Supplementary Material – Appendix A

for

Infrared spectroscopy of natural *Type Ib* diamond: insights into the formation of *Y*-centers and the early aggregation of nitrogen

Maxwell C. Day[1*], Mike Jollands[2], Francesca Innocenzi[1], Davide Novella[1], Fabrizio Nestola[1], and Martha G. Pamato[1]

[1] Dipartimento di Geoscienze, Università di Padova, 35131 Padova, Italy

[2] Gemological Institute of America (GIA), 50W, 47th Street, New York, NY, 10036, U.S.A.

Validation of the ***Y***-center absorption coefficient and statistical analysis of FTIR data

Sections A.1 and A.2 (Figures A.1 – A.7)

A.1. Validation of the *Y*-center absorption coefficient

In the new deconvolution procedure for *Type Ib + IaA* diamonds, the ***Y***-center absorption coefficient ($\epsilon_{\boldsymbol{Y}}$) is set to that of the ***C***-center, 25 at.ppm/cm^{-1}. As it is clear that the ***Y***-center contains N, it is logical to assume that such N is in a single-substitutional form as ***Y***-centers are only observed in *Type Ib + IaA* diamonds and occur in highest concentration in *Type Ib > IaA* specimens. Previous studies dealing with ***Y***-centers broadly associate them with ***C***-centers. For example, based on EPR data, Titkov et al. (2015) suggest that an electronic interaction between N_S^0 (***C***-centers) and N_S^+ (***X***-centers) may form an extended defect that is the ***Y***-center. Other papers rely largely on the findings of Hainschwang et al. (2012) where the positive correlation between ***X***- and ***Y***-centers is used in support of the ***Y***-center being related to some form of single-substitutional N. Hainschwang et al. (2012) also describe the similar absorption "pattern" of ***C***- and ***Y***-centers. As the position of peaks in the IR is a function of the bond strength (i.e., N-C bond energy (Kj/mol)), which is directly related to bond-length (i.e., mean N-C bond-length (Å)), and the mass of each atom participating on the vibrational mode (i.e., reduced mass and N and C), the similar absorption maxima of the ***C***-center (1130 cm^{-1}) and ***Y***-center (1145 cm^{-1}) spectra suggest the dominant N-C vibrational mode of ***C***-centers may be similar to that of ***Y***-centers.

Now, the validity of this assumption can be assessed by deriving a realistic range (error) for $\epsilon_{\boldsymbol{Y}}$. In some previous studies, there is reasonable agreement between N_{tot} determined by SIMS and deconvolution of the IR N-region (without incorporation of ***Y***-centers). Reutsky et al. (2017) report an agreement of 10-15% and Zedgenizov et al. (2017) report an average agreement of ~15% between N_{tot} determined with SIMS and IR. This suggests that $\epsilon_{\boldsymbol{Y}}$ must be at least similar to $\epsilon_{\boldsymbol{C}}$ as during traditional deconvolution (without incorporation of ***Y***-centers), absorption due to ***Y***-centers is erroneously fit with the ***C***-center spectrum. In other words, setting $\epsilon_{\boldsymbol{Y}}$ = 25 at.ppm/cm^{-1} produces

the least possible discrepancy between N_{tot} and $N_{tot}(Y)$. Therefore, if ϵ_Y is significantly larger or smaller than ϵ_C, large discrepancies between N contents determined with FTIR (N_{tot} or $N_{tot}(Y)$) and SIMS (N_{bulk}) would be expected which is not the case. Moreover, larger discrepancies between N_{tot} and $N_{tot}(Y)$ would be expected, on average there is a ~20% discrepancy between N_{tot} and $N_{tot}(Y)$ for the samples studied here (see Table C.2). Consider the hypothetical question; for a given diamond, what is the degree to which ϵ_Y can vary from 25 at.ppm/cm^{-1} such that reasonable errors would be observed between N_{bulk} determined by SIMS, and $N_{tot}(Y)$ determined by IR? Assuming a reasonable error for $N_{tot}(Y)$ of ±15% (as shown by Reutsky et al. 2017; Zedgenizov et al. 2017), the range of ϵ_Y required to produce such errors was calculated for all 178 samples studied here. As shown in Fig. A.1, this calculation shows that a maximum and minimum ϵ_Y of 41 and 11 at.ppm/cm^{-1} is reasonable assuming an agreement between $N_{tot}(Y)$ and N_{bulk} of ±15%.

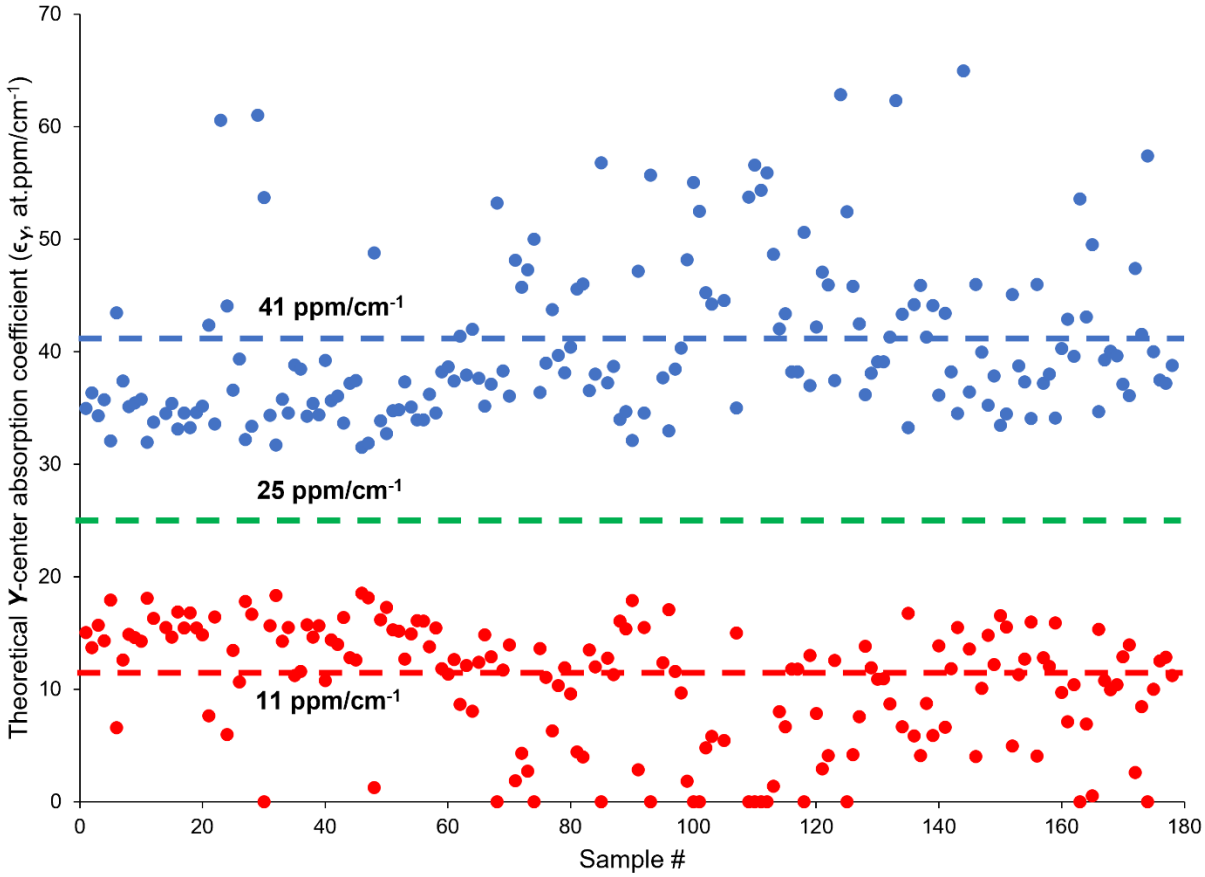

Figure A.1. Theoretical ***Y***-center absorption coefficient (ϵ_Y) calculated for all samples by assuming $N_{tot}(Y)$ that are 15% higher (blue circles) and 15% lower (red circles) than $N_{tot}(Y)$ calculated assuming ϵ_Y = 25 at.ppm/cm^{-1} (green dashed line). The blue and red dashed lines indicate the average respective ϵ_Y values suggesting the approximate maximum and minimum ϵ_Y is 41 and 11 at.ppm/cm^{-1} assuming a ±15% agreement between N_{bulk} and $N_{tot}(Y)$.

A different approach can be taken to estimate the minimum ϵ_Y by considering one of the *Type Ib* diamonds from the Zimmi alluvial locality (West Africa) studied by Smit et al. (2016, 2018). For these diamonds, Re/Os dating of sulphide inclusions gives a diamond formation age of

650 Mya and associated kimberlites in the Man Shield are 150 Mya (Skinner et al., 2004) thus defining a residence time of ~500 Myr. Using this residence time, Smit et al. (2016) determined residence temperatures of 850-900°C in accord with temperatures in the diamond stability field along the local geotherm for the Man Shield. As both residence time and temperature are approximately known, re-analysis of one of the Zimmi diamonds and deconvolution (while incorporating ***Y***-centers) allowed the minimum ϵ_{Y} to be constrained. This was done by iteratively deconvoluting the IR spectra of one Zimmi diamond where ϵ_{Y} was set to 0-100 at.ppm/cm^{-1}. In Fig. A.2., 500 Myr isochrons are plotted for temperatures ranging from 850 to 940°C and it is shown that assuming ϵ_{Y} = 25at.ppm/cm^{-1} (green circle in Fig. A.2.) requires a residence temperature (893°C, green line in Fig. A.2.) within the appropriate range (850-900°C) (Smit et al., 2016). Assuming an error for N_{tot}(Y) of ±15% (as described above), a maximum and minimum ϵ_{Y} (32 and 18 at.ppm/cm^{-1}) is calculated which requires temperatures of 887 and 901°C, respectively (orange dashed lines in Fig. A.2.). This minimum ϵ_{Y} (18 at.ppm/cm^{-1}) accords with the minimum ϵ_{Y} (19 at.ppm/cm^{-1}) based on the maximum residence temperature (900°C) assuming a 500 Myr residence time. Using ϵ_{Y} values below 19 at.ppm/cm^{-1} would require residence temperatures higher that 900°C and larger discrepancies between N_{tot} and N_{tot}(Y), e.g., ~55% if ϵ_{Y} is set at 1 at.ppm/cm^{-1}. Although where ϵ_{Y} = 100 at.ppm/cm^{-1}, realistic residence temperatures (850°C) are required, this would result in unrealistically high N_{tot}(Y) values and discrepancies between N_{tot} and N_{tot}(Y) of ~170%. Based on the above discussion, it is reasonable to assume that the true ϵ_{Y} is reasonably close to 25 at.ppm/cm^{-1} and that using this ϵ_{Y} value during deconvolution will produce more accurate N-defect contents compared to where ***Y***-center absorbance is ignored during deconvolution. It is likely that most of the N_{tot} and %IaA values reported in the literature for *Type*

Ib + *IaA* diamonds are incorrect as ***Y***-centers are neglected during fitting in all the commonly used deconvolution software.

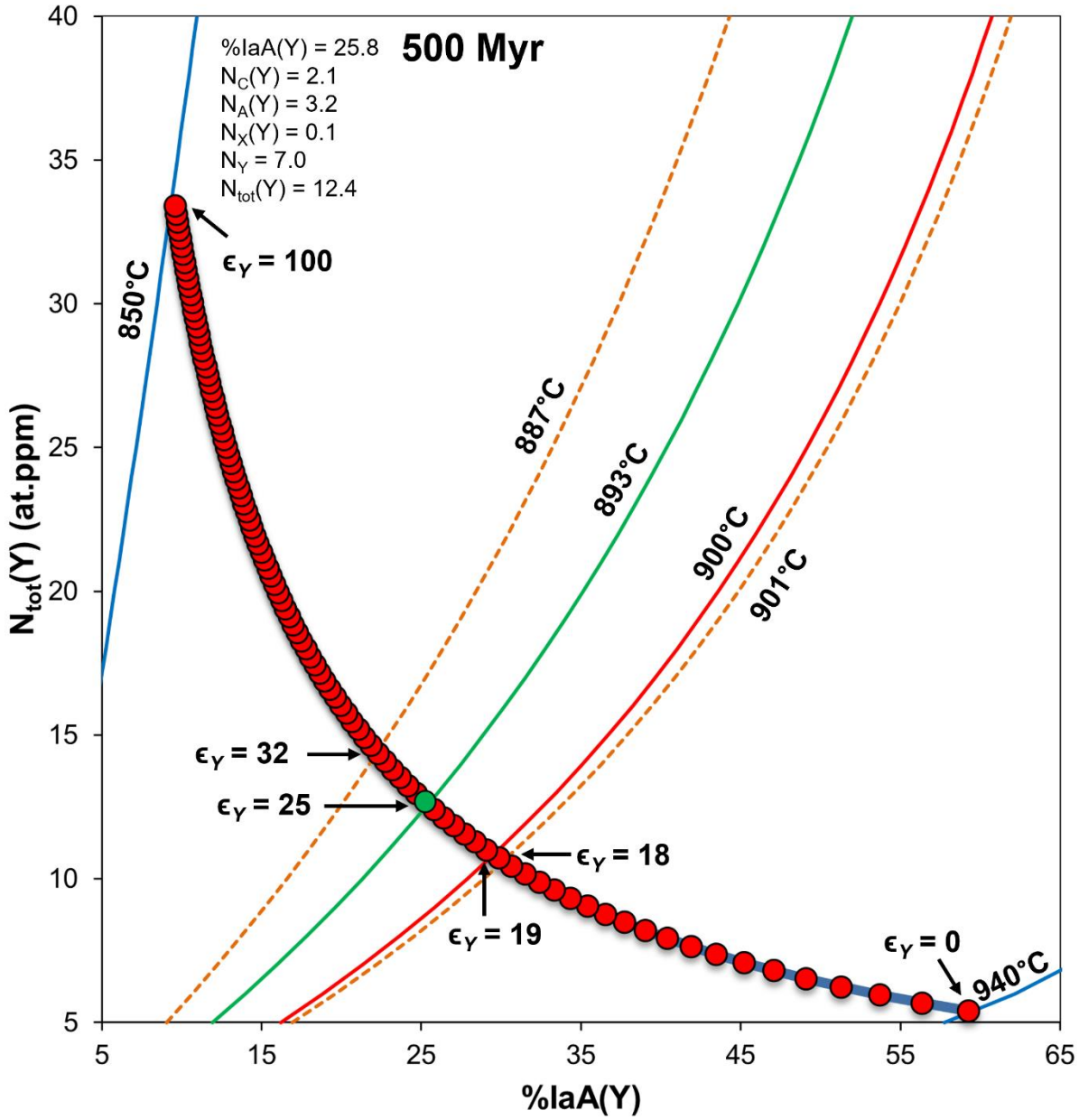

Figure A.2. Plot of $N_{tot}(Y)$ and %IaA(Y) calculated by iterative deconvolution of the IR spectrum recorded from a Zimmi diamond where ϵ_Y was set as 0 to 100 at.ppm/cm^{-1}. 500 Myr isochrons are plotted for temperatures between 850 and 940°C. Assuming a 500 Myr residence time, Smit et al.

(2016) determined these diamonds would require residence temperatures of ~850-900°C assuming cuboid growth morphologies. To allow comparison with data from Smit et al. (2016), isochrons were calculated using the same C→A activation energy, 6.0 eV from Taylor et al. (1996). The average deconvolution results (obtained from six spectra) assuming $\epsilon_{\boldsymbol{Y}}$ = 25 at.ppm/cm^{-1} (green circle) are given in the top left corner and correspond to a residence temperature of 893°C (green line). Assuming an error for N_{tot}(Y) of ±15%, a maximum and minimum $\epsilon_{\boldsymbol{Y}}$ (of 32 and 18 at.ppm/cm^{-1}) is calculated which requires temperatures of 887 and 901°C, respectively (orange dashed lines). The minimum $\epsilon_{\boldsymbol{Y}}$ (18 at.ppm/cm^{-1}) accords with the minimum $\epsilon_{\boldsymbol{Y}}$ (19 at.ppm/cm^{-1}) based on the maximum residence temperature (900°C, red line) for the Zimmi diamonds assuming a 500 Myr residence time. The minimum residence temperature (850°C, blue line) requires an unrealistically high $\epsilon_{\boldsymbol{Y}}$ (100at.ppm/cm^{-1}) and when $\epsilon_{\boldsymbol{Y}}$ = 0 at.ppm/cm^{-1}, a unrealistically high residence temperature is required (940°C, blue line).

As we use the absorption coefficient for ***C***-centers to estimate N_Y, the fitting routine essentially acts to split the ***C***-center absorption signal across a wider frequency range. Consequently, one may consider the new fitting routine as one with a revised ***C***-center spectrum used for fitting *Type Ib* + *IaA* spectra. This is useful if one does not wish to include ***Y***-centers in their deconvolution routine but is interested in improving the fit of the N-region from ~1130-1150 cm^{-1} and thus the accuracy of calculated N-defect contents derived from such a fit.

A.2. Statistical analysis of FTIR data

Deconvolution of the IR spectra of each sample is done by least-squares fitting of the signal in the N-region using the spectra of the ***C***-, ***A***-, ***X***-, and ***Y***-center absorption systems. Fitting results are evaluated by calculating a modified *cosine similarity* (S_C) between the raw spectrum (after normalization and baseline correction) and the fit of the N-region as follows:

$$S_C = \left[1 - \left(\frac{\sum_{i=1}^{n} S_i F_i}{\sqrt{\sum_{i=1}^{n} S_i^2} \cdot \sqrt{\sum_{i=1}^{n} F_i^2}}\right)\right] \times 100$$

Here, S_i and F_i represent the intensity at each wavenumber (from 1000 to 1400 cm^{-1}) of the spectrum and fit, respectively. Calculation of S_C allows assessment of the fit quality independent of N content (unlike the sum of squared residuals, *SSR*), as S_C depends only on the shape of the signal whereas *SSR* depends on both the shape and magnitude of the signal being fit. S_C values close to 0 indicate a high-quality fit and as S_C increases, the fit-quality decreases to a maximum value $S_C = 200$. Here, fits where $S_C > 0.7$ are considered poor-quality. In the *Caxbd_Inherit_2024-Ib* spreadsheet (see Supplementary Material) the modified cosine similarity and squared sum of residuals is automatically calculated. For all comparisons (scatter plots), whether or not a correlation is statistically significant is determined by calculating the Pearson correlation coefficient (*r*) and conducting standard two-tailed *t*-tests at $\alpha = 0.05$ (*i.e.*, 97.5% confidence). Where significant correlation has been found, the R^2 value is also reported.

Interpretation of bulk (averaged) spectra of zoned diamonds

Using the FTIR DRIFT accessory, *bulk* spectra were recorded representing averaged data across the IR beam path. *Type Ib + IaA* diamonds often show complex growth zonation (e.g.,

Zedgenizov et al. 2016; Mashkovtsev et al. 2021) associated with variations in N-defect contents and %IaA and thus our results represent an averaged picture of early N-defect aggregation. The effect of averaged (bulk) N-defect contents recorded from zoned diamonds on calculated mantle residence times must be assessed as such values are used to demonstrate the importance of incorporating ***Y***-centers during decomposition. In general, diamonds may exhibit *primary* and/or *secondary* growth zonation. Primary zones grow during the same diamond-growth event where fluctuations in the contents and types of impurities (e.g., N and H defects) reflect evolution of a single diamond-forming fluid and/or different pulses of fluid from the same source (as indicated by similar C and N isotopic signatures). Secondary growth zones, often observed as an older core and younger rim, reflect two different diamond growth events that may be associated with different diamond-forming fluids and fluid sources. All primary growth zones in a single *Type Ib + IaA* diamond are subject to the same mantle residence history (i.e., time and temperature of mantle residence). It follows that any bulk spectra collected from such a diamond will yield the same mantle residence temperature (for a given known/assumed residence time) as a spectrum collected from any one of the individual primary growth zones.

Secondary growth zones (e.g., coats that form long after primary diamond crystallization) experience relatively shorter residence times and may result in a shifting effect when comparing data from specific zones of a single diamond (Kohn et al. 2016). As access to the diamond specimens was not granted by the GIA, it cannot be ruled out that some of the spectra studied here represent averaged data recorded from both primary and secondary growth zones. In the following examples, we show how data derived from spectroscopic analysis of such samples is not *scattered* but is *linearly shifted* with respect to the composition of each growth zone. Consequently, decomposition of the bulk spectra produces an averaged N-aggregation state that can be linearly

compared to %IaA and N-defect contents (and resultant mantle residence time/temperatures) derived from other bulk spectra as $\boldsymbol{C} \rightarrow \boldsymbol{A}$ proceeds in all zones of a given diamond according to a second-order rate equation. As shown by Kohn et al. (2016), detailed thermochronometric interpretations using bulk spectra is not possible but given a large enough dataset (e.g., 178 diamonds studied in this work), averaged data can serve as a means to extract meaningful trends from such data as was done in this study.

For example, consider the following zoned diamond (sample ISTD-13 from Zedgenizov et al. 2016) with central and outer zones described as follows;

central zone: $N_A = 257$; $N_C = 10$; $N_{tot} = 267$ at.ppm; IaA = 96%

outer zone: $N_A = 40$; $N_C = 18$; $N_{tot} = 58$ at.ppm; IaA = 69%

Using a ***C*** to ***A*** activation energy of 5.5 eV and assuming a mantle residence temperature of 900°C, a mantle residence time of 8.1 Myr and 3.5 Myr is calculated for the central and outer zones, respectively. As this difference cannot be attributed to the central and outer zones being subject to different residence temperatures (Kohn et al. 2016), it is likely that the outer secondary zone formed ~4.6 Myr after the central zone. A hypothetical bulk (DRIFT) analysis of sample ISTD-13 will produce an "averaged" spectrum that reflects the N-defects contents averaged based on the total volumes of the central and outer zones. If we assume the central zone occupies 70, 50, 30 and 5% of the total analytical volume (volume sampled by the IR beam during DRIFT analysis), the following %IaA values are derived, 93.9, 91.4, 87.1, and 74.8 %IaA which correspond to mantle residence times of 6.7, 6.2, 5.4, and 3.6 Ma, respectively. These values, in addition to those determined from spectra recorded from the central and outer zones (taken from Zedgenizov et al. 2016) are plotted in the Fig. A.3. In addition, data from the central and outer zones of sample

ISTII-28 (taken from Zedgenizov et al. 2016) were processed in the same way as described above for sample ISTD-13 and are plotted in Fig. A.4.

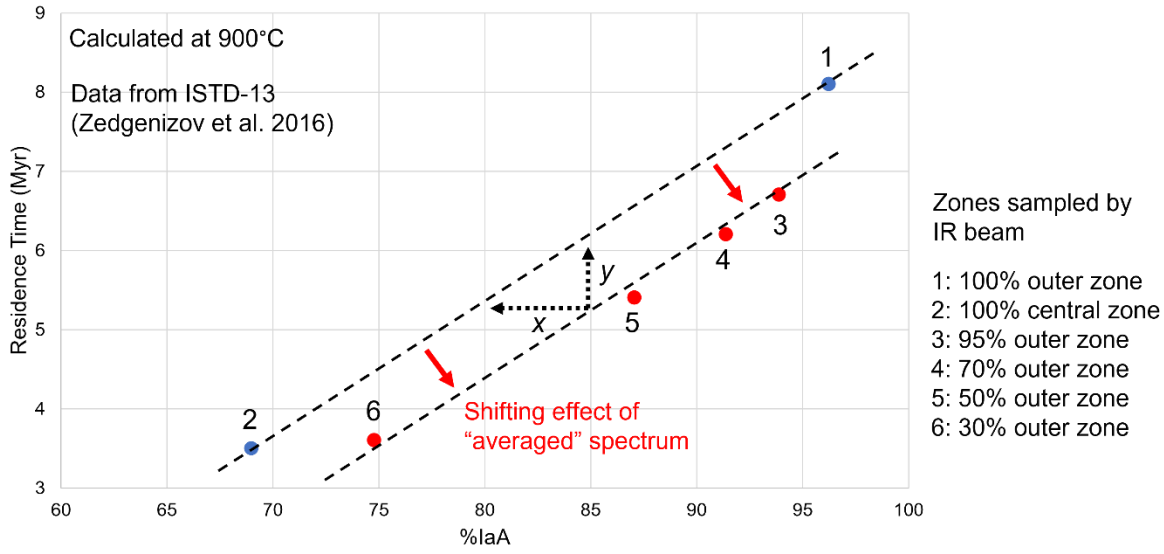

Figure A.3. The %IaA values of the outer and central zones (blue points 1 and 2) of sample ISTD-13 (from Zedgenizov et al. 2016) plotted as a function of mantle residence time. Red points 3-6 represent data from hypothetical "averaged" spectra recorded from sample ISTD-13 where the volume % of central and outer zones sampled by the IR beam varies from 95, 70, 50 and 30% for the outer zone. The shifting effect (red arrows) due to collecting an averaged spectrum (sampling both zones to different degrees) is linearly related (best-fit lines have almost identical slopes) to the line defining the relationship between points 1 and 2. Data was calculated assuming a mantle residence temperature of 900 °C and a ***C*** to ***A*** activation energy of 5.5 eV.

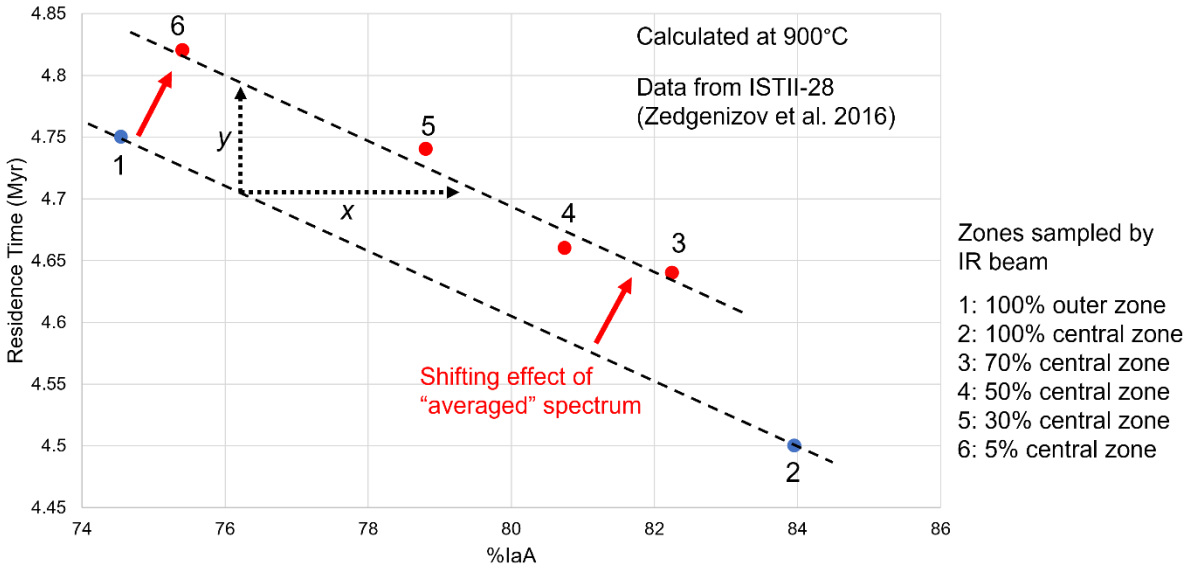

Figure A.4. The %IaA values of the outer and central zones (blue points 1 and 2) of sample ISTII-28 (from Zedgenizov et al. 2016) plotted as a function of mantle residence time. Red points 3-6 represent data from hypothetical "averaged" spectra recorded from sample ISTII-28 where the volume % of central and outer zones sampled by the IR beam varies from 70, 50, 30 and 5% for the central zone. The shifting effect (red arrows) due to collecting an averaged spectrum (sampling both zones to different degrees) is linearly related (best-fit lines have almost identical slopes) to the line defining the relationship between points 1 and 2. Data was calculated assuming a mantle residence temperature of 900 °C and a ***C*** to ***A*** activation energy of 5.5 eV.

The results of collecting bulk (averaged) spectrum from samples ISTD-13 and ISTII-28 is identical, however the sign of the line is negative in Fig. A.4 as ***C*** to ***A*** aggregation is sensitive to not only temperature but N_{tot} and sample ISTII-28 has a significantly lower average N_{tot}. In Figs. A.3 and A.4, the degree of shifting in terms of residence time (indicated as "*y*") and in terms of %IaA (indicated as "*x*") is linearly related to the difference in N_{tot} and %IaA (the N_{tot}/%IaA ratio)

in both the outer and central zones. It follows that mantle residence times derived from bulk spectra are not *scattered* but instead *linearly shifted* with respect to the residence times derived from individual sectors and thus any trends in such data are preserved. Moreover, the examples above show that variations in mantle residence times, due to neglecting ***Y***-centers during decomposition are not scattered by the "bulk" nature of the spectra studied here. The "shifting effect" (as described above) may linearly shift the *uncorrected* and ***Y****-corrected* residence times and the **difference** between the *uncorrected* and ***Y****-corrected* residence times (which is the value of most importance here) cannot be skewed with respect to the true error due to neglecting ***Y***-centers during deconvolution. In other words, estimates of the error in mantle residence times (e.g., 200 Myr) due to neglecting ***Y***-centers during deconvolution will be the same regardless if they derived from averaged spectra or from spectra recorded from discrete growth zones. Of course, data from discrete zones may yield information about local processes that occur within a single diamond that would not be accessible from a bulk spectrum of the diamond. However, our aim is to identify the *general* trends related to ***Y***-centers and other defects and understand the *basic* processes by which early N-aggregation proceeds.

Now the effect of data derived from bulk spectra on comparisons of N-defect contents (e.g., N_Y (at.ppm)) and %IaA can be assessed. Consider the hypothetical diamonds with varying degrees of zonation shown in Fig. A.5.

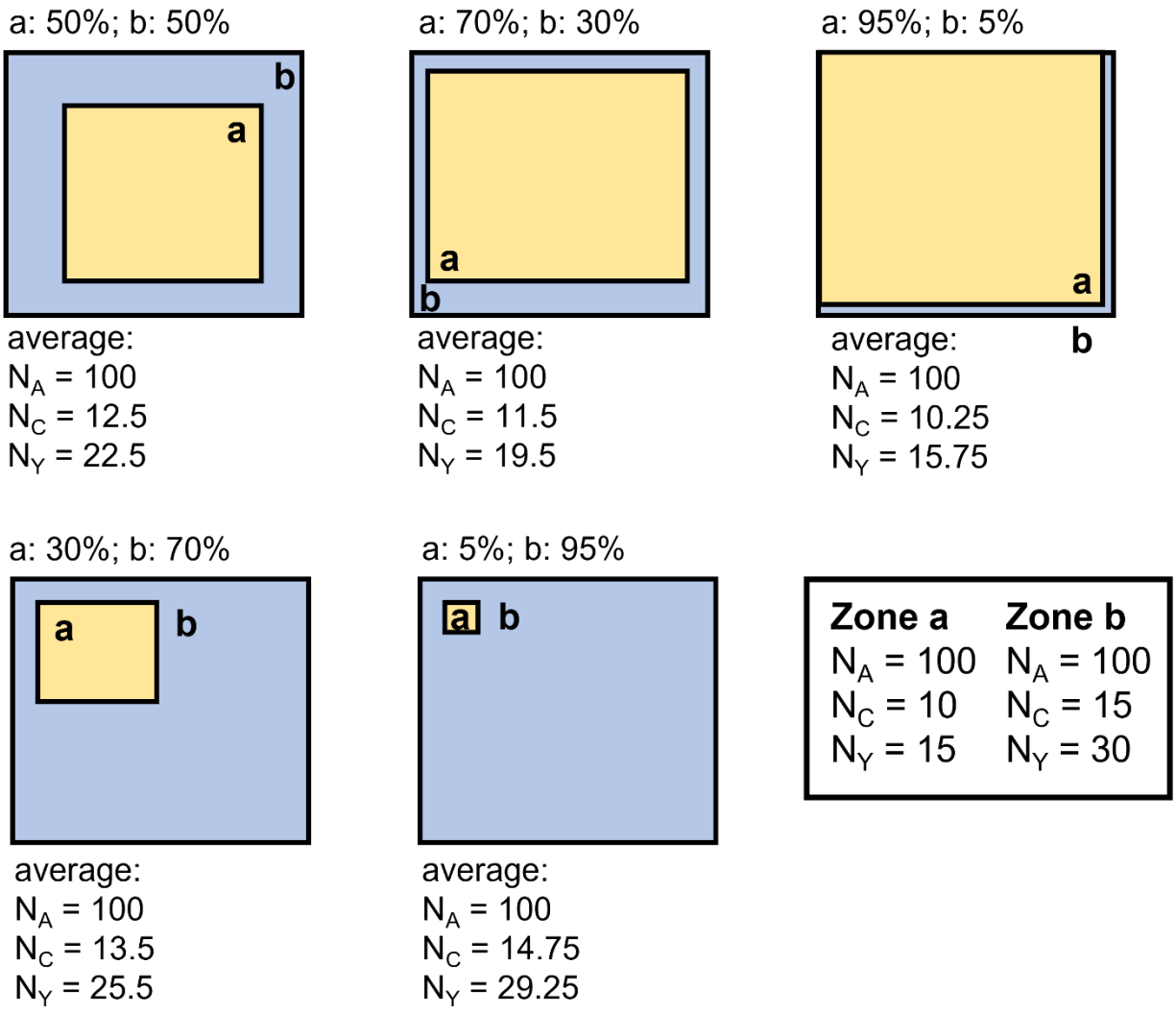

Figure A.5. Hypothetical bi-zoned diamonds composed of varying volumes of growth zones "**a**" and "**b**". Assuming bulk IR spectra are collected on the entirety of each diamond, averaged N-defects contents (N_A, N_C and N_Y) are calculated based on the individual compositions of zones **a** and **b**.

For each of the hypothetical diamonds in Fig. A.5, a bulk composition is calculated assuming the IR beam has sampled the entire diamond. To evaluate this shifting effect, N_Y vs. %IaA is plotted for each diamond in Fig. A.6.

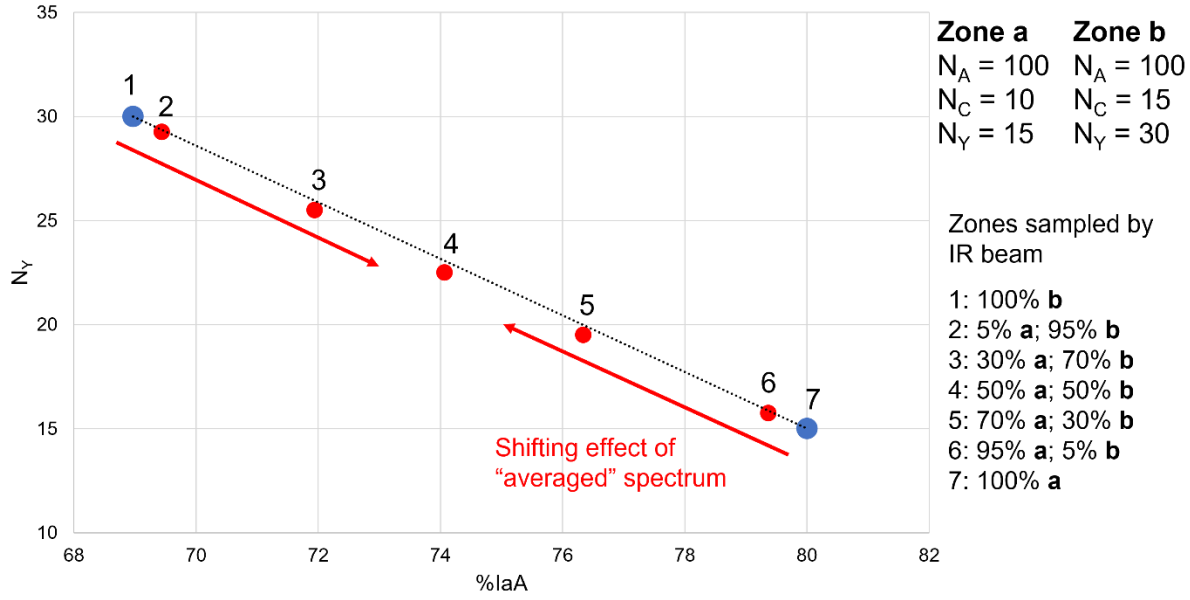

Figure A.6. The %IaA values of pure zones **a** and **b** (blue points 1 and 7) plotted as a function of N_Y. Red points 2-6 represent data from hypothetical "averaged" spectra recorded from each of the five diamonds in Fig. A.5.

In Fig. A.6, pure zones **a** and **b** serve as *end-members* in the N_Y-%IaA compositional space. Point 4 represents data derived from a bulk spectrum recorded from a diamond in which zone **a** and **b** each occupy 50% of the diamond by volume (i.e., or where the diamond-volume sampled by the IR beam is comprised of 50% zone **a** and 50% zone **b**). Decreasing the proportion of zone **b** shifts points *linearly* away from point 1 and decreasing the proportion of zone **a** shifts points *linearly* away from point 7. The dashed line in Fig. A.6 represents the line along which a bulk spectrum may plot depending on the proportion of different zones sampled by the IR beam. In the case where a diamond contains more than two zones, a higher dimensional (2D or 3D) compositional space will be defined, inside which data from bulk spectra may plot. One can envision a series of

hypothetical bi-zoned *Type Ib + IaA* diamonds where the N-defects in each zone control the slope and length of each compositional line as shown in Fig. A.7.

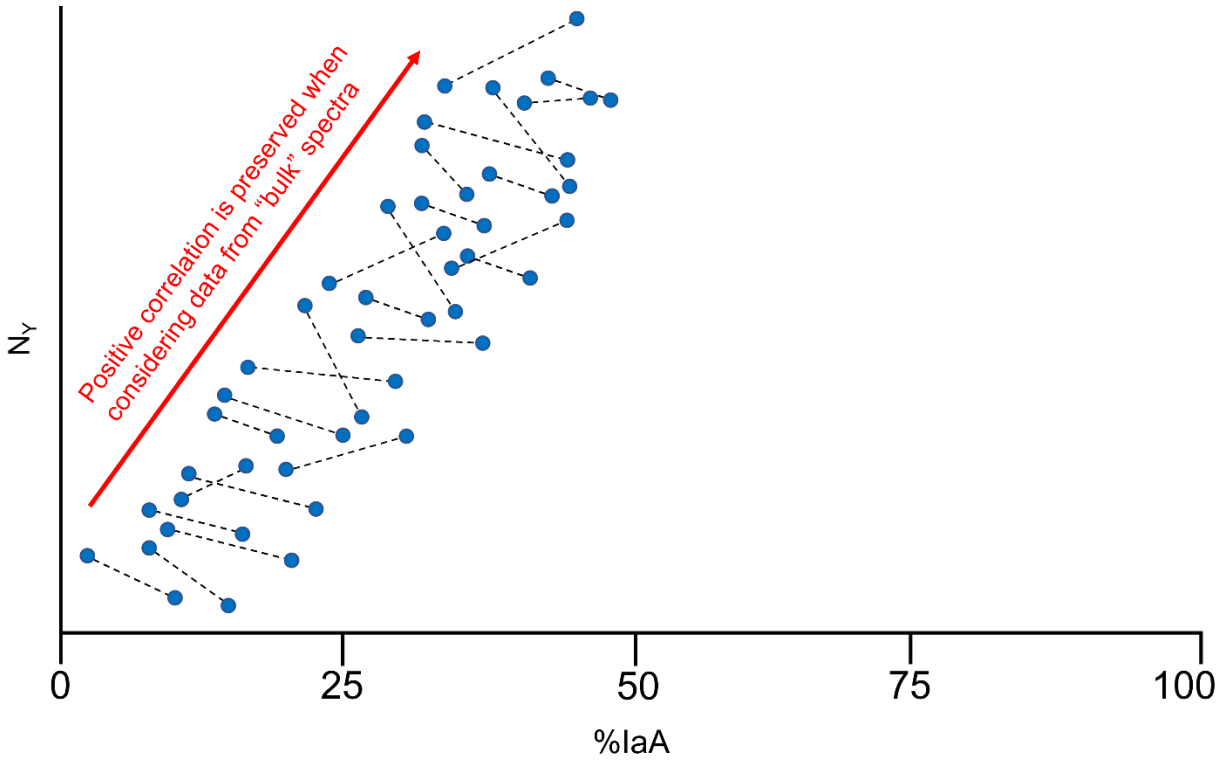

Figure A.7. The N_Y-%IaA compositional lines for a series of hypothetical bi-zoned diamonds, blue points indicate data derived from individual zones in each diamond as shown in Fig. A.6. As data derived from bulk spectra recorded from each diamond must plot along a linear compositional join (dashed lines), general trends in N-defect contents will be preserved.

In Fig. A.7 we show how a hypothetical positive trend would be preserved if data derived from the bulk spectra of a series of bi-zoned diamonds were used instead of the spectra collected from each of the discrete zones in each diamond. In fact, it is apparent how using data from bulk spectra may strengthen correlations as such points will plot between the two end-member points (blue circles)

or in higher dimensional compositional fields depending on the number of discrete growth zones (e.g., a diamond with 5 growth zones will define a 3D compositional space bounded by 5 points). It is apparent from Fig. A.7 that the bulk nature of spectra collected from zoned diamonds does not scatter data but instead linearly shifts data with respect to the composition of each discrete zone (end-member) and thus preserves trends as observed in this study.

References

Hainschwang, T., Fritsch, E., Notari, F., and Rondeau, B. (2012) A new defect center in type Ib diamond inducing one phonon infrared absorption: The Y center. Diamond and Related Materials, *21*, 120–126.

Kohn, S.C., Speich, L., Smith, C.B. and Bulanova, G.P. (2016) FTIR thermochronometry of natural diamonds: A closer look. Lithos, *265*, 148-158. https://doi.org/10.1016/j.lithos.2016.09.021

Mashkovtsev, R. I., Rakhmanova, M. I., & Zedgenizov, D. A. (2021). Specific spectroscopic features of yellow cuboid diamonds from placers in the north-eastern Siberian Platform. *Journal of Geosciences*, 117–126. https://doi.org/10.3190/jgeosci.323

Reutsky, V. N., Shiryaev, A. A., Titkov, S. V., Wiedenbeck, M., & Zudina, N. N. (2017). Evidence for large scale fractionation of carbon isotopes and of nitrogen impurity during crystallization of gem quality cubic diamonds from placers of North Yakutia. *Geochemistry International*, *55*(11), 988–999. https://doi.org/10.1134/S001670291711009X

Skinner, E. M. W., Apter, D. B., Morelli, C., & Smithson, N. K. (2004). Kimberlites of the Man craton, West Africa. *Lithos*, *76*(1–4), 233–259. https://doi.org/10.1016/j.lithos.2004.04.034

Smit, K. V., D'Haenens-Johansson, U. F. S., Howell, D., Loudin, L. C., & Wang, W. (2018). Deformation-related spectroscopic features in natural Type Ib-IaA diamonds from Zimmi (West African craton). *Mineralogy and Petrology*, *112*(S1), 243–257. https://doi.org/10.1007/s00710-018-0587-6

Smit, K. V., Shirey, S. B., & Wang, W. (2016). Type Ib diamond formation and preservation in the West African lithospheric mantle: Re–Os age constraints from sulphide inclusions in Zimmi diamonds. *Precambrian Research*, *286*, 152–166. https://doi.org/10.1016/j.precamres.2016.09.022

Taylor, W. R., Canil, D., & Judith Milledge, H. (1996). Kinetics of Ib to IaA nitrogen aggregation in diamond. *Geochimica et Cosmochimica Acta*, *60*(23), 4725–4733. https://doi.org/10.1016/S0016-7037(96)00302-X

Titkov, S. V., Mineeva, R.M., Zudina, N.N., Sergeev, A.M., Ryabchikov, I.D., Shiryaev, A.A., Speransky, A. V., and Zhikhareva, V.P. (2015) The luminescent nature of orange coloration in natural diamonds: optical and EPR study. Physics and Chemistry of Minerals, 42, 131–141.

Zedgenizov, D. A., Kalinina, V. V., Reutsky, V. N., Yuryeva, O. P., & Rakhmanova, M. I. (2016). Regular cuboid diamonds from placers on the northeastern Siberian platform. *Lithos*, *265*, 125–137. https://doi.org/10.1016/j.lithos.2016.04.012

Zedgenizov, D., Reutsky, V., & Wiedenbeck, M. (2017). The Carbon and Nitrogen Isotope Characteristics of Type Ib-IaA Cuboid Diamonds from Alluvial Placers in the Northeastern Siberian Platform. *Minerals*, *7*(10), 178. https://doi.org/10.3390/min7100178