Supporting Information

Expansion dynamics of supercritical water probed by picosecond time-resolved photoelectron spectroscopy

Thomas Gladytz[a,b], Bernd Abel[a,b] and Katrin R. Siefermann[a,*]

[a] Leibniz-Institute of Surface Modification (IOM), Chemistry Department, Permoserstr. 15, 04318 Leipzig, Germany

[b] University of Leipzig, Wilhelm-Ostwald-Institute for Physical and Theoretical Chemistry, Linnéstr. 3, 04318 Leipzig, Germany

* corresponding author: katrin.siefermann@iom-leipzig.de

S1 Data pretreatment

Time-of-flight (t_{fl}) photoelectron spectra are converted to the binding energy scale with the calibration function E1. For each individual spectrum the position of the $1b_{1,g}$ signal was fixed to the literature value of 12.6 by adding or substracting a binding energy offset $E_{bind,offset}$.

$$E_{Bind}(eV) = -2450000 \cdot t_{fl}^{-2.06} + E_{bind,offset} \qquad \text{(E1)}$$

After calibration the spectra were interpolated to a constant and equidistant energy scale for further analysis with the noise-corrected target transform fitting (cTTF) algorithm.

Fluctuations in the signal intensity are removed by normalizing all spectra to a constant total integral.

S2 Static photoelectron spectra

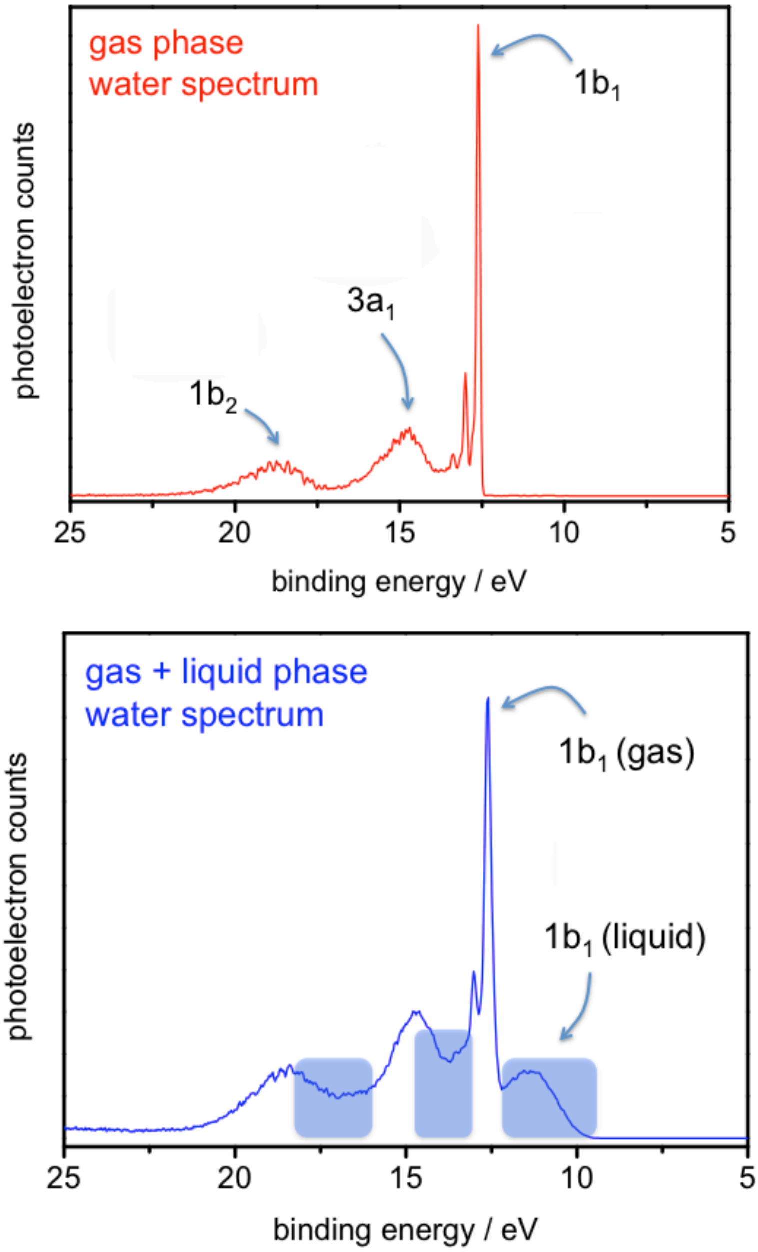

Figure S1: Photoelectron spectrum of gas phase water (red) and of the liquid water microjet (blue). The latter contains contribution from gas and liquid signal. Binding energy regions with liquid signal contribution are shaded in blue. Photoelectron bands are labeled according to the orbitals the photoelectrons originate from. Only the $1b_1$ signal of the liquid spectrum is clearly separated from the gas phase spectrum. Therefore, this signal is particularly useful to track the expansion dynamics of water.

S3 Principal Component Analysis (PCA)

If there is a high degree of correlation among the spectra of a dataset, Principal Component Analysis (PCA)[1] can be used to identify this correlation in order to reduce complexity in the representation of the dataset: A typical spectrum consists of *n* intensity values which are observed at *n* values of an investigated variable x (wavelength, frequency, binding energy). As so, each spectrum can be represented by a single point in an *n*-dimensional space. Every axis of this space represents the intensity at one value of the variable x (e.g. at one wavelength x_i). Due to the high correlation of the different intensity values, all data is found in a small, low-dimensional subspace of the *n*-dimensional coordinate system. PCA finds this subspace (the PC-space) by transforming the strongly correlated intensity values into non-correlated orthogonal variables which are called principal components. The principle components are typically combined as columns of a matrix ***U***. A given spectrum t can be projected into the PC-space using equation E2.

$$\boldsymbol{t_{proj}} = \boldsymbol{U}(\boldsymbol{U^T}\boldsymbol{U})^{-1}\boldsymbol{U^T}\boldsymbol{t} \qquad \text{(E2)}$$

with: $\boldsymbol{t}$ .... column vector containing a given spectrum

$\boldsymbol{t_{proj}}$ ..... column vector of the projection of the given spectrum into the PC-space

$\boldsymbol{U}$ .... matrix containing the significant PCs of the dataset as columns

S4 Determination of r_{noise} for noise-corrected target transform fitting (cTTF)

The following sections describe how r_{noise} can be quantified. As illustrated in Figure 3 r_{noise} consists of contributions from the noise r_j on any principal component, the coordinates k_j in the PC-space and the noise r_{origin} at the origin of the PC-space. Assuming that the noise r_j is uncorrelated to the noise on the other PCs and independent from the noise r_{origin}, the noise contributions will (almost perfectly) add up orthogonally and the noise of an arbitrary point in PC-space can be approximated using equation E3:

$$r_{noise} = \sqrt{r_{origin}^2 + \sum_j (k_j r_j)^2} \qquad \text{(E3)}$$

with: r_{noise} ... noise deviation contained in an arbitrary point in PC space

r_{origin} .. noise contribution to the mean spectrum (i.e. the origin of the PC-space)

r_j .... noise contribution on the j-th principal component

k_j .... coordinate on the j-th principal component

As can be seen from equation E3, r_{noise} grows with increasing contribution k_j of a noisy PC, that is with increasing distance from the origin of the PC-space. This is particularly important for components having a low abundance in the measured spectra, which are expected to lie far away from the origin of the PC-space. (Note that in order to have the origin of the PC-space close to the spectra, the dataset is usually mean-centred, i.e. the mean spectrum is subtracted from all spectra.)

Also, for models with many parameters, which are expected to give a flat optimum, this effect will contribute significantly to the deviation of the TTF results from the true results. The described effect can also be seen in the datasets published by Jandanklang et. al.[2], in which measured data and TTF results of an HPLC analysis of an aqueous salt mixture were compared.

A method to quantify the amount of noise r_j on all principal component vectors was proposed by Shabalin and Nobel[3] based on the Marčenko-Pastur-law[4]. The Marčenko-Pastur-law is a distribution which very universally describes the singular values of matrices containing noise. It was developed for *independent identically distributed* (iid) Gaussian noise. iid-noise is present when every point in a spectrum contains noise of the same variance and the noise of individual points is not correlated. However, in photoelectron spectroscopy and many other spectroscopic techniques counting detectors are used. In these cases the noise is proportional to the square root of the intensity and is therefore not identically distributed. In this case, a modification of the Marčenko-Pastur-law can be used to describe non-iid noise and is therefore very useful to understand the noise pattern of a given dataset. With the modified Marčenko-Pastur-law the model of Shabalin and Nobel[3] can be extended to non-iid noise. The following section will:

1. Summarize the basics of the Marčenko-Pastur-law which describes the singular values of iid noise[4]
2. show how this principle can be extended to datasets with non-iid noise
3. present the model of Shabalin and Nobel[3] and show how this can be used to quantify how the directions of the principal component vectors are influenced by noise.

In 1967 Marčenko and Pastur found that the singular values of a matrix with iid Gaussian random variables follow the probability density function shown in equation E4 [4].

$$p_c(\lambda) = \begin{cases} \frac{1}{2\pi c\sigma^2\lambda}\sqrt{(\lambda - a)(b - \lambda)} & : a \leq \lambda \leq b \\ 0 & : otherwise \end{cases} \tag{E4}$$

$a = \sigma^2\left(1 - \sqrt{c}\right)^2$ and $b = \sigma^2\left(1 + \sqrt{c}\right)^2$
λ .... singular value of a *m*-by-*n* random matrix
c ....$c=m/n$; aspect ratio of the random matrix
σ^2 ....variance of the random variables

This probability density function became known as the Marčenko-Pastur law and is fundamental in the field of random matrix theory of non-square matrices.

From the Marčenko-Pastur law a predictant for the scree plot of a pure random matrix can be derived (Figure S2A). From the perspective of random matrix theory the scree plot of a random matrix is a sorted collection of *N* samples from the Marčenko-Pastur-distribution, where *N* is the smaller value of *m* and *n*[5]. On average this collection of samples will follow the *inverse Cumulative Distribution Function* (iCDF) of the Marčenko-Pastur law which has no analytical expression but can be computed numerically from the probability density function (equation E4). Figure S2A shows the scree plot of a random noise matrix and its corresponding Marčenko-Pastur iCDF.

In case of a dataset matrix with signal and iid noise, the signal-free singular values follow the Marčenko-Pastur law. As can be seen in the scree plot in figure S2B, the signal-free singular values are those beyond the significant dimensions caused by the signal. (In the example of figure S2B for principal component number N > 3.)

In many real datasets, however, the noise is not independent identically distributed (iid). As a result, the singular values in the scree plot deviate from the iCDF.

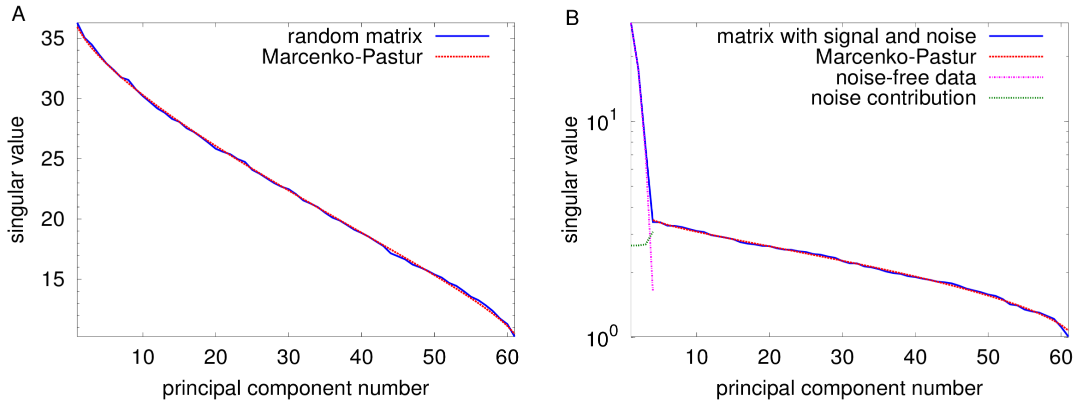

Figure S2: A: Scree-plot of a matrix containing Gaussian noise together with the corresponding Marčenko-Pastur iCDF; B: Scree-plot of a matrix containing a 3-component signal and Gaussian noise (see text for details).

As a consequence, also the aspect ratio *c* of the matrix changes to the aspect ratio *c*=m*/n* of the independent noise. The values of the two fit parameters (height of the noise σ and number of independent noisy points *m**) can further be used as input information for the model of Shabalin and Nobel[3] which describes the noise r_j on the signal singular vectors $\mathbf{u}_j$ (Figure 3, main text). This model delivers an analytic expression for the deviation of the singular vectors $\mathbf{u}_j$*(experimental)* of a data matrix **Y** containing signal and iid noise, from the singular vectors $\mathbf{u}_j$*(noise free)* of a data matrix **A** containing only the signal (eq. E5):

$$\langle \mathbf{u}_j(\mathbf{Y}); \mathbf{u}_j(\mathbf{A}) \rangle^2 = \frac{1 - \frac{c^* \sigma^4}{\lambda_j^4(\mathbf{A})}}{1 + \frac{c^* \sigma^2}{\lambda_j^2(\mathbf{A})}} = \frac{\lambda_j^4(\mathbf{A}) - c^* \sigma^4}{\lambda_j^4(\mathbf{A}) + \lambda_j^2(\mathbf{A}) c^* \sigma^2} \tag{E5}$$

A ....signal data matrix of size $m \times n$
Y .... signal and noise data matrix of size $m \times n$
*c**.........c*=*m*/n*
(for iid noise, c* = c = m/n, aspect ratio of the matrices **A** and **Y**)
λ_j(**A**) ... jth singular value of matrix **A**
σ^2 ....variance of the random variables

Further, the singular values λ_j of the noise-free matrix **A**, which are usually slightly smaller than the known singular values of the real data matrix, can be estimated using expression E5[3].

$$\hat{\lambda}_j^2(\mathbf{A}) = \frac{1}{2}\left[\lambda_j^2(\mathbf{Y}) - \sigma^2(1 + c^*) + \sqrt{\left(\lambda_j^2(\mathbf{Y}) - \sigma^2(1 + c^*)\right)^2 - 4c^* \sigma^4}\right] \tag{E6}$$

Using the law of cosines $\boxed{c^2 = a^2 + b^2 - 2ab\cos\gamma}$ and the fact that the singular vectors are normalized to a length of 1, the dot product of equation E5 can be transformed to the squared deviation of the two vectors as shown in equation E7.

$$r_j^2 = \left|u_j(\mathbf{Y}) - u_j(\mathbf{A})\right|^2 = 2 - 2\sqrt{\frac{\lambda_j^4(\mathbf{A}) - c^* \sigma^4}{\lambda_j^4(\mathbf{A}) + \lambda_j^2(\mathbf{A}) c^* \sigma^2}} \tag{E7}$$

The deviation r_j of the direction of principal components is further used in equation E3.

The second variable that is needed to calculate r_{noise} for every point of the PC-space (see equation E3) is the noise on the origin of the PC-space r_{origin}. This quantity can also be calculated from the Marčenko-Pastur fit of the scree plot (equation E8).

$$r_{origin}^2 = c^* \sigma^2 \qquad \text{(E8)}$$

Using equations E3, E7 and E8 it is now possible to quantify the amount of noise r_{noise} for every point of the spectrum. As shown in equation 1 (main text) the bias of the TTF results is corrected by normalizing the determined residual *d* of the projection of the model spectrum by the amount of noise r_{noise} on the projection.

S4.1 Application of cTTF to simulated data

Here, we demonstrate the principle and the power of the implemented noise correction method for an artificial dataset. The dataset shows simulated time-resolved absorption spectra of a reaction A --> B --> C. The spectra of the pure components of A, B and C consist of Gauss shaped peaks with given positions and full widths at half maxima. The spectra of A, B and C overlap significantly. Additionally, the maximum abundance of B is rather low. It is therefore challenging to extract the concentration profiles and the spectra of pure components unambiguously. Figure S3 summarizes the important features of the given data set.

In order to extract the spectra of the pure components from the dataset we applied TTF with and without noise correction. The results are shown in figure S4.

For component A and C which appear with high abundances in the dataset, the result of the uncorrected TTF method is completely satisfactory. As these two components are highly abundant in the simulated dataset, their spectrum is very close to the origin of the PC-Space. Accordingly, the influence of the noise can be neglected. However the situation is different for compound B. As shown in figure S3B its maximum abundance is only 30%. That means the dataset only contains linear combinations of A, B and C with major contributions of A and/or C. Accordingly, the spectrum of the pure component B is located far from the origin of the PC-space. As explained above, the influence of noise increases drastically with increasing distance from the centre of the PC-space. Therefore the result of the uncorrected

TTF method deviates significantly from the noise corrected results, as can be seen in figure S4B. For this simulated dataset, the results of the noise corrected cTTF and the uncorrected TTF method can be compared with the true pure components. The results are given in Table S1.

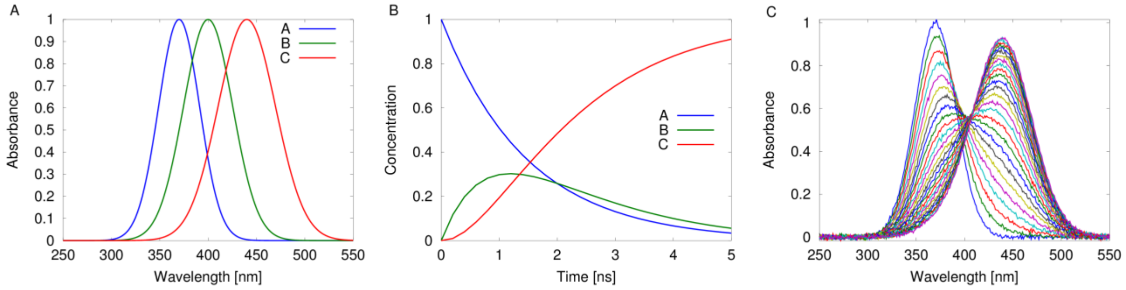

Figure S3: Simulated dataset of a reaction A --> B --> C traced with time resolved absorption spectroscopy. A: Absorption Spectra of the three simulated species A, B and C; B: Concentration profiles of the three simulated species; C: Dataset of time resolved absorption spectroscopy of A --> B --> C.

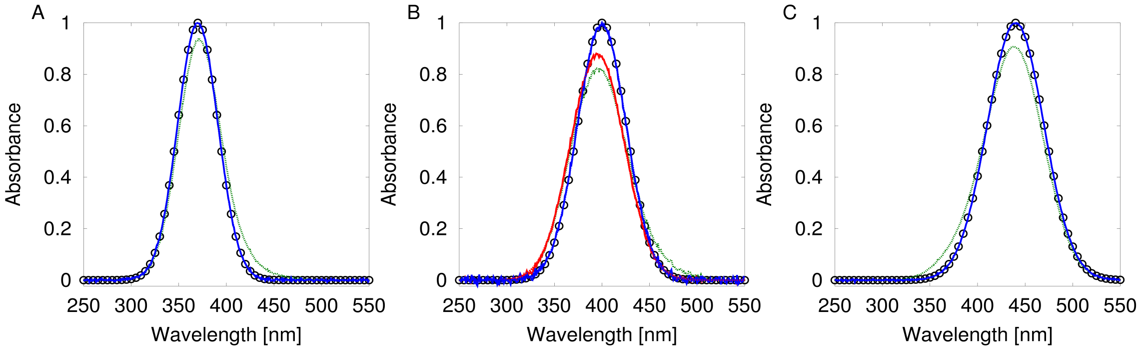

Figure S4: True pure spectra (black open circles) of educt A, intermediate B and product C together with TTF results (red), cTTF results (blue) as well as the initial estimates for the TTF methods (green); For the intermediate B the TTF result deviates from the true spectrum, whereas cTTF retrieves the correct solution.

Table S1: Comparision of the corrected and uncorrected results with input data. Uncorrected TTF results deviate from the true input values for the low abundant compound B. These strong deviations are almost completely removed by the introduction of the noise correction method.

	Input	TTF result		cTTF result	
	peak position [nm]	peak position [nm]	error [nm]	peak position [nm]	error [nm]
Component A	370.0	370.1	0.1	370.0	0.0
Component B	400.0	395.3	-4.7	399.8	-0.2
Component C	440.0	439.8	-0.2	440.0	0.0
	FWHMs [nm]	FWHMs [nm]	error [nm]	FWHMs [nm]	error [nm]
Component A	50.0	50.1	0.1	49.9	-0.1
Component B	60.0	66.4	6.4	60.3	0.3
Component C	70.0	70.3	0.3	70.0	0.0

S5 Identification of the number of significant principal components

The number of significant principal components (PCs) was determined by visual inspection of the projections of the measured spectra into the PC space. It was found that 3 PCs are sufficient to represent all features present in the dataset.

Figure S5 illustrates that a 4th PC does not further improve the representation of the dataset; it shows the spectrum with the highest contribution of the 4th PC together with its projection into a 3-dimensional PC space (A) and into a 4-dimensional PC space (B). This illustrates that the spectrum is well represented in a 3-dimensional PC space, and that 3 PCs are sufficient to represent the dataset.

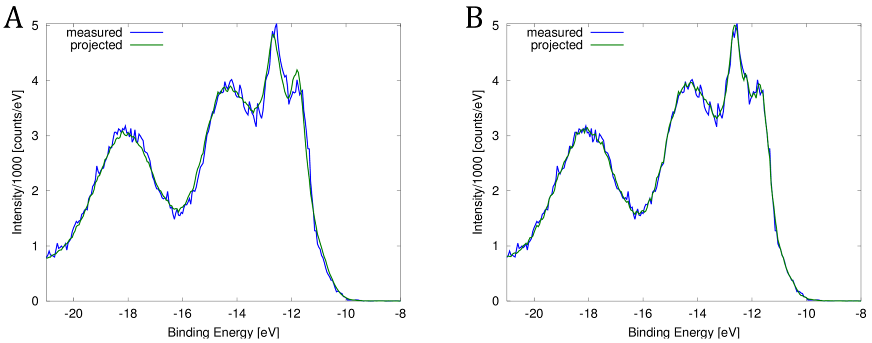

Figure S5: Spectrum with the highest contribution of the 4th PC (blue) projected into a 3-dimensional (A) and a 4-dimensional (B) PC-space.

S6 Determination of initial estimates for pure component spectra with ITTFA

cTTF requires initial estimates for the spectra of pure components (here: liquid water, gaseous water, small clusters, large clusters). These were obtained with ITTFA[6]. To minimize the bias of the ITTFA solution coming from the position of the starting delta-function, the whole $1b_1$ region (Ebind = 10 eV - 13.2 eV) was scanned (figure S6A). The obtained ITTFA spectra cluster into four different spectra which are indicated by vertical lines in figure S6A and shown in figure S6B. The ITTFA spectra suffer from overextrapolation, as some spectra show dips at positions where other spectra have maxima. Nevertheless, the results span the range of possible solutions and are thus good start spectra for the subsequent cTTF analysis.

Figure S6 shows the projection of the datasets (blue and green) and the ITTFA estimates into the PC1-PC2-plane. The ITTFA estimates are represented by open circles and frame the measured datasets. This indicates that they are good starting spectra to determine the pure component spectra with cTTF, as the pure component spectra will also frame the datasets.

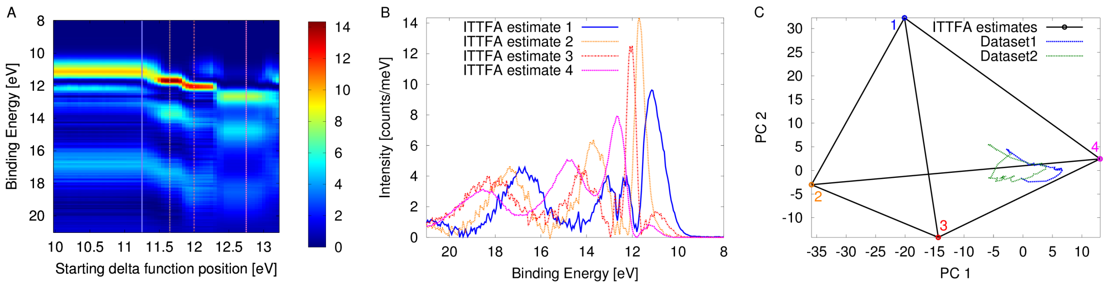

Figure S6: A: ITTFA spectra as a function of different start positions of the delta function between 10 eV and 13.2 eV. The four resulting spectra are indicated by vertical lines. B: The four spectra obtained with ITTFA. C: Projection of the datasets (blue and green) and the ITTFA estimates into the PC1-PC2-plane.

S7 Equations describing the expansion of supercritical water into vacuum

In our simulation, water is approximated as an ideal gas with an adiabatic exponent γ of 1.33. The surface of the water jet is assumed to be flat on the length scale the particles travel during the ultrafast experiment. Heat flow and diffusion are neglected. In other words the expansion is adiabatic and the adiabatic relations for ideal gases apply (E9-11).

$$pV^{\gamma} = const. \tag{E9}$$

$$TV^{\gamma-1} = const. \tag{E10}$$

$$T^{\gamma}p^{1-\gamma} = const. \tag{E11}$$

p …… pressure of the ideal gas
V …… molar volume of the ideal gas
T …… temperature of the ideal gas
γ …… adiabatic exponent of the ideal gas

The speed of sound in the whole system is approximated by E12 as known for ideal gases.

$$a = \sqrt{\gamma \frac{RT}{M}} \tag{E12}$$

γ …… adiabatic exponent of the ideal gas
R …… universal gas constant
T …….temperature of the ideal gas
M …..molar mass of the ideal gas

We expect the ideal gas approximation to underestimate the speed of sound and with this the speed of decompression in the very dense liquid jet. However, the reliability of the ideal gas approximation will improve substantially upon decompression to values around the critical density and below, and should thus describe the metastable gas phase before nucleation reasonably well.

The flow of gas during expansion is assumed to be unhindered by dissipative forces like friction and therefore neglects viscosity. This seems reasonable as the expansion is not hindered by any wall or boundary, which might exert such a force. With this, we have the case of an unsteady inviscid isentropic one-dimensional compressible flow[7].

Due to the pressure drop at the surface of the water jet the molecules are accelerated into an empty space. With conservation of momentum and mass, equation E13 is obtained for the relationship between the pressure drop over an expansion wave and the acceleration of the gas flow[7].

$$dp = -a\rho du \tag{E13}$$

dp … pressure drop at an expansion wave
du … change in flow speed at the expansion wave
a …… speed of sound at the expansion wave
ρ …… density at the expansion wave

With the adiabatic relations for ideal gases (E9-11) and E12, equation E13 can be transformed to E14[7].

$$du = -\frac{2}{\gamma-1}da \tag{E14}$$

du … change in flow speed at the expansion wave
γ …… adiabatic exponent of the ideal gas
da … change in speed of sound due to the temperature drop at the expansion wave

Integration of E14 with the starting flowspeed $u_0=0$ yields:

$$u = \frac{2}{\gamma-1}(a_0 - a) \tag{E15}$$

u ……. flow speed of the fluid in the expansion fan
γ ……. adiabatic exponent of the ideal gas
a ……. speed of sound in the expansion fan
a_0 ….. speed of sound in the unperturbed dense fluid

The expansion waves form at the surface of the water jet upon excitation and travel into the water jet at the speed of sound a. This motion of the waves is superimposed on the laminar flow of the fluid into the vacuum. The position x of an expansion wave relative to the position of the former surface at any time t after the excitation can be expressed by:

$$x = (u - a)t \tag{E16}$$

x …….distance an expansion wave has traveled away from the former surface of the liquid jet
u …….flow speed of the fluid at the expansion wave
a …….speed of sound at the expansion wave
t ……..time elapsed since the start of the expansion

Solving for u, equating with E15 and solving for the speed of sound a yields:

$$\frac{a}{a_0} = \frac{2}{\gamma+1}\left(1 - \frac{\gamma-1}{2}\frac{x}{a_0 t}\right) \tag{E17}$$

a …….speed of sound in the expansion fan
γ …….adiabatic exponent of the ideal gas
a_0 ……speed of sound in the unperturbed dense fluid

x/t ….position in space and time relative to the surface at the beginning of the expansion

With E9-11 the speed of sound can be converted to all thermodynamic variables:

$$\frac{a}{a_0} = \sqrt{\frac{T}{T_0}} = \left(\frac{p}{p_0}\right)^{\frac{\gamma-1}{2\gamma}} = \left(\frac{\rho}{\rho_0}\right)^{\frac{\gamma-1}{2}} \tag{E18}$$

a …….speed of sound in the expansion fan
γ …….adiabatic exponent of the ideal gas
a_0 ……speed of sound in the unperturbed dense fluid
x/t ….position in space and time relative to the surface at the beginning of the expansion

References

1. S. Wold, K. Esbensen and P. Geladi, *Chemom. Intell. Lab. Syst.*, 1987, **2**(1), 37.
2. P. Jandanklang, M. Maeder and A. C. Whitson, *J. Chemometrics*, 2001, **15**(6), 511.
3. A. A. Shabalin and A. B. Nobel, *J. Multivariate Analysis*, 2013, **118**(0), 67.
4. V. A. Marčenko and L. A. Pastur, *Sbornik: Mathematics*, 1967, **1**(4), 457.
5. A. M. Tulino and S. Verdú, *Random matrix theory and wireless communications*, Now Publishers Inc, 2004.
6. B. G. M. Vandeginste, W. Derks and G. Kateman, *Anal. Chim. Acta*, 1985, **173**(0), 253.
7. H. Daneshyar, *One-Dimensional Compressible Flow*, Pergamon Press Ltd., Oxford, 1976.