Reaching Curzon-Ahlborn limit in linear response and Whitney limit in nonlinear response in edge mode quantum thermoelectrics and refrigeration

Sachiraj Mishra[1,2,*] and Colin Benjamin[1,2,†]

[1]*School of Physical Sciences, National Institute of Science Education and Research, HBNI, Jatni-752050, India*
[2]*Homi Bhabha National Institute, Training School Complex, AnushaktiNagar, Mumbai, 400094, India*

Quantum heat engines and quantum refrigerators are proposed in three-terminal quantum Hall (QH) and quantum spin Hall (QSH) setups with a voltage-temperature probe in linear and nonlinear transport regimes. In the linear response regime, we find that efficiency at maximum power approaches the Curzon-Ahlborn limit in both QH and QSH setups. Similarly, in nonlinear response, we find that efficiency at maximum power reaches the Whitney bounds. For the first time, we see that the thermoelectric efficiency limits in linear and nonlinear transport regimes are achieved using quantum point contacts in the same setup.

Introduction: The quantum Hall resistance [1] in 2DEGs (two-dimensional electron gas) in the presence of a normal uniform magnetic field at low temperature arise as a consequence of chiral edge modes [2], which is famously known as quantum Hall effect. These edge modes are robust against disorder and are immune to backscattering [3] being topologically protected. Complimentary to chiral edge modes, helical edge modes are seen in quantum spin Hall experiments with Mercury Telluride/Cadmium Telluride heterostructures [4–6], which are also topologically protected. Recently in QH [7–11] and QSH [12, 13] setups, thermoelectric properties have been investigated using Landaeur-Buttiker formalism [14, 15]. Quantum transport can be described in two regimes. First is linear, wherein the applied voltage and temperature biases are quite small, and thus leads to a linear relation between current and the temperature or voltage biases applied [16]. The second is nonlinear, wherein the applied voltage and temperature biases are not small. In the linear response regime, the efficiency at maximum power ($\eta|_{P_{max}}$) and the maximum efficiency (η_{max}) are closely related to the figure of merit ($Z\theta$), where θ is the reference temperature of the setup. In this regime, there exists a bound to $\eta|_{P_{max}}$, which is given by Curzon-Ahlborn(CA) limit with the efficiency ($\eta_{P_{max}}$)= $\eta_c/2$ [16]. To achieve these values, one needs to have a substantial Seebeck coefficient (S) and low thermal conductance (K). Similarly, in the non-linear transport regime, there exists an upper bound, the Whitney limit [17, 18], i.e., efficiency at maximum power output ($\eta|_{P^{Wh}_{max}}$) = $\eta|_{P^{Wh}_{max}} = \eta_c/(1 + 0.936(1 + \theta_2/\theta_1))$, where θ_1 and θ_2 are the temperatures of the hot and cold reservoirs, while N being the number of edge modes. This is the nonlinear equivalent of the CA efficiency in linear response. The study of two-terminal setups in the nonlinear response was extended to a three-terminal setup with a voltage probe and it was shown that the Whitney limits are universal [19]. Ref. [19] considered boxcar-type transmission, which can be implemented via a chain of quantum dots [17–19]. However, the electron-electron interaction between quantum dots, which can lead to coulomb blockade effects, is neglected in the calculation in Ref. [17–19]. This assumption is quite extreme, as in real systems, having a chain of quantum dots without the coulomb blockade effect is almost impossible to implement [16]. In
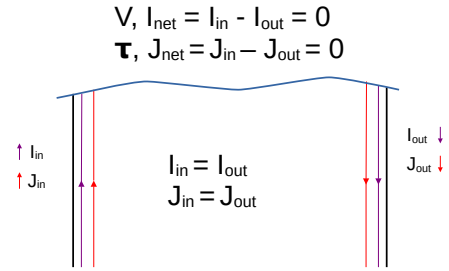

Figure 1: A voltage-temperature probe. V and τ are the applied voltage bias and temperature bias at the reservoir (shown by wavy blue curve), which achieves $I_{net} = 0$ and $J_{net} = 0$ simultaneously. The purple arrows indicate the direction of the charge currents and the red arrows indicate the direction of the heat currents.

contrast, this work considers quantum point contact (QPC) type tunneling, which is easier to implement experimentally [20–23] than the boxcar-type transmission. With QPC-type of tunneling, one also reaches the Whitney bounds in the nonlinear transport regime. Previously, in Refs. [17–19] a QH setup was only considered, however in this letter we consider both QH and QSH setups albeit with the easier to implement QPC type tunneling contacts for both linear as well as non-linear response regimes.

In this letter, our interest lies in looking at the thermoelectric performance of three terminal QH and QSH setups via a voltage-temperature (VT) probe both in linear and nonlinear transport regimes. The VT probe is used to incorporate the inelastic scattering processes that can arise due to electron-electron or electron-phonon scattering, phenomenologically. In any multi-terminal setup with transport via chiral or helical edge modes, the VT probe is that terminal, where the net charge current, as well as heat current, vanishes simultaneously. It is achieved by adjusting the voltage bias as well as temperature bias in that terminal, see Fig. 1. VT probe condition implies no net charge current ($I_{net} = I_{out} - I_{in}$) and no net heat current ($J_{net} = J_{out} - J_{in}$) through the terminal, i.e., $I_{net} = J_{net} = 0$. Similarly, a voltage probe, which is also used

to incorporate inelastic scattering processes in mesoscopic devices, is that terminal where only the net charge current (I_{net}) is zero and the net heat current (J_{net}) is nonzero, which can be achieved only at a fixed voltage bias. But, in principle, if the reservoir is in contact with a scatterer in the steady state, both the net charge current as well as heat current via the same reservoir vanishes [16] and in this situation, a VT probe is better and more accurate than voltage probe. In [24], a VT probe is used to compare the scaling of Lorentz ratio with that of the Luttinger-liquid model [25] via both phase and momentum relaxation. Using voltage probes, thermoelectric coefficients have been calculated in QH [7] and QSH [12] setups to analyze the impact of inelastic scattering. In this letter, we propose three-terminal QH and QSH setups with the third terminal acting as a VT probe that can work as a highly efficient quantum heat engine (QHE) or as a quantum refrigerator (QR) both in the linear and nonlinear response regimes. Further, the efficiency at maximum power approaches the Curzon-Ahlborn limit in linear response and the Whitney limit in nonlinear transport regimes.

This letter is organized as follows. We divide the work into two parts. In the first part of the work, we investigate the thermoelectric transport of the QH and QSH setups within the linear response regime and show that efficiency at maximum power approaches Curzon-Ahlborn in both QH and QSH setups. In the second part, we focus on non-linear transport, we see that the efficiency at maximum power approaches the Whitney limit. The technical details of the calculation are presented in the supplementary material (SM).

Thermoelectricity in QH and QSH setups: We start by discussing the general theory for thermoelectric transport in QH as well as QSH setups within the Landaeur-Buttiker scattering theory in both linear and non-linear regimes [14, 16]. The charge and heat currents in a multiterminal QH sample, see Sec. II of SM, are given as,

$$I_\alpha = \frac{2e}{h} \int_{-\infty}^{\infty} dE \sum_\beta T_{\alpha\beta}(f_\alpha(E) - f_\beta(E)),$$
$$J_\alpha = \frac{2}{h} \int_{-\infty}^{\infty} dE \sum_\beta (E - \mu_\alpha)T_{\alpha\beta}(f_\alpha(E) - f_\beta(E)), \quad (1)$$

where the transmission probability for an electron to scatter from terminal β to terminal α is $T_{\alpha\beta}$. Here, $f_{\alpha(\beta)} = \left(1 + e^{(E-\mu_{\alpha(\beta)})/k_B \theta_{\alpha(\beta)}}\right)^{-1}$, $\mu_{\alpha(\beta)}$ and $\theta_{\alpha(\beta)}$ are the Fermi function, chemical potential and temperature of terminal $\alpha(\beta)$ respectively.

Similarly, for QSH setup, with helical edge modes, the currents carry an extra spin degree of freedom and are written as $I_\alpha = \sum_{\sigma \in \{\uparrow/\downarrow\}} I_\alpha^\sigma$ and $J_\alpha = \sum_{\sigma \in \{\uparrow/\downarrow\}} J_\alpha^\sigma$, where [12],

$$I_\alpha^\sigma = \frac{e}{h} \int_{-\infty}^{\infty} dE \sum_\beta T_{\alpha\beta}^\sigma(f_\alpha(E) - f_\beta(E))],$$
$$J_\alpha^\sigma = \frac{1}{h} \int_{-\infty}^{\infty} dE(E - \mu_\alpha) \sum_\beta T_{\alpha\beta}^\sigma(f_\alpha(E) - f_\beta(E))]. \quad (2)$$
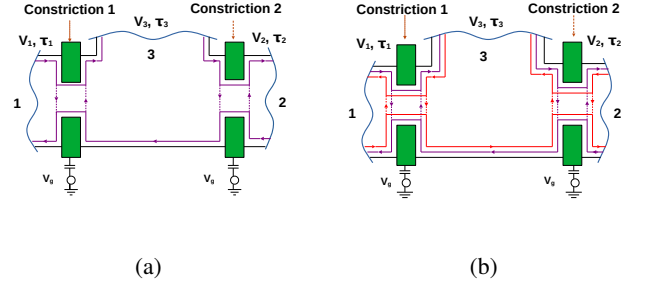

Figure 2: Three terminal (a) QH with chiral edge modes, (b) QSH sample with helical edge modes, and two QPC constrictions connected capacitively to an external gate voltage (V_g). For a QH sample, the edge mode of an electron, which can scatter from constriction, is shown by the purple solid (dashed) line. similarly, for the QSH sample, the edge mode for a spin-up electron is shown by the purple solid (dashed) line, which can scatter from the QPC-type constriction, whereas the edge mode for a spin-down electron is shown by the red solid (dashed) line. We assume QPCs' have a single fixed energy level.

with the transmission probability for an electron to transmit from terminal β with spin ρ to terminal α with spin σ being $T_{\alpha\beta}^\sigma = \sum_{\rho \in \{\uparrow/\downarrow\}} T_{\alpha\beta}^{\sigma\rho}$. In this three-terminal setup (both QH and QSH) as shown in Fig. 2, $\theta_1 = \theta + \tau_1, \theta_2 = \theta + \tau_2$ and $\theta_3 = \theta + \tau_3$, where θ is the reference temperature and τ_1, τ_2 and τ_3 are the temperature biases applied in terminals 1, 2 and 3 respectively. For our setups, we apply temperature bias only at terminal 1, i.e., $\tau_1 = \tau$ and $\tau_2 = 0$. We study the thermoelectric properties for a QHE considering terminal 3 to be a VT probe, where the net charge current (I_3), as well as heat current (J_3) are zero. Additionally, we assume both terminals 1 and 2 to be current probes with $V_1 = -V$ and $V_2 = 0$. Here, we have considered only one type of energy-dependent transmission via constrictions, which is known as the quantum point contact (QPC) type [26], defined mathematically via $T_{QPC} = [1 + \exp(-2\pi(E - E_l)/\hbar\omega)]^{-1}$, where l denotes the constriction index, i.e., $l \in \{1, 2\}$. Below a certain threshold energy level E_l, T_{QPC} vanishes, while above E_l, it is transparent, while $\hbar\omega$ indicates the width of the transmission resonance. An explanation of T_{QPC} is found in sec. I of SM. Following Refs. [27, 28], where a capacitive external gate voltage V_g is connected to the constrictions and, in turn, controls the threshold energy E_l of QPC, where we can explore both linear and nonlinear transport regimes for our setups in Fig. 2(a) and (b).

Linear transport in QH and QSH setups: In the linear response regime for the QH case, i.e., when $\mu_\alpha(= eV_\alpha) \ll k_B\theta$ and $(\theta_\alpha - \theta) \ll \theta$, one can see that the current (both charge and heat) varies linearly with the voltage and temperature biases applied across the sample [16]. Considering terminal 3 to be a VT probe i.e., $I_3 = J_3 = 0$, the charge currents and heat currents in terminal 1 are modified with different Onsager matrix

elements and are written as,

$$\begin{pmatrix} I_1 \\ J_1 \end{pmatrix} = \begin{pmatrix} L_{eV} & L_{e\theta} \\ L_{hV} & L_{h\theta} \end{pmatrix} \begin{pmatrix} -V \\ \tau \end{pmatrix} \qquad (3)$$

where the Onsager matrix elements are $L_{eV}, L_{e\theta}, L_{hV}, L_{h\theta}$, See Eq. (37) in SM. These Onsager matrix elements are used to find the transport coefficients such as conductance (G), Seebeck coefficient (S), Peltier coefficient (Π), and the thermal conductance (K) and are given as,

$$G = L_{eV}, \quad S = \frac{L_{e\theta}}{L_{eV}}, \quad \Pi = \frac{L_{hV}}{L_{eV}}, \quad K = L_{h\theta} - \frac{L_{hV} L_{e\theta}}{L_{eV}}. \quad (4)$$

Using a similar procedure, one can also derive the Onsager matrix elements in the QSH case for charge current I_α^σ and heat current J_α^σ by imposing the VT probe condition ($I_3 = \sum_{\sigma \in \{\uparrow,\downarrow\}} I_3^\sigma = 0$ and $J_3 = \sum_{\sigma \in \{\uparrow,\downarrow\}} J_3^\sigma = 0$) and they can be written in the same form as Eq. (3) (see derivation in Sec. III A 1 in SM) and is written as,

$$\begin{pmatrix} I_1^\sigma \\ J_1^\sigma \end{pmatrix} = \begin{pmatrix} L_{eV}^\sigma & L_{e\theta}^\sigma \\ L_{hV}^\sigma & L_{h\theta}^\sigma \end{pmatrix} \begin{pmatrix} -V \\ \tau \end{pmatrix} \qquad (5)$$

where the spin-components of Onsager matrix elements L_{eV}^σ, $L_{e\theta}^\sigma$, L_{hV}^σ and $L_{h\theta}^\sigma$ for VT probe are derived in Eq. (70) in Sec. III A 1 in SM. Using the above spin-components of the Onsager coefficients, the spin-component of transport coefficients such as Conductance (G^σ), Seebeck Coefficient (S^σ), Peltier Coefficient (Π^σ), and thermal conductance (K^σ) are defined as

$$G^\sigma = L_{eV}^\sigma, \quad S^\sigma = \frac{L_{e\theta}^\sigma}{L_{eV}^\sigma}, \quad \Pi^\sigma = \frac{L_{hV}^\sigma}{L_{eV}^\sigma}, \quad K^\sigma = L_{h\theta}^\sigma - \frac{L_{hV}^\sigma L_{e\theta}^\sigma}{L_{eV}^\sigma}. \qquad (6)$$

The output power in terminal 1, $P = I_1 V$ is maximum at the voltage bias $V = L_{e\theta}\tau/4L_{eV}$ and is given as $P_{max} = L_{e\theta}^2 \tau^2/4L_{eV}$ and this derivation is found in Sec. II A 2 for QH and III A 2 for QSH in SM. At the maximum power, the efficiency, i.e., $\eta|_{P_{max}} = P_{max}/J = \frac{\eta_c}{2} \frac{Z\theta}{Z\theta+2}$ (for derivation see Eq. (33) of Sec. II A 2 for QH and Eq. (73) of Sec. III A 2 for QSH in SM). J is the heat current from the hotter terminal. Other quantities of interest are the maximum efficiency, i.e., $\eta_{max} = \eta_c \frac{\sqrt{Z\theta+1}-1}{\sqrt{Z\theta+1}+1}$ (see the derivation in Eq. (37) of Sec. II A 2 for QH and Eq. (79) of Sec. III A 2 for QSH in SM).

In Fig. 3(a), we can show a parametric plot of the total efficiency (η/η_c) and the power (P/P_{max}). This has been calculated using Eqs. (3-6). For this purpose, we have considered threshold energies of QPC 1 and QPC 2 to be $E_1' = E_1 + eV_g$ and $E_2' = E_1 + eV_g$ respectively. One can see that both $\eta|_{P_{max}}$ and η_{max} in both QH and QSH setups approach the Curzon-Ahlborn efficiency at $eV_g = 84.8 k_B \theta$ as shown in Fig. 3(a).

In Ref. [7], it has been reported that a QH setup cannot work as a QR due to time-reversal symmetry breaking. In contrast, the exact same setup with QSH edge modes can act as both QHEs and QR as time-reversal symmetry is preserved as shown in [12]. This is because in the setup used in [7], the

asymmetric parameter (AP), i.e., the Seebeck to Peltier coefficient ratio, is zero or infinity [7]. AP is closely related to a setup's refrigeration properties, which is why it reduces the capability of certain QH setups to work as a QR [12, 29]. On the other hand, in the QSH setup, AP is always finite since time-reversal symmetry is preserved [12]. However, this observation is not valid for all QH setups as in our QH case, and the QSH setups with a VT probe can act both as a QHE and a QR because the asymmetric parameter in our setups is finite, see Sec. II A 4 of SM.

Similarly, for QR, we calculate the cooling power at the maximum coefficient of performance ($J|_{\eta_{max}^r}$) and the maximum coefficient of performance (η_{max}^r) for both QH and QSH setups. The coefficient of performance (COP = J/P) is generally defined as the ratio of heat taken from the cooler terminal (J) to the power absorbed (P) by the setup. In our setup, the heat current from the cooler terminal is $-J_2$. The maximum COP (η_{max}^r) is achieved using the condition $\frac{d(COP)}{dV} = 0$ and is given as $\eta_{max}^r = \eta_c^r \frac{\sqrt{Z\theta+1}-1}{\sqrt{Z\theta+1}+1}$, where η_c^r is the Carnot coefficient of performance. As shown in Fig. 3(b), η_{max}^r approaches the Carnot limit, at $eV_g = 84.80 k_B \theta$. Similar to the QHE, we have also shown the parametric plot of cooling power J and coefficient of performance η^r in Fig. 3(c) for QH and in Fig. 3(d) for QSH.

In any nano/meso scale device, one of the ambitions is the search for an excellent thermoelectric (TE) material with high efficiency. QH and QSH states of matter are helpful in this context as they have insulating bulk and conducting edge modes. As shown in Sec. II A 2 for QH and Sec. III A 2 for QSH setup in SM, both the maximum efficiency (η_{max}) and efficiency at maximum power ($\eta|_{P_{max}}$) are dependent on the figure of merit ($Z\theta$) = $GS^2\theta/K$ for a QHE. Similarly, for a QR, the maximum coefficient of performance depends on this figure of merit ($Z\theta$). For achieving Curzon-Ahlborn efficiency ($\eta_c/2$) $Z\theta$ needs to be much larger than 1. The key to achieving this is to have the Seebeck coefficient (S) much larger and thermal conductance (K) much smaller both occurring for same configuration. We find that in our QH and QSH setups with a QPC-like constriction, the Seebeck coefficient is very large, while thermal conductivity is very small at $eV_g = 84.8 k_B \theta$.

Non-linear transport in QH and QSH setups: There has been ongoing work in this direction on nonlinear thermoelectric transport [17, 18, 30–32] in both QH and QSH setups. In a two-terminal QH setup, it has been verified that there is a significant departure from the Onsager-Casimir relation in the nonlinear transport regime [30]. A similar investigation has also been performed in a QSH setup with helical edge modes [31]. In a three-terminal setup with chiral edge modes [32] with a voltage probe, it has been verified that it can also work as an efficient QR in the weakly non-linear regime. Further, a normal metal Aharanov-Bohm heat engine has been investigated in the nonlinear regime and it provides better thermoelectric performance [27] compared to the linear regime [28].
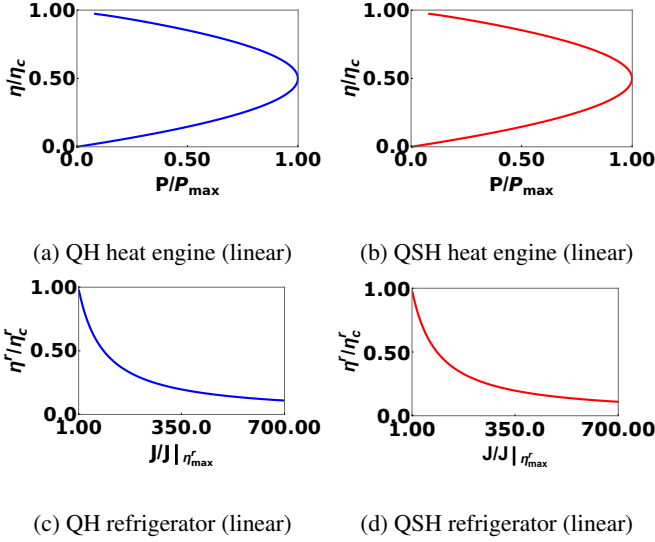
(a) QH heat engine (linear)

(b) QSH heat engine (linear)

(c) QH refrigerator (linear)

(d) QSH refrigerator (linear)

Figure 3: Parametric plot of Efficiency (η) in units of Carnot efficiency (η_c) vs power in units of maximum power (P_{max}) (a) QH, (b) QSH QHE. Parametric plot of coefficient of performance (η^r) in units of η_c^r vs cooling power (J) in units of $J|_{\eta_{max}^r}$ for a (c) QH and (d) QSH QR in the linear response regime. Parameters taken are $eV_g = 84.8 k_B \theta$, $\omega = 0.1 k_B \theta / \hbar$, $\theta_1 = 1.01 K, \theta_2 = 1.0 K, \mu = 0, E_1 = k_B \theta$.

In the linear response regime, the electron transport is only dependent on the electron's kinetic energy, whereas, in the nonlinear regime, it is dependent on both its kinetic energy as well as the interaction potential (U') of the sample where the electron flows. It implies that in the linear response regime, the energy value E_l is shifted by the gate voltage, i.e. $E_l \longrightarrow E_l + eV_g$, whereas it is in the nonlinear response regime, the interaction potential U' shifts the energy values E_l, i.e. $E_l \longrightarrow E_l + eU'$, $U' = U + u_g V_g$ being the interaction potential, see Eq. (56) of SM. u_g is the characteristic potential, which determines the response of U' to the application of gate voltage V_g and the form of U is given below Eq. (54) of SM. The advantage of having a gate voltage is that it helps in nullifying the interaction potential (U) in the nonlinear response regime, the argument for this is given in Sec. II B 2 of SM, see also Ref. [27]. As discussed in Sec. II B 2 of SM, the gate voltage needs to be of the form $V_g = (-U + V_g')/u_g$, where the first term inside the parenthesis completely nullifies the interaction potential U and the second term lifts the energy level E_l, i.e., $E_l \longrightarrow E_l + eV_g'$ for the nonlinear response regime.

For non-linear transport, we start from Eq. (1) for QH and (2) for QSH, respectively. Unlike linear response, this is a hard problem to solve analytically but can be done numerically. The Mathematica code can be found in [33] for both QH and QSH setups. In [33], for the case of QHE, we first numerically solve for V_3 and τ_3 at a particular V and τ by imposing the VT probe condition, i.e., $I_3 = J_3 = 0$ and then use it to find the current I_1 and J_1 (using Eq. 7) and finally, evaluate the power $P = V I_1$ and efficiency $\eta = P/J_1$. Similarly, for the
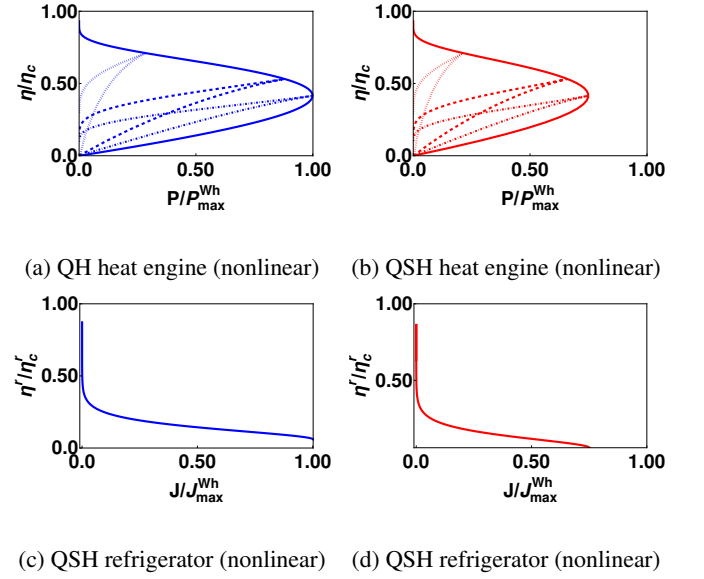

(a) QH heat engine (nonlinear)

(b) QSH heat engine (nonlinear)

(c) QSH refrigerator (nonlinear)

(d) QSH refrigerator (nonlinear)

Figure 4: Parametric plot for efficiency (η) in units of Carnot efficiency (η_c) vs power (P) in units of maximum power (P_{max}^{Wh}) for (a) QH heat engine and (b) QSH heat engine. Parametric plot for the coefficient of performance (η^r) in units of η_c^r vs cooling power (J) in units of J_{max}^{Wh} for (c) QH refrigerator and (d) QSH heat engine in nonlinear response regime. For QH (QSH) heat engine, the blue (red) dot-dashed curve is for $V = 0.14 k_B \theta / e$, the blue (red) dashed curve for $V = 0.77 k_B \theta / e$ and the blue (red) dotted curve for $V = 2.7 k_B \theta / e$. The blue (red) line shows the bound to the power and efficiency. For QH (QSH) refrigerator, blue (red) line is for $V = 20 k_B \theta / e$. The parameters taken are $\omega = \frac{0.1 k_B \theta}{h}$, $\theta_1 = 2K, \theta_2 = 1K, \mu = 0, E_1 = k_B \theta$.

case of QR too, we evaluate V_3 and τ_3 from VT probe condition at a particular V and τ and then use to find the cooling power $J = -J_2$ and the coefficient of performance $\eta^r = J/P$, where P is the power absorbed by the setup. More details regarding the technical details of the calculation can be found in SM.

As shown in Refs. [17, 18] valid for QH setups, there exists a bound to maximum possible power ($P_{max} = P_{max}^{Wh}$), which is given as $0.0642 \, \pi^2 N k_B^2 (\theta_1 - \theta_2)^2 / h$. Similarly, the bound to the efficiency at the maximum power ($\eta|_{P_{max}}$) can be found out at P_{max}^{Wh}, i.e., $\eta|_{P_{max}^{Wh}} = \eta_c / (1 + 0.936(1 + \theta_2/\theta_1))$, where θ_1 and θ_2 are the temperatures of the hot and cold reservoirs, while N is the number of edge modes. In our work, since we consider $\theta_1 = 2K$, $\theta_1 = 1K$ and $N = 1$, therefore $P_{max}^{Wh} = 0.632 k_B^2/h$ and $\eta|_{P_{max}^{Wh}} = 0.41 \eta_c$. At finite power output (P), much smaller than P_{max}^{Wh}, the bound to the maximum efficiency (η^{max}) is given as $(\eta|_{max}^{Wh}) = \eta_c \left(1 - 0.478 \sqrt{\theta_2 P / \theta_1 P_{max}^{Wh}}\right)$. These quantities were derived in a two-terminal QH setup, which has a constriction and boxcar-type transmission (see, Sec. II B 1 of SM and also Refs. [17, 18]). We extend the same approach to a two-terminal QSH setup and prove that

the bounds to P_{max}, $\eta|_{P_{max}}$ and η_{max} are exactly similar to the QH case (see, Sec. II B 1 of SM).

For a QR, as in Refs. [17, 18], the bound to the maximum cooling power (J_{max}) within this non-linear regime is given as J_{max}^{Wh} is $\pi^2 N k_B^2 \theta_1^2/6h$ valid for QH setup. The coefficient of performance (η^r) vanishes at this maximum cooling power. Similarly, at any arbitrary cooling power (J), smaller than J_{max}^{Wh}, the maximum coefficient of performance (η_{max}^r) approaches $(\theta_1/\theta_2 - 1)^{-1} \left(1 - 1.09 \sqrt{\theta_1 J/(\theta_1 - \theta_2)J_{max}^{Wh}}\right)$. For the refrigerator too, we extend the same approach of Refs. [17, 18] and find that the bounds to J_{max} and η^r are same as the QH case (see, Sec. III B 1 of SM).

Table I: Thermoelectric performance of the QH and QSH setup in the linear and nonlinear response regime. P is in units of $P_{max} = L_{e\theta}^2 \tau^2/4L_{eV}$ for linear, and $P_{max}^{Wh} = 0.0642\pi^2 k_B^2 \tau^2/h$ for nonlinear, $\eta|_{P_{max}}$, J is in units of $J|_{\eta_{max}^r}$ for linear and $J_{max}^{Wh} = \pi^2 k_B^2 \theta_1^2/6h$ for nonlinear.

Regime	Setup	QHE			QR	
		P	$\frac{\eta|_{P_{max}}}{\eta_c}$	$\frac{\eta_{max}}{\eta_c}$	J	$\frac{\eta_{max}^r}{\eta_c^r}$
Linear	QH	1 [Fig. 3(a)]	0.50 [Fig. 3(a)]	0.97 [Fig. 3(a)]	1 [Fig. 3(c)]	0.97 [Fig. 3(c)]
	QSH	1 [Fig. 3(b)]	0.50 [Fig. 3(b)]	0.97 [Fig. 3(b)]	1 [Fig. 3(d)]	0.97 [Fig. 3(d)]
Nonlinear	QH	1 [Fig. 4(a)]	0.41 [Fig. 4(a)]	0.93 [Fig. 4(a)]	1 [Fig. 4(c)]	0.86 [Fig. 4(c)]
	QSH	0.76 [Fig. 4(b)]	0.41 [Fig. 4(b)]	0.93 [Fig. 4(b)]	0.75 [Fig. 4(d)]	0.86 [Fig. 4(d)]

Here, we consider the gate voltage in such a way that it nullifies the interaction potential, this in turn shifts the energy level of the QPC-like constriction. This is explained in sec. II B 2 of SM with the help of a two-terminal nonlinear QH setup and a QPC-like constriction between the two terminals, which is capacitively connected to an external gate voltage V_g, and is extended to a three-terminal QH setup (see, Sec. II B 3 of SM). A similar argument also has been made for a two-terminal QSH setup in Sec. III B 2 of SM. We examine the maximum power, efficiency at maximum power, and maximum efficiency at finite power numerically. As shown in Fig. 4(a), for the QH setup, the maximum power approaches the Whitney limit at a particular voltage bias $V = 0.14k_B\theta/e$ (See the blue curve of Fig. 4(a)). Similarly, as one increases the voltage bias V the power reduces and efficiency increases (see black and red curves in Fig. 4(a)). The QHE is more efficient as $\eta|_{P_{max}}$ and η_{max} approaches $0.41\eta_c$ and $0.93\eta_c$ respectively. Similarly, in 4(c), one gets the same maximum cooling power, which approaches the Whitney limit, whereas the maximum coefficient of performance is around 86% of η_c. The method to get the plot in Fig. 4(a) and (c) are discussed in Sec. II B 4 of SM.

In the case of QSH, the maximum possible power achieved is less than the Whitney limit, but the efficiency at this power is $0.41\eta_c$ (see, Fig. 4(b)). Similarly, the maximum efficiency also approaches $0.93\eta_c$ similar to the chiral case (see, Fig. 4(b)). In Fig. 4(d), we observe that the maximum cooling power is less than the Whitney limit, whereas the coefficient of performance is around 86% of η_c. The method to obtain the plots in Fig. 4(b), (d) are discussed in sec. III B 4 of SM.

Experimental Realization and Conclusion: We consider three-terminal QH and QSH setups with QPC-like constriction and observe that the best possible thermoelectric performance is achieved as it approaches the benchmarks set for efficiency as a QHE and coefficient of performance as a QR both in the linear (see, Fig. 3) and non-linear (see, Fig. 4) response regimes. Additionally, for the nonlinear response, the maximum value of power generated as a QHE approaches the Whitney bound, i.e. P_{max}^{Wh} (see Fig. 4(a), Table I). For a QR too, the maximum cooling power approaches J_{max}^{Wh} (see Fig. 4(c), Table I) for QH setup. In case of QSH, the maximum power as a QHE is reduced from P_{max}^{Wh}, but the efficiency at maximum power is still $0.41\eta_c$ (the equivalent of Curzon-Ahlborn efficiency for nonlinear response regime) (see, Fig. 4(b), Table I). Similarly, as a QR, the maximum possible cooling power generated is below J_{max}^{Wh} (see Fig. 4(d), Table I). This suppression of power (both P_{max}^{Wh} and J_{max}^{Wh}) arises as a result of helical edge mode transport with a VT probe as there is a finite spin current in three-terminal QSH setup, which is absent in chiral case. This is another way to distinguish between chiral and helical edge mode transport. However, experimentally, a VT probe is slightly more difficult to design as the probe reservoir can interact with the surroundings, which may initiate heat flow between the reservoir and surroundings and ultimately change its steady state temperature [16]. We have considered QPC-type tunneling, which has been used in Refs. [20–23] to study electron transport in 2DEGs, and is easier to implement experimentally as it can be controlled electrostatically with the help of split gate [34–36].

To conclude, this letter proposes a setup that can achieve the Curzon-Ahlborn (CA) limit in the linear response regime and the corresponding Whitney bound in the nonlinear response regime. This work has been done in three-terminal QH and QSH setups and can be easily generalized to multi-terminal QH and QSH setups with multiple VT probes. Not only that, but one can also perform this study on other edge mode systems such as trivial helical [37–39] and anomalous edge modes [40–42].

---

[1] K. v. Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. **45**, 494 (1980).
[2] B. I. Halperin, Phys. Rev. B **25**, 2185 (1982).
[3] M. Büttiker, Phys. Rev. B **38**, 9375 (1988).
[4] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, Science **314**, 1757 (2006), https://www.science.org/doi/pdf/10.1126/science.1133734.
[5] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. Molenkamp, X. Qi, and S. Zhang, Science **318**, 776770

(2007).

[6] A. Roth, C. Brüne, H. Buhmann, L. W. Molenkamp, J. Maciejko, X.-L. Qi, and S.-C. Zhang, Science **325**, 294 (2009).

[7] R. Sánchez, B. Sothmann, and A. N. Jordan, Phys. Rev. Lett. **114**, 146801 (2015).

[8] D. Sánchez and L. m. c. Serra, Phys. Rev. B **84**, 201307 (2011).

[9] K. Brandner and U. Seifert, New Journal of Physics **15**, 105003 (2013).

[10] R. Sánchez, B. Sothmann, and A. N. Jordan, New Journal of Physics **17**, 075006 (2015).

[11] Z. Sartipi and J. Vahedi, Phys. Rev. B **108**, 195435 (2023).

[12] A. Mani and C. Benjamin, Phys. Rev. E **97**, 022114 (2018).

[13] D. Gresta, M. Real, and L. Arrachea, Phys. Rev. Lett. **123**, 186801 (2019).

[14] S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge Studies in Semiconductor Physics and Microelectronic Engineering (Cambridge University Press, 1995).

[15] S.-Q. Shen, *Topological Insulators* (Springer Singapore, 2017).

[16] G. Benenti, G. Casati, K. Saito, and R. Whitney, Physics Reports **694**, 1 (2017), fundamental aspects of steady-state conversion of heat to work at the nanoscale.

[17] R. S. Whitney, Phys. Rev. Lett. **112**, 130601 (2014).

[18] R. S. Whitney, Phys. Rev. B **91**, 115425 (2015).

[19] R. Whitney, Entropy **18**, 208 (2016).

[20] B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhoven, D. van der Marel, and C. T. Foxon, Phys. Rev. Lett. **60**, 848 (1988).

[21] D. Wharam, T. J. Thornton, R. Newbury, M. Pepper, H. Ahmed, J. Frost, D. Hasko, D. Peacock, D. Ritchie, and G. Jones, Journal of Physics C: solid state physics **21**, L209 (1988).

[22] H. Van Houten, L. Molenkamp, C. Beenakker, and C. Foxon, Semiconductor Science and Technology **7**, B215 (1992).

[23] L. W. Molenkamp, T. Gravier, H. van Houten, O. J. A. Buijk, M. A. A. Mabesoone, and C. T. Foxon, Phys. Rev. Lett. **68**, 3765 (1992).

[24] R. Das and C. Benjamin, Majorana fermion induced power-law scaling in the violation of wiedemann-franz law (2023), arXiv:2309.05492 [cond-mat.mes-hall].

[25] F. Buccheri, A. Nava, R. Egger, P. Sodano, and D. Giuliano, Phys. Rev. B **105**, L081403 (2022).

[26] M. Büttiker, Phys. Rev. B **41**, 7906 (1990).

[27] G. Haack and F. Giazotto, AVS Quantum Science **3**, 10.1116/5.0064936 (2021).

[28] G. Haack and F. Giazotto, Phys. Rev. B **100**, 235442 (2019).

[29] K. Brandner, K. Saito, and U. Seifert, Phys. Rev. Lett. **110**, 070603 (2013).

[30] D. Sánchez and M. Büttiker, Phys. Rev. Lett. **93**, 106802 (2004).

[31] S.-Y. Hwang, R. López, M. Lee, and D. Sánchez, Phys. Rev. B **90**, 115301 (2014).

[32] D. Sánchez, R. Sánchez, R. López, and B. Sothmann, Phys. Rev. B **99**, 245304 (2019).

[33] S. Mishra and C. Benjamin, Mathematica notebook for the numerical calculation can be found in https://github.com/Sachiraj/thermoelectricity.git (2024).

[34] H. Z. Zheng, H. P. Wei, D. C. Tsui, and G. Weimann, Phys. Rev. B **34**, 5635 (1986).

[35] T. J. Thornton, M. Pepper, H. Ahmed, D. Andrews, and G. J. Davies, Phys. Rev. Lett. **56**, 1198 (1986).

[36] C. Beenakker and H. van Houten, in *Solid state physics*, Vol. 44 (Elsevier, 1991) pp. 1–228.

[37] F. Nichele, H. J. Suominen, M. Kjaergaard, C. M. Marcus, E. Sajadi, J. A. Folk, F. Qu, A. J. A. Beukman, F. K. de Vries, J. van Veen, S. Nadj-Perge, L. P. Kouwenhoven, B.-M. Nguyen, A. A. Kiselev, W. Yi, M. Sokolich, M. J. Manfra, E. M. Spanton, and K. A. Moler, New Journal of Physics **18**, 083005 (2016).

[38] A. Mani and C. Benjamin, Scientific Reports **7**, 10.1038/s41598-017-06820-w (2017).

[39] S. Mishra and C. Benjamin, Phys. Rev. B **108**, 115301 (2023).

[40] C.-Z. Chang, W. Zhao, D. Y. Kim, H. Zhang, B. A. Assaf, D. Heiman, S.-C. Zhang, C. Liu, M. H. Chan, and J. S. Moodera, Nature materials **14**, 473 (2015).

[41] X. Kou, S.-T. Guo, Y. Fan, L. Pan, M. Lang, Y. Jiang, Q. Shao, T. Nie, K. Murata, J. Tang, Y. Wang, L. He, T.-K. Lee, W.-L. Lee, and K. L. Wang, Phys. Rev. Lett. **113**, 137201 (2014).

[42] J. G. Checkelsky, R. Yoshimi, A. Tsukazaki, K. S. Takahashi, Y. Kozuka, J. Falson, M. Kawasaki, and Y. Tokura, Nature Physics **10**, 731 (2014).

[43] R. A. Jalabert, A. D. Stone, and Y. Alhassid, Phys. Rev. Lett. **68**, 3468 (1992).

[44] M. Buttiker, Journal of Physics: Condensed Matter **5**, 9361 (1993).

[45] T. Christen and M. Büttiker, Europhysics letters **35**, 523 (1996).

[46] D. Sánchez and R. López, Phys. Rev. Lett. **110**, 026804 (2013).

[47] R. López and D. Sánchez, Phys. Rev. B **88**, 045129 (2013).

[48] J. Meair and P. Jacquod, Journal of Physics: Condensed Matter **25**, 082201 (2013).

[49] B. Z. Rameshti and A. G. Moghaddam, Phys. Rev. B **91**, 155407 (2015).

SUPPLEMENTARY MATERIAL FOR "REACHING CURZON-AHLBORN EFFICIENCY IN LINEAR RESPONSE AND WHITNEY LIMIT IN NONLINEAR RESPONSE IN EDGE MODE QUANTUM THERMOELECTRICS AND REFRIGERATION"

Sachiraj Mishra[1,2], Colin Benjamin[1,2]

[1]School of Physical Sciences, National Institute of Science Education and Research, HBNI, Jatni-752050, India
[2]Homi Bhabha National Institute, Training School Complex, AnushaktiNagar, Mumbai, 400094, India

In this supplementary material, we discuss various aspects of linear and nonlinear thermoelectricity for both quantum Hall (QH) and quantum spin Hall (QSH) systems with voltage-temperature probe. We also discuss the technical details, and methods, and summarize the calculations. In Sec. I, we discuss the QPC type and Boxcar type of transmission function and discuss how implementing QPC-type transmission is easier than implementing boxcar-type transmission experimentally. In Sec. II, we derive the formulae for charge and heat current for the transport via chiral edge mode in a generic multiterminal QH setup and discuss the linear transport regime (See Sec. II A), and derive charge and heat current in terms of Onsager coefficients. In Sec. II A 1, we discuss our setup and the transmission probabilities and also derive the Onsager coefficients with a voltage-temperature probe. In Sec. II A 2, we derive the formula for maximum power (P_{max}), efficiency at maximum power ($\eta|_{P_{max}}$) and maximum efficiency (η_{max}) for QH setup as a quantum heat engine. Further in Sec. II A 3, we derive the formula for the maximum coefficient of performance (η^r_{max}) and maximum cooling power $J|_{\eta^r_{max}}$ and in Sec. II A 4, we discuss why our QH setup can work as a quantum refrigerator even if time-reversal symmetry is broken, whereas a QH setup discussed in [7] cannot work as a refrigerator. In Sec. II B, we discuss the transport in the nonlinear transport regime for a QH setup, whereas in Sec. II B 1, we discuss the QH setup as a nonlinear quantum heat engine and quantum refrigerator using Whitney's approach. Then in Sec. II B 2, we discuss the 2T QH setup both as a quantum heat engine and refrigerator using our approach, where we extend the same study to a 3T QH setup in Sec. II B 3. Finally, in Sec. II B 4, we study the nonlinear thermoelectricity both as a quantum heat engine and refrigerator for the 3T QH setup with a voltage-temperature probe. In Sec. III, we discuss the charge and heat currents for Helical edge mode transport in a generic multiterminal QSH setup in the linear response regime, and Sec. III A discusses the transport in the linear response regime in a generic multiterminal QSH setup. Next, we discuss our QSH setup and its transmission probabilities and in Sec. III A 1 the Onsager coefficients are derived with a voltage-temperature probe. In Sec. III A 2, we derive P_{max}, $\eta|_{P_{max}}$ and η_{max} for the QSH setup as a quantum heat engine. In Sec. III A 3, we further discuss the cooling power and coefficient of performance in our 3T QSH setup as a quantum refrigerator. In Sec. III B, we discuss the nonlinear transport regime in QSH setup, whereas in Sec. III B 1, we discuss the QSH setup as a nonlinear quantum heat engine and quantum refrigerator using Whitney's approach. Then in Sec. III B 2, we discuss the 2T QSH setup both as a quantum heat engine and refrigerator using our approach, where we extend the same study to a 3T QSH setup in Sec. III B 3. Finally, in Sec. III B 4, we study the nonlinear thermoelectricity both as a quantum heat engine and refrigerator for the 3T QSH setup with a voltage-temperature probe.

I. QUANTUM POINT CONTACT VS BOXCAR-TYPE TRANSMISIION

In this section, we discuss the two types of transmission, which we have extensively considered in this work. The first is the quantum point contact (QPC)-type tunneling [26] and the other one is the boxcar-type [17, 18].

A quantum point contact can be modeled by a potential

$$V(x,y,z) = V_0 - \frac{1}{2}m\omega_x^2 x^2 + \frac{1}{2}m\omega_y^2 y^2 + \frac{1}{2}m\omega_z^2 z^2, \qquad (7)$$

where m is the effective mass of the electron and the electron feels the potential $V(x,y,z)$ when the QPC is present in a setup. As derived in [26], the transmission probability (T_{QPC}) for this constriction is given as

$$T_{QPC} = \sum_{n_y,n_z} \frac{1}{1 + \exp\left[-2\pi(E - E(n_y,n_z))/\hbar\omega_x\right]}, \qquad (8)$$

where, $E(n_y,n_z) = V_0 + \hbar\omega_y(n_y + 1/2) + \hbar\omega_z(n_z + 1/2)$, with $n_y, n_z = 0, 1, 2, \ldots$. We denote $E(n_y,n_z)$ as E_n in the 2D limit, where only a few subbands are below the Fermi energy E_F, one can consider $n_y, n_z = 1$ and $\omega_x, \omega_y, \omega_z = \omega$ and we have the transmission probability (T_{QPC}) as,

$$T_{QPC} = \frac{1}{1 + e^{-2\pi(E - E_1)/\hbar\omega}}, \qquad (9)$$

which has been shown in Fig. 5(a), where below a certain energy value E_1, it vanishes, and above this, it is always one. Here, $\hbar\omega$ is the width of the step. In our work, we have capacitively connected a gate voltage V_g to the QPC-type constriction, giving $V(n_x, n_y, n_z) \implies V(n_x, n_y, n_z) + U'$, which implies $E_1 \to E_1 + U'$, where $U' = eV_g$. We denote $E_1 + V'$ as E_1'.
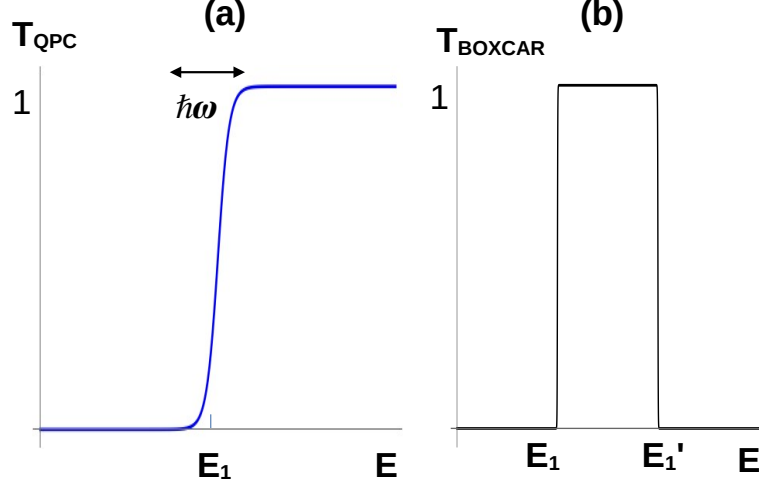
Figure 5: (a) QPC type, (b) Boxcar-type transmission.

Similarly, the boxcar-type of transmission is 1 in the energy range $E_1 < E < E_1'$, otherwise zero [17, 18], See Fig. 5(b). This transmission function will be used while we discuss Whitney's approach to studying the thermoelectrics of a two-terminal QH and QSH setup. As discussed by Whitney, one can generate a boxcar-type of transmission by considering many quantum dots in a region between two terminals [17, 18]. When there is only a single quantum dot with a single energy level, the transmission probability via it will be the resonant-tunneling type [43]. With one quantum dot, the electron will hop into it from one terminal and later hop out into the other terminal. But when one increases the number of quantum dots, each having a single energy level, then the electrons hop repeatedly between the quantum dots while going from one terminal to the other, which gives a boxcar-type transmission, which is 1 in between two arbitrary energy levels E_1 and E_1' as shown in Fig. 5(b). When $E_1' \rightarrow \infty$, the boxcar type transmission becomes QPC-like, always 1 above E_1 and zero otherwise.

However, implementing the boxcar-type transmission is more difficult than implementing the QPC-type transmission. In the derivation of boxcar-type transmission by Whitney [17, 18], the electron-electron interaction is completely ignored, meaning the coulomb blockade effects are also ignored. However, coulomb blockade effects cannot be ignored in real systems with real quantum dots [16]. This would mean that experimentally achieving the boxcar-type transmission by a chain of quantum dots is extremely difficult as the Coulomb blockade effect becomes an obstacle [16]. On the other hand, implementing a QPC-type transmission experimentally is easier. In fact, this was the first type of tunneling used to study the quantum transport in 2DEGs [20–23] such as GaAs-AlGaAs heterostructure. Most of the ballistic or edge mode transport can be done much more conveniently with the help of QPC in mesoscopic samples. For the implementation of QPC, one needs to control electrostatically the local electrostatic potential, which is felt by the electron with the help of metallic split gates [34–36]. Here, in this work, with the help of only the QPC-type of tunneling, one can study the nonlinear transport, which can help reach the Whitney bounds.

II.   CHARGE AND HEAT CURRENTS IN CHIRAL QH SETUP

For a multiterminal QH setup, the charge and heat current in terminal α can be derived using Landauer-Buttiker formalism. First, we focus on the charge current in terminal α, which has a Fermi-Dirac distribution $f_\alpha(E) = \left(1 + e^{\frac{E-\mu_\alpha}{k_B\theta_\alpha}}\right)^{-1}$, where μ_α and θ_α are the equilibrium chemical potential and temperature in terminal α respectively. The current outgoing from terminal α is given as,

$$I_\alpha^{\text{out}} = \frac{e}{L} \sum_k v f_\alpha(E), \quad \text{where the summation is over all '}k\text{' states of electrons.} \tag{10}$$

Electrons can enter terminal α from any other terminal β including itself. The fraction of particles incident from terminal β in edge channel m that scatter into terminal α in edge channel n is $\sum_{\beta,m,n} |s_{\alpha n,\beta m}|^2 f_\beta$, where $s_{\alpha n,\beta m}$ is the amplitude for an electron to scatter from terminal β in the edge channel m into terminal α in the edge channel n. Thus, the incoming current in terminal α

is given as,

$$I_\alpha^{\text{in}} = \frac{e}{L} \sum_k v \sum_{\beta,m,n} |s_{\alpha n, \beta m}|^2 f_\beta(E).$$ (11)

Thus, the total current in terminal α is,

$$I_\alpha = I_\alpha^{\text{out}} - I_\alpha^{\text{in}} = \frac{e}{L} \sum_k v \left(f_\alpha(E) - \sum_{\beta,m,n} |s_{\alpha n, \beta m}|^2 \right) f_\beta(E) = \frac{e}{L} \sum_k v \sum_{\beta,m,n} \left(\delta_{\alpha\beta}\delta_{mn} - |s_{\alpha n, \beta m}|^2 \right) f_\beta(E).$$ (12)

Converting the summation into integral, i.e., $\sum_k = 2 \times \frac{L}{2\pi} \int dk$, Eq. (12) becomes,

$$I_\alpha = 2e \int_{-\infty}^{\infty} \frac{dk}{2\pi} v(k) \sum_{\beta,m,n} \left(\delta_{\alpha\beta}\delta_{mn} - |s_{\alpha n, \beta m}|^2 \right) f_\beta(E).$$ (13)

where, the extra 2 has been taken due to spin degeneracy. Inserting $v(k) = \frac{1}{\hbar}\frac{dE}{dk}$ in Eq. (13), we get,

$$I_\alpha = \frac{2e}{h} \int_{-\infty}^{\infty} dE \sum_\beta f_\beta(E) [N_\alpha \delta_{\alpha\beta} - Tr(s_{\alpha\beta}^\dagger s_{\alpha\beta})],$$ (14)

where, N_α being the number of edge modes in terminal α and $s_{\alpha\beta}^\dagger s_{\alpha\beta} = |s_{\alpha\beta}|^2 = T_{\alpha\beta}$ is the transmission probability for an electron to transmit from terminal β to terminal α. From probability conservation, i.e., $\sum_\beta T_{\alpha\beta} = N_\alpha$, Eq. (14) can also be written as,

$$I_\alpha = \frac{2e}{h} \int_{-\infty}^{\infty} dE \sum_\beta T_{\alpha\beta}(f_\alpha - f_\beta).$$ (15)

This is the general formula for electric charge current irrespective of whether transport is linear or nonlinear. Now, the heat current can be found using the first law of thermodynamics. According to this law, the heat in terminal α with chemical potential μ_α, wherein volume is kept constant, is given as,

$$dQ_\alpha = dE_\alpha - \mu_\alpha dN_\alpha.$$ (16)

where, dE_α is the change in the internal energy and dN_α is the change in particle number in terminal α. Taking the time derivative of the above equation, we get $J_\alpha = J_{\alpha E} - \mu_\alpha J_{\alpha N}$, where J_α is the heat current, $J_{\alpha E}$ is the energy current and $J_{\alpha N}$ is the particle current in terminal α, given as

$$J_{\alpha E} = \frac{2}{h} \int_{-\infty}^{\infty} dE E \sum_\beta f_\beta(E) [N_\alpha \delta_{\alpha\beta} - Tr(s_{\alpha\beta}^\dagger s_{\alpha\beta})], \quad \text{and} \quad J_{\alpha N} = \frac{2}{h} \int_{-\infty}^{\infty} dE \sum_\beta f_\beta(E) [N_\alpha \delta_{\alpha\beta} - Tr(s_{\alpha\beta}^\dagger s_{\alpha\beta})].$$ (17)

The heat current is thus [16],

$$J_\alpha = \frac{2}{h} \int_{-\infty}^{\infty} dE (E - \mu_\alpha) \sum_\beta f_\beta(E) [N_\alpha \delta_{\alpha\beta} - Tr(s_{\alpha\beta}^\dagger s_{\alpha\beta})] = \frac{2}{h} \int_{-\infty}^{\infty} dE (E - \mu_\alpha) \sum_\beta T_{\alpha\beta}(f_\alpha - f_\beta).$$ (18)

This is the general formula for heat current irrespective of whether transport is linear or nonlinear.

A. Transport in the linear response regime in the QH setup

We can write charge and heat current as in Eqs. (15) and (18) in the linear response regime. We denote the chemical potential and temperature of reservoir i as μ_i and θ_i respectively, we is given as,

$$\mu_i = \mu + eV_i, \quad \theta_i = \theta + \tau_i, \quad i \in \text{all terminals},$$ (19)

with Fermi-Dirac distribution $f_i = \left(1 + e^{\frac{E-\mu_i}{k_B \theta_i}}\right)^{-1}$. μ and θ are the equilibrium chemical potential and temperature respectively

for terminal i with Fermi-Dirac distribution $f = \left(1 + e^{\frac{E-\mu}{k_B \theta}}\right)^{-1}$ and V_i and τ_i are the voltage bias and temperature bias applied

in terminals i. Now, doing a Taylor series expansion of f_i up to the first order, we get,

$$f_i = f + \left.\frac{\partial f_i}{\partial \mu_i}\right|_{\mu_i = \mu, \theta_i = \theta} eV_i + \left.\frac{\partial f_i}{\partial \theta_i}\right|_{\mu_i = \mu, \theta_i = \theta} \tau_i. \tag{20}$$

Using the property of f_i,

$$\left.\frac{\partial f_i}{\partial \mu_i}\right|_{\mu_i = \mu, \theta_i = \theta} = \frac{-\partial f}{\partial E} \quad \text{and,} \quad \left.\frac{\partial f_i}{\partial \theta_i}\right|_{\mu_i = \mu, \theta_i = \theta} = -\left(\frac{E-\mu}{k_B \theta}\right)\frac{\partial f}{\partial E} \implies f_i = f - \frac{\partial f}{\partial E}\left(eV_i + \frac{E-\mu}{k_B \theta}\tau_i\right). \tag{21}$$

We can use Eq. (21) in Eqs. (15) and (18) to get the Onsager coefficients. The charge current as in Eq. (15) becomes,

$$I_\alpha = \frac{2e}{h}\int_{-\infty}^{\infty} dE \sum_\beta \left(f - \frac{\partial f}{\partial E}\left(eV_\beta + \frac{E-\mu}{\theta}\tau_\beta\right)\right)[N_\alpha \delta_{\alpha\beta} - T_{\alpha\beta}], \tag{22}$$

and, the heat current as in Eq. (18) becomes,

$$J_\alpha = \frac{2}{h}\int_{\infty}^{\infty} dE (E-\mu) \sum_\beta \left(f - \frac{\partial f}{\partial E}\left(eV_\beta + \frac{E-\mu}{\theta}\tau_\beta\right)\right)[N_\alpha \delta_{\alpha\beta} - T_{\alpha\beta}], \tag{23}$$

where going from Eq. (18) to Eq. (23), we have taken $\mu_\alpha \simeq \mu$ and $T_{\alpha\beta} = Tr(s_{\alpha\beta}^\dagger s_{\alpha\beta})$. Now in Eqs. (22) and (23), the equilibrium distribution f makes zero contribution, since without any bias, there will be zero charge current and heat current, which is why Eqs. (22) and (23) reduce to,

$$I_\alpha = \sum_\beta (G_{\alpha\beta}V_\beta + L_{\alpha\beta}\tau_\beta), \quad \text{and} \quad J_\alpha = \sum_\beta (\Pi_{\alpha\beta}V_\beta + K_{\alpha\beta}\tau_\beta). \tag{24}$$

where, $G_{\alpha\beta} = \frac{2e^2}{h}\int_{-\infty}^{\infty} dE \left(-\frac{\partial f}{\partial E}\right)[N_\alpha \delta_{\alpha\beta} - T_{\alpha\beta}], \quad L_{\alpha\beta} = \frac{2e}{h\theta}\int_{-\infty}^{\infty} dE (E-\mu) \left(-\frac{\partial f}{\partial E}\right)[N_\alpha \delta_{\alpha\beta} - T_{\alpha\beta}],$

$$\Pi_{\alpha\beta} = \frac{2e}{h}\int_{-\infty}^{\infty} dE (E-\mu) \left(-\frac{\partial f}{\partial E}\right)[N_\alpha \delta_{\alpha\beta} - T_{\alpha\beta}], \quad K_{\alpha\beta} = \frac{2}{h\theta}\int_{-\infty}^{\infty} dE (E-\mu)^2 \left(-\frac{\partial f}{\partial E}\right)[N_\alpha \delta_{\alpha\beta} - T_{\alpha\beta}]. \tag{25}$$

The quantities as written in Eq. (25) are Onsager coefficients in linear response regime.

In the subsequent subsections, we derive Onsager coefficients for our QH setup (See, Sec. II A 1) with a three-terminal voltage-temperature probe. In Sec. II A 2, we derive the general formulae for power and efficiency as a quantum heat engine. Further in Sec. II A 3, we discuss the cooling power and coefficient of performance as a quantum refrigerator and in Sec. II A 4, we discuss why our QH setup can work as a quantum refrigerator even if time-reversal symmetry is preserved although, the QH setup used in [7] cannot be used as a quantum refrigerator.

1. Derivation of Onsager coefficients in a three-terminal QH setup in the linear response regime

It is also possible to find the transmission probabilities $T_{\alpha\beta}$ for this three-terminal QH arrangement (see Fig. 6). We talk about one instance, T_{12}, which is the probability that an electron will scatter from terminal 2 to terminal 1. For this process, the electron injects from terminal 2 with probability 1 and transmits via the constriction 2 and 1 with probabilities T_2 and T_1 respectively. Thus, the total probability for the electron to reach terminal 1 from terminal 2 is $T_1 T_2$. Similarly, one can determine other transmission probabilities and they are given as,

$$T_{11} = 1 - T_1, \quad T_{12} = T_1 T_2, \quad T_{13} = T_1(1 - T_2), \quad T_{21} = 0, \quad T_{22} = (1 - T_2), \quad T_{23} = T_2, \quad T_{31} = T_1,$$
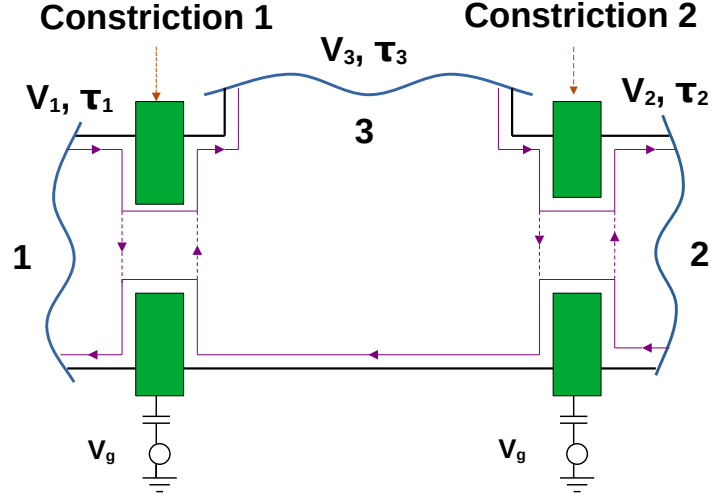$$T_{32} = (1 - T_1)T_2, \quad T_{33} = (1 - T_1)(1 - T_2). \tag{26}$$

Figure 6: Three terminal QH sample with chiral edge modes and two QPC-type constrictions. The edge mode of an electron, which can scatter from constriction, is shown by the purple solid (dashed) line. Terminal 3 is a voltage-temperature probe.

Here, we assume $\tau_1 = \tau$ and $\tau_2 = 0$. For the voltage-temperature probe in our setup (see Fig. 2), we solve for both V_3 and τ_3. According to Landauer-Buttiker formalism as shown in Eq. (24), the charge and heat currents in terminal 3 are given as,

$$I_3 = \sum_\beta G_{3\beta} V_\beta + \sum_\beta L_{3\beta} \tau_\beta, \quad J_3 = \sum_\beta \Pi_{3\beta} V_\beta + \sum_\beta K_{3\beta} \tau_\beta. \tag{27}$$

Putting $I_3 = J_3 = 0$, we get the solution for V_3 and τ_3. They are given as,

$$V_3 = \frac{(-K_{33}L_{31} + L_{33}K_{31})\tau - (-K_{33}G_{31} + L_{33}\Pi_{31})V}{X}, \quad \tau_3 = \frac{(-G_{33}K_{31} + \Pi_{33}L_{31})\tau - (-\Pi_{33}G_{31} + G_{33}\Pi_{31})V}{X}. \tag{28}$$

where, $X = G_{33}K_{33} - L_{33}\Pi_{33}$. Now, the current conservation in this setup implies $I_1^e = -I_2^e$. The charge and heat currents in terminal 1 are given as,

$$I_1 = \sum_\beta G_{1\beta} V_\beta + \sum_\beta L_{1\beta} \tau_\beta, \quad J_1 = \sum_\beta \Pi_{1\beta} V_\beta + \sum_\beta K_{1\beta} \tau_\beta. \tag{29}$$

Now, using V_3 and τ_3 from Eq. (28), Eq. (29) can be rewritten as,

$$\begin{pmatrix} I_1 \\ J_1 \end{pmatrix} = \begin{pmatrix} L_{eV} & L_{e\theta} \\ L_{hV} & L_{h\theta} \end{pmatrix} \begin{pmatrix} -V \\ \tau \end{pmatrix} \tag{30}$$

where,

$$
\begin{aligned}
L_{eV} &= G_{11} + \frac{G_{13}(L_{33}\Pi_{31} - G_{31}K_{33})}{X} - \frac{L_{13}(G_{33}\Pi_{31} - G_{31}\Pi_{33})}{X}, \\
L_{e\theta} &= L_{11} + \frac{G_{13}(L_{33}K_{31} - L_{31}K_{33})}{X} - \frac{L_{13}(G_{33}K_{31} - L_{31}\Pi_{33})}{X}, \\
L_{hV} &= \Pi_{11} + \frac{\Pi_{13}(L_{33}\Pi_{31} - G_{31}K_{33})}{X} - \frac{K_{13}(G_{33}\Pi_{31} - G_{31}\Pi_{33})}{X}, \\
L_{h\theta} &= K_{11} + \frac{\Pi_{13}(L_{33}K_{31} - L_{31}K_{33})}{X} - \frac{K_{13}(G_{33}K_{31} - L_{31}\Pi_{33})}{X}.
\end{aligned}
\tag{31}
$$

Here, Eq. (31) is the expression for Onsager coefficients. These coefficients determine the transport parameters such as charge conductance G, Seebeck coefficient S, Peltier coefficient Π, and thermal conductance K.

2. Power and efficiency in 3T QH setup as a quantum heat engine

Now, the charge power generated in terminal 1 is given as [16],

$$P = VI_1 = V(-L_{eV}V + L_{e\theta}\tau). \tag{32}$$

The maximum charge power can be obtained by finding V, for which $\frac{\partial P}{\partial V} = 0$, i.e., $V = \frac{L_{e\theta}\tau}{2L_{eV}}$. Thus maximum power is $P_{max} = \frac{L_{e\theta}^2\tau^2}{4L_{eV}}$.

Similarly, the efficiency at maximum power ($\eta|_{P_{max}}$) is,

$$\eta|_{P_{max}} = \frac{P_{max}}{J_1} = \frac{\theta\eta_c}{2} \frac{L_{e\theta}^2}{2L_{h\theta}L_{eV} - L_{hV}L_{e\theta}} = \frac{\eta_c}{2} \frac{Z\theta}{Z\theta + 2}. \tag{33}$$

Here, $Z\theta$ is the figure of merit, which is defined as

$$Z\theta = \frac{L_{hV}L_{e\theta}}{L_{eV}L_{h\theta} - L_{e\theta}L_{hV}} = \frac{GS^2\theta}{K}. \tag{34}$$

For $Z\theta \to \infty$, $\eta|_{P_{max}} = \frac{\eta_c}{2}$. The maximum value of efficiency at maximum power ($\eta|_{P_{max}}$) one can attain is half of Carnot efficiency, which is defined as Curzon-Ahlborn efficiency.

Similarly, one can find maximum efficiency. The general expression for efficiency is given as,

$$\eta = \frac{VI_1}{J} = \frac{V(-L_{eV}V + L_{e\theta}\tau)}{-L_{hV}V + L_{h\theta}\tau}. \tag{35}$$

The voltage bias V_{max} required to achieve maximum efficiency is for $\frac{\partial \eta}{\partial V} = 0$,

$$V_{max} = \frac{L_{h\theta}}{L_{hV}} \left(1 - \sqrt{\frac{L_{eV}L_{h\theta} - L_{e\theta}L_{hV}}{L_{eV}L_{h\theta}}}\right) \tau \tag{36}$$

which can be derived from the condition $\frac{d\eta}{dV} = 0$ and one can verify that $\frac{d^2\eta}{dV^2} < 0$. Thus, the maximum efficiency is given as [16]

$$\eta_{max} = \eta_c x \frac{\sqrt{Z\theta + 1} - 1}{\sqrt{Z\theta + 1} + 1}, \tag{37}$$

where, $x = \frac{\theta L_{e\theta}}{L_{hV}} = \frac{\theta S}{\Pi}$ is the asymmetric parameter (AP) of the setup. For $x = 1$ and in the limit $Z\theta \to \infty$, the maximum efficiency approaches Carnot efficiency.

3. Cooling power and coefficient of performance in 3T QH setup as a quantum refrigerator

One can similarly derive the maximum coefficient of performance (η_{max}^r) and cooling power at maximum coefficient of performance ($J|_{\eta_{max}^r}$) for a QH setup to act as a quantum refrigerator since the asymmetric parameter, i.e., the ratio of Seebeck to the Peltier coefficient is 1, which plays a major role in the setup's performance as a heat engine. Here, the heat will be absorbed from the cooler terminals (terminals 2 and 3 with voltage probe condition, and only terminal 2 with voltage-temperature probe) and dumped into the hotter terminal (terminal 1 in either voltage probe or voltage-temperature probe). For a quantum refrigerator, the coefficient of performance (η^r) is defined as the ratio of heat taken from the cooler terminal (J^Q) and the power absorbed (P) by the setup, i.e., $\eta^r = \frac{J^Q}{P}$. Here, $J^Q = -(J_2 + J_3)$ for voltage probe and $-J_2$ for voltage-temperature probe and can be derived from Eq. (30). Similarly, $P = I_1 V$ can also be derived from Eq. (30). Now, η_{max}^r can be found using the condition $\frac{d\eta^r}{dV} = 0$, which gives

$$V|_{\eta_{max}^r} = \frac{L_{h\theta}}{L_{hV}} \left(1 + \sqrt{\frac{L_{eV}L_{h\theta} - L_{hV}L_{e\theta}}{L_{eV}L_{h\theta}}}\right) \tag{38}$$

where, $V|_{\eta_{max}^r}$ is the voltage required to achieve η_{max}^r. From here onwards one can derive η_{max}^r by using Eq. (30) in $\eta^r = \frac{J^Q}{P}$, which yields

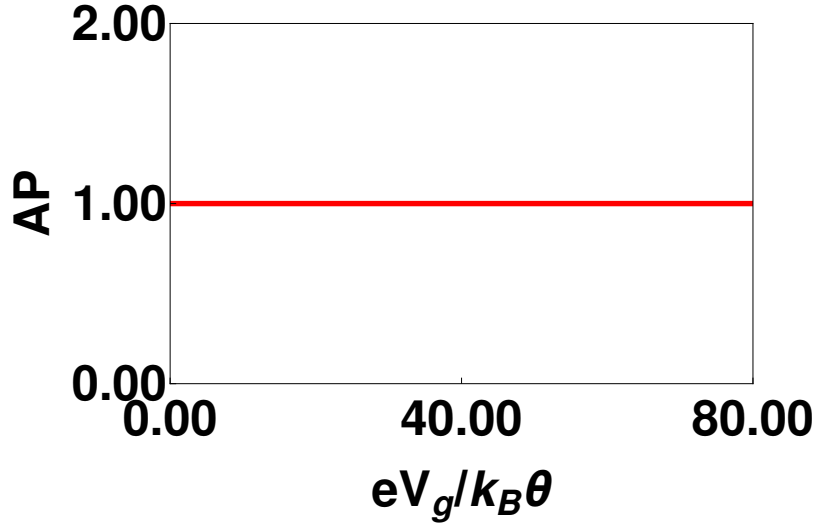$$\eta_{max}^r = \frac{\eta_c^r}{x} \frac{\sqrt{Z\theta + 1} - 1}{\sqrt{Z\theta + 1} + 1} \tag{39}$$

Figure 7: Asymmetric parameter as a function of gate voltage $\frac{eV_g}{k_B\theta}$ of our setup. Parameters taken are $\omega = 0.1 k_B\theta/\hbar$, $\theta = 1.0K$, $\mu = 0$, $E_1 = k_B\theta$.

where η_c^r = Carnot COP = $\frac{\theta}{\Delta\theta} = \eta_c^{-1}$ and x is the asymmetric parameter (AP). when $x = 1$ and $Z\theta \rightarrow \infty$, η_{max}^r approaches η_c^r. Similarly, the cooling power at maximum coefficient of performance ($J|_{\eta_{max}^r}$) is given as,

$$J|_{\eta_{max}^r} = L_{hV}V_{max} + L_{hT}\tau = L_{hT}\left(\sqrt{\frac{L_{eV}L_{h\theta} - L_{hV}L_{e\theta}}{L_{eV}L_{h\theta}}}\right)\tau. \tag{40}$$

4. Why the QH setup in Ref. [7] can not work as a quantum refrigerator, whereas our setup does?

In this section, we discuss why our QH setup can work as a quantum refrigerator, whereas the QH setup, which is considered in [7] cannot work. The ability of a QH setup to work as a quantum refrigerator is closely related to its AP, which is evident from Eq. (39). As shown in [29], when the AP deviates from one, then the coefficient of performance is reduced and can go to zero as AP is further increased, although it can still work as a quantum heat engine. There is a possibility that the coefficient of performance can achieve the Carnot limit η_c^r at AP = 1 only [29]. Now, in the setup considered in [7], the AP is either zero or infinity, resulting from time-reversal symmetry breaking. In this case, according to Ref. [29], the coefficient of performance should be very small, which means the setup's ability to work as a quantum refrigerator is reduced. But in our setup, the AP is always one, shown in Fig. 7, enabling us to use it as a quantum refrigerator. One can also see this analytically using Eq. (31). From the $L_{e\theta}$ expression in Eq. (31), one can see that

$$\theta L_{e\theta} = \Pi_{11} + \frac{G_{13}(\Pi_{33}K_{31} - \Pi_{31}K_{33})}{X} - \frac{\Pi_{13}(G_{33}K_{31} - \Pi_{31}\Pi_{33})}{X} \tag{41}$$

Comparing Eq. (41) with L_{hV} of Eq. (31), one can see that both of them are almost same, which makes AP to be one regardless of any parameter.

B. Transport in the nonlinear response regime in the QH setup

In this section, we turn to the non-linear regime of transport [17, 18] in the same QH setup as shown in Fig. 6. From Landaeur-Buttiker formalism, the charge and heat current in terminal α as derived in Eqs. (15) and (18), at finite temperature are given as,

$$I_\alpha = \frac{2e}{h}\int_{-\infty}^{\infty}dE\sum_\beta T_{\alpha\beta}(f_\alpha - f_\beta), \quad J_\alpha = \frac{2}{h}\int_{-\infty}^{\infty}dE(E - \mu_\alpha)\sum_\beta T_{\alpha\beta}(f_\alpha - f_\beta). \tag{42}$$

As discussed in Sec. II A, when applied biases are small compared to the relevant thermal energy scale ($eV \ll k_B\theta$, $\tau \ll \theta$), then we are in the linear transport regime and when it is not that small i.e., $eV \approx k_B\theta$, $\tau \approx \theta$, then we are in the non-linear transport regime. In case of linear thermoelectrics, electron transport is only dependent on its kinetic energy, whereas, in non-linear regime, electron transport is dependent on both its kinetic energy as well as the potential (U) of the sample where it flows [44, 45]. When the voltage bias and temperature bias are small compared to the relevant thermal energy scale ($k_B\theta$) i.e., in the linear transport regime, then U is much smaller compared to the kinetic energy of the electron and does not play much role in electron transport, and therefore we ignore it in the linear response regime [44, 45]. However, in the nonlinear transport regime, the applied biases are not small and are comparable to $k_B\theta$, and therefore in this case, U cannot be ignored. When the electrons are injected from a terminal into a sample, it perturbs the charge distribution inside the sample, eventually affecting the behavior of electrons via electron-electron interaction. This means the potential energy of the whole system is changed by the voltage and temperature biases applied, which ultimately tells us that the transmission probability $T_{\alpha\beta}$ is not only a function of kinetic energy here, but a function of potential (U) as well, which in turn is dependent on voltage as well as temperature biases applied across the sample. This has to be calculated in a self-consistent manner [44–48]. This is why $T_{\alpha\beta}$ is dependent only on the kinetic energy of the electron for linear transport, while it is dependent on both kinetic energy as well as the potential of the sample in the nonlinear transport regime [44–48].

In the subsequent subsections, we discuss the two-terminal (2T) QH setup as a nonlinear quantum heat engine and a quantum refrigerator in Sec. II B 1 using Whintey's approach. In Sec. II B 2, we discuss the two-terminal QH set up both as a heat engine and refrigerator using our approach, which is an amalgamation of the approaches of Sanchez, Lopez and Büttiker [44–48] and Haack, Giazotto [27]. We then extend our approach to a three-terminal (3T) QH setup in Sec. II B 3. Finally, in Sec. II B 4, we use our approach to discuss the thermoelectric properties of a 3T QH setup with a voltage-temperature probe.

1. Nonlinear thermoelectrics in 2T QH setup (Whitney's approach)

Here, we discuss the nonlinear thermoelectricity of a 2T QH setup as shown in Fig. 8(a) using the approach of Whitney following the Refs. [17, 18]. We consider the voltage biases to be $V_1 = -V, V_2 = 0$ and θ_1 and θ_2 being the temperatures of terminals 1 and 2 respectively. Here, the power generated (P) as a quantum heat engine is given as

$$P = (V_2 - V_1)I_1 = \frac{2eV}{h}\int_{-\infty}^{\infty} dE \quad T_{12}(f_1 - f_2),$$ (43)

where, T_{12} represents the transmission probability that an electron will scatter from terminal 2 to terminal 1, and the f_1 and f_2 are the Fermi-Dirac distributions of reservoirs 1 and 2 respectively. Similarly, from Eq. (18), the heat current out of terminal 1 is

$$J_1 = \frac{2}{h}\int_{-\infty}^{\infty} dE(E + eV)T_{12}(f_1 - f_2),$$ (44)

where we have taken $\mu_1 = -eV$. Using Eqs. (43) and (44), one can evaluate the efficiency i.e., $\eta = P/J_1$.

In [17, 18], Whitney proved that T_{12} needs to be a boxcar-type transmission as shown in Fig. 1(b) to extract the best possible performance as a heat engine. Considering a single edge mode, the boxcar-type transmission probability $T_{12} = 1$ for $E_1 < E < E_1'$, otherwise it is zero. When $E_1' \to \infty$, T_{12} is a QPC type, where it is one above E_1 and zero otherwise as shown in Fig. 1. Here, $E_1 = eV / \left(\frac{\theta_1}{\theta_2} - 1\right)^{-1}$ and $E_1' = eVJ_1'/P'$ [17, 18], where the prime in J_1 and P denotes the first derivatives with respect to V. The maximum possible power (P_{max}) is generated when $P' = 0$, which makes $E_1' = \infty$, and consequently one can also derive the efficiency at the maximum power ($\eta|_{P_{max}}$). Similarly, the maximum efficiency (η_{max}) can be achieved when E_1' is closer to E_1 so that the finite power (P) generated won't be only much lesser than P_{max}, but the maximum efficiency will also be close to Carnot efficiency (η_c). All these quantities of interest such as P_{max}, $\eta|_{P_{max}}$ and η_{max} as derived by Whitney are given as [17, 18],

$$P_{max} = P_{max}^{Wh} = 0.0642\frac{\pi^2 k_B^2 \tau^2}{h}, \quad \eta|_{P_{max}} = \eta|_{P_{max}^{Wh}} = \frac{\eta_c}{1 + 0.936(1 + \theta_2/\theta_1)}, \quad \eta_{max} = \eta_{max}^{Wh} = \eta_c\left(1 - 0.478\sqrt{\frac{\theta_2 P}{\theta_1 P_{max}^{Wh}}}\right), \quad (45)$$

where, $\tau = \theta_1 - \theta_2 =$ temperature bias applied across the sample, $\eta_c =$ Carnot efficiency, $\theta_1 =$ temperature in terminal 1, $\theta_2 =$ temperature in terminal 2, $P =$ Finite output power with $P \ll P_{max}^{Wh}$ and we have taken spin degeneracy into account.

Similarly, for a quantum refrigerator, the cooling power ($J = -J_2$) of this 2T QH setup is given as

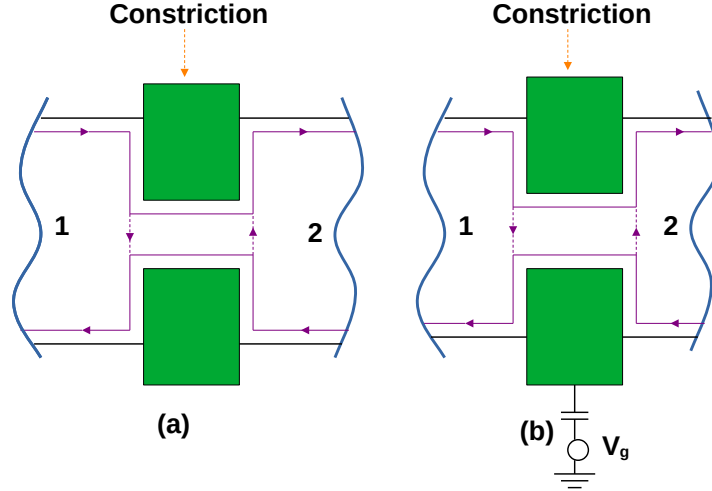$$J = J_2 = \frac{2}{h}\int_{-\infty}^{\infty} dEET_{21}(f_2 - f_1).$$ (46)

Figure 8: (a) Two-terminal QH setup used in Whitney's approach, where the constriction in between has boxcar-type of transmission, (b) Two-terminal QH setup used in our approach, where the constriction has QPC-type transmission and capacitively connected to the external gate voltage.

Using the cooling power J and the power absorbed $P = I_1 V$ (see, Eq. (43)), one can determine the coefficient of performance i.e., $\eta^r = J/P$. Similar to the quantum heat engine also, T_{21} should be a Boxcar-type of transmission probability, which provides the best possible performance as a refrigerator in a two-terminal QH nonlinear setup as proved in Refs. [17, 18], except E_1' is eVJ_2'/P' and with the form of E_1 remaining same. Now the maximum cooling power (J_{max}) can only be extracted when $J_2' = 0$ and the applied voltage bias is large making $E_1' = 0$ and $E_x \to \infty$ [17, 18]. When the maximum cooling power is reached, then due to the large voltage bias, the power absorbed P by the system will be large, which makes the coefficient of performance (η^r) vanish. Similarly, the maximum coefficient of performance (η^r_{max}) can be achieved when E_x is closer to E_1' and the extracted cooling power (J) is much lesser than J_{max}. So, the quantities of interest such as J_{max} and η^r_{max} as derived by Whitney are given as [17, 18]

$$J_{max} = J_{max}^{Wh} = \frac{\pi^2}{6h} N k_B^2 \theta_2^2, \quad \text{and} \quad \eta^r_{max} = \eta^{r,Wh}_{max} = \eta^r_c \left(1 - 1.09 \sqrt{\frac{\theta_2}{\theta_1 - \theta_2} \frac{J}{J_{max}^{Wh}}} \right), \tag{47}$$

where η^r_c = Carnot coefficient of performance, θ_1 = temperature in terminal 1, θ_2 = temperature in terminal 2, J = Finite cooling output power and $J \ll J_{max}^{Wh}$.

2. Nonlinear thermoelectrics in 2T QH setup (Our approach)

Now, we discuss the nonlinear thermoelectrics in a 2T QH setup as shown in Fig. 8(b) by considering a QPC-like constriction between the two terminals. As discussed in the introduction, See Sec. II B, in the nonlinear response regime, the interaction potential (U) comes into the picture because the applied biases are not small. Thus, this interaction potential must be a function of the voltage bias (V) as well as the temperature bias (τ). In that case, the general form of the interaction potential U in a mesoscopic system is [46–48]

$$U = \sum_\alpha u_\alpha V_\alpha + \sum_\alpha z_\alpha \tau_\alpha. \tag{48}$$

where, α is the index for terminals, with $u_\alpha = \frac{\partial U}{\partial V_\alpha}$ and $z_\alpha = \frac{\partial U}{\partial \tau_\alpha}$ are the so-called characteristic potentials, which determine the response of interaction potential U to an applied voltage bias V_α and temperature bias τ_α respectively in terminal α. The characteristic potentials u_α and z_α can be derived from particle injectivity $v_\alpha^p(E)$ and entropic injectivity $v_\alpha^e(E)$ [46–48]. The

expression for $v_\alpha^p(E)$ and $v_\alpha^e(E)$ are given as,

$$v_\alpha^p(E) = \frac{1}{2\pi i}\sum_\beta \text{Tr}\left[s_{\beta\alpha}^\dagger \frac{ds_{\beta\alpha}}{dE}\right], \quad v_\alpha^e(E) = \frac{1}{2\pi i}\sum_\beta \text{Tr}\left[\frac{E}{\theta}s_{\beta\alpha}^\dagger \frac{ds_{\beta\alpha}}{dE}\right]. \tag{49}$$

where $s_{\beta\alpha}$ represents the scattering amplitude that an electron will scatter to terminal β from terminal α.

When the gate voltage is capacitively applied to the constriction, see Fig. 8(b). Therefore, there will be an additional contribution to U due to the gate voltage as derived in Refs. [46–48] and the modified U is,

$$U' = U + u_g V_g = \sum_\alpha u_\alpha V_\alpha + \sum_\alpha z_\alpha \tau_\alpha + u_g V_g, \tag{50}$$

where, u_g is the characteristic potential, which determine the response of U to the external gate voltage V_g.

In [46, 47], the derivations to the characteristic potentials in a two-terminal ballistic setup with a resonant-tunneling-type of constriction in between the two terminals has been done, whereas in [48], the same derivation has been done in presence of QPC-like constriction. For a two-terminal ballistic setup either with QPC-like or resonant tunneling-like, the general form of the interaction potential U' using Eq. (50) is given as

$$U' = u_1 V_1 + u_2 V_2 + z_1 \tau_1 + z_2 \tau_2 + u_g V_g. \tag{51}$$

In the presence of either resonant-tunneling or QPC-like constriction with single-level energy (E'), the form of the characteristic potentials is given as [46–48]

$$u_1 = \frac{e^2 D_1^p}{C+e^2 D}, \quad u_2 = \frac{e^2 D_2^p}{C+e^2 D}, \quad z_1 = \frac{e D_1^e}{C+e^2 D}, \quad z_2 = \frac{e D_2^e}{C+e^2 D}, \quad u_g = \frac{C}{C+e^2 D}. \tag{52}$$

where, $D_1^p = -\int dE v_1^p(E)\frac{\partial f}{\partial E}$, $D_2^p = -\int dE v_2^p(E)\frac{\partial f}{\partial E}$, $D_1^e = -\int dE v_1^e(E)\frac{\partial f}{\partial E}$, $D_2^e = -\int dE v_1^e(E)\frac{\partial f}{\partial E}$, where $D = D_1^p + D_2^p$, with the injectivities $v_1^p, v_2^p, v_1^e, v_2^e$ given as (See, Eq. (49)),

$$v_1^p = \frac{1}{2\pi i}\text{Tr}\left[s_{11}^\dagger \frac{ds_{11}}{dE} + s_{21}^\dagger \frac{ds_{21}}{dE}\right], \quad v_2^p = \frac{1}{2\pi i}\text{Tr}\left[s_{11}^\dagger \frac{ds_{11}}{dE} + s_{12}^\dagger \frac{ds_{12}}{dE}\right],$$

$$v_1^e = \frac{E}{2\pi i\theta}\text{Tr}\left[s_{11}^\dagger \frac{ds_{11}}{dE} + s_{21}^\dagger \frac{ds_{21}}{dE}\right], \quad v_2^e = \frac{E}{2\pi i\theta}\text{Tr}\left[s_{11}^\dagger \frac{ds_{11}}{dE} + s_{12}^\dagger \frac{ds_{12}}{dE}\right], \tag{53}$$

Here, f is the equilibrium Fermi-Dirac distribution, when there are no biases applied across the sample, i.e., $f = \left(1+e^{\frac{E}{k_B\theta}}\right)^{-1}$. D is a Lindhard screening function, which determines the screening charge that is developed inside the conductor due to the interaction of injected charges from the reservoirs with the interaction potential U [44–48], that is developed due to the voltage and temperature biases in terminal α, and the expression for U is given in Eq. (48). For the resonant-tunneling, the s-matrix elements $s_{\beta\alpha}$ are given in [43] and for QPC-type, it is given in [26]. Using the s-matrix elements, one can find the injectivities $v_{1(2)}^p, v_{1(2)}^e$, for a resonant tunneling or QPC and subsequently find $D_{1(2)}^p, D_{1(2)}^e$. . In the case of resonant tunneling [46, 47], $D_{1(2)}^p$ and $D_{1(2)}^e$ are dependent upon the width Γ of the resonant tunneling as well as single energy level E_{RT}'. Similarly, for the QPC-like constriction also, $D_{1(2)}^p, D_{1(2)}^e$ are dependent on the width ω and its single energy level E_{QPC}' [48]. If the energy level of QPC is fixed and the widths of the constrictions, then for a fixed capacitance C, one sees that the characteristic potentials $u_{1(2)}, z_{1(2)}$ are constants. Therefore, u_g will also be a constant.

Now, for our purpose, we will restrict our discussion to a setup with QPC-like constriction [48]. In the presence of interaction potential U, the single energy level of QPC: E_{QPC}' is modified to $E_{QPC}' + eU$. In the linear response regime, the first and second terms of RHS of Eq. (50) are ignored as voltage and temperature biases are very small in that regime. The only term that stays relevant in the linear response regime is the third term of Eq. (50). But, D in the case of linear response regime is zero, which means $u_g = 1$, because there won't be any screening of charges inside the conductor, thus $D = 0$. In the case of linear response, the modified single energy level in the presence of gate voltage is $E_{QPC}'' = E_{QPC}' + eV_g$, but in case of a nonlinear-response regime, the modified single energy level is $E_{QPC}'' = E_{QPC}' + U'$. In our setup, we consider $V_1 = -V, V_2 = 0, \tau_1 = \tau, \tau_2 = 0$ with QPC-like constriction. In this case, the interaction potential U using Eq. (48) is

$$U = u_1 V_1 + z_1 \tau_1 = -u_1 V + z_1 \tau, \tag{54}$$

and, when the constriction is capacitively connected to the gate voltage V_g, the interaction potential U is modified as (See Eq. (51))

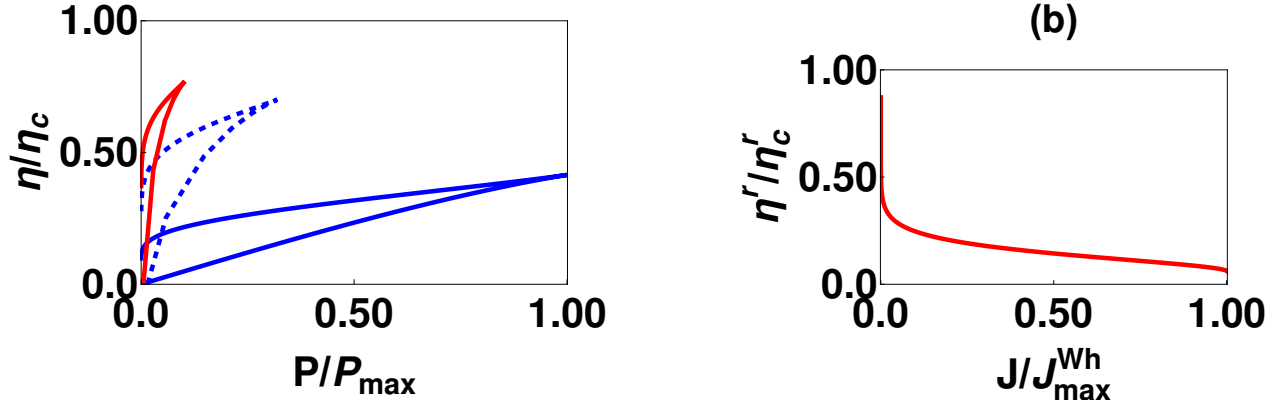$$U' = -u_1 V + z_1 \tau + u_g V_g = U + u_g V_g, \tag{55}$$

Figure 9: (a) Parametric plot of the power (P) and efficiency as a heat engine. The blue thick line corresponds to $V = 0.57 k_B \theta / e$, blue dotted line corresponds to $V = 1.77 k_B \theta / e$, red thick line corresponds to $V = 2.52 k_B \theta / e$, (b) Parametric plot of the cooling power (J) and coefficient of performance (η^r) as a refrigerator of a two-terminal QH (chiral) setup at $V = 20 k_B \theta / e$. Here, we have taken $E'_{QPC} = k_B \theta$, where $\theta_1 = 2\theta_2 = 2K$ and $\tau = \theta_1 - \theta_2 = 1K$. For the QPC, we have taken $\omega = 0.1 k_B \theta / e$. These results are similar to the three-terminal QH setup without any probe with $V_1 = -V, V_2 = V_3 = 0$ and $\tau_1 = \tau, \tau_2 = \tau_3 = 0$.

where, the other characteristic potentials u_2 and z_2 are ignored as there are no applied biases in terminal 2. The form of u_1, z_1 and u_g is same as in Eq. (55). As already discussed, the quantities such as u_1, u_2, and u_g are constant when C and ω are constant. For this setup, it is assumed that capacitance C is constant, while the width of the QPC ω is also constant, which implies that u_1, z_1 and u_g will also be constant throughout the work. This can be done experimentally as well. If the setup is kept at a constant voltage bias V and temperature bias τ, then the interaction potential U will also be constant. At the same time, one can always apply V_g to nullify U as discussed in Ref. [27] depending upon the value of V and τ. Therefore, at a fixed voltage bias V and τ, for the quantum heat engine, we consider the gate voltage to be

$$V_g = ((u_1 V - z_1 \tau) + V'_g)/u_g = (-U + V'_g)/u_g, \tag{56}$$

The first term inside the parenthesis of V_g in Eq. (56) is used to nullify the interaction potential $U = -(u_1 V - z_1 \tau)$ and the second term V'_g is used to lift the energy level of QPC, which means

$$E''_{QPC} = E'_{QPC} + eU + eu_g V_g = E'_{QPC} + eV'_g. \tag{57}$$

Now, the transmission probability of the QPC becomes

$$T_{QPC} = \frac{1}{1 + e^{-2\pi(E - E''_{QPC})/\hbar\omega}} = \frac{1}{1 + e^{-2\pi(E - E'_{QPC} - eV'_g)/\hbar\omega}}, \tag{58}$$

and this form is always maintained at any voltage bias V and V'_g. With this, the power delivered ($P = I_1 V$) and efficiency ($\eta = P/J_1$) can be found using the formulae given in Eq. (42). Now, in this work, we fix the values of V and τ and evaluate the power and efficiency with different values of V'_g. We plot them parametrically in Fig. 9(a). Now, we fix $V = 1.14 k_B \theta / e$, $\tau = 1K$, we find that at a particular V_g, the maximum power approaches P^{Wh}_{max} as shown in Fig. 9(a).

Similarly, for the refrigerator, again we consider $V_g = -U + V'_g$. We now calculate the cooling power $J = J_2$ and coefficient of performance ($\eta^r = J/P$), where $P = I_1 V$ for different values of V. The parametric plot of the cooling power J and η^r is shown in Fig. 5(b). The maximum cooling power approaches J^{Wh}_{max} and the best coefficient of performance reach is 86% of η^r_c. Here, we conclude that one can achieve the best thermoelectric performance as a refrigerator in our two-terminal setup with maximum possible cooling power approaching J^{Wh}_{max} and 87% of η^r_c.

3. Nonlinear thermoelectrics in 3T QH setup (Our approach)

Here, we discuss the power and efficiency of a simple three-terminal QH setup without any voltage probe or voltage-temperature probes with two QPC-like constrictions as shown in Fig. 6 by controlling the transmission of QPC 1 and QPC

2 by an applied gate voltage using the conditions: $E'_1 = E_1 + eV_g$ and $E'_2 = E_2 + eV_g$, with $E_1 = E_2$. We consider a simple case with $V_1 = -V, V_2 = V_3 = 0$, and $\tau_1 = \tau, \tau_2 = \tau_3 = 0$. Similar to the 2T QH case as in Sec. II B 2, the interaction potential in the presence of a gate voltage of the same form as Eq. (54). Therefore, the interaction potential can be nullified by considering the gate voltage V_g of the same form as in Eq. (56). For the heat engine, the power delivered ($P = I_1 V$) and efficiency ($\eta = P/J_1$) can be found by using the formula given in Eq. (42) for different values of voltage bias (V), the performance exactly matches with the result of two-terminal heat engine as shown in Fig. 9(a). Similarly, for the refrigerator, we take $V_g = -U + V'_g$ and cooling power ($J = (J_2 + J_3)$) and coefficient of performance ($\eta^r = J/P$) exactly matches the result of the two-terminal refrigerator as in Fig. 9(b).

4. Nonlinear thermoelectrics in 3T QH setup with voltage-temperature probe (Our approach)

Next, we extend our approach to a 3T nonlinear QH setup as shown in Fig. 6 with a voltage-temperature probe by controlling the transmission of QPC 1 and QPC 2 via a gate voltage. We consider the threshold of energies of QPC's $E'_1 = E_1 + eV_g$ and $E'_2 = E_2 + eV_g$ above which the transmission is 1 with $E_1 = E_2$. We consider terminal 3 to be a voltage-temperature probe and calculate V_3 and τ_3 from the voltage-temperature condition $I_3 = J_3 = 0$, where $V_1 = -V, V_2 = 0, \tau_1 = \tau, \tau_2 = 0$. This is hard to solve analytically, therefore we do it numerically using Mathematica. The Mathematica code is uploaded to GitHub [33]. Here, in the presence of gate voltage V_g, the interaction potential U' is given as

$$U' = -u_1 V + z_1 \tau + u_3 V_3 + z_3 \tau_3 + u_g V_g = U + u_g V_g, \tag{59}$$

where, $U = -u_1 V + z_1 \tau + u_3 V_3 + z_3 \tau_3$ is the interaction potential in the absence of the gate voltage V_g. As discussed in Sec. II B 2, the characteristic potentials are constant if one assumes the capacitance C and the width of the QPC ω are constant. The solutions V_3 and τ_3 will also be constants if the voltage bias V and temperature bias τ are kept constant. Now, one can consider the gate voltage V_g to be

$$V_g = (-U + V'_g)/u_g, \tag{60}$$

Which will nullify the interaction potential U and the energy level of the QPC becomes $E''_{QPC} = E_{QPC} + eV'_g$. Considering this, one can calculate power and efficiency as a quantum heat engine by changing V'_g at fixed values of V, τ and we observe that the maximum power approaches the Whitney bound, whereas the maximum efficiency is 0.93 η_c as shown in Fig. 4(a) of the main text.

Similar to the quantum heat engine, we investigate the three-terminal QH setup as a quantum refrigerator with both voltage probe and voltage-temperature probe by our approach and considering identical values of E'_1 and E'_2 same as used for quantum heat engine. The coefficient of performance $\eta^r = \frac{J}{P}$ and the cooling power J is $(J_2 + J_3)$ are evaluated with either probe and the parametric plot are shown in Fig. 4(c) of the main text. With these conditions, in the three-terminal QH setup for a voltage-temperature probe, we observe that J_{max} reaches J^{Wh}_{max}, while η^r reaches about 86% of η^r_c as shown in Fig. 4(c) of the main text.

III. CHARGE AND HEAT CURRENT IN QSH SETUP

For a multiterminal QSH setup, the charge and heat current in terminal α can be derived using Landauer-Buttiker formalism [12, 49]. The total current in terminal α is a sum of currents due to spin-up as well as spin-down electrons. The charge current due to an electron with spin $\sigma = \uparrow$ or $\downarrow$ is,

$$I^\sigma_\alpha = \frac{e}{h} \int_{-\infty}^{\infty} dE \sum_\beta f_\beta(E)[N_\alpha \delta_{\alpha\beta} - T^\sigma_{\alpha\beta}], \tag{61}$$

where, $T^\sigma_{\alpha\beta} = \sum_{\rho=\uparrow/\downarrow} T^{\sigma\rho}_{\alpha\beta}$ with $T^{\sigma\rho}_{\alpha\beta}$ represents the transmission probability that an electron will scatter from terminal β with spin ρ to terminal α with spin σ. Similarly, the heat current is given as,

$$J^\sigma_\alpha = \frac{1}{h} \int_{-\infty}^{\infty} dE(E - \mu_\alpha) \sum_\beta f_\beta(E)[N_\alpha \delta_{\alpha\beta} - T^\sigma_{\alpha\beta}]. \tag{62}$$

A. Transport in the linear response regime in the QSH setup

Similar to the chiral case, in the linear response regime, both charge and heat currents can be derived as,

$$I_\alpha^\sigma = \sum_\beta (G_{\alpha\beta}^\sigma V_\beta + L_{\alpha\beta}^\sigma \tau_\beta), \quad J_\alpha^\sigma = \sum_\beta (\Pi_{\alpha\beta}^\sigma V_\beta + K_{\alpha\beta}^\sigma \tau_\beta), \tag{63}$$

where,

$$G_{\alpha\beta}^\sigma = \frac{e^2}{h} \int_{-\infty}^{\infty} dE \left(-\frac{\partial f}{\partial E}\right) [N_\alpha \delta_{\alpha\beta} - T_{\alpha\beta}^\sigma], \quad L_{\alpha\beta}^\sigma = \frac{e}{h\theta} \int_{-\infty}^{\infty} dE (E - \mu) \left(-\frac{\partial f}{\partial E}\right) [N_\alpha \delta_{\alpha\beta} - T_{\alpha\beta}^\sigma],$$

$$\Pi_{\alpha\beta}^\sigma = \frac{e}{h} \int_{-\infty}^{\infty} dE (E - \mu) \left(-\frac{\partial f}{\partial E}\right) [N_\alpha \delta_{\alpha\beta} - T_{\alpha\beta}^\sigma], \quad K_{\alpha\beta}^\sigma = \frac{1}{h\theta} \int_{-\infty}^{\infty} dE (E - \mu)^2 \left(-\frac{\partial f}{\partial E}\right) [N_\alpha \delta_{\alpha\beta} - T_{\alpha\beta}^\sigma]. \tag{64}$$

The quantities as written in Eq. (64) are the spin-polarized Onsager coefficients for helical edge mode transport.

In the subsequent subsections, we derive Onsager coefficients for our QH setup (See, Sec. III A 1) with a three-terminal voltage-temperature probe. In Sec. III A 2, we derive the general formulae for power and efficiency as a quantum heat engine. Further in Sec. III A 3, we discuss the cooling power and coefficient of performance as a quantum refrigerator.

1. Derivation of Onsager coefficients in a three-terminal QSH setup in the linear response regime

In this section, we discuss the thermoelectric properties of a three-terminal QSH setup as shown in Fig. 4 with terminal 3 acting as either a voltage probe or a voltage-temperature probe. For this setup, we consider $\tau_2 = \tau_3 = 0$, i.e., $\theta_2 = \theta_3 = \theta$ and $\theta_1 = \theta + \tau_1$, wherein $\tau_1 = \tau$. Similarly for the QSH setup as shown in Fig. 8, there will be electron motion via both spin-up and spin-down electrons in the opposite directions because of helical edge mode transport. Thus, for $s = \uparrow / \downarrow$ and $\sigma = \uparrow / \downarrow$, the transmission probabilities ($T_{\alpha\beta}^{s\sigma}$) that an electron will scatter from terminal β with initial spin σ to enter terminal α with final spin s are as follows,

$$T_{11}^{\uparrow\uparrow} = T_{11}^{\downarrow\downarrow} = (1 - T_1), \quad T_{12}^{\uparrow\uparrow} = T_{21}^{\downarrow\downarrow} = T_1 T_2, \quad T_{12}^{\downarrow\downarrow} = T_{21}^{\uparrow\uparrow} = 0, \quad T_{13}^{\uparrow\uparrow} = T_{31}^{\downarrow\downarrow} = T_1(1 - T_2), \quad T_{13}^{\downarrow\downarrow} = T_{31}^{\uparrow\uparrow} = T_1$$

$$T_{22}^{\uparrow\uparrow} = T_{22}^{\downarrow\downarrow} = (1 - T_2), \quad T_{23}^{\uparrow\uparrow} = T_{32}^{\downarrow\downarrow} = T_2, \quad T_{23}^{\downarrow\downarrow} = T_{32}^{\uparrow\uparrow} = (1 - T_1)T_2, \quad T_{33}^{\uparrow\uparrow} = T_{33}^{\downarrow\downarrow} = (1 - T_1)(1 - T_2). \tag{65}$$

Similar to the previous case, we consider $\tau_1 = \tau$, $\tau_2 = 0$, but we consider terminal 3 to be a voltage-temperature probe, i.e., both the charge current I_3 and the heat current J_3 through terminals are zero, and terminal 2 is a current probe, with, $V_2 = 0$. Here, chemical potential μ is taken to be zero. We derive the Onsager coefficients, as follows from Eq. (63), the charge and heat currents in terminal 3 is given as,

$$I_3 = \sum_{\sigma \in \{\uparrow,\downarrow\}} I_3^\sigma = \sum_{\sigma \in \{\uparrow,\downarrow\}} \left(\sum_\beta G_{3\beta}^\sigma V_\beta + \sum_\beta L_{3\beta}^\sigma \tau_\beta\right), \quad J_3 = \sum_{\sigma \in \{\uparrow,\downarrow\}} J_3^\sigma = \sum_{\sigma \in \{\uparrow,\downarrow\}} \left(\sum_\beta \Pi_{3\beta}^\sigma V_\beta + \sum_\beta K_{3\beta}^\sigma \tau_\beta\right). \tag{66}$$

Substituting $I_3 = J_3 = 0$, we get solutions for V_3 and τ_3, as,

$$V_3 = \frac{(-K_{33}L_{31} + L_{33}K_{31})\tau - (-K_{33}G_{31} + L_{33}\Pi_{31})V}{X}, \quad \tau_3 = \frac{(-G_{33}K_{31} + \Pi_{33}L_{31})\tau - (-\Pi_{33}G_{31} + G_{33}\Pi_{31})V}{X}. \tag{67}$$

where, $X = (G_{33}K_{33} - L_{33}\Pi_{33})$ and $G_{ij} = \sum_{\sigma \in \{\uparrow,\downarrow\}} G_{ij}^\sigma, L_{ij} = \sum_{\sigma \in \{\uparrow,\downarrow\}} L_{ij}^\sigma, \Pi_{ij} = \sum_{\sigma \in \{\uparrow,\downarrow\}} \Pi_{ij}^\sigma$ and $K_{ij} = \sum_{\sigma \in \{\uparrow,\downarrow\}} K_{ij}^\sigma$ for $i, j = 1, 2, 3$. The spin-polarized charge and heat currents in terminal 1 is given as,

$$I_1^\sigma = \sum_\beta G_{1\beta}^\sigma V_\beta + \sum_\beta L_{1\beta}^\sigma \tau_\beta, \quad \text{and} \quad J_1^\sigma = \sum_\beta \Pi_{1\beta}^\sigma V_\beta + \sum_\beta K_{1\beta}^\sigma \tau_\beta. \tag{68}$$

Using V_3 and τ_3 from Eq. (67), I_1^σ and J_1^σ can be written as,

$$\begin{pmatrix} I_1^\sigma \\ J_1^\sigma \end{pmatrix} = \begin{pmatrix} L_{eV}^\sigma & L_{e\theta}^\sigma \\ L_{hV}^\sigma & L_{h\theta}^\sigma \end{pmatrix} \begin{pmatrix} -V \\ \tau \end{pmatrix}, \tag{69}$$
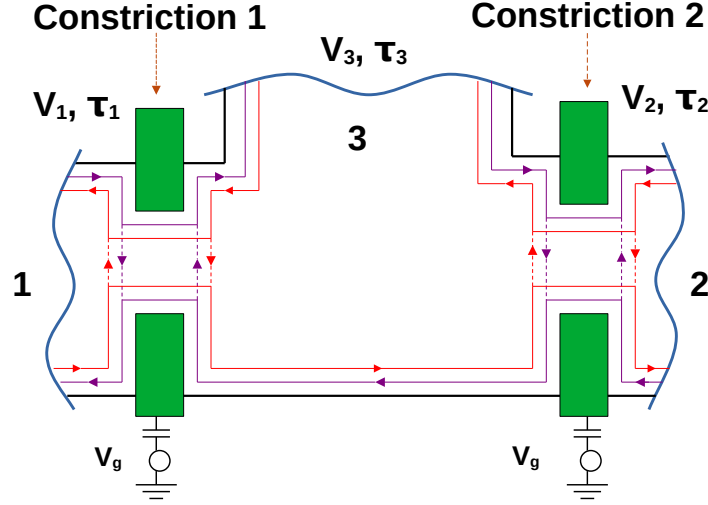
Figure 10: Three terminal QSH sample with helical edge modes and two constrictions. The red solid (dashed) line represents the edge mode for spin-down electrons, whereas the purple solid (dashed) line represents the edge mode for spin-up electrons, which can scatter from the constrictions. Terminal 3 is a voltage-temperature probe.

where,

$$
\begin{aligned}
L_{eV}^{\sigma} &= G_{11}^{\sigma} + \frac{G_{13}^{\sigma}(L_{33}\Pi_{31} - G_{31}K_{33})}{X} - \frac{L_{13}^{\sigma}(G_{33}\Pi_{31} - G_{31}\Pi_{33})}{X}, \\
L_{e\theta}^{\sigma} &= L_{11}^{\sigma} + \frac{G_{13}^{\sigma}(L_{33}K_{31} - L_{31}K_{33})}{X} - \frac{L_{13}^{\sigma}(G_{33}K_{31} - L_{31}\Pi_{33})}{X}, \\
L_{hV}^{\sigma} &= \Pi_{11}^{\sigma} + \frac{\Pi_{13}^{\sigma}(L_{33}\Pi_{31} - G_{31}K_{33})}{X} - \frac{K_{13}^{\sigma}(G_{33}\Pi_{31} - G_{31}\Pi_{33})}{X}, \\
L_{h\theta}^{\sigma} &= K_{11}^{\sigma} + \frac{\Pi_{13}^{\sigma}(L_{33}K_{31} - L_{31}K_{33})}{X} - \frac{K_{13}^{\sigma}(G_{33}K_{31} - L_{31}\Pi_{33})}{X}.
\end{aligned}
\tag{70}
$$

Eq. (70) are the Onsager coefficients for 3T QSH system with voltage-temperature probe. The total conductance G, Seebeck coefficient S, Peltier coefficient Π and thermal conductance K are given as,

$$
G = \sum_{\sigma\in\{\uparrow,\downarrow\}} G^{\sigma}, \quad S = \frac{1}{2}\sum_{\sigma\in\{\uparrow,\downarrow\}} S^{\sigma}, \quad \Pi = \frac{1}{2}\sum_{\sigma\in\{\uparrow,\downarrow\}} \Pi^{\sigma}, \quad K = \sum_{\sigma\in\{\uparrow,\downarrow\}} K^{\sigma}.
\tag{71}
$$

2. Power and efficiency in 3T QSH setup as a quantum heat engine

Now, the charge power generated in terminal 1 is given as,

$$
P = VI_1 = V\sum_{\sigma\in\{\uparrow,\downarrow\}} I_1^{\sigma} = V\sum_{\sigma\in\{\uparrow,\downarrow\}} (-L_{eV}^{\sigma}V + L_{e\theta}^{\sigma}\tau) = V(-L_{eV}V + L_{e\theta}\tau)
\tag{72}
$$

The maximum charge power can be obtained from $\frac{\partial P}{\partial V} = 0$, i.e., $V = \frac{L_{e\theta}\tau}{2L_{eV}}$. Thus maximum power is $P_{max} = \frac{L_{e\theta}^2\tau^2}{4L_{eV}}$. Similarly, the efficiency at maximum power is given as,

$$
\eta|_{P_{max}} = \frac{P_{max}}{J} = \frac{\theta\eta_c}{2}\frac{L_{e\theta}^2}{2L_{h\theta}L_{eV} - L_{hV}L_{e\theta}} = \frac{\eta_c}{2}\frac{Z\theta}{Z\theta + 2}.
\tag{73}
$$

Here, $Z\theta$ is the figure of merit, which is defined as

$$
Z\theta = \frac{L_{hV}L_{e\theta}}{L_{eV}L_{h\theta} - L_{e\theta}L_{hV}} = \frac{GS^2\theta}{K}.
\tag{74}
$$

For $Z\theta \to \infty$, $\eta|_{P_{max}} = \frac{\eta_c}{2}$. The maximum value of efficiency at maximum power ($\eta|_{P_{max}}$) one can attain is half of Carnot efficiency, which is defined as Curzon-Ahlborn efficiency [16].

Similarly, one can find maximum efficiency. The general expression for efficiency is given as

$$\eta = \frac{VI_1}{J} = \frac{V(-L_{eV}V + L_{e\theta}\tau)}{L_{hV}V + L_{h\theta}\tau}. \tag{75}$$

The voltage bias V_{max} is given as,

$$V_{max} = \frac{L_{h\theta}}{L_{hV}}\left(1 - \sqrt{\frac{L_{eV}L_{h\theta} - L_{e\theta}L_{hV}}{L_{eV}L_{h\theta}}}\right)\tau \tag{76}$$

which can be derived from the condition $\frac{d\eta}{dV} = 0$. Thus, the maximum efficiency is given as

$$\eta_{max} = \eta_c x \frac{\sqrt{Z\theta+1}-1}{\sqrt{Z\theta+1}+1}, \quad \text{where} \quad Z\theta = \frac{GS^2\theta}{K} = \quad \text{Figure of merit.} \tag{77}$$

where, $x = \frac{\theta L_{e\theta}}{L_{hV}} = \frac{\theta S}{\Pi}$ is the asymmetric parameter (AP) of the setup. For $x = 1$ and in the limit $Z\theta \to \infty$, the maximum efficiency approaches Carnot efficiency.

3. Cooling Power and coefficient of performance in 3T QSH setup as a quantum refrigerator

One can similarly derive the maximum coefficient of performance (η_{max}^r) and cooling power at maximum coefficient of performance ($J|_{\eta_{max}^r}$) for a QSH setup to act as a quantum refrigerator. Here, the heat will be absorbed from the cooler terminals (terminals 2 and 3 with voltage probe condition, and only terminal 2 with voltage-temperature probe) and dumped into the hotter terminal (terminal 1 in either voltage probe or voltage-temperature probe). For a quantum refrigerator, the coefficient of performance (η^r) is defined as the ratio of heat taken from the cooler terminal (J^Q) and the power absorbed (P) by the setup, i.e., $\eta^r = \frac{J^Q}{P}$. Here, $J^Q = -(J_2 + J_3)$ for voltage probe and $-J_2$ for voltage-temperature probe and can be derived from Eq. (63). Similarly, $P = I_1V$ can also be derived from Eq. (52). Now, η_{max}^r can be found using the condition $\frac{d\eta^r}{dV} = 0$, which gives

$$V|_{\eta_{max}^r} = \frac{L_{h\theta}}{L_{hV}}\left(1 + \sqrt{\frac{L_{eV}L_{h\theta} - L_{hV}L_{e\theta}}{L_{eV}L_{h\theta}}}\right) \tag{78}$$

where, $V|_{\eta_{max}^r}$ is the voltage required to achieve η_{max}^r. From here onwards one can derive η_{max}^r by using Eq. (72) in $\eta^r = \frac{J^Q}{P}$, which yields

$$\eta_{max}^r = \frac{\eta_c^r}{x}\frac{\sqrt{Z\theta+1}-1}{\sqrt{Z\theta+1}+1} \tag{79}$$

where η_c^r = Carnot COP = $\frac{\theta}{\Delta\theta} = \eta_c^{-1}$. Similarly, the cooling power at maximum coefficient of performance ($J|_{\eta_{max}^r}$) is given as,

$$J|_{\eta_{max}^r} = L_{hV}V_{max} + L_{hT}\tau = L_{hT}\left(\sqrt{\frac{L_{eV}L_{h\theta} - L_{hV}L_{e\theta}}{L_{eV}L_{h\theta}}}\right)\tau. \tag{80}$$

B. Transport in the Nonlinear response regime in the QSH setup

The charge and heat current in a QSH setup is given as

$$I_\alpha^\sigma = \frac{e}{h}\int_{-\infty}^{\infty}dE\sum_\beta f_\beta(E)[N_\alpha\delta_{\alpha\beta} - T_{\alpha\beta}^\sigma], \quad J_\alpha^\sigma = \frac{1}{h}\int_{-\infty}^{\infty}dE(E - \mu_\alpha)\sum_\beta f_\beta(E)[N_\alpha\delta_{\alpha\beta} - T_{\alpha\beta}^\sigma]. \tag{81}$$

Here too, the theory of nonlinear transport is similar as discussed in Sec. II B. In the subsequent subsections, we follow the same pattern as QH, where we first discuss 2T nonlinear QSH setup as a heat engine and refrigerator using Whitney's approach in Sec. III B 1, then do the same thing with our approach in 2T setup in Sec. III B 2 and in 3T setup without any probes in Sec. III B 3. Finally, we conclude this section by addressing the nonlinear transport property of QSH setup with both voltage and voltage-temperature probe in Sec. III B 4.
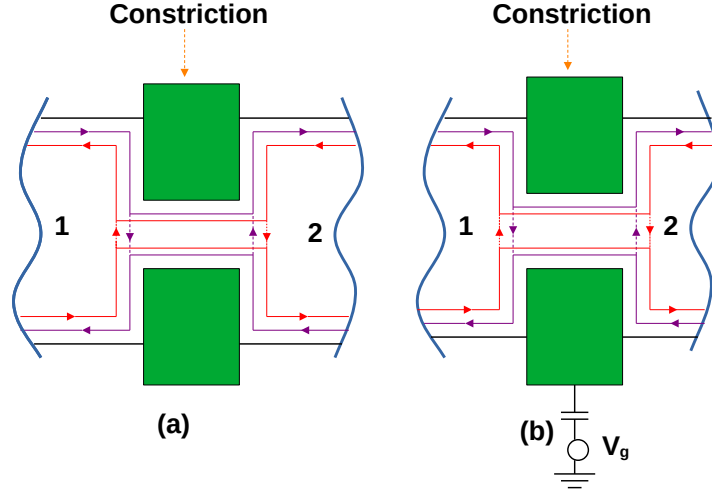
Figure 11: (a) Two-terminal QSH setup used to analyze Whitney's approach, wherein the constriction has a boxcar-type transmission, (b) Two-terminal QSH setup used in our approach, where the constriction has QPC-type transmission and is capacitively connected an external gate voltage.

1. Nonlinear thermoelectrics in 2T QSH setup (Whitney's approach)

Here, we discuss the nonlinear thermoelectricity of a 2T QSH setup as shown in Fig. 11(a) following the approach of Whitney [17, 18] both as a quantum heat engine as well as a quantum refrigerator. This analysis is novel as Whitney only used his approach for a 2T QH setup. Similar to the QH case, we consider $V_1 = -V, V_2 = 0$, and θ_1 and θ_2 are the temperatures of terminals 1 and 2 respectively. First, we discuss the quantum heat engine and later, we will discuss the quantum refrigerator.

Using the formula given in Eq. (81), the power generated (P) for a quantum heat engine is

$$P = V(I_1^\uparrow + I_1^\downarrow) = \frac{eV}{h} \int_{-\infty}^{\infty} dE(T_{12}^\uparrow + T_{12}^\downarrow)(f_1 - f_2), \tag{82}$$

where, $T_{12}^\uparrow(T_{12}^\downarrow)$ represents the transmission probability that a spin-up (spin-down) electron will scatter from terminal 2 to terminal 1. The Fermi-Dirac distributions for terminals 1 and 2, respectively, are f_1 and f_2. Similarly, the heat current out of terminal 1 using Eq. (81) is given as,

$$J_1 = J_1^\uparrow + J_1^\downarrow = \frac{1}{h} \int_{-\infty}^{\infty} dE(E+eV)(T_{12}^\uparrow + T_{12}^\downarrow)(f_1 - f_2) = \frac{1}{h} \int_{-\infty}^{\infty} dE(E)(T_{12}^\uparrow + T_{12}^\downarrow)(f_1 - f_2) + P, \tag{83}$$

wherein we have taken $\mu_1 = -eV$. Using Eqs. (82) and (83), one can evaluate the efficiency, i.e., $\eta = P/J_1$. For a heat engine, the power generated is always positive. From Eq. (82), one can see that the positive contribution to P comes, when $f_1 - f_2 > 0$. Now, from Fig. 12, one sees that $f_1 - f_2$ is positive above an energy value E_x, which is given as,

$$E_x = \frac{eV}{\left(\frac{\theta_1}{\theta_2} - 1\right)}, \quad \text{(Derived from the condition } f_1 = f_2\text{)}. \tag{84}$$

As proposed by Whitney [17, 18], one can maximize the power generated by a heat engine by blocking the transmission of the electrons with energy E below E_x, which means only the electrons above energy E_x will contribute to the power P. Now, by comparing Eqs. (81) and (82), one can see that for higher energy values i.e., for $E > E_x$, the rate of increase of J_1 is higher than that of P as J_1 is always more than P, See Eq. (82). This means that high-energy electrons contribute to power (P) less efficiently than it does for the heat current (J_1). So, one can guess to have a transmission probability in the energy range $E_x < E < E_x'$, where E_x' is the voltage-dependent energy just like E_x, above which the transmission is zero. Here the analysis is exactly similar to the QH case as in Sec. II B 1 as the helical edge modes is a composition of two chiral edge modes each for spin-up and spin-down electrons, then both $T_{12}^\uparrow$ and $T_{12}^\downarrow$ are boxcar type as shown in Fig. in some energy range $E_x < E < E_x'$ to have the best possible
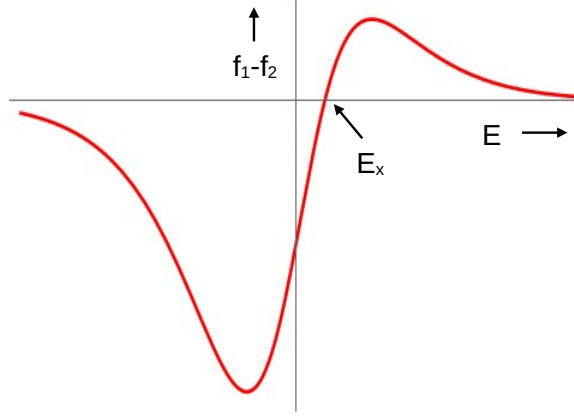
Figure 12: Plot of $f_1 - f_2$ in the 2T QSH setup.

performance as a quantum heat engine, where $E_x' = eVJ_1'/P'$. The expression for E_x' is derived in [18]. Considering single edge mode for spin-up electron, $T_{12}^{\uparrow}$ will be 1 for $E_x < E < E_x'$, otherwise zero. Similarly, in the same energy range $E_X < E < E_x'$, $T_{12}^{\downarrow}$ will be 1, otherwise zero. At some voltage bias, when $E_x' \to \infty$, then both $T_{12}^{\uparrow}$ as well as $T_{12}^{\downarrow}$ will become QPC-like transmission as shown in Fig. 1. Here, $E_x' = eVJ_1'/P'$ [17, 18], where the prime in J_1 and P denotes the first derivatives of J_1 and P with respect to the applied voltage bias V. Now, the maximum possible power (P_{max}) is generated, when $P' = 0$, which makes $E_x' = \infty$, and consequently one can derive the efficiency at the maximum power ($\eta|_{P_{max}}$). The maximum power is given as

$$P_{max} = \frac{2e}{h} \int_{E_x}^{\infty} dE(f_1 - f_2), \tag{85}$$

and the heat current from the terminal 1, when $P' = 0$ is

$$J|_{P_{max}} = \frac{2e}{h} \int_{E_x}^{\infty} dE(f_1 - f_2). \tag{86}$$

One can derive the efficiency at maximum power by using the formula $\eta|_{P_{max}} = P_{max}/J|_{P_{max}}$. Following the Ref. [18], one can similarly derive the expressions for the maximum power (P_{max}) and efficiency at maximum power ($\eta|_{P_{max}}$) and they are given as

$$P_{max} = P_{max}^{Wh} = 0.0642 \frac{\pi^2 k_B^2 \tau^2}{h}, \quad \eta|_{P_{max}} = \eta|_{P_{max}^{Wh}} = \frac{\eta_c}{1 + 0.936(1 + \theta_2/\theta_1)}. \tag{87}$$

Similarly, when E_x and E_x' are close to each other the finite power generated will be low, and then the expression for maximum efficiency η_{max} can be derived for generic values of E_x and E_x'. Now, following Whitney's result [17, 18], η_{max} is given as

$$\eta_{max} = \eta_{max}^{Wh} = \eta_c \left(1 - 0.478 \sqrt{\frac{\theta_2 P}{\theta_1 P_{max}^{Wh}}} \right), \tag{88}$$

where, $\tau = \theta_1 - \theta_2$ = temperature bias applied across the sample, η_c = Carnot efficiency, θ_1 = temperature in terminal 1, θ_2 = temperature in terminal 2, P = Finite output power with $P \ll P_{max}^{Wh}$.

Similarly, for a quantum refrigerator, the cooling power ($J = J_2$) of this 2T QSH setup is given as

$$J^Q = J_2 = (J_2^{\uparrow} + J_2^{\downarrow}) = \frac{1}{h} \int_{-\infty}^{\infty} dEE(T_{21}^{\uparrow} + T_{21}^{\downarrow})(f_2 - f_1). \tag{89}$$

From Eq. (89), the positive contribution to the cooling power J comes when $(f_2 - f_1) > 0$, means when $f_1 - f_2 < 0$. Now, from Fig. 12, it is clear that, $f_1 - f_2$ is negative below the energy value E_x, which is given in Eq. (83). According to the argument of Whitney [17, 18], one needs to maximize the cooling power by allowing the transmission of electrons only below energy E_x and blocking above them. Similarly, one can also look into maximizing the coefficient of performance $\eta^r = J/P_1$, where $P = I_1/V$ is

the power delivered on the system, which is given in Eq. (76). Now, one can guess a transmission function in the energy range $E_{x2} < E < E_x$, where E_{x2} is also a voltage-dependent energy, above which the transmission is zero. From now onwards, the analysis is similar to the QH case as in Sec. II B 2 as the helical edge mode is a composition of two chiral edge modes each for spin-up and spin-down electrons. Then, $T_{21}^{\uparrow}$ and $T_{21}^{\downarrow}$ are boxcar-type transmission, which is 1 in the energy range $E_{x2} < E < E_x$, where $E_{x2} = eVJ_2'/P'$, as derived by Whitney [18]. So, the cooling power J is given as

$$J = J_2 = \frac{2}{h} \int_{E_{x2}}^{E_x} dE E(f_2 - f_1), \tag{90}$$

where the cooling power is maximum when $J_2' = 0$, which makes the lower limit of integration $E_{x2} = 0$. Now, to have a large cooling power, the upper limit of the integration E_x needs to be large, which means $E_x \to \infty$. Now, from Eq. (84), it is clear that $E_x \to \infty$, when $V \to \infty$, which will make $f_1 \to 0$. Then the maximum cooling power (J_{max}^{Wh}) following Whitney's approach is given as

$$J_{max}^{Wh} = \frac{2}{h} \int_0^{\infty} dE E(f_2) = \frac{\pi^2}{6h} k_B^2 \theta_2^2 \tag{91}$$

Similarly, the power delivered on the system ($P = I_1 V$) is large, when the maximum cooling power (J_{max}^{Wh}) is reached. So, the coefficient of performance at J_{max}^{Wh} is J_{max}^{Wh}/P is zero as P is very large. Similarly, one achieves the maximum coefficient of performance when the finite cooling power is much smaller than J_{max}^{Wh}, which can be achieved when E_x and E_{x2} are close. Then one can calculate the cooling power and power delivered at generic values of E_x and E_{x2}, which are quite close as done by Whitney [17, 18]. In that case, the coefficient of performance is large, which is given as

$$\eta_{max}^r = \eta_{max}^{r,Wh} = \eta_c^r \left(1 - 1.09 \sqrt{\frac{\theta_2}{\theta_1 - \theta_2} \frac{J}{J_{max}^{Wh}}} \right). \tag{92}$$

where $J \ll J_{max}^{Wh}$.

One can conclude that the best possible performance of a 2T QSH setup is exactly similar to that of 2T QH setup as done by Whitney.

2. Nonlinear thermoelectrics in 2T QSH setup (Our approach)

Here, we discuss nonlinear thermoelectrics for a 2T QSH setup by considering a QPC-like constriction in between, See Fig. 11(b), where we have applied a gate voltage (V_g) to the constriction. Here, the gate voltage controls the energy level of QPC i.e., $E_1' = E_1 + eU$, where U is the interaction potential. For this case, we consider $V_1 = -V, V_2 = 0, \tau_1 = \tau$ and $\tau_2 = 0$. Here, also we follow the same argument used for the 2T QH setup as in Sec. II B 2. This can be understood in the following way.

The general form of U (interaction potential) in any multiterminal QSH setup is [31] can be written as,

$$U = \sum_{\sigma=\uparrow,\downarrow} U_\sigma = U_{eq} + \sum_{\alpha,\sigma} u_{\alpha\sigma} V_\alpha + \sum_{\alpha,\sigma} z_{\alpha\sigma} \tau_\alpha, \tag{93}$$

where U_{eq} is the equilibrium potential. $u_{\alpha\sigma} = \partial U_\sigma / \partial V_\alpha$ and $z_{\alpha\sigma} = \partial U_\sigma / \partial \tau_\alpha$ are spin-dependent characteristic potentials, which determine the response of potential U to a voltage bias and temperature bias respectively [31]. The characteristic potentials $u_{\alpha\sigma}$ and $z_{\alpha\sigma}$ can be derived from particle injectivity $v_\alpha^p(E,\sigma)$ and entropic injectivity $v_\alpha^e(E,\sigma)$ [31]. The expression for $v_\alpha^p(E,\sigma)$ and $v_\alpha^e(E,\sigma)$ are given as,

$$v_\alpha^p(E,\sigma) = \frac{1}{2\pi i} \sum_\beta \text{Tr}\left[s_{\beta\alpha}^\dagger \frac{ds_{\beta\alpha}}{dE} \right], \quad v_\alpha^e(E,\sigma) = \frac{1}{2\pi i} \sum_\beta \text{Tr}\left[\frac{E}{\theta} s_{\beta\alpha}^\dagger \frac{ds_{\beta\alpha}}{dE} \right]. \tag{94}$$

where, $s_{\beta\alpha}$ represents the scattering amplitude that an electron will scatter from terminal α to terminal β.

When the gate voltage is capacitively applied to the constriction, see Fig. 11(b). Therefore, there will be an additional contribution to U due to the gate voltage as derived in Refs. [46–48] and the modified U is,

$$U' = U + u_g V_g = \sum_{\alpha,\sigma} u_{\alpha\sigma} V_\alpha + \sum_\alpha z_{\alpha\sigma} \tau_\alpha + \sum_\sigma u_{g\sigma} V_g, \tag{95}$$

where, $u_{g\sigma}$ is the characteristic potential, which determine the response of U_σ to the external gate voltage V_g.

For a two-terminal QSH setup with QPC-like constriction, the general form of the interaction potential U' using Eq. (95) is given as

$$U' = \sum_{\sigma=\uparrow,\downarrow} u_{1\sigma}V_1 + \sum_{\sigma=\uparrow,\downarrow} u_{2\sigma}V_2 + \sum_{\sigma=\uparrow,\downarrow} z_{1\sigma}\tau_1 + \sum_{\sigma=\uparrow,\downarrow} z_{2\sigma}\tau_2 + \sum_{\sigma=\uparrow,\downarrow} u_{g\sigma}V_g. = u_1V_1 + u_2V_2 + z_1\tau_1 + z_2\tau_2 + u_gV_g, \qquad (96)$$

where, $u_1 = \sum_{\sigma=\uparrow,\downarrow} u_{1\sigma}, u_2 = \sum_{\sigma=\uparrow,\downarrow} u_{2\sigma}, z_1 = \sum_{\sigma=\uparrow,\downarrow} z_{1\sigma}, z_2 = \sum_{\sigma=\uparrow,\downarrow} z_{2\sigma}, u_g = \sum_{\sigma=\uparrow,\downarrow} u_{g\sigma}$ are the characteristic potentials. Similar to the 2T QH case as discussed in Sec. II B 2, these characteristic potentials will be constant if the capacitance C and width ω of the QPC are assumed to be constant.

In the presence of interaction potential U, the single energy level of QPC: E'_{QPC} is modified to $E'_{QPC} + eU$. In the presence of gate potential it is further modified as $E''_{QPC} = E'_{QPC} + U'$. In our setup, we consider $V_1 = -V, V_2 = 0, \tau_1 = \tau, \tau_2 = 0$ with QPC-like constriction. In this case, the interaction potential U using Eq. (96) is

$$U = u_1V_1 + z_1\tau_1 = -u_1V + z_1\tau, \qquad (97)$$

and, when the constriction is capacitively connected to the gate voltage V_g, the interaction potential U is modified as (See Eq. (96))

$$U' = -u_1V + z_1\tau + u_gV_g = U + u_gV_g, \qquad (98)$$

where, the other characteristic potentials u_2 and z_2 are ignored as there are no applied biases in terminal 2. The form of u_1, z_1 and u_g is same as in Eq. (52). As already discussed, the quantities such as u_1, u_2, and u_g are constant when C and ω are constant. For this setup, it is assumed that capacitance C is constant, while the width of the QPC ω is also constant, which implies that u_1, z_1 and u_g will also be constant throughout the work. This can be done experimentally as well. If the setup is kept at a constant voltage bias V and temperature bias τ, then the interaction potential U will also be constant. At the same time, one can always apply V_g to nullify U as discussed in Ref. [27] depending upon the value of V and τ. Therefore, at a fixed voltage bias V and τ, for the quantum heat engine, we consider the gate voltage to be

$$V_g = ((u_1V - z_1\tau) + V'_g)/u_g = (-U + V'_g)/u_g, \qquad (99)$$

The first term inside the parenthesis of V_g in Eq. (99) is used to nullify the interaction potential $U = -(u_1V - z_1\tau)$ and the second term V'_g is used to lift the energy level of QPC, which means

$$E''_{QPC} = E'_{QPC} + eU + eu_gV_g = E'_{QPC} + eV'_g. \qquad (100)$$

Now, the transmission probability of the QPC becomes

$$T_{QPC} = \frac{1}{1 + e^{-2\pi(E - E''_{QPC})/\hbar\omega}} = \frac{1}{1 + e^{-2\pi(E - E'_{QPC} - eV'_g)/\hbar\omega}}, \qquad (101)$$

and this form is always maintained at any voltage bias V and V'_g. With this, the power delivered ($P = I_1V$) and efficiency ($\eta = P/J_1$) can be found using the formulae given in Eqs. (81). Now, in this work, we fix the values of V and τ and evaluate the power and efficiency with different values of V'_g. We plot them parametrically in Fig. 3(a). Now, we fix $V = 0.14k_B\theta/e$, $\tau = 1K$, we find that at a particular V_g, the maximum power approaches P^{Wh}_{max} as shown in Fig. 9(a).

Similarly, for the refrigerator, again we consider $V_g = -U + V'_g$. We now calculate the cooling power $J = J_2$ and coefficient of performance ($\eta^r = J/P$), where $P = I_1V$ for different values of V. The parametric plot of the cooling power J and η^r is shown in Fig. 9(b). The maximum cooling power approaches J^{Wh}_{max} and the best coefficient of performance reach is 86% of η^r_c. Here, we conclude that one can achieve the best thermoelectric performance as a refrigerator in our two-terminal setup with maximum possible cooling power approaching J^{Wh}_{max} and 87% of η^r_c.

3. Nonlinear thermoelectrics in 3T QSH setup (Our approach)

Here, we discuss the power and efficiency of a simple three-terminal QSH setup without any voltage probe or voltage-temperature probes with two QPC-like constrictions as shown in Fig. 8 by controlling the transmission of QPC 1 and QPC 2 by an applied gate voltage using the conditions: $E'_1 = E_1 + eV_g$ and $E'_2 = E_2 + eV_g$, with $E_1 = E_2$. We consider a simple case with $V_1 = -V, V_2 = V_3 = 0$, and $\tau_1 = \tau, \tau_2 = \tau_3 = 0$. Similar to the 2T QSH case as in Sec. III B 2, the interaction potential in the presence of a gate voltage of the same form as Eq. (98). Therefore, the interaction potential can be nullified by considering the gate voltage V_g of the same form as in Eq. (99). For the heat engine, the power delivered ($P = I_1V$) and efficiency ($\eta = P/J_1$) can be found by using the formula given in Eqs. (81) for different values of voltage bias (V), the performance exactly matches with the result of two-terminal heat engine as shown in Fig. 9(a). Similarly, for the refrigerator, we take $V_g = -U + V'_g$ and cooling power ($J = (J_2 + J_3)$) and coefficient of performance ($\eta^r = J/P$) exactly matches the result of the two-terminal refrigerator as in Fig. 9(b).
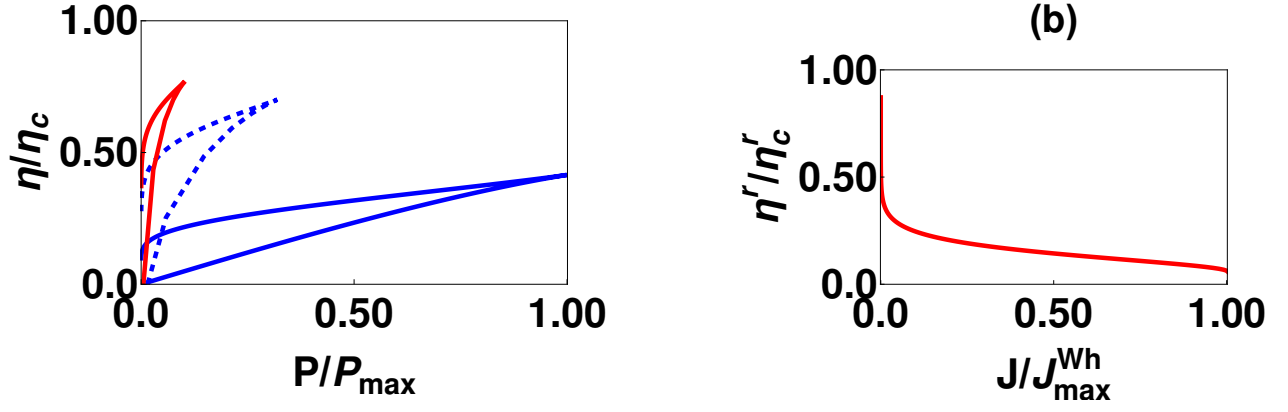
Figure 13: (a) Parametric plot of the power (P) and efficiency as a heat engine. The blue thick line corresponds to $V = 0.57 k_B \theta / e$, blue dotted line corresponds to $V = 1.77 k_B \theta / e$, red thick line corresponds to $V = 2.52 k_B \theta / e$, (b) Parametric plot of the cooling power (J) and coefficient of performance (η^r) as a refrigerator of a two-terminal QSH (helical) setup at $V = 20 k_B \theta / e$. Here, we have taken $E'_{QPC} = k_B \theta$, where $\theta_1 = 2\theta_2 = 2K$ and $\tau = \theta_1 - \theta_2 = 1K$. For the QPC, we have taken $\omega = 0.1 k_B \theta / e$. These results are similar to the three-terminal QH setup without any probe with $V_1 = -V, V_2 = V_3 = 0$ and $\tau_1 = \tau, \tau_2 = \tau_3 = 0$.

4. Nonlinear thermoelectrics in 3T with a voltage-temperature probe (Our approach)

Next, we extend our approach to a 3T nonlinear QSH setup as shown in Fig. 4 with a voltage-temperature probe by controlling the transmission of QPC 1 and QPC 2 via a gate voltage. We consider the threshold of energies of QPC's $E'_1 = E_1 + eV_g$ and $E'_2 = E_2 + eV_g$ above which the transmission is 1 with $E_1 = E_2$. We consider terminal 3 to be a voltage-temperature probe and calculate V_3 and τ_3 from the voltage-temperature condition $I_3 = J_3 = 0$, where $V_1 = -V, V_2 = 0, \tau_1 = \tau, \tau_2 = 0$. This is hard to solve analytically, therefore we do it numerically using Mathematica. The Mathematica code is uploaded to GitHub [33]. Here, in the presence of gate voltage V_g, the interaction potential U' is given as

$$U' = -u_1 V + z_1 \tau + u_3 V_3 + z_3 \tau_3 + u_g V_g = U + u_g V_g, \tag{102}$$

where, $U = -u_1 V + z_1 \tau + u_3 V_3 + z_3 \tau_3$ is the interaction potential in the absence of the gate voltage V_g and $u_1 = \sum_{\sigma=\uparrow,\downarrow} u_{1\sigma}, u_2 = \sum_{\sigma=\uparrow,\downarrow} u_{2\sigma}, u_3 = \sum_{\sigma=\uparrow,\downarrow} u_{3\sigma}, z_1 = \sum_{\sigma=\uparrow,\downarrow} z_{1\sigma}, z_2 = \sum_{\sigma=\uparrow,\downarrow} z_{2\sigma}, z_3 = \sum_{\sigma=\uparrow,\downarrow} z_{3\sigma} u_g = \sum_{\sigma=\uparrow,\downarrow} u_{g\sigma}$. As discussed in Sec. III B 2, the characteristic potentials are constant if one assumes the capacitance C and the width of the QPC ω are constant. The solutions V_3 and τ_3 will also be constants if the voltage bias V and temperature bias τ are kept constant. Now, one can consider the gate voltage V_g to be

$$V_g = (-U + V'_g)/u_g, \tag{103}$$

Which will nullify the interaction potential U and the energy level of the QPC becomes $E''_{QPC} = E_{QPC} + eV'_g$. Considering this, one can calculate power and efficiency as a quantum heat engine by changing V'_g at fixed values of V, τ.

Similar to the quantum heat engine, we investigate the three-terminal QSH setup as a quantum refrigerator with both voltage probe and voltage-temperature probe by our approach and considering identical values of E'_1 and E'_2 same as used for quantum heat engine. The coefficient of performance $\eta^r = \frac{J}{P}$ and the cooling power J is $(J_2 + J_3)$ are evaluated with either probe and the parametric plot are shown in Fig. 4(b) of the main text. With these conditions, in the three-terminal QH setup for a voltage-temperature probe, we observe that J_{max} reaches J_{max}^{Wh}, while η^r reaches about 86% of η_c^r as shown in Fig. 13(b).

Like the heat engine, we investigate the refrigeration property of the three-terminal QSH setup with the voltage-temperature probe by considering the same values of E'_1 and E'_2 as in the heat engine. The coefficient of performance $\eta^r = \frac{J}{P}$ and the cooling power J is $(J_2 + J_3)$ can also be evaluated with both types of probes, and the parametric plot is shown in Fig. 4(b) in the main text. Here, too, we consider $V_g = (-U + V'_g)/u_g$. With these considerations, in the three-terminal QSH setup with a voltage-temperature probe, we observe that J_{max} reaches $0.75 J_{max}^{Wh}$ in the case of voltage-temperature probe, while η^r reaches about 86% of η_c^r as shown in Fig. 4(b) of the main text. Here, one can also distinguish between chiral and helical edge mode transport as a refrigerator at its best possible performance. For the chiral edge mode with the voltage-temperature probe, the maximum cooling power approaches J_{max}^{Wh}, whereas for helical edge mode transport, it is well below J_{max}^{Wh}.