



HAL
open science

Regressing bubble cluster dynamics as a disordered many-body system

Kazuki Maeda, Daniel Fuster

► **To cite this version:**

Kazuki Maeda, Daniel Fuster. Regressing bubble cluster dynamics as a disordered many-body system. Journal of Fluid Mechanics, 2024, 985, pp.A23. 10.1017/jfm.2024.313 . hal-04558779

HAL Id: hal-04558779

<https://hal.science/hal-04558779>

Submitted on 25 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Banner appropriate to article type will appear here in typeset article

Regressing bubble cluster dynamics as a disordered many-body system

Kazuki Maeda¹†, and Daniel Fuster²

¹School of Aeronautics and Astronautics, Purdue University, West Lafayette 47907, USA

²Sorbonne Université, CNRS, UMR 7190, Institut Jean Le Rond D'Alembert, F-75005 Paris, France

(Received xx; revised xx; accepted xx)

The coherent dynamics of bubble clusters are of fundamental and industrial importance and are elusive due to the complex interactions of disordered bubble oscillations. Here we introduce and demonstrate a method for decomposition of the Lagrangian time series of bubble dynamics data by combining theory and principal component analysis. The decomposition extracts coherent features of bubble oscillations based on their energy, in a way similar to Proper Orthogonal Decomposition of Eulerian flow field data. This method is applied to a data set of spherical clusters under harmonic excitation at different amplitudes, with various nuclei density and polydispersity parameters. Results indicate that the underlying correlated mode of oscillations is isolated in a single dominant feature in cavitating regimes, independent of nuclei's parameters. A systematic data analysis procedure further suggests that this feature is globally controlled by the dynamic cloud interaction parameter of Maeda and Colonius (*J. Fluid Mech.*, vol. 862, 2019, pp. 1105–1134) that quantifies the mean-field interactions, regardless of initial polydispersity or nonlinearity. The method provides a simplified and comprehensive representation of complex bubble dynamics as well as a new path to reduced-order modeling of cavitation and nucleation.

Key words: Bubble dynamics, Cavitation

MSC Codes (*Optional*) Please enter your MSC Codes here

1. Introduction

Cavitation bubble clusters nucleate when the liquid pressure rapidly falls below a certain threshold. These clusters coherently oscillate and violently collapse to cause extreme energy concentration that leads to various critical consequences and use in applications as diverse as injectors and pumps (e.g., Plesset & Ellis 1955; Blander & Katz 1975; Mørch 1980; Chen & Heister 1994; Hashimoto *et al.* 1997; Prosperetti 2017), hydraulic machines (Arndt 1981), underwater propulsion and hydrofoils (e.g., Kubota *et al.* 1992; Merkle *et al.* 1998; Kunz *et al.* 2000; Schnerr & Sauer 2001; Gnanaskandan & Mahesh 2016; Ganesh *et al.*

† Email address for correspondence: kmaeda@purdue.edu

2016; Venning *et al.* 2022), medical ultrasound (e.g., Ikeda *et al.* 2006; Pishchalnikov *et al.* 2011; Maxwell *et al.* 2011; Maeda *et al.* 2015; Movahed *et al.* 2016), surface cleaning (Verhaagen & Rivas 2016; Yamashita & Ando 2019), chemical synthesis (Suslick *et al.* 1999; Cairós & Mettin 2017), and bio-inspired devices (Tang & Staack 2019). Characterizing the dynamics is a challenge due to the complex interactions of bubbles involving disorders and stochasticity. Nuclei are typically micro-sized and polydisperse and randomly distributed. Their rapid, nonlinear oscillations are, except in controlled experiments (Bremond *et al.* 2006), practically not measurable. Molecular and hydrodynamics simulations can provide detailed insights into nucleation (Angélil *et al.* 2014; Gallo *et al.* 2021), while their time- and spatial scales have not reached those of practical cluster oscillations. Analyses have been made on the interaction dynamics in various regimes (Brennen 2014), yet no common knowledge has been established if the many-body coherence globally exists and if so scaling is possible, beyond the consensus that polydispersity induces strong disorders.

For the past decades, the Rayleigh-Plesset (R-P) equation and its variations have been actively explored to investigate the dynamics of single bubbles (Plesset 1949; Plesset & Prosperetti 1977). Relatively few studies addressed the theory of clusters. By using mean field approach to interacting bubbles modeled by the R-P equation, d'Agostino and Brennen (d'Agostino & Brennen 1989) derived a nondimensional parameter that dictates the linear coherent oscillations of monodisperse clusters, the so-called "cloud interaction parameter". Zeravcic *et al.* (2011) used the coupled R-P equations and identified disorders represented by the Anderson localization of acoustic energy in polydisperse, lattice-like clouds under weak excitation. We have recently extended the interaction parameter to the non-equilibrium, cavitating clusters under strong excitation by considering the effective interaction at excited states (Maeda & Colonius 2019; Maeda *et al.* 2018; Maeda & Maxwell 2021). To recall, we scale the mean kinetic energy of liquid induced by $N(\gg 1)$ bubbles as (Maeda & Colonius 2019)

$$\langle K \rangle \sim \langle K_s \rangle (1 + B_d), \quad (1.1)$$

where K_s is the energy of a single bubble: $K_s = 2\pi\rho R_{b,c}^3 \dot{R}_{b,c}^2$. B_d is the parameter controlling the effective contribution of hydrodynamic inter-bubble interaction: $B_d = N\overline{\langle R(t) \rangle} / R_C$, and $R_{b,c}$ and R_C denote the characteristic (reference) bubble radius and the cluster radius. ρ_l is the liquid density. $\langle \cdot \rangle$ and $\overline{(\cdot)}$ denote time average during a period in which bubble dynamics are statistically stationary and the mean value about the bubbles ($i = 1, 2, \dots, N$), respectively. The scaling can be simply derived from the coupled R-P equation for the correlated (synchronized) limit of monodisperse bubbles ($R_1 = R_2 \dots = R_{b,c}$). Although realistic correlations are imperfect due to polydispersity and nonlinearity, B_d was found to control well both the coherent dynamics of polydisperse cavitating clusters and their acoustic emission in numerical simulations and experiments (Maeda & Colonius 2019; Maeda & Maxwell 2021). Overall, previous studies indicate that the coherence can depend on both polydispersity and nonlinearity in a non-separable manner, posing perplexing questions about the universality of scaling. The theoretical characterization of the nonlinear dynamics of disordered many-body systems is in general not a simple task. Meanwhile, greater computing power has enabled learning physics by analysing big data. Principal component analysis (PCA) is a powerful method for unsupervised learning which has seen recent success in characterizing the coherent physics of many-body and high-dimensional systems in fields ranging from quantum information to fluid dynamics (Lloyd *et al.* 2014; Holmes 2012; Taira *et al.* 2017; Milano & Koumoutsakos 2002).

In this study, we introduce and demonstrate a method for unsupervised data decomposition to study the coherent bubble cluster dynamics by combining theory and PCA. PCA extracts

80 dominant states and dynamical features, such as coherent quantum states and turbulent
 81 structures, and their amplitudes as the eigenfunction (feature) and the eigenvalue (variance)
 82 of the co-variance matrix of physical data. When applied to spatio-temporal data of dynamical
 83 systems, PCA is often denoted as proper-orthogonal decomposition (POD). Those data can
 84 be properly weighted prior to PCA such that the variance becomes consistent with the
 85 norm induced by an energetic inner product of state variables (e.g., kinetic energy) (Lall
 86 *et al.* 1999; Rowley 2005). This weighting allows a physical interpretation that resulting
 87 features associated with a large variance are energetically dominant coherent structures.
 88 Proper weighting of Lagrangian bubble dynamics data is non-trivial since the linear variance
 89 of extracted features need to account for the nonlinear interaction energy. For meaningful
 90 analysis, we introduce strategic pre-processing of the data prior to PCA such that the PC-
 91 variance becomes theoretically consistent with the energy modeled by the coupled R-P
 92 equation. Analysing simulation data sets of clusters, we show that the PCA can systematically
 93 extract not only coherent but also incoherent features whose magnitudes are respectively
 94 measured by the PC-variance and the entropy. We discover that the coherence is lost by
 95 disorders induced by polydispersity and nonlinearity, while under strong excitation the
 96 underlying correlations are globally isolated in a single coherent feature whose variance
 97 (energy) is scaled by B_d , regardless of the disorders.

98 The remainder of this paper is as follows. In §2, we describe the method. The first PC
 99 variance, spectral entropy, and the coherence measure are introduced as quantifiable measures
 100 to characterize the coherent bubble dynamics from extracted features. In §3, we verify and
 101 demonstrate the method using a numerical dataset of bubble clouds with different density and
 102 polydispersity parameters under various amplitudes of harmonic excitation. The deviation
 103 of the PC-variance from the physical energy is quantified for two weighting methods. The
 104 amplitude dependencies of the measures are quantified. The PC-spectra and their correlations
 105 with the coherent energy are analyzed. Moreover, the extracted coherent dynamics is related
 106 to B_d and its universality is discussed for cavitating clouds. In §4, the physical significance
 107 of the method is discussed. In §5, we state conclusions.

108 2. Methods

109 2.1. Principal component analysis of bubble dynamics data

110 For clusters modeled by the coupled R-P equations, observable dynamical variables are
 111 bubbles' radial velocities and radii. Consider a data matrix \mathbf{Q} containing the N_t snapshots of
 112 the radial velocities with a constant temporal interval:

$$113 \quad \mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{N_t}], \quad (2.1)$$

114 where \mathbf{q}_k denotes the vector containing the radial velocities of the N bubbles at time t_k (in
 115 the k -th snapshot):

$$116 \quad \mathbf{q}_k = [\dot{R}_1(t_k), \dot{R}_2(t_k), \dots, \dot{R}_N(t_k)]^T. \quad (2.2)$$

117 For later convenience, we also define the vector \mathbf{r}_k , containing the radii of the bubbles at the
 118 same instances:

$$119 \quad \mathbf{r}_k = [R_1(t_k), R_2(t_k), \dots, R_N(t_k)]^T. \quad (2.3)$$

120 PCA can be performed on \mathbf{Q} by using the singular value decomposition (SVD) (e.g., Abdi
 121 & Williams 2010; Jolliffe & Cadima 2016):

$$122 \quad \mathbf{Q} = \mathbf{U}\mathbf{\Sigma}\mathbf{X}^*. \quad (2.4)$$

123 The i -th principal component (feature) is stored in the i -th score matrix:

$$124 \quad \mathbf{\Pi}_i = \mathbf{U}\mathbf{\Sigma}_i, \quad (2.5)$$

125 where $\mathbf{\Sigma}_i$ contains the i -th largest singular value and zeros elsewhere.

126 2.2. Weighted principal component analysis

127 Although this procedure for PCA is simple and straightforward, the physical meaning of the
 128 extracted features are obscure since PCA itself is not informed on the underlying dynamics
 129 of the system. A related issue of PCA for fluid flow data has been addressed in the context of
 130 POD. In POD of the snapshots of Eulerian fluid flow data, the state vectors consisting the data
 131 matrix are often weighted such that the corresponding weighted inner product of the state
 132 becomes consistent with the kinetic energy of the original system. The dominant features
 133 (POD modes) can then be interpreted as energetically dominant coherent flow structures.
 134 Moreover, in our previous studies (Maeda & Colonius 2019; Maeda & Maxwell 2021), the
 135 scaling of bubble cloud dynamics by B was successfully demonstrated based on the total
 136 kinetic energy of liquid induced by interacting bubbles. These considerations motivate us to
 137 relate the energy with PCA for addressing the physics of bubble cloud.

138 Inspired by the POD, we consider weighting the present bubble dynamics data prior to
 139 PCA. To find an appropriate weight, we revisit the potential theory behind eq. (1.1). The
 140 kinetic energy of the fluid induced by the spherical bubble oscillations is explicitly expressed
 141 as

$$142 \quad K = 2\pi\rho_l \sum_{i=1}^N \left[R_i^3 \dot{R}_i^2 + \sum_{j \neq i}^N \frac{R_i^2 R_j^2 \dot{R}_i \dot{R}_j}{r_{ij}} \right] + (H.O.T), \quad (2.6)$$

143 where R_i , \dot{R}_i , and r_{ij} are the radius and the radial velocity of bubble i , and distance between
 144 the centers of bubble i and j , respectively. The second term in the bracket represents the
 145 contribution of the long-range interactions. At each instant, K can be expressed as a weighted
 146 inner product of \mathbf{q} :

$$147 \quad K = \mathbf{q}^T \mathbf{T} \mathbf{q} = (\mathbf{W} \mathbf{q})^T (\mathbf{W} \mathbf{q}). \quad (2.7)$$

148 where

$$149 \quad T_{ij}(\mathbf{r}_k) = \begin{cases} 2\pi\rho R_i^3(t_k) & (i = j), \\ 2\pi\rho \frac{R_i^2(t_k) R_j^2(t_k)}{r_{ij}} & (i \neq j) \end{cases} \quad (2.8)$$

150 The weight matrix, \mathbf{W} , can be obtained through the Cholesky factorization of $\mathbf{T}(\mathbf{r})$:

$$151 \quad \mathbf{T}(\mathbf{r}) = \mathbf{W} \mathbf{W}^*. \quad (2.9)$$

152 This expression suggests the weighted data, $\mathbf{W} \mathbf{Q}$, is appropriate for used in PCA. However,
 153 $\mathbf{T}(\mathbf{r})$ and \mathbf{W} are time dependent since \mathbf{r} can change in time during large-amplitude
 154 oscillations. The choice of \mathbf{r}_k in the time series to define \mathbf{W} is unclear. A straightforward
 155 choice is to use the temporal mean of the radius for each bubble, $\langle \mathbf{r} \rangle$: $\langle \mathbf{r} \rangle = \sum_{k=1}^{N_t} \mathbf{r}_k / N_t$,
 156 but $T(\langle \mathbf{r} \rangle)$ is not a favorable approximation for $T(\mathbf{r})$ for bubbles with large-amplitude
 157 oscillations.

158 We address this obstacle by variable transformation as a means of pre-processing the data.
 159 The schematic is shown in figure 1. First, we transform the variables from (\mathbf{q}, \mathbf{r}) to $(\boldsymbol{\xi}, \boldsymbol{\eta})$,
 160 where $\boldsymbol{\xi} = [\xi_1, \xi_2, \dots, \xi_N]^T$ and $\boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_N]^T$, and $(\xi_i, \eta_i) = (q_i r_i, r_i) = (\dot{R}_i R_i, R_i)$
 161 for $i \in [1, N]$. ξ_i is nothing but the velocity potential evaluated at the surface of bubble i .
 162 Supplemental discussions and justifications for this transformation are provided in Appendix

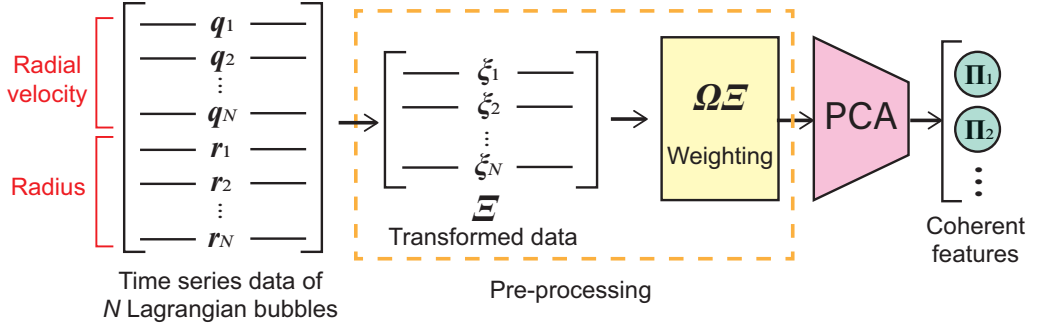


Figure 1: Schematic of the feature extraction by PCA from Lagrangian bubble dynamics data, \mathbf{Q} , after pre-processing. The variance of the resulting features is consistent with the interaction energy predicted by the coupled Rayleigh-Plesset equation.

163 A. The instantaneous energy of the system is expressed by a weighted inner product as

$$164 \quad K = \xi^T \mathbf{P}(\eta) \xi = (\Omega \eta)^T (\Omega \eta). \quad (2.10)$$

165 Ω is the weight matrix satisfying

$$166 \quad \mathbf{P}(\mathbf{r}) = \Omega \Omega^*, \quad (2.11)$$

167 where

$$168 \quad P_{ij}(\eta) = \begin{cases} 2\pi\rho\eta_i & i = j, \\ 2\pi\rho\frac{\eta_i\eta_j}{r_{ij}} & i \neq j. \end{cases} \quad (2.12)$$

169 Second, we partially replace η with $\langle \eta \rangle$ to approximate the system. Using $(\hat{\xi}, \hat{\eta})$, the
170 instantaneous energy of the approximate system is expressed as

$$171 \quad \hat{K} = \hat{\xi}^T \mathbf{P}(\hat{\eta}) \hat{\xi}. \quad (2.13)$$

172 The temporal mean of the energy of the original system can then be approximated as

$$173 \quad \langle K \rangle \approx \langle \hat{K} \rangle = \langle \hat{\xi}^T \mathbf{P}(\hat{\eta}) \hat{\xi} \rangle \approx \langle \xi^T \mathbf{P}(\langle \eta \rangle) \xi \rangle. \quad (2.14)$$

174 This is a critical result in the present context of PCA, since the energy is now related to the
175 weighted inner product with the constant weight, $\mathbf{P}(\langle \eta \rangle)$. Quantitative verification of this
176 approximation is addressed through numerical experiments in the following section. Using
177 the new set of variables, the PC decomposition is performed as

$$178 \quad \Omega \Xi = \mathbf{U}_\xi \Sigma_\xi \mathbf{X}_\xi^*. \quad (2.15)$$

179 The i -th PC is stored in $\Pi_{\xi,i}$: $\Pi_{\xi,i} = \mathbf{U}_\xi \Sigma_{\xi,i}$. We denote the i -th largest singular value of
180 Σ_ξ as σ_i . The degree of coherence can be measured by the normalized variance of the first
181 PC:

$$182 \quad \hat{\sigma}_1^2 = \sigma_1^2 / \text{tr}(\Sigma^2), \quad (2.16)$$

183 which represents the ratio of the energy occupied by the first PC to the total energy of
184 the system. In the following sections, to distinguish the present method from the standard
185 PCA, we denote (2.15) as the weighted principal component analysis of transformed data
186 (WPCA-TD).

187 2.3. Spectral entropy and coherence measure

188 We introduce two additional key measures to characterize the coherent physics of bubble
189 clusters through WPCA-TD. First, to quantify the degree of incoherent bubble oscillations,

190 we define the spectral entropy of the weighted co-variance matrix, $\mathbf{\Omega}\mathbf{\Xi}$, as

$$191 \hat{S}_{vN} = - \sum \frac{\hat{\sigma}_k^2 (\ln \hat{\sigma}_k^2)}{\ln(N)}, \quad (2.17)$$

193 where $\ln_2(N)$ is the normalization factor. The spectral entropy of data is a discrete analogue
 194 of the Shannon entropy based on the spectrum of data and has been used to quantify the
 195 randomness of data in various applications (Kullback 1997; Alter *et al.* 2000; Hu *et al.* 2005;
 196 De Domenico & Biamonte 2016). Aubry *et al.* (1991) introduced a similar definition of
 197 entropy for spatio-temporal signals of canonical fluid flows from their POD eigenvalues and
 198 used the entropy to characterize flow instabilities. In the present context, when bubbles are
 199 in perfect correlation, we expect to excite only the first PC capturing the entire energy of the
 200 system: $\hat{\sigma}_1^2 = 1$ and $\hat{\sigma}_k^2 = 0$ for $k : k \in [2, N]$, yielding $\hat{S}_{vN} = 0$. In contrast, if the energy
 201 is equi-partitioned into all PCs, $\hat{\sigma}_k^2 = 1/N$ for all $k : k \in [1, N]$ and $\hat{S}_{vN} = 1$. Second, we
 202 define the coherence measure C .

$$203 C = \frac{B_1}{B_d}, \quad (2.18)$$

204 where

$$205 B_1 = \frac{\sigma_1^2}{\sigma_1'^2} - 1. \quad (2.19)$$

207 $\sigma_1'^2$ is the first PC-variance excluding the contribution of interactions, obtained from the
 208 WPCA-TD of the same data using a diagonal weight matrix $\mathbf{\Omega}'$ whose diagonal entries
 209 are those of $\mathbf{\Omega}$. Further details of $\mathbf{\Omega}'$ are provided in Appendix B. In the limit of perfect
 210 correlation (also see eq. (1.1)),

$$211 \langle K \rangle \approx \sigma_1^2 \approx (1 + B_d)\sigma_1'^2 \quad (2.20)$$

212 and

$$213 B_d \approx B_1. \quad (2.21)$$

214 Therefore $C \approx 1$. This condition is typically realized in monodisperse clusters under weak
 215 (linear) oscillations (d'Agostino & Brennen 1989). For real clusters under strong excitation,
 216 bubbles are not perfectly correlated and the energy can be distributed in broad features. In
 217 this regime, approximation (2.20) is not necessarily expected to hold and C can take values
 218 far from unity. Phenomenologically speaking, C quantifies from data the degree to which the
 219 most coherent mode of oscillation is represented by the cluster's mean-field interaction. A
 220 schematic of the process to obtain C from data is provided in Appendix B.

221 3. Numerical experiments

222

3.1. Data sets

223 To verify and demonstrate the WPCA-TD, we use data sets of spherical clusters excited by
 224 40 cycles of harmonic pressure excitation. Each cluster contains $O(10 - 10^3)$ bubbles with
 225 their initial radii following log-normal distributions with a reference radius of $R_{ref} = O(10)$
 226 μm ; $\ln(R_0/R_{ref}) = N(0, s_d)$, where s_d is the lognormal standard deviation as the measure
 227 of polydispersity (Maeda & Colonius 2019). We address $s_d = [0.1, 0.3, 0.5, 0.7]$. In real
 228 bubble clouds, nuclei are expected to be polydisperse. $s_d = 0.7$ may be a representative
 229 estimation based on previous studies (Katz 1978; Ando *et al.* 2012; Maeda & Colonius
 230 2019), and is used unless noted. The smaller values of s_d are used in some cases to quantify
 231 the effect of polydispersity. The bubbles are randomly distributed in the spherical region

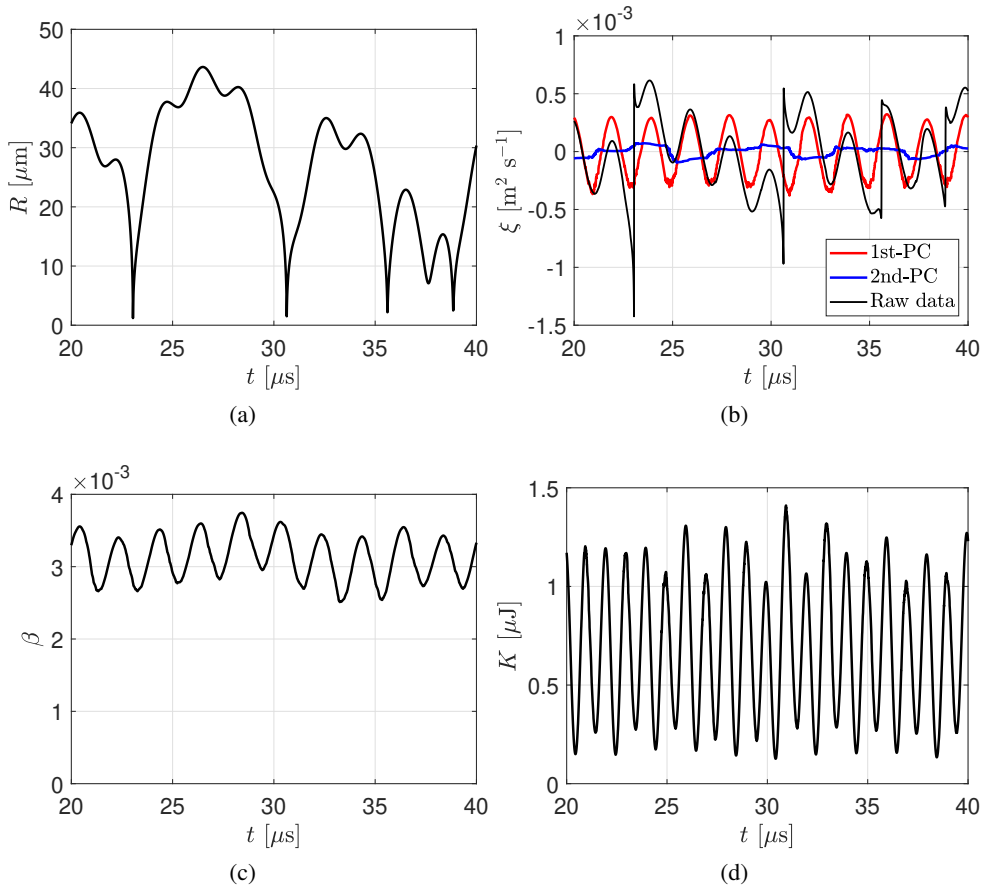


Figure 2: Evolution of representative quantities of a bubble cloud with $(B_0, s_d, N) = (0.5, 0.1, 100)$ excited at $A = 20$, during a stationary state. (a) Radius of a representative bubble. (b) For the same bubble, ξ obtained from the raw data, and those from the first and second dominant features extracted using WPCA-TD. (c) Void fraction. (d) Kinetic energy of the fluid induced by the bubble cloud.

232 with a specified cluster radius. Similar parameters of clusters were previously simulated
 233 to compare with experiments (Maeda & Colonius 2019; Maeda & Maxwell 2021). The density
 234 of bubbles is characterized by B_0 , the value of B_d at rest. The far-field pressure is given
 235 as $p_\infty(t) = p_0[1 + A\sin(2\pi ft)]$. The frequency of excitation is $f = 500$ kHz unless noted,
 236 near the adiabatic resonant frequency of the reference bubble. The amplitude of excitation is
 237 defined by A , relative to the ambient pressure at $p_0 = 1.0$ atm. For each set of parameter, we
 238 compute an ensemble average by taking a mean of the results from 20 bubble clouds with
 239 distinct spatial placements of bubbles in the clouds. For data generation, we use mesh-free,
 240 coupled Keller-Miksis equations by modifying previous methods. Details of this method are
 241 provided in Appendix D.

242

3.2. Visualization of representative data

243 In figure 2 we show evolution of representative quantities of a bubble cloud in the dataset
 244 with $(B_0, s_d, N) = (0.5, 0.1, 100)$ excited at $A = 20$, during a stationary . Figure 2 (a)
 245 shows the radius of a representative bubble. The plot presents familiar features of cavitation

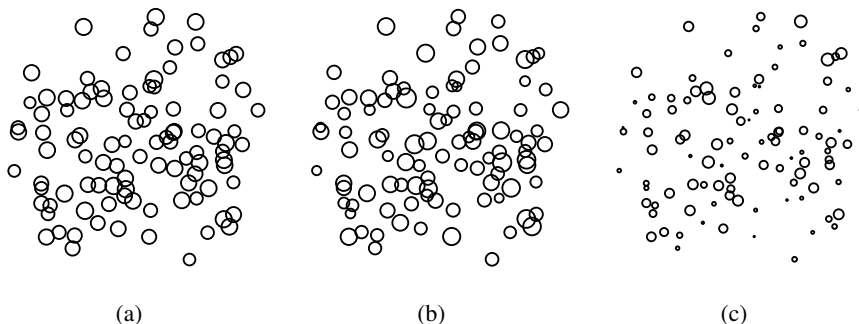


Figure 3: Projected side-views of the three-dimensional bubble cloud of figure 2. The size of the spheres denotes the root-mean-square amplitude of the velocity potential evaluated at the bubble surface at corresponding locations during the stationary state of oscillations, for (a) the raw data, (b) the first principal feature, and (c) the second principal feature.

246 bubbles forced by continuous, strong excitation, including fast events of collapse/rebound
 247 and slow growth/decay between them. During the slow phase, small-amplitude oscillations at
 248 excitation amplitude are evident. For the same bubble, figure 2(b) shows the evolution of the
 249 velocity potential obtained from the raw data, and those from the first and second dominant
 250 features extracted through the WPCA-TD. Compared to the radius, the raw data of the
 251 velocity potential looks much more symmetric about zero, even around the collapse events.
 252 The potential of the first feature presents a sinusoidal profile at the excitation frequency
 253 and with a constant amplitude which is close to the peak amplitude of the original data.
 254 The collapse events are not captured in this feature. The potential of the second feature
 255 is symmetric but has a much lower amplitude compared to the first feature. There is no
 256 clear similarity between the original data and the second feature, unlike that between the
 257 original data and the first-PC. Figure 2(c) shows the void fraction. The fraction oscillates
 258 with an amplitude of around 5.0×10^{-4} near 3.0×10^{-3} at the excitation frequency. Slow,
 259 small amplitude of variations are also observed. Figure 2(d) shows the kinetic energy of
 260 fluid induced by bubble oscillations. The energy oscillations around $0.75 \mu\text{J}$ at doubled
 261 the excitation frequency. Although the frequency is as expected, the peak amplitude largely
 262 fluctuates in the window as well as the waveform is not symmetric, unlike the void fraction.
 263 The fluctuation and asymmetry indicate can be associated with the incoherent oscillations to
 264 the kinetic energy. For instance, if one bubble is expanding and another bubble is collapsing
 265 out of phase, their net contribution may not appear in the void fraction due to mutual
 266 cancellation but can appear in the energy. Overall, the averaged quantities, void fraction and
 267 the energy, are much smoother than the individual bubble dynamics. This can be trivially
 268 explained by the incoherence of violent collapse events among bubbles and coherence of
 269 the linear response against fundamental frequencies. Meanwhile, the quantitative nature of
 270 coherent response is not predictable or easily analyzable from these plots due to strong
 271 non-linearity, especially under inter-bubble interactions with disordering factors including
 272 randomness of bubble position and polydispersity.

273 Figure 3(a-c) show the projected side-views of the three-dimensional bubble cloud of
 274 figure 2. The size of the spheres denotes the root-mean-square amplitude of the velocity
 275 potential evaluated at the bubble surface at corresponding locations during the stationary
 276 state of oscillations, for (a) the raw data, (b) the first principal feature, and (c) the second
 277 principal feature. The size of spheres can be interpreted as the mean energy of oscillations
 278 of bubbles at those locations in each feature. The overall distribution of the spheres of the

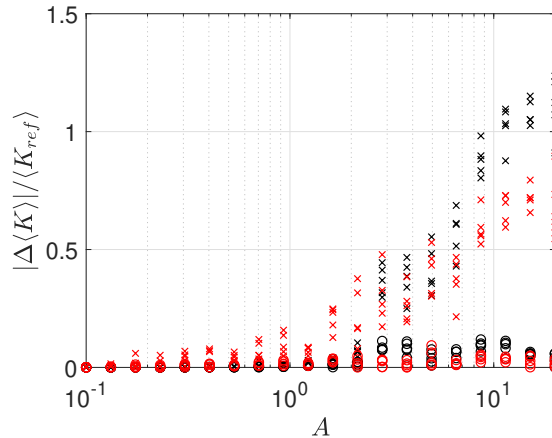


Figure 4: Relative error of the mean kinetic energy of the fluid induced by clusters for the two approximations, (×): $\langle K \rangle \approx \langle \mathbf{q}^T \mathbf{T}(\langle \mathbf{q} \rangle) \mathbf{q} \rangle$ and (○): $\langle K \rangle \approx \langle \xi^T \mathbf{P}(\langle \eta \rangle) \xi \rangle$, against the excitation pressure amplitude with various polydispersities and values of B_0 . Black: $(s_d, B_0) = (0.1, 0.5)$ and red: $(s_d, B_0) = (0.7, 5.0)$.

279 1st feature resembles that of the raw data, while the energy of the bubbles in the 2nd feature
 280 is much smaller than that of the 1st feature. The plots therefore visually confirm that the
 281 1st-feature represents the most energetic mode of oscillations in the original data. This result
 282 also agrees with the observation of figure 3(b).

283

3.3. Error analysis

284 To show the effectiveness of the pre-processing, in figure 4, we plot the relative errors
 285 of the time-averaged kinetic energy of bubble clusters, one approximated using the original
 286 variable ($\langle \mathbf{q}^T \mathbf{T}(\langle \mathbf{r} \rangle) \mathbf{q} \rangle$) and the other using the transformed variables ($\langle \xi^T \mathbf{P}(\langle \eta \rangle) \xi \rangle$), against
 287 the excitation amplitude. At $A < 10^{-1}$, the error is nearly zero for both approximations. At
 288 $A > 10^{-1}$, at which bubble dynamics become nonlinear, the error grows with A for the
 289 former, while it remains small for the latter. This result confirms the improved approximation
 290 by the pre-processing.

291

3.4. Amplitude dependence of the key measures

292 The coherent dynamics critically depend on the excitation amplitude. Figure 5 shows the
 293 dependence of $\hat{\sigma}_1^2$, $\hat{S}_{v,N}$, and C against A for various density and polydispersity parameters
 294 of clouds. This dependence is best highlighted in the result of sparse, weakly polydisperse
 295 cluster ($B_0 = 0.5$, $s_d = 0.1$) in figure 5(a). The relative importance of the first PC, $\hat{\sigma}_1^2$,
 296 decays nearly monotonically from 0.9 to 0.2 through three distinct regimes. For $A \lesssim 0.2$,
 297 $\hat{\sigma}_1^2 \approx 1$ meaning the entire energy is captured by the first PC. Then $\hat{\sigma}_1^2$ rapidly decays to 0.4
 298 and stays nearly constant up to $A \approx 3$. At $A > 3$, $\hat{\sigma}_1^2$ decays again and stay nearly constant
 299 around 0.2. The decay indicates the decrease in the coherence with increasing A and can be
 300 explained by the excitation of the nonlinear oscillations and cavitation triggered at $A \approx 1$
 301 and above. $\hat{S}_{v,N}$ has a profile vertically mirrored to $\hat{\sigma}_1^2$; $\hat{S}_{v,N}$ increases from around 0.1 to
 302 0.5 through the three regimes, indicating more partitioning of the energy into multiple PCs
 303 and increase of incoherence, by increasing A . The mirrored profiles of $\hat{\sigma}_1^2$ and $\hat{S}_{v,N}$ suggest
 304 that these parameters are complementary.

305

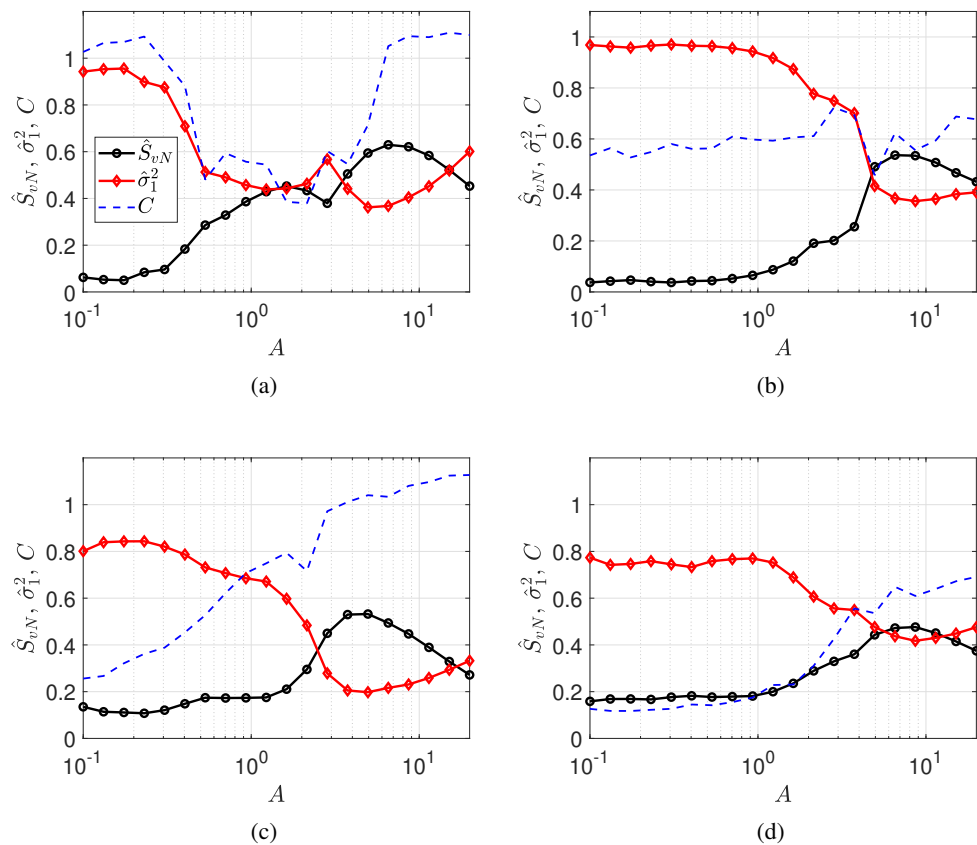


Figure 5: The first PC-variance ($\hat{\sigma}_1^2$), the normalized von Neumann entropy (\hat{S}_{vN}), and the coherence measure (C), against the excitation amplitude with various initial density and polydispersity parameters, with $(B_0, s_d) =$ (a) (0.5, 0.1), (b) (0.5, 0.7), (c) (5.0, 0.1), and (d) (5.0, 0.7).

306 Remarkably, C draws a square-well like profile, where $C \approx 1$ for both low ($A < 0.2$) and
 307 large ($A > 5$) amplitudes reaching a minimum value ($C \approx 0.5$) at $A \approx 1$. This counter-
 308 intuitive result suggests that the energy of the first PC is scaled by the mean-field parameter
 309 B_d regardless of the increase of incoherence for large A , and the scaling is lost only for the
 310 intermediate range at $1 < A < 5$. Comparisons with the result of dense, weakly polydisperse
 311 clusters $(B_0, s_d) = (5.0, 0.1)$ shown in figure 5(b) highlights the effect of the density of
 312 bubbles. In figure 5(b), $\hat{\sigma}_1^2$ takes values near unity at small A and decays at $A \approx 1.0$ to around
 313 0.4 and stays nearly constant at $A > 5.0$. At $A > 10$, $\hat{\sigma}_1^2$ slightly grows against A . \hat{S}_{vN} has a
 314 profile vertically mirrored to $\hat{\sigma}_1^2$. The features of $\hat{\sigma}_1^2$ and \hat{S}_{vN} are similar to those observed
 315 in figure 5(a), except that the transition occurs at a larger range of A . The sudden decay and
 316 growth of $\hat{\sigma}_1^2$ and \hat{S}_{vN} can likewise be associated the linear-to-nonlinear transition of bubble
 317 dynamics which results in incoherence. Meanwhile, the positive shift of the transition range
 318 of A indicates that the dense bubble clouds tend to behave more coherently than sparse bubble
 319 clouds, agreeing with the previously theory (d'Agostino & Brennen 1989; Maeda & Colonius
 320 2019). C is greater than 0.5 at almost all values of A , and is relatively more insensitive against
 321 A compared to figure 5(a), indicating that the cluster dynamics is moderately controlled by

322 B_d . Figure 3(c) and (d) respectively corresponds to sparse and dense clusters with strongly
 323 polydisperse nuclei; $(B_0, s_d) = (0.5, 0.7)$ and $(5.0, 0.7)$. These plots show clear differences
 324 from those of weakly polydisperse clusters in figure 5(a) and figure 5(b). For the sparse
 325 clusters (figure 5(c)), $\hat{\sigma}_1^2$ mildly decays from 0.8 to 0.7 at $0.1 < A < 0.2$, sharply decays to
 326 the minimum of 0.2 at $A \approx 0.4$ and then mildly grows to 0.3 at $A = 1.2$. The sharp decay
 327 resembles that in figure 5(b), while the slope is milder. \hat{S}_{vN} has also a profile mirrored to
 328 $\hat{\sigma}_1^2$, but it is less symmetric than figure 5(a) and figure 5(b). \hat{S}_{vN} is nearly constant around
 329 0.2 at $0.1 < A < 2.0$ and draws a concave curve with its maximum of around 0.5 at $A \approx 5.0$
 330 followed by a smooth decay to 0.3 at $A = 20$. C almost constantly increases from 0.2 to $O(1)$
 331 throughout the plot, indicating that the coherent dynamics is controlled by B_d , only at large
 332 A , unlike weakly polydisperse case of figure 5(a). The overall trend of the plots in figure 5(b)
 333 is similar to figure 5(a), although the changes of variables against A are milder in figure 5(b).
 334 Overall, the mutual trends of variables at $A > 5$ are similar between figure 5(a) and 5(c) and
 335 between figure 5(b) and 5(d), indicating the decreasing influence of the initial polydispersity
 336 in the nonlinear regime. At $A < O(1)$, the polydisperse clouds tend to have smaller values of
 337 $\hat{\sigma}_1^2$, and larger values of \hat{S}_{vN} and C . This can be explained by the enhancement of incoherent
 338 dynamics induced by polydispersity.

339 3.5. Principal component spectrum

340 To gain deeper insights on the meaning of C , in figure 6(a-d) we show the PC-variances
 341 obtained at $A = 2 \times 10^{-2}$, 1.2, and 20, for the first 10 PCs, obtained from the sparse,
 342 weakly polydisperse clouds blueplotted in figure 5(a). The insets show the evolution of the
 343 square-root of the normalized total energy (\sqrt{K}) and those of the first and the second PC-
 344 variances ($\sqrt{K_1}$ and $\sqrt{K_2}$), during the four periods of excitation in statistically stationary
 345 states. As expected, in the linear regime (figure 6(a)), the first PC occupies nearly the entire
 346 energy. $\sqrt{K_1}$ evolves at the fundamental frequency. In the transition regime (figure 6(b)), the
 347 first PC occupies 40% of the energy and the rest is partitioned into the sub-dominant PCs
 348 with a smooth decay. \sqrt{K} evolves more chaotic than the linear regime, as expected due to
 349 the nonlinear response of bubbles. Both $\sqrt{K_1}$ and $\sqrt{K_2}$ evolve with similar quasi-periodic
 350 profiles. We interpret that both PCs represent the coherent part of the energy.

351 Interestingly, in the nonlinear regime (figure 6(c)) energy partitioning is non-smooth; the
 352 first PC occupies 40% of the total energy similar to the transition regime, but the rest of the
 353 energy is broadly distributed into the other PCs with much smaller amplitudes. The evolution
 354 of \sqrt{K} is non-periodic with noisy, fine structures of spikes. These spikes are expected due
 355 to the incoherent collapse events. The evolution of $\sqrt{K_1}$ is, in contrast, highly periodic, and
 356 somewhat resembles that of figure 6(b). The evolution of $\sqrt{K_2}$ is more chaotic and less
 357 smooth than $\sqrt{K_1}$. The difference between the first-PC and the rest of PCs suggest that, in this
 358 regime, only the first PC captures a major coherent feature and the rest of PCs represent more
 359 the incoherent dynamics as broadband noise. Figure 6(d) shows the result of the nonlinear
 360 regime with a stronger excitation amplitude ($A = 20$). Overall, both the spectrum and the
 361 evolution of K look similar to that of figure 6(c), other than that the amplitude of the first
 362 PC-variance is increased to 60%. The resemblance of figure 6(c) and figure 6(d) indicates
 363 that the dynamical features identified from these two plots are common in the non-linear
 364 regimes.

365 The resemblance of the evolution of the first PCs in the linear and the nonlinear regimes
 366 can explain the recovery of C at $A > 5$ in figure 6(a). Although the overall dynamics are
 367 much more chaotic in the nonlinear regime, the contribution of the coherent interactions to
 368 the system's energy effectively appears only in the first PC in both regimes and therefore the
 369 relative contribution of the interaction to the first PC is commonly scaled by B_d . This result

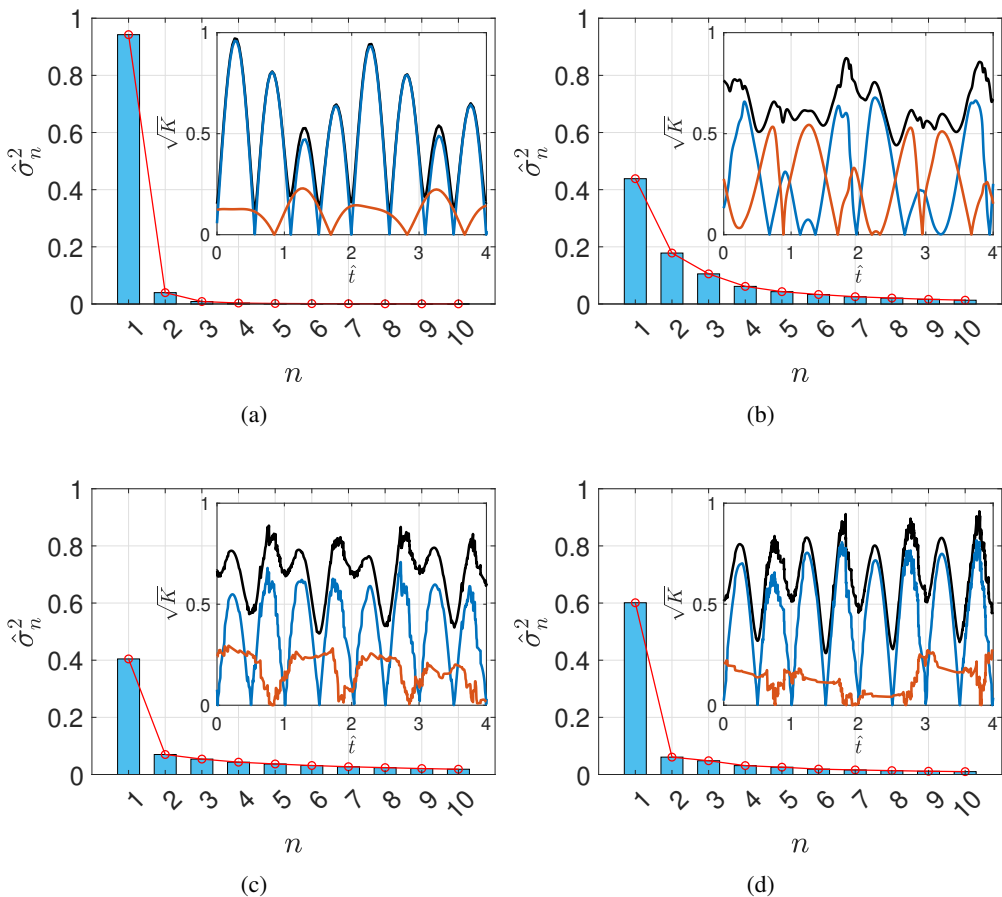


Figure 6: (a-d) The PC spectral profiles at $A = 2 \times 10^{-2}$, 1.2, 8.6, and 20. Insets show the evolution of the square-root of the normalized total energy (\sqrt{K} , black) and those of the first (blue) and the second (red) PC-variances ($\sqrt{K_1}$ and $\sqrt{K_2}$), during the four periods of harmonic excitation, with \hat{t} being a non-dimensional time $\hat{t} = tf$. The y-axis of each inset is normalized by the maximum value of \sqrt{K} .

370 also implies that the underlying coherence in the nonlinear regime represents the perfect
 371 correlation (synchronized oscillations) like that of linear, monodisperse clouds.

372 3.6. Evaluation of the coherence measure

373 To assess the variation of the coherence measure dependent on the three regimes, in figure
 374 7(a-c) we plot B_1 against B_d for $O(10^3)$ clusters with various values of s_d , N , and R_C , at the
 375 three distinct excitation amplitudes ($A = 2 \times 10^{-2}$, 1.2, and 20). Appendix D summarizes the
 376 parameters used. With the weak excitation (figure 7(a)), bubble oscillations are in a linear
 377 regime and $B_d \approx B_0$. The data points are collapsed on the line of $C = 1$ for $s_d = 0.1$, while
 378 data points are scattered for the other values of s_d . This result is expected as B_d was originally
 379 defined to scale the coherence of the monodisperse, perfectly correlated bubbles in the linear
 380 regime (d'Agostino & Brennen 1989; Zeravcic *et al.* 2011). With the intermediate excitation
 381 (figure 7(b)), the bubble oscillations are nonlinear. The data points are scattered from $C = 1$,
 382 regardless of the value of s_d . In this regime, the result indicates that the coherence is lost due to

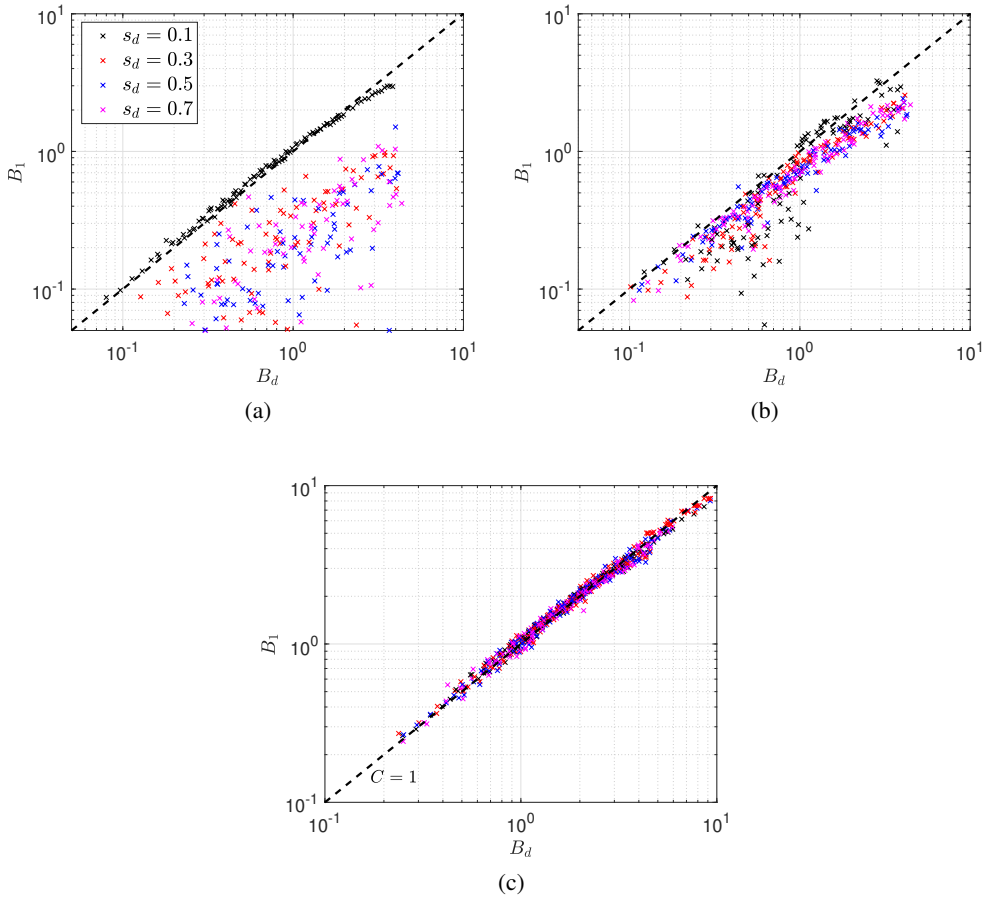


Figure 7: B_1 against B_d , for $O(10^3)$ clusters with various values of s_d , N , R_C , and other physical parameters (Appendix D) in the linear (a: $A = 2 \times 10^{-2}$), transition (b: $A = 1.2$), and nonlinear (c: $A = 20$) regimes.

383 the nonlinear dynamics, regardless of polydispersity. Surprisingly, with the strong excitation,
 384 (figure 7(c)), the data points are collapsed on the line of $C = 1$, meaning that the variance of
 385 the first PC is scaled by B_d , regardless of the parameters. This collapse is not observed for the
 386 second PC, regardless of the parameters (Appendix E). The results suggest that the scaling
 387 of B_1 is universal for cavitating clusters, which are typically excited at $O(1)$ MPa and above.
 388 Physically speaking, the scaling indicates that in cavitating clusters, the energy is partitioned
 389 into a single coherent mode of correlated (synchronized) oscillations and incoherent modes
 390 generalizing the aforementioned interpretation of figure 6(c) and figure 6(d). The coherent
 391 energy is controlled by the mean field originally derived for monodisperse, near-equilibrium
 392 bubbles. It is suggested that this partitioning and scaling are universal regardless of nuclei's
 393 polydispersity and the degree of nonlinearity. This finding explains successful use of B_d in
 394 characterizing and controlling seemingly disordered clusters in our previous studies (Maeda
 395 & Colonius 2019; Maeda & Maxwell 2021). The isolation of coherence implies that the
 396 details of the microscopic scale as well as of the many-body interactions represented by
 397 higher-order PCs could be modeled as fast variables, which force the macroscopic (relevant)

398 scales in the form of noise. A rigorous way to corroborate this hypothesis might be the
399 Mori-Zwanzig projection operator formalism (Mori 1965) by assuming that the fluctuations
400 of fast and slow variables are uncorrelated. Although the choice of relevant slow variables
401 may be unclear, the present approach is promising since the WPCA-TD can be seen as a
402 prompter for more refined models of the evolution of slow variables.

403 **4. Significance of the WPCA-TD to inform on cavitation physics**

404 In general, PCA can be considered a mathematical technique for data decomposition,
405 and the physical meaning of extracted features can be left to one's interpretation. To the
406 authors' knowledge, studies of complex fluid flows employing data decomposition and feature
407 extraction techniques often rely on one's intuition to discuss the physical meaning of features.
408 In fact, in §3.5, the analyses of the PC spectra required rigorous interpretations based on
409 previous knowledge on cavitation physics. Meanwhile, the correlation between B_1 and B_d
410 identified in the last section shows that the present WPCA-TD can directly extract B_d from
411 data without additional signal processing like spectral filtering, which are often required for
412 Fourier-based analysis, despite the presence of noisy and incoherent features. To generalize,
413 WPCA-TD is not only a tool for data decomposition and feature extraction but also can be
414 used to directly inform on the non-dimensional number that controls the coherent physics
415 of interests. This informative aspect of the WPCA-TD can provide a meaningful shortcut to
416 address the physics of cloud cavitation.

417 To perform WPCA-TD, bubble dynamics data can be obtained from experimental mea-
418 surements as well as other numerical approaches (e.g., Kameda & Matsumoto 1996; Maeda
419 & Colonius 2018; Pishchalnikov *et al.* 2019). The cluster's shape can be arbitrary. Although
420 bubble's translation and deformation are neglected in the present numerical experiments,
421 WPCA-TD can incorporate dynamical variables controlling these effects (e.g., Ilinskii *et al.*
422 2007; Murakami *et al.* 2020). Physically meaningful data decomposition requires the fine
423 temporal resolution of individual bubble dynamics. In practice, such information would be
424 difficult to obtain in experiments except for a small number of bubbles in a highly controlled
425 environment. We thus emphasize that WPCA-TD would primarily be useful to process fine
426 temporal resolution of numerical data.

427 **5. Conclusion**

428 In conclusion, to corroborate the coherent dynamics of bubble clusters, we introduced and
429 demonstrated WPCA-TD, a method of PCA to comprehensively decompose the time series
430 of Lagrangian bubble dynamics data into coherent dynamical features, in way similar to
431 the modal decomposition of Eulerian flow field data. The data are pre-processed such that
432 the PC-variance of the features becomes consistent with the hydrodynamic potential energy
433 induced by bubble oscillations that is predicted by the R-P equation. Analyzing simulation
434 data sets of clusters under harmonic excitation, we demonstrated that the coherent energy and
435 the degree of incoherence are respectively quantified by the variance and the spectral entropy.
436 The coherence was lowered by disorders induced by nuclei's polydispersity and nonlinear
437 response of bubbles, as expected. Meanwhile, in cavitating regimes, underlying, correlated
438 mode of oscillations were isolated in a single dominant feature. The variance of this feature
439 was found to be controlled by the previously identified mean-field parameter, B_d , regardless
440 of the disordering factors, indicating that the underlying coherent dynamics may be universal
441 in cavitating clusters. These results suggest that the method can provide a simplified and
442 comprehensive representation of complex bubble dynamics. Analogous to the use of POD
443 for reduced-order modeling (ROM) of various flows, the dynamical features extracted by

444 PCA may be used for ROM, without directly solving many-body interactions. Such a model
 445 may be promising for controlling cavitation and nucleation without tracking individual nuclei.

446

447 **Acknowledgement.** Both authors are appreciative of Professor Tim Colonius's mentoring
 448 during their days at Caltech. We are pleased to offer this paper as a small token of our
 449 appreciation on the occasion of his 60th birthday.

450

451 **Funding.** K.M. acknowledges support from SRB Co. Inc. and Purdue University. Some of
 452 the computation presented here utilized the Extreme Science and Engineering Discovery
 453 Environment (XSEDE), which is supported by NSF under grant TG-CTS190009, as
 454 well as Anvil at Purdue University through allocation PHY220130 from the Advanced
 455 Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program,
 456 which is supported by National Science Foundation grants #2138259, #2138286, #2138307,
 457 #2137603, and #2138296.

458

459 **Declaration of Interests.** The authors report no conflict of interest.

460

461 **Author ORCID.**

462 Kazuki Maeda <https://orcid.org/0000-0002-5729-6194>;

463 Daniel Fuster <https://orcid.org/0000-0002-1718-7890>.

464

465 **Appendix A. Variable transformation for a single bubble**

466 In this section, we provide supplemental discussions and justification on the variable
 467 transformation used in the WPCA-TD, by considering the single bubble dynamics. By
 468 transformation $(\xi, \eta) = (R\dot{R}, R)$, with ξ being nothing but the velocity potential evaluated at
 469 the bubble surface, the R-P equation can be expressed as

$$470 \quad \dot{\xi} = -\frac{1}{2} \left(\frac{\xi}{\eta} \right)^2 + \frac{p_b - p_\infty(t)}{\rho}, \quad (\text{A } 1)$$

$$471 \quad \dot{\eta} = \frac{\xi}{\eta}. \quad (\text{A } 2)$$

473 p_b and p_∞ are the pressure inside the bubble and that in infinity, respectively. The
 474 instantaneous energy of the system is expressed as

$$475 \quad K = 2\pi\rho\dot{R}^2R^3 = 2\pi\rho\xi^2\eta. \quad (\text{A } 3)$$

476 Next, we approximate the system (A 1). Given the time series data, the temporal average of
 477 η can be computed as $\langle \eta \rangle$. Using $\langle \eta \rangle$, we approximate the system of eq. (A 1) on a space
 478 spanned by a set of new variables $(\hat{\xi}, \hat{\eta})$.

$$479 \quad \dot{\hat{\xi}} = -\frac{1}{2} \left(\frac{\hat{\xi}}{\langle \eta \rangle} \right)^2 + \frac{p_b - p_\infty(t)}{\rho}, \quad (\text{A } 4)$$

$$480 \quad \dot{\hat{\eta}} = \frac{\hat{\xi}}{\langle \eta \rangle}. \quad (\text{A } 5)$$

482 This system models well the dynamical features of the R-P equation including bifurcation,
 483 although $\hat{\xi}$ and $\hat{\eta}$ are now partially decoupled. Our explanation is the following. In the
 484 linear regime, the quadratic term of eqn. (A 4) can be neglected, and this system becomes

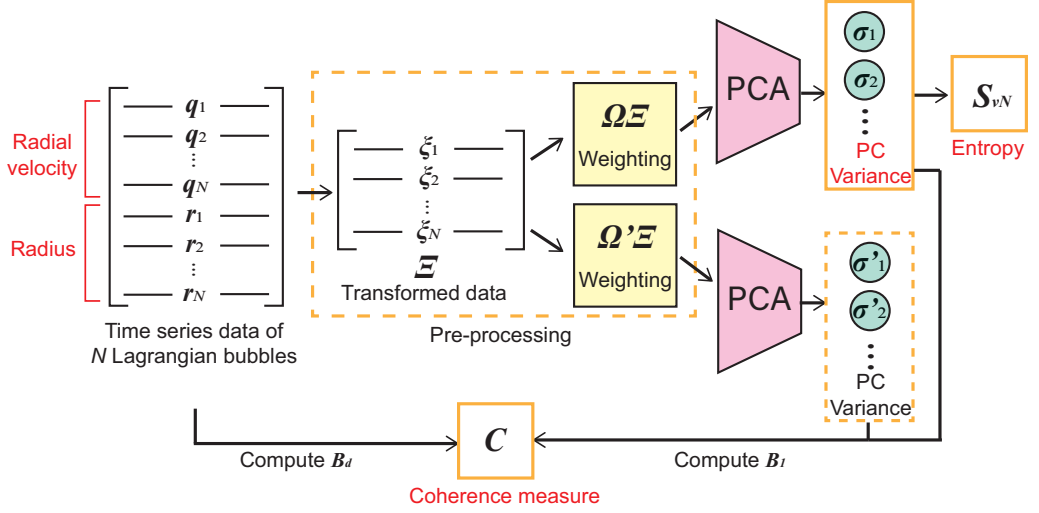


Figure 8: Comprehensive schematic of the present procedure to obtain the spectral entropy, PC-variance, and coherence measure, from input data.

485 identical with the original one. In nonlinear, cavitating regime, the bubble size (η) varies
 486 slowly near its peak after the explosive cavitation growth. Sporadic, fast collapse events
 487 do not influence much the temporal mean of the bubble size. Therefore, η is close to $\langle \eta \rangle$
 488 except at collapse events and, even if η changes rapidly during the collapse and rebound, this
 489 change has a relatively small influence on the mean behavior of the system at the timescale
 490 of (statistically) stationary bubble oscillations.

491 We can express the instantaneous energy of this approximate system as

$$492 \quad \hat{K} = 2\pi\rho\hat{\xi}^2\hat{\eta}. \quad (\text{A } 6)$$

493 With the same initial condition ($\xi = \hat{\xi}$ and $\eta = \hat{\eta}$ at $t = 0$), the temporal mean of the energy
 494 of the original system is approximated as

$$495 \quad \langle K \rangle \approx \langle \hat{K} \rangle = 2\pi\rho\langle \hat{\xi}^2\hat{\eta} \rangle \approx 2\pi\rho\langle \hat{\xi}^2 \rangle \langle \eta \rangle. \quad (\text{A } 7)$$

496

497 Note that Preston *et al.* (2007) used a POD-based analysis of single bubble dynamics for
 498 reduced-order modeling of heat and mass diffusion across bubble interface. In the study, the
 499 temperature and concentration fields were obtained by solving partial differential equations
 500 and then represented by POD modes, following the POD/Galerkin framework. This is distinct
 501 from the present PCA (POD) of Lagrangian bubble dynamics data based on the direct
 502 transformation and projection of the Rayleigh-Plesset equation.

503 **Appendix B. Schematic of the input-output procedure**

504 Figure 8 shows the schematic of the present procedure to obtain the spectral entropy, PC-
 505 variance, and coherence measure, from the input data including the time series of the radius
 506 and radial velocity of Lagrangian bubbles.

507 For computing C , we introduced the alternative weight, Ω' . Ω' is defined through \mathbf{P}' :
 508 $\mathbf{P}'(\langle s \rangle) = \Omega' \Omega'^*$, where \mathbf{P}' is the diagonal matrix with its entries from \mathbf{P} :

$$509 \quad P'_{ij}(\eta) = \begin{cases} 2\pi\rho\eta_i & i = j, \\ 0 & i \neq j. \end{cases} \quad (\text{B } 1)$$

510 The corresponding inner product, $\xi^T \mathbf{P}' \xi$, represents the portion of the energy of the system
 511 excluding the contribution of inter-bubble interactions.

$$512 \quad K' = \xi^T \mathbf{P}' \xi, \quad (\text{B } 2)$$

513 where

$$514 \quad K' = 2\pi\rho_l \sum_{i=1}^N R_i^3 \dot{R}_i^2 + (\text{H.O.T}). \quad (\text{B } 3)$$

515

516 **Appendix C. Details of the K-M equation used for data preparation**

517 In this section, we provide details of the formulation of the Keller-Miksis (K-M) equation for
 518 multiple bubbles, which is used to generate the Lagrangian bubble dynamics data. Various
 519 extensions of the Rayleigh-Plesset equation and its applications to a system of multiple
 520 bubbles are available (Doinikov 2004; Takahira *et al.* 1994; Ilinskii *et al.* 2007; Yasui *et al.*
 521 2008). Our formulation can be derived from the K-M equation extended for multiple bubbles.
 522 First, we recall the general formulation for the oscillations of interacting spherical bubbles
 523 in weakly compressible liquid with arbitrary inter-bubble distances, whose derivation is, for
 524 instance, provided in Appendix 2.4 of Fuster & Colonius (2011). In the present study, we
 525 use a simplified version of this derivation. To recall, in Fuster & Colonius (2011), the radial
 526 evolution of bubble i is described as

$$527 \quad \ddot{R}_i \left(R_i \left(1 - \frac{\dot{R}_i}{c} \right) \right) + \frac{3}{2} \dot{R}_i^2 \left(1 - \frac{\dot{R}_i}{3c} \right) = F^* + I^*, \quad (\text{C } 1)$$

528 where F^* and I^* represent the forcing due to the external potential and the inter-bubble
 529 interaction, respectively, and are expressed as

$$530 \quad F^* = \frac{\partial \phi_\infty}{\partial t} \left(1 - \frac{\dot{R}_i}{c} \right) + \frac{R_i}{c} \frac{\partial^2 \phi_\infty}{\partial t^2} + H_i \left(1 + \frac{\dot{R}_i}{c} \right) + \frac{R_i \dot{H}_i}{c}, \quad (\text{C } 2)$$

531 and

$$532 \quad I^* = \sum_{j \neq i}^N \left[\left(1 + \frac{\dot{R}_i}{c} \right) \frac{\partial \phi_j(R_i)}{\partial t} \right] + \frac{R_i}{c} \sum_{j \neq i}^N \left[\left(1 + \frac{\dot{R}_j(t')}{c} \right) \left(1 - \frac{\dot{R}_j}{c} \right) \frac{\partial^2 \phi_j(t' - R_j(t')/c)}{\partial t'^2} \right] \quad (\text{C } 3)$$

$$533 \quad - \sum_{j \neq i}^N \left[\left(1 + \frac{\dot{R}_j(t')}{c} \right) \frac{\partial \phi_j(t' - R_j(t')/c)}{\partial t'} \frac{R_i}{R_j} \frac{\dot{R}_j}{c} \right]. \quad (\text{C } 4)$$

534

535 ϕ_∞ is the velocity potential of liquid at infinity, and $\phi_i(R_i)$ and H_i are the potential and
 536 the enthalpy of liquid evaluated at the surface of bubble i . t' is the retarded time defined as
 537 $t' = t - (d_{ij} - R_j)/c$, where $(d_{ij} - R_j)/c$ represents the travel time for the pressure wave to
 538 reach bubble i from the surface of the bubble j and d_{ij} is the distance between the centers of
 539 those bubbles. In the sparse limit, the equation recovers the original Keller-Miksis equation
 540 (Keller & Miksis 1980).

541 To close the equations, several relations are considered. Using the Bernoulli's equation,
 542 the potential for the bubble i can be expressed as

$$543 \quad \frac{\partial \phi_i(R_i)}{\partial t} = - \left(\frac{1}{2} \dot{R}_i^2 + H_i + \sum_{j \neq i}^N \frac{\partial \phi_j(R_i)}{\partial t} + \frac{\partial \phi_\infty}{\partial t} \right). \quad (\text{C } 5)$$

544

545 The velocity potentials at the bubble i and bubble j satisfy the following relation.

$$546 \quad \frac{\partial \phi_i(R_i)}{\partial t} = \frac{R_j(t')}{d_{ij}} \frac{\partial \phi_j(t' - R_j/c)}{\partial t'}. \quad (\text{C } 6)$$

548 The enthalpy and the potential derivative are approximated as

$$549 \quad H_i \approx \frac{p_i - p_0}{\rho_{l,0}}. \quad (\text{C } 7)$$

$$551 \quad \frac{\partial \phi_\infty}{\partial t} \approx \frac{p_\infty - p_0}{\rho_{l,0}}. \quad (\text{C } 8)$$

552 Finally, p_∞ is obtained using the information of the background Eulerian field computed on
553 a mesh.

554 In the present study, we simplify F^* and I^* by invoking further approximations. First,
555 consider that the bubble cluster size is smaller than the characteristic length-scale of the
556 pressure wave in the field, λ_c . The inter-bubble distance in the cluster is naturally smaller
557 than λ_c :

$$558 \quad d_{ij} - R_j \ll \lambda_c. \quad (\text{C } 9)$$

559 Dividing both sides with c ,

$$560 \quad \frac{d_{ij} - R_j}{c} = t - t' \ll \frac{\lambda_c}{c} = \frac{1}{f_c} = T_c, \quad (\text{C } 10)$$

561 where f_c and T_c are the characteristic frequency and the period of the wave in the field.
562 Therefore, the difference between t and t' is much smaller than the characteristic timescale
563 of the dynamics. Given this knowledge, we approximate that $t' \approx t$. Second, we neglect the
564 terms in the order of $(\dot{R}/c)^2$. Using these approximations, we can simplify F^* and I^* as

$$565 \quad F^* = \frac{p_i - p_\infty}{\rho_l} \left(1 + \frac{\dot{R}_i}{c} \right) + \frac{R_i}{\rho_l c} \frac{\partial (p_i - p_\infty)}{\partial t} \quad (\text{C } 11)$$

566 and

$$567 \quad I^* = \quad (\text{C } 12)$$

$$568 \quad - \sum_{j \neq i}^N \frac{R_j}{d_{ij}} \left(\frac{1}{2} \dot{R}_j^2 + \dot{R}_j \left(R_j \left(1 - \frac{\dot{R}_j}{c} \right) \right) + \frac{3}{2} \dot{R}_j^2 \left(1 - \frac{\dot{R}_j}{3c} \right) - \frac{p_j - p_\infty}{\rho_l} \frac{\dot{R}_j}{c} - \frac{R_j}{\rho_l c} \frac{\partial (p_j - p_\infty)}{\partial t} \right).$$

$$569 \quad (\text{C } 13)$$

570 For harmonic excitation,

$$571 \quad p_\infty = p_a \sin(\omega t). \quad (\text{C } 14)$$

572 We use polytropic law to describe the pressure of the gas inside each bubble (Brennen 2014).

$$573 \quad p_i = p_{i,0} \left(\frac{R_{i,0}}{R_i} \right)^{3\gamma}, \quad (\text{C } 15)$$

574 where γ is the constant polytropic exponent. Eqns (C 1), (C 11) and (C 13), together with
575 relations (C 14) and (C 15) provide a complete system of ODEs for the radius of N interacting
576 bubbles, which can be readily solved with a given initial condition.

N	R_c [mm]	R_{ref} [μm]	s_d	f [kHz]	N	R_c [mm]	R_{ref} [μm]	s_d	f [kHz]
16	0.33-2.0	10	0.1-0.7	500					
32	0.33-2.0	10	0.1-0.7	500					
64	0.33-2.0	10	0.1-0.7	500	16	0.33-2.0	10	0.1-0.7	500
128	0.33-2.0	10	0.1-0.7	500	32	0.33-2.0	10	0.1-0.7	500
256	0.33-2.0	10	0.1-0.7	500	64	0.33-2.0	10	0.1-0.7	500
512	0.8-2.0	10	0.1-0.7	500	128	0.33-2.0	10	0.1-0.7	500
64	0.33-2.0	5	0.1-0.7	500					
64	0.33-2.0	10	0.1-0.7	250					

Table 1: Summary of the parameters used for the data set of clusters analyzed in the main manuscript.

577 Appendix D. Summary of the parameters

578 Table 1 summarizes a set of parameters used to construct the database analyzed the main
579 manuscript. The number of bubbles N , cluster radius R_c , reference bubble radius R_{ref} ,
580 polydispersity measure s_d , and the forcing frequency were varied. The left table of parameters
581 are used for the clusters in figure 4 ($s_d = 0.1, 0.7$), figure 7(a) (all values of s_d), and figure
582 9(a) (all values of s_d). The right table of parameters are used for the clusters in figure 7 (b),
583 figure 7 (c), figure 9 (b), and figure 9 (c).

584 Appendix E. Scaling of the second PC

585 Figure 9 shows the scaling of the interaction energy of the second dominant feature of bubble
586 clouds in the data set, $B_2 = \sigma_2^2/\sigma_2'^2 - 1$, for various values of excitation amplitude and
587 nuclei polydispersity. For all parameters, data points are widely scattered and there is no
588 clear correlation observed between B_2 and B_d , indicating that the second dominant feature
589 do not represent coherent oscillations like those of the principal feature.

REFERENCES

- 590 ABDI, HERVÉ & WILLIAMS, LYNNE J 2010 Principal component analysis. *Wiley Interdiscip. Rev. Comput.*
591 *Stat.* **2** (4), 433–459.
- 592 ALTER, ORLY, BROWN, PATRICK O & BOTSTEIN, DAVID 2000 Singular value decomposition for genome-wide
593 expression data processing and modeling. *Proceedings of the National Academy of Sciences* **97** (18),
594 10101–10106.
- 595 ANDO, K, LIU, A-Q & OHL, C-D 2012 Homogeneous nucleation in water in microfluidic channels. *Phys.*
596 *Rev. Lett.* **109** (4), 044501.
- 597 ANGÉLIL, RAYMOND, DIEMAND, JÜRIG, TANAKA, KYOKO K & TANAKA, HIDEKAZU 2014 Bubble evolution
598 and properties in homogeneous nucleation simulations. *Phys. Rev. E* **90** (6), 063301.
- 599 ARNDT, ROGER EA 1981 Cavitation in fluid machinery and hydraulic structures. *Annual Review of Fluid*
600 *Mechanics* **13** (1), 273–326.
- 601 AUBRY, NADINE, GUYONNET, RÉGIS & LIMA, RICARDO 1991 Spatiotemporal analysis of complex signals:
602 theory and applications. *Journal of Statistical Physics* **64**, 683–739.
- 603 BLANDER, M & KATZ, JL 1975 Bubble nucleation in liquids. *AIChE Journal* **21** (5), 833–848.
- 604 BREMOND, N, ARORA, M, OHL, CD & LOHSE, D 2006 Controlled multibubble surface cavitation. *Phys. Rev.*
605 *Lett.* **96** (22), 224501–224501.
- 606 BRENNEN, CE 2014 *Cavitation and bubble dynamics*. Cambridge University Press.
- 607 CAIRÓS, C & METTIN, R 2017 Simultaneous high-speed recording of sonoluminescence and bubble dynamics
608 in multibubble fields. *Phys. Rev. Lett.* **118** (6), 064301.
- 609 CHEN, YONGLIAN & HEISTER, STEPHEN D 1994 A numerical treatment for attached cavitation. *Journal of*
610 *fluids engineering* **116** (3), 613–618.

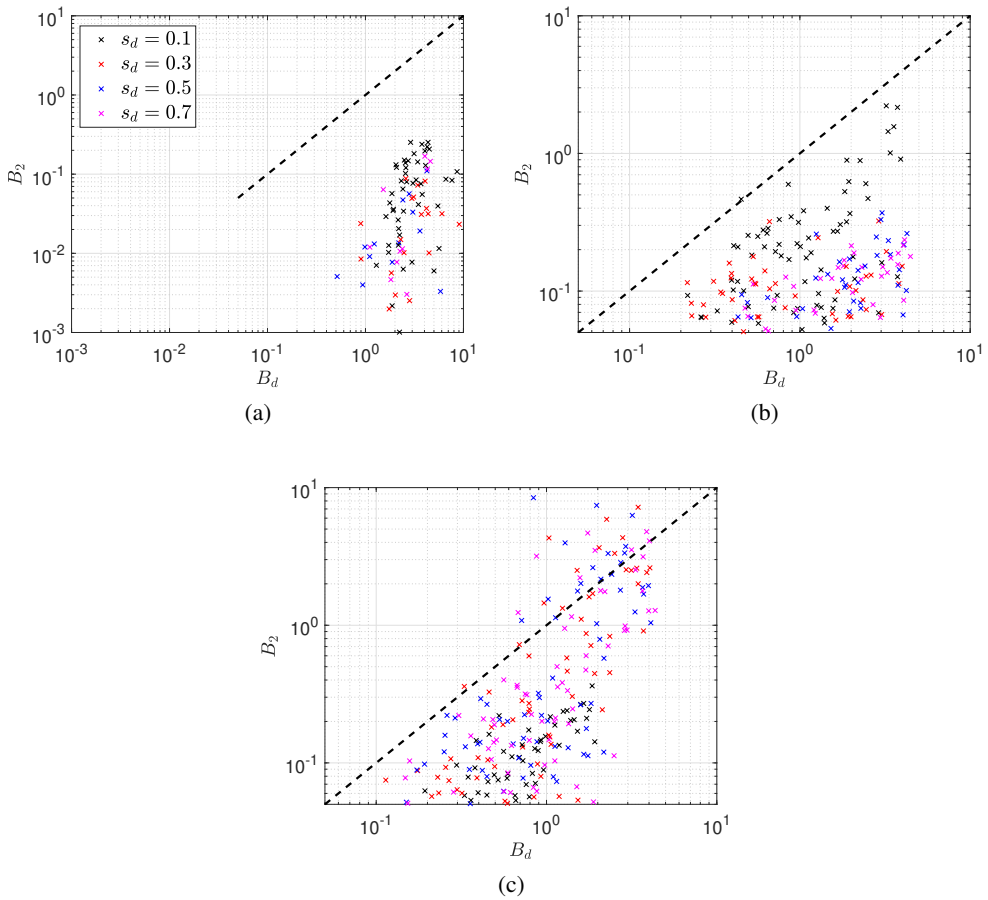


Figure 9: B_2 against B_d , for $O(10^3)$ clusters with various values of s_d , N , R_C , and other physical parameters (Appendix D) in the linear (a: $A = 2 \times 10^{-2}$), transition (b: $A = 1.2$), and nonlinear (c: $A = 20$) regimes.

- 611 D'AGOSTINO, L & BRENNEN, CE 1989 Linearized dynamics of spherical bubble clouds. *J. Fluid. Mech.* **199**,
612 155–176.
- 613 DE DOMENICO, MANLIO & BIAMONTE, JACOB 2016 Spectral entropies as information-theoretic tools for
614 complex network comparison. *Physical Review X* **6** (4), 041062.
- 615 DOINIKOV, ALEXANDER A 2004 Mathematical model for collective bubble dynamics in strong ultrasound
616 fields. *The Journal of the Acoustical Society of America* **116** (2), 821–827.
- 617 FUSTER, DANIEL & COLONIUS, TIM 2011 Modelling bubble clusters in compressible liquids. *J. Fluid Mech.*
618 **688**, 352–389.
- 619 GALLO, MIRKO, MAGALETTI, FRANCESCO & CASCIOLA, CARLO MASSIMO 2021 Heterogeneous bubble
620 nucleation dynamics. *J. Fluid Mech.* **906**.
- 621 GANESH, HARISH, MÄKI HARJU, SIMO A & CECCIO, STEVEN L 2016 Bubbly shock propagation as a
622 mechanism for sheet-to-cloud transition of partial cavities. *Journal of Fluid Mechanics* **802**, 37–
623 78.
- 624 GNANASKANDAN, ASWIN & MAHESH, KRISHNAN 2016 Large eddy simulation of the transition from sheet
625 to cloud cavitation over a wedge. *International Journal of Multiphase Flow* **83**, 86–102.
- 626 HASHIMOTO, TOMOYUKI, YOSHIDA, MAKOTO, WATANABE, MITSUO, KAMIJO, KENJIRO & TSUJIMOTO,
627 YOSHINOBU 1997 Experimental study on rotating cavitation of rocket propellant pump inducers.
628 *Journal of Propulsion and Power* **13** (4), 488–494.

- 629 HOLMES, PHILIP 2012 *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge
630 university press.
- 631 HU, JIANHUA, YIN, GUOSHENG, MORRIS, JEFFREY S, ZHANG, LI & WRIGHT, FRED A 2005 Entropy and
632 survival-based weights to combine affymetrix array types and analyze differential expression and
633 survival. *Methods of Microarray Data Analysis* pp. 95–108.
- 634 IKEDA, T, YOSHIKAWA, S, TOSAKI, M, ALLEN, JS, TAKAGI, S, OHTA, N, KITAMURA, T & MATSUMOTO, Y
635 2006 Cloud cavitation control for lithotripsy using high intensity focused ultrasound. *Ultrasound*.
636 *Med. Biol.* **32** (9), 1383–1397.
- 637 ILINSKII, YURII A, HAMILTON, MARK F & ZABOLOTSKAYA, EVGENIA A 2007 Bubble interaction dynamics
638 in lagrangian and hamiltonian mechanics. *J. Acoust. Soc. Am.* **121** (2), 786–795.
- 639 JOLLIFFE, IAN T & CADIMA, JORGE 2016 Principal component analysis: a review and recent developments.
640 *Philos. Trans. R. Soc. A* **374** (2065), 20150202.
- 641 KAMEDA, MASAHARU & MATSUMOTO, YOICHIRO 1996 Shock waves in a liquid containing small gas bubbles.
642 *Physics of Fluids* **8** (2), 322–335.
- 643 KATZ, J 1978 Determination of solid nuclei and bubble distributions in water by holography. *Calif. Inst. of*
644 *Tech., Eng. and Appl. Sci. Div. Rep. No. 183* **3**.
- 645 KELLER, JB & MIKSI, M 1980 Bubble oscillations of large amplitude. *J. Acoust. Soc. Am.* **68** (2), 628–633.
- 646 KUBOTA, AKIHIRO, KATO, HIROHARU & YAMAGUCHI, HAJIME 1992 A new modelling of cavitating flows:
647 a numerical study of unsteady cavitation on a hydrofoil section. *Journal of fluid Mechanics* **240**,
648 59–96.
- 649 KULLBACK, SOLOMON 1997 *Information theory and statistics*. Courier Corporation.
- 650 KUNZ, ROBERT F, BOGER, DAVID A, STINEBRING, DAVID R, CHYCZEWSKI, THOMAS S, LINDAU, JULES W,
651 GIBELING, HOWARD J, VENKATESWARAN, SANKARAN & GOVINDAN, TR0972 2000 A preconditioned
652 navier–stokes method for two-phase flows with application to cavitation prediction. *Computers &*
653 *Fluids* **29** (8), 849–875.
- 654 LALL, SANJAY, MARSDEN, JERROLD E & GLAVAŠKI, SONJA 1999 Empirical model reduction of controlled
655 nonlinear systems. *IFAC Proc. Volumes* **32** (2), 2598–2603.
- 656 LLOYD, SETH, MOHSENI, MASOUD & REBENTROST, PATRICK 2014 Quantum principal component analysis.
657 *Nat. Phys.* **10** (9), 631–633.
- 658 MAEDA, K & COLONIUS, T 2018 Eulerian–lagrangian method for simulation of cloud cavitation. *J. Comput.*
659 *Phys.* **371**, 994–1017.
- 660 MAEDA, K & COLONIUS, T 2019 Bubble cloud dynamics in an ultrasound field. *J. Fluid Mech.* **862**,
661 1105–1134.
- 662 MAEDA, KAZUKI, KREIDER, WAYNE, MAXWELL, ADAM, CUNITZ, BRYAN, COLONIUS, TIM & BAILEY,
663 MICHAEL 2015 Modeling and experimental analysis of acoustic cavitation bubbles for burst wave
664 lithotripsy. In *Journal of Physics: Conference Series*, , vol. 656, p. 012027. IOP Publishing.
- 665 MAEDA, K, MAXWELL, AD, COLONIUS, T, KREIDER, W & BAILEY, MR 2018 Energy shielding by cavitation
666 bubble clouds in burst wave lithotripsy. *J. Acoust. Soc. Am.* **144** (5), 2952–2961.
- 667 MAEDA, KAZUKI & MAXWELL, ADAM D 2021 Controlling the dynamics of cloud cavitation bubbles through
668 acoustic feedback. *Phy. Rev. Applied.* **15** (3), 034033.
- 669 MAXWELL, AD, WANG, T-Y, CAIN, CA, FOWLKES, JB, SAPOZHNIKOV, OA, BAILEY, MR & XU, Z 2011
670 Cavitation clouds created by shock scattering from bubbles during histotripsy. *J. Acoust. Soc. Am.*
671 **130** (4), 1888–1898.
- 672 MERKLE, CL, FENG, J & BUELOW, P 1998 Computational modeling of the dynamics of sheet cavitation. In
673 *Proc. 3rd International Symposium on Cavitation*, , vol. 2, pp. 307–311.
- 674 MILANO, MICHELE & KOUMOUTSAKOS, PETROS 2002 Neural network modeling for near wall turbulent flow.
675 *J. Comput. Phys.* **182** (1), 1–26.
- 676 MØRCH, KA 1980 On the collapse of cavity clusters in flow cavitation. In *Cavitation and Inhomogeneities*
677 *in Underwater Acoustics*, pp. 95–100. Springer.
- 678 MORI, HAZIME 1965 Transport, collective motion, and brownian motion. *Progress of theoretical physics*
679 **33** (3), 423–455.
- 680 MOVAHED, P, KREIDER, W, MAXWELL, AD, HUTCHENS, SB & FREUND, JB 2016 Cavitation-induced damage
681 of soft materials by focused ultrasound bursts: A fracture-based bubble dynamics model. *J. Acoust.*
682 *Soc. Am.* **140** (2), 1374–1386.
- 683 MURAKAMI, KAZUYA, GAUDRON, RENAUD & JOHNSEN, ERIC 2020 Shape stability of a gas bubble in a soft
684 solid. *Ultrason Sonochem* **67**, 105170.

- 685 PISHCHALNIKOV, YA, WILLIAMS JR, JC & McATEER, JA 2011 Bubble proliferation in the cavitation field of
686 a shock wave lithotripter. *J. Acoust. Soc. Am.* **130** (2), EL87–EL93.
- 687 PISHCHALNIKOV, YURI A, BEHNKE-PARKS, WILLIAM M, SCHMIDMAYER, KEVIN, MAEDA, KAZUKI,
688 COLONIUS, TIM, KENNY, THOMAS W & LASER, DANIEL J 2019 High-speed video microscopy and
689 numerical modeling of bubble dynamics near a surface of urinary stone. *The Journal of the Acoustical*
690 *Society of America* **146** (1), 516–531.
- 691 PLESSET, MS & ELLIS, AT 1955 On the mechanism of cavitation damage. *Transactions of the ASME* **77**,
692 1055–1064.
- 693 PLESSET, MS & PROSPERETTI, A 1977 Bubble dynamics and cavitation. *Annu. Rev. Fluid Mech.* **9** (1),
694 145–185.
- 695 PLESSET, MILTON S 1949 The dynamics of cavitation bubbles. *J. Appl. Mech.* **16**, 277–282.
- 696 PRESTON, AT, COLONIUS, T & BRENNEN, CE 2007 A reduced-order model of diffusive effects on the
697 dynamics of bubbles. *Physics of Fluids* **19** (12).
- 698 PROSPERETTI, ANDREA 2017 Vapor bubbles. *Annual review of fluid mechanics* **49**, 221–248.
- 699 ROWLEY, CLARENCE W 2005 Model reduction for fluids, using balanced proper orthogonal decomposition.
700 *Int. J. Bifurc. Chaos Appl. Sci. Eng.* **15** (03), 997–1013.
- 701 SCHNERR, GÜNTER H & SAUER, JÜRGEN 2001 Physical and numerical modeling of unsteady cavitation
702 dynamics. In *Fourth international conference on multiphase flow*, , vol. 1. ICMF New Orleans New
703 Orleans, LO, USA.
- 704 SUSLICK, KS, DIDENKO, Y, FANG, MM, HYEON, T, KOLBECK, KJ, McNAMARA III, WB, MDLELENI, MM
705 & WONG, M 1999 Acoustic cavitation and its chemical consequences. *Philos. Trans. R. Soc. A*
706 **357** (1751), 335–353.
- 707 TAIRA, KUNIHICO, BRUNTON, STEVEN L, DAWSON, SCOTT TM, ROWLEY, CLARENCE W, COLONIUS, TIM,
708 MCKEON, BEVERLEY J, SCHMIDT, OLIVER T, GORDEYEV, STANISLAV, THEOFILIS, VASSILIOS &
709 UKEILEY, LAWRENCE S 2017 Modal analysis of fluid flows: An overview. *AIAA J* **55** (12), 4013–4041.
- 710 TAKAHIRA, HIROYUKI, AKAMATSU, TERUAKI & FUJIKAWA, SHIGEO 1994 Dynamics of a cluster of bubbles in
711 a liquid: Theoretical analysis. *JSME International Journal Series B Fluids and Thermal Engineering*
712 **37** (2), 297–305.
- 713 TANG, XIN & STAACK, DAVID 2019 Bioinspired mechanical device generates plasma in water via cavitation.
714 *Sci. Adv.* **5** (3), eaau7765.
- 715 VENNING, JAMES A, PEARCE, BRYCE W & BRANDNER, PAUL A 2022 Nucleation effects on cloud cavitation
716 about a hydrofoil. *Journal of Fluid Mechanics* **947**, A1.
- 717 VERHAAGEN, BRAM & RIVAS, DAVID FERNÁNDEZ 2016 Measuring cavitation and its cleaning effect. *Ultrason*
718 *Sonochem* **29**, 619–628.
- 719 YAMASHITA, TATSUYA & ANDO, KEITA 2019 Low-intensity ultrasound induced cavitation and streaming
720 in oxygen-supersaturated water: Role of cavitation bubbles as physical cleaning agents. *Ultrason*
721 *Sonochem* **52**, 268–279.
- 722 YASUI, KYUICHI, IIDA, YASUO, TUZIUTI, TORU, KOZUKA, TERUYUKI & TOWATA, ATSUYA 2008 Strongly
723 interacting bubbles under an ultrasonic horn. *Phys. Rev. E* **77** (1), 016609.
- 724 ZERAVCIC, ZORANA, LOHSE, DETLEF & VAN SAARLOOS, WIM 2011 Collective oscillations in bubble clouds.
725 *J. Fluid. Mech.* **680**, 114.