Diamond Nanothermometry using a machine learning approach

Dylan G. Stone,[1] Yongliang Chen,[2] Evgeny A. Ekimov,[3,4] Toan Trong Tran,[2,5] Carlo Bradac[1]

1. Trent University, Department of Physics & Astronomy, 1600 West Bank Drive, Peterborough, ON, K9L 0G2, Canada.
2. School of Mathematical and Physical Sciences, University of Technology Sydney, Ultimo, NSW, 2007, Australia.
3. Institute for High Pressure Physics, Russian Academy of Sciences, Troitsk 108840, Moscow, Russia.
4. Lebedev Physics Institute, Russian Academy of Sciences, Moscow 117924, Russia.
5. School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, NSW, 2007, Australia

Abstract

With applications ranging from biomedicine to high-power microelectronics, nanothermometry has become a powerful tool for monitoring and controlling temperature at the nanoscale. Most of the nanothermometry techniques developed to date utilize secondary nanothermometers that require the calibration of each individual nanosensor prior to use, ideally both off- and in-situ. Here, we propose an alternative method that addresses this practical limitation. The method utilizes fluorescent nanodiamonds co-hosting germanium-vacancy and silicon-vacancy centers, and is based on a machine learning, multi-feature regression algorithm. The technique is attractive for practical scenarios where the calibration of each nanothermometer before deployment is difficult or unfeasible. The algorithm has also the merit to be general and suitable for any nanothermometry technique that utilizes nanosensors with at least two temperature-dependent observables.

1. Introduction

With its ability to measure and control temperature at the nanoscale, nanothermometry has become a powerful tool for the characterization and understanding of submicrometric systems whose dynamics and performance are heavily influenced by temperature. The access to this knowledge has fostered a wide range of applications in fields such as nano-medicine,[1–8] microfluidic[9] and nanoelectronics,[10] with fundamental and practical realizations ranging, from thermally-driven gene expression[11,12] and cancer therapy[13,14] to temperature management in high-power microelectronics.[15]

The number of available nanothermometry techniques is large. This is mainly due to the widespread variety of existing and engineered nanomaterials that display physico-chemical properties highly sensitive to thermal variations, and that are suitable to operate at specific regimes and temperature ranges. Nanoscale thermal probes include organic dyes, fluorescent proteins, polymers, and inorganic nanoparticles such as quantum dots, gold nanostructures, upconversion nanoparticles and color centers in nanodiamonds.[1,2,16–20]

Most of these temperature probes are secondary nanothermometers. They are relative rather than absolute and require calibration against a reference system or known set of temperatures. This is a key aspect—and a rather time-consuming one—as performing practical measurements

requires the direct calibration of every individual nanothermometer. Also—ideally—said calibration should be performed both off- and in-situ as the environment surrounding the nanothermometers can significantly affect their optical properties and result in so-called thermal equivalent noise (TEN).[16,21] Here we propose an all-optical nanothermometry technique that addresses these limits. The technique utilizes fluorescent nanodiamonds and is based on a machine learning (ML) multi-feature regression (MFR) algorithm. Whilst a similar method based on the multiparametric analysis of fluorescent nanodiamonds has already been demonstrated,[22] our technique bears a few key differences. We use fluorescent nanodiamonds that are co-doped to host both germanium-vacancy (GeV) and silicon-vacancy (SiV) centers, in high concentrations.[23] This allows for the monitoring of potentially twice as many temperature-dependent physical quantities or features (intensity, zero-phonon line emission wavelength and linewidth) as the GeV and SiV centers are excited simultaneously. We also show that once the algorithm has been trained with as few as five nanodiamonds and at as few as three reference calibration temperatures, the ML-MFR algorithm can make temperature determinations with a resolution $\sim 1\ \mathrm{K} \cdot \mathrm{Hz}^{-1/2}$ on any uncalibrated nanodiamond. This is a modest value, yet the main advantage of our approach is that it overcomes the need for calibrating every specific nanothermometer prior to use. Also, we show that the resolution and accuracy of the algorithm can be improved by increasing the number of nanodiamonds and calibration temperatures used in the training phase, and that the algorithm can be customized to automatically select any subset of features to minimize uncertainty. This matters in practical scenarios where control over these parameters can be directly leveraged to streamline the calibration process and the execution of temperature measurements. Finally, in this work we deliberately deviate from traditional investigations that characterize the performance of a nanothermometry technique via the resolution, accuracy and sensitivity of the nanothermometers on which the calibration was carried out. Instead, we estimate their *generalized* values—determined from uncalibrated nanothermometers. The resulting generalized resolution, accuracy and sensitivity are thus not best-case-scenario values, but rather lowest-bound limits that are realistic in practical applications where it is difficult—or unfeasible—to calibrate each specific nanosensor before use.

2. Results and Discussion

For our experimental characterization, we use nanodiamonds (NDs) containing both germanium-vacancy (GeV) and silicon-vacancy (SiV) color centers in high concentration (cf. Methods, §4.1 and Figure 1a). We monitor their photoluminescence and spectral properties via a custom-built confocal microscope integrated with an open-loop, temperature-controlled cryostat (cf. Methods, §4.2). Briefly, the nanodiamonds are dropcasted on a silicon dioxide substrate mounted on the scanning piezo-stage of the cryostat. A single continuous-wave laser at $532\ \mathrm{nm}$, focused through a high numerical aperture objective ($NA = 0.9$), is used to optically excite the nanodiamonds. The photoluminescence from the GeV and SiV centers is then back-collected through the same objective and sent to a spectrometer that measures the centers' photoemission intensity and spectra (Figure 1b) and their change as the temperature of the sample is adjusted to several target values.[24,25]

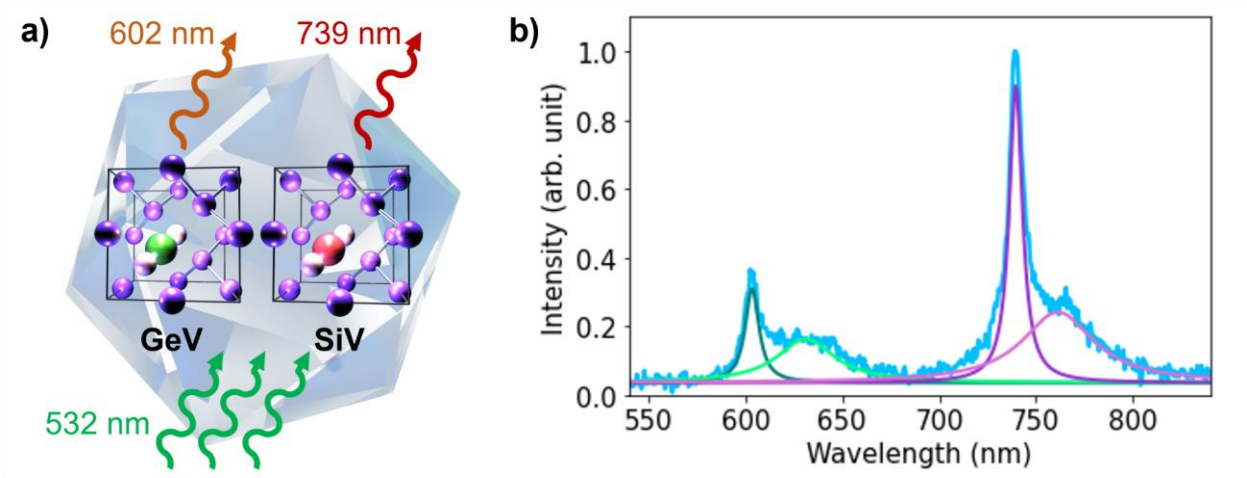


Figure 1. Photoluminescence properties of diamond particles used as nanothermometers. **a)** Schematic of diamond nano/microparticles containing both germanium-vacancy (GeV) and silicon-vacancy (SiV) centers. The GeV and SiV centers are co-excited by the same 532-nm laser and emit photons at ~602 nm and ~739 nm, respectively. **b)** Spectrum of a typical nanodiamond acquired at 80 °C using the 532-nm excitation laser with 150 μW of power and an integration time of 1 s. The zero-phonon line (ZPL) and phonon-sideband (PSB) of both the GeV at ~603.5 nm and ~631.2 nm and of the SiV centers at ~739.7 nm and ~761.4 nm, respectively, are clearly distinguishable.

Note that because the nanodiamonds in our experiment host GeV and SiV centers, a total of up to 12 physical quantities can be monitored, simultaneously, as temperature is varied. These are: intensity, position and linewidth of the zero-phonon line (ZPL) wavelength and of the phonon-sideband (PSB), for both the GeV and SiV centers. Desirably, the emission lines of the two centers are well separated in wavelength—which simplifies the data extraction and fitting procedures. For reference, figures 2a–d show the measured shifts in the photoemission ZPL wavelength and full-width at half maximum (FWHM) of the GeV and SiV centers of a representative nanodiamond as temperature is varied in the range $25 - 80$ °C.

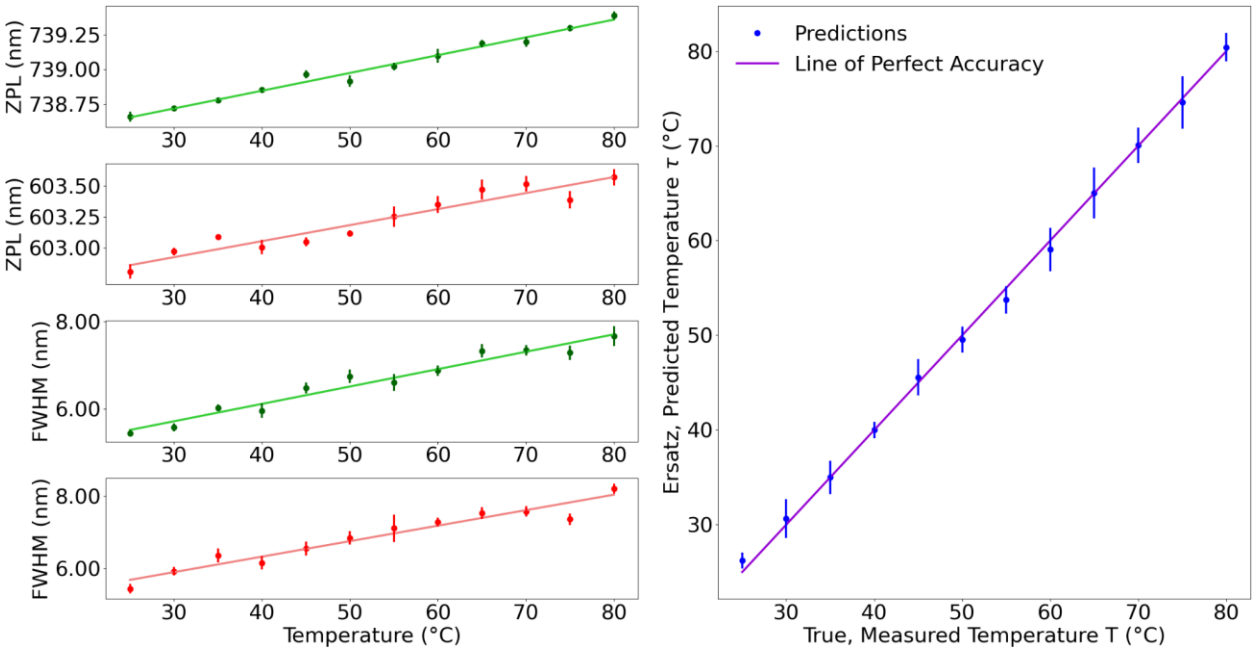


Figure 2. Temperature-dependence of the SiV and GeV photoluminescence spectral properties. **a–d)** Temperature-dependent changes of the zero-phonon line (ZPL) emission wavelength (a, b) and of the full-width at half-maximum (FWHM) linewidth (c, d) for the SiV (green traces) and GeV (red traces) centers, respectively, for a representative nanodiamond. Raw data points are fitted with the line of best fit; error bars are given as standard deviation of the measurement for each temperature. **e)** Accuracy of the multi-feature regression model shown as the difference between the true (i.e., measured) value of the temperature (x-axis) and the value predicted by the model (y-axis). The straight line is the line of 'perfect' accuracy. Note that whilst the error bars in (e) seem greater that those in (a)–(d), they are a factor ~1.2 × smaller: the relative errors are, in fact, ~3.3% in (e) vs. ~3.9% in (a)–(d).

The ability to simultaneously monitor such a large number of temperature-dependent quantities —from here on referred to as *features*—makes multi-feature regression an attractive strategy to determine the temperature of the sample. The hypothesis is that a multi-feature model could be less prone to correlations between fitting parameters and to (experimental) fluctuations or errors—thanks to averaging—especially if such fluctuations affect just one or a few features.[22]

Our multi-feature regression (MFR) model, is based on classical machine-learning (ML) multiple linear regression (cf. Methods, §4.3).[26,27] Briefly, a sets of known spectroscopy-vs-temperature data from the GeV-SiV co-doped nanodiamonds is used to train and test the MFR algorithm, which can then be used to determine the temperature of any nanodiamond. It is worth noting that the ML strategy of building a predictive model based on training datasets inherently mirrors that of nanothermometry measurements based on calibration datasets—making ML-based MFR a natural way of approaching the task. Expectedly, one of our goals is to determine how the ML-MFR-based algorithm fares against traditional all-optical nanothermometry methods based on monitoring the change of a single parameter (e.g., ZPL wavelength or FWHM) as a function of temperature.[1,2,16–19] We specifically focus on a few aspects of the algorithm, mainly, on its generalized resolution, accuracy and sensitivity (cf. Methods, §4.4), and on how its performance

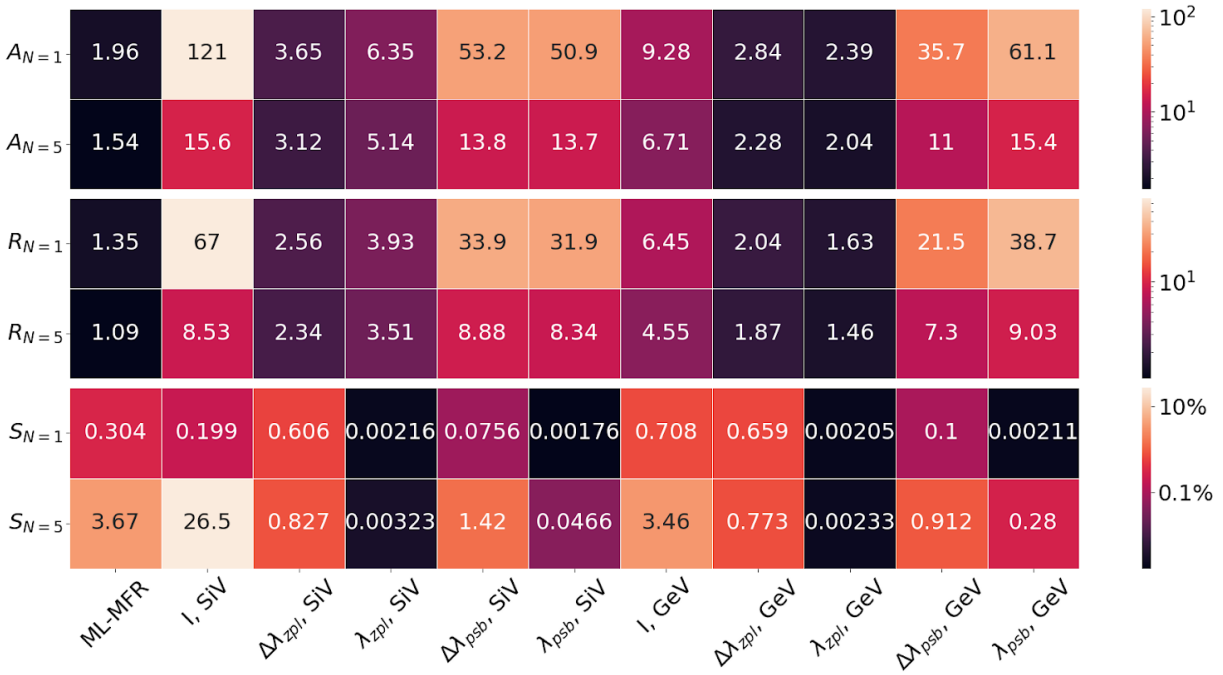
can be tuned by choosing the optimal subset of features, number of training/test data, as well as number of temperature calibration steps.

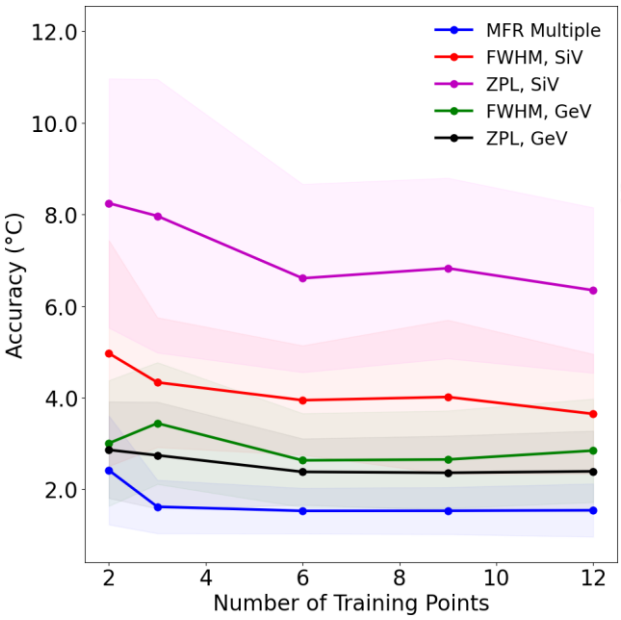
Beyond the traditionally quoted values of resolution, accuracy and sensitivity, in this work we are interested in determining their *generalized* values, which take into account the results of uncalibrated nanothermometers (see below and §4.4). This is motivated by a series of observations. First, the traditional resolution and sensitivity of our technique are comparable to those of any other all-optical nanothermometry approach. This follows from the fact that we are monitoring the thermal dependence of the same physical observables (Intensity, ZPL, FWHM, etc.)—simply more of them, simultaneously. Notably, the resolution of our ML-MFR model shows a modest improvement, by factors $\sim 1.2-8.3\times$, compared to methods that monitor the temperature-related changes of a single observable. This is consistent with previous reports and is attributed to the multi-feature model suppressing the correlations between fitting parameters and to noise averaging.[22] Second, the standard resolution, accuracy and sensitivity are optimized figures of merit that are only partially suitable to the scenario we are interested in—one where the temperature calibration is performed on nanosensors (the training set) that are *not* the ones used as the actual nanothermometers. When the temperature calibration is done on the same nanothermometer, its sensitivity and resolution can simply be extracted from the noise/standard deviation of the individual measurements (cf. Methods, §4.4). Yet, this overlooks the fact that the overall fitting model might have a larger error, i.e., a larger deviation from the 'true' value, which can become even larger when the model is applied to the data of uncalibrated nanosensors. This instead matters when the aim is to determine a universal algorithm that works on any nanothermometer, even those that were not utilized for calibration.
Our intent is deliberately practical. We want to measure how accurately any nanothermometer can determine a certain temperature value, given that the only information we have is a temperature-calibration measurement obtained—once—from a reference training set of nanothermometers. The generalized values of resolution, accuracy and sensitivity that include the data of uncalibrated nanosensors are therefore the figures of merit we are mostly interested in.

Figure 3a shows a comparison between the resolutions, accuracies and sensitivities of the various methods. In the heatmap, better (i.e., smaller) values of resolution and accuracy are indicated by darker colors, whilst better (i.e., higher) values of sensitivity are indicated by lighter colors. Note that to highlight the difference in resolution, accuracy and sensitivity of the various models, the intensity color scale of the heatmap is logarithmic.
In the heatmap of figure 3a, there are two values reported for the resolution, $R_{N=1}$ and $R_{N=5}$, the accuracy, $A_{N=1}$ and $A_{N=5}$, and the sensitivity, $S_{N=1}$ and $S_{N=5}$, for each model (cf. Methods, §4.4). The $R_{N=1}$ and $A_{N=1}$ values are obtained by calibrating (or 'training') each model on a single nanodiamond and determining how well—i.e., with what resolution and accuracy—it predicts the temperature of the remaining five (test) uncalibrated nanodiamonds. The values of $R_{N=1}$ and $A_{N=1}$ thus set the lowest limit for the resolution and accuracy of each model, for these are obtained from a single calibration curve done on a single training nanodiamond that is not one of the test nanodiamonds. Conversely, $R_{N=5}$ and $A_{N=5}$, are obtained by calibrating (or 'training') each model on $N = 5$ nanodiamonds and determining how well—i.e., with what resolution and accuracy—it predicts the temperature of the remaining (test) uncalibrated nanodiamond. To estimate the

values of $R_{N=5}$ and $A_{N=5}$ we average all the six possible combinations of 5-training/1-test sets of nanodiamonds. Notably, as the size N of the training set increases, the values of the resolution and accuracy improve (i.e., they get smaller). Similarly, $S_{N=1}$ and $S_{N=5}$ are the values of sensitivities estimated for each model trained on a single nanodiamond or on five nanodiamonds, respectively (more details can be found in the Methods section, §4.4).

From the analysis of these generalized figures of merit, our ML-MFR model displays the best resolution, $R_{N=1} = 1.35\ \mathrm{K} \cdot \mathrm{Hz}^{-1/2}$ and $R_{N=5} = 1.09\ \mathrm{K} \cdot \mathrm{Hz}^{-1/2}$, and accuracy, $A_{N=1} = 1.96\ \mathrm{K}$ and $A_{N=5} = 1.54\ \mathrm{K}$. It is followed in performance, in our measurements, by the single-feature model that tracks the temperature dependence of the GeV's ZPL, for which $R_{N=1} = 1.63\ \mathrm{K} \cdot \mathrm{Hz}^{-1/2}$, $R_{N=5} = 1.46\ \mathrm{K} \cdot \mathrm{Hz}^{-1/2}$, $A_{N=1} = 2.34\ \mathrm{K}$ and $A_{N=5} = 2.09\ \mathrm{K}$.


$A_{N=1}$ 1.96 121 3.65 6.35 53.2 50.9 9.28 2.84 2.39 35.7 61.1
$A_{N=5}$ 1.54 15.6 3.12 5.14 13.8 13.7 6.71 2.28 2.04 11 15.4
$R_{N=1}$ 1.35 67 2.56 3.93 33.9 31.9 6.45 2.04 1.63 21.5 38.7
$R_{N=5}$ 1.09 8.53 2.34 3.51 8.88 8.34 4.55 1.87 1.46 7.3 9.03
$S_{N=1}$ 0.304 0.199 0.606 0.00216 0.0756 0.00176 0.708 0.659 0.00205 0.1 0.00211
$S_{N=5}$ 3.67 26.5 0.827 0.00323 1.42 0.0466 3.46 0.773 0.00233 0.912 0.28
10^2
10^1
10^1
10%
0.1%
ML-MFR
I, SiV
$\Delta\lambda_{zpl}$, SiV
λ_{zpl}, SiV
$\Delta\lambda_{psb}$, SiV
λ_{psb}, SiV
I, GeV
$\Delta\lambda_{zpl}$, GeV
λ_{zpl}, GeV
$\Delta\lambda_{psb}$, GeV
λ_{psb}, GeV


MFR Multiple
FWHM, SiV
ZPL, SiV
FWHM, GeV
ZPL, GeV
Accuracy (°C)
Number of Training Points

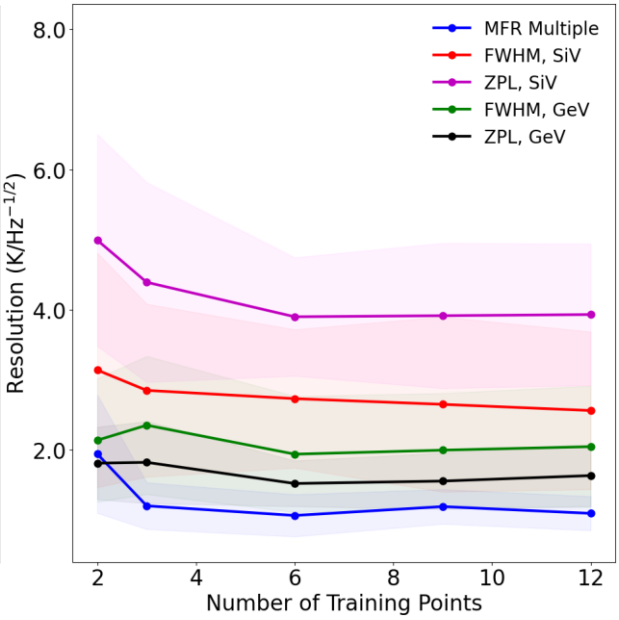

MFR Multiple
FWHM, SiV
ZPL, SiV
FWHM, GeV
ZPL, GeV
Resolution (K/Hz$^{-1/2}$)
Number of Training Points

Figure 3. Performance of the tested nanothermometry models. **a)** Heatmap of the *generalized* accuracy, resolution and sensitivity of the tested models: each column in the heatmap corresponds to a different model. The label ML-MFR indicates the values for our multi-feature regression model and is compared to a selection of other single-feature models based on monitoring the temperature dependence of observables such as the intensity, I, linewidth, $\Delta\lambda$, and wavelength, λ, of the zero-phonon line (ZPL) and phonon sideband (PSB) of either GeV or SiV centers. In the heatmap, darker colors indicate better (i.e., smaller) accuracies and resolutions while lighter colors indicate better (i.e., higher) sensitivities. To better highlight the difference in performance of the various models the color scale of the heatmap is logarithmic. For each model two values of accuracy ($A_{N=1}$ and $A_{N=5}$), resolution ($R_{N=1}$ and $R_{N=5}$) and sensitivity ($S_{N=1}$ and $S_{N=5}$) are listed (see main text). The ML-MFR model displays the best accuracy and resolution of all models, and the second-highest sensitivity; its sensitivity is only inferior to that of the model based on monitoring the intensity of the SiV's ZPL, which however has some of the worst resolution and accuracy. **b, c)** Accuracy (b) and resolution (c) of the top five performing models as a function of the number of temperatures used for their calibration. As expected, a higher number of calibration points for the temperature results in better (i.e., smaller) accuracy and resolution. However, and interestingly, as few as 3 calibration temperatures already produce relatively good values for both figures of merit, with only a small improvement as the number of calibration temperatures is increased to 4 or more.

Analysis of the overall accuracy heatmap reveals a few key points. First, in general a multi-feature approach produces higher accuracies by factors $\sim 1.3-10.1\times$, which is likely a consequence of the averaging effect affecting the noise. It however warrants an important caveat. A MFR algorithm intrinsically assumes a linear relationship between the label (i.e., the temperature) and the features (i.e., the GeV's and SiV's photoluminescence intensities, ZPLs and FWHMs). Over the relatively small temperature range considered in this study ($25-80$ ℃) the assumption of linearity is reasonable—yet it would not necessarily hold for larger temperature ranges for which the spectral features of diamond color centers are known to depend non-linearly on temperature.[24,28] From a practical standpoint, the nonlinearity would result in lowered accuracy. Fortunately, this problem can be mitigated by pre-linearizing the data through their functional dependence with temperature, provided that this dependence is known (which is usually the case for the spectroscopy features of diamond color centers and, in general, of other optical nanothermometers). Also—commonly to ML—one should be wary of the ability of the algorithm to make reliable, generalized predictions beyond the range spanned by the training set.

Second, during the training-test phase of the models, one can select the subset of features that gives the highest resolution, accuracy and sensitivity simply verifying all possible combinations of features. In our case for instance, we found that the best values for $R_{N=1}$, $A_{N=1}$ and $S_{N=1}$ were obtained by choosing the combination of the following 3 features: FWHM of the SiV, and ZPL and FWHM of the Gev. At the same time, the best values for $R_{N=5}$, $A_{N=5}$ and $S_{N=5}$ were obtained by choosing the combination of the following 5 features: FWHM of the SiV, and intensity, ZPL, FWHM and PSB of the GeV.

Importantly, and third, while this process of multiple-feature selection might seem time-consuming, it is completely automated, and it needs to be done only once. In fact, and from a practical standpoint, our analysis shows that even just training the ML-MFR model on a few nanodiamonds produces a general predictive model that can be applied to any other nanodiamond. For reference, in this study we tested over 10^3 possible combinations of features to find the MFR model with the highest resolution and accuracy, but this was done automatically,

running a simple script. It is worth noting that the achieved resolution and accuracy on uncalibrated nanodiamonds are not a merit, solely, of the MFR algorithm. Rather, they underscore how the spectral properties of color centers in diamond are, remarkably, only marginally affected by inhomogeneities in the local environment of the solid-state host nanoparticle.
Fourth, and again typical of ML approaches, the algorithm has the desirable draw of becoming more accurate over time, i.e., as more data is processed by the algorithm (figures 3b, c).

Finally, we are interested in determining the relationship between accuracy and number of calibration points. In other words, we want to determine how many target temperatures the calibration should be performed at to achieve a desired level of accuracy and resolution. Again, this is motivated by the fact that acquiring calibration curves is a time-consuming step for many nanothermometry methods in practical settings. Figure 3b shows the accuracy of the best five models (our MFR model and the four next best ones) as a function of the number of calibration temperatures utilized to train the model. As one expects, in general, increasing the number of temperatures at which the various features are measured, results in a higher overall accuracy of each model. However—and interestingly for practical cases—for the best-performing models the accuracy does not increase significantly when they are trained over more than 3 target calibration temperature values. In our ML-MFR model the accuracy increases significantly by $\sim 33\%$ going from 2 to 3 calibration temperatures, but the improvement drops to a much more modest $\sim 5\%$ going from 3 to 6 calibration temperatures. Therefore, calibrating the nanothermometers against more than a few target temperatures might be an avoidable endeavor in practical scenarios for which time and efficiency are key. A similar conclusion can be obtained for the resolution analysis, which is shown in figure 3c. Again, in our ML-MFR model the resolution increases by $\sim 38\%$ going from 2 to 3 calibration temperatures, but the improvement drops to much more modest $\sim 11\%$ going from 3 to 6 calibration temperatures.

3. Conclusions

To conclude, we have shown an all-optical nanothermometry technique that utilizes fluorescent nanodiamond co-hosting GeV and SiV centers as thermosensors. The technique utilizes a classical machine-learning (ML), multi-feature regression (MFR) algorithm. The motivation for our work is mostly practical with focus on determining the best combination of observable/features and, simultaneously, the smallest number of calibration temperatures one can rely on to determine the temperature of a diamond nanosensor within a desired level of accuracy and resolution. We find that using a subset of 3-5 features we can train the ML-MFR algorithm with as few as 3 calibration temperatures on a set of 5 nanodiamonds and produce a model that can predict the temperature of any other diamond nanosensor with $1.5\ \mathrm{K}$ accuracy and $1.1\ \mathrm{K} \cdot \mathrm{Hz}^{-1/2}$ resolution. The method is also appealing as it is general and, besides nanodiamonds, can be applied to any nanothermometer possessing at least two temperature-dependent observables or features.

4. Materials and Methods

In this work we focus on fluorescent nanodiamonds as our test system—specifically nanodiamonds containing both germanium-vacancy (GeV) and silicon-vacancy (SiV) centers. We perform nanothermometry measurements based on detecting temperature-dependent changes

in the centers' photoluminescence spectra including for instance changes of the zero-phonon line (ZPL) wavelength and full width at half maximum (FWHM).[22,24,25,29,30] Note however, that our ML-MFR model can be generalized to other nanothermometry techniques, including those based on other nanoprobes as well as non-all-optical ones.

4.1 Sample preparation

The diamond micro- and nanoparticles co-doped with germanium and silicon were produced via high pressure and temperature in a C-H-Si(0.19 at%)-Ge(0.2 at%) growth system. Microdiamonds were synthesized at a pressure of $8\ \mathrm{GPa}$ and at temperatures in the range $1800-2000\ ^{\circ}\mathrm{C}$, while nanodiamonds at a pressure of $9\ \mathrm{GPa}$ and at temperatures in the range $1500-1600\ ^{\circ}\mathrm{C}$, for about $60\ \mathrm{s}$. For the synthesis, Adamantane $C_{10}H_{16}$ ($300\ \mathrm{mg}$, $>99\%$, Sigma-Aldrich), Tetraphenylgermane $C_{24}H_{20}Ge$ ($15\ \mathrm{mg}$, 96%, Aldrich) and Tetraphenylsilane $C_{24}H_{20}Si$ ($18\ \mathrm{mg}$, 96%, Aldrich) were mixed with a jasper mortar and pestle for about 5 minutes, pressed into a pellet ($65\ \mathrm{mg}$) and placed inside a titanium capsule ($6\ \mathrm{mm}$ in diameter, $4\ \mathrm{mm}$ in height, with a $0.2\ \mathrm{mm}$ wall thickness). A toroid-type, high-pressure chamber was used to generate the target pressure and temperature in the reaction cell.[31] After the treatment, samples were quenched under pressure to room temperature. The concentration of diamond SiV and GeV centers obtained by this synthesis process is estimated to be ~p.p.b.

For characterization, the nanodiamonds (sizes $\sim 300-500\ \mathrm{nm}$) were then dispersed in isopropanol (IPA) at the concentration of 0.1% (w/w), dropcast on a clean $0.5\times 0.5\ \mathrm{cm}^2$ silicon substrate and left to dry on a hotplate at $60\ ^{\circ}\mathrm{C}$ to remove the residual solvent.

4.2 Experimental setup

The photoluminescence properties of the nanodiamonds were investigated using a lab-built confocal setup described elsewhere.[30] Briefly, the sample was mounted on a three-dimensional piezo-stage (ANPx series; Attocube Inc.) inside an open-loop cryostat (custom-adapted from a ST500 cryostat; Janis) with flowing liquid nitrogen. Temperature was controlled via a cryogenic temperature controller (335; Lakeshore) and optical access to the sample was achieved through a thin quartz window. The excitation laser (CW at $532\ \mathrm{nm}$) was focused via a high-numerical aperture air objective ($\mathrm{NA}=0.90;\ 100\times$; TU Plan Fluor; Nikon). The emitted light was back-collected through the same objective, spectrally filtered, and sent to a spectrometer (QEPro, Ocean Optics).

4.3 Machine learning multi-feature regression model

The philosophy of most machine learning approaches is to predict certain outcomes based on past observations. This can be done by using a so-called *training dataset* to devise a predictive model and a *test dataset* to verify its accuracy. Note that both the training and the test sets contain known data, so the accuracy of the model can be determined quantitatively before it is applied to an unknown dataset. The ML approach we use in this work is a simple multi-feature regression model, where the value of the temperature T is predicted based on the simultaneous measurement of n *features* $\boldsymbol{x}_i=[x_{i1},x_{i2},\dots,x_{in}]$ that all vary with temperature. These can include observables such as photoluminescence intensity, zero phonon line (ZPL) wavelength and full width at half maximum (FWHM) of color centers in nanodiamonds. The predicted temperature ersatz τ is thus given by:

$$\tau_i(1, x_{i1}, x_{i2}, \dots, x_{in}) = \sum_{j=0}^{n} w_{ij} x_{ij} = w_{i0} + w_{i1} x_{i1} + w_{i2} x_{i2} + \cdots + w_{in} x_{in} \quad (1)$$

Which is a generalization of the slope-intercept linear regression for $n > 1$ features. The term 'ersatz' here is used to indicate that τ_i is the function defined by equation (1) to predict the true temperature as measured by the cryostat sensor. Note that we use the extended vector $\boldsymbol{x}_i = [1, x_{i1}, x_{i2}, \dots, x_{in}]$, whose first element is '1,' to account for the bias or intercept value w_0. The index $i = 1, 2, \dots, N$ runs over the total number N of observations. The values of the weight coefficients $\boldsymbol{w}_i = [w_{i0}, w_{i1}, \dots, w_{in}]$ are extracted from the training dataset via minimization of a defined *cost function*. A standard cost function is the least squares error (LSE), or L2-norm loss function:

$$LSE = \sum_{i=1}^{N} (\boldsymbol{w}_i \boldsymbol{x}_i - \tau_i)^2 \quad (2)$$

Alternatively, the L1-norm loss function or least absolute deviations (LAD) can be used:

$$LAD = \sum_{i=1}^{N} |\boldsymbol{w}_i \boldsymbol{x}_i - \tau_i| \quad (3)$$

We used both LSE and LAD in this work. The former is one of the most common ways of rating how well a model fits the data, but the latter has the advantage of being a direct measure of a model's overall accuracy to predict the temperature from the measured feature(s). Note that in equations (1)–(3), each coefficient w_{ij}, with $j = 1, 2, \dots, n$, has its specific units determined by the corresponding feature x_{ij}. Also, in our analysis, the range over which each feature can vary as a function of temperature is normalized to the interval $[0, 1]$ to avoid over- or under-weighting the contributions of features with larger or smaller absolute values, respectively.

4.4 Data analysis

In this section we discuss the figures of merit—namely resolution, accuracy and sensitivity—used to assess the performance of the various techniques. Our characterization is based on a series of key elements.

i) In order to compare the performance of our ML-MFR algorithm to that of any other established single-feature method we use our own measurements for all methods, rather than using the literature values. This is so that we can make a relative comparison, independent of a particular measurement system's detection efficiency, signal to noise ratio, resolution, etc.

ii) The values we report for resolution, accuracy and sensitivity are what we refer to as *generalized* values. This so-called generalized approach follows a few steps. We first build a model using data from known calibration, training dataset(s)—this is both the case for the ML-MFR algorithm and for any of the single-feature comparative models. We then feed to each model values of features (ZPL, intensity, FWHM, etc.) from unseen (yet known) test nanothermometry datasets, and measure with what resolution, accuracy and sensitivity the model can predict the true temperature (see point **(iv)** below). The difference from many traditional approaches is that this 'generalized' figure of merits are not estimated from the data used to determine the fitting model itself.

iii) As explained in the Results and Discussion section of the paper, for each model we quote two values for the accuracy, $A_{N=1}$ and $A_{N=5}$, resolution, $R_{N=1}$ and $R_{N=5}$, and sensitivity, $S_{N=1}$ and $S_{N=5}$. The subscripts $N = 1$ and $N = 5$ distinguish each corresponding figure of merit based on how the predictive model is determined. The subscript $N = 1$ indicates that the model is extracted fitting the data (of any feature of interest vs. temperature) of a single nanodiamond (ND). In our

study, we have six NDs so each value for $R_{N=1}$, $A_{N=1}$ and $S_{N=1}$ listed in the heatmap of figure 3a is an average over the six sets of available data. Conversely, for $N = 5$ the model is extracted by fitting, simultaneously, the data from five nanodiamonds. The dataset from the sixth ND is then used to measure the accuracy and resolution of each model, as per point **(ii)** above.
iv) In measuring the resolution and accuracy of each technique we compare the temperatures predicted by the various models to the 'true' temperatures. We consider 'true' the values of temperature measured by the cryostat. The rigorous—albeit impractical—definition would require absolute knowledge of the temperature rather than its values measured by a reference instrument.

Resolution. The resolution of a nanothermometry technique is usually defined as $\sigma \cdot \sqrt{t_m}$ where σ is the standard deviation of the measured observable or feature x_{ij} (Intensity, ZPL, FWHM, etc.) and t_m is the measurement integration time. In this work, we measure the resolution as the standard deviation of the absolute differences between the true temperatures and the temperatures predicted by the model, as per points **(ii)**, **(iii)** above. This standard deviation is then multiplied by the integration time t_m (for convenience, $t_m = 1$ s in most of our acquisitions). Also, in the heatmap in figure 3a, the listed resolutions of any single-feature method are determined from the average standard deviation σ obtained at the various temperatures (note, e.g., that in figure 2a–d the error bars are different at different temperatures). In similar studies, the resolution is often determined by selecting the worst (i.e., the largest) or the best (i.e., the smallest) standard deviation σ amongst all the surveyed temperatures.

Accuracy. Accuracy is commonly defined as the absolute difference between the measured (average) value and the 'true' value of an observable—temperature in this case. In this work, the measured value is the value determined by a model, for the model itself is obtained from fitting the experimental data. To determine the accuracy (figure 3a), we therefore directly compare the true temperatures with the ones predicted by each model, as per point **(ii)** above.

Sensitivity. The relative sensitivity of a nanothermometry technique is defined as $|(\partial O/\partial T)/O|$ where O is the measured observable and T is the temperature. The values of sensitivity for each single-feature model listed in figure 3a are obtained using this definition, and as per point **(iii)** above. The value of the sensitivity for our ML-MFR model, is obtained instead as a weighted linear combination of the single-feature sensitivities used by the model. Specifically, as in our model we use up to $n = 5$ features x_{ij} (FWHM of the SiV, and intensity, FWHM, ZPL and PSB of the GeV), we calculate the sensitivity $S_{MFR,i}$ for each i-th dataset as:

$$S_{MFR,i} = \sum_{j=0}^{n} \alpha_{ij} \left(\frac{\partial x_{ij}}{\partial T}\right) / x_{ij} \tag{4}$$

where the individual sensitivities for each feature, $s_{ij} = \left(\partial x_{ij}/\partial T\right)/x_{ij}$, are each weighted by the corresponding coefficient α_{ij}. These coefficients α_{ij} are obtained from the coefficients w_{ij} in equation (1), via normalization: $\alpha_{ij} = w_{ij}/\sum_{j=1}^{n} w_{ij}$. Note that the index j runs over the number of features (from 1 to 5) used in the ML-MFR model, while the index i refers to a dataset (performed on the corresponding i-th nanodiamond and up to $N = 5$).

5. Author Contributions

C.B. and D.G.S. conceived the idea of the project and performed the data interpretation and analysis. Y.C. and T.T.T. carried out the experimental measurements. E.A.E. synthesized the diamond samples. All authors discussed the results and commented on the manuscript.

6. Funding

C.B. thanks the NSERC (RGPIN-2021-03059 and DGECR-2021-00234) and CFI JELF (#41173) for financial support. D.G.S also thanks the NSERC for their support through the Canada Graduate Scholarships-Master's Program. E.A.E. acknowledges support from the Russian Science Foundation, Grant No. 19-12-00407. T. T. T. acknowledges the support from the Australian Research Council (DE220100487). T. T. T. and Y. C. thank the UTS node of Optofab ANFF for the assistance with nanofabrication.

References

1. Bai, T. & Gu, N. Micro/Nanoscale Thermometry for Cellular Thermal Sensing. *Small* **12**, 4590–4610 (2016).
2. del Rosal, B., Ximendes, E., Rocha, U. & Jaque, D. In Vivo Luminescence Nanothermometry: from Materials to Applications. *Adv. Opt. Mater.* **5**, 1600508 (2017).
3. Zhou, H., Sharma, M., Berezin, O., Zuckerman, D. & Berezin, M. Y. Nanothermometry: From Microscopy to Thermal Treatments. *ChemPhysChem* **17**, 27–36 (2016).
4. Nakano, M. & Nagai, T. Thermometers for monitoring cellular temperature. *Spec. Issue Bioimaging* **30**, 2–9 (2017).
5. Sotoma, S., Epperla, C. P. & Chang, H.-C. Diamond Nanothermometry. *ChemNanoMat* **4**, 15–27 (2018).
6. Tsai, P.-C. *et al.* Measuring Nanoscale Thermostability of Cell Membranes with Single Gold–Diamond Nanohybrids. *Angew. Chem. Int. Ed.* **56**, 3025–3030 (2017).
7. Tsuji, T., Ikado, K., Koizumi, H., Uchiyama, S. & Kajimoto, K. Difference in intracellular temperature rise between matured and precursor brown adipocytes in response to uncoupler and β-adrenergic agonist stimuli. *Sci. Rep.* **7**, 12889 (2017).
8. Carrasco, E. *et al.* Intratumoral Thermal Reading During Photo-Thermal Therapy by Multifunctional Fluorescent Nanoparticles. *Adv. Funct. Mater.* **25**, 615–626 (2015).
9. Wu, J. *et al.* Mapping three-dimensional temperature in microfluidic chip. *Sci. Rep.* **3**, 3321 (2013).
10. Yue, Y. & Wang, X. Nanoscale thermal probing. *Nano Rev.* **3**, 11586 (2012).
11. Kamei, Y. *et al.* Infrared laser–mediated gene induction in targeted single cells in vivo. *Nat. Methods* **6**, 79–81 (2009).
12. Kumar, S. V. & Wigge, P. A. H2A.Z-Containing Nucleosomes Mediate the Thermosensory Response in Arabidopsis. *Cell* **140**, 136–147 (2010).
13. Schroeder, A. *et al.* Treating metastatic cancer with nanotechnology. *Nat. Rev. Cancer* **12**, 39–50 (2012).
14. O'Neal, D. P., Hirsch, L. R., Halas, N. J., Payne, J. D. & West, J. L. Photo-thermal tumor ablation in mice using near infrared-absorbing nanoparticles. *Cancer Lett.* **209**, 171–176 (2004).
15. Chen, Y. *et al.* Optical Thermometry with Quantum Emitters in Hexagonal Boron Nitride. *ACS Appl. Mater. Interfaces* **12**, 25464–25470 (2020).

16. Bradac, C., Lim, S. F., Chang, H.-C. & Aharonovich, I. Optical Nanoscale Thermometry: From Fundamental Mechanisms to Emerging Practical Applications. *Adv. Opt. Mater.* **8**, 2000183 (2020).
17. Brites, C. D. S. *et al.* Thermometry at the nanoscale. *Nanoscale* **4**, 4799–4829 (2012).
18. Jaque, D. & Vetrone, F. Luminescence nanothermometry. *Nanoscale* **4**, 4301–4326 (2012).
19. Quintanilla, M. & Liz-Marzán, L. M. Guiding Rules for Selecting a Nanothermometer. *Nano Today* **19**, 126–145 (2018).
20. Liang, Z., Wu, J., Cui, Y., Sun, H. & Ning, C.-Z. Self-optimized single-nanowire photoluminescence thermometry. *Light Sci. Appl.* **12**, 36 (2023).
21. Labrador-Páez, L. *et al.* Reliability of rare-earth-doped infrared luminescent nanothermometers. *Nanoscale* **10**, 22319–22328 (2018).
22. Choi, S., Agafonov, V. N., Davydov, V. A. & Plakhotnik, T. Ultrasensitive All-Optical Thermometry Using Nanodiamonds with a High Concentration of Silicon-Vacancy Centers and Multiparametric Data Analysis. *ACS Photonics* **6**, 1387–1392 (2019).
23. Chen, Y. *et al.* Real-Time Ratiometric Optical Nanoscale Thermometry. *ACS Nano* (2023) doi:10.1021/acsnano.2c10974.
24. Fan, J.-W. *et al.* Germanium-Vacancy Color Center in Diamond as a Temperature Sensor. *ACS Photonics* **5**, 765–770 (2018).
25. Nguyen, C. T. *et al.* All-optical nanoscale thermometry with silicon-vacancy centers in diamond. *Appl. Phys. Lett.* **112**, 203102 (2018).
26. Brooks, R. A. Elephants don't play chess. *Des. Auton. Agents* **6**, 3–15 (1990).
27. Murphy, K. P. *Machine learning: a probabilistic perspective*. (MIT press, 2012).
28. Neu, E. *et al.* Low-temperature investigations of single silicon vacancy colour centres in diamond. *New J. Phys.* **15**, 043005 (2013).
29. Alkahtani, M. *et al.* Tin-vacancy in diamonds for luminescent thermometry. *Appl. Phys. Lett.* **112**, 241902 (2018).
30. Tran, T. T. *et al.* Anti-Stokes excitation of solid-state quantum emitters for nanoscale thermometry. *Sci. Adv.* **5**, eaav9180.
31. Kondrina, K. M., Kudryavtsev, O. S., Vlasov, I. I., Khmelnitskiy, R. A. & Ekimov, E. A. High-pressure synthesis of microdiamonds from polyethylene terephthalate. *Diam. Relat. Mater.* **83**, 190–195 (2018).