are many matrices $U$ that fulfill identities (36). One possible choice is

$$U = \begin{pmatrix} 1 & 0 \\ 0 & i \tau_y \end{pmatrix}_{TR}. \quad (37)$$

In the case of usual rather than Dirac fermions, an analogous equivalence between the sigma model for the density of states in a superconducting hybrid structure and the orthogonal class sigma model is discussed in Ref. [14].

### Topological term

So far we have discussed the derivation of the real part of the sigma-model action (8) and also proved the equivalence of this model to the symplectic class sigma model for Dirac fermions (up to the boundary conditions involving the term $\Delta$). The latter model is known to possess the $\mathbb{Z}_2$ topological term as an imaginary part of its action, see Ref. [12]. This allows us to include the topological term in the action (8) and thus complete its derivation.

The explicit form of the topological term is written in the main text, Eq. (19), in a noninvariant form, using the parameterization (5) – (7). In this section we will discuss an indirect but explicitly invariant representation of this topological term.

The target manifold of the symplectic class sigma model is fixed by the representation $Q = T^{-1} \Lambda T$ with the constraint $Q = \bar{Q}$. If we neglect the Grassmann parameters and consider the central part of $Q$, see Eq. (19), these conditions yield $Q_F \in O(4)/O(2) \times O(2)$ and $Q_B \in Sp(2, 2)/Sp(2) \times Sp(2)$. The topological term arises in the compact F sector of the model. An explicit expression for the topological term can be written within a construction similar to the Wess-Zumino-Witten term, cf. Ref. [18].
We first extend the target manifold by relaxing the condition $Q^2 = 1$. Let us introduce the matrix $Q = \bar{T} \Lambda T$. The only restriction on this matrix is $Q = \bar{Q}$. The F sector of the extended manifold is $Q_F \in O(4)$. It has trivial second homotopy group, $\pi_2[O(4)] = 0$. This allows us to introduce the third, auxiliary, coordinate $0 \leq t \leq 1$, such that the real 2D space corresponds to $t = 1$, and continuously extend the matrix field according to

$$Q(r, t) = \begin{cases} Q(r), & t = 1, \\ \Lambda, & t = 0. \end{cases}$$  \hspace{1cm} (38)$$

Such an extension assumes that the physical space has no boundary and can be viewed as a surface of a 3D ball with $t$ being the radial coordinate. With this definition of $Q$, the Wess-Zumino-Witten term has the form

$$S_0[Q] = \frac{i\epsilon_{abc}}{24\pi} \int_0^1 dt \int d^2r \text{str} \left[ Q^{-1}(\nabla_a Q)Q^{-1}(\nabla_b Q)Q^{-1}(\nabla_c Q) \right].$$  \hspace{1cm} (39)$$

Here the indices $a$, $b$, and $c$ take three values corresponding to the three coordinates $t$, $x$, and $y$, and $\epsilon_{abc}$ is the unit antisymmetric tensor.

The integrand in this expression is a total derivative therefore the result of the integration depends only on the value of $Q$ at the boundary of integration domain, i.e., on the physical field $Q = Q|_{t=1}$. This can be checked by an explicit calculation of the variation of $S_0$:

$$\delta S_0[Q] = \frac{i\epsilon_{abc}}{8\pi} \int_0^1 dt \int d^2r \nabla_a \text{str} \left[ Q^{-1}\delta Q Q^{-1}(\nabla_b Q)Q^{-1}(\nabla_c Q) \right] = \frac{i}{8\pi} \int d^2r \text{str} \left\{ Q^{-1}\delta Q \left[ Q^{-1}(\nabla_x Q), Q^{-1}(\nabla_y Q) \right] \right\}.$$  \hspace{1cm} (40)$$

A topological term does not change under continuous variations of the field $Q$ but takes different values in different (disconnected) topological sectors of the model. The general Wess-Zumino-Witten term has a nonzero variation and hence does not obey this property. However, the term (39) is constrained by $Q^2 = 1$. Under this condition, the variation (40) is identically zero. Hence the Wess-Zumino-Witten term (39) indeed plays the role of a topological term.

The value of the topological term is quantized, yielding the topological charge of the field configuration, only in a system without boundary. This is not the case for the geometry considered in the paper. The matrix $Q$ is defined in the finite region inside the hole (see Fig. 1 of the main text) with the boundary conditions fixed by the $\bar{\Delta}$ term in the Hamiltonian (1). The value of the topological term is not completely fixed in this case. In particular, the expression depends not only on physical values of $Q$ but also on the way it is extended in the third dimension, Eq. (38). However, this uncertainty leads only to an uncontrolled imaginary constant in the action. This constant is the same in both topological sectors of the model and hence does not alter any observable quantities.

An alternative derivation of the sigma-model action, including the topological term in the form Eq. (39) is possible within the non-abelian bosonization formalism. Let us for a moment assume, that the Hamiltonian has an extra chiral symmetry, $H = -s_z H s_z$. This situation corresponds to the symmetry class DIII and can be realized, e.g., at the Dirac point in the presence of a random velocity disorder. The sigma model of the class DIII has an extended target space, corresponding to the manifold of $Q$ introduced above. This sigma model possesses the Wess-Zumino-Witten term (39), see Ref. [14], and is the result of the non-abelian bosonization of the initial fermionic problem. Non-zero chemical potential drives the system away from the Dirac point and breaks the chiral symmetry. This reduces the model to the symplectic symmetry class and restricts the field by the condition $Q^2 = 1$. The resulting action has the topological term in the form of the restricted Wess-Zumino-Witten term (39) as discussed above.

**Tunneling coupling term**

In this section we discuss the appearance of the imaginary contribution to the energy due to the presence of the tunneling tip. In order to include the coupling to an external metallic probe, we extend the Hamiltonian as

$$H = \begin{pmatrix} H & t \\ t^\dagger & H_M \end{pmatrix}.$$  \hspace{1cm} (41)$$

Diagonal blocks of this matrix are the Hamiltonian $H$ of the topological insulator surface (1) and the Hamiltonian $H_M$ describing electron states in the metallic tip. Off-diagonal elements $t$ and $t^\dagger$ are tunneling amplitudes between the two subsystems. We assume these tunneling amplitudes are local in space and couple the states at point $r_0$ on the sample surface to the states at the tip of the metallic probe.
Deriving the sigma model as described in previous sections we arrive at the expression similar to Eq. (31) but with an extended matrix structure:

\[ S = \frac{\pi \nu}{16\pi} \text{str} Q^2 + \frac{\pi \nu M}{16\pi M} \text{str} Q_M^2 - \frac{1}{2} \text{str} \ln \left[ \frac{(E\Lambda - \tau_2 H + iQ/2\tau)}{0} \right]. \]  

(42)

The quantities \( \nu_M, \tau_M, \) and \( Q_M \) refer to the metallic tip. In the absence of tunneling, \( t = 0 \), the two subsystems decouple and we obtain independent sigma models describing topological insulator and the tip. We will assume the density of states \( \nu_M \) to be large, that will fix \( Q_M = \Lambda \). Tunneling coupling is assumed to be weak. This allows us to expand the logarithm in powers of \( t \) and \( t^\dagger \). The lowest non-vanishing contribution appears in the second order of the perturbation theory and yields the tunneling term in the sigma model,

\[ S_t = \frac{1}{2} \text{str} \left[ (E\Lambda - \tau_2 H + iQ/2\tau)^{-1} \tau_2 t (E\Lambda - \tau_2 H M + iQ_M/2\tau M)^{-1} \tau_2 t^\dagger \right] = -\frac{\pi^2}{8} \nu_M \text{str} \left( Q\tau_2 t Q_M \tau_2 t^\dagger \right). \]  

(43)

In the last expression we have used the saddle-point relation between \( Q \) and the Green function \( (E\Lambda - \tau_2 H + iQ/2\tau)^{-1} \) in coincident points. We introduce normal dimensionless (in units \( c^2/h \)) tunnel conductance of the junction \( G_t = \pi^2 \nu_M \text{tr}(t) \) (note that \( \nu \) and \( \nu_M \) are total rather than spinless densities of states at the Fermi energy). Assuming \( Q_M = \Lambda \), we rewrite the action as

\[ S_t = -\frac{G_t}{8} \text{str} \Lambda Q. \]  

(44)

Thus the tunneling term has the same structure as the energy term and produces an imaginary contribution to \( E \), see Eq. (8) of the main text.

**Calculation of the density of states**

**Low energies \( E \ll E_{Th} \)**

The local density of states on the surface of the topological insulator is given by Eq. (13) of the main text. The integral over matrix \( Q \) is to be calculated over the saddle manifold fixed by the solution of the Usadel equation (10). There are two distinct solutions \( \theta_{1,2} \) [see Eqs. (11), (12) of the main text] for the angle \( \theta_Q \). In the F sector of the model, all other parameters are fixed and thus \( \theta_{1,2} \) represent two disjoint parts of the integration manifold. At the same time, only the solution \( \theta_1 = \theta_2 \) is allowed in the B sector, and two other angles, \( k_B \) and \( \chi_B \), are free forming a hyperboloid (the parameter \( k_B \) must be imaginary in order to ensure convergence of the integral).

To calculate the density of states from Eq. (13), we have to express \( \text{Str}(k\Lambda Q(r)) \), \( S[Q] \) and \( DQ \) in terms of \( k_B, \chi_B, \eta, \) and \( \zeta \) for the two disjoint parts of the manifold. However, a direct use of our parameterization leads to an uncertainty in the integral over the part corresponding to \( \theta_f = \theta_2 \). The integral over \( d\eta d\zeta \) is zero, while the integral over \( k_B \) diverges. We resolve this uncertainty by a trick proposed in Ref. [15].

The two disjoint parts of the integration manifold can be mapped onto each other by the similarity transformation

\[ Q \mapsto T^{-1} QT, \quad T = \begin{pmatrix} \sigma_x & 0 \\ 0 & 1 \end{pmatrix}_{FB}. \]  

(45)

We use the parametrization \( \theta \) in the domain with \( \theta_f = \theta_1 \) and then apply the transformation \( (45) \) to cover the second topological sector with \( \theta_f = \theta_2 \). Owing to the relation \( (45) \), integration measure is the same in both sectors. It is given by the superdeterminant of the metric tensor in the space of \( Q \) matrices, see Ref. [17]. The integration measure reads

\[ DQ = \frac{1}{2\pi} \tanh \frac{\kappa}{2} d\kappa d\chi_B d\eta d\zeta, \]  

(46)

with \( \kappa = ik_B \) (\( k_B \) is imaginary to ensure the convergence of the integral). The integration runs over the hyperboloid \( \kappa > 0, 0 < \chi_B < 2\pi \).

The action in the two topological sectors \( \theta_f = \theta_{1,2} \) has the form

\[ S_1 = -4i\pi \nu \int d^2r \epsilon \cos \theta_1 \left( \sinh^2 \frac{\kappa}{2} - i\eta \cosh^2 \frac{\kappa}{2} \right) - \frac{i\pi}{2} = -2i\pi \hat{\epsilon} \left( \sinh^2 \frac{\kappa}{2} - i\eta \cosh^2 \frac{\kappa}{2} \right) - \frac{i\pi}{2}, \]  

(47)

\[ S_2 = -4i\pi \nu \int d^2r \epsilon \cos \theta_1 \cosh^2 \frac{\kappa}{2} - \frac{i\pi}{2} = -2i\pi \hat{\epsilon} \cosh^2 \frac{\kappa}{2} + \frac{i\pi}{2}. \]  

(48)
Here we have introduced the complex dimensionless energy parameter $\tilde{x} = \int d^2r \epsilon \cos \theta_1 = x + i\gamma$, with the real part $x = E/\omega_0$ [see Eq. (15) in the main text] and the imaginary part $\gamma = Gi\delta(r_0)/2\pi$. The latter appears due to the coupling to the tunneling tip. The terms $\pm i\pi/2$ in Eqs. (17), (18) appear due to the topological term $S_\theta[Q]$, see Eq. (19). With the above expressions for $DQ$ and the action we can calculate the total partition function of the theory,

$$\int DQ e^{-S[Q]} = \int_0^\infty \tanh \frac{\kappa}{2} d\kappa d\eta d\zeta \exp \left[ 2i\pi \tilde{x} \left( \sinh^2 \frac{\kappa}{2} - i\eta \cosh \frac{\kappa}{2} \right) + i\pi \right] = 1. \quad (49)$$

This is exactly what one expects for the partition function of a supersymmetric theory. Note that the topologically nontrivial sector does not contribute to this result since the action $S_2$ contains no Grassmann variables.

In order to calculate the density of states, we need the pre-exponential factor from Eq. (13). In our parameterization, it has the following form in the two sectors:

$$\frac{\nu}{8} \text{Str}(k\Lambda Q) = \nu \cos \theta_1 \times \begin{cases} \cosh^2 \frac{\kappa}{2}, & \theta_f = \theta_1, \\ \sinh^2 \frac{\kappa}{2} - i\eta \cosh \frac{\kappa}{2}, & \theta_f = \theta_2. \end{cases} \quad (50)$$

$$\rho = \nu \cos \theta_1 \Re \int_0^\infty \tanh \frac{\kappa}{2} d\kappa d\eta d\zeta \left\{ \cosh^2 \frac{\kappa}{2} \exp \left[ 2i\pi \tilde{x} \left( \sinh^2 \frac{\kappa}{2} - i\eta \cosh \frac{\kappa}{2} \right) + i\frac{\pi}{2} \right] \\ + \left( \sinh^2 \frac{\kappa}{2} - i\eta \cosh \frac{\kappa}{2} \right) \exp \left[ 2i\pi \tilde{x} \cosh \frac{\kappa}{2} - i\frac{\pi}{2} \right] \right\}$$

$$= \nu \cos \theta_1 \Re \int_0^\infty d\kappa \sinh \frac{\kappa}{2} \left[ -2i\pi \tilde{x} \cosh \frac{\kappa}{2} \exp \left( 2i\pi \tilde{x} \sinh \frac{\kappa}{2} \right) + \exp \left( 2i\pi \tilde{x} \cosh \frac{\kappa}{2} \right) \right]$$

$$= \nu \cos \theta_1 \left( 1 - \Re \frac{1 + e^{2i\pi \tilde{x}}}{2i\pi \tilde{x}} \right) = \nu \cos \theta_1 \left[ 1 + \frac{\gamma}{\pi(x^2 + \gamma^2)} - \frac{\sin(2\pi x)}{2\pi x} \right]. \quad (51)$$

In the last expression we have used the condition $\gamma \ll 1$. The result (51) coincides with Eqs. (15) – (17) of the main text. In the tunneling limit $\gamma \to 0$ the lorentzian term yields a delta function. This implies that the Majorana state is robust in a closed system and disorder does not smear it.

Note, that for an alternative model of the symmetry class D, when the topological term is absent, the density of states does not contain the Majorana delta-peak. Two topological sectors contribute with the same sign yielding $\nu$.

$$\rho_D = \nu \cos \theta_1 \left[ 1 + \frac{\sin(2\pi x)}{2\pi x} \right]. \quad (52)$$

High energies $E_{Th} \ll E \ll \Delta$

In the previous section, we have calculated the density of states at low energies. The result (51) exhibits a delta-peak at zero energy and oscillations which decay at the characteristic energy scale $\omega_0$. These oscillations appear due to repulsion between low-lying levels with energies $E$ and $-E$. The scale $\omega_0$ is the global mean level spacing inside the hole. At higher energies $E \gg \omega_0$, the effect of level repulsion can be neglected and the density of states is given by the mean-field expression

$$\rho(r, E) = \nu \Re \cos \theta(r, E), \quad (53)$$

where $\theta$ is a solution of the Usadel equation (10) with $k = 0$. The result (53) can be obtained from the sigma model identity (13) by integrating over $Q$ around the unique global minimum of the action given by the Usadel equation. Supersymmetry ensures that the integration with respect to small fluctuations around the minimum yields unity and hence Eq. (53) follows.

The Usadel equation is a complicated non-linear equation that cannot be solved analytically for arbitrary $E$. At relatively low energies $\omega_0 \ll E \ll E_{Th}$, the solution can be found perturbatively in $E$. While linear correction in $E$ to $\theta$ is purely imaginary and does not change the density of states, the second order calculation up to terms $\sim E^2$ is
necessary to obtain the observable result. The spatial profile of this small energy correction does not factorize as in the main term, see Eq. (15) in the main text. Second order perturbation theory amounts to solving linear differential equations, obtained by linearizing Eq. (10), and leads to very complicated functions. We will not discuss any further details of this approach here.

As the energy is increased further and exceeds the Thouless energy, \( E_{\text{Th}} \ll E \ll \Delta \), the Usadel equation admits another approximate solution. The angle \( \theta \) is small everywhere except a narrow vicinity of the hole boundary. In this vicinity we can neglect the curvature of the superconductor edge and write an approximate one-dimensional Usadel equation

\[
D \frac{\partial^2 \theta}{\partial r^2} + 2iE \sin \theta = 0. \tag{54}
\]

With the boundary conditions \( \theta(R) = \pi/2 \) and \( \theta(r \ll R) \to 0 \), solution to this equation reads

\[
\theta = 4 \arctan \left[ (\sqrt{2} - 1) \exp \sqrt{-2i \frac{E}{E_{\text{Th}}} \left( \frac{r}{R} - 1 \right)} \right]. \tag{55}
\]

The density of states is given by Eq. (53) with the above result for \( \theta \). We see that \( \theta \) decreases from \( \pi/2 \) down to exponentially small values in a narrow strip of the width \( \sim R \sqrt{E_{\text{Th}}/E} \) near the hole boundary. The density of states takes its normal value \( \nu \) everywhere inside the hole except for this narrow strip where it is depleted down to zero. Spatial integration yields global density of states

\[
N(E \gg E_{\text{Th}}) = \pi \nu R^2 \left( 1 - (2 - \sqrt{2}) \sqrt{\frac{E_{\text{Th}}}{E}} \right). \tag{56}
\]

This is the result (20) of the main text.

We have also solved the Usadel equation numerically for the whole range of energies below \( \Delta \). The spatial profile of the density of states is shown in Fig. 3 and the energy dependence of global DOS is depicted in Fig. 2 in the main text. It perfectly matches both the low-energy limit (up to delta peak and oscillatory term) of Eq. (15) and high-energy asymptotics (56).

**Tunneling current and noise**

Tunneling current and noise can be found using the Landauer formalism instead of the sigma model. Electrons incident from the metallic probe are either reflected normally or experience Andreev reflection. Amplitudes of these processes can be found from the microscopic Hamiltonian of the system and used to calculate the current and higher...
cumulants, such as noise. The current $I(V)$ found this way coincides with the result derived from the sigma model and presented in the main text, Eq. (21). The noise, Eq. (23), was found by means of the Landauer approach since the sigma-model consideration of this problem is rather cumbersome. A detailed derivation of the Landauer formalism for our problem and the calculation of current and noise will be presented elsewhere [16].

[17] K. B. Efetov, *Supersymmetry in Disorder and Chaos* (Cambridge University Press, New York, 1997).
[18] E. J. König, P. M. Ostrovsky, I. V. Protopopov, A. D. Mirlin, arXiv:1201.6288
[19] E. Witten, Commun. Math. Phys. 92, 455 (1984).