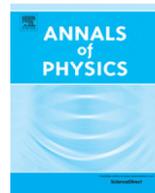




ELSEVIER

Contents lists available at ScienceDirect

## Annals of Physics

journal homepage: [www.elsevier.com/locate/aop](http://www.elsevier.com/locate/aop)

# Energy relaxation rate and its mesoscopic fluctuations in quantum dots

Vladyslav A. Kozii<sup>a,\*</sup>, Mikhail A. Skvortsov<sup>b,c,d</sup>

<sup>a</sup> Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>b</sup> Skolkovo Institute of Science and Technology, Skolkovo 143026, Russia

<sup>c</sup> L. D. Landau Institute for Theoretical Physics, Chernogolovka 142432, Russia

<sup>d</sup> Moscow Institute of Physics and Technology, Dolgoprudny 141700, Russia

## ARTICLE INFO

### Article history:

Received 31 March 2016

Accepted 5 April 2016

Available online 13 April 2016

### Keywords:

Quantum dot

Electron–electron interaction

Many-body localization

Mesoscopic fluctuations

Non-equilibrium system

## ABSTRACT

We analyze the applicability of the Fermi-golden-rule description of quasiparticle relaxation in a closed diffusive quantum dot with electron–electron interaction. Assuming that single-particle levels are already resolved but the initial stage of quasiparticle disintegration can still be described by a simple exponential decay, we calculate the average inelastic energy relaxation rate of single-particle excitations and its mesoscopic fluctuations. The smallness of mesoscopic fluctuations can then be used as a criterion for the validity of the Fermi-golden-rule description. Technically, we implement the real-space Keldysh diagram technique, handling correlations in the quasi-discrete spectrum non-perturbatively by means of the non-linear supersymmetric sigma model. The unitary symmetry class is considered for simplicity. Our approach is complementary to the lattice-model analysis of Fock space: though we are not able to describe many-body localization, we derive the exact lowest-order expression for mesoscopic fluctuations of the relaxation rate, making no assumptions on the matrix elements of the interaction. It is shown that for the quasiparticle with the energy  $\varepsilon$  on top of the thermal state with the temperature  $T$ , fluctuations of its energy width become large and the Fermi-golden-rule description breaks down at  $\max\{\varepsilon, T\} \sim \Delta\sqrt{g}$ , where  $\Delta$  is the mean level spacing in the quantum dot, and  $g$  is its dimensionless conductance.

© 2016 Elsevier Inc. All rights reserved.

\* Corresponding author.

E-mail address: [koziiva@mit.edu](mailto:koziiva@mit.edu) (V.A. Kozii).

## 1. Introduction

Understanding the structure of wave functions in Fock space is a hot topic in many-body physics, intimately related to the concepts of ergodicity and thermalization in complex interacting systems. The natural way to think of the interaction is as a cause of transitions between single-particle states, which are no longer eigenstates of a quantum system in the presence of interaction. Single-particle excitations decay into three-particle states (two electrons and one hole), which then further decay into five-particle states, etc. The simplest way to characterize the process of quantum-mechanical spreading of an excitation in Fock space is to study its energy relaxation rate (inverse lifetime).

The study of the inelastic relaxation in a confined geometry has been pioneered by Sivan, Imry and Aronov (SIA) [1], who calculated the quasiparticle lifetime in a diffusive quantum dot with chaotic electron dynamics. Working within the conventional Fermi golden rule (FGR) picture, they calculated the relaxation rate of an excitation with the energy  $\varepsilon$ , induced by the screened Coulomb interaction with the small momentum transfer:

$$\gamma_0(\varepsilon) \sim \lambda^2 \Delta(\varepsilon/E_{\text{Th}})^2, \quad (1)$$

where  $\Delta$  is the mean single-particle level spacing in the dot, and  $E_{\text{Th}} = D/L^2$  is the Thouless energy determined by the inverse time of electron diffusion across the system ( $D$  is the diffusion coefficient, and  $L$  is the typical size of the quantum dot). For generality, in Eq. (1) we introduced the dimensionless interaction strength  $\lambda$ , taking its maximal value,  $\lambda = 1$ , for the screened Coulomb potential. The result (1) was derived in the hot-electron regime (negligible temperature,  $T \ll \varepsilon$ ) under the assumption of the zero-dimensional geometry,  $\varepsilon \ll E_{\text{Th}}$ . It provides the diffusive contribution to the relaxation rate originating from the processes with momentum transfer of the order of the inverse system size,  $1/L$ . The total relaxation rate is then the sum of  $\gamma_0(\varepsilon)$  given by Eq. (1) and the Fermi-liquid contribution due to the processes with large-momentum transfer,  $\gamma_*(\varepsilon) \sim \varepsilon^2/E_F$  ( $E_F$  is the Fermi energy) [2]. The latter can be neglected for sufficiently large quantum dots,  $L \gg k_F l^2$  ( $k_F$  is the Fermi momentum, and  $l$  is the mean free path) [1,3], that will be assumed thereafter.

Remarkably, Eq. (1) derived in the zero-dimensional limit,  $\varepsilon \ll E_{\text{Th}}$ , shows that in this regime the single-particle spectrum is well resolved,  $\gamma_0(\varepsilon) \ll \Delta$ . Though this conclusion is in a good agreement with the experimental results on tunneling spectroscopy of a disordered quantum dot [4], it raises the question of consistency of the derivation. Indeed, the FGR can be safely applied only in the case of continuous spectrum, while the result (1) implies that it is not. This subtle point was recognized already by SIA, who argued that their approach might be correct as summation over many final states is performed. This idea was elaborated in a seminal paper by Altshuler, Gefen, Kamenev, and Levitov (AGKL) [5], who emphasized that it is the spectrum of final states that should be continuous for the FGR picture to be applicable. In the problem of the hot-electron decay, final states are three-particle states with two electrons and one hole, and the corresponding mean level spacing can be estimated as

$$\Delta_3(\varepsilon) \sim \Delta^3/\varepsilon^2. \quad (2)$$

Comparing  $\gamma_0(\varepsilon)$  and  $\Delta_3(\varepsilon)$ , AGKL came to the conclusion that the FGR description should be valid for sufficiently large energies,  $\varepsilon > \varepsilon_{\text{FGR}}$ , where

$$\varepsilon_{\text{FGR}} = \Delta\sqrt{g}, \quad (3)$$

and  $g = E_{\text{Th}}/\lambda\Delta \gg 1$  sets the scale,  $\Delta/g$ , of the interaction matrix elements (for the screened Coulomb interaction with  $\lambda = 1$ ,  $g$  coincides with the dimensionless conductance of the dot).

In order to study the spreading of a single-particle excitation over the space of many-particle states beyond the FGR approximation, AGKL proposed a remarkable mapping of the initial problem to a tight-binding Anderson model on a hierarchical lattice, treating each site as a basis vector in the many-particle Fock space. Motivated by the observation that the structure of this lattice for the quantum dot problem locally looks like a tree, AGKL approximated it by the Bethe lattice with a large branching number. Then using the known solution for the Anderson model on the Bethe lattice [6], AGKL predicted the localization transition in Fock space of an interacting quantum dot at the energy

$\varepsilon_{\text{MBL}} \sim \Delta \sqrt{g/\ln g}$ , which appeared to be parametrically smaller than the FGR breaking energy scale  $\varepsilon_{\text{FGR}}$ . Thus, in the AGKL picture of many-body localization one should distinguish between three regimes [5,7]. In the localized regime realized at  $\varepsilon < \varepsilon_{\text{MBL}}$ , the Slater determinants of one-particle states are very close to the exact many-body states, and the width of quasiparticle states is exactly zero. In the intermediate region,  $\varepsilon_{\text{MBL}} < \varepsilon < \varepsilon_{\text{FGR}}$ , quasiparticle states are delocalized, but they are strongly fractal and non-ergodic, with the spectral weight given by a number of slightly broadened lines. The spectral weight acquires a Lorentzian form with a well defined width  $\gamma$  (corresponding to a simple exponential decay  $e^{-\gamma t}$  in the time domain) only in the regime  $\varepsilon \gg \varepsilon_{\text{FGR}}$ , where the FGR description finally sets in.

The AGKL paper has triggered a boost of activity in the field of many-body localization. The suggested mapping onto the lattice model has been proved to be extremely fruitful: instead of studying an interacting problem one now can deal with a non-interacting quantum mechanics, but on a very complicated lattice with an exponentially large number of sites. (Though all states in a finite system are localized by definition, even an extremely weak coupling to the external reservoir providing a level width larger than the exponentially small distance between the many-body states, which can be estimated as  $\exp(-\alpha\sqrt{\varepsilon}/\Delta)$  with  $\alpha \sim 1$  [8–10], renders the spectrum of delocalized states effectively continuous.) Since the complexity of the problem is encoded in the lattice topology, the crucial point is to identify the relevant features responsible for many-body localization.

Quite soon it was recognized that the Bethe lattice with a constant branching number is an oversimplified model of Fock space. Firstly, the coordination number of the lattice decreases with the number of generations [8,11]. Secondly, the actual lattice is not a tree, and the presence of loops essentially modifies combinatorics of the perturbative expansion in the localized region [8], increasing the AGKL estimate for  $\varepsilon_{\text{MBL}}$ . Numerical studies [12,13] demonstrated gradual delocalization with the growth of the quasiparticle energy. Anyway, despite the lack of a rigorous theory of many-body localization in a quantum dot, the general understanding achieved in the beginning of 2000s was that in finite systems one should expect a localization/delocalization crossover, though its position was not firmly established.

Later, the concept of many-body localization developed in the quantum dot problem was applied to extended systems of interacting electrons with spatially localized single-particle states [14,15], where the transition between the localized and delocalized many-body phases occurs at a finite temperature. The related issues of ergodicity and thermalization are currently being actively investigated (for a review, see Ref. [16]).

Very recently, many-body localization in a quantum dot was reconsidered by Gornyi, Mirlin, and Polyakov [17]. Working in the framework of the lattice model, they found an additional factorial contribution in the perturbative series in the number of involved generations, which renders coupling to distant generations less efficient, thus acting in favor of localization. For the hot-electron decay problem, this leads to the estimate for the threshold energy  $\varepsilon_{\text{MBL}} \sim g\Delta/\ln g$ , which is much larger than the original AGKL estimate. For a thermal many-body state with the temperature  $T$  and the total energy  $\varepsilon \sim T^2/\Delta$ , the FGR temperature  $T_{\text{FGR}} \sim \Delta\sqrt{g}$  was shown to be logarithmically larger than the localization transition point,  $T_{\text{MBL}} \sim \Delta\sqrt{g/\ln g}$ . This result formally coincides with the AGKL energy threshold and is in full agreement with Ref. [14].

Since, contrary to AGKL, Refs. [8,17] place the many-body localization threshold  $\varepsilon_{\text{MBL}}$  for the hot-electron problem above the FGR-breaking scale  $\varepsilon_{\text{FGR}}$ , one can ask how to reconcile the FGR description with many-body localization. This question was answered by Silvestrov [18] who studied the temporal decay of a quasiparticle with  $\varepsilon \gg \varepsilon_{\text{FGR}}$ . He showed that the FGR relaxation rate  $\gamma_0(\varepsilon)$  describes the initial stage of exponential relaxation, that slows down at larger times due to smaller relaxation rate of descendant states, and eventually due to many-body localization.

In this paper, we address the initial temporal stage of energy relaxation in a quantum dot, when quantum many-body localization effects are not yet visible. Assuming the FGR approach is applicable, we calculate mesoscopic fluctuations of the energy relaxation rate. The smallness of fluctuations compared to the average relaxation rate provides an *a posteriori* condition for the FGR applicability. In our analysis, we do not use the lattice model of Fock space and work in terms of Keldysh diagram technique in real space, where non-perturbative disorder averaging is performed by means of the non-linear supersymmetric sigma model. Such an approach allows us to derive the

expression for the variance of the relaxation rate without any simplifications concerning the nature of the electron–electron interaction.

The paper is organized as follows. In Section 2 we introduce the model and summarize the results. Section 3 describes the Keldysh kinetic approach along with its modification simplifying the study of weak nonequilibrium. In Section 4 we rederive the SIA result and generalize it to the case of an arbitrary temperature and interaction radius. Section 5 is devoted to the discussion of the general strategy for the calculation of mesoscopic fluctuations of the energy relaxation rate in terms of the exact (disorder-dependent) electron Green functions. The product of several Green functions is averaged non-perturbatively in Section 6. The final step of the calculation of mesoscopic fluctuations is performed, for not very short-range interaction, in Section 7 and, in the general case, in Section 8. Our results are summarized in Section 9. Numerous technical details are relegated to several Appendices.

We use the system of units with  $\hbar = k_B = 1$ .

## 2. Model description and results

### 2.1. Model

We consider an isolated chaotic quantum dot with the broken time-reversal symmetry (unitary symmetry class). Impurity scattering in the dot is supposed to be strong enough to completely randomize the electron trajectories establishing the diffusive regime with  $l \ll L$ , where  $l$  is the elastic mean free path, and  $L$  is the characteristic size of the dot. For generality, we consider the case of an arbitrary spin degeneracy,  $N_s$ , which plays the role of the number of independent fermionic flavors (the case  $N_s = 1$  corresponds to completely spin-polarized electrons).

We will discuss two models of electron–electron interactions: (i) the long-range Coulomb interaction with a small gas parameter  $r_s \equiv e^2/\epsilon v_F \ll 1$  (where  $\epsilon$  is the dielectric constant of the medium, and  $v_F$  is the Fermi velocity), and (ii) an arbitrary weak interaction. In the case of the Coulomb interaction, its screening should be taken into account, whereas for a weak interaction this procedure is unnecessary. In both cases, the statically screened interaction in the real space can be written in the form

$$V(\mathbf{r}) = \frac{\lambda}{N_s v} \delta_\kappa(\mathbf{r}), \quad (4)$$

where  $\lambda$  is the dimensionless interaction strength, and  $\delta_\kappa(\mathbf{r})$  is the delta function smeared on the scale of the interaction radius  $1/\kappa$  [the latter is assumed to be much smaller than the size of the quantum dot, justifying the notion of the smeared delta function in Eq. (4)]. For the long-range Coulomb interaction,  $\lambda = 1$  and  $1/\kappa$  is the Thomas–Fermi screening length, see Eq. (31). Besides the dimensionless strength  $\lambda$ , the only characteristic of the potential (4) that will be important in the following is the ratio,  $F$ , of the Hartree and Fock diagrams, which is a measure of the interaction range [19,20]. In the three- and two-dimensional cases it is given by (see Appendix A for details)

$$F = \int_0^1 v_\kappa(2p_F x) \varphi(x) dx, \quad \varphi(x) = \begin{cases} 2x, & \text{in 3D,} \\ (2/\pi)(1-x^2)^{-1/2}, & \text{in 2D,} \end{cases} \quad (5)$$

where  $v_\kappa(\mathbf{q})$  is the Fourier transform of  $\delta_\kappa(\mathbf{r})$  in Eq. (4). The limiting cases of  $F = 0$  and  $F = 1$  correspond to long-range ( $\kappa \ll k_F$ ) and point-like screened interactions, respectively.

We will assume that interaction can be taken into account perturbatively, that can be justified in the two partially overlapping limits:

$$\lambda \ll 1 \quad \text{or} \quad F \ll 1, \quad (6)$$

i.e., in the limit of weak interaction or for sufficiently large interaction radius ( $\kappa \ll k_F$ ). The two models of electron–electron interaction discussed above conform to the condition (6): For the Coulomb interaction with  $\lambda = 1$ , its applicability is provided by the smallness of the gas parameter  $r_s$  since  $F \sim r_s \ln 1/r_s$  [see Eq. (A.6)]; for weak interaction with an arbitrary radius  $1/\kappa$ , it is justified as long as  $\lambda \ll 1$ .

## 2.2. Average energy relaxation rate

We start with generalizing the SIA result (1) to the case of a finite temperature  $T$ , arbitrary spin degeneracy  $N_s$  and arbitrary interaction radius measured by the parameter  $F$  [assuming that the restriction (6) holds]. The temperature and excitation energy are supposed to be smaller than the Thouless energy,  $\max\{\varepsilon, T\} \ll E_{\text{Th}}$ . The average energy relaxation rate calculated in the second order in the statically screened interaction (4) is given by

$$\gamma_0(\varepsilon, T) = \frac{\lambda^2 c(N_s, F)}{\pi} \frac{\Delta}{E_2^2} (\varepsilon^2 + \pi^2 T^2), \quad (7)$$

where  $\Delta = (\nu V)^{-1}$  is the mean level spacing in the dot ( $\nu$  is the single-particle density of states at the Fermi level per one spin projection, and  $V$  is volume of the dot), and the energy scale  $E_2$  is determined by the diffusion inside the dot:

$$E_n \equiv \left[ \sum'_m \frac{1}{(Dq_m^2)^n} \right]^{-1/n} \sim E_{\text{Th}}. \quad (8)$$

Here  $D$  is the diffusion coefficient, and  $q_m^2$  are non-zero eigenvalues of the Laplace operator in the dot,  $-\nabla^2 \psi_m(\mathbf{r}) = q_m^2 \psi_m(\mathbf{r})$ , with the von Neumann boundary conditions. The magnitude of  $E_n$  defined by Eq. (8) is set by the Thouless energy,  $E_{\text{Th}} = D/L^2$ , where  $L$  is the typical size of the quantum dot.

The factor  $c(N_s, F)$  in Eq. (7) takes into account spin degeneracy and the spatial structure of the interaction potential (4):

$$c(N_s, F) = \frac{1}{N_s^2} (N_s - 2F + N_s F^2). \quad (9)$$

The function  $c(N_s, F)$  has an important property  $c(1, 1) = 0$ , which ensures vanishing of interaction effects for polarized ( $N_s = 1$ ) fermions with a contact interaction, as a consequence of the Pauli exclusion principle. It should be noted that the anticipated cancellation of  $c(1, 1)$  takes place only if the Hartree-type diagrams yielding the terms with  $F$  in Eq. (9) are taken into account (see Section 4.3). This issue is often overlooked in literature, leading to a variety of claimed prefactors in Eq. (1).

## 2.3. Mesoscopic fluctuations of the energy relaxation rate

In the FGR description, the relaxation rate  $\gamma_i$  of a given single-particle state  $|i\rangle$  is a function of its energy only, implying that the states close in energy should have nearly the same inelastic width. The validity of this approximation can be verified *a posteriori* by studying mesoscopic, i. e., level-to-level fluctuations of  $\gamma_i$ . We calculate them under the same assumptions of the zero-dimensional geometry,  $\max\{\varepsilon, T\} \ll E_{\text{Th}}$ , used for the determination of the average  $\gamma_0(\varepsilon, T)$ .

We describe the relaxation rate in the framework of the quantum kinetic equation for electrons in dirty metals [21]. This approach differs significantly from the lattice model proposed by AGKL [5] and extensively used in subsequent publications [7,8,11,12,17]. The method of kinetic equation is not suitable for studying many-body localization, but is sufficiently simple to unveil the limits of applicability of the FGR description. Nevertheless, the kinetic approach, working well for interacting systems with continuous spectrum, meets significant difficulties in the considered region  $\max\{\varepsilon, T\} \ll E_{\text{Th}}$ , where individual energy levels are well-resolved,  $\gamma_0(\varepsilon, T) \ll \Delta$ . Discreteness of energy spectrum requires non-perturbative averaging over disorder, which is performed by means of the non-linear supersymmetric sigma model [22].

The main idea behind our calculations is the following. In the zeroth approximation (which is equivalent to the FGR), energy levels acquire an inelastic width  $\gamma_0(\varepsilon, T)$ . The inverse of this quantity yields the temporal scale at which single-particle coherence is maintained. This corresponds to the appearance of the ‘mass’ of the order  $\gamma_0(\varepsilon, T)$  for the zero-dimensional diffusons (diffusons with zero momentum). At the next stage, when loop corrections to the FGR result are considered, this ‘mass’ will

regularize the otherwise divergent contributions, producing  $\gamma_0(\varepsilon, T)$  in the denominator, precisely in the way it occurs in the case of the non-interacting quantum dot in an external field [23–25] and in the high-temperature phase in the problem of many-body localization [14]. Thus, with decreasing the energy of excitations and/or temperature, the loop corrections to the quasiclassical rate increase, and at some energy scale they become comparable, indicating the breakdown of the FGR description.

Fluctuations of the energy relaxation rate  $\gamma(\varepsilon, T)$  around its FGR average value  $\gamma_0(\varepsilon, T)$  given by Eq. (7) are characterized by the irreducible average:

$$\langle \gamma^2(\varepsilon, T) \rangle = \gamma_0^2(\varepsilon, T) + \langle \langle \gamma^2(\varepsilon, T) \rangle \rangle. \quad (10)$$

We calculate the leading contribution to  $\langle \langle \gamma^2(\varepsilon, T) \rangle \rangle$  in the delocalized regime and obtain

$$\begin{aligned} \langle \langle \gamma^2(\varepsilon, T) \rangle \rangle &= \frac{\lambda^4 \Delta^5}{4\pi^3} \left[ \frac{c_2(N_s, F)}{E_2^4} + \frac{c_4(N_s, F)}{E_4^4} \right] \\ &\times \int d\varepsilon_1 d\omega_1 \frac{(\mathcal{F}_{\varepsilon_1} - \mathcal{F}_{\varepsilon_1 - \omega_1})^2 (\mathcal{B}_{\omega_1} + \mathcal{F}_{\varepsilon - \omega_1})^2}{\gamma_0(\varepsilon - \omega_1, T) + \gamma_0(\varepsilon_1, T) + \gamma_0(\varepsilon_1 - \omega_1, T)}, \end{aligned} \quad (11)$$

where

$$\mathcal{F}_\varepsilon = \tanh \frac{\varepsilon}{2T}, \quad \mathcal{B}_\omega = \coth \frac{\omega}{2T} \quad (12)$$

are equilibrium fermionic and bosonic distribution functions, respectively, while the energies  $E_2$  and  $E_4$  (both of the order of  $E_{\text{Th}}$ ) are defined in Eq. (8). The functions  $c_2(N_s, F)$  and  $c_4(N_s, F)$  in Eq. (11) are given by

$$c_2(N_s, F) = \frac{1}{N_s^4} \left[ (3N_s^2 + 1) - 16N_s F + 2(5N_s^2 + 7)F^2 - 16N_s F^3 + (3N_s^2 + 1)F^4 \right], \quad (13a)$$

$$c_4(N_s, F) = \frac{1}{N_s^4} \left[ 2N_s^2 - 8N_s F + 4(N_s^2 + 2)F^2 - 8N_s F^3 + 2N_s^2 F^4 \right]. \quad (13b)$$

They obey an important relation  $c_2(1, 1) = c_4(1, 1) = 0$ , which guarantees vanishing of mesoscopic fluctuations [as well as the inelastic width itself, see Eq. (9)] for spin-polarized fermions with a contact interaction.

To single out the energy and temperature dependence of Eq. (11), we rewrite it in the form

$$\langle \langle \gamma^2(\varepsilon, T) \rangle \rangle = \frac{\lambda^2 \Delta^4}{4\pi^2} \frac{E_2^2}{c(N_s, F)} \left[ \frac{c_2(N_s, F)}{E_2^4} + \frac{c_4(N_s, F)}{E_4^4} \right] \Upsilon \left( \frac{\varepsilon}{2T} \right), \quad (14)$$

where the dimensionless function  $\Upsilon(x)$  plotted in Fig. 1 is defined as

$$\Upsilon(x) = \int dy dz \frac{[\tanh(y) - \tanh(y - z)]^2 [\coth(z) - \tanh(z - x)]^2}{(x - z)^2 + y^2 + (y - z)^2 + 3\pi^2/4} = \begin{cases} 0.248, & x = 0; \\ 16.95, & x = \infty. \end{cases} \quad (15)$$

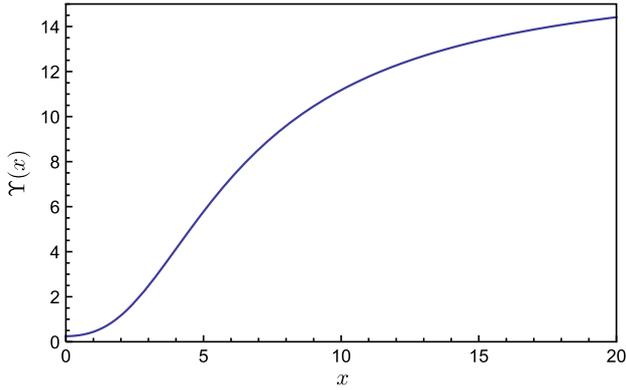
Hence, the magnitude of mesoscopic fluctuations of the relaxation rate can be roughly estimated as

$$\langle \langle \gamma^2(\varepsilon, T) \rangle \rangle \sim \lambda^2 \frac{\Delta^4}{E_{\text{Th}}^2}. \quad (16)$$

Note however that, due to a huge variation of the function  $\Upsilon(x)$ , fluctuations at the Fermi energy ( $\varepsilon \ll T$ ) are nearly 100 times smaller than the naive estimate (16).

Expression (11) for mesoscopic fluctuations of the energy relaxation rate in the FGR regime is the main result of our work. The relative strength of mesoscopic fluctuations can be estimated as

$$\frac{\langle \langle \gamma^2(\varepsilon, T) \rangle \rangle}{\gamma_0^2(\varepsilon, T)} \sim \frac{\Delta_3(\max\{\varepsilon, T\})}{\gamma_0(\varepsilon, T)} \sim \left( \frac{\varepsilon_{\text{FGR}}}{\max\{\varepsilon, T\}} \right)^4, \quad (17)$$



**Fig. 1.** The function  $\Upsilon(x)$  [Eq. (15)], which determines the energy and temperature dependence of  $\langle\langle\gamma^2(\varepsilon, T)\rangle\rangle$ , see Eq. (14).

where  $\Delta_3(\varepsilon)$  is the mean three-particle level spacing defined in Eq. (2). The ratio (17) becomes of the order of unity as  $\max\{\varepsilon, T\}$  approaches the FGR-breaking scale  $\varepsilon_{\text{FGR}} = \Delta\sqrt{g}$ . Hence for the validity of the FGR description of the initial stage of quasiparticle disintegration, the temperature  $T$  of electrons in the dot and the excitation energy  $\varepsilon$  at which the decay rate is studied play nearly the same role. They will act differently at larger time scales, when many-body localization may have enough time to show up [17,18]. Physically, at  $\max\{\varepsilon, T\} \gg \varepsilon_{\text{FGR}}$  the state has many available routes to decay into three-particle excitations. Fluctuations of the number of such routes are small and are determined by Eq. (17).

### 3. General expression for the energy relaxation rate

#### 3.1. Keldysh technique

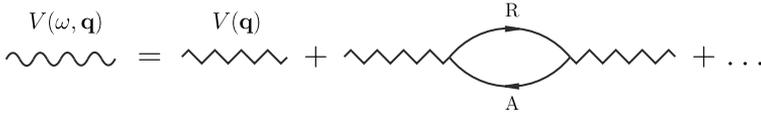
In order to express the inelastic energy relaxation rate, we use the Keldysh technique [26] in the representation of the functional integral [27,28]. For an interacting system, the Keldysh action is a functional of the fermionic Grassmann fields  $\psi_\sigma(\mathbf{r}, t)$  ( $\sigma = 1, \dots, N_s$  counts spin projections) and the bosonic plasmon field  $\phi(\mathbf{r}, t)$ :

$$S = \int dt \left\{ \sum_{\sigma=1}^{N_s} \int d\mathbf{r} \psi_\sigma^+(\mathbf{r}, t) \left( \hat{G}_0^{-1} + \phi_a(\mathbf{r}, t)\gamma^a \right) \psi_\sigma(\mathbf{r}, t) + \int (d\mathbf{q}) V_0^{-1}(\mathbf{q}) \phi^T(\mathbf{q}, t) \gamma_2 \phi(-\mathbf{q}, t) \right\}. \quad (18)$$

The fields  $\psi_\sigma(\mathbf{r}, t)$  and  $\phi(\mathbf{r}, t)$  are two-component vectors in the Keldysh space. In the Keldysh-rotated basis [27,28],

$$\hat{G}_0^{-1} = \left( i \frac{\partial}{\partial t} + \frac{\nabla^2}{2m} - U_{\text{dis}}(\mathbf{r}) \right) \gamma_1 \equiv G_0^{-1} \gamma_1, \quad \gamma^1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \gamma^2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (19)$$

where  $U_{\text{dis}}(\mathbf{r})$  is the disorder potential with the correlation function  $\langle U_{\text{dis}}(\mathbf{r}) U_{\text{dis}}(\mathbf{r}') \rangle = \delta(\mathbf{r} - \mathbf{r}') / (2\pi \nu \tau)$  ( $\tau$  is the elastic mean free time, and  $\nu$  is the density of states per one spin projection at the Fermi energy). We assume no spin-dependent interactions so that all matrices act as a unit matrix in the spin space. The second term in Eq. (18) is the action of the plasmon field, with  $V_0(\mathbf{q})$  being the bare interaction potential, and  $(d\mathbf{q}) \equiv d^d q / (2\pi)^d$  is the momentum integration measure in  $d$  dimensions.



**Fig. 2.** The dynamically screened interaction  $V(\omega, \mathbf{q})$  (wavy line) expressed in terms of the statically screened interaction (SSI)  $V(\mathbf{q})$  (zigzag line) and the dynamic polarization operator  $G^R G^A$ . For a quantum-dot at  $(\epsilon, T) \ll E_{\text{Th}}$ , each dynamic polarization bubble contains a small factor of  $\omega/E_{\text{Th}}$ .

The Green functions are defined as ( $\xi_i$  denotes the pair  $\mathbf{r}_i, t_i$ ):

$$\hat{G}(\xi_1, \xi_2) = -i\langle \psi(\xi_1) \psi^+(\xi_2) \rangle = -i \int D\psi^* D\psi D\phi e^{iS} \psi(\xi_1) \psi^+(\xi_2), \tag{20}$$

$$\hat{V}(\xi_1, \xi_2) = -2i\langle \phi(\xi_1) \phi^T(\xi_2) \rangle = -2i \int D\psi^* D\psi D\phi e^{iS} \phi(\xi_1) \phi^T(\xi_2). \tag{21}$$

The electron Green function has the triangular structure in the Keldysh space:

$$\hat{G} = \begin{pmatrix} G^R & G^K \\ 0 & G^A \end{pmatrix}, \quad G^K = G^R \circ \mathcal{F} - \mathcal{F} \circ G^A, \tag{22}$$

where  $\mathcal{F}$  is the fermion distribution function, and the symbol “ $\circ$ ” denotes the convolution over intermediate spatial and time indices. At the equilibrium with the temperature  $T$ ,  $\mathcal{F}_\epsilon = \tanh(\epsilon/2T)$ .

### 3.2. Interaction propagator

In the derivation below we will mainly assume that electrons in the dot interact via the Coulomb potential  $V_0(\mathbf{r}) = e^2/\epsilon r$ , where  $\epsilon$  is the dielectric constant of the medium. Due to the long-range nature of the Coulomb interaction,  $V_0^{-1}(\mathbf{q} \rightarrow 0) = 0$ , the propagator  $\hat{V}$  should be calculated taking screening into account (this procedure is not required for an arbitrary weak and not long-range potential). As usual, we do this in the random phase approximation (RPA) [29] justified in the case of a small gas parameter:

$$r_s = \frac{e^2}{\epsilon v_F} \ll 1. \tag{23}$$

Under this condition the ratio of the Hartree and Fock diagrams,  $F$ , introduced in Eq. (5) is small too, see Eq. (A.6). Then one should sum only the bubble contributions to the effective propagator, with independent averaging over disorder in each bubble.

The RPA-screened interaction propagator has the standard structure:

$$\hat{V} = \begin{pmatrix} V^K & V^R \\ V^A & 0 \end{pmatrix}, \quad V^K = V^R \circ \mathcal{B} - \mathcal{B} \circ V^A, \tag{24}$$

where  $\mathcal{B}$  is the boson distribution function defined as [27]

$$\mathcal{B}_\omega = \frac{1}{2\omega} \int_{-\infty}^{\infty} (1 - \mathcal{F}_{\epsilon'} \mathcal{F}_{\epsilon'-\omega}) d\epsilon'. \tag{25}$$

At the thermal equilibrium,  $\mathcal{B}_\omega = \coth(\omega/2T)$ .

The dynamically screened interaction can be written in the form (see Fig. 2)

$$V^{R(A)}(\omega, \mathbf{q}) = [V^{-1}(\mathbf{q}) + \Pi_{\text{dyn}}^{R(A)}(\omega, \mathbf{q})]^{-1}, \tag{26}$$

where the real function  $V(\mathbf{q})$  is the statically screened interaction (SSI):

$$V(\mathbf{q}) = [V_0^{-1}(\mathbf{q}) + N_s v]^{-1}, \tag{27}$$

and  $\Pi_{\text{dyn}}$  is the dynamic part of the polarization operator originating from the bubble  $G^R G^A$ :

$$\Pi_{\text{dyn}}^{\text{R(A)}}(\omega, \mathbf{q}) = N_s v \frac{\pm i\omega}{Dq^2 \mp i\omega}. \quad (28)$$

The simple form of the polarization operator  $\Pi(\mathbf{q}, \omega)$  in Eqs. (27) and (28) corresponds to the limit  $q \ll p_F$  (and  $\omega \ll E_F$ ). In the case of the Coulomb interaction with a small  $r_s$ , the spatial dispersion of  $\Pi(\mathbf{q}, 0)$  is irrelevant as the interaction in Eq. (5) is determined by  $q \sim \kappa \ll p_F$ . In the case of a weak short-scale interaction,  $\lambda \ll 1$ , static screening can be neglected and the form of  $\Pi(\mathbf{q}, 0)$  is not important.

The SSI defined in Eq. (27) can be conveniently rewritten in the form [Fourier transform of Eq. (4)]

$$V(\mathbf{q}) = \frac{\lambda}{N_s v} v_\kappa(\mathbf{q}), \quad (29)$$

where  $v_\kappa(\mathbf{q})$  is the momentum representation of the  $\delta$  function smeared over the interaction radius,  $\delta_\kappa(\mathbf{r})$ . By definition,  $v_\kappa(0) = 1$ . In the important case of the Coulomb interaction,  $\lambda^{\text{Coul}} = 1$  and

$$v_\kappa^{\text{Coul}}(\mathbf{q}) = \begin{cases} \kappa_{3D}^2 / (q^2 + \kappa_{3D}^2), & \text{in 3D,} \\ \kappa_{2D}^2 / (q + \kappa_{2D}), & \text{in 2D,} \end{cases} \quad (30)$$

where  $1/\kappa$  is the Thomas–Fermi screening length given by

$$\kappa_{3D}^2 = 4\pi N_s v_3 e^2 / \epsilon, \quad \kappa_{2D} = 2\pi N_s v_2 e^2 / \epsilon. \quad (31)$$

### 3.3. Kinetic equation

In this Section we sketch the main steps in the derivation of the quantum kinetic equation in the Keldysh formalism (for a detailed discussion see, e.g., Refs. [21,30]). From the Dyson equation for the exact (disorder-dependent) electron Green function  $\hat{G}$  we obtain

$$[\hat{G}_0^{-1}, \hat{G}] = [\hat{\Sigma}, \hat{G}], \quad (32)$$

where  $[A, B] = A \circ B - B \circ A$ , and  $\hat{\Sigma}$  is the irreducible self-energy due to interaction, bearing the same triangular structure as the electron Green function:

$$\hat{\Sigma} = \begin{pmatrix} \Sigma^{\text{R}} & \Sigma^{\text{K}} \\ 0 & \Sigma^{\text{A}} \end{pmatrix}. \quad (33)$$

As we are interested in the slow dynamics of the system, it is convenient to switch to the mixed energy–time (Wigner) representation,

$$f_\varepsilon(t) = \int d\tau f\left(t + \frac{\tau}{2}, t - \frac{\tau}{2}\right) e^{i\varepsilon\tau}, \quad (34)$$

which allows us to get rid of the convolution on the right-hand side of Eq. (32). The kinetic equation is obtained from the Keldysh component of Eq. (32) by tracing over the space. We also assume that the distribution function does not depend on the coordinates. Therefore the left-hand side of Eq. (32) produces only the time derivatives of  $\mathcal{F}$ , and we arrive at the standard form of the kinetic equation,

$$\partial_t \mathcal{F}_\varepsilon = \text{St}[\mathcal{F}_\varepsilon]. \quad (35)$$

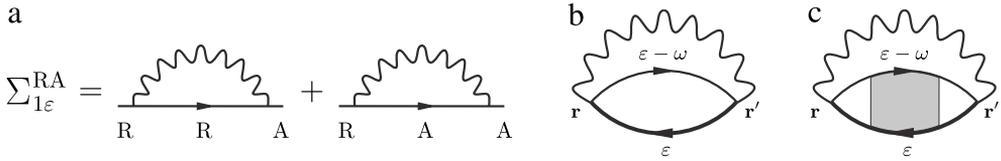
The collision integral  $\text{St}[\mathcal{F}_\varepsilon]$  is given by

$$\text{St}[\mathcal{F}_\varepsilon] \int d\mathbf{r} \Delta G_\varepsilon(\mathbf{r}, \mathbf{r}) = -i \int d\mathbf{r} d\mathbf{r}' \{ \Delta \Sigma_\varepsilon(\mathbf{r}, \mathbf{r}') G_\varepsilon^{\text{K}}(\mathbf{r}', \mathbf{r}) - \Sigma_\varepsilon^{\text{K}}(\mathbf{r}, \mathbf{r}') \Delta G_\varepsilon(\mathbf{r}', \mathbf{r}) \}, \quad (36)$$

where all functions have an implicit central-time argument  $t$ , and we denote

$$\Delta G = G^{\text{R}} - G^{\text{A}}, \quad \Delta \Sigma = \Sigma^{\text{R}} - \Sigma^{\text{A}}. \quad (37)$$





**Fig. 3.** (a) The simplest (and the most important for small  $F$ ) contribution to the self-energy  $\Sigma_{\epsilon}^{RA}$ . (b) The graphic representation of the product  $\Sigma_{\epsilon}^{RA} \Delta G_{\epsilon}$  which determines the collision integral (46). The thick line standing for  $\Delta G_{\epsilon} = G_{\epsilon}^R - G_{\epsilon}^A$  corresponds to the electron state whose decay is studied. (c) The diagram (b) averaged over disorder, with the shaded block representing the diffuson.

Eq. (45) is the starting point for the calculation of the energy relaxation rate. This equation contains exact disorder-dependent Green functions, and averaging its  $n$ th power over the random potential generates the  $n$ th moment of  $\gamma(\epsilon, T)$ . The simplest is the first moment,  $\langle \gamma(\epsilon, T) \rangle$ , related to the average collision integral  $\langle \text{St}[\mathcal{F}_{\epsilon}] \rangle$ . In this case, the left-hand side of Eq. (43) gives the average density of states,  $\langle \Delta G_{\epsilon}(\mathbf{r}, \mathbf{r}) \rangle = -2\pi i\nu$ , and one arrives at

$$\langle \text{St}[\mathcal{F}_{\epsilon}] \rangle = \frac{\Delta}{2\pi} \int d\mathbf{r} d\mathbf{r}' \langle \Sigma_{\epsilon}^{RA}(\mathbf{r}, \mathbf{r}') \Delta G_{\epsilon}(\mathbf{r}', \mathbf{r}) \rangle. \tag{46}$$

In the next Section we show how this equation reproduces Sivan, Imry and Aronov result (7) for the average inelastic rate in the limit  $F \rightarrow 0$ , and generalize it to the case of an arbitrary parameter  $F$ . The second moment of  $\gamma(\epsilon, T)$  will be considered in Sections 5–8.

#### 4. Average energy relaxation rate

In this Section we rederive the result of Sivan, Imry and Aronov [1] and generalize it to the case of an arbitrary temperature  $T$ , spin degeneracy  $N_s$  and interaction radius characterized by the parameter  $F$ . Our treatment closely follows the standard derivation of the kinetic equation in the RPA approximation [21], but within the modified Keldysh technique introduced in Section 3.4. The purpose of this Section is to illustrate the usage of this technique and to prepare the ingredients for the analysis of mesoscopic fluctuations of  $\gamma(\epsilon, T)$  in Section 5.

##### 4.1. Derivation of the Sivan, Imry and Aronov result

The simplest diagram for the self-energy  $\Sigma_{\epsilon}^{RA}$  is shown in Fig. 3(a). In the limit  $F \ll 1$  it gives the leading contribution to the relaxation rate (a more general situation is discussed in Section 4.3). The corresponding analytic expression can be easily written in terms of the elements  $Y$  introduced in Eq. (41):

$$\Sigma_{1\epsilon}^{RA}(\mathbf{r}, \mathbf{r}') = \frac{i}{2} \int (d\omega) \{ G_{\epsilon-\omega}^R(\mathbf{r}, \mathbf{r}') Y_{RR,RA}(\epsilon, \epsilon, \omega, \mathbf{r} - \mathbf{r}') + G_{\epsilon-\omega}^A(\mathbf{r}, \mathbf{r}') Y_{RA,AA}(\epsilon, \epsilon, \omega, \mathbf{r} - \mathbf{r}') \}, \tag{47}$$

with  $(d\omega) \equiv d\omega/2\pi$ . The interaction propagators are assumed to be RPA screened and averaged over disorder (see Section 3.1), while the electron Green functions are still exact in a given realization of disorder. Using Eqs. (B.2) and (B.8), and employing the analyticity properties we obtain

$$\Sigma_{1\epsilon}^{RA}(\mathbf{r}, \mathbf{r}') = \frac{i}{2} \int (d\omega) \Phi(\epsilon, \omega) \Delta G_{\epsilon-\omega}(\mathbf{r}, \mathbf{r}') \Delta V(\omega, \mathbf{r} - \mathbf{r}'), \tag{48}$$

where  $\Delta V = V^R - V^A$ , and

$$\Phi(\epsilon, \omega) = (\mathcal{F}_{\epsilon} - \mathcal{F}_{\epsilon-\omega}) \mathcal{B}_{\omega} - (1 - \mathcal{F}_{\epsilon} \mathcal{F}_{\epsilon-\omega}) \tag{49}$$

is a combination of the fermionic and bosonic distribution functions which vanishes at the equilibrium (detailed balance).

The collision integral (46) involves the product  $\Sigma_{\varepsilon}^{\text{RA}} \Delta G_{\varepsilon}$ , which is to be averaged over disorder. To make this calculation more intuitive, we represent it diagrammatically on a separate graph in Fig. 3(b), where the thick line stands for the Green function  $\Delta G_{\varepsilon}$  of the initial state. To avoid confusion we emphasize that this picture is just a simple graphical notation for  $\Sigma_{\varepsilon}^{\text{RA}} \Delta G_{\varepsilon}$ , since the vertices with a thick line are not described by the rule of Eq. (41). Substituting the self-energy (48) into Eq. (46) we obtain for the averaged collision integral:

$$\langle \text{St}[\mathcal{F}_{\varepsilon}] \rangle = \frac{i\Delta}{4\pi} \int d\mathbf{r} d\mathbf{r}' (d\omega) \Phi(\varepsilon, \omega) \langle \Delta G_{\varepsilon}(\mathbf{r}', \mathbf{r}) \Delta G_{\varepsilon-\omega}(\mathbf{r}, \mathbf{r}') \rangle \Delta V(\omega, \mathbf{r} - \mathbf{r}'). \quad (50)$$

The averaged product of two Green functions,

$$\langle \Delta G_{\varepsilon}(\mathbf{r}', \mathbf{r}) \Delta G_{\varepsilon-\omega}(\mathbf{r}, \mathbf{r}') \rangle = -4\pi \nu \text{Re} D_0^{\text{R}}(\omega, \mathbf{r} - \mathbf{r}') \quad (51)$$

is expressed in terms of the particle–hole ladder (diffuson),  $D_0^{\text{R}}(\omega, \mathbf{q}) = 1/(Dq^2 - i\omega)$ , see Fig. 3(c). Then one arrives at the well-known result for the collision integral in dirty metals (see, for example, Refs. [21,33,34]):

$$\langle \text{St}[\mathcal{F}_{\varepsilon}] \rangle = 2 \int (d\mathbf{q}) (d\omega) \Phi(\varepsilon, \omega) \text{Re} D_0^{\text{R}}(\omega, \mathbf{q}) \text{Im} V^{\text{R}}(\omega, \mathbf{q}), \quad (52)$$

where  $(d\mathbf{q}) \equiv d^d q / (2\pi)^d$ , and  $d$  is the space dimensionality.

According to Eq. (26), the imaginary part of the fluctuation propagator is determined by the dynamic polarization operator  $\Pi_{\text{dyn}}$ . For a quantum dot in the zero-dimensional regime ( $\omega \ll E_{\text{Th}}$ ),  $\Pi_{\text{dyn}}$  is a small correction to  $V^{-1}(\mathbf{q})$ , and  $\text{Im} V^{\text{R}}(\omega, \mathbf{q})$  can be written as

$$\text{Im} V^{\text{R}}(\omega, \mathbf{q}) \approx -V^2(\mathbf{q}) \text{Im} \Pi_{\text{dyn}}^{\text{R}}(\omega, \mathbf{q}). \quad (53)$$

Using Eq. (28) and assuming that the system size,  $L$ , is larger than the interaction radius,  $1/\kappa$ , we obtain in the zero-dimensional regime:

$$\langle \text{St}[\mathcal{F}_{\varepsilon}] \rangle = -\frac{2\lambda^2}{N_s \nu} \int (d\omega) (d\mathbf{q}) \frac{\omega \Phi(\varepsilon, \omega)}{(Dq^2)^2} = -\frac{2\lambda^2 \Delta}{N_s} \sum'_m \int (d\omega) \frac{\omega \Phi(\varepsilon, \omega)}{(Dq_m^2)^2}, \quad (54)$$

where  $q_m^2$  are the eigenvalues of the operator  $-\nabla^2$  in the dot with von Neumann boundary conditions, with the zero mode being excluded due to electroneutrality [1,3,5,35,36]. Performing summation over discrete momenta, we get

$$\langle \text{St}[\mathcal{F}_{\varepsilon}] \rangle = -\frac{2\lambda^2 \Delta}{N_s E_2^2} \int (d\omega) \omega \Phi(\varepsilon, \omega), \quad (55)$$

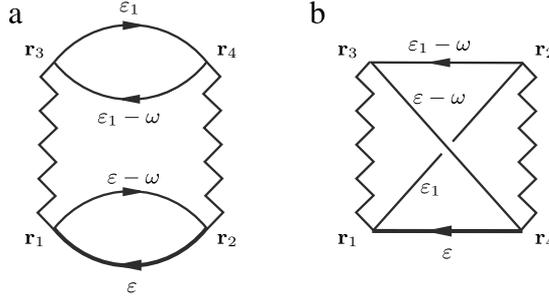
where the energy  $E_2 \sim E_{\text{Th}}$  is defined in Eq. (8). The inelastic energy relaxation rate at the equilibrium can be extracted with the help of Eq. (44):

$$\gamma_0^{\text{RPA}}(\varepsilon, T) = \frac{2\lambda^2 \Delta}{N_s E_2^2} \int (d\omega) \omega \left\{ \coth\left(\frac{\omega}{2T}\right) + \tanh\left(\frac{\varepsilon - \omega}{2T}\right) \right\}. \quad (56)$$

Integrating over  $\omega$ , we arrive at Eq. (7) for the average energy relaxation rate  $\gamma_0(\varepsilon, T)$  in the limit  $F \rightarrow 0$ .

#### 4.2. Physical interpretation and the other diagram

It is instructive to rederive the result for  $\gamma_0^{\text{RPA}}(\varepsilon, T)$  in a slightly different manner to clarify the physics of the decay process responsible for the width of an one-electron level. Real decay processes are described by  $\text{Im} V^{\text{R}}(\omega, \mathbf{q})$  [see Eq. (52)], which is determined by the dynamic screening of the bare interaction. In the zero-dimensional limit,  $(E, T, \omega) \ll E_{\text{Th}}$ , the dynamic part of the polarization bubble,  $\Pi_{\text{dyn}}(\omega, \mathbf{q})$ , contains a small factor  $\omega/E_{\text{Th}}$  [see Eq. (28)]. This means that with the accuracy of



**Fig. 4.** The diagrams for the collision integral with two SSI lines prior to disorder averaging. (a) The RPA diagram 3(b) redrawn in terms of the SSI lines (29) and the dynamic polarization bubble  $G^R G^A$ . The corresponding contribution to the relaxation rate is given by the  $X_4$  term in Eq. (59). (b) The other, non-RPA diagram discarded in Section 4.1, leading to the  $Y_4$  term in Eq. (59). Its contribution to the mean relaxation rate can be neglected at small  $F$  [see Eq. (63)].

order  $\omega/E_{\text{Th}}$  we may consider  $\Pi_{\text{dyn}}(\omega, \mathbf{q})$  as a small perturbation and retain only the first dynamic bubble in the expansion of  $V(\omega, \mathbf{q})$  near the statically screened interaction (SSI)  $V(\mathbf{q})$ , see Eq. (53).

Therefore with the accuracy of order  $\omega/E_{\text{Th}}$  the diagram in Fig. 3(a) for the average collision integral with one dynamically screened interaction can be redrawn as the diagram in Fig. 4(a) with two SSI lines. The advantage of the diagrammatic representation of Fig. 4(a) is that it elucidates the physics of the inelastic collision: The cross-section of this diagram corresponds to the decay of an electron with energy  $\varepsilon$  into an electron with energy  $\varepsilon - \omega$  and an electron-hole pair with energies  $\varepsilon_1$  and  $\varepsilon_1 - \omega$ . The corresponding contribution to the self-energy (before disorder averaging) can be easily read off from Eq. (41):

$$\begin{aligned} \Sigma_a^{\text{RA}} &= -N_s \left( \frac{i}{2} \right)^2 \sum_{abc} G_{\varepsilon-\omega}^a(\mathbf{r}_1, \mathbf{r}_2) G_{\varepsilon_1-\omega}^b(\mathbf{r}_4, \mathbf{r}_3) G_{\varepsilon_1}^c(\mathbf{r}_3, \mathbf{r}_4) Y_{\text{Rabc}}(\varepsilon, \varepsilon_1, \omega, \mathbf{r}_1, \mathbf{r}_3) \\ &\quad \times Y_{\text{cbaA}}(\varepsilon_1, \varepsilon, \omega, \mathbf{r}_2, \mathbf{r}_4), \end{aligned} \quad (57)$$

where here and in what follows the interaction line  $Y_{abcd}$  corresponds to the SSI  $V(\mathbf{q})$ . The factor  $-N_s$  in Eq. (57) comes from the upper closed electron loop [the bottom bubble containing a highlighted Green function at the external energy  $\Delta G_\varepsilon$  is not a closed loop in the diagrammatic sense (see Fig. 3) and does not contribute an extra factor  $-N_s$ ].

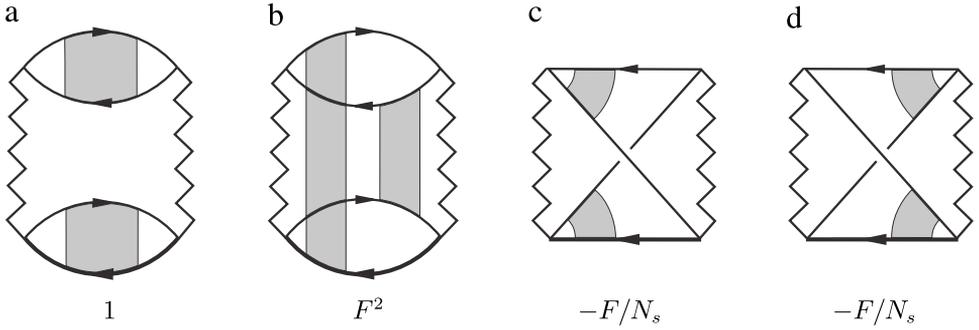
The diagram 4(a) is not the only one with two SSI lines that describes quasiparticle decay process. The other diagram is shown in Fig. 4(b). Though it is not as intuitive as the diagram 4(a), it also makes a contribution to the relaxation rate. That contribution is usually discarded since it vanishes in the limit  $F \rightarrow 0$  (see below). However as we show in Section 7.4, the diagram 4(b) should be taken into account in calculating mesoscopic fluctuations of the relaxation rate, since its contribution is comparable to that of the diagram 4(a) even in the limit  $F \rightarrow 0$ . The self-energy for the diagram 4(b) is given by

$$\begin{aligned} \Sigma_b^{\text{RA}} &= \left( \frac{i}{2} \right)^2 \sum_{abc} G_{\varepsilon_1}^a(\mathbf{r}_1, \mathbf{r}_2) G_{\varepsilon_1-\omega}^b(\mathbf{r}_2, \mathbf{r}_3) G_{\varepsilon-\omega}^c(\mathbf{r}_3, \mathbf{r}_4) Y_{\text{Rabc}}(\varepsilon, \varepsilon - \omega, \varepsilon - \varepsilon_1, \mathbf{r}_1, \mathbf{r}_3) \\ &\quad \times Y_{\text{abcA}}(\varepsilon_1, \varepsilon, \omega, \mathbf{r}_2, \mathbf{r}_4). \end{aligned} \quad (58)$$

Due to the absence of the closed electron loop in the diagram 4(b),  $\Sigma_b^{\text{RA}}$  does not contain an extra factor  $-N_s$  compared to  $\Sigma_a^{\text{RA}}$ .

The energy relaxation rate  $\gamma(\varepsilon, T)$  (in a given realization of disorder) due to processes with two SSI lines can be obtained from Eq. (45) with  $\Sigma^{\text{RA}} = \Sigma_a^{\text{RA}} + \Sigma_b^{\text{RA}}$ . After some algebra involving analyticity properties we obtain

$$\begin{aligned} \gamma(\varepsilon, T) \int d\mathbf{r} \Delta G_\varepsilon(\mathbf{r}, \mathbf{r}) &= -\frac{i\lambda^2}{4N_s^2 v^2} \int d\mathbf{r}_1 \dots d\mathbf{r}_4 \int (d\varepsilon_1)(d\omega) (\mathcal{F}_{\varepsilon_1} - \mathcal{F}_{\varepsilon_1-\omega}) \\ &\quad \times (\mathcal{F}_{\varepsilon-\omega} + \mathcal{B}_\omega) [N_s X_4 - Y_4], \end{aligned} \quad (59)$$



**Fig. 5.** Possible ways to average the diagrams for the relaxation rate shown in Fig. 4 over disorder (other types of averaging with two diffusons exist but are suppressed at least by the factor  $l/L \ll 1$ , where  $l$  is the mean free path). (a) RPA averaging of the diagram 4(a) reproducing the SIA result for  $\gamma_0^{\text{RPA}}(\varepsilon, T)$  [Eq. (56)]. (b) Non-RPA averaging of the diagram 4(a) with the contribution  $F^2\gamma_0^{\text{RPA}}(\varepsilon, T)$ . (c) and (d) Two ways to average the diagram 4(b), each producing the contribution  $-(F/N_s)\gamma_0^{\text{RPA}}(\varepsilon, T)$  to the relaxation rate.

where

$$X_4 = \delta_\kappa(\mathbf{r}_1 - \mathbf{r}_3)\delta_\kappa(\mathbf{r}_2 - \mathbf{r}_4) \Delta G_\varepsilon(\mathbf{r}_2, \mathbf{r}_1) \Delta G_{\varepsilon-\omega}(\mathbf{r}_1, \mathbf{r}_2) \Delta G_{\varepsilon_1-\omega}(\mathbf{r}_4, \mathbf{r}_3) \Delta G_{\varepsilon_1}(\mathbf{r}_3, \mathbf{r}_4) \quad (60)$$

and

$$Y_4 = \delta_\kappa(\mathbf{r}_1 - \mathbf{r}_3)\delta_\kappa(\mathbf{r}_2 - \mathbf{r}_4) \Delta G_\varepsilon(\mathbf{r}_4, \mathbf{r}_1) \Delta G_{\varepsilon_1}(\mathbf{r}_1, \mathbf{r}_2) \Delta G_{\varepsilon_1-\omega}(\mathbf{r}_2, \mathbf{r}_3) \Delta G_{\varepsilon-\omega}(\mathbf{r}_3, \mathbf{r}_4). \quad (61)$$

In Eq. (59), the term  $N_s X_4$  comes from  $\Sigma_a^{\text{RA}}$  [diagram 4(a)], and the term  $-Y_4$  comes from  $\Sigma_b^{\text{RA}}$  [diagram 4(b)]. An expression similar to Eq. (59) has been derived in Ref. [37], where only the contribution from the diagram 4(a) has been considered. In the limit when the interaction radius  $1/\kappa$  exceeds the Fermi wavelength, i.e. at  $F \ll 1$ , disorder averaged  $\langle \gamma(\varepsilon, T) \rangle$  reproduces the result (56) for  $\gamma_0^{\text{RPA}}(\varepsilon, T)$  [see Eq. (63)].

### 4.3. Disorder averaging and the role of the parameter $F$

Eq. (59) is the starting point for evaluating moments of the relaxation rate. In order to find its mean value one has to average the combinations of four Green functions in Eqs. (60) and (61).

The RPA result of Section 4.1 is reproduced if one takes only the diagram 4(a) (the term  $X_4$ ) into account and average each bubble independently [see Fig. 5(a)]:

$$\langle X_4 \rangle_{\text{RPA}} = \delta_\kappa(\mathbf{r}_1 - \mathbf{r}_3)\delta_\kappa(\mathbf{r}_2 - \mathbf{r}_4) \langle \Delta G_\varepsilon(\mathbf{r}_2, \mathbf{r}_1) \Delta G_{\varepsilon-\omega}(\mathbf{r}_1, \mathbf{r}_2) \rangle \langle \Delta G_{\varepsilon_1-\omega}(\mathbf{r}_4, \mathbf{r}_3) \Delta G_{\varepsilon_1}(\mathbf{r}_3, \mathbf{r}_4) \rangle. \quad (62)$$

Averaging with the help of Eq. (51) and integrating over  $\varepsilon_1$ , one arrives at Eq. (56) for  $\gamma_0^{\text{RPA}}(\varepsilon, T)$ .

Along with the RPA averaging (62), there exists another way to average  $X_4$  over disorder shown in Fig. 5(b). It has the same structure of diffusons as the RPA diagram 5(a), but with the interaction lines taken at fast momenta  $q \sim \min(\kappa, p_F)$ . The corresponding contribution to the relaxation rate,  $F^2\gamma_0^{\text{RPA}}(\varepsilon, T)$ , differs from the RPA-contribution by a factor  $F^2$ , where  $F$  is the standard notation for the ratio of the Hartree and Fock diagrams [19,20,38] discussed in Appendix A.

Finally, consider disorder averaging of the diagram 4(b) corresponding to the term  $Y_4$  in Eq. (59). For this diagram, there are also two ways to draw two-diffuson configurations shown in Fig. 5(c) and (d). Their contributions are equal and can be calculated similar to that of the diagram 5(b), but now only one of the two SSI lines is taken at fast momentum making the result proportional to the first power of  $F$ . The overall correction of the diagram 4(b) to the average relaxation rate is given by  $-(2F/N_s)\gamma_0^{\text{RPA}}(\varepsilon, T)$ .

Thus we see that among four possible contributions to the average relaxation rate shown in Fig. 5, all non-RPA diagrams [(b)–(d)] are proportional to some power of  $F$  and hence are suppressed if the interaction radius is larger than the Fermi wavelength. The smallness of the non-RPA diagrams

in this limit should be attributed to the Friedel oscillations which suppress the contribution of the corresponding process [29]. Such a situation was considered, e.g., in Refs. [21,33,34]. On the other hand, for the point-like interaction (corresponding to  $F = 1$ ), the diagrams 5(a) and (b) give the same contribution. This short-range limit was considered in Refs. [3,5,35,39].

Collecting the contributions of all diagrams with two SSI lines and two diffusons (Fig. 5), we obtain the final result for the average energy relaxation rate  $\gamma_0(\varepsilon, T) = \langle \gamma(\varepsilon, T) \rangle$ :

$$\gamma_0(\varepsilon, T) = (1 - 2F/N_s + F^2) \gamma_0^{\text{RPA}}(\varepsilon, T), \quad (63)$$

leading to Eq. (7). Note that for spinless electrons with point-like interaction, the inelastic relaxation rate vanishes,  $\gamma = 0$ , which is a consequence of the Pauli exclusion principle. It is essential that such a cancellation takes place only if the diagram 4(b) is taken into account.

## 5. Mesoscopic fluctuations of the energy relaxation rate: general consideration

In this Section we discuss the general approach to the calculation of mesoscopic fluctuations of the energy relaxation rate  $\gamma(\varepsilon, T)$ . The starting point is the diagrams for the collision integral shown in Fig. 4(a) and (b). The corresponding expression for  $\gamma(\varepsilon, T)$  is given by Eq. (59) which is written for a particular realization of impurities. Mesoscopic fluctuations of the relaxation rate are determined by the square of  $\gamma(\varepsilon, T)$  averaged over disorder:

$$\begin{aligned} & \langle \gamma^2(\varepsilon, T) \rangle \int d\mathbf{r} d\mathbf{r}' \langle \Delta G_\varepsilon(\mathbf{r}, \mathbf{r}) \Delta G_\varepsilon(\mathbf{r}', \mathbf{r}') \rangle \\ &= -\frac{\lambda^4}{16N_s^4 v^4} \int d\mathbf{r}_1 \dots d\mathbf{r}_8 \int (d\varepsilon_1)(d\varepsilon_2)(d\omega_1)(d\omega_2) (\mathcal{F}_{\varepsilon_1} - \mathcal{F}_{\varepsilon_1 - \omega_1})(\mathcal{F}_{\varepsilon - \omega_1} + \mathcal{B}_{\omega_1}) \\ & \quad \times (\mathcal{F}_{\varepsilon_2} - \mathcal{F}_{\varepsilon_2 - \omega_2})(\mathcal{F}_{\varepsilon - \omega_2} + \mathcal{B}_{\omega_2}) \langle (N_s X_4 - Y_4)(N_s X'_4 - Y'_4) \rangle, \end{aligned} \quad (64)$$

where the objects  $X_4$  and  $Y_4$  containing different products of four Green functions and two SSI lines are defined in Eqs. (60) and (61), while  $X'_4$  and  $Y'_4$  are obtained from them through the replacement

$$\{\mathbf{r}_i\} \rightarrow \{\mathbf{r}_{i+4}\}, \quad \varepsilon_1 \rightarrow \varepsilon_2, \quad \omega_1 \rightarrow \omega_2. \quad (65)$$

Eq. (64) contains three different products of eight Green functions,  $X_4 X'_4$ ,  $X_4 Y'_4$  and  $Y_4 Y'_4$ , which need to be separately averaged over disorder (the term  $Y_4 X'_4$  reduces to  $X_4 Y'_4$  after an obvious change of variables). For example, the explicit form of  $X_4 X'_4$  is given by

$$\begin{aligned} X_4 X'_4 &= \delta_\kappa(\mathbf{r}_1 - \mathbf{r}_3) \delta_\kappa(\mathbf{r}_2 - \mathbf{r}_4) \delta_\kappa(\mathbf{r}_5 - \mathbf{r}_7) \delta_\kappa(\mathbf{r}_6 - \mathbf{r}_8) \Delta G_\varepsilon(\mathbf{r}_2, \mathbf{r}_1) \Delta G_{\varepsilon - \omega_1}(\mathbf{r}_1, \mathbf{r}_2) \\ & \quad \times \Delta G_{\varepsilon_1 - \omega_1}(\mathbf{r}_4, \mathbf{r}_3) \Delta G_{\varepsilon_1}(\mathbf{r}_3, \mathbf{r}_4) \Delta G_\varepsilon(\mathbf{r}_6, \mathbf{r}_5) \Delta G_{\varepsilon - \omega_2}(\mathbf{r}_5, \mathbf{r}_6) \Delta G_{\varepsilon_2 - \omega_2}(\mathbf{r}_8, \mathbf{r}_7) \Delta G_{\varepsilon_2}(\mathbf{r}_7, \mathbf{r}_8). \end{aligned} \quad (66)$$

Since  $\Delta G = G^R - G^A$ , each of the products  $X_4 X'_4$ ,  $X_4 Y'_4$  and  $Y_4 Y'_4$  contains  $2^8 = 256$  different elementary contributions in terms of  $G^R$  and  $G^A$ .

Our task is to calculate the irreducible part  $\langle\langle \gamma^2(\varepsilon, T) \rangle\rangle$  defined in Eq. (10). In a usual situation, mesoscopic fluctuations of a random quantity  $x$  are determined by irreducible averaging over disorder, when two copies of  $x$  are connected by at least one impurity line. In the present case the situation is different as the left-hand side of the basic Eq. (64) contains the pairwise correlator  $\langle \Delta G_\varepsilon \Delta G_\varepsilon \rangle$ , which deviates significantly from its reducible part  $\langle \Delta G_\varepsilon \rangle^2$  (see Section 6.1). Therefore, even extracting the square of the average,  $\langle \gamma(\varepsilon, T) \rangle^2$ , from Eq. (64) should be done with care as it involves irreducible disorder averaging of  $\langle (N_s X_4 - Y_4)(N_s X'_4 - Y'_4) \rangle$  needed for non-perturbative account for correlations between two Green functions with the energy  $\varepsilon$ . The details of this calculation are presented in Appendix C. Such a complication is the price one has to pay for extracting the property of a single discrete level with the technique well suited for describing continuous spectra. Keeping that in mind we proceed with evaluation of the irreducible part of  $\langle \gamma^2(\varepsilon, T) \rangle$ .

In order to determine  $\langle \gamma^2(\varepsilon, T) \rangle$  one has to compute  $\langle X_4 X'_4 \rangle$ ,  $\langle X_4 Y'_4 \rangle$  and  $\langle Y_4 Y'_4 \rangle$ , and evaluate four remaining energy integrals in Eq. (64). This calculation is rather nontrivial and will be performed in Sections 6–8. Meanwhile we discuss the main idea and outline the principal technical steps of the derivation.

The most complicated task of this procedure is to perform disorder averaging which becomes rather involved in the low-energy limit,  $(\varepsilon, T) \ll E_{\text{Th}}$ , we are interested in. Indeed, the single-particle spectrum is well resolved in this case,  $\langle \gamma(\varepsilon, T) \rangle \ll \Delta$ , indicating that the averaging is to be performed non-perturbatively. That can be achieved with the help of the Efetov's nonlinear supersymmetric sigma model technique [22], properly generalized to the case of several Green functions with different energies.

The averaging of two Green functions on the left-hand side of Eq. (64) is straightforward and can be done exactly in the zero-dimensional geometry, see Section 6.1.

The calculation of  $\langle X_4 X'_4 \rangle$ ,  $\langle X_4 Y'_4 \rangle$  and  $\langle Y_4 Y'_4 \rangle$ , containing eight Green functions is much more complicated and cannot be performed exactly in the general case. Fortunately, the exact expression is not required for determination of the leading contribution to mesoscopic fluctuations of the relaxation rate. A significant simplification is suggested by analyzing the physics of the decay process. According to the FGR,  $\gamma(\varepsilon, T)$  can be considered as a sum of the decay processes of a single-particle excitation into all possible final three-particle states allowed by the energy conservation:  $\varepsilon \rightarrow (\varepsilon - \omega, \varepsilon', -\varepsilon' + \omega)$ . Consequently,  $\gamma^2(\varepsilon, T)$  given by Eq. (64) contains, in principle, six different final states:  $(\varepsilon - \omega_1, \varepsilon_1, -\varepsilon_1 + \omega_1, \varepsilon - \omega_2, \varepsilon_2, -\varepsilon_2 + \omega_2)$ .

However, as it was first demonstrated in Ref. [14], the leading contribution to mesoscopic fluctuations of the relaxation rate comes from those configurations that describe *the square of the same decay process*, i.e., from the terms with the identical set of the final states. Such a situation can be realized with two choices:

$$(a) \varepsilon_1 \approx \varepsilon_2 \quad \text{and} \quad \omega_1 \approx \omega_2, \quad (b) \varepsilon_1 \approx \varepsilon - \omega_2 \quad \text{and} \quad \varepsilon_2 \approx \varepsilon - \omega_1, \quad (67)$$

where the energies should coincide with the accuracy of the single level width  $\gamma$ . Technically, the importance of such configurations is related to the appearance of the formally divergent delta function of zero frequency,  $\delta(0)$ , if conditions (67) are considered as strict equalities, corresponding to the use of non-interacting Green functions in  $\langle X_4 X'_4 \rangle$ ,  $\langle X_4 Y'_4 \rangle$  and  $\langle Y_4 Y'_4 \rangle$ . In order to take into account the single-particle level broadening due to interaction one should add an imaginary part to the energy argument  $E$  of the Green function, replacing it by  $E_+$  (for  $G^R$ ) and  $E_-$  (for  $G^A$ ):

$$E_{\pm} = E \pm i\gamma(E)/2, \quad (68)$$

where  $\gamma(E) \equiv \gamma_0(E, T)$  is the average value of the relaxation rate obtained in the FGR approximation [Eq. (7)]. The substitution (68) is equivalent to including the elastic part of the electron–electron interaction in the zero-dimensional diffuson (see Section 6.4 and Appendix D). A finite imaginary part cuts the singularity in  $\delta(0)$  which should be replaced roughly by  $1/\gamma$ , making fluctuations finite but divergent as  $\varepsilon$  and  $T$  are decreased. This enhancement of fluctuations at low energies and temperatures is precisely the effect we are looking for.

Such a mechanism of enhancement of mesoscopic fluctuations of the inelastic width with decreasing temperature was suggested in Ref. [14] for a model when the matrix elements of the interaction  $V_{\alpha\beta\gamma\delta}$  in the basis of exact single-particle states were assumed to be independently distributed Gaussian variables. Our task is more complicated as we do not make any assumptions about the structure of the matrix elements in Fock space and calculate them for the real Coulomb interaction. Though we do not use the language of the matrix elements  $V_{\alpha\beta\gamma\delta}$ , they are effectively generated after proper averaging over fast diffusive modes [3,5].

Having discussed the main idea, we now outline the crucial steps in the calculation of  $\langle X_4 X'_4 \rangle$ ,  $\langle X_4 Y'_4 \rangle$  and  $\langle Y_4 Y'_4 \rangle$ :

- Each of  $X_4 X'_4$ ,  $X_4 Y'_4$  and  $Y_4 Y'_4$  contains in general  $2^8 = 256$  elementary products of  $G^R$  and  $G^A$ . As we discuss in Sections 6.4 and 7.1, the principal contribution to  $\langle \gamma^2(\varepsilon, T) \rangle$  originates from the terms with four  $G^R$  and four  $G^A$ . Each of these  $C_8^4 = 70$  terms may be presented as a functional integral over the 16-component superfields which is then averaged over disorder following the standard technique [22]. The resulting supersymmetric nonlinear sigma model is formulated in terms of the functional integral over the  $16 \times 16$  superfield  $Q(\mathbf{r})$ , see Section 6.2.
- As the SSI lines carry nonzero momenta (otherwise the diagram is zero due to electroneutrality), momentum conservation requires the presence of some number of fast ( $q \neq 0$ ) diffusive modes.

Since each fast diffuson contributes a small factor of  $\Delta/E_{\text{Th}}$ , their number should be minimized. For  $\langle \gamma(\varepsilon, T) \rangle$  the minimal number was two [see Fig. 5], and for  $\langle \langle \gamma^2(\varepsilon, T) \rangle \rangle$  it is four, producing the factors  $1/E_2^4$  and  $1/E_4^4$ . Following the method developed in Ref. [40], we integrate over these fast modes perturbatively and derive an effective action for the zero-mode (spatially-uniform supermatrix  $Q$ ). This procedure is described in detail in Section 6.3, with the total number of emergent contributions  $7!! = 105$  for every given elementary term from  $X_4 X_4', X_4 Y_4'$  and  $Y_4 Y_4'$  with four  $G^R$  and four  $G^A$ .

- The resulting zero-dimensional sigma model is still too complicated because of the large size of the  $Q$  matrix. However, as described above, the leading contribution comes from configurations where among energy arguments of eight Green functions, four are pairwise equal (with an uncertainty of  $\gamma$ ). Since the typical energy difference between pairs is  $(\varepsilon, T) \gg \Delta$ , our extended sigma model splits into four blocks, each corresponding to the standard Efetov sigma model for  $\langle G^R G^A \rangle$ . The integrals over these sigma models are then evaluated non-perturbatively using the standard technique [22].

The detailed discussion of this step in connection with the choice of pairs is presented in Sections 6.4 and 6.5.

- As a result of the described procedure, quite a few terms will be generated. Fortunately, most of them will be either zero or less singular than expected. Only a small number of terms will effectively contribute to fluctuations of the relaxation rate. This selection is discussed in Section 7 in the simplest case of  $F \rightarrow 0$ . The final evaluation of  $\langle \langle \gamma^2(\varepsilon, T) \rangle \rangle$  in the general case of arbitrary parameter  $F$  is performed in Section 8.

The announced program will be realized step by step in Sections 6–8.

## 6. Nonperturbative averaging of eight Green functions

### 6.1. Pairwise correlator $\langle \Delta G_\varepsilon \Delta G_\varepsilon \rangle$

We start with discussing the pair correlation function on the left-hand side of Eq. (64). It is instructive to introduce a small energy mismatch  $\omega$  between the energies of two Green functions and to consider a more general correlation function

$$\int d\mathbf{r} d\mathbf{r}' \langle \Delta G_\varepsilon(\mathbf{r}, \mathbf{r}) \Delta G_{\varepsilon-\omega}(\mathbf{r}', \mathbf{r}') \rangle = -4\pi^2 v^2 R_\gamma(\omega). \quad (69)$$

As usual, the terms  $G^R G^R$  and  $G^A G^A$  are averaged trivially, each contributing  $1/4$  to  $R_\gamma(\omega)$ . Averaging of the cross term  $G^R G^A$  is performed with the help of Efetov's zero-dimensional supersymmetric sigma model [22], where in accordance with Eq. (68) one has to introduce a complex frequency

$$\Omega = \varepsilon_+ - (\varepsilon - \omega)_- = \omega + i[\gamma(\varepsilon) + \gamma(\varepsilon - \omega)]/2. \quad (70)$$

Evaluating the standard integrals with the help of the machinery developed in Appendix E, we obtain

$$R_\gamma(\omega) = 1 - \text{Re} X(-i\pi \Omega / \Delta), \quad (71)$$

where the function  $X(a)$  is given by Eq. (E.4):

$$X(a) = \frac{1 - e^{-2a}}{2a^2}. \quad (72)$$

In the limit of vanishing width ( $\gamma \rightarrow 0$ ),  $R_0(\omega)$  reproduces the pair correlation function for the Gaussian unitary ensemble in the random matrix theory [41]:  $R_0^{\text{RMT}}(\omega) = 1 - (\sin x/x)^2 + \pi \delta(x)$ , where  $x = \pi\omega/\Delta$ . A finite but small width acts as a regularizer of the  $\delta$  function, accounting for the contribution of the same broadened level into the correlation function. In the relevant limit of a resolved discrete spectrum,  $(\omega, \gamma) \ll \Delta$ ,  $R_\gamma(\omega)$  can be written as

$$R_\gamma(\omega) \approx \Delta \delta_\gamma(\omega), \quad (73)$$

where  $\delta_\gamma(\omega)$  is a Lorentzian approximation of the delta function:

$$\delta_\gamma(\omega) = \frac{1}{\pi} \frac{\gamma(\varepsilon)}{\gamma^2(\varepsilon) + \omega^2}. \tag{74}$$

It is worth noting that Eq. (73) is non-perturbative. Though it looks just like a one-diffuson contribution,  $R_\gamma(\omega) \approx -\text{Im} 1/(\omega + i\gamma)$ , it is an asymptotic expansion of the exact non-perturbative expression (71).

The left-hand side of Eq. (64) can be expressed with the help of Eq. (69), with  $R_\gamma(0) = \Delta/\pi\gamma(\varepsilon)$ .

### 6.2. Supersymmetric sigma model for eight Green functions (the case $F = 0$ )

The objects  $X_4X'_4$ ,  $X_4Y'_4$  and  $Y_4Y'_4$  contain the products of eight  $\Delta G = G^R - G^A$  which should be averaged over disorder. To introduce the method we start with the simplest case of  $F \rightarrow 0$ . Generalization to the case of an arbitrary  $F$  will be performed in Section 8. To be specific, we focus on calculating  $\langle X_4X'_4 \rangle$  (disorder averaging of other products is performed analogously and the result is presented in Section 7.4). As we will see in Sections 6.4 and 7.1, only terms with the equal number of retarded and advanced Green functions make the leading contribution to mesoscopic fluctuations. Consider, e.g., a particular choice of  $G^R$  and  $G^A$  from  $X_4X'_4$  and calculate

$$K = \langle G_\varepsilon^R(\mathbf{r}_2, \mathbf{r}_1) G_\varepsilon^A(\mathbf{r}_6, \mathbf{r}_5) G_{\varepsilon-\omega_2}^R(\mathbf{r}_5, \mathbf{r}_6) G_{\varepsilon-\omega_1}^A(\mathbf{r}_1, \mathbf{r}_2) \times G_{\varepsilon_1-\omega_1}^R(\mathbf{r}_4, \mathbf{r}_3) G_{\varepsilon_2-\omega_2}^A(\mathbf{r}_8, \mathbf{r}_7) G_{\varepsilon_2}^R(\mathbf{r}_7, \mathbf{r}_8) G_{\varepsilon_1}^A(\mathbf{r}_3, \mathbf{r}_4) \rangle. \tag{75}$$

Disorder averaging of other relevant terms from  $X_4X'_4$  [listed in Eq. (102)] can be performed analogously (see Sections 7.2 and 7.3).

In order to calculate  $K$  we follow the standard line of Efetov’s supersymmetric sigma model [22], generalizing it to the case of eight Green functions. We group the Green functions into four RA pairs in the sequence they appear in Eq. (75):

	1	2	3	4	
R	$\varepsilon$	$\varepsilon - \omega_2$	$\varepsilon_1 - \omega_1$	$\varepsilon_2$	
A	$\varepsilon$	$\varepsilon - \omega_1$	$\varepsilon_2 - \omega_2$	$\varepsilon_1$	(76)

thereby introducing a new space of pairs, that will be referred to as 1234. The way this space is introduced is consistent with the pairing (a) in Eq. (67) shown in Fig. 6(a). The leading contribution to mesoscopic fluctuations from this pairing will come from configurations when the energies in each pair nearly coincide, corresponding to  $Q$  matrices which are block-diagonal in the space 1234 [see Eq. (88)]. The other relevant contribution originates from the pairing (b) in Eq. (67) shown in Fig. 6(b). In principle, it can be also handled using the 1234 structure of Eq. (76), however the corresponding  $Q$  matrices will have a cumbersome structure in this basis. Therefore in the study of the pairing (b) in Section 7.3 we will introduce 1234 space in a different way consistent with that pairing.

The resulting sigma model is formulated in terms of the  $16 \times 16$  supermatrix field  $Q(\mathbf{r})$  acting in the tensor product of the spaces  $\text{FB} \otimes \text{RA} \otimes 1234$ , where FB stands for the superspace. The matrix  $Q$  can be written as  $Q = T^{-1} \Lambda T$ , where  $T$  spans the supersymmetric coset  $U(8|4, 4)/U(4|4) \times U(4|4)$ , and  $\Lambda = \sigma_3^{\text{RA}}$ . The sigma-model action is given by (we adopt the fermion-dominated notation of Ref. [22])

$$S[Q] = \frac{\pi\nu}{4} \int d\mathbf{r} \text{str} [D (\nabla Q(\mathbf{r}))^2 + 4i\hat{E}Q(\mathbf{r})], \tag{77}$$

where  $\hat{E}$  is the diagonal matrix made of the energy arguments [with widths, according to Eq. (68)] of the corresponding Green functions:

$$\hat{E} = \text{diag}\{\varepsilon_+, (\varepsilon - \omega_2)_+, (\varepsilon_1 - \omega_1)_+, (\varepsilon_2)_+, \varepsilon_-, (\varepsilon - \omega_1)_-, (\varepsilon_2 - \omega_2)_-, (\varepsilon_1)_-\} \otimes 1_{\text{FB}}. \tag{78}$$

In the sigma-model language, the average product of eight Green functions in Eq. (75) transforms to

$$K = (\pi \nu)^8 \int Q_{21}^{AR}(\mathbf{r}_1) Q_{12}^{RA}(\mathbf{r}_2) Q_{21}^{RA}(\mathbf{r}_5) Q_{12}^{AR}(\mathbf{r}_6) Q_{43}^{AR}(\mathbf{r}_3) Q_{34}^{RA}(\mathbf{r}_4) Q_{43}^{RA}(\mathbf{r}_7) Q_{34}^{AR}(\mathbf{r}_8) e^{-S[Q]} DQ, \quad (79)$$

where all the elements of  $Q$  matrices in the preexponent are taken from the BB sector (omitted for brevity). Eq. (79) holds only in the limit  $F \rightarrow 0$ , when the interaction range is much larger than the Fermi wave length, and additional terms in the preexponent are suppressed by Friedel oscillations. In the general case discussed in Section 8, Eq. (79) should be replaced by Eq. (119).

### 6.3. Integration over fast modes

Besides eight Green functions, the block  $X_4 X'_4$  given by Eq. (66) contains four SSI lines, each carrying a non-zero momentum. Therefore in calculating  $K$  in Eq. (79) we should (i) allow fast ( $q \neq 0$ ) diffuson modes to make  $\langle X_4 X'_4 \rangle$  nonzero, (ii) minimize their number as every fast mode brings a small factor of  $\Delta/E_{Th}$ , and (iii) be able to handle the resulting zero-dimensional integral non-perturbatively in order to resolve discrete levels. This task can be accomplished following the strategy of Ref. [40], where non-universal corrections to the random-matrix level statistics were calculated beyond the zero-dimensional limit. For this purpose we write

$$Q(\mathbf{r}) = T^{-1} Q'(\mathbf{r}) T, \quad (80)$$

where the matrix  $T$  is spatially uniform, and  $Q'(\mathbf{r})$  describes all fast modes with non-zero momenta. Since the latter are to be accounted perturbatively, we expand  $Q'(\mathbf{r})$  near the origin:

$$Q'(\mathbf{r}) = \Lambda [1 + W(\mathbf{r}) + W^2(\mathbf{r})/2 + \dots], \quad \{W(\mathbf{r}), \Lambda\} = 0. \quad (81)$$

To get the leading contribution to  $\langle X_4 X'_4 \rangle$  we extract one fast  $W$  from each of the eight  $Q$  matrices in the preexponent of Eq. (79), and average it with the Gaussian action

$$S^{(2)}[W] = -\frac{\pi \nu D}{4} \int d\mathbf{r} \text{str} [\nabla W(\mathbf{r})]^2, \quad (82)$$

coming from the gradient term of Eq. (77). Using Wick's theorem, the correlator of eight  $W$  fields can be expressed as the sum of all possible products of four pairwise correlators. The latter can be easily calculated with the help of the following contraction rule valid for arbitrary matrices  $P$  and  $R$ :

$$\langle \text{str} P W(\mathbf{r}) \text{str} R W(\mathbf{r}') \rangle_W = D(\mathbf{r}, \mathbf{r}') \text{str} (P A R A - P R), \quad D(\mathbf{r}, \mathbf{r}') = \frac{\langle \mathbf{r} | (-\nabla^2)^{-1} | \mathbf{r}' \rangle}{\pi \nu D}. \quad (83)$$

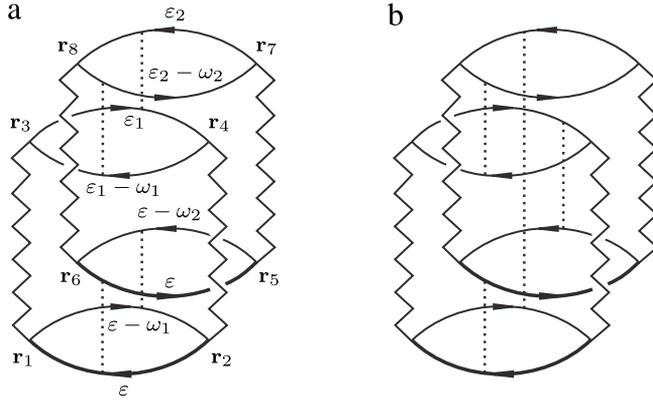
According to Wick's theorem, averaging over fast modes generates  $7!! = 105$  different terms in the expression for  $K$ . Not all of them are equally important. To pick up the relevant terms one should understand how to perform further integration over the zero mode. This procedure will be discussed in Section 6.4, and in Section 6.5 we will proceed with the derivation based on Eq. (83).

### 6.4. Block structure of the zero-dimensional sigma model and energy pairs

Having integrated out fast diffusive modes, we end up with the effective zero-dimensional sigma model for the  $16 \times 16$  supermatrix  $Q = T^{-1} \Lambda T$  with the action

$$S_0[Q] = \frac{\pi i}{\Delta} \text{str} \hat{E} Q. \quad (84)$$

Owing to a large size of the matrix  $Q$ , this is still a complicated theory. Fortunately, we do not need its exact solution since, as discussed in Section 5, the most singular contribution to  $\langle \langle \gamma^2(\varepsilon, T) \rangle \rangle$  comes from pairwise coinciding energies, see Eq. (67). Since we are interested in the limit  $(\varepsilon, T) \gg \Delta$ , energy difference between pairs is typically large,  $|\varepsilon_i - \varepsilon_j| \gg \Delta$ , and hence correlations between



**Fig. 6.** Two types of arranging eight Green functions in the product  $X_4 X'_4$  [Eq. (66)] into four pairs that produce the most singular contributions  $\langle X_4 X'_4 \rangle_{\text{sing}}^{(a)}$  and  $\langle X_4 X'_4 \rangle_{\text{sing}}^{(b)}$  in Eq. (85). Green functions from the same pair are marked by a dotted line. The pairings (a) and (b) correspond to the two ways to match energies in Eq. (67). All energy and coordinate indices in (b) are the same as in (a).

different pairs can be neglected (configurations with  $|\varepsilon_i - \varepsilon_j| \lesssim \Delta$  which require non-perturbative treatment exist but their weight is small). On the other hand, energies from the same pair match with the accuracy of  $\gamma$  and should be treated non-perturbatively like in Section 6.1. In the language of the zero-dimensional sigma model, large energy difference between pairs suppresses degrees of freedom in the supermatrix  $Q$  which are off-diagonal in the space of pairs. As a result,  $Q$  becomes block-diagonal in the space of pairs, and the full theory splits into a product of four standard Efetov's supermatrix sigma models for  $\langle G^R G^A \rangle$ , see Eq. (89).

The next step is to understand what are the possible ways to group eight Green functions into four pairs to maximize their contribution to fluctuations of  $\gamma(\varepsilon, T)$ . First of all, it is clear that the Green functions with the energies  $\varepsilon$  always need to be in a pair, as they immediately produce the smeared delta function of zero argument,  $\delta_\gamma(0) = 1/\pi\gamma$  [cf. Eq. (74)]. It is this large factor that compensates the analogous contribution from  $\langle G_\varepsilon^R G_\varepsilon^A \rangle$  on the left-hand side of Eq. (64). Our aim then is to identify those pairings that may introduce an additional large factor of  $\delta_\gamma(0)$  after integration over all intermediate energies. Such a situation can be realized only if coincidence of the energies in two pairs automatically implies coincidence of the energies in the third pair. That can be achieved only in two physically relevant cases, (a) and (b), listed in Eq. (67) and shown diagrammatically in Fig. 6 (there exist three other unphysical pairings which should be discarded as demonstrated in Appendix F). We emphasize that possible choices of pairing are dictated by the energy arguments of electron Green functions and are not specific for the particular arrangement of  $G^R$  and  $G^A$ . The only natural requirement is that each pair contains one retarded and one advanced Green function (otherwise, there are no correlations within a pair at all).

According to this general logic, each diagram for  $\langle X_4 X'_4 \rangle$ ,  $\langle X_4 Y'_4 \rangle$  and  $\langle Y_4 Y'_4 \rangle$  can be written in the form

$$\langle A \rangle = \langle A \rangle_{\text{sing}}^{(a)} + \langle A \rangle_{\text{sing}}^{(b)} + \langle A \rangle_{\text{reg}}, \tag{85}$$

where  $\langle \dots \rangle_{\text{sing}}^{(a)}$  and  $\langle \dots \rangle_{\text{sing}}^{(b)}$  denote singular (in the limit  $\gamma \rightarrow 0$ ) contributions from the pairings (a) and (b). We proceed with calculating these singular contributions for the term  $K$  [a particular element of  $\langle X_4 X'_4 \rangle$  defined in Eq. (75)] from the zero-dimensional sigma model (84). The introduction of the 1234 space in Eq. (76) is consistent with the pairing (a): in the 1234 basis the corresponding matrix  $Q$  becomes diagonal. On the contrary, the pairing (b) is described by non-diagonal matrices in the 1234 space. To calculate  $K_{\text{sing}}^{(b)}$  we find it more convenient to change the basis and introduce a new 1234' space consistent with the pairing (b), in which the matrix  $Q$  is diagonal. Then the calculation of  $K_{\text{sing}}^{(b)}$

becomes completely analogous to the calculation of  $K_{\text{sing}}^{(a)}$ . To illustrate the technique, we focus on the contribution of  $K$  due to the pairing (a) [the contribution of the pairing (b) is discussed in Section 7.3].

### 6.5. Singular contribution of the pairing (a)

Now we are ready to proceed with the calculation started in Section 6.3. In order to apply the contraction rules (83) for averaging the preexponent in Eq. (79) over fast modes, it is convenient to introduce projectors onto different sectors of the  $Q$  manifold. By definition, we write  $Q_{ij}^{ab}(\mathbf{r}) = \text{str } P_{ij}^{ab} Q(\mathbf{r})$ , where  $a$  and  $b$  refer to the RA space, whereas  $i$  and  $j$  refer to the 1234 space (BB sector is also implied). For example, the projector  $P_{21}^{\text{AR}}$  is given by

$$P_{21}^{\text{AR}} = \begin{pmatrix} 0 & P^{\text{AR}} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}_{1234}, \quad P^{\text{AR}} = \begin{pmatrix} 0 & P_{\text{BB}} \\ 0 & 0 \end{pmatrix}_{\text{RA}}, \quad P_{\text{BB}} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}_{\text{FB}}. \quad (86)$$

As we discussed in Section 6.4, the singular contribution  $K_{\text{sing}}^{(a)}$  originates from the matrices  $Q$  which are diagonal in 1234 space. Such a structure of the  $Q$  matrix guarantees that only four out of  $7!! = 105$  terms appearing after averaging over fast modes [see Eq. (80)] turn out to be non-zero:

$$K_{\text{sing}}^{(a)} = (\pi\nu)^8 \{ D(\mathbf{r}_1, \mathbf{r}_2) D(\mathbf{r}_5, \mathbf{r}_6) \text{str} (P_{21}^{\text{AR}} Q P_{12}^{\text{RA}} Q - P_{21}^{\text{AR}} P_{12}^{\text{RA}}) \text{str} (P_{21}^{\text{RA}} Q P_{12}^{\text{AR}} Q - P_{21}^{\text{RA}} P_{12}^{\text{AR}}) \\ + D(\mathbf{r}_1, \mathbf{r}_6) D(\mathbf{r}_2, \mathbf{r}_5) \text{str} (P_{21}^{\text{AR}} Q P_{12}^{\text{AR}} Q - P_{21}^{\text{AR}} P_{12}^{\text{AR}}) \text{str} (P_{21}^{\text{RA}} Q P_{12}^{\text{RA}} Q - P_{21}^{\text{RA}} P_{12}^{\text{RA}}) \} \\ \times \{ D(\mathbf{r}_3, \mathbf{r}_4) D(\mathbf{r}_7, \mathbf{r}_8) \text{str} (P_{43}^{\text{AR}} Q P_{34}^{\text{RA}} Q - P_{43}^{\text{AR}} P_{34}^{\text{RA}}) \text{str} (P_{43}^{\text{RA}} Q P_{34}^{\text{AR}} Q - P_{43}^{\text{RA}} P_{34}^{\text{AR}}) \\ + D(\mathbf{r}_3, \mathbf{r}_8) D(\mathbf{r}_4, \mathbf{r}_7) \text{str} (P_{43}^{\text{AR}} Q P_{34}^{\text{AR}} Q - P_{43}^{\text{AR}} P_{34}^{\text{AR}}) \text{str} (P_{43}^{\text{RA}} Q P_{34}^{\text{RA}} Q - P_{43}^{\text{RA}} P_{34}^{\text{RA}}) \}, \quad (87)$$

where  $\langle \dots \rangle$  denotes averaging over the  $16 \times 16$  zero-dimensional sigma-model manifold with the action (84).

Since  $Q$  is diagonal in the 1234 space, it can be represented as

$$Q = \text{diag}\{Q_1, Q_2, Q_3, Q_4\}, \quad (88)$$

where  $Q_j$  are independent  $4 \times 4$  supermatrices in the  $\text{RA} \otimes \text{FB}$  space spanning the standard supermanifold of the sigma model for  $\langle G^{\text{R}G^{\text{A}}} \rangle$ . The action (84) then splits into a sum of four separate Efetov's sigma-model actions:

$$S_0[Q] = \sum_{j=1}^4 S_{0j}[Q_j] = \frac{\pi i}{\Delta} \sum_{j=1}^4 \text{str} \hat{E}_j Q_j, \quad (89)$$

where the elements of the diagonal matrix  $\hat{E}_j$  are taken from the corresponding sector of  $\hat{E}$  defined in Eq. (78). Now averaging over all four  $Q_j$  is performed independently and can be done exactly in the standard way [22]. To handle the supertrace structure in Eq. (87), it is convenient to use the zero-dimensional contraction rules (E.3) derived in Appendix E. The result can be expressed in a compact form in terms of the quantities

$$X_j \equiv X(-i\pi \Omega_j / \Delta), \quad Z_j \equiv Z(-i\pi \Omega_j / \Delta), \quad (90)$$

where the functions  $X(a)$  and  $Y(a)$  are given by Eqs. (E.4) and (E.5):

$$X(a) = \frac{1 - e^{-2a}}{2a^2}, \quad Z(a) = \frac{1}{a}, \quad (91)$$

and  $\Omega_j$  is the energy difference between the arguments of  $G^{\text{R}}$  and  $G^{\text{A}}$  in the corresponding pair [cf. Eq. (76)]:

$$\Omega_1 = \varepsilon_+ - \varepsilon_-, \quad \Omega_2 = (\varepsilon - \omega_2)_+ - (\varepsilon - \omega_1)_-, \\ \Omega_3 = (\varepsilon_1 - \omega_1)_+ - (\varepsilon_2 - \omega_2)_-, \quad \Omega_4 = (\varepsilon_2)_+ - (\varepsilon_1)_-. \quad (92)$$

To demonstrate the technique, consider for example one of the four terms in Eq. (87):

$$K_{\text{sing}}^{(a)} = (\pi\nu)^8 D(\mathbf{r}_1, \mathbf{r}_2) D(\mathbf{r}_5, \mathbf{r}_6) D(\mathbf{r}_3, \mathbf{r}_4) D(\mathbf{r}_7, \mathbf{r}_8) L_1 + \dots \quad (93)$$

Substituting  $Q$  in the form (88) and tracing over the 1234 space we obtain

$$L_1 = \left( \text{str} (P^{\text{AR}} Q_2 P^{\text{RA}} Q_1 - P_{21}^{\text{AR}} P_{12}^{\text{RA}}) \text{str} (P^{\text{RA}} Q_2 P^{\text{AR}} Q_1 - P_{21}^{\text{RA}} P_{12}^{\text{AR}}) \right. \\ \left. \times \text{str} (P^{\text{AR}} Q_4 P^{\text{RA}} Q_3 - P_{43}^{\text{AR}} P_{34}^{\text{RA}}) \text{str} (P^{\text{RA}} Q_4 P^{\text{AR}} Q_3 - P_{43}^{\text{RA}} P_{34}^{\text{AR}}) \right). \quad (94)$$

Sequentially applying the contraction rules (E.3) for averaging over all  $Q_j$ , we arrive at the exact non-perturbative expression

$$L_1 = (4 + 2X_1 + 2X_2 + 4X_1X_2)(4 + 2X_3 + 2X_4 + 4X_3X_4). \quad (95)$$

Contributions of the three other terms in Eq. (93) are calculated analogously, and we obtain finally the singular part of  $K$  due to pairing (a):

$$K_{\text{sing}}^{(a)} = (\pi\nu)^8 \{ D(\mathbf{r}_1, \mathbf{r}_2) D(\mathbf{r}_5, \mathbf{r}_6) (4 + 2X_1 + 2X_2 + 4X_1X_2) + 4D(\mathbf{r}_1, \mathbf{r}_6) D(\mathbf{r}_2, \mathbf{r}_5) Z_1 Z_2 \} \\ \times \{ D(\mathbf{r}_3, \mathbf{r}_4) D(\mathbf{r}_7, \mathbf{r}_8) (4 + 2X_3 + 2X_4 + 4X_3X_4) + 4D(\mathbf{r}_3, \mathbf{r}_8) D(\mathbf{r}_4, \mathbf{r}_7) Z_3 Z_4 \}. \quad (96)$$

Though the singular contribution due to the energy pairing (a) may originate only from Eq. (96), not all terms in this equation do actually produce the singular contribution. We analyze this expression and perform the final step of calculation in the next Section.

## 7. Mesoscopic fluctuations at $F = 0$

### 7.1. General recipe

The most singular contribution from Eq. (96) originates from the terms which contain the maximal number (four) of  $X_j$  and  $Z_j$ . We have already explained that this choice is dictated by the necessity to obtain two smeared delta-functions (74) with zero argument, contributing a large factor  $\Delta/\gamma$  each. This is also the reason why we have to choose each pair in the 1234 space to consist of one  $G^{\text{R}}$  and one  $G^{\text{A}}$ : large factors  $X_j$  and  $Z_j$  appear as a result of averaging  $\langle G^{\text{R}} G^{\text{A}} \rangle$  over the zero-dimensional diffusive model.

In calculating the energy integrals in Eq. (64), the important contribution comes from the vicinity of the poles of  $X_j$  and  $Z_j$ . This observation allows us to work in the limit  $\Omega_j \sim \gamma \ll \Delta$  [with  $\Omega_j$  introduced in Eq. (92) being the energy mismatch within a pair] and use the leading-order asymptotics of Eq. (91):

$$X_j \approx Z_j = \frac{i\Delta}{\pi\Omega_j}. \quad (97)$$

The chosen sequence of  $G^{\text{R}}$  and  $G^{\text{A}}$  in the correlator  $K$  [Eq. (75)] also ensures that the product  $1/\Omega_2\Omega_3\Omega_4$  has poles both in the upper and lower half-planes of the energy variables  $\varepsilon_1, \varepsilon_2, \omega_1$  and  $\omega_2$ . Hence integration over intermediate energies does not vanish and indeed produces the desired delta-function-like contribution.

Now the recipe for extracting the leading singular term from Eq. (96) and similar expressions for other diagrams can be formulated as follows. One should substitute all  $X_j$  by  $Z_j$  and take the term proportional to  $\mathcal{Z} = Z_1 Z_2 Z_3 Z_4$ . Then the coefficient  $\mathcal{Z}$  should be replaced by

$$\mathcal{Z} \rightarrow \frac{4\Delta^4}{\pi^2\gamma(\varepsilon)} \frac{\delta(\varepsilon_1 - \varepsilon_2)\delta(\omega_1 - \omega_2)}{\gamma(\varepsilon - \omega_1) + \gamma(\varepsilon_1) + \gamma(\varepsilon_1 - \omega_1)}, \quad (98)$$

where the factor  $\Delta/\pi\gamma(\varepsilon)$  originates from  $Z_1$  and will be canceled by a similar factor from the left-hand side of Eq. (64) (see Section 6.1), while the coefficient in front of the delta functions can be easily obtained by calculating, e.g.,  $\int Z_2 Z_3 Z_4 d\varepsilon_2 d\omega_2$ . The last step is to integrate four diffuson propagators in Eq. (96) over  $\mathbf{r}_i$  [see Eq. (64)]. In doing that,  $\delta_{\mathbf{r}}(\mathbf{r} - \mathbf{r}')$  in the expressions for  $X_4 X_4', X_4 Y_4'$  and  $Y_4 Y_4'$  can be considered as the usual zero-range delta function. Then depending on the term under consideration,

diffusons combine either into  $(\text{tr } D^2)^2$  or  $\text{tr } D^4$ . These traces evaluated in the Fourier space produce either  $1/E_2^4$  or  $1/E_4^4$ , where  $E_n \sim E_{\text{Th}}$  are defined in Eq. (8). As a result, the contribution to mesoscopic fluctuations of the inelastic rate can be written in the form

$$\langle\langle \gamma^2(\varepsilon, T) \rangle\rangle_A^{(i)} = \frac{\lambda^4 \Delta^5}{4\pi^3} c_A^{(i)} \int d\varepsilon_1 d\omega_1 \frac{(\mathcal{F}_{\varepsilon_1} - \mathcal{F}_{\varepsilon_1 - \omega_1})^2 (\mathcal{B}_{\omega_1} + \mathcal{F}_{\varepsilon - \omega_1})^2}{\gamma_0(\varepsilon - \omega_1, T) + \gamma_0(\varepsilon_1, T) + \gamma_0(\varepsilon_1 - \omega_1, T)}, \quad (99)$$

where  $A$  is a particular diagram or a set of diagrams considered, and  $i$  labels the type of the energy pairing (a or b).

### 7.2. Contribution from $X_4 X_4'$ due to the energy pairing (a)

Applying this general recipe to expression (96) for  $K_{\text{sing}}^{(a)}$ , we first reduce it to the form

$$K_{\text{sing}}^{(a)} = 16(\pi\nu)^8 \mathcal{Z} \{ D(\mathbf{r}_1, \mathbf{r}_2)D(\mathbf{r}_5, \mathbf{r}_6) + D(\mathbf{r}_1, \mathbf{r}_6)D(\mathbf{r}_2, \mathbf{r}_5) \} \\ \times \{ D(\mathbf{r}_3, \mathbf{r}_4)D(\mathbf{r}_7, \mathbf{r}_8) + D(\mathbf{r}_3, \mathbf{r}_8)D(\mathbf{r}_4, \mathbf{r}_7) \}, \quad (100)$$

and then, tracing the diffusion propagators, obtain the coefficient  $c_K^{(a)}$  in Eq. (99):

$$c_K^{(a)} = \frac{1}{2N_s^2} \left( \frac{1}{E_2^4} + \frac{1}{E_4^4} \right). \quad (101)$$

To complete the analysis of the contribution from  $X_4 X_4'$  due to type-(a) energy pairing, one has to consider other arrangements of  $G^R$  and  $G^A$ . The requirement of having one  $G^R$  and one  $G^A$  in each pair limits the number of various possibilities to  $2^4 = 16$ . However, not all of them should be taken into account. Consider, for example, the correlator

$$K' = \langle G_\varepsilon^R(\mathbf{r}_2, \mathbf{r}_1) G_\varepsilon^A(\mathbf{r}_6, \mathbf{r}_5) G_{\varepsilon - \omega_2}^R(\mathbf{r}_5, \mathbf{r}_6) G_{\varepsilon - \omega_1}^A(\mathbf{r}_1, \mathbf{r}_2) \\ \times G_{\varepsilon_1 - \omega_1}^R(\mathbf{r}_4, \mathbf{r}_3) G_{\varepsilon_2 - \omega_2}^A(\mathbf{r}_8, \mathbf{r}_7) G_{\varepsilon_2}^A(\mathbf{r}_7, \mathbf{r}_8) G_{\varepsilon_1}^R(\mathbf{r}_3, \mathbf{r}_4) \rangle,$$

which differs from  $K$  [Eq. (75)] by changing  $R \leftrightarrow A$  in the last two Green functions. The expression analogous to Eq. (96) will contain  $X_4^*$  instead of  $X_4$  and  $Z_4^*$  instead of  $Z_4$ , which renders the poles of both  $X_3 X_4^*$  and  $Z_3 Z_4^*$  lying in the upper half-plane of  $\varepsilon_2$ . Therefore, the contribution of this term is non-singular and should be disregarded. A simple analysis demonstrates that there are only four possibilities to arrange  $G^R$  and  $G^A$  in  $X_4 X_4'$ :

$$\text{RARARARA, ARRARARA, RAARARAR, ARARARAR}, \quad (102)$$

where all energy and coordinate indices follow Eq. (75). Each choice gives the same contribution given by Eq. (101). As a result, the total contribution of the term  $X_4 X_4'$  due to type-(a) energy pairing (shown in Fig. 6(a)) is described by the coefficient

$$c_{XX}^{(a)} = \frac{2}{N_s^2} \left( \frac{1}{E_2^4} + \frac{1}{E_4^4} \right). \quad (103)$$

### 7.3. Contribution from $X_4 X_4'$ due to the energy pairing (b)

Following the same line we can analyze the singular part of  $K$  due to pairing (b) shown in Fig. 6(b). Instead of Eq. (100) we obtain:

$$K_{\text{sing}}^{(b)} = 16(\pi\nu)^8 \mathcal{Z} D(\mathbf{r}_1, \mathbf{r}_2)D(\mathbf{r}_3, \mathbf{r}_4)D(\mathbf{r}_5, \mathbf{r}_6)D(\mathbf{r}_7, \mathbf{r}_8). \quad (104)$$

Following the procedure described in Section 7.1 and utilizing the same four possibilities (102) to arrange  $G^R$  and  $G^A$ , we arrive at

$$c_{XX}^{(b)} = \frac{1}{N_s^2 E_2^4}. \quad (105)$$

#### 7.4. Contributions from $X_4Y'_4$ and $Y_4Y'_4$

It is left to discuss the other two terms,  $X_4Y'_4$  and  $Y_4Y'_4$ , in Eq. (64). Following the same steps we obtain that the contributions of the cross term  $\langle X_4Y'_4 \rangle$  from both the pairings (a) and (b) vanish after integration over fast diffusive modes due to the diagonal structure of the  $Q$  matrix in the 1234 space, while in calculating  $\langle Y_4Y'_4 \rangle$  only the pairing (b) should be taken into account for the same reason.

Consider, e.g., a particular realization of  $G^R$  and  $G^A$  from  $\langle Y_4Y'_4 \rangle$  [compare with  $K$  defined in Eq. (75)]:

$$\begin{aligned} \tilde{K} = & \langle G_{\varepsilon}^R(\mathbf{r}_4, \mathbf{r}_1) G_{\varepsilon_1}^A(\mathbf{r}_1, \mathbf{r}_2) G_{\varepsilon_1 - \omega_1}^R(\mathbf{r}_2, \mathbf{r}_3) G_{\varepsilon - \omega_1}^A(\mathbf{r}_3, \mathbf{r}_4) \\ & \times G_{\varepsilon}^A(\mathbf{r}_8, \mathbf{r}_5) G_{\varepsilon_2}^R(\mathbf{r}_5, \mathbf{r}_6) G_{\varepsilon_2 - \omega_2}^A(\mathbf{r}_6, \mathbf{r}_7) G_{\varepsilon - \omega_2}^R(\mathbf{r}_7, \mathbf{r}_8) \rangle. \end{aligned} \quad (106)$$

Its singular contribution due to the pairing (b) is given by

$$\langle \tilde{K} \rangle_{\text{sing}}^{(b)} = 16(\pi\nu)^8 \mathcal{Z} D(\mathbf{r}_1, \mathbf{r}_2) D(\mathbf{r}_3, \mathbf{r}_4) D(\mathbf{r}_5, \mathbf{r}_6) D(\mathbf{r}_7, \mathbf{r}_8), \quad (107)$$

Tracing the diffusons and taking into account four possibilities (102), we get

$$c_{YY}^{(b)} = \frac{1}{N_s^4 E_2^4}. \quad (108)$$

Finally, adding the contributions (103), (105) and (108) we obtain the final expression for mesoscopic fluctuations of the inelastic rate in the limit  $F \rightarrow 0$ :

$$\begin{aligned} \langle\langle \gamma^2(\varepsilon, T) \rangle\rangle = & \frac{\lambda^4 \Delta^5}{4\pi^3 N_s^4} \left( \frac{3N_s^2 + 1}{E_2^4} + \frac{2N_s^2}{E_4^4} \right) \\ & \times \int d\varepsilon_1 d\omega_1 \frac{(\mathcal{F}_{\varepsilon_1} - \mathcal{F}_{\varepsilon_1 - \omega_1})^2 (\mathcal{B}_{\omega_1} + \mathcal{F}_{\varepsilon - \omega_1})^2}{\gamma_0(\varepsilon - \omega_1, T) + \gamma_0(\varepsilon_1, T) + \gamma_0(\varepsilon_1 - \omega_1, T)}. \end{aligned} \quad (109)$$

## 8. Mesoscopic fluctuations at arbitrary $F$

### 8.1. Account for a finite $F$ in the sigma-model language

In this Section we generalize the result (109) to the case of an arbitrary interaction parameter  $F$  and derive the general expression (11). First, we explain the main idea with an example of the type-(a) energy pairing in the diagram for  $X_4X'_4$ , and then apply it to the other diagrams ( $X_4Y'_4$  and  $Y_4Y'_4$ ) and energy pairing (b).

When the range of the SSI (4) characterized by the function  $\delta_{\kappa}(\mathbf{r})$  becomes comparable to the Fermi wave length (i. e.,  $F$  becomes non-negligible), a simple sigma-model expression (79) for the correlator  $K$  breaks down. To modify it one has to turn back to the intermediate step in the derivation of the sigma model, prior to the final integration over the 16-component supervector  $\psi$  used to represent Green functions in the functional form. At this stage, the correlator  $K$  [Eq. (75)] is written as

$$\begin{aligned} K = & \int \langle \psi_{R1}(\mathbf{r}_2) \psi_{R1}^*(\mathbf{r}_1) \psi_{A1}(\mathbf{r}_6) \psi_{A1}^*(\mathbf{r}_5) \psi_{R2}(\mathbf{r}_5) \psi_{R2}^*(\mathbf{r}_6) \psi_{A2}(\mathbf{r}_1) \psi_{A2}^*(\mathbf{r}_2) \\ & \times \psi_{R3}(\mathbf{r}_4) \psi_{R3}^*(\mathbf{r}_3) \psi_{A3}(\mathbf{r}_8) \psi_{A3}^*(\mathbf{r}_7) \psi_{R4}(\mathbf{r}_7) \psi_{R4}^*(\mathbf{r}_8) \psi_{A4}(\mathbf{r}_3) \psi_{A4}^*(\mathbf{r}_4) \rangle_{\psi} e^{-S[Q]} DQ, \end{aligned} \quad (110)$$

where the bosonic components of the superfield  $\psi$  are implied. Averaging over  $\psi$  should be performed with the help of Wick's theorem with the correlation function

$$\langle \psi(\mathbf{r}) \psi^*(\mathbf{r}') \rangle = -ig(\mathbf{r}, \mathbf{r}') \Lambda, \quad (111)$$

where  $g$  is the supermatrix Green function defined as

$$g(\mathbf{r}, \mathbf{r}') = \langle \mathbf{r} | (\hat{E} - H_0 + iQ/2\tau)^{-1} | \mathbf{r}' \rangle. \quad (112)$$

The result of the  $\psi$ -averaging is sensitive to the distance between the points  $\mathbf{r}_i$  controlled by the spatial range of the SSI propagators through the factors  $\delta_\kappa(\mathbf{r}_i - \mathbf{r}_j)$  in Eq. (66). Since the interaction is assumed to be short-range on the scale of the system size  $L$ , only correlations between  $\psi$ 's coupled to the same SSI propagator should be taken into account. Therefore the  $\psi$ -averaging in Eq. (110) factorizes into four contributions corresponding to four SSI propagators in Eq. (66):

$$K = \int K^{(1,3)} K^{(2,4)} K^{(5,7)} K^{(6,8)} e^{-S[Q]} DQ. \quad (113)$$

Consider for example the group of four  $\psi$ 's coupled by the interaction with  $\delta_\kappa(\mathbf{r}_1 - \mathbf{r}_3)$ :

$$K^{(1,3)} = \langle \psi_{R1}^*(\mathbf{r}_1) \psi_{A2}(\mathbf{r}_1) \psi_{R3}^*(\mathbf{r}_3) \psi_{A4}(\mathbf{r}_3) \rangle_\psi. \quad (114)$$

Application of Wick's theorem generates two terms:

$$K^{(1,3)} = -g_{A2,R1}(\mathbf{r}_1, \mathbf{r}_1) g_{A4,R3}(\mathbf{r}_3, \mathbf{r}_3) - g_{A2,R3}(\mathbf{r}_1, \mathbf{r}_3) g_{A4,R1}(\mathbf{r}_3, \mathbf{r}_1). \quad (115)$$

The first (diagonal) term in Eq. (115) reduces to the  $Q$  matrix due to the self-consistency equation  $g(\mathbf{r}, \mathbf{r}) = -i\pi\nu Q(\mathbf{r})$  [22], and then one immediately recovers Eq. (79) used previously in the analysis of the  $F = 0$  case. In contrast, the second (off-diagonal) term in Eq. (115) cannot be so easily expressed in terms of the  $Q$  matrix. But here we can use the knowledge that at the later stage, in the process of averaging over fast modes (see Section 6.3), each  $Q$  will be expanded to the first power of  $W$  [see Eq. (80), (81)]. The linear-in- $W$  contribution to  $g$  is given by:

$$\delta g(\mathbf{r}_1, \mathbf{r}_3) = -\frac{i}{2\tau} \int d\mathbf{r}' \langle \mathbf{r}_1 | (-H_0 + iQ/2\tau)^{-1} | \mathbf{r}' \rangle T^{-1} \Lambda W(\mathbf{r}') T \langle \mathbf{r}' | (-H_0 + iQ/2\tau)^{-1} | \mathbf{r}_3 \rangle, \quad (116)$$

where we write  $Q = T^{-1} \Lambda T$ , and neglect the term  $\hat{E}$  as we are working in the diffusive limit  $\max\{\varepsilon, T\}\tau \ll 1$ . Since  $W$  anticommutes with  $\Lambda$ , we can rewrite Eq. (116) in the form

$$\begin{aligned} \delta g(\mathbf{r}_1, \mathbf{r}_3) &= -\frac{i}{2\tau} \int d\mathbf{r}' T^{-1} \langle \mathbf{r}_1 | (-H_0 + i\Lambda/2\tau)^{-1} | \mathbf{r}' \rangle \langle \mathbf{r}' | (-H_0 - i\Lambda/2\tau)^{-1} | \mathbf{r}_3 \rangle \Lambda W T \\ &= -\frac{i\delta Q}{2\tau} (G^R G^A)(\mathbf{r}_1 - \mathbf{r}_3), \end{aligned} \quad (117)$$

where  $\delta Q = T^{-1} \Lambda W T$ , and  $G^R G^A$  is a scalar function given by Eq. (A.3).

Multiplying the second term of Eq. (115) by  $\delta_\kappa(\mathbf{r}_1 - \mathbf{r}_3)$  and integrating over the difference  $\mathbf{r}_1 - \mathbf{r}_3$ , we immediately recover the factor  $F$  introduced in Eq. (A.2). Hence, Eq. (115) can be written as

$$K^{(1,3)} = (\pi\nu)^2 [Q_{21}^{AR}(\mathbf{r}_1) Q_{43}^{AR}(\mathbf{r}_3) + F Q_{23}^{AR}(\mathbf{r}_1) Q_{41}^{AR}(\mathbf{r}_3)], \quad (118)$$

where with our accuracy we do not distinguish between  $\mathbf{r}_1$  and  $\mathbf{r}_3$ . Generally speaking, Eq. (118) is incorrect. We write it in such a form for brevity, assuming that it will be further subject to the procedure of fast mode extraction described in Section 6.3.

Performing the same analysis for other  $K^{(i,j)}$  from Eq. (113), one concludes that for the purpose of calculation of mesoscopic fluctuations of the relaxation rate at  $F \neq 0$ , Eq. (79) should be modified as

$$\begin{aligned} K &= (\pi\nu)^8 \int [Q_{21}^{AR}(\mathbf{r}_1) Q_{43}^{AR}(\mathbf{r}_3) + F Q_{23}^{AR}(\mathbf{r}_1) Q_{41}^{AR}(\mathbf{r}_3)] [Q_{12}^{RA}(\mathbf{r}_2) Q_{34}^{RA}(\mathbf{r}_4) + F Q_{14}^{RA}(\mathbf{r}_2) Q_{32}^{RA}(\mathbf{r}_4)] \\ &\quad \times [Q_{21}^{RA}(\mathbf{r}_5) Q_{43}^{RA}(\mathbf{r}_7) + F Q_{23}^{RA}(\mathbf{r}_5) Q_{41}^{RA}(\mathbf{r}_7)] \\ &\quad \times [Q_{12}^{AR}(\mathbf{r}_6) Q_{34}^{AR}(\mathbf{r}_8) + F Q_{14}^{AR}(\mathbf{r}_6) Q_{32}^{AR}(\mathbf{r}_8)] e^{-S[Q]} DQ, \end{aligned} \quad (119)$$

where the difference between the coordinates within the same bracket can be neglected. Similar expressions originate in the analysis of other contributions from  $X_4 X'_4$ ,  $X_4 Y'_4$  and  $Y_4 Y'_4$ .

## 8.2. Contributions of different diagrams and energy pairings

To find mesoscopic fluctuations at a finite  $F$ , one should use Eq. (119) as a starting point and perform all the steps outlined in Sections 6 and 7. This is a routine procedure leading to the following results. In the case  $F \neq 0$ , there are finite contributions from all the terms  $X_4X'_4$ ,  $X_4Y'_4$ , and  $Y_4Y'_4$ , and from both energy pairings (a) and (b). These six contributions are characterized by six coefficients  $c_A^{(i)}$  in Eq. (99). After some algebra, we obtain (the coefficients  $c_{XY}^{(i)}$  already contain the factor 2 accounting for the mirror term  $Y_4X'_4$ ):

$$c_{XX}^{(a)} = \frac{2(1+F^2)^2}{N_s^2 E_2^4} + \frac{2(1+F^4)}{N_s^2 E_4^4}, \quad c_{XX}^{(b)} = \frac{1+6F^2+F^4}{N_s^2 E_2^4} + \frac{4F^2}{N_s^2 E_4^4}, \quad (120)$$

$$c_{XY}^{(a)} = c_{XY}^{(b)} = -\frac{8(F+F^3)}{N_s^3 E_2^4} - \frac{4(F+F^3)}{N_s^3 E_4^4}, \quad (121)$$

$$c_{YY}^{(a)} = \frac{8F^2}{N_s^4 E_2^4} + \frac{4F^2}{N_s^4 E_4^4}, \quad c_{YY}^{(b)} = \frac{1+6F^2+F^4}{N_s^4 E_2^4} + \frac{4F^2}{N_s^4 E_4^4}. \quad (122)$$

Summing the contributions (120)–(122), we come to the final result (11) valid for an arbitrary  $F$ .

## 9. Discussion and conclusion

In this work we have analyzed the applicability of the Fermi-golden-rule description of the initial stage of quasiparticle decay in diffusive quantum dots in the regime when the single-particle levels are already resolved. Approaching the problem from the high energy/temperature side, where each energy level can be characterized by a Lorentzian width  $\gamma_0(\varepsilon, T)$ , we have calculated mesoscopic fluctuations of the energy relaxation rate. The leading contribution to fluctuations comes from the diagrams which describe the square of the same decay process, i.e. have the same set of final states. The resulting expression is non-perturbative in  $\gamma_0$ , which appears in the denominator of Eq. (11), ensuring the growth of fluctuations with the decrease of the excitation energy and/or temperature.

Quantum relaxation of the initial state  $|i\rangle$  can be described by the return probability

$$P(t) = \overline{|\langle i|e^{-iHt}|i\rangle|^2}, \quad (123)$$

where  $H$  is the Hamiltonian of the interacting quantum dot, and the bar stands for the thermal average. In the semiclassical FGR picture this is just a pure exponential decay:

$$P_{\text{FGR}}(t) = e^{-\gamma(\varepsilon, T)t}, \quad (124)$$

where the rate  $\gamma(\varepsilon, T)$  depends on a particular disorder realization. Its average value is given by Eq. (7), and we focused on its mesoscopic fluctuations. We found that the FGR description of the initial stage of quasiparticle decay is applicable as long as  $\max\{\varepsilon, T\} \gg \varepsilon_{\text{FGR}}$ . In this limit, each level is characterized by a well-defined energy width  $\gamma(\varepsilon, T)$ , which weakly fluctuates near the FGR mean:

$$\frac{\langle\langle \gamma^2(\varepsilon, T) \rangle\rangle}{\gamma_0^2(\varepsilon, T)} \sim \left( \frac{\varepsilon_{\text{FGR}}}{\max\{\varepsilon, T\}} \right)^4. \quad (125)$$

The temperature and the excitation energy enter the result in a similar way [note, however, the presence of the factor  $\mathcal{T}(\varepsilon/2T)$  in Eq. (14), that can change by two orders of magnitude]. This fact is a consequence of the lowest-order approximation. In higher orders their role is expected to be different, in accordance with the difference between the relaxation dynamics in the hot-electron and thermal problems [17].

It is important that in the range of applicability of the FGR description,  $\max\{\varepsilon, T\} \gg \varepsilon_{\text{FGR}}$ , the hybridization with distant generations and eventually many-body localization effects (if any) become relevant at sufficiently large time scales. The characteristic time  $t_*$  when  $P_{\text{FGR}}(t)$  crosses over to a

weaker dependence is determined by the initial state. In the limit  $\max\{\varepsilon, T\} \gg \varepsilon_{\text{FGR}}$ , the scale  $t_*$  satisfies  $\gamma_0(\varepsilon, T)t_* \gg 1$ , indicating that almost all the quasiparticle weight is lost during the FGR exponential relaxation (for the hot-electron problem,  $\gamma_0(\varepsilon, 0)t_* \sim \ln(\varepsilon/\varepsilon_{\text{FGR}})$  [18]).

Our approach is conceptually similar to the one used by Basko, Aleiner and Altshuler (BAA) [14], and hence it is instructive to compare the two results. BAA considered inelastic relaxation in a chaotic quantum dot (in Ref. [14], referred to as the localization cell), working in the basis of exact one-particle states  $|\alpha\rangle$  and treating electron–electron interaction phenomenologically. Interaction matrix elements,  $\langle\alpha\beta|V|\gamma\delta\rangle$ , were assumed to be independent normally distributed random variables with the standard deviation  $\lambda_{\text{BAA}}\Delta$  for energy difference smaller than the ultraviolet cutoff  $M\Delta$ , and zero otherwise. The FGR relaxation rate is then given by  $\gamma_{\text{BAA}}(T) \sim \lambda_{\text{BAA}}^2 MT$ , and BAA obtained the following expression for mesoscopic fluctuations:

$$\frac{\langle\langle\gamma^2(T)\rangle\rangle_{\text{BAA}}}{\gamma_{\text{BAA}}^2(T)} \sim \frac{\lambda_{\text{BAA}}^4 M \Delta^2 T}{\gamma_{\text{BAA}}^3(T)} \sim \frac{\Delta^2}{\lambda_{\text{BAA}}^2 M^2 T^2}. \quad (126)$$

In order to apply these results to the case of a diffusive quantum dot, one should put  $\lambda_{\text{BAA}} \sim \lambda\Delta/E_{\text{Th}}$  [5] and  $M \sim T/\Delta$ . Then  $\gamma_{\text{BAA}}(T)$  coincides with the Sivan–Imry–Aronov relaxation rate (7), while the estimate (126) reproduces our result (125) for mesoscopic fluctuations.

Finally, we emphasize that our result (11) for the leading contribution to mesoscopic fluctuations provides an exact account for the (screened) electron–electron interaction in a diffusive quantum dot. It has an important implementation for the statistics of the interaction matrix elements  $\langle\alpha\beta|V|\gamma\delta\rangle$  in the lattice-model language. Since Ref. [5], it is usually assumed for simplicity that this statistics is Gaussian. However, our result (11) demonstrates that for a real diffusive quantum dot such an assumption is generally incorrect. Indeed, for the Gaussian statistics all correlators of the four matrix elements in  $\langle\langle\gamma^2(\varepsilon, T)\rangle\rangle$  can be expressed through pairwise correlators (the Wick theorem), which are known to be determined by  $E_2$  only [3,5]. Therefore the presence of the quantity  $E_4$  in Eq. (11) indicates that the statistics of the interaction matrix elements in a quantum dot is essentially non-Gaussian. This fact should be taken into account in constructing the theory of many-body localization in quantum dots.

## Acknowledgments

We thank I. Aleiner, Ya.M. Blanter, M.V. Feigel'man, I.V. Gornyi, V.E. Kravtsov, A.D. Mirlin, and A. Silva for useful discussions. This work was supported by the Russian Science Foundation under Grant No. 14-42-00044 (M.A.S.).

## Appendix A. Non-RPA averaging of the relaxation rate

In this Appendix we consider the non-RPA diagrams 5(b)–(d) for the average energy relaxation rate. The main difference from the RPA diagram 5(a) is that now one [5(c) and (d)] or both [5(b)] SSI lines are coupled to diffusons via the square box

$$I = \begin{array}{c} \text{R} \quad \text{A} \\ \diagdown \quad \diagup \\ \text{A} \quad \text{R} \end{array} = \int d\mathbf{r} [(G^{\text{R}}G^{\text{A}})(r)]^2 V(r), \quad (A.1)$$

where solid lines denote electron Green functions and the zigzag line stands for SSI. Contrary to the RPA diagram 5(a), where the momentum of the SSI line coincides with the slow ( $q \sim 1/L$ ) diffuson momentum, in the box configuration the interaction momentum can be comparable to  $p_F$  typically carried by the Green functions. The ratio of the diagram with the square box (A.1) to a similar diagram with the RPA averaging is characterized by the parameter

$$F = \frac{I}{(2\pi\nu\tau)^2 V(q=0)} = \frac{1}{(2\pi\nu\tau)^2} \int d\mathbf{r} [(G^{\text{R}}G^{\text{A}})(r)]^2 \delta_k(\mathbf{r}). \quad (A.2)$$

where the smeared delta function  $\delta_\kappa(\mathbf{r})$  is defined in Eq. (4).

The real-space representation of  $(G^R G^A)(r)$  in the limit of a good metal,  $p_F l \gg 1$ , is given by

$$(G^R G^A)(r) = \int \frac{e^{i\mathbf{p}\mathbf{r}}}{\xi^2 + 1/4\tau^2} \frac{d\mathbf{p}}{(2\pi)^d} = 2\pi v\tau j_d(p_F r) e^{-r/2l}, \quad (\text{A.3})$$

where  $\tau$  and  $l$  are the mean free time and mean free path correspondingly,  $\xi = (p^2 - p_F^2)/2m$ , and the function  $j_d(x) = \langle e^{ixn_1} \rangle_{\mathbf{n}}$  describes Friedel oscillations in  $d$  dimensions:  $j_2(x) = J_0(x)$  and  $j_3(x) = \sin x/x$  [ $J_0(x)$  is the Bessel function]. Substituting Eq. (A.3) into Eq. (A.2) we obtain for  $F$ :

$$F = \int j_d^2(p_F r) e^{-r/l} \delta_\kappa(\mathbf{r}) d\mathbf{r} = \int v_\kappa(\mathbf{q}) \frac{d\mathbf{q}}{(2\pi)^d} \int j_d^2(p_F r) e^{i\mathbf{q}\mathbf{r}} e^{-r/l} d\mathbf{r}. \quad (\text{A.4})$$

Then we average over angles and take the leading term in  $p_F l \gg 1$ :

$$F = \int v_\kappa(\mathbf{q}) \frac{d\mathbf{q}}{(2\pi)^d} \int j_d^2(p_F r) j_d(qr) d\mathbf{r}. \quad (\text{A.5})$$

Evaluating the integral over  $\mathbf{r}$  we arrive at Eq. (5).

In the case of the Coulomb interaction,  $v_\kappa(\mathbf{q})$  is given by Eq. (30) and we obtain [38,42]

$$F_{3D}^{\text{Coul}} = \left( \frac{\kappa}{2p_F} \right)^2 \ln \left[ 1 + \left( \frac{2p_F}{\kappa} \right)^2 \right], \quad F_{2D}^{\text{Coul}} = \frac{2 \operatorname{arccosh}(2p_F/\kappa)}{\pi \sqrt{(2p_F/\kappa)^2 - 1}}. \quad (\text{A.6})$$

## Appendix B. Expressions for $Y_{abcd}(\varepsilon, \varepsilon', \omega, \mathbf{q})$

Here we present explicit expressions for the effective interaction propagators  $Y_{abcd}(\varepsilon, \varepsilon', \omega, \mathbf{q})$  defined in Eq. (41). All components of  $V = V(\omega, \mathbf{q})$  are taken at frequency  $\omega$  and momentum  $\mathbf{q}$ .

$$Y_{RR,RR} = V^K + \mathcal{F}_{\varepsilon' - \omega} V^R + \mathcal{F}_\varepsilon V^A, \quad (\text{B.1})$$

$$Y_{RR,RA} = (\mathcal{F}_{\varepsilon'} - \mathcal{F}_{\varepsilon' - \omega})(V^K + \mathcal{F}_\varepsilon V^A) - (1 - \mathcal{F}_{\varepsilon'} \mathcal{F}_{\varepsilon' - \omega}) V^R, \quad (\text{B.2})$$

$$Y_{RR,AR} = -V^R, \quad (\text{B.3})$$

$$Y_{RR,AA} = V^K - \mathcal{F}_{\varepsilon'} V^R + \mathcal{F}_\varepsilon V^A, \quad (\text{B.4})$$

$$Y_{RA,RR} = -(\mathcal{F}_\varepsilon - \mathcal{F}_{\varepsilon - \omega})(V^K + \mathcal{F}_{\varepsilon' - \omega} V^R) - (1 - \mathcal{F}_\varepsilon \mathcal{F}_{\varepsilon - \omega}) V^A, \quad (\text{B.5})$$

$$Y_{RA,RA} = -(\mathcal{F}_\varepsilon - \mathcal{F}_{\varepsilon - \omega})(\mathcal{F}_{\varepsilon'} - \mathcal{F}_{\varepsilon' - \omega}) V^K + (\mathcal{F}_\varepsilon - \mathcal{F}_{\varepsilon - \omega})(1 - \mathcal{F}_{\varepsilon'} \mathcal{F}_{\varepsilon' - \omega}) V^R \\ - (1 - \mathcal{F}_\varepsilon \mathcal{F}_{\varepsilon - \omega})(\mathcal{F}_{\varepsilon'} - \mathcal{F}_{\varepsilon' - \omega}) V^A, \quad (\text{B.6})$$

$$Y_{RA,AR} = (\mathcal{F}_\varepsilon - \mathcal{F}_{\varepsilon - \omega}) V^R, \quad (\text{B.7})$$

$$Y_{RA,AA} = -(\mathcal{F}_\varepsilon - \mathcal{F}_{\varepsilon - \omega})(V^K - \mathcal{F}_{\varepsilon'} V^R) - (1 - \mathcal{F}_\varepsilon \mathcal{F}_{\varepsilon - \omega}) V^A, \quad (\text{B.8})$$

$$Y_{AR,RA} = -(\mathcal{F}_{\varepsilon'} - \mathcal{F}_{\varepsilon' - \omega}) V^A, \quad (\text{B.9})$$

$$Y_{AR,AR} = 0, \quad (\text{B.10})$$

$$Y_{AR,AA} = -V^A, \quad (\text{B.11})$$

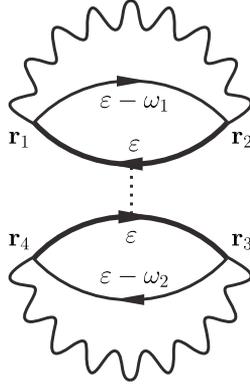
$$Y_{AR,RR} = -V^A, \quad (\text{B.12})$$

$$Y_{AA,RA} = (\mathcal{F}_{\varepsilon'} - \mathcal{F}_{\varepsilon' - \omega})(V^K - \mathcal{F}_{\varepsilon - \omega} V^A) - (1 - \mathcal{F}_{\varepsilon'} \mathcal{F}_{\varepsilon' - \omega}) V^R, \quad (\text{B.13})$$

$$Y_{AA,AA} = V^K - \mathcal{F}_{\varepsilon'} V^R - \mathcal{F}_{\varepsilon - \omega} V^A, \quad (\text{B.14})$$

$$Y_{AA,RR} = V^K - \mathcal{F}_{\varepsilon - \omega} V^A + \mathcal{F}_{\varepsilon' - \omega} V^R, \quad (\text{B.15})$$

$$Y_{AA,AR} = -V^R. \quad (\text{B.16})$$



**Fig. 7.** The reducible part of mesoscopic fluctuations of the relaxation rate,  $\langle \gamma^2(\varepsilon, T) \rangle_{\text{red}}$ , in terms of the dynamically screened interaction [cf. Fig. 3(b)]. The dotted line denotes non-perturbative averaging of  $G_\varepsilon G_\varepsilon$  over disorder.

### Appendix C. Reducible part of the correlator $\langle \gamma^2(\varepsilon, T) \rangle$

In this methodological Appendix we show how to extract the reducible part of the correlator  $\langle \gamma^2(\varepsilon, T) \rangle$  from the general expression (64) and demonstrate that it coincides with the square  $\gamma_0^2(\varepsilon, T)$  of the average relaxation rate given by Eq. (7). For simplicity, we consider the case of a sufficiently long-range interaction with  $F = 0$ , when only the term  $X_4 X_4'$  contributes to Eq. (64).

In the formalism of kinetic equation used in our analysis, extraction of the reducible part should be done with care, and a naive independent averaging of  $X_4$  and  $X_4'$  would give a wrong answer. The reason is that the product  $\Delta G_\varepsilon \Delta G_\varepsilon$  contained both on the left- and right-hand sides of Eq. (64) requires non-perturbative treatment. Disorder averaging on the left-hand side of Eq. (64) is discussed in Section 6.1, while on the right-hand side one has to take into account non-perturbative averaging between Green functions with the energies  $\varepsilon$  only. In the language of the sigma model developed in Section 6.2, this means that after averaging over fast modes non-perturbative integration over the sigma-model manifold should be performed only in the sector 1 of the 1234 space, whereas in all other sectors one can simply take the saddle-point value  $Q_j = \Lambda$ .

This observation significantly simplifies further calculation, as now we can use the representation of Eqs. (45) and (48) in terms of the dynamically screened RPA interaction  $\Delta V = V^R - V^A = 2i \text{Im} V^R$  instead of that of Eq. (64) with the statically screened interaction. The corresponding reducible contribution to  $\langle \gamma^2(\varepsilon, T) \rangle$  then should be extracted from

$$\begin{aligned} & \langle \gamma^2(\varepsilon, T) \rangle_{\text{red}} \int d\mathbf{r} d\mathbf{r}' \langle \Delta G_\varepsilon(\mathbf{r}, \mathbf{r}) \Delta G_\varepsilon(\mathbf{r}', \mathbf{r}') \rangle \\ &= \frac{1}{4} \int d\mathbf{r}_1 \dots d\mathbf{r}_4 (d\omega_1)(d\omega_2) (\mathcal{B}_{\omega_1} + \mathcal{F}_{\varepsilon-\omega_1}) (\mathcal{B}_{\omega_2} + \mathcal{F}_{\varepsilon-\omega_2}) \Delta V(\omega_1, \mathbf{r}_1 - \mathbf{r}_2) \\ & \quad \times \Delta V(\omega_2, \mathbf{r}_3 - \mathbf{r}_4) \langle \tilde{X}_4 \rangle, \end{aligned} \quad (\text{C.1})$$

where

$$\tilde{X}_4 = \Delta G_\varepsilon(\mathbf{r}_2, \mathbf{r}_1) \Delta G_{\varepsilon-\omega_1}(\mathbf{r}_1, \mathbf{r}_2) \Delta G_\varepsilon(\mathbf{r}_4, \mathbf{r}_3) \Delta G_{\varepsilon-\omega_2}(\mathbf{r}_3, \mathbf{r}_4), \quad (\text{C.2})$$

and  $\langle \dots \rangle$  implies now averaging over minimal number (two) of fast modes and exact non-perturbative averaging of  $G_\varepsilon(\mathbf{r}_2, \mathbf{r}_1) G_\varepsilon(\mathbf{r}_4, \mathbf{r}_3)$ , see the diagram in Fig. 7.

Before we proceed, we highlight the important subtlety regarding R–A counting. Consider first independent averaging over disorder in each bubble, when correlations between  $G_\varepsilon(\mathbf{r}_2, \mathbf{r}_1)$  and  $G_\varepsilon(\mathbf{r}_4, \mathbf{r}_3)$  are neglected. This is precisely the situation encountered in Section 4.1, see Eq. (51) and Fig. 3(c). To organize a diffuson in each bubble, one needs one  $G^R$  and one  $G^A$ . Therefore only four out of sixteen terms would contribute to  $\langle \Delta G_\varepsilon(\mathbf{r}_2, \mathbf{r}_1) \Delta G_{\varepsilon-\omega_1}(\mathbf{r}_1, \mathbf{r}_2) \rangle \langle \Delta G_\varepsilon(\mathbf{r}_4, \mathbf{r}_3) \Delta G_{\varepsilon-\omega_2}(\mathbf{r}_3, \mathbf{r}_4) \rangle$ :

RARA, RAAR, ARRA, ARAR. The situation changes if we consider non-perturbative correlations between  $G_\varepsilon(\mathbf{r}_2, \mathbf{r}_1)$  and  $G_\varepsilon(\mathbf{r}_4, \mathbf{r}_3)$ . In this case, only RAAR and ARRA from the previous four contributions survive. At the same time, now we have to take into account six additional terms: RRAR, RAAA, RRAA, ARRR, AARA, AARR. It turns out that each of these terms admits configuration with two fast diffusons, and, as a result, contributes equally to  $\langle \gamma^2(\varepsilon, T) \rangle_{\text{red}}$ .

To perform averaging over disorder, we generalize the technique developed in Sections 6 and 7. The pair coupling introducing the 12 space is obvious:  $\{G_\varepsilon(\mathbf{r}_2, \mathbf{r}_1), G_\varepsilon(\mathbf{r}_4, \mathbf{r}_3)\}$  and  $\{G_{\varepsilon-\omega_1}(\mathbf{r}_1, \mathbf{r}_2), G_{\varepsilon-\omega_2}(\mathbf{r}_3, \mathbf{r}_4)\}$ . After perturbative averaging over fast modes and non-perturbative averaging over zero modes we obtain

$$\langle G^R G^R G^A G^R \rangle = -2(\pi\nu)^2 X_1 D(\mathbf{r}_1, \mathbf{r}_2) D(\mathbf{r}_3, \mathbf{r}_4), \quad (\text{C.3})$$

$$\langle G^R G^A G^A G^R \rangle = (\pi\nu)^2 [(4 + 2X_1 + 2X_2 + 4X_1 X_2) D(\mathbf{r}_1, \mathbf{r}_2) D(\mathbf{r}_3, \mathbf{r}_4) + 4Z_1 Z_2 D(\mathbf{r}_1, \mathbf{r}_4) D(\mathbf{r}_2, \mathbf{r}_3)], \quad (\text{C.4})$$

$$\langle G^R G^R G^A G^A \rangle = (\pi\nu)^2 [(4 + 2X_1 + 2X_2^* + 4X_1 X_2^*) D(\mathbf{r}_1, \mathbf{r}_2) D(\mathbf{r}_3, \mathbf{r}_4) + 4Z_1 Z_2^* D(\mathbf{r}_1, \mathbf{r}_4) D(\mathbf{r}_2, \mathbf{r}_3)], \quad (\text{C.5})$$

$$\langle G^R G^A G^A G^A \rangle = -2(\pi\nu)^2 X_1 D(\mathbf{r}_1, \mathbf{r}_2) D(\mathbf{r}_3, \mathbf{r}_4). \quad (\text{C.6})$$

Here the order of Green functions is the same as in Eq. (C.2), and the factors  $X_j, Z_j$  are defined in Eqs. (90)–(92). There are also four equations which are complex conjugate to Eqs. (C.3)–(C.6) (obtained by replacing  $G^R \leftrightarrow G^A$ ).

As the energies in the second pair,  $\{\varepsilon - \omega_1, \varepsilon - \omega_2\}$ , are not supposed to coincide, the terms with  $X_2$  and  $Z_2$  describing correlations between  $G_{\varepsilon-\omega_1}(\mathbf{r}_1, \mathbf{r}_2)$  and  $G_{\varepsilon-\omega_2}(\mathbf{r}_3, \mathbf{r}_4)$  contain additional smallness of  $\Delta / \max\{\varepsilon, T\}$  and thus can be safely neglected. Retaining only the terms with  $X_1 = \Delta / \pi \gamma(\varepsilon) \gg 1$ , we obtain that each average (C.3)–(C.6) as well as its complex conjugate, in the leading order, contributes  $\gamma_0^2(\varepsilon, T) / 8$  to  $\langle \gamma^2(\varepsilon, T) \rangle_{\text{red}}$ . Hence their sum equals  $\langle \gamma^2(\varepsilon, T) \rangle_{\text{red}} = \gamma_0^2(\varepsilon, T)$ , reproducing the square of the average relaxation rate, as expected.

## Appendix D. Diffuson with the electron–electron interaction

In Section 5, we introduced the inelastic single-particle level broadening in Eq. (68). This energy level width plays an important role as a regularizer of the otherwise divergent delta functions in the process of non-perturbative averaging over disorder [see, e.g., Eq. (98)]. As has been pointed out in Ref. [43], the same quantity  $\gamma_0(\varepsilon, T)$  determines the width of the single-particle electron Green function (that was calculated in Section 4) and the mass of two-electron particle–particle (cooperon) [44,45] and particle–hole (diffuson) [3,43,46] propagators in the presence of interaction. To make this paper comprehensive, we rederive this result in the framework of the modified Keldysh technique introduced in Section 3.1. We consider the zero-momentum diffuson and assume Coulomb interaction in the limit  $F \rightarrow 0$ .

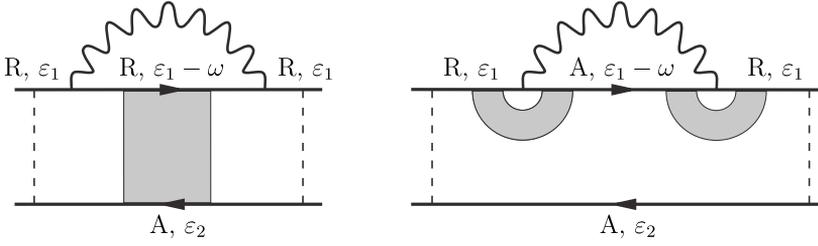
The important contribution comes from the elastic processes only (processes that do not lead to the energy exchange between  $G_{\varepsilon_1}^R$  and  $G_{\varepsilon_2}^A$ ). As a result, the diffuson  $D_{\text{el}}(\varepsilon_1, \varepsilon_2)$  can be found from the algebraic (not integral) Dyson equation:

$$D_{\text{el}}^R(\varepsilon_1, \varepsilon_2) = D_0^R(\varepsilon_1 - \varepsilon_2, 0) + D_0^R(\varepsilon_1 - \varepsilon_2, 0) \Pi(\varepsilon_1, \varepsilon_2) D_{\text{el}}^R(\varepsilon_1, \varepsilon_2). \quad (\text{D.1})$$

Here  $D_0^R(\omega, \mathbf{q})$  is the bare diffuson, defined in Eq. (51), and  $\Pi(\varepsilon_1, \varepsilon_2)$  is the elastic part of the self-energy (also at zero momentum). In the lowest order,  $\Pi(\varepsilon_1, \varepsilon_2)$  is given by the diagrams shown in Fig. 8 and their counterparts where the interaction line is coupled to  $G^A$ .

The contribution of the diagrams in Fig. 8 equals

$$\begin{aligned} \Pi_1(\varepsilon_1, \varepsilon_2) = & -\frac{i}{2} \int (d\omega)(d\mathbf{q}) D_0^R(\varepsilon_1 - \varepsilon_2 - \omega, \mathbf{q}) (V^K + \mathcal{F}_{\varepsilon_1 - \omega} V^R + \mathcal{F}_{\varepsilon_1} V^A) \\ & + \frac{i}{2} \int (d\omega)(d\mathbf{q}) [Dq^2 - i(\varepsilon_1 - \varepsilon_2 + \omega)] (D_0^R(\omega, \mathbf{q}))^2 (\mathcal{F}_{\varepsilon_1} - \mathcal{F}_{\varepsilon_1 - \omega}) V^R. \end{aligned} \quad (\text{D.2})$$



**Fig. 8.** Diagrams for the diffuson self-energy  $\Pi(\varepsilon_1, \varepsilon_2)$ , with the interaction line inserted into  $D_0^R$  (there are mirror counterparts with the interaction inserted into  $G^A$ ). The bare diffuson  $D_0(\omega, \mathbf{q})$  is depicted by the shaded block. The dashed lines denote the beginning of external diffusons which are not included into the self-energy part. The diagram (b) admits insertion of an impurity line in two ways (not shown), producing the standard Hikami box as an interaction vertex of four diffusons [47].

The contribution of the analogous diagrams with the interaction line inserted to  $G^A$  can be written as

$$\begin{aligned} \Pi_2(\varepsilon_1, \varepsilon_2) = & -\frac{i}{2} \int (d\omega)(d\mathbf{q}) D_0^R(\varepsilon_1 - \varepsilon_2 - \omega, \mathbf{q}) (V^K - \mathcal{F}_{\varepsilon_2} V^A - \mathcal{F}_{\varepsilon_2 + \omega} V^R) \\ & + \frac{i}{2} \int (d\omega)(d\mathbf{q}) [Dq^2 - i(\varepsilon_1 - \varepsilon_2 + \omega)] (D_0^R(\omega, \mathbf{q}))^2 (\mathcal{F}_{\varepsilon_2 + \omega} - \mathcal{F}_{\varepsilon_2}) V^R. \end{aligned} \quad (\text{D.3})$$

In Eqs. (D.2) and (D.3) all interaction propagators are taken at frequency  $\omega$  and momentum  $\mathbf{q}$ . Employing analyticity properties, we obtain for the diffuson self-energy  $\Pi = \Pi_1 + \Pi_2$ :

$$\begin{aligned} \Pi(\varepsilon_1, \varepsilon_2) = & -\frac{i}{2} \int (d\omega)(d\mathbf{q}) D_0^R(\omega, \mathbf{q}) \left\{ 2V^K(\varepsilon_1 - \varepsilon_2 - \omega, \mathbf{q}) \right. \\ & \left. + (\mathcal{F}_{\omega - \varepsilon_1} + \mathcal{F}_{\omega + \varepsilon_2}) [V^R(\varepsilon_1 - \varepsilon_2 - \omega, \mathbf{q}) - V^R(\omega, \mathbf{q}) + i(\varepsilon_1 - \varepsilon_2) D_0^R(\omega, \mathbf{q}) V^R(\omega, \mathbf{q})] \right\}. \end{aligned} \quad (\text{D.4})$$

In the universal limit of Coulomb interaction,  $V^{R(A)}(\omega, \mathbf{q}) = (Dq^2 \mp i\omega)/N_s \nu Dq^2$ , we obtain

$$\begin{aligned} \Pi(\varepsilon_1, \varepsilon_2) = & -\frac{1}{N_s \nu} \int \frac{(d\omega)(d\mathbf{q}) \omega (2\mathcal{B}_\omega - (\mathcal{F}_{\omega - \varepsilon_2} + \mathcal{F}_{\omega + \varepsilon_1}))}{Dq^2 (Dq^2 - i\omega - i(\varepsilon_1 - \varepsilon_2))} \\ & + \frac{\varepsilon_1 - \varepsilon_2}{N_s \nu} \int \frac{(d\omega)(d\mathbf{q}) (\mathcal{F}_{\omega - \varepsilon_2} + \mathcal{F}_{\omega + \varepsilon_1})}{Dq^2 (Dq^2 - i\omega - i(\varepsilon_1 - \varepsilon_2))}. \end{aligned} \quad (\text{D.5})$$

Two terms in this expression have a different nature. In the first term, the integration over  $\omega$  converges at the scale  $\max\{T, \varepsilon_1, \varepsilon_2\}$ , and the summation over momenta converges at  $Dq^2 \sim E_{\text{Th}}$ . In the second term, the main contribution comes from  $E_{\text{Th}} \ll Dq^2 \ll \omega \ll \tau^{-1}$  ( $\tau$  is the mean free time). Assuming the equilibrium distribution function, we obtain

$$\Pi(\varepsilon_1, \varepsilon_2) = -\frac{\Delta}{2\pi N_s E_2^2} (2\pi^2 T^2 + \varepsilon_1^2 + \varepsilon_2^2) + \frac{2(\varepsilon_1 - \varepsilon_2)}{N_s \nu} \int \frac{(d\omega)(d\mathbf{q}) \text{sign } \omega}{Dq^2 (Dq^2 - i\omega)}. \quad (\text{D.6})$$

Using the result (7) for  $\gamma_0(\varepsilon, T)$  (with  $\lambda = 1$  and  $F = 0$ ), we can write

$$\Pi(\varepsilon_1, \varepsilon_2) = -\frac{\gamma_0(\varepsilon_1, T) + \gamma_0(\varepsilon_2, T)}{2} + i(\varepsilon_1 - \varepsilon_2)(Z - 1), \quad (\text{D.7})$$

where  $Z$  is the wave-function renormalization factor [48]:

$$Z - 1 = \frac{2}{\pi N_s \nu} \int_{1/L}^{1/l} \frac{(d\mathbf{q})}{Dq^2} \ln \frac{1}{Dq^2 \tau}, \quad (\text{D.8})$$

$L$  is the linear size of the sample,  $l$  is the mean free path.

Finally, the zero-dimensional diffuson acquires the form

$$D_{\text{el}}^R(\varepsilon_1, \varepsilon_2) = \frac{1}{-i(\varepsilon_1 - \varepsilon_2)Z + [\gamma_0(\varepsilon_1, T) + \gamma_0(\varepsilon_2, T)]/2}. \quad (\text{D.9})$$

Precisely the same form of the diffuson follows from the sigma model (77), provided the wave function renormalization can be neglected,  $Z - 1 \ll 1$ .

### Appendix E. Contraction rules for the zero-dimensional sigma model

In Efetov's parametrization [22], the sigma-model action for  $\langle G_{\varepsilon_1}^R G_{\varepsilon_2}^A \rangle$  [where the Green functions may contain a finite width  $\gamma(E)$ , see Eq. (68)] depends only on the 'radial' variables  $\lambda_F$  and  $\lambda_B$ :

$$S_0 = a (\lambda_B - \lambda_F), \quad (\text{E.1})$$

where

$$a = -\frac{\pi i}{\Delta} \{(\varepsilon_1)_+ - (\varepsilon_2)_-\}, \quad (\text{E.2})$$

and  $\varepsilon_{\pm}$  are defined in Eq. (68). In order to calculate correlation functions, the full form of the parametrization should be employed. After some superalgebra, one can verify that the following non-perturbative contraction rules hold for averaging over the zero-dimensional sigma model with the action (E.1):

$$\langle \text{str } PQ \text{ str } RQ \rangle = \text{str } P\Lambda \text{ str } R\Lambda + Z \text{ str } (PR - P\Lambda R\Lambda) - X (\text{str } P \text{ str } R - \text{str } P\Lambda \text{ str } R\Lambda), \quad (\text{E.3a})$$

$$\langle \text{str } PQRQ \rangle = \text{str } P\Lambda R\Lambda + Z (\text{str } P \text{ str } R - \text{str } P\Lambda \text{ str } R\Lambda) - X \text{ str } (PR - P\Lambda R\Lambda), \quad (\text{E.3b})$$

where  $P$  and  $Q$  are arbitrary supermatrices, and the functions  $X(a)$  and  $Z(a)$  are defined as

$$X(a) = \frac{1}{2} \int_0^1 d\lambda_F \int_1^\infty d\lambda_B e^{-a(\lambda_B - \lambda_F)} = \frac{1 - e^{-2a}}{2a^2}, \quad (\text{E.4})$$

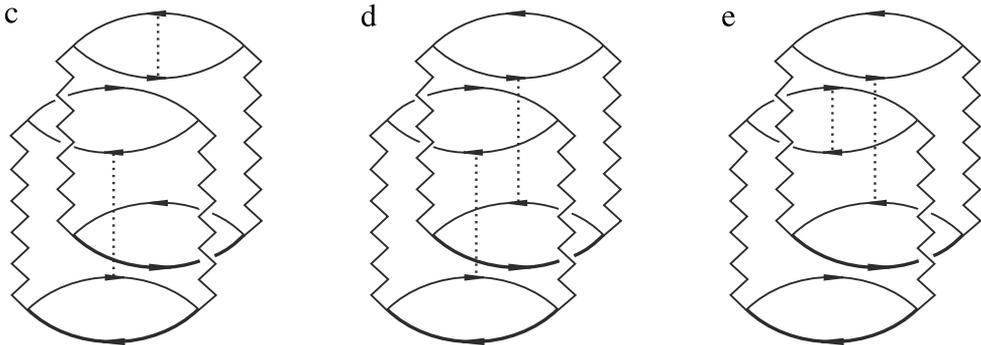
$$Z(a) = \frac{1}{2} \int_0^1 d\lambda_F \int_1^\infty d\lambda_B \frac{\lambda_B + \lambda_F}{\lambda_B - \lambda_F} e^{-a(\lambda_B - \lambda_F)} = \frac{1}{a}. \quad (\text{E.5})$$

### Appendix F. Spurious pairings of eight Green functions

As we explained in Section 6.4, the maximal number of the zero-argument delta-functions  $\delta_\gamma(0)$  in the expression for  $\langle \gamma^2(\varepsilon, T) \rangle$  originates if two Green functions with energy  $\varepsilon$  are paired together, and the remaining six Green functions are grouped such that energy coincidence in two pairs automatically guarantees the energy coincidence in the third pair. Direct analysis shows, however, that apart from two physical possibilities (a) and (b), given by Eq. (67) and shown in Fig. 6, there are three others that can formally satisfy this condition:

$$(c) \varepsilon_1 \approx \varepsilon \quad \text{and} \quad \omega_2 \approx 0, \quad (d) \varepsilon_1 \approx \varepsilon_2 \approx \varepsilon, \quad (e) \varepsilon_2 \approx \varepsilon \quad \text{and} \quad \omega_1 \approx 0. \quad (\text{F.1})$$

However, these new possibilities shown in Fig. 9 appear to be too restrictive: In each of these diagrams, the energies of four Green functions are pinned to the external energy  $\varepsilon$ , while the rest four are grouped in two pairs (shown by the dotted line). The possibilities listed in Eq. (F.1) are unphysical as they do not describe any real decay process (the final state consists of a single electron instead of an electron and an electron-hole pair) and one can expect that they do not contribute to the relaxation rate. Indeed, using expressions from Appendix B one can immediately check that spurious contributions with either  $\varepsilon_1 = \varepsilon$  or  $\varepsilon_2 = \varepsilon$  nullify the collision integral. Consequently, the only possibility to acquire an additional large factor  $\delta_\gamma(0)$  is to analyze two pairings (a) and (b), which is performed in the main text.



**Fig. 9.** Spurious ways to match the energies of eight Green functions listed in Eq. (F.1). All energy and coordinate indices are the same as in Fig. 6(a). Two pairs are shown by the dotted lines, while the rest four Green functions have the energy  $\varepsilon$  of the decaying particle. Such configurations automatically nullify the collision integral and do not contribute to mesoscopic fluctuations of  $\gamma(\varepsilon, T)$ .

## References

- [1] U. Sivan, Y. Imry, A.G. Aronov, *Europhys. Lett.* 28 (1994) 115.
- [2] D. Pines, P. Nozieres, *The Theory of Quantum Liquids*, Vol. 1, Benjamin, New York, 1966.
- [3] Ya.M. Blanter, *Phys. Rev. B* 54 (1996) 12807.
- [4] U. Sivan, F.P. Milliken, K. Milkove, S. Rishton, Y. Lee, J.M. Hong, V. Boegli, D. Kern, M. deFranza, *Europhys. Lett.* 25 (1994) 605.
- [5] B.L. Altshuler, Y. Gefen, A. Kamenev, L.S. Levitov, *Phys. Rev. Lett.* 78 (1997) 2803.
- [6] R. Abou-Chacra, P.W. Anderson, D.J. Thouless, *J. Phys. C: Solid State Phys.* 6 (1973) 1734.
- [7] A.D. Mirlin, Y.V. Fyodorov, *Phys. Rev. B* 56 (1997) 13393.
- [8] P.G. Silvestrov, *Phys. Rev. Lett.* 79 (1997) 3994.
- [9] H.A. Bethe, *Phys. Rev.* 50 (1936) 332.
- [10] A. Bohr, B.R. Mottelson, *Nuclear Structure*, Vol. 1, Benjamin, New York, 1969, p. 284.
- [11] Ph. Jacquod, D.L. Shepelyansky, *Phys. Rev. Lett.* 79 (1997) 1837.
- [12] X. Leyronas, J. Tworzydło, C.W.J. Beenakker, *Phys. Rev. Lett.* 82 (1999) 4894.
- [13] A.M.F. Rivas, E.R. Mucciolo, A. Kamenev, *Phys. Rev. B* 65 (2002) 155309.
- [14] D.M. Basko, I.L. Aleiner, B.L. Altshuler, *Ann. Phys.* 321 (2006) 1126.
- [15] I.V. Gornyi, A.D. Mirlin, D.G. Polyakov, *Phys. Rev. Lett.* 95 (2005) 206603.
- [16] A. Polkovnikov, K. Sengupta, A. Silva, M. Vengalattore, *Rev. Modern Phys.* 83 (2011) 863.
- [17] I.V. Gornyi, A.D. Mirlin, D.G. Polyakov, *Phys. Rev. B* 93 (2016) 125419.
- [18] P.G. Silvestrov, *Phys. Rev. B* 64 (2001) 113309.
- [19] B.L. Altshuler, A.G. Aronov, *Electron–Electron Interaction in Disordered Systems*, Elsevier, Amsterdam, 1985.
- [20] E. Akkermans, G. Montambaux, *Mesoscopic Physics of Electrons and Photons*, Cambridge University Press, New York, 2007.
- [21] J. Rammer, H. Smith, *Rev. Modern Phys.* 58 (1986) 323.
- [22] K.B. Efetov, *Supersymmetry in Disorder and Chaos*, Cambridge University Press, New York, 1997.
- [23] D.M. Basko, M.A. Skvortsov, V.E. Kravtsov, *Phys. Rev. Lett.* 90 (2003) 096801.
- [24] M.A. Skvortsov, *Phys. Rev. B* 68 (2003) 041306.
- [25] D.A. Ivanov, M.A. Skvortsov, *Nuclear Phys. B* 737 (2006) 304.
- [26] L.V. Keldysh, *Zh. Eksp. Teor. Fiz.* 47 (1964) 1945; *Sov. Phys.—JETP* 20 (1977) 1018.
- [27] A. Kamenev, A. Andreev, *Phys. Rev. B* 60 (1999) 2218.
- [28] A. Kamenev, A. Levchenko, *Adv. Phys.* 58 (2009) 197.
- [29] G.D. Mahan, *Many-Particle Physics*, Plenum Press, New York, 1990.
- [30] B.L. Altshuler, *Zh. Eksp. Teor. Fiz.* 75 (1978) 1330; *Sov. Phys.—JETP* 48 (1978) 670.
- [31] A. Schmid, *Z. Phys.* 259 (1973) 421.
- [32] M.Yu. Reizer, *Phys. Rev. B* 39 (1989) 1602.
- [33] B.L. Altshuler, A.G. Aronov, *Zh. Eksp. Teor. Fiz.* 75 (1978) 1610; *Sov. Phys.—JETP* 48 (1978) 812.
- [34] A. Schmid, *Z. Phys.* 271 (1974) 251.
- [35] I. Aleiner, L. Glazman, *Phys. Rev. B* 57 (1998) 9608.
- [36] Ya.M. Blanter, A.D. Mirlin, *Phys. Rev. E* 55 (1997) 6514.
- [37] D.M. Basko, *Phys. Rev. Lett.* 91 (2003) 206801.
- [38] B.L. Altshuler, A.G. Aronov, P.A. Lee, *Phys. Rev. Lett.* 44 (1980) 1288.
- [39] I.L. Aleiner, P.W. Brouwer, L.I. Glazman, *Phys. Rep.* 358 (2002) 309.
- [40] V.E. Kravtsov, A.D. Mirlin, *Pis'ma Zh. Eksp. Teor. Fiz.* 60 (1994) 645; *JETP Lett.* 60 (1994) 656.
- [41] M.L. Mehta, *Random Matrices*, Elsevier Academic Press, San Diego, 2004.
- [42] T.F. Rosenbaum, K. Andres, G.A. Thomas, P.A. Lee, *Phys. Rev. Lett.* 46 (1981) 568.
- [43] J.M.B. Lopes dos Santos, *Phys. Rev. B* 28 (1983) 1189.
- [44] H. Fukuyama, E. Abrahams, *Phys. Rev. B* 27 (1983) 5976.
- [45] H. Fukuyama, *J. Phys. Soc. Japan* 53 (1984) 3299.
- [46] C. Castellani, C. Di Castro, G. Kotliar, P.A. Lee, *Phys. Rev. Lett.* 56 (1986) 1179.
- [47] S. Hikami, *Phys. Rev. B* 24 (1981) 2671.
- [48] C. Castellani, C. Di Castro, P.A. Lee, M. Ma, *Phys. Rev. B* 30 (1986) 527.