Vacancy-induced low-energy states in undoped graphene

Sambuddha Sanyal,[1] Kedar Damle,[2] and Olexei I. Motrunich[3]

[1]*International Center for Theoretical Sciences, Tata Institute of Fundamental Research, Bengaluru 560089, India*
[2]*Department of Theoretical Physics, Tata Institute of Fundamental Research, Mumbai 400005, India*
[3]*Department of Physics, California Institute of Technology, Pasadena, California 91125, USA*

We demonstrate that a nonzero concentration n_v of static, randomly-placed vacancies in graphene leads to a density w of zero-energy quasiparticle states at the band-center $\epsilon = 0$ within a tight-binding description with nearest-neighbour hopping t on the honeycomb lattice. We show that w remains generically nonzero in the compensated case (exactly equal number of vacancies on the two sublattices) even in the presence of hopping disorder, and depends sensitively on n_v and correlations between vacancy positions. For low, *but not-too-low* $|\epsilon|/t$ in this compensated case, we show that the density of states (DOS) $\rho(\epsilon)$ exhibits a strong divergence of the form $\rho_{\rm 1D}(\epsilon) \sim |\epsilon|^{-1}/[\log(t/|\epsilon|)]^{(y+1)}$, which crosses over to the universal low-energy asymptotic form expected on symmetry grounds $\rho_{\rm GW}(\epsilon) \sim |\epsilon|^{-1} e^{-b[\log(t/|\epsilon|)]^{2/3}}$ below a crossover scale $\epsilon_c \ll t$. ϵ_c is found to decrease rapidly with decreasing n_v, while y decreases much more slowly.

PACS numbers: 71.23.-k;73.22.Pr;71.23.An;72.15.Rn

Static impurities, which give rise to random time-independent terms in the single-particle Hamiltonian for quasiparticle excitations of a condensed matter system, can lead to the phenomenon of Anderson localization, whereby quasiparticle wavefunctions lose their plane-wave character and become localized [1]. Such localization transitions and universal low-energy properties of the localized phase have been successfully described in many cases using effective field-theories [2, 3] whose form depends on symmetry properties of the quasiparticle Hamiltonian in the presence of impurities. In some cases [4, 5], it has also been possible to refine these field theoretical predictions using real-space strong-disorder renormalization group ideas [6].

In this Letter, we study the effects of a nonzero concentration n_v of static, randomly-located vacancies in graphene. We use a tight-binding description for electronic states of graphene, with hopping amplitude t between nearest-neighbour sites on a honeycomb lattice, and model vacancies by the deletion of the corresponding site in this tight-binding model [7–11]. We focus on the compensated case, *i.e.*, exactly equal numbers of vacancies on the two sublattices of the honeycomb lattice, and demonstrate that vacancies generically lead to a nonuniversal density w of zero-energy quasiparticle states at the band-center $\epsilon = 0$ even in this compensated case, including in the presence of hopping disorder. For low, *but not-too-low* $|\epsilon|/t$ in this compensated case, the density of states (DOS) $\rho(\epsilon)$ exhibits a strong divergence of the form:

$$\rho_{\rm 1D}(\epsilon) \sim |\epsilon|^{-1}/[\log(t/|\epsilon|)]^{(y+1)} \, , \qquad (1)$$

familiar in the context of various random-hopping problems in one dimension [12–20]. At still lower energies, below a crossover scale ϵ_c that is several orders of magnitude smaller than t even for moderately small values of n_v (0.05–0.1), we show that the DOS crosses over to the low-energy asymptotic behaviour [4–6, 21] of the chiral orthogonal universality class (to which our tight-binding model belongs on symmetry grounds):

$$\rho_{\rm GW}(\epsilon) \sim |\epsilon|^{-1} e^{-b[\log(t/|\epsilon|)]^{2/3}} \, . \qquad (2)$$

The density of zero-energy states w depends sensitively on correlations between vacancies and decreases as n_v is lowered. The crossover energy ϵ_c is found to decrease rapidly with decreasing w, while y (in fits to Eq. (1) for $|\epsilon| > \epsilon_c$) decreases much more slowly. On comparing the corresponding crossover length scale l_c, defined as the mean spatial separation between nonzero energy modes with $|\epsilon| < \epsilon_c$, with $l_w \equiv w^{-1/2}$, the mean spatial separation between zero-energy states, we find that l_c tracks l_w up to a nonuniversal prefactor. Thus, our results imply that the $w \to 0$ limit of the DOS is singular and does not commute with the $\epsilon \to 0$ limit: For any $w > 0$, the true asymptotic form $\rho_{\rm GW}(\epsilon)$ cannot be obtained from an extrapolation of results obtained for $\epsilon_c < \epsilon \ll t$, which instead reflect the intermediate-energy physics encoded in the form $\rho_{\rm 1D}(\epsilon)$.

Our work sheds light on an interesting question motivated by the results of Willans *et. al.*, who found a vacancy-induced DOS of the form $\rho_{\rm 1D}(\epsilon)$ at not-too-low energies in their study of Majorana excitations of Kitaev's honeycomb model [22]: Does a nonzero vacancy density lead to low-energy properties qualitatively different from the asymptotic behaviour expected in the chiral orthogonal universality class of quasiparticle localization? In recent work that addressed this question in the context of graphene [23, 24], it was argued that vacancies lead to a new term in the low-energy field theory, which causes the DOS to take on the form $\rho_{\rm 1D}(\epsilon)$, Eq. (1), with $y = 1/2$ at asymptotically low energies, rather than the asymptotic form $\rho_{\rm GW}(\epsilon)$, Eq. (2), expected on symmetry grounds.

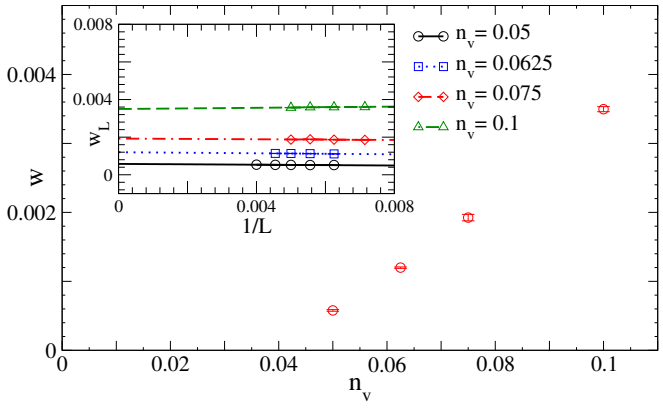


FIG. 1: w_L, the density of zero modes in an $L \times L$ sample, tends to a nonzero thermodynamic limit w that depends on n_v, the concentration of vacancies.

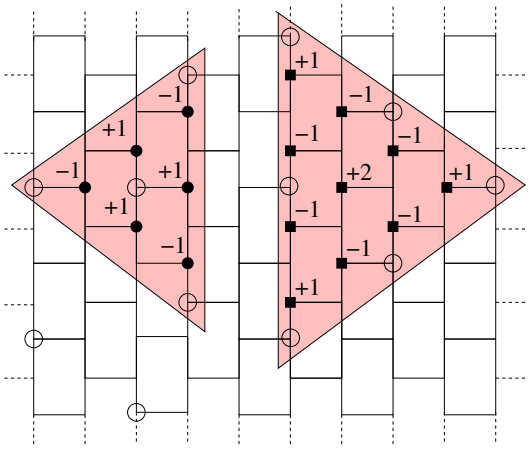


FIG. 2: Four B-sublattice (six A-sublattice) vacancies arranged in a "4-triangle" pattern ("$\mathcal{R}_6$" motif) give rise to a zero mode of H living on A-sublattice sites (B-sublattice sites) within the 4-triangle ($\mathcal{R}_6$ motif). While hopping disorder eliminates the 4-triangle zero mode, it only changes the wavefunction of the $\mathcal{R}_6$ zero mode *without* changing its energy.

Clearly, our conclusion is quite different, and raises two perhaps more interesting questions: When $\epsilon_c \ll t$, are the crossover exponent y and crossover energy ϵ_c "universally" determined by the zero-mode density w, although the function $w(n_v)$ itself depends sensitively on microscopic details such as correlations between vacancies? Can this crossover be understood within a renormalization group description of the low-energy physics? Leaving these interesting questions for future work, we devote the remainder of this Letter to an account of the calculations that lead us to our results, and thence, to these questions.

We choose the lattice spacing of the honeycomb lattice as our unit of length and measure all energies in terms of the hopping amplitude t, which is set by the bandwidth of the π-band of undoped graphene. We focus on the compensated case, with exactly $n_v L^2$ vacancies placed randomly on *each* sublattice of a finite $L \times L$ honeycomb lattice with L^2 unit cells ($2L^2$ sites). The spectrum of single-particle states can be obtained by diagonalizing the real symmetric matrix H

$$H = \begin{pmatrix} 0 & T_{AB} \\ T_{AB}^\dagger & 0 \end{pmatrix} \qquad (3)$$

where T_{AB} is the $(1-n_v)L^2$-dimensional matrix of amplitudes for hopping from the undeleted sites of the B sublattice to their undeleted A sublattice neighbours, and $T_{AB}^\dagger$ is the transpose of this matrix (the spin label of the electronic quasiparticles is dropped since we do not study magnetic properties or sources of spin-flip scattering in this Letter).

The purely off-block-diagonal form of H reflects the "chiral" symmetry of the problem, corresponding to the bipartite structure of the honeycomb lattice, which guarantees that every eigenstate with energy $\epsilon > 0$ has a corresponding eigenstate at energy $-\epsilon$. In order to eliminate zero modes of H in the pure $L \times L$ lattice [25–27], we choose even values of L and impose antiperiodic boundary conditions along the $\hat{x}$ direction, while terminating the lattice in the $\hat{y}$ direction in a pair of armchair edges. We also impose a nearest-neighbour and next-nearest-neighbour exclusion constraint on the vacancies, and do not allow them to interrupt the armchair edges. These restrictions, along with the compensated nature of the vacancy disorder, eliminate all previously studied and well-understood sources of vacancy-induced [9, 28] zero modes in the spectrum of H.

We find it convenient to focus on the symmetric matrix $T_{AB}^\dagger T_{AB}$, which has a single eigenvalue ϵ^2 for every pair of nonzero eigenvalues $(\epsilon, -\epsilon)$ of H. Zero modes of $T_{AB}^\dagger T_{AB}$, with wavefunction living entirely on the B sublattice, map on to exactly half of the zero modes in the spectrum of H, while zero modes of the symmetric matrix $T_{AB} T_{AB}^\dagger$, with wavefunction living entirely on the A sublattice, make up the other half of the null space of H. We use the ALGOL [29] routines of Martin and Wilkinson [30] to compute the number $\mathcal{N}_\Lambda$ of eigenvalues of the banded matrix $T_{AB}^\dagger T_{AB}$ which are smaller in magnitude than some positive number $t^2 \times 10^{-\Lambda}$. Our implementation [31] uses calls to the GNU multiprecision library [32] for all arithmetic operations, including comparison of the magnitudes of two numbers, and has been benchmarked against routines from the LAPACK library [33] as well as C-translations (used in earlier work [6]) of the ALGOL routines of Martin and Wilkinson.

Anticipating that the physics of interest to us spans many orders of magnitude in energy ϵ, we define the 'log-energy' $\Gamma = \log_{10}(t/|\epsilon|)$, and compute $N_{\rm tot}^{(i)}(\Gamma, L) \equiv \mathcal{N}_{\Lambda=2\Gamma}^{(i)}/L^2$ for the i^{th} $L \times L$ random sample using values of log-energy drawn from an equispaced grid ranging from $\Gamma \sim 1$ to $\Gamma \sim 100$. For large enough Γ, $N_{\rm tot}^{(i)}(\Gamma, L)$

plateaus out to a constant value which represents the density of zero modes $w_L^{(i)}$ of that sample. For not-too-small n_v ($n_v \geq 0.05$) for which we are able to access this plateau, we separately keep track of $w_L^{(i)}$ and $N_L^{(i)}(\Gamma) \equiv N_{\rm tot}^{(i)}(\Gamma, L) - w_L^{(i)}$. From the position, $\Gamma_{\rm g}^{(i)}(L)$, of the last downward step in $N_{\rm tot}^{(i)}(\Gamma, L)$, we also obtain the spectral gap $\epsilon_{\rm g}^{(i)}(L) \equiv t \times 10^{-\Gamma_{\rm g}^{(i)}(L)}$ corresponding to the lowest pair of nonzero eigenvalues $\pm\epsilon_{\rm g}^{(i)}$ for that sample. Analyzing this data for up to 3000 samples for each value of L and n_v, we obtain statistically reliable estimates of the corresponding disorder-averaged quantities w_L and $N_L(\Gamma)$. The density of states $\rho_L(\epsilon)$ can then be obtained from N_L using the relation $\rho_L(\epsilon) \equiv \frac{1}{2\epsilon}\frac{dN_L}{d\Gamma}$. Additionally, we estimate f_L, the probability that an $L \times L$ sample has at least one pair of zero modes, and measure the histogram of $\Gamma_{\rm g}(L)$. The position of the peak in the latter provides us an estimate of $\Gamma_{\rm g}^*(L)$, the most probable value of $\Gamma_{\rm g}(L)$. For the smallest values of n_v, which require multiprecision computation at impracticably large Γ in order to access the plateau in $N_{\rm tot}^{(i)}(\Gamma, L)$ (and thence, w_L^i), we instead compute $\frac{dN_L}{d\Gamma}$ by numerical differentiation of $N_{\rm tot}^{(i)}(\Gamma, L)$.

Extrapolating our results for f_L (Supplemental Material Section) and w_L (Fig. 1) to obtain $f \equiv \lim_{L\to\infty} f_L$ and $w \equiv \lim_{L\to\infty} w_L$, we find that $f = 1$, and that w depends sensitively on n_v (Fig. 1). To understand these results, we observe that $T_{AB}T_{AB}^\dagger$ ($T_{AB}^\dagger T_{AB}$) *must* have a zero mode, with wavefunction shown in Fig. 2, if four of the B-sublattice vacancies (six of the A-sublattice vacancies) are arranged in the specific "4-triangle" pattern ("$\mathcal{R}_6$ motif") shown in Fig. 2, with no restrictions on the positions of the other vacancies. H must therefore have a *pair of zero modes* if a *single* 4-triangle or $\mathcal{R}_6$ motif occurs anywhere in the sample on either sublattice. Since there is a nonzero probability of finding a 4-triangle at a given location, this already implies that a large enough sample will certainly have at least one zero mode, *i.e.*, $f = 1$. Additionally, one has an elementary lower-bound on $w_L^{(i)}$ in terms of the numbers $N_{\Delta_{4A}}^{(i)}$ and $N_{\Delta_{4B}}^{(i)}$ of 4-triangles on A and B lattices in a given sample: $w_L^{(i)} \geq \left[\max(N_{\Delta_{4A}}^{(i)}, N_{\Delta_{4B}}^{(i)})\right]/L^2$, implying $w \geq n_{\Delta_4}$, where n_{Δ_4} is the ensemble averaged concentration of 4-triangles in the thermodynamic limit. When the vacancies obey the exclusion constraints described earlier, it is not possible to produce a similar zero mode with fewer than four vacancies (Supplemental Material Section). Thus, we expect $w \sim n_v^4$ in the $n_v \to 0$ limit.

While our lower bound can be strengthened somewhat by including larger versions of the 4-triangle motif (Supplemental Material Section), they do not change this limiting behaviour. However, our results (Fig. 1) suggest that this limiting behaviour sets in only for $n_v \ll 0.05$, for which a direct computation of w would require access

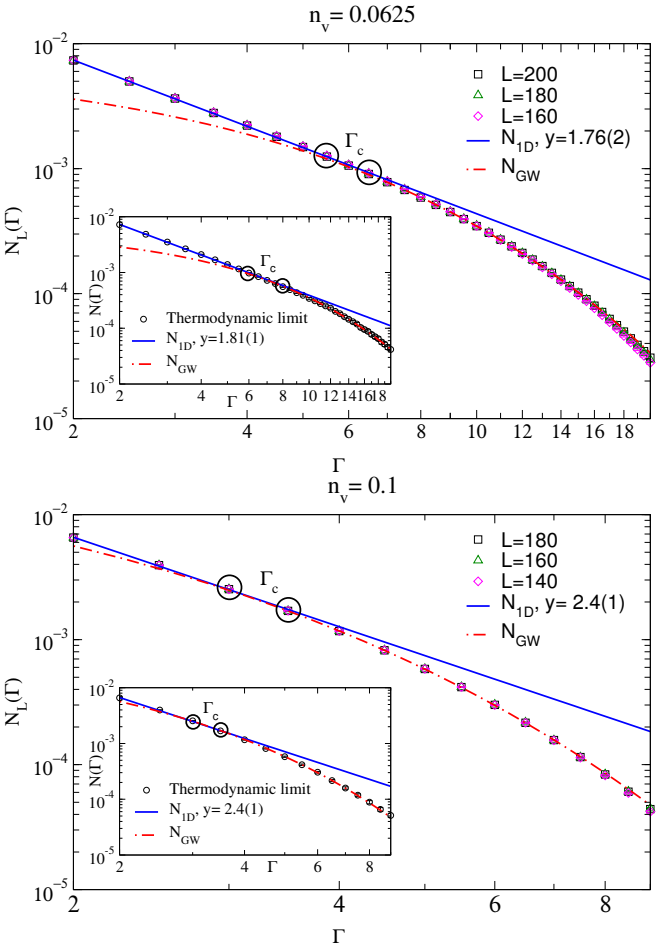


FIG. 3: $N_L(\Gamma)$ at the three largest values of L studied for $n_v = 0.0625$ and $n_v = 0.1$. Insets show $N(\Gamma)$ obtained by extrapolation to the thermodynamic limit. Circles demarcate the crossover region centered at the crossover scale Γ_c. Data for $\Gamma \lesssim \Gamma_c$ fits well to power-law form $N_{\rm 1D}(\Gamma)$ with the values of y indicated in each case, while the large-Γ regime fits well to the modified Gade-Wegner form $N_{\rm GW}(\Gamma)$.

to impracticably large Γ. For $n_v \gtrsim 0.05$, 4-triangles are *not* the dominant contribution to w (Supplemental Material Section), which we expect arises instead from generalizations of the $\mathcal{R}_6$ motif: Such "$\mathcal{R}$-type" regions have more undeleted sites belonging to one sublattice than the other, but are connected to the rest of the lattice only via sites belonging to the other sublattice. Like the $\mathcal{R}_6$ zero mode, all such $\mathcal{R}$-type zero modes are robust to disorder in the nearest-neighbour hopping amplitudes (Supplemental Material Section). Unlike zero modes associated with specific patterns like 4-triangles, these $\mathcal{R}$-type zero modes cannot be eliminated by any additional local constraints on the vacancy positions. They are therefore a *generic* feature of the diluted graphene lattice. Thus we see that a nonzero concentration n_v of vacancies leads to a density w of zero modes of H, where w depends sensitively on n_v, and on correlations in the positions of

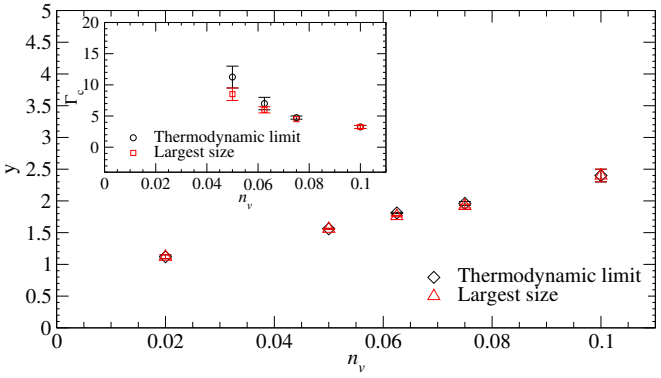


FIG. 4: n_v dependence of crossover scale Γ_c and power-law exponent y for samples with compensated random dilution.

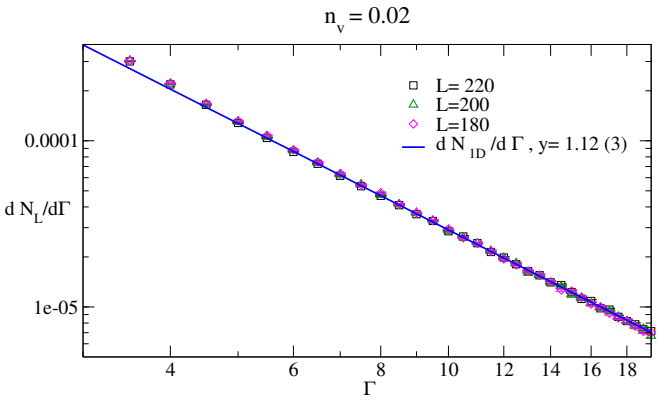


FIG. 5: $\frac{dN_L(\Gamma)}{d\Gamma}$ at $n_v = 0.02$ in the crossover regime converges to the thermodynamic limit for $L \sim 200$ and fits well to the form $\frac{dN_{1D}(\Gamma)}{d\Gamma}$, with a value of y consistent with the trends established at larger n_v for Γ_c and y (Fig. 4). Based on these trends, we expect $N(\Gamma)$ to cross over to the asymptotic form N_{GW} at *much larger* values of Γ, for which we are unable to reliably compute $N(\Gamma)$ due to computational constraints.

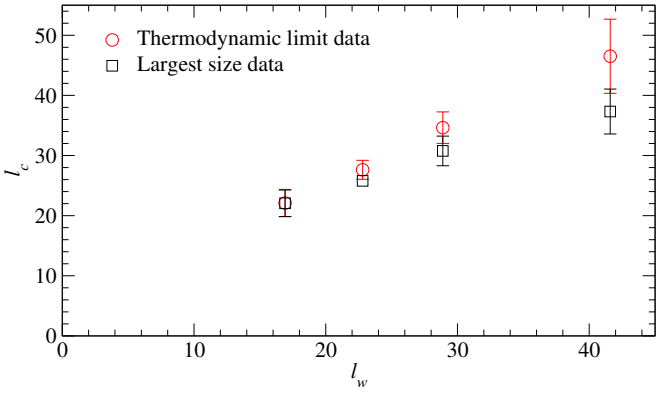


FIG. 6: The crossover length-scale $l_c \equiv N(\Gamma_c)^{-1/2}$ tracks the mean spatial separation $l_w \equiv w^{-1/2}$ between zero modes reasonably well for compensated random dilution. From left to right, the exhibited data points correspond to vacancy densities 0.1, 0.075, 0.0625, and 0.05.

vacancies, but remains generically nonzero even in the compensated case, including in the presence of hopping disorder.

Figure 3 displays $N_L(\Gamma)$ for $n_v = 0.0625$ and $n_v = 0.1$ for the three largest sizes used in our extrapolations to the thermodynamic limit. Since we expect finite-size effects to dominate for $\Gamma > \Gamma_g^*(L)$, we estimate $\Gamma_g^*(L)$ from histograms of $\Gamma_g(L)$ (Supplemental Material Section) and restrict attention to $\Gamma < \Gamma_g^*(L_{\min})$, where $L_{\min}$, the smallest of the sizes used in our extrapolations, is chosen large enough that $f_{L_{\min}} \approx 1$ in order to ensure that the physics of zero modes is correctly captured in all our analysis. In this range of Γ, we can reliably extrapolate (see Supplemental Material Section) from our data to obtain the thermodynamic limit $N(\Gamma)$ displayed in the inset of Fig. 3. Up to a fairly well-defined and readily-identified crossover scale $\Gamma_c(L) \equiv \log_{10}(t/|\epsilon_c(L)|)$, $N_L(\Gamma)$ is found to fit well to a power-law form $N_{1D}(\Gamma) \equiv c\Gamma^{-y}$. However, for larger Γ beyond Γ_c, the asymptotic fall-off is clearly faster than a power law. $\Gamma_c(L)$ increases slightly with L over the range of L studied, but saturates at large L to a finite thermodynamic limit Γ_c that marks the presence of the same crossover in the limiting curve $N(\Gamma)$. Thus, $N(\Gamma)$ is again fit well by the power-law form N_{1D} for $\Gamma \lesssim \Gamma_c$, but falls off much faster in the large-Γ regime.

Given that H belongs to the chiral orthogonal universality class, standard universality arguments predict that $N(\Gamma)$ and $N_L(\Gamma)$ should, at large enough Γ, follow the modified Gade-Wegner form [4–6, 21] $N_{GW}(\Gamma) \equiv a\Gamma^{1/3}e^{-b\Gamma^{2/3}}$. From Fig. 3, we see that this form indeed provides a very good fit in the asymptotic large-Γ regime. The same crossover is also visible at $n_v = 0.05$ and $n_v = 0.075$. From Fig. 4, we see that y decreases gradually with n_v, while Γ_c increases extremely rapidly as we go to smaller values of n_v, thereby limiting our ability to directly study this crossover for $n_v \lesssim 0.05$. However, one can nevertheless reliably compute the exponent y that characterizes the behaviour of $\rho(\epsilon)$ in the intermediate regime $t \gg |\epsilon| \gg \epsilon_c$ (Fig. 5), and confirm that its value evolves smoothly (Fig. 4) down to these small values of n_v. This strongly suggests that the crossover identified by us is an intrinsic and generic feature of the density of states for any nonzero n_v.

The corresponding crossover length scale $l_c \equiv N(\Gamma_c)^{-1/2}$, which represents the mean spatial separation between nonzero energy modes with $|\epsilon|/t < 10^{-\Gamma_c}$, grows relatively slowly (Fig. 6) as w is decreased, with $l_c \lesssim 50$ lattice units even at the smallest value of w studied (corresponding to $n_v = 0.05$). This explains why our extrapolations to the thermodynamic limit using finite-size data with $L \sim 200$ remain reliable for all n_v studied. From Fig. 6, which compares l_c for the randomly diluted samples with $l_w \equiv w^{-1/2}$, the mean spatial separation between zero modes, we also see that l_c tracks l_w (up to a nonuniversal prefactor). This suggests that the

crossover identified in this Letter is controlled primarily by the density of zero modes. Additional support for this idea comes from our study of samples diluted with an equal number of randomly placed 4-triangles (instead of individual vacancies) on each sublattice (Supplemental Material Section), which show the same crossover, but with very different values of ϵ_c and y that are better predicted by the zero mode density w as opposed to the vacancy density. This then leads us to the questions identified earlier: Is the physics of this crossover "universally controlled" by the value of w (*i.e.*, independent of correlations between vacancy-positions and other microscopic details) in the limit of small w, and can it be understood via a renormalization group description of the low-energy physics?

Acknowledgements We thank M. Barma and D. Dhar for useful comments on a previous draft, and gratefully acknowledge use of computational resources funded by DST (India) grant DST-SR/S2/RJN-25/2006, in addition to departmental computational resources of the Dept. of Theoretical Physics of the TIFR. KD and OM gratefully acknowledge hospitality of ICTS-TIFR (Bengaluru) and IISc (Bengaluru) during completion of part of this work. SS gratefully acknowledges funding from DST (India) and DAE -SRC (India) and support from IISc (Bengaluru) during completion of part of this work. OM also acknowledges support by the NSF through grant DMR-1206096.

[1] P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
[2] A. Altland, B. D. Simons, and M. R. Zirnbauer, Phys. Rep. **359**, 283 (2002).
[3] F. Evers and A. D. Mirlin, Rev. Mod. Phys. **80**, 1355 (2008).
[4] R. Gade, Nucl. Phys. B **398**, 499 (1993).
[5] R. Gade and F. Wegner, Nucl. Phys. B **360**, 213 (1991).
[6] O. I. Motrunich, K. Damle, and D. A. Huse, Phys. Rev. B **65**, 064206 (2002).
[7] P. T. Araujo, M. Terrones, M. S. Dresselhaus, Materials Today **15**, 98 (2012).
[8] G. Forte, A. Grassi, G. M. Lombardo, A. La Magna, G. G. N. Angilella, R. Pucci, R. Vilardi, Phys. Lett. A **372**, 6168 (2008).
[9] V. M. Pereira, J. M. B. Lopes dos Santos, and A. H. Castro Neto, Phys. Rev. B **77**, 115109 (2008).
[10] V. M. Pereira, F. Guinea, J. M. B. Lopes dos Santos, N. M. R. Peres, and A. H. Castro Neto, Phys. Rev. Lett. **96**, 036801 (2006).
[11] T. O. Wehling, S. Yuan, A. I. Lichtenstein, A. K. Geim, and M. I. Katsnelson, Phys. Rev. Lett. **105**, 056802 (2010).
[12] F. J. Dyson, Phys. Rev. **92**, 1331 (1953).
[13] G. Theodorou and M. H. Cohen, Phys. Rev. B **13**, 4597 (1976).
[14] T. P. Eggarter and R. Riedinger, Phys. Rev. B **18**, 569 (1978).
[15] O. Motrunich, K. Damle, and D. A. Huse, Phys. Rev. B **63**, 134424 (2001).
[16] O. Motrunich, K. Damle, and D. A. Huse Phys. Rev. B **63**, 224204 (2001).
[17] I. A. Gruzberg, N. Read, and S. Vishveshwara, Phys. Rev. B **71**, 245124 (2005).
[18] P. W. Brouwer, A. Furusaki, I. A. Gruzberg, and C. Mudry, Phys. Rev. Lett. **85**, 1064 (2000).
[19] P. W. Brouwer, C. Mudry, and A. Furusaki, Phys. Rev. Lett. **84**, 2913 (2000).
[20] M. Titov, P. W. Brouwer, A. Furusaki, and C. Mudry, Phys. Rev. B **63**, 235318 (2001).
[21] C. Mudry, S. Ryu, A. Furusaki, Phys. Rev. B **67**, 064202 (2003).
[22] A. J. Willans, J. T. Chalker, and R. Moessner, Phys. Rev. B **84**, 115146 (2011)
[23] V. Hafner, J. Schindler, N. Weik, T. Mayer, S. Balakrishnan, R. Narayanan, S. Bera, and F. Evers, Phys. Rev. Lett. **113**, 186802 (2014).
[24] P. M. Ostrovsky, I. V. Protopopov, E. J. Konig, I. V. Gornyi, A. D. Mirlin, and M. A. Skvortsov, Phys. Rev. Lett. **113**, 186803 (2014).
[25] E. H. Lieb and M. Loss, Duke Math. J. **71**, 337 (1993).
[26] S. Ryu and Y. Hatsugai, Phys. Rev. Lett. **89**, 077002 (2002).
[27] L. Brey and H. A. Fertig, Phys. Rev. B **73**, 235411 (2006).
[28] P. W. Brouwer, E. Racine, A. Furusaki, Y. Hatsugai, Y. Morita, and C. Mudry, Phys. Rev. B **66**, 014204 (2002).
[29] https://en.wikipedia.org/wiki/ALGOL
[30] R. S. Martin and J. H. Wilkinson in *Handbook for Automatic Computation, Vol. II: Linear Algebra*, J. H. Wilkinson and C. Reinsch (eds.), Springer-Verlag (Berlin, 1971).
[31] S. Sanyal, Ph.D thesis, Tata Institute of Fundamental Research, Mumbai, 2014, http://theory.tifr.res.in/Research/Thesis/
[32] https://en.wikipedia.org/wiki/GNU_Multiple_Precision_Arithmetic_Library
[33] https://en.wikipedia.org/wiki/LAPACK

Supplemental Material for "Vacancy-induced low-energy states in undoped graphene"

In this Supplemental Material, we present additional numerical evidence and analytical arguments which support the key findings described in the main text.

Additional numerical evidence

Other concentrations

Figure 7 displays $N_L(\Gamma)$ for $n_v = 0.05$, for the three largest sizes studied. The corresponding extrapolation to the thermodynamic limit is shown in Fig. 8. The corresponding results for $n_v = 0.075$ are displayed in Fig. 9. In all these figures, we focus on $\Gamma < \Gamma_g^*(L_{\rm min})$, where $L_{\rm min}$ is the smallest size for which data is displayed, and Γ_g^* is read off from the peak in the histograms of Γ_g shown in Fig. 10 and Fig. 11. The corresponding histograms for $n_v = 0.0625$ and $n_v = 0.1$ are displayed in Fig. 12

As is clear from these results for $n_v = 0.05$ and $n_v = 0.075$, $N_L(\Gamma)$ is found to fit well to a power-law form $N_{\rm 1D} \equiv c\Gamma^{-y}$ up to a fairly well-defined and readily-identified crossover scale $\Gamma_c(L) \equiv \log_{10}(t/|\epsilon_c(L)|)$. However, beyond Γ_c, the asymptotic fall-off is clearly faster than a power law. While the increase of $\Gamma_c(L)$ with L is more significant at the smallest concentration studied ($n_v = 0.05$), it is nevertheless clear that $\Gamma_c(L)$ does saturate to a finite value even in this case. This is clear from the fact that the extrapolated thermodynamic density of states $N(\Gamma)$ (Fig. 8) also displays the same crossover seen in the finite-size data. In the large-Γ regime beyond this crossover, the modified Gade-Wegner form $N_{GW}(\Gamma) \equiv a\Gamma^{1/3}e^{-b\Gamma^{2/3}}$ is seen to provide a very good fit of the data for both these concentrations. The corresponding values of $\Gamma_c(L)$ and Γ_c, and of the best fit values of y, provide us additional points that fill in the curves shown in Fig. 4 and Fig. 6 of the main text, which display the n_v dependence of y and Γ_c, and the close relationship between $l_c \equiv N(\Gamma_c)^{-1/2}$ and $l_w \equiv w^{-1/2}$. Finally, we re-emphasize a point made already in the main text: Our computational constraints prevent us from accessing the thermodynamic limit for the much larger values of Γ at which we expect to see the same crossover for the lowest concentration $n_v = 0.02$.

Extrapolations

Since states at any finite Γ (*i.e.*, away from the band center $\epsilon = 0$) in such particle-hole symmetric hopping problems are not critical, one expects the leading corrections to the thermodynamic limit $N(\Gamma)$ at any finite Γ to be regular rather than singular, similar to the finite-size corrections expected in noncritical phases of matter (*i.e.*, away from critical points or critical lines). Guided by this rationale, the thermodynamic limit $N(\Gamma)$ is obtained from $N_L(\Gamma)$ by performing a polynomial extrapolation in $1/L$ (note that we expect that the leading finite-size corrections are $\mathcal{O}(1/L)$ rather than $\sim \exp(-L/\xi)$ because of "surface" contributions associated with the semi-open boundary conditions we employ). Since we are careful to only use large enough sizes for which almost every sample has at least one zero mode ($f_L \approx 1$), our finite-size data is already rather close to the thermodynamic limit, leading to a rather small secular drift with increasing L. In most cases, given the size of our error bars relative to the magnitude of this secular drift with L, the inclusion of the next-order term c/L^2 only results in an over-interpretation of statistical fluctuations. Therefore, a simple linear (in $1/L$) extrapolation $a + b/L$ has been used in most cases.

We have also tested the stability of this extrapolation procedure to the inclusion of data at larger sizes. For the representative case of $n_v = 0.0625$, this is shown Fig. 13, which is devoted to a comparison of the thermodynamic limit obtained in the main text using sizes $L = 200, 180, 160$ with two other alternatives: A linear extrapolation from three sizes $L = 220, 200, 180$, and a linear extrapolation from four sizes $L = 220, 200, 180, 160$. As is clear from this figure, all three extrapolations (*i.e.*, the one used in the main text as well as the other alternatives which use data at a larger size) yield extrapolated values that lie within the error-bars of each other. Further, there is no systematic trend that suggests that any one of these extrapolations yields a consistently higher or lower value of $N(\Gamma)$ at all Γ. Details of all three extrapolations, for each value of Γ, are also shown as a separate multi-page figure (Fig. 26) placed at the end of this supplemental section for ease of inspection. Some examples of extrapolations used to arrive at $N(\Gamma)$ from data for $N_L(\Gamma)$ at other concentrations are also shown in Figs. 14, 15, 16, and 17. From this careful and detailed study, we conclude that our approach indeed allows us to reliably obtain the thermodynamic limit curve $N(\Gamma)$.

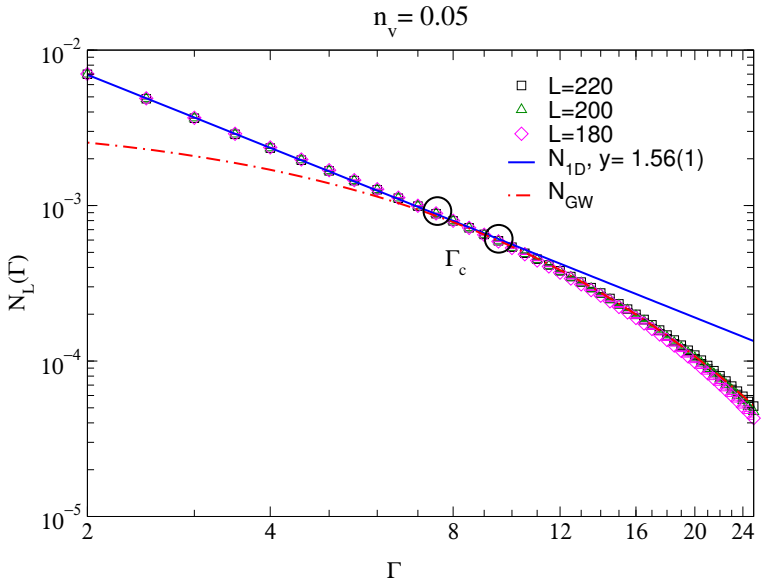


FIG. 7: $N_L(\Gamma)$ at the three largest values of L studied for $n_v = 0.05$. Circles demarcate the crossover region centered at the crossover scale Γ_c. Data for $\Gamma \lesssim \Gamma_c$ fits well to power-law form $N_{1D}(\Gamma)$ (see text for details) with the value of y indicated in the figure, while the large-Γ regime fits well to the modified Gade-Wegner form $N_{GW}(\Gamma)$.

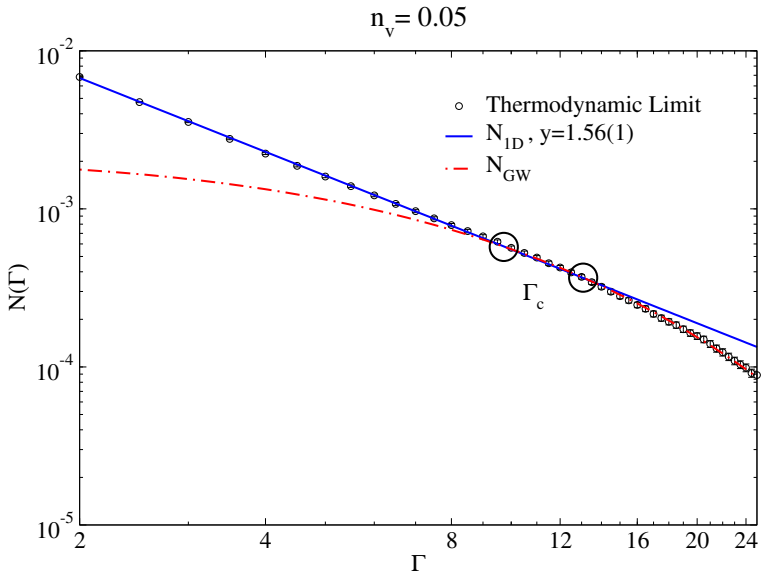


FIG. 8: $N(\Gamma)$, the extrapolation to the thermodynamic limit of the finite-size data from the previous figure. Again, circles demarcate the crossover region centered at the crossover scale Γ_c. Data for $\Gamma \lesssim \Gamma_c$ fits well to power-law form $N_{1D}(\Gamma)$ with the value of y indicated in the figure, while the large-Γ regime fits well to the modified Gade-Wegner form $N_{GW}(\Gamma)$.

Further analysis of zero modes

Our data for f_L, the probability that an $L \times L$ sample has at least one zero mode, is shown in Fig. 18. Clearly, f_L tends to 1 as $L \to \infty$, as already mentioned in the main text. This is consistent with the analytical argument in the main text, which also provides a simple rigorous lower bound for the density of zero modes. The 4-triangle zero mode used in this argument is the first term in an infinite series in n_v, with higher powers of n_v arising from bigger patterns consisting of a larger number of impurities in specific locations relative to each other. In Fig. 19 and Fig. 20, we show a few examples of zero mode constructions that contribute to this series. However, as already noted in the main text, terms in this series do not give the dominant contribution to w at the not-too-small values of n_v studied

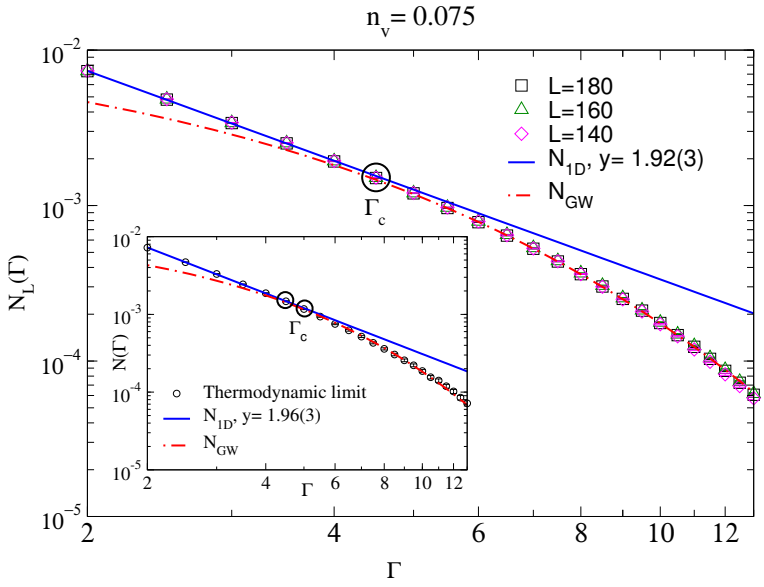


FIG. 9: $N_L(\Gamma)$ at the three largest values of L studied for $n_v = 0.075$. Circles demarcate the crossover region centered at the crossover scale Γ_c. Data for $\Gamma \lesssim \Gamma_c$ fits well to power-law form $N_{1D}(\Gamma)$ with the value of y indicated in the figure, while the large-Γ regime fits well to the modified Gade-Wegner form $N_{GW}(\Gamma)$. Inset shows the extrapolation to the thermodynamic limit, in which the same crossover is clearly visible.

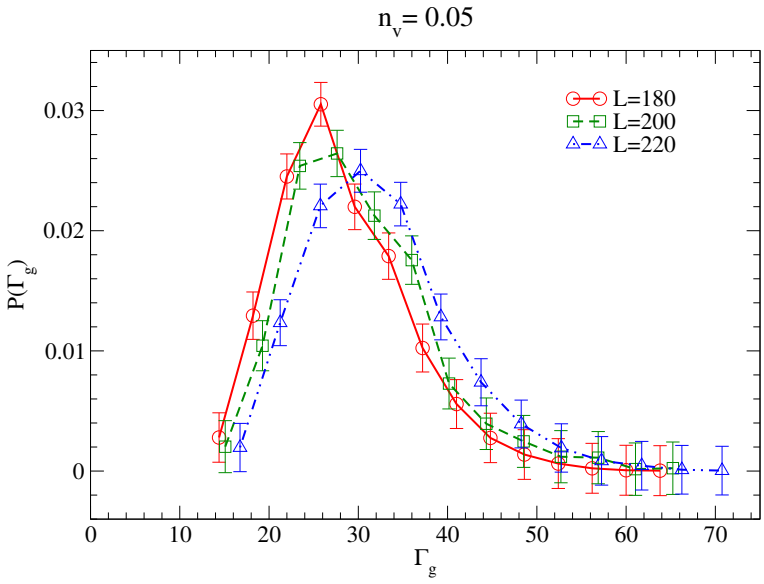


FIG. 10: Histogram of Γ_g, corresponding to the lowest nonzero gap for the three largest sizes studied at $n_v = 0.05$.

by us in this work. Indeed, we have explicitly measured the density of 4-triangles and checked that it is significantly smaller than the density of zero modes for all n_v at which we have computed w (including $n_v = 0.05$). Additionally, we have enumerated all possible clusters of fewer than four impurities and verified that it is not possible to produce a similar zero mode with fewer than four vacancies in a cluster so long as the exclusion constraints outlined in the main text are in place.

The zero mode associated with the $\mathcal{R}_6$ motif described in the main text also generalizes in an obvious way to yield a series of zero modes that all survive the effects of bond disorder in a manner completely analogous to the $\mathcal{R}_6$ zero mode. These $\mathcal{R}_n$ zero modes ($n > 6$) live on larger and larger equilaterial triangles (with zig-zag edges) which are connected to the rest of the lattice only via B (A) sublattice sites but have more undeleted A (B) sublattice sites than B (A) sublattice sites, allowing a zero mode to exist within the triangle for generic realizations of bond-disorder.

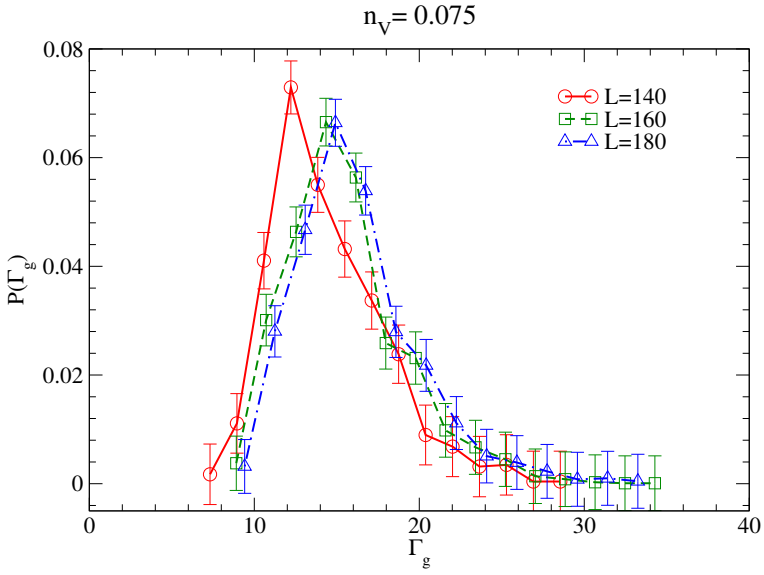


FIG. 11: Histogram of Γ_g, corresponding to the lowest nonzero gap for the three largest sizes studied at $n_v = 0.075$.

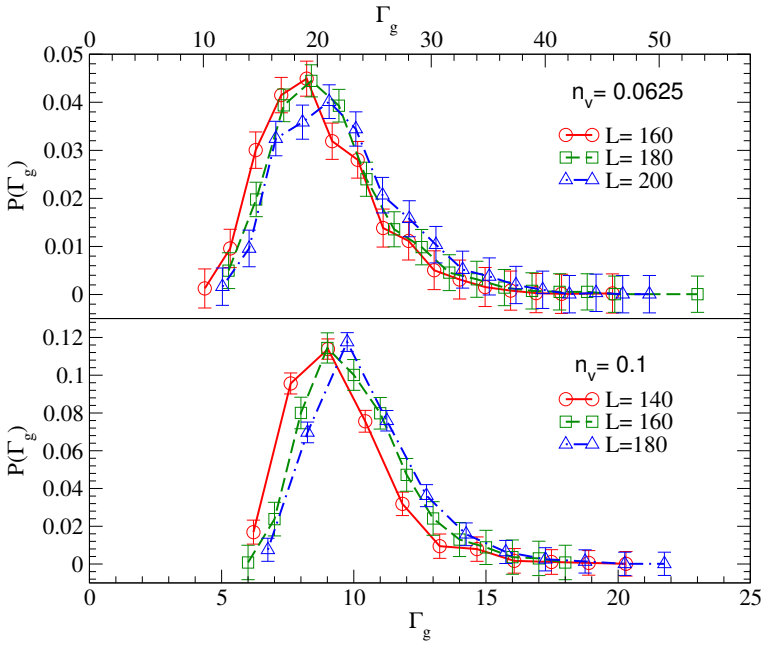


FIG. 12: Histograms of Γ_g at the three largest values of L studied for $n_v = 0.0625$ and $n_v = 0.1$.

As in the case of the $\mathcal{R}_6$ zero mode described in the main text, this robustness to disorder follows from the fact that the number of free components of the wavefunction of any such mode is one more than the number of zero-energy equations that they must satisfy.

We have also found other simple examples of such "$\mathcal{R}$-type" zero modes that live near the armchair boundary and are not associated with a specific regular arrangement of vacancies. Instead, as already mentioned earlier, these modes appear to generically live in a region $\mathcal{R}$ which connects to the rest of the lattice only via B (A) sublattice sites belonging to $\mathcal{R}$, although it has more undeleted A (B) sublattice sites than B (A) sublattice sites. In such a region, $T_{AB}T_{AB}^\dagger$ ($T_{AB}^\dagger T_{AB}$) has a zero mode living on the A (B) sublattice sites, simply because the number of constraints that need to be satisfied by this zero mode wavefunction is smaller than the number of A (B) sublattice sites on which this zero mode lives. As already noted, this feature also guarantees that such zero modes survive the effects of disorder in the nearest-neighbour hopping amplitudes. One example of such a mode is shown in Fig. 21. We believe

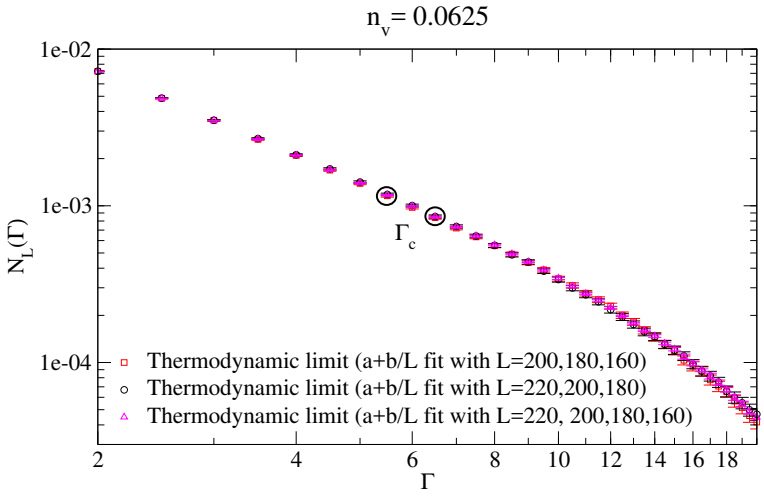


FIG. 13: Three different extrapolations yield results for the thermodynamic limit $N(\Gamma)$ that fall within the error bars of each other, confirming the reliability and stability of our procedure to obtain the thermodynamic limit for the representative case of $n_v = 0.0625$.

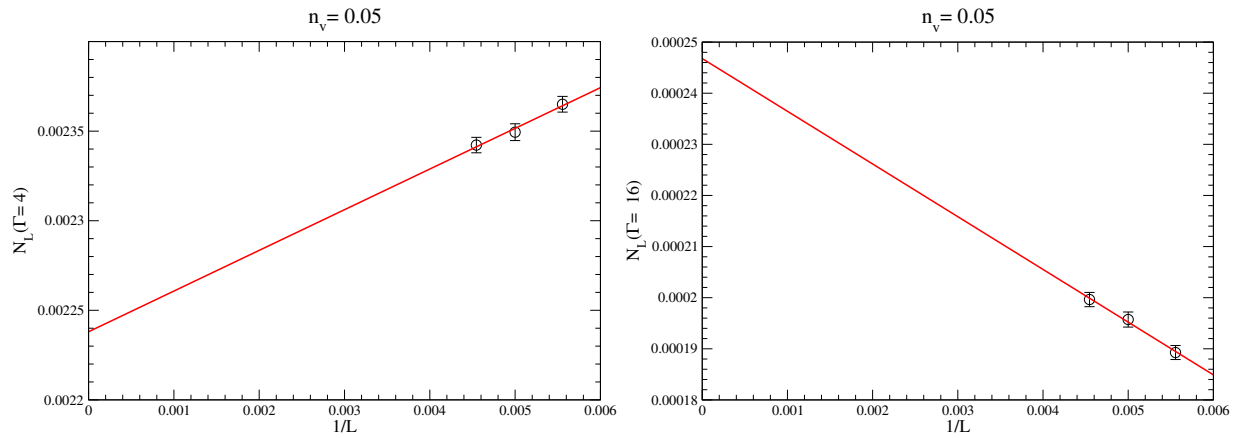


FIG. 14: Examples of extrapolation of $N_L(\Gamma)$ to the thermodynamic limit at $n_v = 0.05$. For this concentration, $\Gamma_c \approx 11$ (see Fig. 4 in the main text), and the left panel illustrates the extrapolation for $\Gamma < \Gamma_c$, while the right panel is for $\Gamma > \Gamma_c$. Note in particular that our extrapolation for $\Gamma > \Gamma_c$ is very likely an overstimate, so one can be fairly confident that $N(\Gamma)$ in the thermodynamic limit drops below $N_{1D}(\Gamma)$, ruling out a fit to this form for $\Gamma > \Gamma_c$.

that bulk versions of such more general $\mathcal{R}$-type zero modes provide the dominant contribution to w for the values of n_v studied by us, which is why our lower-bound on w (obtained by thinking in terms of Fig. 2 in the main text) substantially underestimates w at such not-too-small values of n_v. Clearly, no additional local correlations among impurities can entirely eliminate such more general $\mathcal{R}$-type zero modes . Therefore, a non-zero density of zero-energy modes is expected to be a generic feature of such systems. However, we have been unable to convert this observation into an improved lower-bound.

Dilution by 4-triangles

Finally, we provide an illustration of the importance of spatial correlations between vacancies via a simple toy model in which vacancies enter the sample only in groups of four, arranged as a 4-triangle at random locations in the sample (as in Fig. 3 of the main text). In Figs. 22 and 23, we respectively display the density of states of $L \times L$ samples with $L = 160$ and $L = 180$. The $L = 160$ sample is diluted by 25 4-triangles placed at random on each sublattice, while the $L = 180$ sample is diluted with 40 4-triangles placed at random on each sublattice. The former sample corresponds to a "bare" value of $n_v \approx 0.0039$, while the latter sample corresponds to a bare value of $n_v \approx 0.0049$.

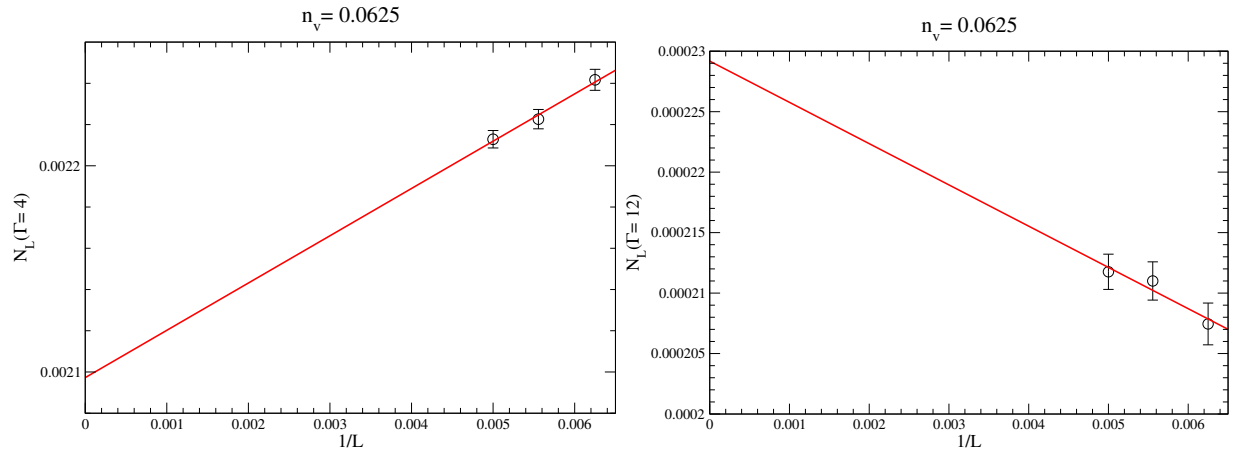


FIG. 15: Examples of extrapolation of $N_L(\Gamma)$ to the thermodynamic limit at $n_v = 0.0625$. For this concentration, $\Gamma_c \approx 7$ (see Fig. 4 in the main text), and the left panel illustrates the extrapolation for $\Gamma < \Gamma_c$, while the right panel is for $\Gamma > \Gamma_c$. Note in particular that our extrapolation for $\Gamma > \Gamma_c$ is very likely an overstimate, so one can be fairly confident that $N(\Gamma)$ in the thermodynamic limit drops below $N_{1D}(\Gamma)$, ruling out a fit to this form for $\Gamma > \Gamma_c$. Extrapolations at other values of Γ, as well as extrapolations including a larger size ($L = 220$) are shown in Fig. 26.

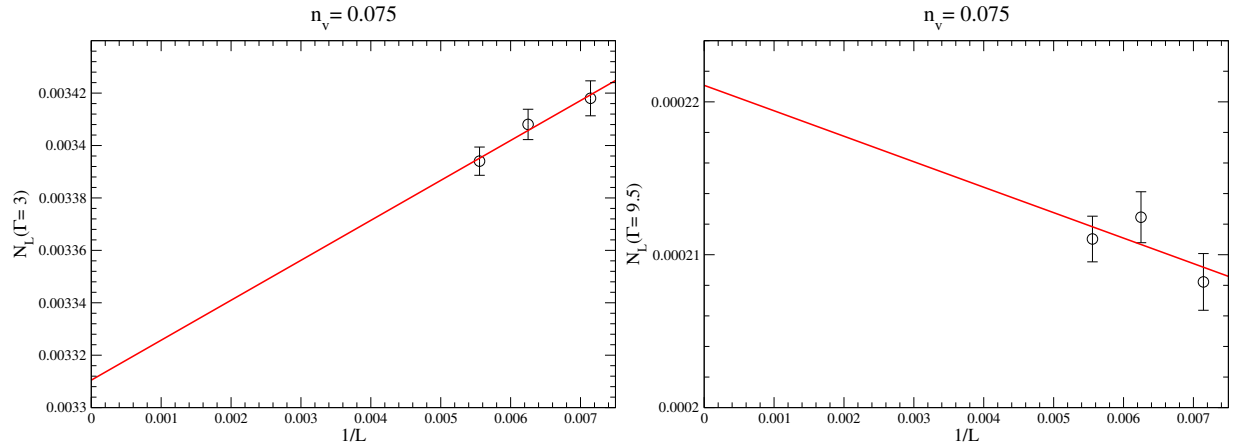


FIG. 16: Examples of extrapolation of $N_L(\Gamma)$ to the thermodynamic limit at $n_v = 0.075$. For this concentration, $\Gamma_c \approx 5$ (see Fig. 4 in the main text), and the left panel illustrates the extrapolation for $\Gamma < \Gamma_c$, while the right panel is for $\Gamma > \Gamma_c$. Note in particular that our extrapolation for $\Gamma > \Gamma_c$ is very likely an overstimate, so one can be fairly confident that $N(\Gamma)$ in the thermodynamic limit drops below $N_{1D}(\Gamma)$, ruling out a fit to this form for $\Gamma > \Gamma_c$.

These values of n_v are an order of magnitude different from the values of n_v studied by us in the main part of our work (in which the impurities are uncorrelated except for exclusion constraints designed to prevent the occurrence of "trivial" zero modes). However, since all vacancies go in as part of a 4-triangle, $w \approx 9.76 \times 10^{-4}$ for the $L = 160$ sample and $w \approx 1.23 \times 10^{-3}$ for the $L = 180$ sample. The values of w are thus very similar to those obtained in our independently diluted samples with n_v in the range 0.05—0.06.

From Figs. 22 and 23, we see that the density of states again undergoes a crossover that is qualitatively the same as the crossover identified in our main study. However, the corresponding Γ_c is much smaller (*i.e.*, the energy scale $|\epsilon_c|$ is much larger) than one would have expected based on the value of the overall vacancy concentration n_v (had the vacancies been independent as in the main study). Similarly, the value of y is also very different from the (extrapolated) value of y one would have expected at such small n_v. The corresponding histograms of Γ_g are shown in Figs. 24 and 25. From these figures, we see that Γ_g^*, corresponding to the position of the peak in the histogram of Γ_g, is significantly smaller than one would have expected based on the overall vacancy concentration n_v (had the vacancies been independent, as in the main study). This provides a simple illustration of the importance of spatial correlations between vacancies in setting the lowest gap scale Γ_g^*, and the density of zero modes w. It also emphasizes that the crossover identified by us is a robust and generic aspect of the low-energy physics of vacancy-disorder.

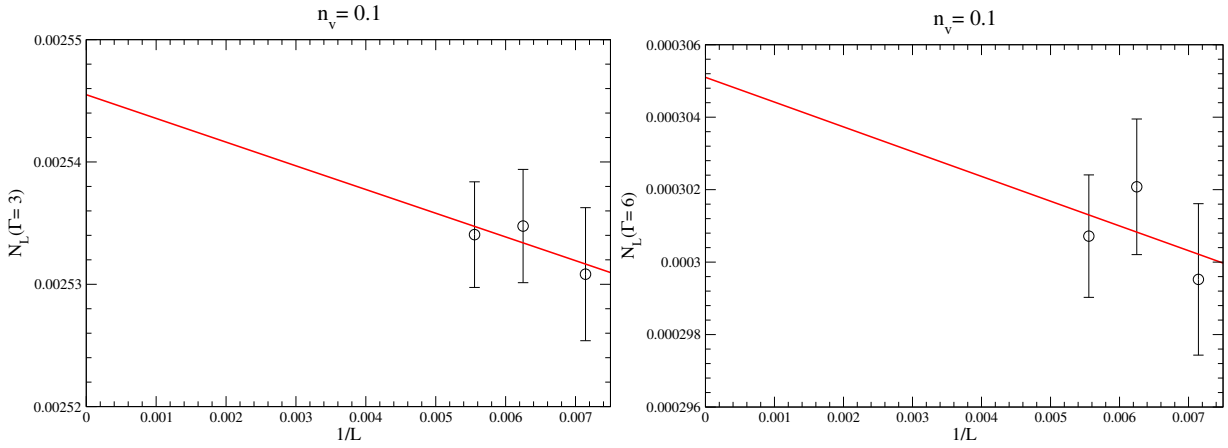


FIG. 17: Examples of extrapolation of $N_L(\Gamma)$ to the thermodynamic limit at $n_v = 0.1$. For this concentration, $\Gamma_c \approx 4$ (see Fig. 4 in the main text), and the left panel illustrates the extrapolation for $\Gamma < \Gamma_c$, while the right panel is for $\Gamma > \Gamma_c$. Note in particular that our extrapolation for $\Gamma > \Gamma_c$ is very likely an overstimate, so one can be fairly confident that $N(\Gamma)$ in the thermodynamic limit drops below $N_{1D}(\Gamma)$, ruling out a fit to this form for $\Gamma > \Gamma_c$.

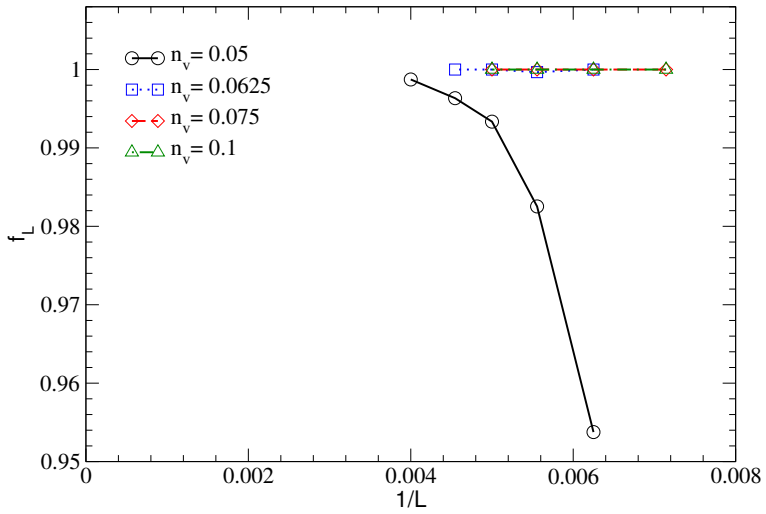


FIG. 18: Probability f_L that an $L \times L$ sample has at least one pair of zero modes tends to 1 in the thermodynamic limit for each concentration studied. Due to our computational constraints, we have been unable to obtain such data at $n_v = 0.02$, where we expect the density of zero modes to be much lower, but f_L to still tend to 1 in the thermodynamic limit (based on the analytical arguments given in the main text).

Finally, we note that the values of Γ_c and y in the case of dilution by 4-triangles are apparently predicted much better by the value of w (as opposed to the n_v). This raises the interesting questions already alluded to in the main text: Are Γ_c and y determined in a "universal" way (*i.e.*, independent of short-ranged correlations between vacancies and other such microscopic details) by the value of the zero-mode density w in the limit of small but nonzero w? Can this dependence be understood in terms of a low-energy effective theory or renormalization group approach?

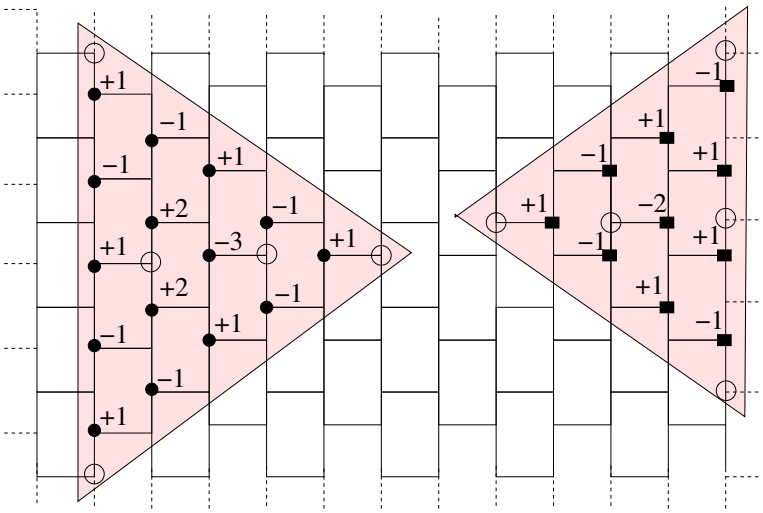


FIG. 19: Two kinds of 5-vacancy clusters ("5-triangles") that host an exact zero mode, with the corresponding wavefunction marked. Open circles correspond to vacancies.

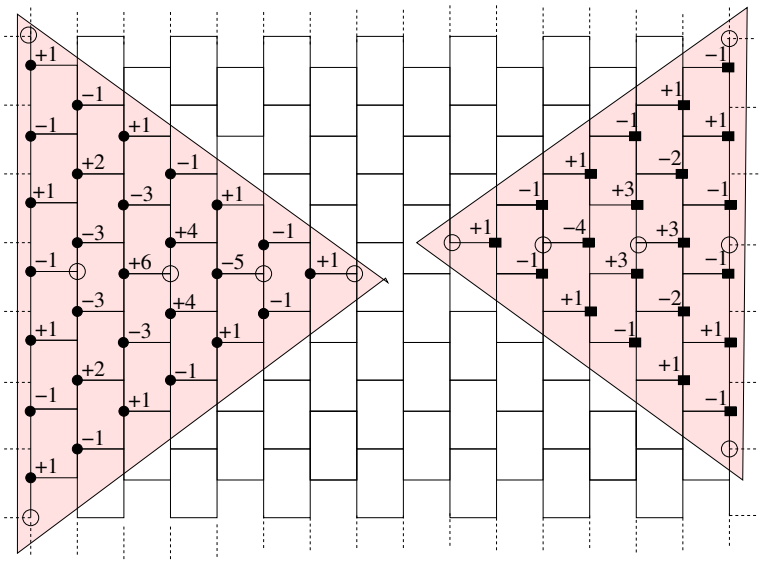


FIG. 20: Two kinds of 6-vacancy clusters ("6-triangles") that host an exact zero mode, with the corresponding wavefunction marked. Open circles correspond to vacancies.

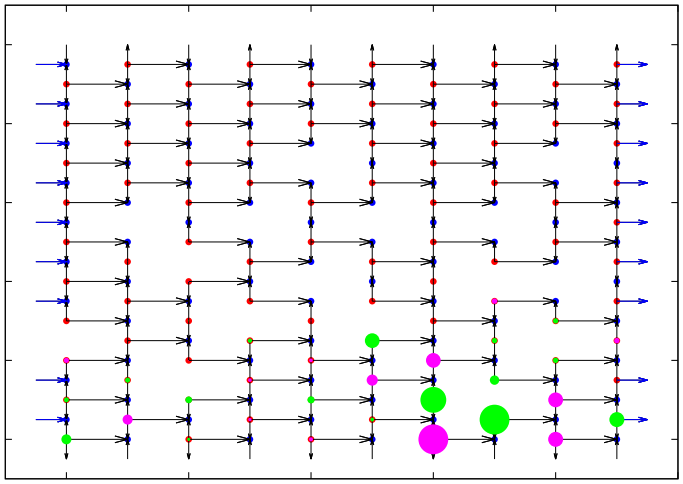


FIG. 21: An $L \times L$ semi-open sample (of the type consistently used in all our numerical work) with $L = 10$, with vacancies represented by missing lattice sites. This sample provides a simple example of a zero mode that does not seem to arise from any of the regular arrangements of vacancies used in our zero mode constructions. The actual wavefunction of this zero mode is represented by color-coded circles. The size of the circle corresponds to the magnitude of the wavefunction at the corresponding site, while the two different colors represent opposite signs for the wavefunction at the corresponding sites.

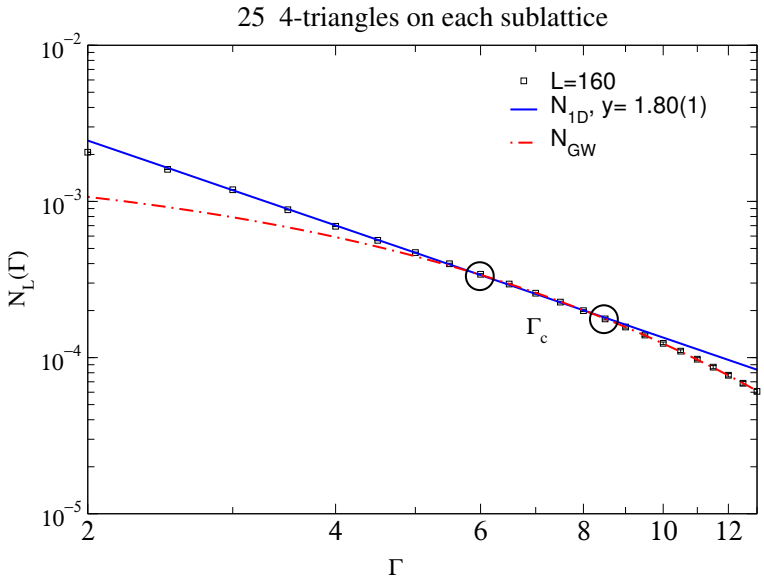


FIG. 22: $N_L(\Gamma)$ in the toy model in which a $L = 160$ sample is diluted with 25 randomly placed 4-triangles on each sublattice. Circles demarcate the crossover region centered at the crossover scale Γ_c. Data for $\Gamma \lesssim \Gamma_c$ fits well to power-law form $N_{1D}(\Gamma)$ with the value of y indicated in the figure, while the large-Γ regime fits well to the modified Gade-Wegner form $N_{GW}(\Gamma)$.

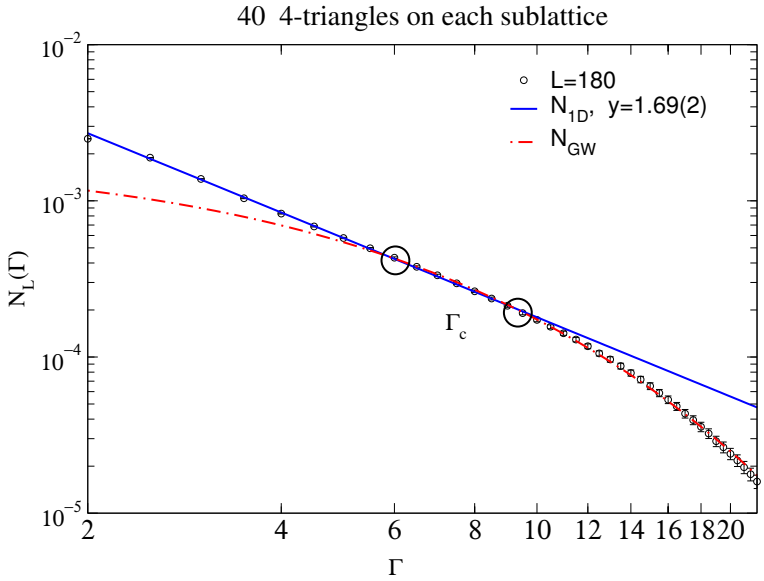


FIG. 23: $N_L(\Gamma)$ in the toy model in which a $L = 180$ sample is diluted with 40 randomly placed 4-triangles on each sublattice. Circles demarcate the crossover region centered at the crossover scale Γ_c. Data for $\Gamma \lesssim \Gamma_c$ fits well to power-law form $N_{\rm 1D}(\Gamma)$ with the value of y indicated in the figure, while the large-Γ regime fits well to the modified Gade-Wegner form $N_{\rm GW}(\Gamma)$.

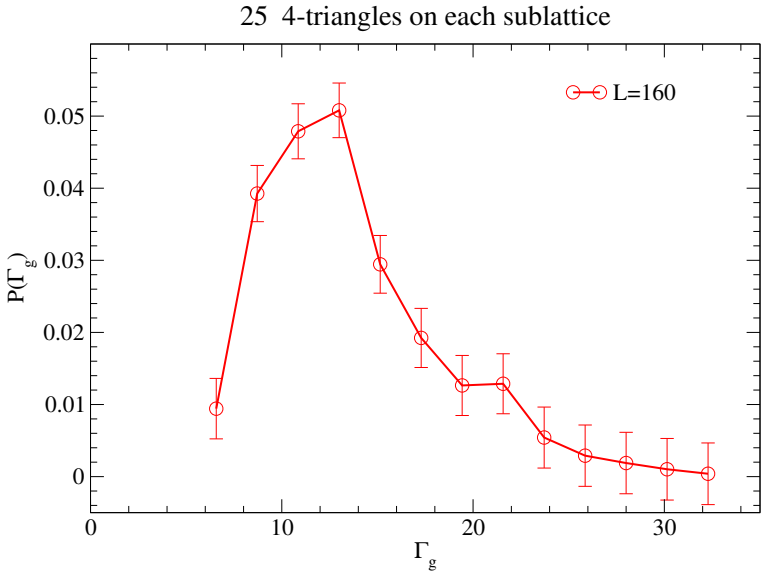


FIG. 24: Histogram of Γ_g, corresponding to the lowest nonzero gap for the $L = 160$ sample diluted with 25 randomly placed 4-triangles on each sublattice.

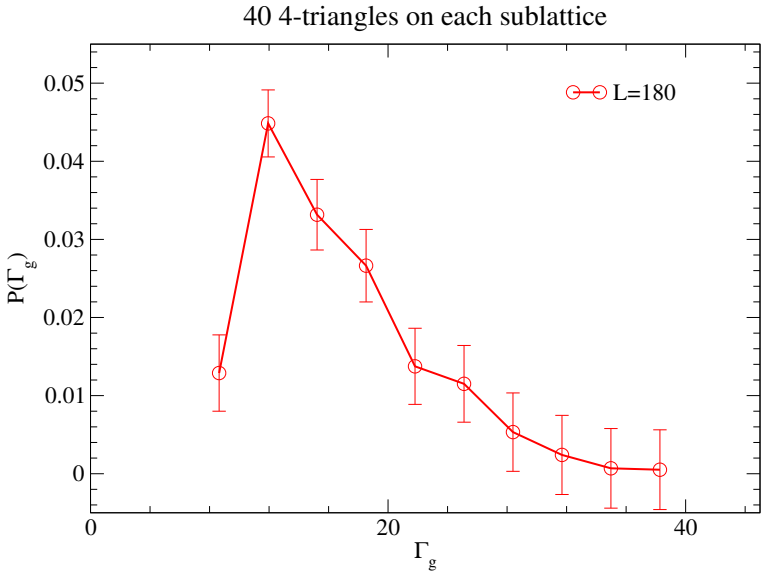
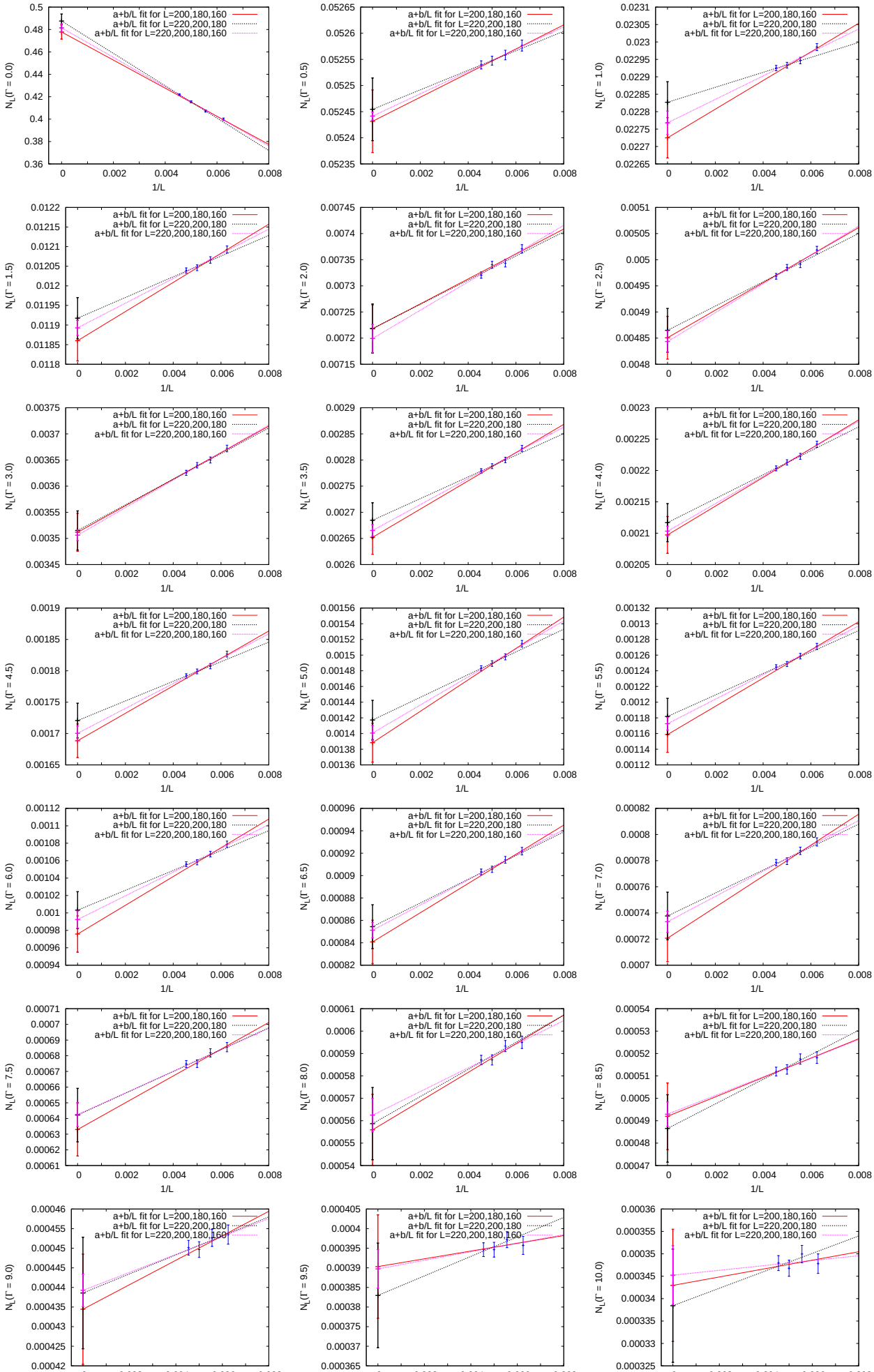
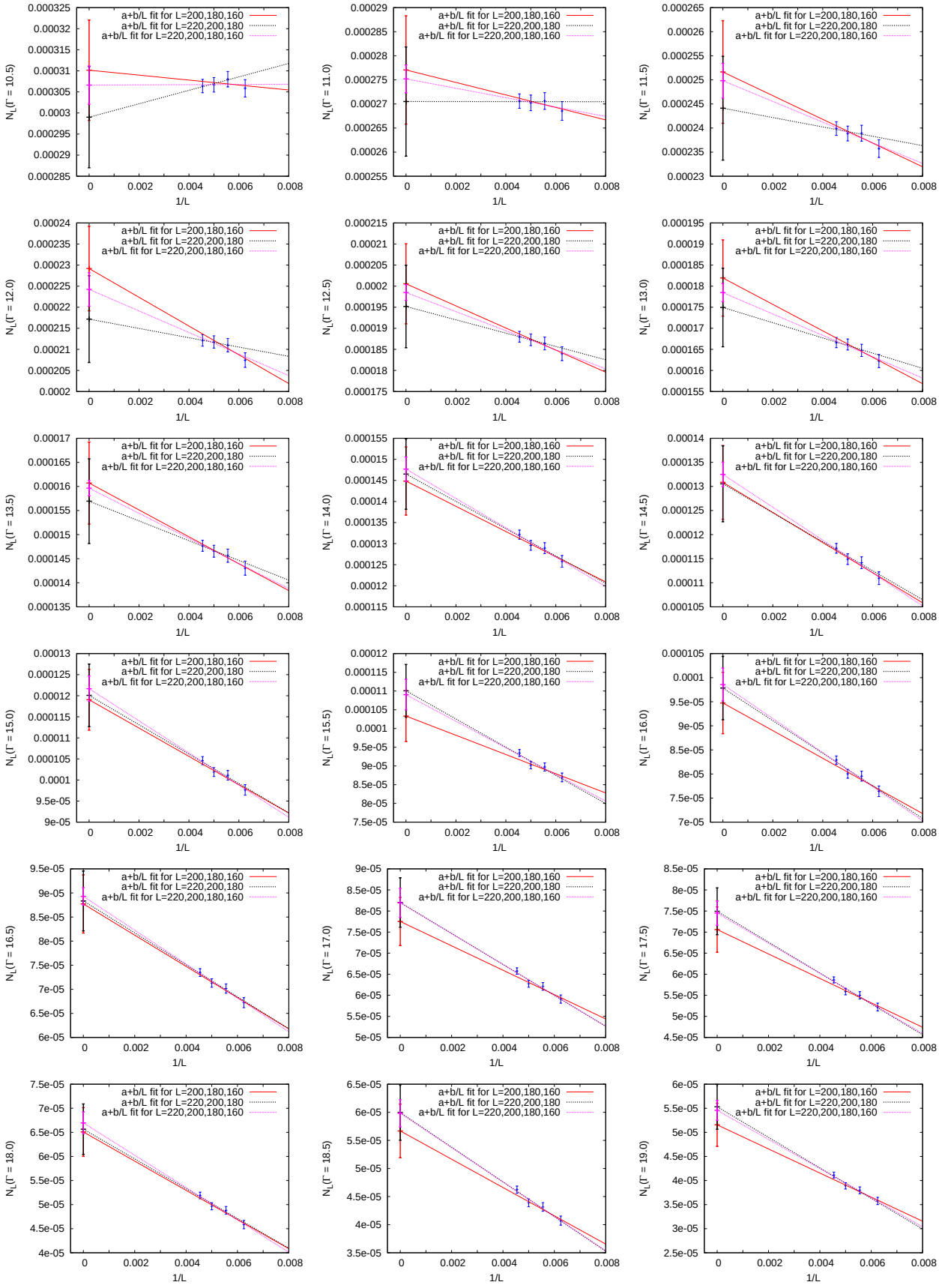


FIG. 25: Histogram of Γ_g, corresponding to the lowest nonzero gap for the $L = 180$ sample diluted with 40 randomly placed 4-triangles on each sublattice.


a+b/L fit for L=200,180,160
a+b/L fit for L=220,200,180
a+b/L fit for L=220,200,180,160
N_L(Γ = 0.0)
N_L(Γ = 0.5)
N_L(Γ = 1.0)
N_L(Γ = 1.5)
N_L(Γ = 2.0)
N_L(Γ = 2.5)
N_L(Γ = 3.0)
N_L(Γ = 3.5)
N_L(Γ = 4.0)
N_L(Γ = 4.5)
N_L(Γ = 5.0)
N_L(Γ = 5.5)
N_L(Γ = 6.0)
N_L(Γ = 6.5)
N_L(Γ = 7.0)
N_L(Γ = 7.5)
N_L(Γ = 8.0)
N_L(Γ = 8.5)
N_L(Γ = 9.0)
N_L(Γ = 9.5)
N_L(Γ = 10.0)
1/L


a+b/L fit for L=200,180,160
a+b/L fit for L=220,200,180
a+b/L fit for L=220,200,180,160
N_L(Γ = 10.5)
N_L(Γ = 11.0)
N_L(Γ = 11.5)
N_L(Γ = 12.0)
N_L(Γ = 12.5)
N_L(Γ = 13.0)
N_L(Γ = 13.5)
N_L(Γ = 14.0)
N_L(Γ = 14.5)
N_L(Γ = 15.0)
N_L(Γ = 15.5)
N_L(Γ = 16.0)
N_L(Γ = 16.5)
N_L(Γ = 17.0)
N_L(Γ = 17.5)
N_L(Γ = 18.0)
N_L(Γ = 18.5)
N_L(Γ = 19.0)
1/L

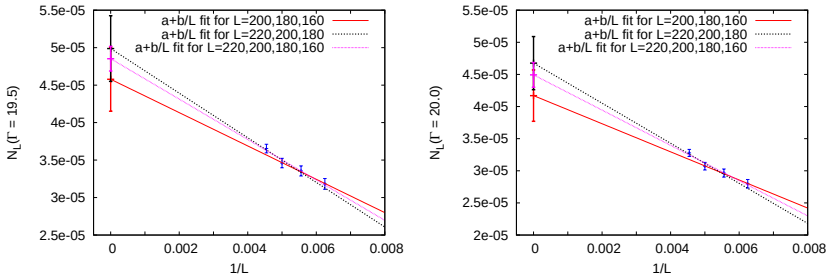


FIG. 26: Comparison of thermodynamic limit of $N_L(\Gamma)$ at $n_v = 0.0625$, taken with and without data at a larger size. At $n_v = 0.0625$, the estimated value of Γ_c is $\Gamma_c \approx 7$. Note that some correlation in the relative ordering of the three extrapolated values is expected over short intervals of Γ since these data are correlated (obtained from the same set of random samples). However, over the range of Γ from $\Gamma = 7$ to $\Gamma = 18$, we already see that there is no consistent ordering of the three extrapolated values, *i.e.*, the red points are not always higher than the black points or vice-versa. Additionally, the three different extrapolations fall within (or lie at the edge of) each other's error bars. Also, the results quoted in the main text (values of Γ_c and y in fits to N_{1D} and the quality of the different fits) do not change significantly if our analysis is performed on the thermodynamic limit $N(\Gamma)$ obtained by including data at the larger size. All this, taken together, provides compelling evidence that our extrapolations to the thermodynamic limit are reliable.