Supporting Information

Whispering Gallery Mode Resonators in Continuous Flow: Spectral Assignments and Sensing with Monodisperse Microspheres

Bryan C. Paulus, Jenny K. Banh, Kirk, D. Rector, Benjamin W. Stein, Laura M. Lilley*

Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

Email: llilley@lanl.gov

Anal. Methods

Table of Contents

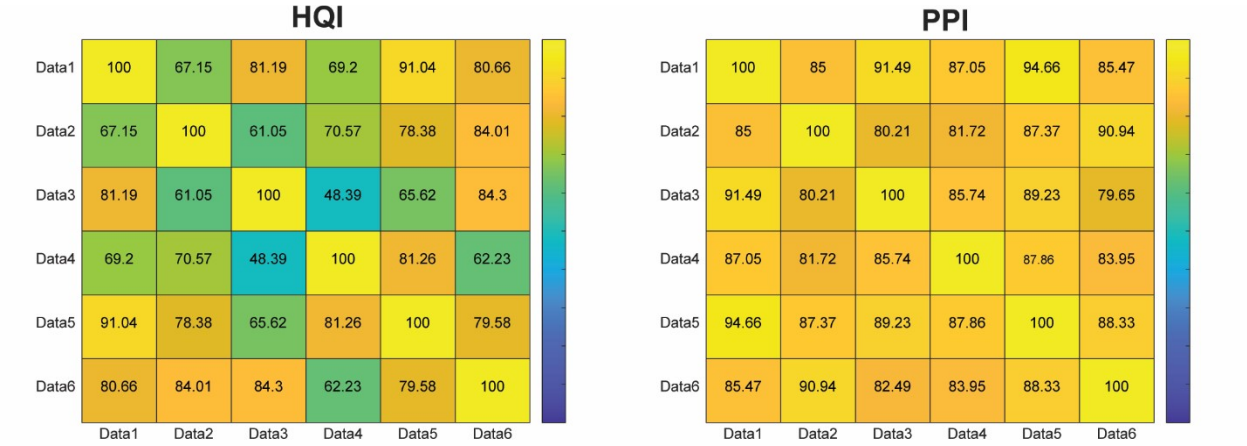

Fig. S1 Comparison of HQI and PPI results on six spectra collected from the same ~8 μm polystyrene microsphere in air. The correlation based HQI is sensitive to the intensity differences found in the six spectra, and yields several low matching indices. The peak location-based PPI, however yields high matching rates for each dataset.

Description of spectral similarity metrics

When analyzing large quantities of WGM spectra it can be helpful to have a means of discriminating between different spectra. Since a sphere can linger in the illuminated region of the capillary tube used for flowing microsphere suspensions it can sometimes contribute to multiple collected spectra. We used two metrics for assessing the similarity of baseline corrected spectra so that only a single dataset pertaining to a given microsphere were saved and kept for spectral fitting. The first is an area matching/ correlation-based approach called the hit quality index (HQI), calculated by taking the square of the dot product of two spectra, ***A*** and ***B***, as seen in equation S1.

$$HQI = 100\% * (A \cdot B)^2 \qquad \text{S1}$$

Since the collected spectra do not necessarily cover the same spectral window, duplicate spectra are generated for the purpose of comparison making and the endpoints of each duplicated spectrum are cut to ensure they cover the same range of wavelengths prior to application of S1. The HQI ranges from 0 to 100 % with 100 % representing a perfect match between the two spectra. The HQI is an attractive metric due to its computational simplicity and efficiency, however it is not always the best suited for comparing all WGM spectra due to its low sensitivity towards discriminating between spectra containing similarly located peaks.

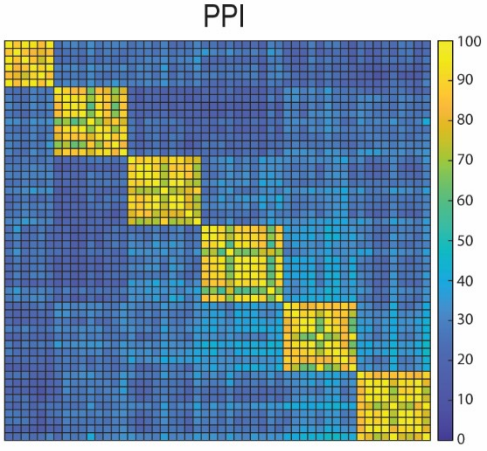

Fig. S2 PPI results for multiples of spectra collected on six different ~8 μm polystyrene microspheres in air. The high matching indices for spectra originating from the same sphere and low matching indices for spectra from different spheres illustrate how the PPI can adequately match spectra originating from the same sphere and discriminate between spectra from different microspheres.

For this, we also apply an index that utilizes peak locations to assess spectral similarity. First, peaks in a given spectrum are located using a peak finding algorithm developed by Yoder.[1] Next, peaks of one spectrum are uniquely matched to the peaks of another spectrum. To be considered a match, peaks of both spectra must be within 0.2 nm of each other and if there are multiple peaks in one spectrum that satisfy this condition with a peak in another spectrum, only the peak that is closest may be deemed a matching peak. For spectra ***A*** and ***B***, with total number of peaks $|\boldsymbol{A}|$ and $|\boldsymbol{B}|$ respectively, let *L* be the number of uniquely matched peaks. We define the peak proximity index (PPI) according to equation S2 where $\lambda_{\boldsymbol{S},l}$ denotes the *l*th peak of spectrum ***S***.

$$PPI = 100\% * \left(\frac{\left(\sum \lambda_{A,l} * \lambda_{B,l} \right)^2}{\sum \lambda_{A,l}^2 * \sum \lambda_{B,l}^2} - \left(1 - \frac{L}{2}\left(\frac{1}{|A|} + \frac{1}{|B|} \right) \right) \right) \quad \text{S2}$$

The first term in equation S2 is a normalized dot product of all the matched peak locations. The second term represents the average number of unmatched peaks between the two spectra being compared. This term acts as a penalty for having unmatched peaks in one or both spectra and is similar to a term found in the mass spectrum matching algorithm found in the work of Hansen and Smedsgaard.[2] Similar to HQI, PPI also has a range of 0-100 % with 100 % representing a perfect match. While both metrics are useful, PPI tends to be better at both discriminating between spectra of different spheres and showing that spectra collected from the same microsphere are similar. Both of these concepts are illustrated in Fig. S1 and S2.

Assessment of clustering performance

The performance of the shared nearest neighbor spectral clustering algorithm (SC-nSNN) was evaluated using the normalized mutual information (NMI)[3] metric which has a range of [0,100 %] where an NMI of 100 % denotes perfect clustering performance (*i.e.* when all data points are assigned to the correct clusters). NMI is defined in equation S3 where *I(X,Y)* is the mutual information

$$NMI(X,Y) = 100\% * \frac{2I(X,Y)}{H(X) + H(Y)} \quad \text{S3}$$

between two random variables *X* and *Y*, and *H*(*X*) and *H*(*Y*) are the entropy of *X* and *Y*. Let $C = \{c_1,c_2,\ldots,c_K\}$ be the clusters obtained from the SC-nSNN algorithm and let $C' = \{c'_1,c'_2,\ldots,c'_K\}$ be the true clustering of the data. The mutual information and entropy are calculated using equations S4-S7,

$$I(C,C') = \sum_{i=1}^{K}\sum_{j=1}^{K} P(c_i \cap c'_j) log \frac{P(c_i \cap c'_j)}{P(C_i)P(c'_j)} \quad \text{S4}$$

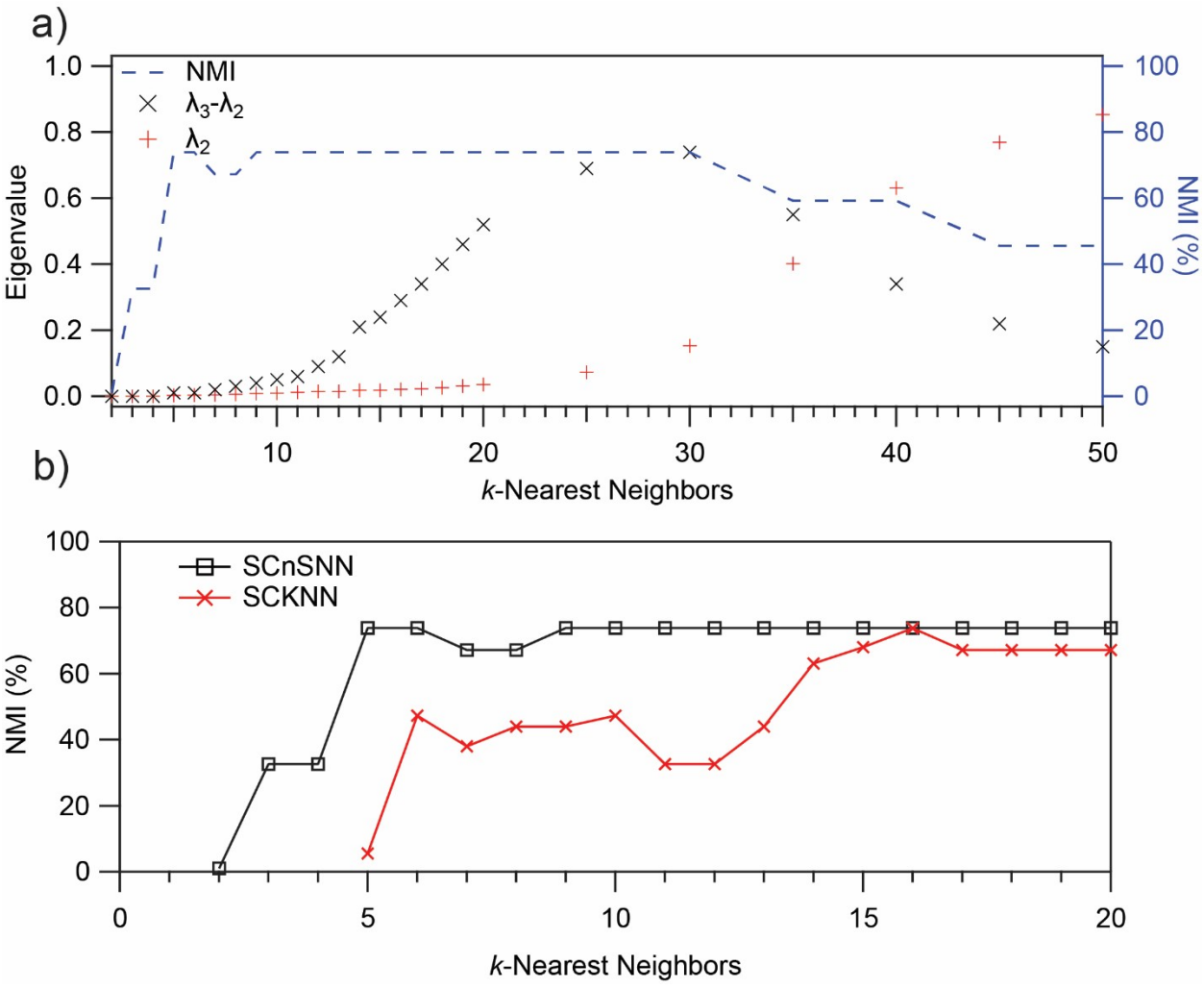

Fig. S3 (a) Plot of the $\lambda_3 - \lambda_2$ eigengap and λ_2 eigenvalue (left axis) and NMI (right axis) for a range of *k*-nearest neighbors used to generate the similarity graph and Laplacian matrix. (b) Comparison of the clustering performance of the number of shared nearest neighbors spectral clustering algorithm (SC-nSNN) and a more traditional undirected *k*-nearest neighbors spectral clustering (SC-KNN) algorithm.

$$H(C) = -\sum_{i=1}^{K} P(c_i) log P(c_i) \quad \text{S5}$$

$$P(c_i \cap c_j) = |c_i \cap c'_j|/n \quad \text{S6}$$

$$P(c_i) = |c_i|/n \quad \text{S7}$$

where $|\cdot|$ is the cardinality of the cluster, and n is the total number of data points. $P(c_i)$ represents the probability that the data points belong to cluster c_i and likewise, $P(c_i \cap c'_j)$ represents the probability that the data points belong to the intersection of clusters c_i and c'_j.

As outlined in the main text, the eigengap heuristic[4] was leveraged to identify how many clusters are identifiable by the SC-nSNN algorithm. In the eigengap heuristic, the goal is to choose the number of clusters, K, such that eigenvalues $\lambda_1,\ldots,\lambda_K$ are close to zero but λ_{K+1} is relatively large. The magnitude of these eigenvalues, and by extension the clustering performance, is dependent on the number of nearest neighbors, k, searched for to generate the similarity graph and subsequently the graph Laplacian matrix from which eigenvalues are calculated. Thus, k must be tuned to optimize the clustering performance of the SC-nSNN algorithm. Fig. S3a shows the λ_3 - λ_2 eigengap and λ_2 eigenvalue alongside the NMI for a range of k-nearest neighbors. When k is small (*i.e.,* $k < 20$), λ_2 is close to zero. As k is increased further, the intercluster connectivity in the similarity graph increases and the visibility of the clusters is diminished, manifesting in larger λ_2 eigenvalues. Similarly, while k is increased over the range of [2, 20] the $\lambda_3 - \lambda_2$ eigengap grows, indicating the presence of two well defined clusters. Eventually, as k is further increased, this eigengap begins to narrow suggesting the cluster visibility becomes lower at large k. The NMI calculated over these ranges also shows sustained, high clustering performance for $5 < k < 30$ which track well with the eigengap heuristic.

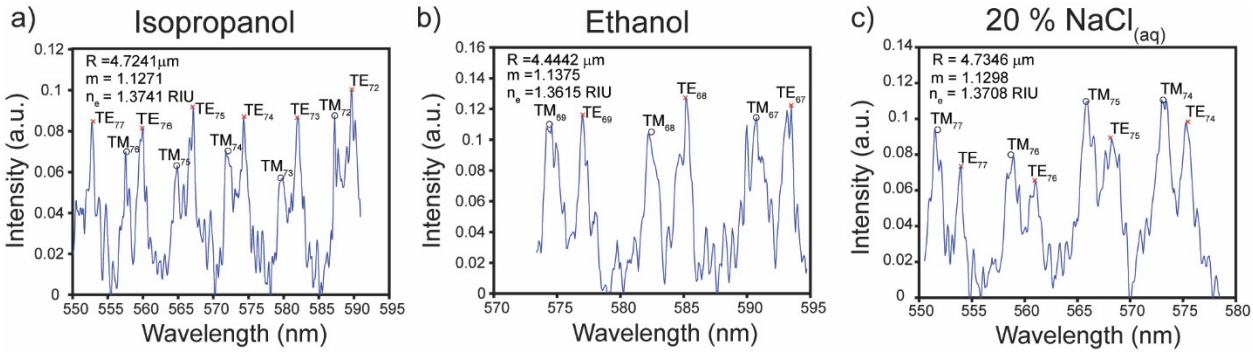

Fig. S4 Examples of noisy data sets acquired in a) isopropanol, b) ethanol, and c) 20 % $NaCl_{(aq)}$ solution which have been fit to acquire useful refractive index and microsphere size information. Each spectrum represents only a single data point and may differ from the values listed in Table 1 in the main text.

Spectral fits on low quality data

Sometimes the data collected by the methods outlined in this text are not of very high quality, *i.e.*, it can be noisy, of such low intensity that only a few modes are observable, or can have multiple microspheres contributing to the overall spectrum. While any of these scenarios might cause problems for correlation based fitting methods, the peak location-based fit program used here can largely ignore such imperfections in the data insofar as the peak locations arising from a single microsphere can accurately be determined from the spectrum. Fig. S4 shows several poor-quality datasets from which useful refractive index and size information was extracted.

Figures of merit

Equation 4 was used to estimate the sensitivity of a WGM spectrum towards the size of the resonator and the refractive index of its surroundings for a typical microspherical resonator used in this study. The simulated resonance position of a first order $l = 60$ TE mode of a nominally 4 μm radius sphere submerged in a 1.3326 RIU fluid (water) was tracked as a function of small changes to the radius and external refractive index. Over the small range of values tested, the response to changes in the radius or

refractive index (Fig. S5) was approximately linear with slopes of 58.47 ± nm/RIU and 150.30 ± 0.03 nm/μm respectively. Given the above sensitivities and the approximate 0.081 nm/pixel pitch of the spectrometer used in this study, the minimum discernable changes in the radius and external refractive index are approximately 0.5 nm and 1.4 × 10^{-3} RIU respectively. The effect of temperature on the WGM spectrum is more complicated as it has an impact on both the refractive indices of the submersion fluid and the sphere material as well as the size of the microsphere. The temperature dependence of the microsphere radius was first estimated using the volumetric thermal expansion coefficient, α_V, as seen in equation S8

$$\alpha_V = \frac{1\delta V}{V\delta T} \qquad \text{S8}$$

Where $V = 4pa^3/3$ is the volume of the microsphere of radius a and δV is the volume change when heated by δT. The linear thermal expansion coefficient for polystyrene is 70 × 10^{-6}/°C.[5] Next, the thermo-optic constants $-dn/dT$ of water (15.82 × 10^{-5} RIU/°C)[6] and polystyrene (17.336 × 10^{-5} RIU/°C)[7] were used to estimate the change in the refractive indices of both substances. Subsequent use of equation 4 afforded the predicted resonance positions of the WGM of interest. The net effect of temperature was approximately linear over the temperature range simulated with a sensitivity of 1.35 × 10^{-2} nm/°C. Thus, assuming the minimum detectable change in the resonance position is equal to the spectrometer pitch, the minimum detectable change in temperature was 6 °C.

References

1 N. Yoder, Peakfinder(X0, Sel, Thresh, Extrema, Include Endpoints, Interpolate). MATLAB Central File Exchange 2021.

2 M. E. Hansen and J. Smedsgaard, *J. Am. Soc. Mass Spectrom.*, 2004, **15**, 1173–1180.

3 X. Ye and T. Sakurai, *ETRI J.*, 2016, **38**, 540–550.

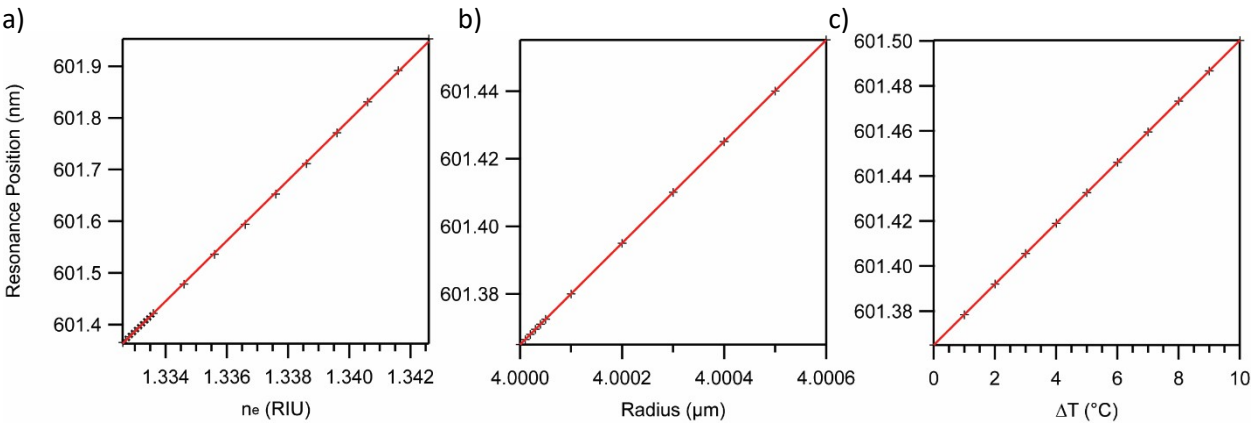

Fig. S5 Simulated response of a first order l = 60 TE whispering gallery mode of a nominally 4 μm radius polystyrene microsphere submerged in water as a function of changes to (a) the external refractive index, (b) the radius of the microsphere, and (c) the temperature of the solution.

4 U. Von Luxburg, *Stat. Comput.*, 2007, **17**, 395–416.

5 P. S. Turner, *J. Res. Natl. Bur. Stand.*, 1946, **37**, 239-250.

6 H. El-Kashef, *Physica B*, 2000, **279**, 295-301.