Torres VJB, Hajj Hussein R, Pages O, Rayson MJ. [Clustering/anticlustering effects on the GeSi Raman spectra at moderate \(Ge,Si\) contents: Percolation scheme vs. ab initio calculations](#). *Journal of Applied Physics* 2017, 121(8), 085704.

Copyright:

© 2017 American Institute of Physics. This article may be downloaded for personal use only. Any other use requires prior permission of the author and the American Institute of Physics. The following article appeared in *Journal of Applied Physics* 2017, **121**(8), 085704. and may be found at <https://doi.org/10.1063/1.4976861>

Date deposited:

24/05/2017

1 **Clustering/anticlustering effects on the GeSi Raman spectra at moderate (Ge,Si) contents:**
2 **Percolation scheme vs. *ab initio* calculations**

3
4 V.J.B. Torres,¹ R. Hajj Hussein², O. Pagès^{2,*} and M. J. Rayson³

5
6 ¹ *Departamento de Física and I3N, Universidade de Aveiro, Campus Santiago, 3810-193, Aveiro,*
7 *Portugal*

8 ² *Université de Lorraine, LCP-A2MC, Institut Jean Barriol, Metz, 57078, France*

9 ³ *School of Electrical, Electronic and Computer Engineering, Newcastle University,*
10 *Newcastle Upon Tyne NE1 7RU, United Kingdom*

11
12
13
14 **Abstract**

15
16 We test a presumed ability behind the phenomenological percolation scheme used for the
17 basic description of the multi-mode Raman spectra of mixed crystals at one dimension along the
18 linear chain approximation, to determine, via the Raman intensities, the nature of the atom
19 substitution, as to whether this is random or due to local clustering/anticlustering. For doing so we
20 focus on the model percolation-type $\text{Ge}_y\text{Si}_{1-y}$ system characterized by six oscillators $\{1 \times (\text{Ge} -$
21 $\text{Ge}), 3 \times (\text{Ge} - \text{Si}), 2 \times (\text{Si} - \text{Si})\}$, and place the study around the critical compositions $y \sim (0.16,$
22 $0.71, 0.84)$ corresponding to nearly matching of intensities between the like Raman modes from a
23 given multiplet ($\text{Ge} - \text{Si}$ triplet or $\text{Si} - \text{Si}$ doublet). The interplay between the $\text{Ge}_y\text{Si}_{1-y}$ Raman
24 intensities predicted by the percolation scheme depending on a suitable order parameter κ of local
25 clustering/anticlustering is found consistent with *ab initio* calculations of the $\text{Ge}_y\text{Si}_{1-y}$ Raman spectra
26 done with the *Ab Initio* Modeling PROgram (AIMPRO) code using large (64-, 216- and 512-atom)
27 disordered cubic supercells matching the required (y, κ) values. The actual ‘percolation vs. *ab initio*’
28 comparative insight at moderate/dilute-(Ge,Si) limits, with an emphasis on the κ -induced intra-bond
29 transfer of oscillator strength, extends a pioneering one earlier achieved at intermediate composition
30 ($y \sim 0.50$) by using small (32-atom) supercells [1], mainly concerned with the inter-bond transfer of
31 oscillator strength, providing altogether a complete picture.

* Author to whom correspondence should be addressed
Email address: olivier.pages@univ-lorraine.fr

I. INTRODUCTION

Generally, when dealing with semiconductor mixed crystals, the crucial issue is to determine whether the atom substitution is ideally random or due to local clustering/anticlustering.^{2,3} In the latter case, the next issue is to estimate the deviation with respect to the random substitution by introducing a relevant order parameter κ .

Vibrational spectroscopies are interesting with respect to the raised issues because they probe directly the force constant of a chemical bond,⁴ which is a local physical property. As such, they are expected to be sensitive to the local environment of a bond. In particular the Raman scattering is attractive because it is conveniently operated at the laboratory scale, non destructive, fast, and relates to the crystal in bulk (the Raman signal is averaged over the penetration depth of the laser probe, scaling up to several micrometers out of resonance condition).

Still, if we refer to the two phenomenological models traditionally used for the basic description at one dimension (1D), i.e. along the linear chain approximation, of the well-documented Raman spectra of the common semiconductor $A_{1-y}B_yC$ mixed crystals with zincblende structure, namely the modified-random-element-isodisplacement (MREI) model⁶ and the cluster model⁷, they both fail to provide a reliable insight into the nature of the atom substitution, for different reasons.

The MREI model presumes that the like bonds of a given species vibrate at the same frequency at a given composition, irrespectively of their local environment. As such the MREI model falls short of explaining already the 1-bond→multi-mode behavior apparent in the Raman spectra of most zincblende-type random mixed crystals (an overview is given, e.g., in Ref. 8), not to mention about the non random ones. In contrast, the cluster model distinguishes between the like bonds of a given species depending on their first-neighbor environment. As such, the latter model actually supports a 1-bond→multi-mode description of the Raman pattern of a mixed crystal, in apparent conformity with experimental findings.⁸ The problem is that the local environments in question are defined at three-dimension (3D), i.e. in the real crystal, and not at 1D as the equations of motions per atom. We have discussed elsewhere⁸ that such 1D-3D ambivalence behind the cluster model introduces a bias in the discussion of the Raman spectra of a mixed crystal, eventually resulting in a misleading insight into the nature of the atom substitution.

Over the past decade we have introduced a novel phenomenological model operating at 1D, i.e. the so-called percolation model (detailed, e.g., in Ref. 8), in order to explain the multi-mode behavior per bond apparent in the Raman spectra of most zincblende-type mixed crystals. This model distinguishes between different environments of a bond, in contrast with the MREI model, and, moreover, the environments in question are defined at 1D as the equations of motion per atom, and not at 3D as with the cluster model. The main drawbacks of both the MREI model and the cluster model thus seem to be overcome with the percolation model. This does not mean for all that, neither that the percolation model is generally valid for the description of the Raman spectra of the mixed crystals, nor that it is likely to produce a reliable insight into the nature of the atom substitution in a mixed crystal, the central issue in this work.

Concerning the “validity” aspect, we mention that over the past decade the percolation model has lead to a unified understanding of the long-standing MREI/cluster-based classification into four main types (two-mode, modified-two-mode, one-mode, multi-mode) of the abundant Raman (and infrared) spectra available in the literature for the common II-VI and III-V zincblende-type $A_{1-y}B_yC$ semiconductor mixed crystals (see Ref. 8 and Refs. therein). Moreover the percolation model has recently been shown to apply also to $\text{Ge}_y\text{Si}_{1-y}$ the leading mixed crystal from column IV in the

periodic table, with diamond structure. The proposed six-oscillator $[1 \times (Ge - Ge), 3 \times (Ge - Si), 2 \times (Si - Si)]$ version of the 1D-like percolation scheme for Ge_ySi_{1-y} (see detail below) solved all the persisting anomalies in the so far admitted $[1 \times (Ge - Ge), 1 \times (Ge - Si), 4 \times (Si - Si)]$ 3D-like description.⁸ Generally, the apparent universality of the percolation model inspires us with confidence regarding its validity.

The second issue, addressed in this work, is whether the percolation model is likely to provide a reliable insight into the nature of the atom substitution in a mixed crystal via the Raman intensities (the Raman frequencies are left by because the percolation model falls short of formalizing their κ -dependence), as we believe, or not. Our approach to investigate this issue, a theoretical one, is to confront the predictions of a proper κ -equipped version of the (1D) percolation model for a given mixed crystal taken as representative, with corresponding Raman spectra calculated *ab initio* using (3D) supercells matching the required composition (y) and κ values, the latter calculations constituting a sort of ultimate reference.

We have discussed elsewhere¹ why a diamond-type mixed crystal, such as Ge_ySi_{1-y} , is more relevant than a $A_{1-y}B_yC$ zincblende-type one for the test. The main reason is that in a zincblende-type mixed crystal the number of A-C and B-C bonds is fixed by the composition y , and thus also the intensities of the overall AC-like and BC-like Raman signals. In this case trends towards local clustering/anticlustering are only likely to modify the relative intensities of the like Raman features due to a given bond (provided a more refined description than the 1-bond $\rightarrow$ 1-mode MREI one applies). This we discuss in terms of a κ -induced *intra*-bond transfer of oscillator strength within a given 1-bond $\rightarrow$ multi-mode multiplet. Now, in a diamond-type mixed crystal such as Ge_ySi_{1-y} , the local clustering/anticlustering has also an effect on the individual bond fractions (Ge-Ge, Ge-Si and Si-Si in this case) at a given composition y . This leads to a κ -induced *inter*-bond transfer of oscillator strength between the Raman signals due to different bonds, with concomitant impact on the Raman intensities. Note that in the case of a diamond-type mixed crystal the MREI model can be used besides the percolation model to provide a quantitative insight into the nature of the atom substitution, i.e. a rather crude one though, since the MREI model omits by construction any fine structuring behind the Raman signal due to a given bond in a mixed crystal.

In brief, in a diamond-type mixed crystal the κ -induced inter-bond transfer of oscillator strength superimposes on top of the κ -induced intra-bond transfer of oscillator strength, whereas in a zincblende-type one only the latter is likely to occur.

A pioneering ‘percolation (1D) vs. (3D) *ab initio*’ comparative test, also incorporating the MREI insight for reference purpose, has earlier been done in the above spirit with the particular diamond-type $Ge_{0.5}Si_{0.5}$ (composition $y_0 = 0.5$) system by applying the *Ab Initio* Modeling PROgram (AIMPRO) code⁹ to small (32-atom) supercells.¹ The focus on composition y_0 was justified due to the related abundant data on the subject of clustering/anticlustering in the literature, experimental¹⁰⁻²⁰ as well as theoretical¹⁸⁻²³ ones.

In fact, the composition y_0 is especially interesting with respect to the κ -induced inter-bond transfer of oscillator strength. This is because the overall Ge-Ge, Ge-Si and Si-Si Raman signals exhibit comparable intensities in the random crystal ($\kappa = 0$, see Ref. 8, and refer also to the **Appendix Sec.**). In this case, any trend towards clustering/anticlustering, even a minor one, immediately generates an appreciable drop/emphasis of the Raman signal(s) due to the bond species driven to minority/majority.

An intuitive rule with clustering that the overall Raman signals due to the *homo* Ge-Ge and Si-Si bonds should reinforce to the detriment of the Raman signal due to the *hetero* Ge-Si bond (and

vice-versa with anticlustering) was actually observed in the $\text{Ge}_{0.5}\text{Si}_{0.5}$ *ab initio* Raman spectra, and also reproduced by the κ -equipped version of the $\text{Ge}_{0.5}\text{Si}_{0.5}$ percolation scheme worked out for the occasion.¹ A further intuitive trend with clustering that, among the like modes due to a given bond, those bond vibrations stemming from *homo*-like environments would be favored to the detriment of those taking place in *hetero*-like environments (and vice versa for anticlustering), in reference to the κ -induced intra-bond transfer of oscillator strength, was also validated by the percolation-type $\text{Ge}_{0.5}\text{Si}_{0.5}$ Raman lineshapes. The κ -dependence of the latter lineshapes was further found to be globally consistent with the existing sparse experimental data in the literature (Ge-Si triplet) and with the current *ab initio* calculations (Si-Si doublet).¹ However, the latter calculations are merely indicative and not conclusive owing to the rather poor statistics on the bond counting behind the concerned *ab initio* Raman peaks (recall the small supercell size).

Additional *ab initio* insight, notably outside y_0 , is definitely needed to secure the trends, especially in what regards the κ -induced intra-bond transfer of oscillator strength. Generally the effect of clustering/anticlustering on the $\text{Ge}_y\text{Si}_{1-y}$ Raman spectra outside y_0 has attracted little attention so far. We are only aware of one recent work by Reparaz *et al.* concerned with the κ -dependence of the Raman frequencies at moderate/small Ge content ($0 \leq y \leq 0.5$).²⁴ We recall that in our case the attention is centered on the Raman intensities.

In this work we extend outside y_0 , and with special attention to the κ -induced intra-bond transfer of oscillator strength, the ‘percolation (1D) vs. (3D) *ab initio*’ comparative test of the κ -dependent $\text{Ge}_y\text{Si}_{1-y}$ Raman lineshapes earlier accomplished at y_0 in Ref. 1. The role of *ab initio* calculations (done with the AIMPRO code) considerably increases in the present study, as much larger supercells (64-, 216- and 512-atom ones) have been constructed, in order to scan a much larger number of “realistic” configurations than in Ref. 1. This is essential to judge about the transferability of the conclusions drawn. The enlarged supercell size is further interesting in that it offers the possibility to displace the analysis towards moderate/dilute compositions y of most relevance with respect to the κ -induced intra-bond transfer of oscillator strength, while preserving a reliable statistics on the bond counting behind the like Raman modes from a given multiplet (Ge-Si or Si-Si). From its part, the κ -equipped version of the percolation model, entirely worked out in Ref. 1, offers an empirical yet robust systematization of trends within a 1D paradigm of the $\text{Ge}_y\text{Si}_{1-y}$ Raman spectra.

For the present study we have selected three critical compositions identifying remarkable intensity interplays (*RI*’s) within the percolation-type Ge-Si and Si-Si Raman multiplets of the random ($\kappa=0$) $\text{Ge}_y\text{Si}_{1-y}$ crystal. These correspond to nearly matching of intensities between the central Ge-Si Raman line and either its lower (RI_1 , $y_1 \sim 0.16$) or upper (RI_3 , $y_3 \sim 0.84$) satellite, and between the two Si-Si Raman lines (RI_2 , $y_2 \sim 0.71$).⁸

Last, for the sake of completeness, we repeat in the **Appendix Section** our *ab initio* calculations earlier done in Ref. 1 for representative κ values at composition y_0 (~ 0.50) using 32-atom supercells by applying the AIMPRO code to enlarged 64-, 216- and 512-atom supercells. Besides the novel *ab initio* insight searched at compositions (y_1 , y_2 , y_3), the *ab initio* update at y_0 is useful in view to complete a consistent comparison between the percolation and *ab initio* κ -dependent $\text{Ge}_y\text{Si}_{1-y}$ Raman lineshapes at all critical compositions y of the $\text{Ge}_y\text{Si}_{1-y}$ percolation scheme, in relation to both the κ -induced intra- and inter-bond transfers of oscillator strength.

The paper is organized as follows. In **Sec. II** we give details concerning the theoretical methods. The **sub-Sec. II-1** is used to recall how the definition of clustering/anticlustering enters the percolation model and which predictions follow from the latter concerning the intensities of the

Raman lines associated with the six basic oscillators included into the $\text{Ge}_y\text{Si}_{1-y}$ version of the percolation model, as originally formalized in Ref. 1. The **sub-Sec. II-2** explains the construction of supercells for probing different (y, κ) combinations, and specifies the technical details of the *ab initio* calculations of Raman spectra done on these supercells. **Sec. III** contains the general discussion accompanying the comparative ‘percolation vs. *ab initio*’ tests on the Raman intensities realized at the three compositions (y_1, y_2, y_3) . For the sake of completeness, we conclude the **Sec.** by a brief comparison of our current *ab initio* data with experiment – in reference to the recent work of Reparaz *et al.*,²⁴ in fact related to the effect of local clustering/anticlustering on the GeSi Raman frequencies. The conclusions drawn are reported in **Sec. IV**. An **Appendix Sec.** is introduced to repeat on a larger scale the *ab initio* calculations earlier done at composition y_0 in Ref. 1, for comparison of the *ab initio* data on equal footing at all (re-)examined compositions y .

II. THEORETICAL METHODS

1. κ –dependent version of the $\text{Ge}_y\text{Si}_{1-y}$ percolation scheme (1D)

In this **Sec.** we recall briefly how κ was originally introduced in the $\text{Ge}_y\text{Si}_{1-y}$ version of the percolation scheme worked out in Ref. 1, to facilitate the reading and to avoid extensive resort to the cited work.

At the term of a careful re-examination of the available experimental and theoretical $\text{Ge}_y\text{Si}_{1-y}$ Raman data in the literature, we have proposed that $\text{Ge}_y\text{Si}_{1-y}$ basically fits into a six-oscillator version of the percolation scheme.⁸ Out of the six main oscillators, only one relates to the long Ge-Ge bond, whereas the hetero Si-Ge bond and the short Si-Si one exhibit distinct three- and two-mode fine structures, respectively. The six oscillators in question are $[(\text{Ge} - \text{Ge}), (\text{Ge} - \text{Si})_2^{\text{Ge}}, (\text{Ge} - \text{Si})_2^{\text{GeSi}}, (\text{Ge} - \text{Si})_2^{\text{Si}}, (\text{Si} - \text{Si})_1^{\text{Ge}}, (\text{Si} - \text{Si})_1^{\text{Si+GeSi}}]$, ranked in order of increasing frequency. In this notation, each oscillator refers to a given bond-stretching (main label) taking place in a given environment whose composition (more Ge-, GeSi- or Si-like) and length scale (nearest-neighbor: 1, next-nearest-neighbor: 2) are specified via a superscript and a subscript, respectively. We stress that the latter “bond + environment” oscillators are defined at 1D along the linear chain approximation, and not in the real 3D crystal (as with the cluster model).

In working out the κ -dependent version of the $\text{Ge}_y\text{Si}_{1-y}$ percolation scheme, the crucial step was to express the individual fractions f of the six oscillators, which directly control the related Raman intensities, via a relevant order parameter κ of local clustering/anticlustering. In fact κ was introduced so as to pinpoint a trend towards clustering/anticlustering of the like substituting atoms from a given species (Ge or Si) when the latter gets minor in the crystal.¹ In doing so we followed an approach earlier proposed by Verleur and Barker in their cluster model⁷ with a basic difference that κ operates at 1D in the percolation model, and at 3D in the cluster model. For example, the probability of finding one Si atom nearby another Si atom in a 1D description of the $\text{Ge}_y\text{Si}_{1-y}$ diamond-type lattice is $P_{\text{SiSi}} = (1 - y) + \kappa \times y$, and likewise for P_{GeGe} . The random $\text{Ge} \leftrightarrow \text{Si}$ substitution corresponds to $\kappa=0$, whereas $\kappa<0$ and $\kappa>0$ refer to anticlustering and to clustering, respectively. At $y = y_0 = 0.5$, the extreme κ values are 1 ($P_{\text{SiSi}} = P_{\text{GeGe}} = 1$) and -1 ($P_{\text{SiSi}} = P_{\text{GeGe}} = 0$), corresponding to full clustering (only Ge-Ge and Si-Si bonds are present in the crystal) and full anticlustering (the crystal is only constituted of hetero bonds), respectively. We emphasize that the full anticlustering cannot fully develop outside y_0 . This is because the dominant atom species has to remain segregated to some

extent, at least locally. Technically, the anticlustering limit is reached when the fraction of the minor bond species vanishes to zero. The individual fractions of oscillators eventually express as follows versus P_{SiSi} and P_{GeGe} ,¹ which contain the κ -dependence,

$$f_1 = f\{(Ge - Ge)\} = y \times P_{GeGe} \quad (1a)$$

$$f_2 = f\{(Ge - Si)_2^{Ge}\} = 2y^2 \times (1 - P_{GeGe}) \times [P_{GeGe} - y \times (1 - P_{GeGe}^2)] \quad (1b)$$

$$f_3 = f\{(Ge - Si)_2^{GeSi}\} = 2(1 - y) \times (1 - P_{SiSi}) - [f\{(Si - Ge)_2^{Ge}\} + f\{(Si - Ge)_2^{Si}\}] \quad (1c)$$

$$f_4 = f\{(Ge - Si)_2^{Si}\} = 2(1 - y)^2 \times (1 - P_{SiSi}) \times [P_{SiSi} - (1 - y) \times (1 - P_{SiSi}^2)] \quad (1d)$$

$$f_5 = f\{(Si - Si)_1^{Ge}\} = (1 - y)^2 \times (1 - P_{SiSi})^2 \quad (1e)$$

$$f_6 = f\{(Si - Si)_1^{GeSi+Si}\} = 2(1 - y)^2 \times P_{SiSi} \times (1 - P_{SiSi}) + (1 - y) \times [P_{SiSi} - (1 - y) \times P_{SiSi}^2]. \quad (1f)$$

The phenomenological contour modeling of the Ge_ySi_{1-y} Raman spectra within the percolation scheme is realized by taking the imaginary part of the six-oscillator dielectric function of the Ge_ySi_{1-y} mixed crystal expressed in its classical form, in which each oscillator is represented by a Lorentzian function. The above (y, κ) -dependent f_i –terms come at the numerator of the six Lorentzian functions, and, as such, govern the intensities of the corresponding Raman lines. In fact the f_i –terms act as weighting factors on the intrinsic Raman efficiencies of the (related) $Ge - Ge$, $Ge - Si$ and $Si - Si$ bonds. The latter were determined by *ab initio* calculation of the Raman spectra applying the AIMPRO code to a pure-Ge supercell, a zincblende-type SiGe supercell (containing Si-Ge bonds only), and a pure-Si supercell, correspondingly.⁸ As for the denominator of each Lorentzian function, this contains the information on both the frequency (corresponding to a divergence of the Lorentzian function) and the linewidth at half maximum (damping term) of the Raman line.

The composition dependence of the relevant Raman frequencies of the random Ge_ySi_{1-y} mixed crystal ($\kappa=0$) are well documented in the literature, both experimentally and theoretically. Representative values taken from Refs. 21 and 25-32 were conveniently regrouped in Ref. 8 (see Fig. 2 therein). As for the damping terms we use the values carefully measured by Brya.²⁸ The Raman frequencies used in this work for the studied y_0 (refer to the **Appendix Sec.**) y_1 , y_2 and y_3 compositions are [~ 291 (~ 10), ~ 395 (~ 8), ~ 408 (~ 8), ~ 423 (~ 8), ~ 472 (~ 6), ~ 488 (~ 6)], [~ 283 (~ 13), ~ 392 (~ 10), ~ 405 (~ 10), ~ 435 (~ 10), ~ 490 (~ 4), ~ 510 (~ 4)], [~ 295 (~ 8), ~ 390 (~ 8), ~ 404 (~ 8), ~ 423 (~ 8), ~ 461 (~ 6), ~ 473 (~ 6)] and [~ 298 (~ 5), ~ 388 (~ 8), ~ 398 (~ 8), ~ 418 (~ 8), ~ 453 (~ 6), ~ 462 (~ 6)] given in cm^{-1} , respectively, with corresponding damping values specified in brackets (taken identical for the like modes of a given multiplet), also expressed in cm^{-1} . In a crude approximation we assume that at any studied composition y neither the Raman frequency nor the phonon damping are κ –dependent, for simplicity.

The κ -dependent version of the Ge_ySi_{1-y} percolation scheme uses, as its main parameters, the (κ -independent) frequencies of individual oscillators and their (κ -dependent) Raman intensities (via the f_i –terms). Both series of parameters are summarized in **Fig. 1**. In fact, the latter consists of a mere variant of Fig. 1 from Ref. 1, and may seem redundant. However, it is needed to set the basis of the present study while avoiding extensive resort to our previous work. Note that in **Fig. 1** the Raman frequencies are reported without link to the different sources (these are given in Ref. 8 – see Fig. 2). The dominant (thick-plain lines) and minor (thin-dashed lines) Raman features in each composition domain are schematically distinguished, for clarity, emphasizing (gray rectangles) those critical compositions studied in this work, corresponding to remarkable intensity interplays (*RI*'s) between the like sub-modes from a given multiplet ($Ge-Si$ and $Si-Si$).

2. *Ab initio* calculations using (3D) supercells

The *ab initio* calculations of the $\text{Ge}_y\text{Si}_{1-y}$ Raman spectra are done with a pseudopotential spin density-functional supercell code (AIMPRO),^{9,33} using the local density approximation for the exchange-correlation potential by Perdew and Wang,³⁴ taking the pseudopotentials for Si and Ge as proposed by Hartwigsen *et al.*³⁵ The calculation setup is the same as detailed in Ref. 36, notably in what regards the norm-conserving pseudopotentials and the basis functions. The novelty in the actual extension lies in the larger size of the used $\text{Ge}_y\text{Si}_{1-y}$ cubic supercells (generally 64-atom and 216-atom ones, and occasionally 512-atom ones,³⁷ against 32-atom ones in our pioneering contribution in Ref. 1). This offers a possibility to explore moderate/small substitution rates while preserving a reliable statistics on the minor-bond counting at all addressed (y, κ) values.

One critical issue is the fine adjustment of the desired κ value (κ_d) for a given $\text{Ge}_y\text{Si}_{1-y}$ supercell prior to realizing the *ab initio* calculation of its Raman spectrum. For doing so we implemented a simulated annealing algorithm³⁸ operating as follows. The initial step is to generate a preliminary $\text{Ge}_y\text{Si}_{1-y}$ supercell at the required composition y by assigning the Ge sites from random numbers. The $\text{Ge}-\text{Ge}$, $\text{Si}-\text{Ge}$ and $\text{Si}-\text{Si}$ bond fractions counted in the as-obtained supercell of the first generation are then compared with those expected at the considered κ value, given by,

$$f\{(\text{Ge}-\text{Ge})\} = f_1, \quad (2a)$$

$$f\{(\text{Ge}-\text{Si})\} = f_2 + f_3 + f_4, \quad (2b)$$

$$f\{(\text{Si}-\text{Si})\} = f_5 + f_6, \quad (2c)$$

respectively, being clear that the f_i -terms have to fulfill the basic summation rule $\sum_{i=1}^6 f_i = 1$, which ensures the conservation of the total number of bonds in the supercell.

Consider, e.g., that we search for a random $\text{Ge} \leftrightarrow \text{Si}$ substitution ($\kappa_d = 0$), and that the actual bond counting in the first-generation supercell gives $\kappa = \kappa_1 = 0.2$ thus indicating a (minor) trend towards local clustering. This means that the homo $\text{Ge}-\text{Ge}$ and $\text{Si}-\text{Si}$ bonds in the supercell are over-represented with respect to the hetero $\text{Ge}-\text{Si}$ ones. In order to converge onto the desired κ_d value, our approach is, first, to replace at random one Ge atom by a Si atom, and, at the same time, one Si atom by a Ge atom, thus preserving the same stoichiometry. This new configuration obtained so far has, in most of the cases, a different κ parameter, say κ_2 . This configuration is accepted if κ_2 is closer to the desired κ_d than κ_1 , otherwise it is accepted with $P(T) = e^{-|\kappa_2 - \kappa_d|/T}$ probability. We repeat this procedure the times needed to achieve the desired κ_d . The T parameter is chosen in such a way to produce high acceptance probability $P(T) \sim 1$ in the first steps, and is forced to decrease slowly as the number of steps increases. As the κ parameter has discrete values, we achieve a $\text{Ge}_y\text{Si}_{1-y}$ configuration with the desired κ_d value by just imposing a suitable κ cut-off. The picture naturally extrapolates to anticlustering, and, more generally, to any desired κ_d value at any composition y .

The *ab initio* κ -dependent $\text{Ge}_y\text{Si}_{1-y}$ Raman spectra have been calculated at three representative κ values for each of the considered y_1 (~ 0.16), y_2 (~ 0.71) and y_3 (~ 0.84) compositions using both 64-atom and 216-atom supercells, plus additional 512-atom supercells for the reference insights at $\kappa = 0$ (random substitution). While the calculations were systematically duplicated with the 64-atom supercells, the duplication was only exceptional with the 216-atom supercells, and generally absent with the 512-atom ones (except at composition $y_0 \sim 0.50$, see [Appendix Sec.](#)) due to the forbidding computational time. The comparison between the like *ab initio* insights obtained for a given pair of (y, κ) values using 64-, 216- and 512-atom supercells helps to distinguish between fortuitous Raman features due to a particular atom distribution in the supercell, that are not worth

to discuss, and robust/reproducible trends, that deserve an explanation. In fact, the Raman spectra calculated by using the 64-atom supercells (an example is given in the **Appendix Sec.**) appear to be contaminated by spurious features which disappear in the Raman spectra obtained with the 216- and 512-atom supercells. Moreover the fluctuations in the Ge-Ge, Ge-Si and Si-Si Raman frequencies appear to be rather large when changing the atom arrangement inside the small (64-atom) supercells at fixed (y, κ) values – up to 10 cm^{-1} at, e.g., composition y_0 (not shown), but reduce by almost one order of magnitude when using large (216- and 512-atom) supercells (detail is given below). Therefore, in the following discussion, we focus on the *ab initio* data obtained with the latter supercells, of a better quality than those generated by the small (64-atom) ones.

Note that the simulated annealing method that we use to generate our random supercell ($\kappa=0$) at a given composition y , based on a mere bond-counting procedure (see above) and thus operative at the first-neighbor scale only, significantly differs from that used by Zunger *et al.*³⁹ to design their own Special Quasirandom Structures (SQS's) that closely mimic even with a small number of atoms the pair-correlation functions characteristic of a perfectly random mixed crystal at a large length scale. If we refer, e.g., to the $\text{A}_{0.5}\text{B}_{0.5}$ diamond-type random mixed crystal, the 8-atom SQS suffices to reproduce the first- and second-neighbors radial correlation functions of the infinite crystal. The agreement extends to third-neighbors with the 64-atom SQS.³⁹ Certainly the agreement is not as good in our case, but still remains of high quality. For a direct insight we report the pair correlation functions of our large (216- and 512-atom) random ($\kappa \sim 0$) supercells up to seventh-neighbors calculated along the method given by Venezuela *et al.*⁴⁰ as **Supplementary Material**.

Further we compare in the **Appendix Sec.** (see **Fig. A1** therein) the *ab initio* Raman spectra of the random $\text{Ge}_{0.5}\text{Si}_{0.5}$ mixed crystal obtained by using, on the one hand, our large-size (216- and 512-atom) supercells (plain curves) optimized via the simulated annealing algorithm, and, on the other hand, a smaller (64-atom) SQS (dotted curve) build up by expanding the 16-atom SQS used by Chroneos *et al.*⁴¹ by $1 \times 2 \times 2$ along the procedure used by the cited authors in Ref. **42**. In fact, the two series of *ab initio* Raman spectra compare fairly well in terms of the Raman intensities – our main concern in this work, if not in terms of the Raman frequencies (recall the inherent error in the Raman frequencies when using small supercells). We conclude that our large-size supercells designed at $\kappa=0$ via the simulated annealing algorithm constitute a reasonable approximation of the random mixed crystals, at least for our use. The simulated annealing algorithm is further attractive for the current investigation of the effect of local clustering/anticlustering on such spectra since it can be implemented to achieve any κ value at any composition y . In contrast, the SQS's are meant for $\kappa=0$ only, as can be inferred from their name.

III. COMPARATIVE 'PERCOLATION Vs. AB INITIO' INSIGHT

1. κ -dependent percolation-type $\text{Ge}_y\text{Si}_{1-y}$ Raman spectra

Representative κ -dependent Raman lineshapes calculated with the κ -equipped version of the $\text{Ge}_y\text{Si}_{1-y}$ percolation scheme (thick lines) at the critical compositions y_1 (a, $\sim 16 \text{ at.}\% \text{ Ge}$), y_2 (b, $\sim 71 \text{ at.}\% \text{ Ge}$) and y_3 (c, $\sim 84 \text{ at.}\% \text{ Ge}$), corresponding to nearly matching intensities (refer to the dotted boxes) between the central Ge-Si mode and its upper satellite (a, RI_1), between the two Si-Si modes (b, RI_2) and between the central Ge-Si mode and its lower satellite (c, RI_3) in the random

mixed crystals ($\kappa = 0$), are displayed in **Fig. 2** (thick lines). The corresponding MREI curves (thin lines) are added, for reference purpose.

In **Fig. 2**, The Raman intensities are directly comparable within each panel and also between the different panels, provided the corrective factors indicated at the top of each panel (applying to all the spectra) and nearby a given peak (specific to the peak then) are taken into account. Well-spanned κ values throughout the accessible κ -domain are considered at each composition y , from full clustering ($\kappa = 1$) down to the accessible anticlustering limit corresponding to disappearance of the minor bond species. Critical κ values associated with extinction of a given oscillator are emphasized on the right hand side of each panel (refer to the f_i -terms that have vanished to zero), for a better appreciation as how the microstructure of the crystal modifies with κ .

As already mentioned, the basic trend with clustering is that the Raman features due to the homo $Ge - Ge$ and $Si - Si$ bonds reinforce to the detriment of those due to the hetero $Ge - Si$ bond, and vice versa for anticlustering. This is already apparent in the crude MREI-like three-oscillator [$1 \times (Ge - Ge)$, $1 \times (Ge - Si)$, $1 \times (Si - Si)$] κ -dependent Raman lineshapes (thin curves) which do not take into account the Ge-Si and Si-Si fine structures. In such contour modeling the dominant feature in each multiplet (either Ge-Si or Si-Si) was attributed the whole oscillator strength available for the relevant bond at the considered composition y , using the f_i -terms given by **Eq. (2)**. Such MREI-like ‘de-zooming’ into the Ge_ySi_{1-y} Raman spectra is interesting, in that, by omitting the $Ge - Si$ and $Si - Si$ fine structures, it reveals how the inter-bond transfer of oscillator strength, taken alone, develops with clustering/anticlustering.

If we ‘zoom’ into the detail of the Ge-Si and Si-Si fine structures in their κ -dependence, a further intuitive trend in case of clustering is that, within a given percolation-type multiplet (Ge-Si or Si-Si), those particular Raman features due to the host environments rich of the homo Ge-Ge and Si-Si bonds should be favored to the detriment of those dominantly formed with the hetero Ge-Si species, and vice versa for anticlustering. Such κ -induced intra-bond transfer of oscillator strength besides the MREI-like κ -induced inter-bond transfer of oscillator strength, is well reproduced in the refined percolation-type six-oscillator [$1 \times (Ge - Ge)$, $3 \times (Ge - Si)$, $2 \times (Si - Si)$] κ -dependent Raman lineshapes (thick curves) derived on the basis of **Eq. (1)**, at any y value.

As far as the Si-Si doublet is concerned, the κ -induced intra-bond transfer of oscillator strength is especially clear at small/moderate Si content (i.e. at y_2 and y_3 – see **Figs. 2b** and **2c**). At this limit, a mere change in sign of κ suffices to generate a dramatic inversion between the two Si-Si sub-modes – the dominant mode at positive κ values becoming minor at negative ones, before co-extinction of the two modes at the anticlustering limit. At large Si content (y_1 – see **Fig. 2a**) the Si-Si bonds have to vibrate dominantly in their own Si-like environment at any κ value. With this, the Si-Si doublet basically reduces to its upper component, the Raman intensity of which remains quasi stable when κ changes. As for the lower component, due to Si-Si bonds vibrating in presence of Ge, this is hardly detectable owing to the small Ge content, even close to the anticlustering limit.

The κ -dependence of the $Ge - Si$ Raman triplet also conforms to intuition. With anticlustering ($\kappa < 0$), the central feature due to Si-Ge bonds vibrating in the (Si,Ge)-mixed environment is favored to the detriment of its lower and upper satellites which stem from the Ge- and Si-like environments, respectively. In contrast, with clustering ($\kappa > 0$) the satellites reinforce, and even tend to dominate the central Ge-Si feature that progressively collapses, just before extinction of the three features at maximum clustering ($\kappa = 1$). In fact, out of the two Ge-Si satellites, mainly one is activated throughout a given κ -domain, namely that due to the environment rich of the dominant atom species at the considered composition y . The alternative satellite, due to environments formed

with the minor species, hardly shows up, and only as a transient feature in the clustering regime before the extinction of the Ge-Si signal.

The above discussed trends concerning the κ -induced inter-bond (straight arrows, outside the frame) and intra-bond (curved arrows, inside the frame) transfers of oscillator strength under clustering/anticlustering are captured by explicit arrows in **Fig. 1**. Additional detail concerning the κ -dependence of the percolation-type Raman lineshapes is given in the course of the comparison with the *ab initio* Raman spectra in the next **Sec.**

2. κ -dependent *ab initio* $\text{Ge}_y\text{Si}_{1-y}$ Raman spectra

The *ab initio* κ -dependent $\text{Ge}_y\text{Si}_{1-y}$ Raman spectra calculated at compositions y_1 (**a**), y_2 (**b**) and y_3 (**c**) by using the 216- and 512-atom supercells (only at $\kappa = 0$ for the largest supercells) are displayed in **Fig. 3**. Additional insight at composition y_0 , is reported in the **Appendix Sec.** Large simple/double arrows are used to emphasize the κ -induced changes in the Raman intensities near the pointed frequencies with respect to the reference Raman signals at $\kappa = 0$. A line broadening of 4 cm^{-1} has been retained for all compositions y and κ values. This is large enough to be realistic and at the same time sufficiently small to allow the resolution of neighboring percolation-type Raman features, the closest ones being separated by $\sim 10 \text{ cm}^{-1}$ (see **Figs. 1** and **2**). As ideally expected the nearly matching of intensities (emphasized by dashed frames in **Fig. 2**) between, on the one hand, the central Ge-Si feature and its lower (y_1 , RI_1 – see **Figs. 2a** and **3a**) and upper (y_3 , RI_3 – see **Figs. 2c** and **3c**) satellites, and, on the other hand, the two Si-Si features (y_2 , RI_2 – see **Figs. 2b** and **3b**), are well reproduced in the *ab initio* Raman spectra of the random mixed crystals ($\kappa = 0$), and moreover at the required Raman frequencies (within less than $\sim 5 \text{ cm}^{-1}$), if we refer to the used values for the percolation-type contour modeling (given in **Sec. II-1**).

A clear separation between the Ge-Ge, Ge-Si and Si-Si spectral ranges, needed in view of the fore coming discussion of the *ab initio* spectra in their κ -dependence (next **sub-Secs.**), was achieved by calculating, via the square of the atomic vibration amplitude, the atomic localization of critical vibration modes presumably situated at the frontier between the spectral ranges of different bonds.

One clear frontier, transverse to all examined compositions, referred to as the lower frontier below, occurs around $\sim 350 \text{ cm}^{-1}$ between the low-frequency vibrations of the heavy-Ge atoms and the high-frequency vibrations involving the light-Si atoms. While the Ge atoms can vibrate only against alternative Ge atoms, the Si ones may vibrate as well against Ge atoms, the latter remaining (quasi) inert then on account of their larger mass, or against like vibrating Si atoms. An additional frontier at high frequency, referred to as the upper frontier hereafter, is thus needed to further separate the latter two kinds of vibrations involving Si.

A decisive insight with respect to the upper frontier is obtained by placing the analysis at intermediary Ge content, concentrating on composition y_2 ($\sim 71 \text{ at.}\% \text{ Ge}$). At large clustering ($\kappa \sim +0.6$) the sharp/dominant Raman feature showing up around 488 cm^{-1} is due to pairs of neighboring Si atoms vibrating in their own (Si-like) environment at the first-neighbor scale. Each atom of the pair is surrounded by at least three, but usually four, Si atoms. Interestingly, none of the Si atom surrounding the central Si-pair vibrates, while the two paired Si atoms vibrate with nearly identical amplitudes (within less than 0.5%). Close to the anticlustering limit ($\kappa \sim -0.2$), the minor feature located at $\sim 460 \text{ cm}^{-1}$ likewise originates from the vibration of like Si-Si pairs, an unique one in fact in the considered supercell. Remarkably the latter takes place in an ultimate Ge-like environment in that each of the Si atom is surrounded by three Ge atoms besides its Si partner in the pair. The as-

identified ‘atom+environment’ vibrating units in the clustering and anticlustering limits, sketched out in **Fig. 3b**, help to delimit the spectral range due to Si atoms vibrating against like Si atoms in our *ab initio* spectra, thus providing an insight into the upper frontier from the high frequency side. The concerned spectral range identifies with $455 - 490 \text{ cm}^{-1}$. The intermediary spectral range between $\sim 350 \text{ cm}^{-1}$ (lower frontier) and $\sim 455 \text{ cm}^{-1}$ (upper frontier) is attributed to Si atoms vibrating against foreign Ge atoms. In fact, we have checked that the sharp feature pointing at $\sim 400 \text{ cm}^{-1}$ that dominates the Raman spectrum near the anticlustering limit is due to isolated vibrations of Si atoms surrounded by four Ge atoms. This is consistent with this mode being assigned in terms of Si vibrations against Ge, thus validating the upper frontier from the low frequency side.

We consider the same two frontiers for compositions y_3 (84 at.% Ge) and y_1 (16 at.% Ge) because, in view of **Fig. 1**, the frontiers do not appear to be largely composition-dependent.

When shifting the above discussion of atom (Ge or Si) vibrations in the real 3D crystal towards the 1D percolation scheme, in which the relevant vibrating unit consists of the stretching of a chemical bond, the considered two frontiers at $\sim 350 \text{ cm}^{-1}$ and $\sim 455 \text{ cm}^{-1}$ separate, in fact, the Ge-Ge, Ge-Si and Si-Si spectral ranges, in order of increasing frequency (as schematically indicated by dotted double-arrows at the top of **Fig. 3b**). On this basis, we may readily assign the prototypal vibrating units of Si-Si pairs in Ge-rich and Si-rich environments discussed above as being due to the percolation-type $(\text{Si} - \text{Si})^{\text{Ge}}$ and $(\text{Si} - \text{Si})^{\text{GeSi, Si}}$ oscillators. As for the remaining three $(\text{Ge} - \text{Si})$ percolation-type oscillators, these will be identified in the course of the ‘percolation vs. *ab initio*’ discussion hereafter.

Generally, the discussion of the upper Ge-Si satellite, generally a very minor one, is a complicated issue at large Ge contents ($y_0 \sim 0.5$, $y_2 \sim 0.71$ and $y_3 \sim 0.84$) because it apparently decomposes into a doublet (see **Fig. 1**) whose y -dependence could not be properly formalized within the percolation scheme of the random $\text{Ge}_y\text{Si}_{1-y}$ crystal ($\kappa=0$), not to mention about its κ -dependence. Calculations based on larger supercells than those used in this work, spanning various y and κ values, would be needed for a reliable discussion of this mode.

a- Composition y_1 (~ 16 at.% Ge)

If we refer to the percolation-type Raman lineshapes (see **Fig. 2a**), the discussion of the *ab initio* data (see **Fig. 3a**) should be especially simple at composition y_1 (~ 16 at.% Ge). This is because the lower components of the Ge-Si triplet and of the Si-Si doublet are expected to vanish at this composition, with the consequence that the latter multiplets should reduce to a doublet and to a singlet, respectively. Moreover, the intensity of the Si-Si singlet ($\sim 515 \text{ cm}^{-1}$), by far dominant in the Raman spectra, is expected to remain quasi insensitive to any change in the κ value. As for the Ge-Ge mode, nominally a pure singlet, this should simply reinforce with clustering and fall down with anticlustering. The latter basic trends, revealed in **Fig. 2a**, are actually observed in the *ab initio* spectra (see **Fig. 3a**).

A more sensitive issue relates to the κ -dependence of the Ge-Si doublet. If we refer to the f_i –terms in **Eq. 1**, the two observable Ge-Si bands, i.e. the central Ge-Si feature and its upper satellite, are expected to exhibit comparable Raman intensities whether the mixed crystal is random ($\kappa \sim 0$, recall RI_1) or close to the anticlustering limit ($\kappa \sim 0.2$). When κ enlarges the two sub-modes should fall down, but the central feature more rapidly than the upper satellite, eventually resulting in the dominance of the latter close to extinction of the Ge-Si signal at large clustering ($\kappa \sim 0.4$). The latter trends (see **Fig. 2a**) are fairly reproduced in the *ab initio* spectra (see **Fig. 3a**), beyond

expectations in fact, when realizing the rather large dilution of the Ge-Si bonds, leading to a moderate/small intensity of the related Raman signal, moreover an overdamped one.

b- Composition y_2 (~71 at.% Ge)

The next examined composition is the intermediary one, i.e. y_2 (~71 at.% Ge). We may readily fix the case of the unique/dominant Ge-Ge feature. As apparent in the *ab initio* data (see Fig. 3b), the Ge-Ge mode roughly doubles in strength when departing from the anticlustering regime ($\kappa \sim -0.2$) and entering the strong-clustering regime ($\kappa \sim +0.6$), consistently with the percolation-type findings. In fact, y_2 is especially interesting with respect to the Si-Ge and Si-Si multiplets. The reason is that out of the five sub-modes constituting the Ge-Si and Si-Si fine structures, only one is expected to survive at the anticlustering limit (see Fig. 2b), namely the central Ge-Si one. This provides an ideal starting point to discuss how the Ge-Si and Si-Si fine structures develop with increasing κ values.

First we consider the Si-Si doublet. According to the percolation scheme the lower Si-Si mode should emerge first and progressively reinforce with clustering while preserving a dominance over its upper counterpart, activated at a later stage, until the random substitution is achieved. At this limit the two Si-Si Raman sub-modes should ideally exhibit comparable intensities (cf. RI_2 at $\kappa=0$). By increasing further the κ value the lower mode should progressively vanish while the upper mode should keep reinforcing until remaining the only surviving Si-Si feature at large clustering ($\kappa \sim +0.6$). The trend is ideally reproduced in the *ab initio* data.

No we turn to the Si-Ge multiplet. The predicted trend within the percolation scheme is that, when increasing κ from the anticlustering limit ($\kappa \sim -0.4$), the lower Ge-Si satellite should emerge and reinforce to the detriment of the original/central Ge-Si sub-mode, being clear that in the same time the overall Ge-Si signal should progressively decline. The prevalence of the central Ge-Si feature should persist up to moderate clustering ($\kappa \sim +0.2$, see Fig. 2b), beyond which limit the balance should eventually turn to the advantage of the lower satellite, slightly before extinction of the entire Ge-Si signal at full clustering ($\kappa=1$). Note that when entering the clustering regime the upper satellite transiently emerge before the overall Ge-Si collapse ($\kappa \sim +0.8$, see Fig. 2b).

In fact, at the considered κ value close to the anticlustering limit ($\kappa \sim -0.4$) a dominant Ge-Si feature, naturally identified as the central Ge-Si sub-mode, emerges strong and sharp near $\sim 410 \text{ cm}^{-1}$ in the *ab initio* spectra. When κ increases the original/unique Ge-Si feature progressively broadens on its low-frequency side, giving rise to a distinct low-frequency asymmetry at $\kappa = 0$. With further increase in the κ value the latter asymmetry eventually evolves into a proper feature, a broad but well-defined one slightly red-shifted (by $\sim 20 \text{ cm}^{-1}$) with respect to the original/native peak. As expected the emerging feature eventually takes the advantage over the central/native one at large clustering ($\kappa \sim +0.6$).

Altogether, this conforms to the percolation-type predictions apparent in Fig. 2b.

c- Composition y_3 (~84 at.% Ge)

The κ -dependent percolation-type GeSi Raman lineshapes at y_3 are only marginally modified with respect to those observed at y_2 , owing to the close Ge content. Regarding the Ge-Ge signal, the main difference is that its intensity has become nearly κ -invariant at y_3 , while it was slightly κ -dependent at y_2 . This is clearly ascertained in the *ab initio* data. As for the Si-Si *ab initio* doublet at y_3 , this develops with κ as earlier discussed at y_2 .

In fact, y_3 is especially interesting with respect to the Ge-Si fine structure. According to the percolation scheme, the Ge-Si signal close to the anticlustering limit ($\kappa \sim -0.2$) should be bimodal at y_3 , with the central Ge-Si feature slightly dominating its lower satellite, and not unimodal, reduced to the sole central feature, as observed at y_2 . Apart from that, the κ -dependence of the y_3 -like Ge-Si doublet is expected to mimic the y_2 -like one in every respect, leading in this case to the nearly matching of intensities between the above two features in the random case (cf. RI_3 , $\kappa=0$), and to a slight inversion of dominance at large clustering essentially due to the collapse of the central Ge-Si feature.

In fact, at the anticlustering limit ($\kappa \sim -0.2$) the *ab initio* Ge-Si signal consists of a rather sharp peak with a distinct shoulder on its low frequency side (marked by a dot in Fig. 3c). Interestingly such shoulder was absent at y_2 . Altogether this is consistent with the peak and its shoulder being assigned as the central/dominant Ge-Si feature and its lower satellite, respectively. In the random crystal ($\kappa=0$), the shoulder has reinforced into a distinct feature (also marked by a dot). Ultimately, i.e. at large clustering ($\kappa \sim +0.4$), the only surviving feature seems to be the lower satellite, now emerging under the form of a broad/overdamped band located at a lower frequency than the central Ge-Si feature. Apparently the latter has totally vanished in the *ab initio* spectra.

d- Summary

In brief, our κ -dependent *ab initio* calculations of the $\text{Ge}_y\text{Si}_{1-y}$ Raman spectra done at critical compositions corresponding to nearly matching of intensities between the like Raman peaks from the Ge-Si and Si-Si multiplets, by using large (64-, 216- and 512-atom) supercells, are consistent with the predictions of the κ -dependent six-oscillator $[1 \times (\text{Ge} - \text{Ge}), 3 \times (\text{Ge} - \text{Si}), 2 \times (\text{Si} - \text{Si})]$ version of the (1D) percolation scheme for $\text{Ge}_y\text{Si}_{1-y}$, with respect to both the κ -induced inter- and intra-bond (Ge-Si, Si-Si) transfers of oscillator strength.

3. Comparison with experiment

In our original contribution at y_0 ,¹ care was taken to compare the κ -dependent Raman spectra calculated *ab initio* and within the percolation scheme with available experimental data from the literature. In the present study developed outside y_0 , the available experimental data are very sparse. We are aware of only one contribution, due to Reparaz *et al.*,²⁴ related to a careful Raman study of the effect of annealing on $\text{Ge}_y\text{Si}_{1-y}$ epilayers with small-moderate y values in the range 0.20 – 0.47 grown on Si substrates by molecular beam epitaxy (MBE). Only the main Ge-Ge, Ge-Si and Si-Si Raman features were discussed, with no mention of any fine structures behind such signals. What emerged is that under cumulative annealing treatments the main Ge-Ge, Ge-Si and Si-Si Raman peaks were all progressively shifted towards low frequency, at any y value (detail is given below). Remarkably, no change was detected in the relative Ge-Ge, Ge-Si and Si-Si Raman intensities (see Fig. 1b of Ref. 24).

Several mechanisms have been considered by Reparaz *et al.* to explain the Raman shifts with annealing. A possible annealing-induced Si-interdiffusion from the substrate towards the epilayers, leading to a lowering of the y values after annealing, has been ruled out, based on observation of similar Raman shifts with thin and thick $\text{Ge}_y\text{Si}_{1-y}$ epilayers. The gradual relaxation under annealing of the substrate-induced strain inside the epilayers has likewise been discarded since almost full strain relaxation was already achieved after the first annealing step, as testified by x-ray diffraction

measurements of the in-plane and out-of-plane lattice constants of the $\text{Ge}_y\text{Si}_{1-y}$ epilayers. As a last resort, and based on the work of Tersoff,⁴³ Reparaz *et al.* advanced an idea that their as-grown $\text{Ge}_y\text{Si}_{1-y}$ epilayers did, in fact, exhibit a significant amount of local clustering ($\kappa > 0$) of the Ge atoms (and thereby also of the Si ones) on account that the MBE is a non-equilibrium growth technique. The Raman shifts were then attributed to a redistribution of the Ge and Si atoms at a fixed composition y stimulated by annealing, until an ideally random Ge-to-Si substitution ($\kappa = 0$) was eventually achieved at the term of the annealing cycles.

It follows that the planned *ab initio* vs. experiment comparison in this work has to be done for positive κ values (clustering-to-random regime) at small-moderate Ge content ($y < y_0$), and only in relation to the frequencies of the main Ge-Ge, Ge-Si and Si-Si Raman peaks, not to their intensities. We have already emphasized that the percolation scheme is not equipped to address the κ -dependence of the Raman frequencies (see **Sec. I**). This suffices to rule out the percolation scheme from the planned comparison with experiment. In contrast the *ab initio* calculations provide clear trends regarding the κ -dependence of both the Raman intensities and the Raman frequencies. An *ab initio* vs. experiment comparison is thus relevant, in relation to the latter aspect at least.

When κ decreases (from positive to negative values) the *ab initio* trend is that the main Raman features due to the homo (Ge-Ge and Si-Si bonds) and hetero (Ge-Si) bonds shift in opposite directions, i.e. towards low and high frequency, respectively, at any composition y . The trend is best appreciated at composition y_0 owing to strong emergence of the main Ge-Ge, Ge-Si and Si-Si Raman features at all considered κ values (see **Appendix Sec.**). y_0 is further attractive in that the *ab initio* calculations were repeated several times at each κ value (considering the duplicated 216-atom supercells plus the unique 512-atom one), offering the possibility to estimate the natural fluctuation in the Raman frequency of a given peak when the atom arrangement changes at fixed (y, κ) values. In fact the fluctuation in the Raman frequency hardly reaches 2.5 cm^{-1} for the main Ge-Ge and Si-Si peaks ($\kappa = +0.31$), and about 2.0 cm^{-1} for the main Ge-Si one ($\kappa = -0.31$), if we refer to the κ values corresponding to the strongest emergence of these peaks (specified in brackets) – emphasized by pairs of opposite/horizontal arrows in **Fig. A1 (Appendix Section)**. Such fluctuations are small in comparison with the observed κ -induced shifts of the Ge-Ge, Ge-Si and Si-Si *ab initio* Raman peaks, corresponding to -15 , $+13$ and -10 cm^{-1} at composition y_0 (in the sense of decreasing κ values), respectively. The observed *ab initio* trend of opposite κ -induced shifts for the Raman peaks due to the homo and hetero bonds thus seems reliable. Now, the latter trend opposes to the experimental findings of Reparaz *et al.* which indicate a shift in the same direction of all Raman features.²⁴

If we omit the problem of the directions of the shifts, an additional difficulty arises in view of the magnitudes of the experimental Ge-Ge, Ge-Si and Si-Si shifts reported by Reparaz *et al.* (in reference to Fig. 3 of Ref. **24**). The latter amount to ~ 3.0 , ~ 3.5 and $\sim 4.5 \text{ cm}^{-1}$ at y_0 , respectively, corresponding to roughly half the current *ab initio* shifts detected at y_0 when κ reduces from $+0.31$ to 0 (of ~ 6.0 , ~ 7.0 and $\sim 4.0 \text{ cm}^{-1}$, respectively). We have difficulties to conceive that such significant experimental shifts may occur without being accompanied by any noticeable change in the Ge-Ge, Ge-Si and Si-Si Raman intensities, since the change is so dramatic in our *ab initio* data.

Summarizing, the experiment vs. *ab initio* comparison is not really conclusive. Apparent discrepancies as to whether the Raman peaks due to the homo and hetero bonds shift in like (experiment) or opposite (*ab initio*) directions when κ varies, and, also, as to whether the κ -induced change in the Raman frequencies should be accompanied (*ab initio*) or not (experiment) by a noticeable change in the Raman intensities, might be due, in fact, to a smaller range of κ values spanned in the experiment, i.e. much smaller than that considered in the current *ab initio* data.

Additional *ab initio* calculations using large supercells covering a narrow domain of small positive κ values would be needed to decide whether the Raman shifts anticipate the change in the Raman intensities when κ varies (as suggested by the experimental data of Reparaz *et al.*) or whether both effects come together (as we believe). Such meticulous *ab initio* calculations fall beyond the scope of this work.

IV. CONCLUSION

The predictions of the (1D) percolation scheme concerning the $\text{Ge}_y\text{Si}_{1-y}$ Raman intensities in their dependence on a relevant order parameter κ of local clustering/anticlustering earlier introduced in Ref. 1 are tested with *ab initio* calculations done by applying the AIMPRO code to large (64-, 216- and 512-atom) fully relaxed (atom positions and lattice constant) disordered (3D) supercells. The comparison is realized at three critical moderate/dilute Ge and Si contents corresponding to nearly matching of Raman intensities within the Si-Si doublet ($y_2 \sim 0.71$), and between the central mode of the Si-Ge triplet and either its lower ($y_3 \sim 0.84$) or its upper ($y_1 \sim 0.16$) satellite. Such compositions are well suited to explore in particular the κ -induced intra-bond transfer of oscillator strength within a given multiplet (the Ge-Si or Si-Si ones). As such, the actual 'percolation vs. *ab initio*' comparative insight completes an earlier one achieved by using small (32-atom) disordered supercells at intermediary Ge/Si content ($y_0 \sim 0.50$),¹ corresponding to comparable intensities of the main Ge-Ge, Ge-Si and Si-Si Raman features, thus best suited to investigate the κ -induced inter-bond transfer of oscillator strength.

The 'percolation vs. *ab initio*' test appears to be conclusive in the positive sense. With this, the κ -equipped version of the $\text{Ge}_y\text{Si}_{1-y}$ percolation scheme emerges as a relevant and easy-to-handle tool likely to provide a refined estimate, i.e. beyond the MREI one, of the nature of the atom substitution in the $\text{Ge}_y\text{Si}_{1-y}$ mixed crystals, i.e. as to whether the latter is random or due to local clustering/anticlustering, from a direct insight into the Raman intensities.

SUPPLEMENTARY MATERIAL

See [supplementary material](#) for a detailed insight into the pair correlation functions up to seventh-neighbors of the used 216- and 512-atom random ($\kappa \sim 0$) supercells in this work.

ACKNOWLEDGMENTS

The authors would like to thank A. V. Postnikov for useful discussions and careful reading of the manuscript. This work is funded by FEDER funds through the COMPETE 2020 Programme and National Funds through FCT - Portuguese Foundation for Science and Technology under the project UID/CTM/50025/2013, and by the European funding of Region Lorraine under Project FEDER-Percolloy n° presage 34619.

Appendix Section

In this **Sec.** we provide an *ab initio* insight into the κ -dependence of the $\text{Ge}_{0.5}\text{Si}_{0.5}$ Raman pattern by using large (216- and 512-atom) supercells. This re-actualizes an earlier insight achieved in the same spirit in Ref. **1** by using small (32-atom) $\text{Ge}_{0.5}\text{Si}_{0.5}$ supercells.

The corresponding MREI- (thin curves) and percolation-type (thick curves) $\text{Ge}_{0.5}\text{Si}_{0.5}$ Raman lineshapes at three representative κ values (-0.31 , 0.00 , $+0.31$), earlier calculated in Ref. **1** (see Fig. 3 therein), are reproduced in **Fig. A1**, for reference purpose (the used Raman frequencies and damping parameters are recalled in **Sec. II-1**). The nearly matching of intensities between the main Ge-Ge, Ge-Si and Si-Si Raman features, in reference to RI_0 , is emphasized (dotted box) in case of a random substitution ($\kappa=0$). With clustering the signals due to the homo-like (Ge-Ge, Si-Si) bonds reinforce to the detriment of that stemming from the hetero-like (Ge-Si) one, and vice versa with anticlustering. If we zoom into the percolation-type fine structures of the Ge-Si and Si-Si signals, a similar trend occurs, but in relation to the homo- and hetero-like host environments of a given bond then. This generates opposite trends in the Ge-Si and Si-Si spectral ranges regarding the κ -dependence of the dominant feature at $\kappa=0$.

The κ -dependent ($\kappa=-0.31$, 0.00 , $+0.31$) *ab initio* Raman spectra calculated with the 216- (thin curves) and 512-atom (thick curves) supercells are displayed in **Figs. A2**. An additional *ab initio* Raman spectrum calculated by using a 64-atom SQS ($\kappa=0$, see main text **Sec. II-2**) is added (dotted curve) for comparison. The crucial issue is whether the percolation-type trends concerning the κ -dependencies of the Ge-Si (three-mode pattern) and Si-Si (two-mode pattern) fine structures, summarized above, are consistent with the reported *ab initio* calculations, or not.

The upper Ge-Si satellite, an intrinsically complicated feature that decomposes into a pseudodoublet from moderate Ge content onwards (greater than ~ 30 at.% Ge, see Fig. 2a of Ref. **8**) – as emphasized by an oval in **Fig. 1**, is not stable when changing the atom arrangement at a given κ value, whichever supercell size is used within our presently accessible limit of 512 atoms. Therefore, as far as the Ge-Si signal is concerned, only the central feature and its lower satellite are discussed below.

At $\kappa=0$ the overall Ge-Si Raman signal generated *ab initio* by using the large (216- and 512-atom) supercells exhibits a small asymmetry on its low frequency tail. The *ab initio* data further reveal a collapse of the central Ge-Si feature accompanied by a reinforcement of the low frequency asymmetry at large clustering ($\kappa=+0.31$) and vice versa at large anticlustering ($\kappa=-0.31$). Altogether such *ab initio* trends are consistent with the predictions of the percolation model.

Now, we turn to the Si-Si doublet. The reference percolation-type situation at $\kappa=0$ consists of the lower Si-Si mode being marginally activated on the tail of its upper counterpart showing up as the dominant feature. Such minor/dominant balance between the lower/upper Si-Si modes is apparent in the *ab initio* data. At large clustering ($\kappa=+0.31$), the percolation model predicts an emphasis of the strength balance due to antagonist variations of the two Si-Si modes. The *ab initio* data are consistent with the latter prediction. The anticlustering case ($\kappa=-0.31$) is somehow more critical. Based on the percolation model, one expects a slight inversion of the strength balance between the two Si-Si modes with respect to $\kappa=0$, the Raman intensities remaining comparable though. While the expected (slight) inversion is not evidenced in the reported *ab initio* data, the strong asymmetry of the overall Si-Si *ab initio* signal is at least consistent with the Raman intensities of its two components being comparable.

Summarizing, the present *ab initio* calculations based on large 216- and 512-atom $\text{Ge}_{0.5}\text{Si}_{0.5}$ supercells provide a more refined insight into the κ -dependence of the $\text{Ge}_{0.5}\text{Si}_{0.5}$ Raman spectra than earlier achieved in Ref. 1 by using small 32-atom supercells. Altogether the observed *ab initio* trends appear to be consistent with the predictions of the percolation model, not only with respect to the κ -driven intra-bond transfer of oscillator strength (within the Ge-Si and Si-Si multiplets) but also with respect to the inter-bond transfer of oscillator strength (between the Ge-Ge, Ge-Si and Si-Si modes). Additional *ab initio* calculations based on larger supercells are nevertheless needed to decide about the κ -dependence of the upper Ge-Si satellite.

Figure captions

Fig. 1: Composition dependence of the Raman frequencies (symbols) for the six-oscillator $[1x(\text{Ge-Ge}), 3x(\text{Ge-Si}), 2x(\text{Si-Si})]$ $\text{Ge}_y\text{Si}_{1-y}$ random mixed crystal ($\kappa=0$), based on the available experimental and theoretical data in the literature (thin symbols) - including our own *ab initio* data (thick symbols), as regrouped in Ref. 8 (see Fig. 2 therein). The related oscillator fractions (f_i -terms), also specified, determine which Raman modes are dominant (thick line), minor (dashed line) or comparable within the Si-Si and Ge-Si multiplets (*RI* labels) at a given composition y . Curved (in-frame) and straight (out-of-frame) arrows schematically indicate the intra- and inter-bond transfers of oscillator strength driven by clustering/anticlustering, respectively, with concomitant impact on the Raman intensities.

Fig. 2: Predicted κ -dependence of the $\text{Ge}_y\text{Si}_{1-y}$ Raman intensities at $y_1=0.16$ (a), $y_2=0.71$ (b) and $y_3=0.84$ (c) corresponding to remarkable intensity interplays (*RI*'s, emphasized by dotted boxes) within the six-oscillator $[1x(\text{Ge-Ge}), 3x(\text{Ge-Si}), 2x(\text{Si-Si})]$ version of the $\text{Ge}_y\text{Si}_{1-y}$ percolation scheme (thick lines) and within the more crude three-oscillator $[1x(\text{Ge-Ge}), 1x(\text{Ge-Si}), 1x(\text{Si-Si})]$ MREI scheme (thin lines). The Raman intensities compare directly in the different insets provided the main and peak-related normalizing factors (specified in brackets) are taken into account.

Fig. 3: *Ab initio* κ -dependent $\text{Ge}_y\text{Si}_{1-y}$ Raman spectra calculated at $y_1 \sim 0.16$ (a), $y_2 \sim 0.71$ (b) and $y_3 \sim 0.84$ (c) by using 216- and 512-atom (thin/thick curves) disordered cubic supercells (SC's) designed via a simulated annealing method (see text). The Raman intensities compare directly in the different panels provided the main and peak-related normalizing factors (specified in brackets) are taken into account. In panel (c) dots are used to emphasize a shoulder on the low frequency tail of the central peak. Large simple/double arrows indicate significant/dramatic variations in the Raman intensities with respect to $\kappa=0$.

Fig. A1: Predicted κ -dependence of the $\text{Ge}_{0.5}\text{Si}_{0.5}$ Raman intensities for three representative κ values calculated within the six-oscillator $[1x(\text{Ge-Ge}), 3x(\text{Ge-Si}), 2x(\text{Si-Si})]$ version of the $\text{Ge}_y\text{Si}_{1-y}$ percolation scheme (thick lines) and within the more crude three-oscillator $[1x(\text{Ge-Ge}), 1x(\text{Ge-Si}), 1x(\text{Si-Si})]$ MREI scheme (thin lines). The Raman intensities compare directly to those reported in Fig. 2.

Fig. A2: *Ab initio* κ -dependent $\text{Ge}_{0.5}\text{Si}_{0.5}$ Raman spectra calculated at three representative κ values by using 216- (thin curves) and 512-atom (thick curves) disordered cubic supercells (SC's) designed via a simulated annealing method. A Raman spectrum obtained at $\kappa=0$ by using a 64-atom SQS (dotted curve) is added, for comparison. The Raman intensities compare directly with those reported in Fig. 3. The stars mark spurious features. Large simple/double vertical arrows indicate significant/dramatic variations in the Raman intensities with respect to $\kappa=0$. Small double/opposite horizontal arrows at the top/bottom refer to fluctuations in the frequencies of the main Raman features due to the homo/hetero bonds.

- 752 ¹ O. Pagès, R. Hajj Hussein and V. J. B. Torres, J. Appl. Phys. **114**, 033513 (2013).
- 753 ² A. Mascarenhas, H.M. Cheong, M.J. Seong and F. Alsina, in '*Spontaneous Ordering in Semiconductor*
- 754 *Alloys*', edited by A. Mascarenhas (Kluwer Academic/Plenum Publishers, New York, 2002).
- 755 ³ S. Adachi, *Properties of Semiconductor Alloys, Group-IV, III-V and II-VI Semiconductors*, (John Wiley
- 756 & Sons Ltd, Chichester, U.K., 2009), Chapter 4.
- 757 ⁴ M. Born and K. Huang, in *Dynamical Theory of Crystal Lattices*, (Clarendon Press, Oxford, 1954), p.
- 758 105.
- 759 ⁶ I. F. Chang and S. S. Mitra, Adv. Phys. **20**, 359 (1971).
- 760 ⁷ H. W. Verleur and A. S. Barker, Jr., Phys. Rev. **149**, 715 (1966).
- 761 ⁸ O. Pagès, J. Souhabi, V.J.B. Torres, A.V. Postnikov and K.C. Rustagi, Phys. Rev. B **86**, 045201 (2012).
- 762 ⁹ R. Jones and P. R. Briddon, chapter 6 in '*Identification of Defects in Semiconductors*', volume 51A of
- 763 Semiconductors and Semimetals (Academic Press, Boston) 1998.
- 764 ¹⁰ S. Kanakaraju, A.K. Sood and S. Mohan, Phys. Rev. B **61**, 8334 (2000).
- 765 ¹¹ N. Nakano, L. Marville, S.-M. Jang, K. Liao, C. Tsai, J. Tsai, H.-W. Kim and R. Reif, J. Appl. Phys. **73**,
- 766 414 (1993).
- 767 ¹² J.S. Reparaz, A. Bernardi, A.R. Gõni, M.I. Alonso and M. Garriga, Appl. Phys. Lett. **92**, 081909 (2008).
- 768 ¹³ J.C. Tsang, V.P. Kesan, J.L. Freeouf, F.K. LeGoues and S.S. Iyer, Phys. Rev. B **46**, 6907 (1992).
- 769 ¹⁴ F. Cerdeira, M.I. Alonso, D. Niles, M. Garriga, M. Cardona, E. Kasper and H. Kibbel, Phys. Rev. B **40**,
- 770 1361 (1989).
- 771 ¹⁵ C. Sheng, T.-C. Zhou, Q. Cai, D. Gong, M.-R. Yu, X.-J. Zhang and X. Wang, Phys. Rev. B **53**, 10771
- 772 (1996).
- 773 ¹⁶ R. Schorer, E. Friess, K. Eberl and G. Abstreiter, Phys. Rev. B **44**, 1772 (1991).
- 774 ¹⁷ R. Schorer, G. Abstreiter, H. Kibbel and H. Presting, Phys. Rev. B **50**, 18211 (1994).
- 775 ¹⁸ M.I. Alonso, F. Cerdeira, D. Niles, M. Cardona, E. Kasper and H. Kibbel, J. Appl. Phys. **66**, 5645
- 776 (1989).
- 777 ¹⁹ S. de Gironcoli, E. Molinari, R. Schorer and G. Abstreiter, Phys. Rev. B **48**, 8959 (1993).
- 778 ²⁰ R. Schorer, G. Abstreiter, S. de Gironcoli, E. Molinari, H. Kibbel and H. Presting, Phys. Rev. B **49**,
- 779 5406 (1994).
- 780 ²¹ M.I. Alonso and K. Winer, Phys. Rev. B **39**, 10056 (1989).
- 781 ²² J. Zi, K. Zhang and X. Xie, Phys. Rev. B **47**, 9937 (1993).
- 782 ²³ J. Zi, W. Ludwig, K. Zhang and X. Xie, Phys. Rev. B **51**, 7886 (1995).
- 783 ²⁴ J.S. Reparaz, I.C. Marcus, A.R. Gõni, M. Garriga and M.I. Alonso, J. Appl. Phys. **112**, 023512 (2012).
- 784 ²⁵ M. Franz, K. F. Dombrowski, H. Rücker, B. Dietrich, K. Pressel, A. Barz, U. Kerat, P. Dold and K. W.
- 785 Benz, Phys. Rev. B **59**, 10614 (1999).
- 786 ²⁶ F. Pezzoli, L. Martinelli, E. Grilli, M. Guzzi, S. Sanguinetti, M. Bollani, H. D. Chrastina, G. Isella, H. von
- 787 Känel, E. Wintersberger, J. Stangl and G. Bauer, Mat. Sci. Eng. B **124-125**, 127 (2005).
- 788 ²⁷ S. Rath, M. L. Hsieh, P. Etchegoin and R. A. Stradling, Semicond. Sci. Technol. **18**, 566 (2003).
- 789 ²⁸ W. J. Brya, Solid State Commun. **12**, 253 (1973).
- 790 ²⁹ C. H. Grein and M. Cardona, Phys. Rev. B **45**, 8328 (1992).
- 791 ³⁰ A. A. Maradudin, Solid State Phys. **18**, 273 (1966).
- 792 ³¹ S. de Gironcoli, Phys. Rev. B **46**, 2412 (1992).
- 793 ³² P. G. Dawber and R. J. Elliott, Proc. Roy. SOC. (London) **A273**, 222 (1963).
- 794 ³³ M. J. Rayson and P. R. Briddon, Comp. Phys. Commun. **178**, 128 (2008).

795 ³⁴ J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).
796 ³⁵ C. Hartwigsen, S. Goedecker and J. Hutter, Phys. Rev. B **58**, 3641 (1998).
797 ³⁶ V. J. B. Torres, J. Coutinho, P. R. Briddon and M. Barroso, Thin Solid Films **517**, 395 (2008).
798 ³⁷ The Brillouin zone of the 64-, 216- and 512-atom cubic supercells was sampled by 4×4×4, 2×2×2
799 and 1×1×1 special k-point mesh, respectively, as proposed by H. J. Monkhorst and J. D. Pack, Phys.
800 Rev. B **13**, 5188 (1976).
801 ³⁸ S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, Science **220**, 671 (1983).
802 ³⁹ A. Zunger, S.-H. Wei, L. G. Ferreira and J. E. Bernard, Phys. Rev. Lett. **65**, 353 (1990).
803 ⁴⁰ P. Venezuela, G. M. Dalpian, A. J. R. da Silva and A. Fazzio, Phys. Rev. B **64**, 193202 (2001).
804 ⁴¹ A. Chroneos, H. Bracht, C. Jiang, B. P. Uberuaga and R. W. Grimes, Phys. Rev. B **78**, 195201 (2008).
805 ⁴² A. Chroneos, E. N. Sgourou and C. A. Londos, J. Mater. Sci.: Mater Electron **24**, 2772 (2013).
806 ⁴³ J. Tersoff, Phys. Rev. Lett. **77**, 2017 (1996).