Regressing bubble cluster dynamics as a disordered many-body system

Kazuki Maeda, Daniel Fuster

Regressing bubble cluster dynamics as a disordered many-body system

Kazuki Maeda[1]†, and Daniel Fuster[2]

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

[2]Sorbonne 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)

$$\langle K \rangle \sim \langle K_s \rangle (1 + B_d), \tag{1.1}$$

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: $B_d = N\langle\overline{R(t)}\rangle/R_C$, 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 ($i = 1, 2, ...N$), respectively. The scaling can be simply derived from the coupled R-P equation for the correlated (synchronized) limit of monodisperse bubbles ($R_1 = R_2 \cdots = R_{b,c}$). Although realistic correlations are imperfect due to polydispersity and nonlienarity, 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 nonliearity 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

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

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

2. Methods

2.1. *Principal component analysis of bubble dynamics data*

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

$$\boldsymbol{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_{Nt}], \tag{2.1}$$

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

$$\boldsymbol{q}_k = [\dot{R}_1(t_k), \dot{R}_2(t_k), \ldots, \dot{R}_N(t_k)]^T. \tag{2.2}$$

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

$$\boldsymbol{r}_k = [R_1(t_k), R_2(t_k), \ldots, R_N(t_k)]^T. \tag{2.3}$$

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

$$\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{X}^*. \tag{2.4}$$

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

$$\boldsymbol{\Pi}_i = \boldsymbol{U}\boldsymbol{\Sigma}_i, \tag{2.5}$$

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

2.2. *Weighted principal component analysis*

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

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

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

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

$$K = \boldsymbol{q}^T\boldsymbol{T}\boldsymbol{q} = (\boldsymbol{W}\boldsymbol{q})^T(\boldsymbol{W}\boldsymbol{q}). \tag{2.7}$$

where

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

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

$$\boldsymbol{T}(\boldsymbol{r}) = \boldsymbol{W}\boldsymbol{W}^*. \tag{2.9}$$

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

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

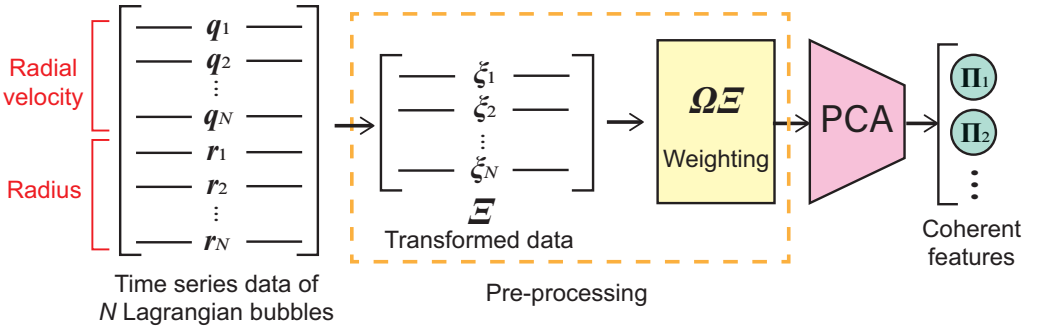


Figure 1: Schematic of the feature extraction by PCA from Lagrangian bubble dynamics data, $\boldsymbol{Q}$, after pre-processing. The variance of the resulting features is consistent with the interaction energy predicted by the coupled Rayleigh-Plesset equation.

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

$$K = \boldsymbol{\xi}^T \boldsymbol{P}(\boldsymbol{\eta})\boldsymbol{\xi} = (\boldsymbol{\Omega\eta})^T(\boldsymbol{\Omega\eta}). \tag{2.10}$$

$\boldsymbol{\Omega}$ is the weight matrix satisfying

$$\boldsymbol{P}(\boldsymbol{r}) = \boldsymbol{\Omega\Omega}^*, \tag{2.11}$$

where

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

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

$$\hat{K} = \hat{\boldsymbol{\xi}}^T \boldsymbol{P}(\hat{\boldsymbol{\eta}})\hat{\boldsymbol{\xi}}. \tag{2.13}$$

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

$$\langle K\rangle \approx \langle\hat{K}\rangle = \langle\hat{\boldsymbol{\xi}}^T \boldsymbol{P}(\hat{\boldsymbol{\eta}})\hat{\boldsymbol{\xi}}\rangle \approx \langle\boldsymbol{\xi}^T \boldsymbol{P}(\langle\boldsymbol{\eta}\rangle)\boldsymbol{\xi}\rangle. \tag{2.14}$$

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

$$\boldsymbol{\Omega\Xi} = \boldsymbol{U}_\xi\boldsymbol{\Sigma}_\xi\boldsymbol{X}_\xi^*. \tag{2.15}$$

The i-th PC is stored in $\boldsymbol{\Pi}_{\xi,i}$: $\boldsymbol{\Pi}_{\xi,i} = \boldsymbol{U}_\xi\boldsymbol{\Sigma}_{\xi,i}$. We denote the i-th largest singular value of $\boldsymbol{\Sigma}_\xi$ as σ_i. The degree of coherence can be measured by the normalized variance of the first PC:

$$\hat{\sigma}_1^2 = \sigma_1^2/\mathrm{tr}(\boldsymbol{\Sigma}^2), \tag{2.16}$$

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

2.3. *Spectral entropy and coherence measure*

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

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

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

where $\ln_2(N)$ is the normalization factor. The spectral entropy of data is a discrete analogue of the Shannon entropy based on the spectrum of data and has been used to quantify the randomness of data in various applications (Kullback 1997; Alter *et al.* 2000; Hu *et al.* 2005; De Domenico & Biamonte 2016). Aubry *et al.* (1991) introduced a similar definition of entropy for spatio-temporal signals of canonical fluid flows from their POD eigenvalues and used the entropy to characterize flow instabilities. In the present context, when bubbles are in perfect correlation, we expect to excite only the first PC capturing the entire energy of the 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 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 define the coherence measure C.

$$C = \frac{B_1}{B_d}, \tag{2.18}$$

where

$$B_1 = \frac{\sigma_1^2}{\sigma_1'^2} - 1. \tag{2.19}$$

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

$$\langle K \rangle \approx \sigma_1^2 \approx (1 + B_d)\sigma_1'^2 \tag{2.20}$$

and

$$B_d \approx B_1. \tag{2.21}$$

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

3. Numerical experiments

3.1. *Data sets*

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

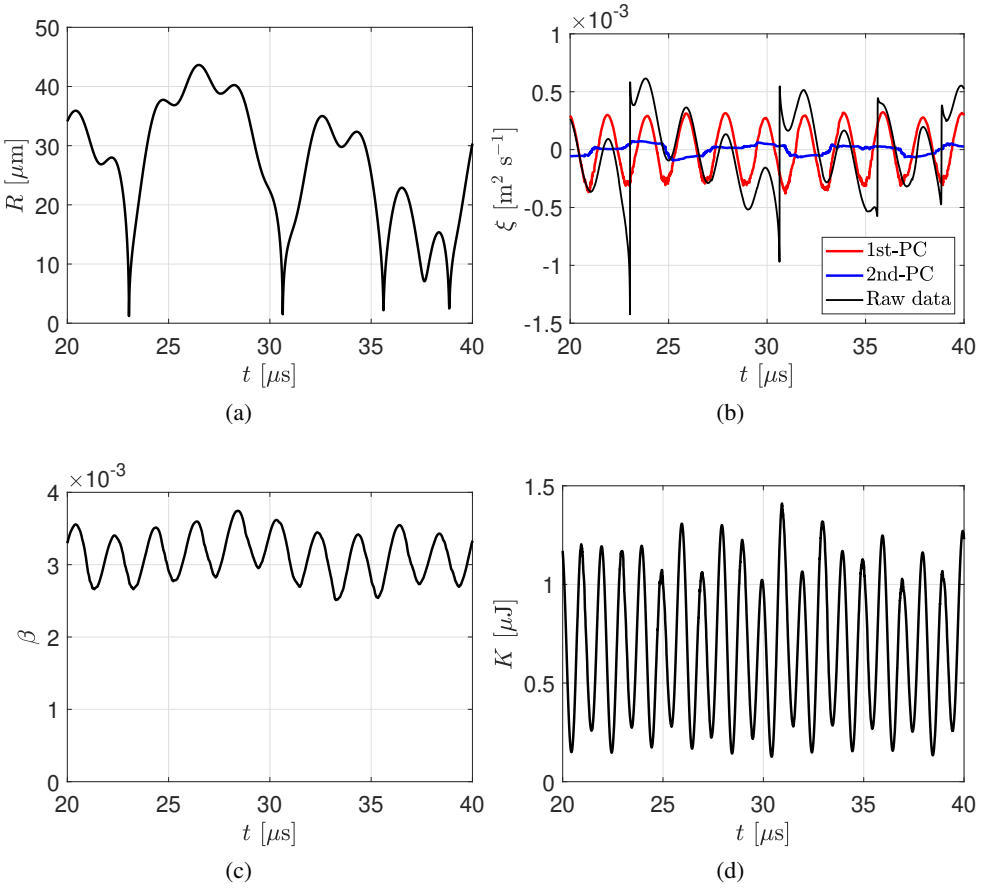
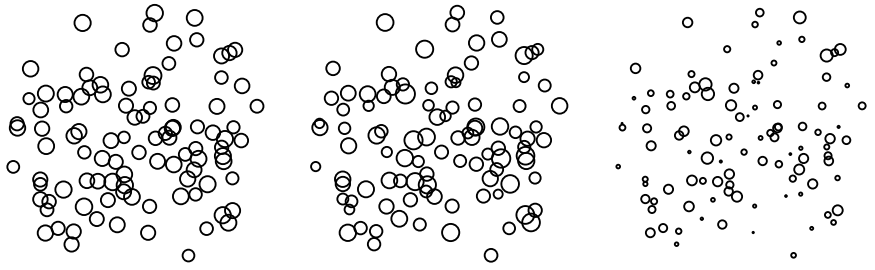


Figure 2: Evolution of representative quantities of a bubble cloud with $(B_0, s_d, N) = (0.5, 0.1, 100)$ excited at $A = 20$, during a stationary state. (a) Radius of a representative bubble. (b) For the same bubble, ξ obtained from the raw data, and those from the first and second dominant features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

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

3.2. *Visualization of representative data*

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



(a) (b) (c)

Figure 3: Projected side-views of the three-dimensional bubble cloud of figure 2The 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.

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

Figure 3(a-c) show the 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. The size of spheres can be interpreted as the mean energy of oscillations of bubbles at those locations in each feature. The overall distribution of the spheres of the

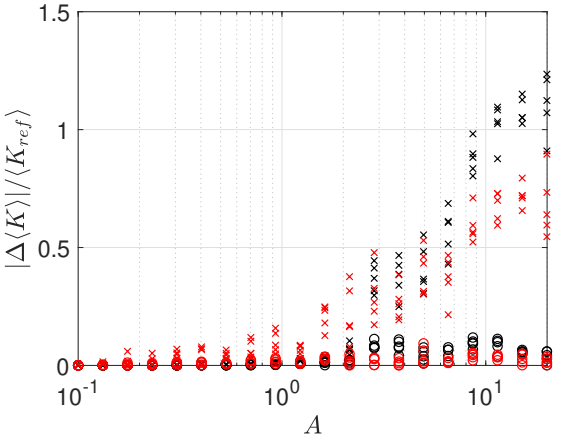


Figure 4: Relative error of the mean kinetic energy of the fluid induced by clusters for the two approximations, (×):$\langle K\rangle \approx \langle \boldsymbol{q}^T \boldsymbol{T}(\langle \boldsymbol{q}\rangle)\boldsymbol{q}\rangle$ and (○):$\langle K\rangle \approx \langle \boldsymbol{\xi}^T \boldsymbol{P}(\langle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

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

3.3. *Error analysis*

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

3.4. *Amplitude dependence of the key measures*

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

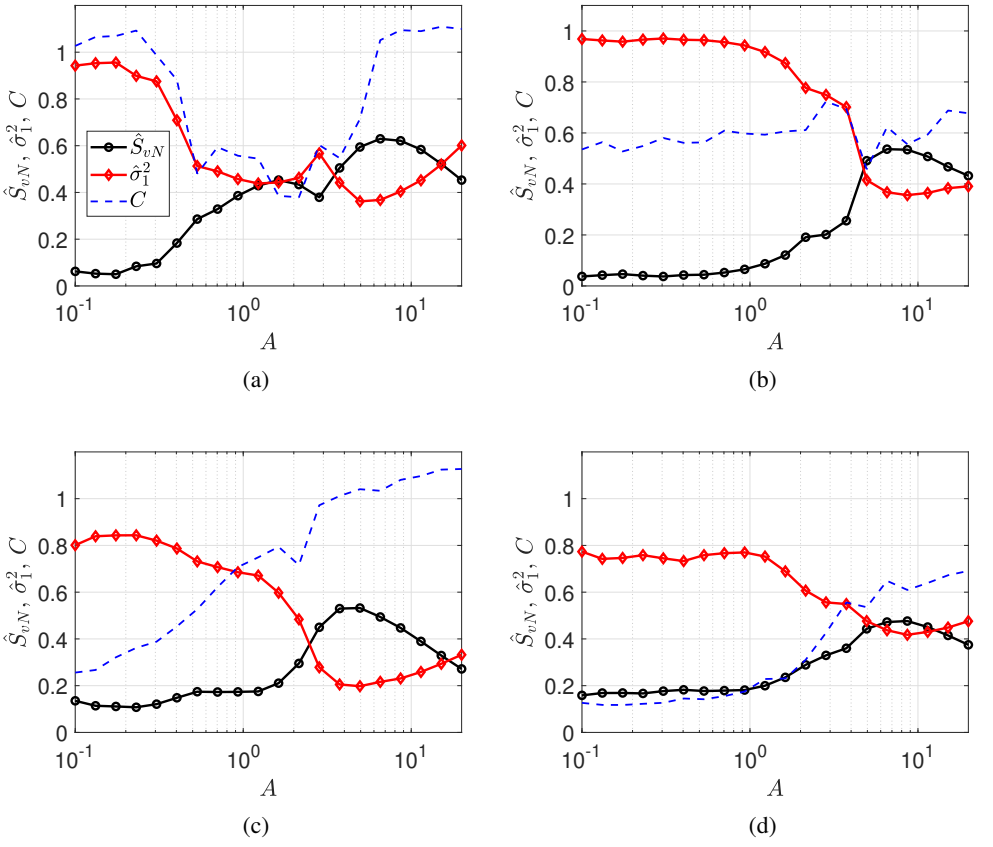


Figure 5: The first PC-variance ($\hat{\sigma}_1^2$), the normalized von Neumann entropy ($\hat{S}_{vN}$), and the coherence measure (C), against the excitation amplitude with various initial density and polydispersity parameters, with (B_0, s_d)= (a) (0.5,0.1), (b) (0.5,0.7), (c) (5.0,0.1), and (d)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

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

B_d. Figure 3(c) and (d) respectively corresponds to sparse and dense clusters with strongly polydisperse nuclei; $(B_0, s_d) = (0.5, 0.7)$ and $(5.0, 0.7)$. These plots show clear differences from those of weekly polydisperse clusters in figure 5(a) and figure 5(b). For the sparse 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 the minimum of 0.2 at $A \approx 0.4$ and then mildly grows to 0.3 at $A = 1.2$. The sharp decay resembles that in figure 5(b), while the slope is milder. $\hat{S}_{vN}$ has also a profile mirrored to $\hat{\sigma}_1^2$, but it is less symmetric than figure 5(a) and figure 5(b). $\hat{S}_{vN}$ is nearly constant around 0.2 at $0.1 < A < 2.0$ and draws a concave curve with its maximum of around 0.5 at $A \approx 5.0$ followed by a smooth decay to 0.3 at $A = 20$. C almost constantly increases from 0.2 to $O(1)$ throughout the plot, indicating that the coherent dynamics is controlled by B_d, only at large A, unlike weakly polydisperse case of figure 5(a). The overall trend of the plots in figure 5(b) is similar to figure 5(a), although the changes of variables against A are milder in figure 5(b).

Overall, the mutual trends of variables at $A > 5$ are similar between figure 5(a) and 5(c) and between figure 5(b) and 5(d), indicating the decreasing influence of the initial polydispersity in the nonlinear regime. At $A < O(1)$, the polydisperse clouds tend to have smaller values of $\hat{\sigma}_1^2$, and larger values of $\hat{S}_{vN}$ and C. This can be explained by the enhancement of incoherent dynamics induced by polydispersity.

3.5. *Principal component spectrum*

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

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

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

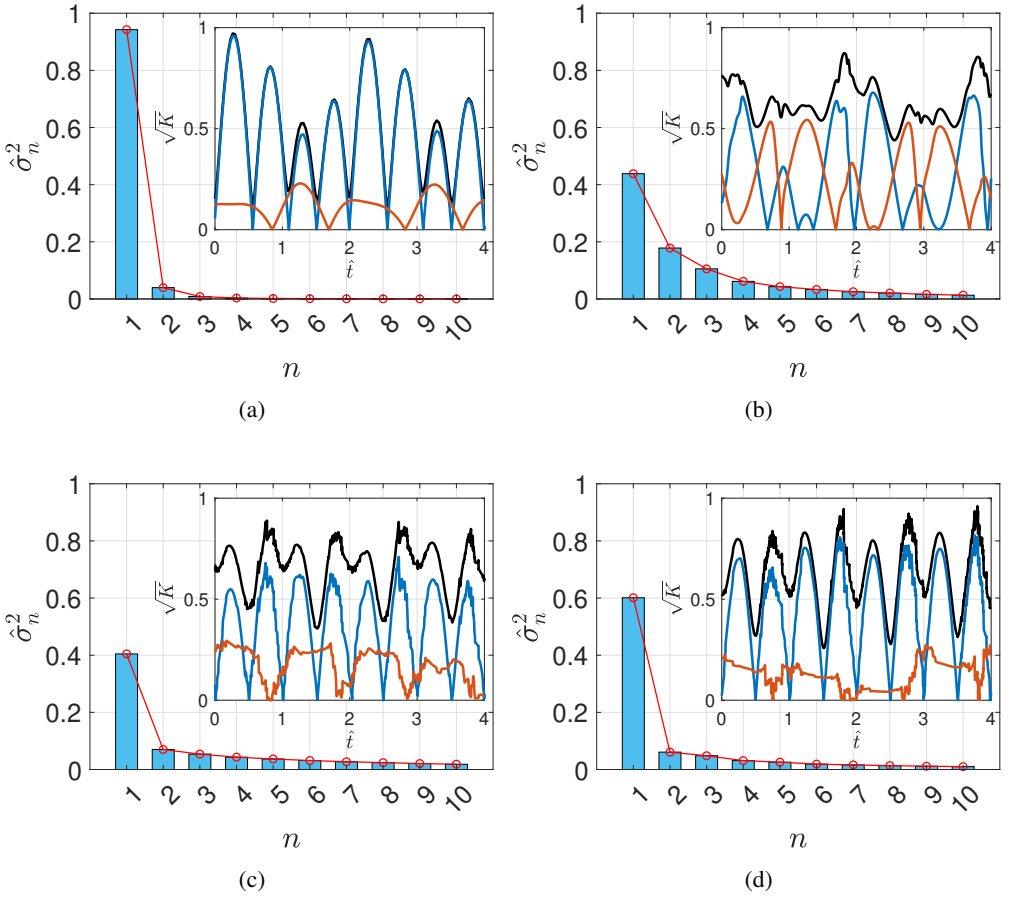


Figure 6: (a-d) The PC spectral profiles at $A = 2 \times 10^{-2}$, 1.2, 8.6, and 20. Insets show the evolution of the square-root of the normalized total energy ($\sqrt{K}$, black) and those of the first (blue) and the second (red) PC-variances ($\sqrt{K_1}$ and $\sqrt{K_2}$), during the four periods 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 $\sqrt{K}$.

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

3.6. *Evaluation of the coherence measure*

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

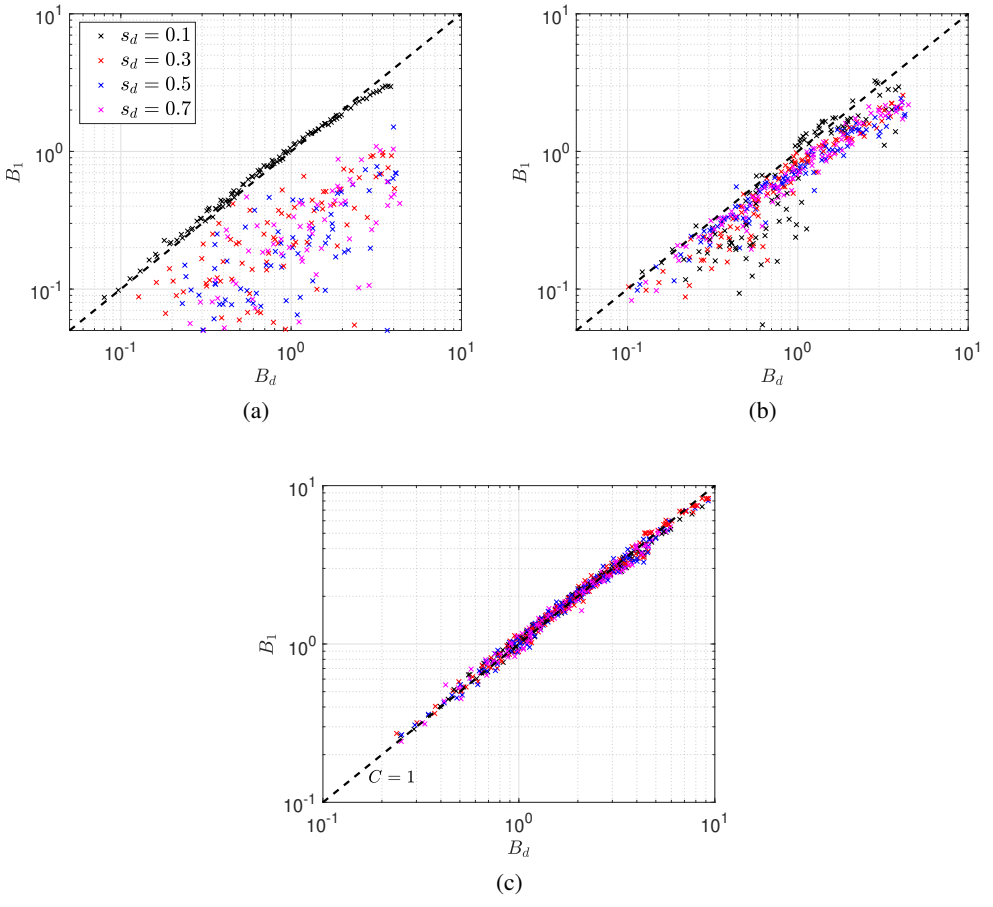


Figure 7: B_1 against B_d, for $O(10^3)$ clusters with various values of s_d, N, R_C, and other physical parameters (Appendix D) in the linear (a: $A = 2 \times 10^{-2}$), transition (b: $A = 1.2$), and nonlinear (c: $A = 20$) regimes.

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

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

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

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

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

5. Conclusion

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

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

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

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

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

Author ORCID.
Kazuki Maeda https://orcid.org/0000-0002-5729-6194;
Daniel Fuster https://orcid.org/0000-0002-1718-7890.

Appendix A. Variable transformation for a single bubble

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

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

$$\dot{\eta} = \frac{\xi}{\eta}. \tag{A 2}$$

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

$$K = 2\pi\rho\dot{R}^2R^3 = 2\pi\rho\xi^2\eta. \tag{A 3}$$

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

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

$$\dot{\hat{\eta}} = \frac{\hat{\xi}}{\hat{\eta}}. \tag{A 5}$$

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

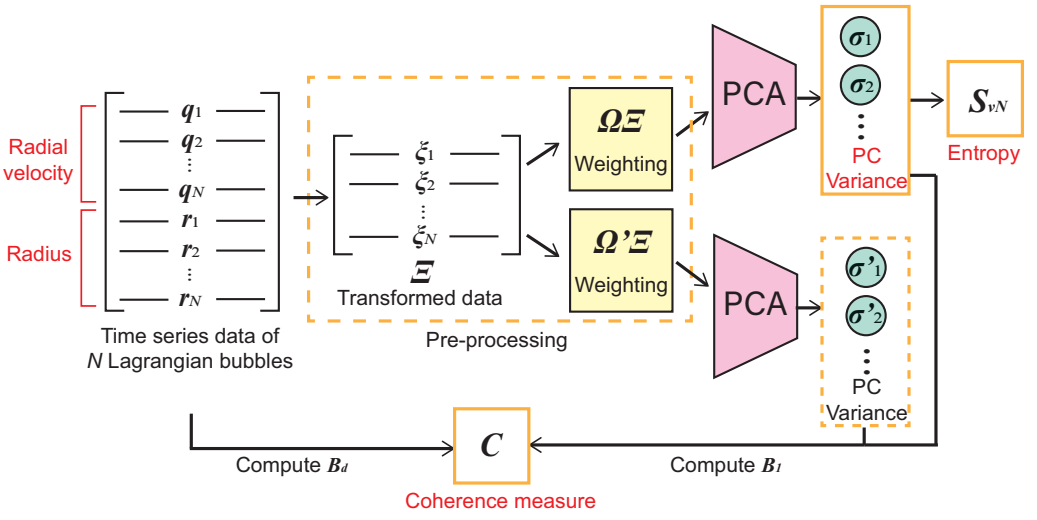


Figure 8: Comprehensive schematic of the present procedure to obtain the spectral entropy, PC-variance, and coherence measure, from input data.

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

We can express the instantaneous energy of this approximate system as

$$\hat{K} = 2\pi\rho\hat{\xi}^2\hat{\eta}. \tag{A 6}$$

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

$$\langle K\rangle \approx \langle\hat{K}\rangle = 2\pi\rho\langle\hat{\xi}^2\hat{\eta}\rangle \approx 2\pi\rho\langle\hat{\xi}^2\rangle\langle\eta\rangle. \tag{A 7}$$

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

Appendix B. Schematic of the input-output procedure

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

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

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

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

$$K' = \boldsymbol{\xi}^T \boldsymbol{P}'\boldsymbol{\xi}, \tag{B 2}$$

where

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

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

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

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

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

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

and

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

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

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

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

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

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

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

The enthalpy and the potential derivative are approximated as

$$H_i \approx \frac{p_i - p_0}{\rho_{l,0}}. \tag{C 7}$$

$$\frac{\partial \phi_\infty}{\partial t} \approx \frac{p_\infty - p_0}{\rho_{l,0}}. \tag{C 8}$$

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

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

$$d_{ij} - R_j \ll \lambda_c. \tag{C 9}$$

Dividing both sides with c,

$$\frac{d_{ij} - R_j}{c} = t - t' \ll \frac{\lambda_c}{c} = \frac{1}{f_c} = T_c, \tag{C 10}$$

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

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

and

$$I^* = \tag{C 12}$$

$$-\sum_{j \neq i}^{N} \frac{R_j}{d_{ij}}\left(\frac{1}{2}\dot{R}_j^2 + \ddot{R}_j\left(R_j\left(1 - \frac{\dot{R}_j}{c}\right)\right) + \frac{3}{2}\dot{R}_j^2\left(1 - \frac{\dot{R}_j}{3c}\right) - \frac{p_j - p_\infty}{\rho_l}\frac{\dot{R}_j}{c} - \frac{R_j}{\rho_l c}\frac{\partial (p_j - p_\infty)}{\partial t}\right). \tag{C 13}$$

For harmonic excitation,

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

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

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

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

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
128	0.33-2.0	10	0.1-0.7	500
256	0.33-2.0	10	0.1-0.7	500
512	0.8-2.0	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

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
128	0.33-2.0	10	0.1-0.7	500

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

Appendix D. Summary of the parameters

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

Appendix E. Scaling of the second PC

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

REFERENCES

Abdi, Hervé & Williams, Lynne J 2010 Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2** (4), 433–459.

Alter, Orly, Brown, Patrick O & Botstein, David 2000 Singular value decomposition for genome-wide expression data processing and modeling. *Proceedings of the National Academy of Sciences* **97** (18), 10101–10106.

Ando, K, Liu, A-Q & Ohl, C-D 2012 Homogeneous nucleation in water in microfluidic channels. *Phys. Rev. Lett.* **109** (4), 044501.

Angélil, Raymond, Diemand, Jürg, Tanaka, Kyoko K & Tanaka, Hidekazu 2014 Bubble evolution and properties in homogeneous nucleation simulations. *Phys. Rev. E.* **90** (6), 063301.

Arndt, Roger EA 1981 Cavitation in fluid machinery and hydraulic structures. *Annual Review of Fluid Mechanics* **13** (1), 273–326.

Aubry, Nadine, Guyonnet, Régis & Lima, Ricardo 1991 Spatiotemporal analysis of complex signals: theory and applications. *Journal of Statistical Physics* **64**, 683–739.

Blander, M & Katz, JL 1975 Bubble nucleation in liquids. *AIChE Journal* **21** (5), 833–848.

Bremond, N, Arora, M, Ohl, CD & Lohse, D 2006 Controlled multibubble surface cavitation. *Phys. Rev. Lett.* **96** (22), 224501–224501.

Brennen, CE 2014 *Cavitation and bubble dynamics*. Cambridge University Press.

Cairós, C & Mettin, R 2017 Simultaneous high-speed recording of sonoluminescence and bubble dynamics in multibubble fields. *Phys. Rev. Lett.* **118** (6), 064301.

Chen, Yonglian & Heister, Stephen D 1994 A numerical treatment for attached cavitation. *Journal of fluids engineering* **116** (3), 613–618.

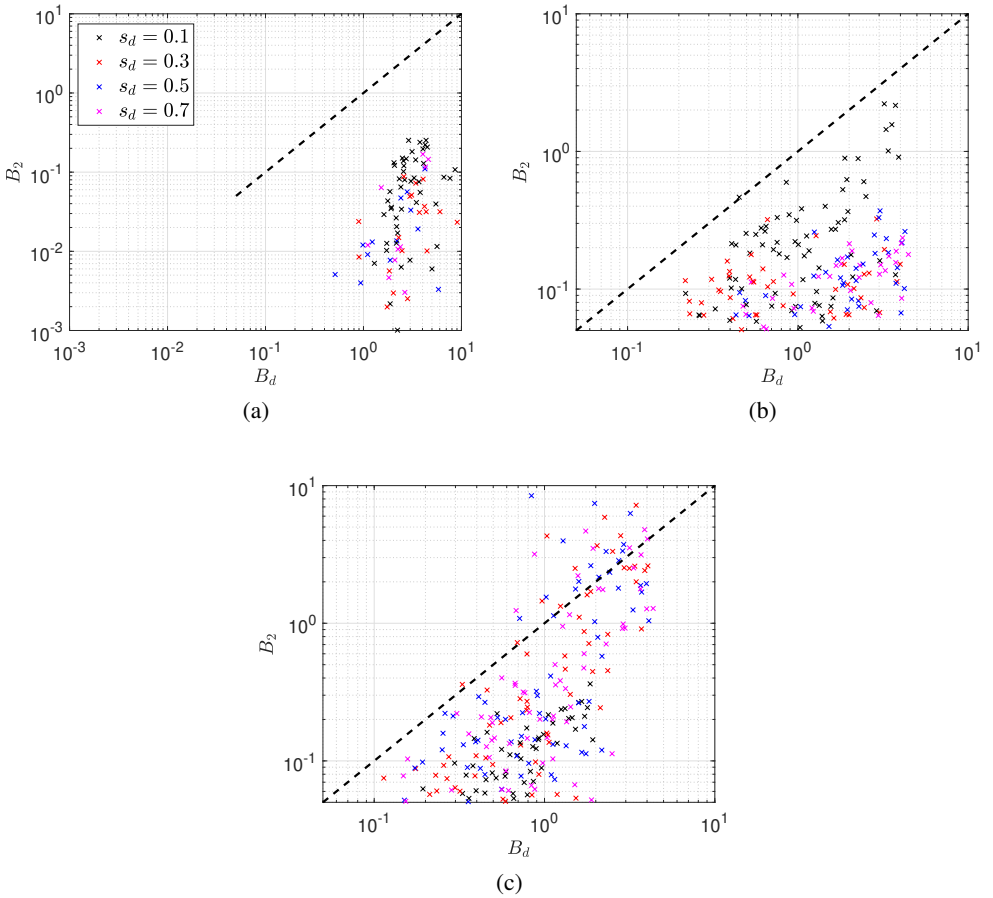


Figure 9: B_2 against B_d, for $O(10^3)$ clusters with various values of s_d, N, R_C, and other physical parameters (Appendix D) in the linear (a: $A = 2 \times 10^{-2}$), transition (b: $A = 1.2$), and nonlinear (c: $A = 20$) regimes.

d'Agostino, L & Brennen, CE 1989 Linearized dynamics of spherical bubble clouds. *J. Fluid. Mech.* **199**, 155–176.

De Domenico, Manlio & Biamonte, Jacob 2016 Spectral entropies as information-theoretic tools for complex network comparison. *Physical Review X* **6** (4), 041062.

Doinikov, Alexander A 2004 Mathematical model for collective bubble dynamics in strong ultrasound fields. *The Journal of the Acoustical Society of America* **116** (2), 821–827.

Fuster, Daniel & Colonius, Tim 2011 Modelling bubble clusters in compressible liquids. *J. Fluid Mech.* **688**, 352–389.

Gallo, Mirko, Magaletti, Francesco & Casciola, Carlo Massimo 2021 Heterogeneous bubble nucleation dynamics. *J. Fluid Mech.* **906**.

Ganesh, Harish, Mäkiharju, Simo A & Ceccio, Steven L 2016 Bubbly shock propagation as a mechanism for sheet-to-cloud transition of partial cavities. *Journal of Fluid Mechanics* **802**, 37–78.

Gnanaskandan, Aswin & Mahesh, Krishnan 2016 Large eddy simulation of the transition from sheet to cloud cavitation over a wedge. *International Journal of Multiphase Flow* **83**, 86–102.

Hashimoto, Tomoyuki, Yoshida, Makoto, Watanabe, Mitsuo, Kamijo, Kenjiro & Tsujimoto, Yoshinobu 1997 Experimental study on rotating cavitation of rocket propellant pump inducers. *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, Yoshizawa, 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 & Miksis, 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, Kunihiko, 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.