*Note for reviewers - red text indicates revisions containing **new** material, blue text indicates revisions relocating **existing** material.*

Surface Reconstruction Amendment to the Intrinsic Sampling Method

Francis G.J. Longford,[1] Jonathan W. Essex,[1] Chris-Kriton Skylaris,[1] and Jeremy G. Frey[1]

University of Southampton, Southampton, UK, SO17 1BJ

(Dated: 12 November 2018)

The intrinsic sampling method (ISM) is a powerful tool that allows the exploration of interfacial properties from molecular simulations by fitting a function that represents the local boundary between two phases. However, owing to the non-physical nature of an 'intrinsic' surface, there remains an ambiguity surrounding the comparison of theoretical properties with the physical world. It is therefore important that the ISM remains internally consistent when reproducing simulated properties found to match experiment, such as the surface tension or interfacial density distribution. We show that the current ISM procedure causes an over-fitting of the surface to molecules in the interface region, leading to a biased distribution of curvature at these molecular coordinates. We assert that this biased distribution is a cause of the disparity between predicted interfacial densities upon convolution to a laboratory frame, an artefact which has been known to exist since the development of the ISM. We present an improvement to the fitting procedure of the ISM in an attempt to alleviate the ambiguity surrounding the true nature of an intrinsic surface. Our "surface reconstruction" method is able to amend the shape of the interface so as to reproduce the global curvature distribution at all sampled molecular coordinates. We present the effects that this method has on the ISM predicted structure of a simulated Lennard-Jones fluid air-liquid interface. Additionally, we report an unexpected relationship between surface thermodynamic predictions of our reconstructed ISM surfaces and those of extended capillary wave theory, which is of current interest.

I. OVERVIEW

It is commonly agreed that the shape of molecular distributions in liquid-vapour and liquid-liquid interfaces should be described as locally sharp, but corrugated by capillary waves or thermal fluctuations[1–4]. Recently, there has been substantial progress in computational analysis techniques that are sensitive to the instantaneous position of interface regions in simulations containing multiple media or phases. These tools seek to resolve the molecular scale structural features[5–13], dynamical processes[14–17] and thermodynamic properties[18–20] of interfaces that become spatially and temporally averaged if observed from a laboratory frame. It is hoped that investigation of these features may help to explain anomalous observations of interfaces found using high resolution optical techniques such as ellipsometry[21–23]. Additionally, new research pathways have been opened up, pushing the boundaries of capillary wave theory and hydrodynamics to the nanoscopic scale[24–27]. Several different methodologies have been investigated, including the Surface Layer Identification (SLI) of Chaowdhary and Ladani[28], the Identification of Truly Interfacial Molecules (ITIM) of Pártay *et al.*[29], the Grid-Based Intrinsic Profile (GIP) of Jorge and Cordeiro[30] and the Intrinsic Sampling Method (ISM) of Chacón and Tarazona[5]. A detailed comparison of these methods has been made by Jorge *et al.*[31,32], which generally highlights that the ISM is the most self-consistent approach, since it avoids coarse-graining and is more easily optimised with suitable parametrisation, though at the cost of being highly computationally intensive. Since publication, there have been additional amendments to the speed of calculation[33] and optimisation routine[10,15]. However, for all methodologies there remains a general ambiguity regarding the definition of an interface at the molecular scale, since a macroscopic boundary between two phases does not have an atomistic physical form[6], and so any mathematical object intending to describe it remains resolution dependent[31]. This ambiguity leads to properties being reported for atomistic and molecular systems which do not appear to consistently reproduce experimental data. In particular, although the predicted frequency-dependent surface tension can recover macroscopic observations in the low frequency limit, it diverges with increasing resolution at the high frequency limit[5,34], and the intrinsic density distribution has been notoriously difficult to convolute back onto the global frame[5,8,12,28,34].

Nevertheless, recent attempts by Chacón and Tarazona and Bresme[35,36] to investigate simulated lipid membranes using the ISM have successfully reproduced experimental esti-

nglies of interfacial properties. Their results provide a thermodynamically consistent way to calculate the membrane area per lipid from a single simulation without *a priori* knowledge of the true bending moduli, or a numerical estimation of errors introduced due to finite size effects, in contrast to previous methodologies[37,38]. It is hoped that general development of other internally consistent approaches to applying the ISM could lead to a clearer understanding of how to estimate experimentally reproducible properties from molecular simulations. Ongoing discussions over the physical nature of surface tension at nanoscopic scales using capillary wave theories[39–41] would also benefit from knowledge of how to simulate and describe these environments with greater confidence.

Here we propose a more robust fitting procedure for the ISM to the interfacial particle coordinates of a molecular dynamics (MD) simulation in an attempt to alleviate the ambiguity surrounding the true nature of an intrinsic surface. We show that the current ISM fitting procedure causes an over-fitting of the surface function to particles in the interface region, leading to a biased distribution of curvature at these particle coordinates. Our "surface reconstruction" routine is able to amend the shape of the interface so as to reproduce the global curvature distribution at all sampled particle coordinates. This significantly reduces the ISM overestimation of density at the interface region upon convolution to a global frame, which has been known to exist since its development[5,8]. Consequently, we improve the self consistency of the ISM, allowing for simulated structural features at higher resolutions to be reported with increased confidence.

II. BACKGROUND

We employ a laboratory frame $\mathbf{r}(x, y, z)$ (see supplementary material) with two phases divided in the xy plane, lying along the normal z axis. We define the density distribution along the normal axis $\rho(z)$, spatially averaged across the plane $\mathbf{r}_{\parallel}(x, y)$. The usual definition of the Gibbs dividing range assumes a flat interface broadened by thermal motion centred at position $z = z_G$. The requirement of z_G is that it fulfils an equipartition of $\rho(z)$ between the vapour ρ_v and liquid ρ_l phases, in accordance with equation (1).

$$\int_{-\infty}^{z_G} (\rho(z) - \rho_v)\, dz = \int_{z_G}^{\infty} (\rho_l - \rho(z))\, dz \tag{1}$$

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

However, this definition is insensitive to the instantaneous fluctuations of particle density across the interface region, only describing them in terms of an average broadening effect. To resolve these fluctuations a model of the local structure of the phase boundary is typically required. Capillary wave theory[1,2] (CWT) describes a fluctuating interface by a two-dimensional function representing the displacement $\xi(\mathbf{r}_\parallel)$ from the surface plane $z = z_G$. A comprehensive overview of CWT is given by Blokhuis[42], which we attempt to summarise below.

A. Capillary Wave Theory

Classical CWT uses a wave-like description of an interface, depicted by the Fourier transform $\xi(\mathbf{q})$ of the surface height function $\xi(\mathbf{r}_\parallel)$.

$$\xi(\mathbf{r}_\parallel) = \int \hat{\xi}(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{r}_\parallel}\, d\mathbf{q} \tag{2}$$

This description enforces a normal distribution of fluctuations around the interface plane, resulting in $\left\langle \xi(\mathbf{r}_\parallel) \right\rangle_{\mathbf{r}_\parallel} = 0$. Therefore taking the frequency vector $\mathbf{q}$ to the macroscopic limit $|\mathbf{q}| \to 0$ will also recover $\hat{\xi}(0) = 0$ whereas the microscopic limit of the capillary length $|\mathbf{q}| \to L_c^{-1}$ is dependent on any restoring forces acting upon the interface (typically the surface tension γ and gravity g). As a consequence, classical CWT successfully captures the effect of long wavelength interfacial fluctuations, but requires a higher level of theory to accurately describe fluctuations on the molecular scale, such as density functional theory (DFT). The intrinsic density distribution $\tilde{\rho}(z)$, a measure of the internal structure of the interface, can be defined in terms of a perturbation to the laboratory frame density so that $\rho(\mathbf{r}) = \tilde{\rho}(z - \xi(\mathbf{r}_\parallel))$. A fundamental assumption of CWT is the statistical independence[1] of $\tilde{\rho}(z)$ and $\xi(\mathbf{r}_\parallel)$, allowing for the series expansion[43] of $\tilde{\rho}(z - \xi(\mathbf{r}_\parallel))$,

$$\rho(\mathbf{r}) \equiv \tilde{\rho}(z) - \tilde{\rho}'(z)\xi(\mathbf{r}_\parallel) + \tilde{\rho}_H(z)\xi''(\mathbf{r}_\parallel) + O(\xi^4)... \tag{3}$$

The inclusion of second derivative terms $\xi''(\mathbf{r}_\parallel)$ corresponding to the curvature of the intrinsic surface is known as "extended CWT"[44], and also results in a correction[43,44] to the density profile $\tilde{\rho}_H(z)$. Considering the normal distribution of $\xi(\mathbf{r}_\parallel)$ in equation (2), we can assert both[45] $\left\langle \xi(\mathbf{r}_\parallel) \right\rangle_{\mathbf{r}_\parallel} = \left\langle \xi''(\mathbf{r}_\parallel) \right\rangle_{\mathbf{r}_\parallel} = 0$ and therefore demonstrate the equivalence $\tilde{\rho}(z) = \left\langle \rho(\mathbf{r}) \right\rangle_{\mathbf{r}_\parallel}$. Integration of the normal vector length of $\xi(\mathbf{r}_\parallel)$ over $\mathbf{r}_\parallel$ yields the intrinsic surface area, and

ings the surface area excess ΔA, allowing for the determination of the change in free energy $\Delta F = \gamma \Delta A$ using γ as the sole restoring force (ie. ignoring gravity). For surfaces containing relatively smooth curvature, this integral can be approximated (see the supplementary material) in a way that allows integration by substitution using equation (2).

$$\Delta F = \frac{1}{8\pi^2} \int \mathbf{q}^2 |\xi(\mathbf{q})^2| \gamma(\mathbf{q}) \, d\mathbf{q} \tag{4}$$

A feature of extended CWT is the effect that higher order terms in equation (3) have on ΔF, which is typically encapsulated by a frequency-dependent surface tension[46] $\gamma(\mathbf{q})$. Each surface mode $\mathbf{q}$ will therefore possess a free energy contribution that has a quadratic relation to the square of its amplitude $\mathbf{q}^2|\xi(\mathbf{q})^2|$, fulfilling the equipartition theorem.

$$\gamma(\mathbf{q}) = \frac{k_B T}{\langle |\xi(\mathbf{q})^2| \rangle \, \mathbf{q}^2} \tag{5}$$

Additionally, the form of equation (3) is sensitive to the behaviour of the system interaction potential at long distances[44,47,48]. Therefore, if the interaction potential is dominated by short-range terms (CWT-SR) then the expansion of $\gamma(\mathbf{q})$ in $\mathbf{q}$ is regular, demonstrating harmonic behaviour[46].

$$\gamma(\mathbf{q}) = \gamma_0 + \kappa \mathbf{q}^2 + O(\mathbf{q}^4)... \tag{6}$$

Here, the scalar γ_0 gives the macroscopic value of the interfacial tension upon taking $\mathbf{q} \to 0$ and the coefficient κ is the interface bending rigidity. However, if strong long-range dispersion forces are present (CWT-LR) then $\gamma(\mathbf{q})$ displays different behaviour at the asymptote as $\mathbf{q} \to 0$, with an extra logarithmic contribution related to a bending length l_k[44].

$$\gamma(\mathbf{q}) = \gamma_0 + \kappa \mathbf{q}^2 \ln(\mathbf{q} l_k) + O(\mathbf{q}^4)... \tag{7}$$

It is of current debate which of these enhanced behaviours should reflect the true description of an interface in the mesoscopic region[24,39–41], since equation (7) predicts a minimum of $\gamma(\mathbf{q})$ for $|\mathbf{q}| > 0$. Consequently, the most relaxed (and therefore experimentally observable) state is not expected to be found at the macroscopic limit. Theoretical clarity over this discussion has become important since claims of experimental observations[49–51] of the microscopic surface tension spectrum $\gamma(\mathbf{q})$, have been problematic to interpret, in part due to technical difficulties surrounding the separation of scattering from the bulk and surface regions[52,53]. It is important to remember at this point that intrinsic surfaces are not expected to be physically interpretable at atomistic length scales, whereas intrinsic density profiles

could theoretically be observed from a flat interface in the absence of spatial averaging[54]. Consequently, transformation between the density distribution along the laboratory frame normal axis $\rho(z)$ and the local intrinsic frame $\tilde{\rho}(z)$ (the latter typically reported in molecular simulations) can be performed by the convolution in equation (8)[5].

$$\rho(z) = \int \tilde{\rho}(z - \xi(\mathbf{r}_\parallel)) P[\xi(\mathbf{r}_\parallel)] \, d\,\xi(\mathbf{r}_\parallel) \tag{8}$$

Assuming the thermodynamics of the intrinsic surface as ergodic allows us to model the probability distribution $P(\xi(\mathbf{r}_\parallel))$ as a Gaussian, with a variance, related to the surface roughness, that follows Boltzmann statistics $\Delta\xi(\mathbf{r}_\parallel) \propto k_B T$.

B. The Intrinsic Sampling Method

The ISM was developed as a discrete version of the continuum CWT model for atomistic simulations by representing equation (2) as Fourier series involving q_u frequencies and neglecting any gravitational effects[5].

$$\xi(\mathbf{r}_\parallel, q_u) = \sum_{|\mathbf{q}| \leq q_u} \hat{\xi}_{\mathbf{q}} e^{2\pi i \mathbf{r}_\parallel \cdot \mathbf{q}} \tag{9}$$

Therefore, the range of frequencies $\mathbf{q}\left(\frac{\mu}{L_x}, \frac{\nu}{L_y}\right)$ is determined by the simulation cell cross-section dimensions L_x and L_y and particle size σ, since μ and ν are integers that run from $-q_u$ to q_u, where $|q_u| \leq q_m$, a resolution limit set by $q_m = \sqrt{L_x L_y}/\sigma$. For atoms or small molecules possessing a single Lennard-Jones site, a suitable value of σ would be the hard sphere radius, since this ensures the minimum wavelength is never smaller then that of the surface particle radius; a constraint enforced by using γ as the restoring force. Chacón and Tarazona went on to optimise the solutions of the Fourier coefficients $\hat{\xi}_{\mathbf{q}}$ to represent an interface in a molecular dynamics simulation by restraining the function to pass through a set of n_0 "pivots" at $\mathbf{r}(x_j, y_j, z_j)$, which give the spatial coordinates of particles that are considered to be part of the interface region. Optimising the coefficients requires minimising the scalar W (equation (10)), representing the residual distance between the intrinsic surface and the pivot particles, as well as the intrinsic surface area across the plane of the interface where $A = L_x L_y$.

$$W = \frac{1}{2} \sum_{j=1}^{n_0} (z_j - \xi(x_j, y_j, q_m))^2 + \frac{\phi A}{2} \sum_{|\mathbf{q}| \leq q_m} |\hat{\xi}_{\mathbf{q}}|^2 \mathbf{q}^2 \tag{10}$$

The weighting scalar ϕ can be used to control the influence of surface area term and therefore the smoothness of the fitted intrinsic surface. However, typically it is set to a small number $\phi \approx 5 \times 10^{-8}$ during the optimisation routine, so that the number of waves included in the summation of (9) can then be varied afterwards to investigate intrinsic surfaces of differing resolution[5]. It is important to remember that the intrinsic surface remains a "soft" description of reality and therefore there is no definition of the parameters that should always represent the true physical state of the interface. Consequently the choice of pivots is not trivial, since the interface region is not defined until the position of the surface has been ascertained. Therefore a self consistent method is employed that both optimises $\xi(\mathbf{r}_\parallel, q_m)$ and defines a set of pivots simultaneously (see the supplementary material).

The ISM has so far produced descriptions of intrinsic surfaces from interfacial MD simulations that support both CWT-SR[5,10,24,34] and CWT-LR[33] behaviour of $\gamma(\mathbf{q})$. However, whether either of these are accurate reflections of physical nature depends on the ISM reproducing the correct form of the higher order term $\tilde{\rho}_H(z)\xi''(\mathbf{r}_\parallel)$ in equation (3), relating to the distribution of density with surface curvature. We report on and confirm a bias in the surface density at regions of high curvature that is introduced during the current ISM fitting procedure and appears to enforce an inaccurate description of $\tilde{\rho}_H(z)\xi''(\mathbf{r}_\parallel)$. To demonstrate this we shall now describe in detail how the ISM is able to map intrinsic surfaces from MD simulations trajectories.

C. Linear Algebra Solution

Despite being conceptually appealing, there are some difficulties involved in incorporating the ISM methodology with modern computing techniques. Most notably, inbuilt fast Fourier transforms do not provide the flexibility of including non-uniform grid approaches and tend to restrict the number of Fourier components to the number of discrete data elements (in our case, pivot sites). In order to work around these complications, it became usual to transform the Fourier series to a set of linear equations of trigonometric functions[8,9] with a set of coefficients $a_{\mu\nu}$.

$$\xi(\mathbf{r}_\parallel, q_u) = \sum_{\mu,\nu=-q_u}^{q_u} a_{\mu\nu} f_\mu(x) f_\nu(y) \tag{11}$$

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

Where μ and ν range from $-q_u..-1,0,1..q_u$, and the functions f_μ are defined as:

$$f_\mu(x) = \cos\left(\frac{2\pi\mu x}{L_x}\right) \tag{12a}$$

$$f_{-\mu}(x) = \sin\left(\frac{2\pi\mu x}{L_x}\right) \tag{12b}$$

Combining equations (10) and (11) yields a set of linear equations with respect to the coefficients $a_{\mu\nu}$, forming the general matrix equation $\mathbf{Ax} = \mathbf{b}$ (equation (13))[8]. Possible solutions to these can be solved in a self consistent way using an efficient forward and backward Gaussian elimination algorithm, following a LU decomposition, with more pivots added at each cycle until n_0 pivots are selected, which can be optimised by minimising the pivot diffusion rate between sampled frames[10].

$$\begin{aligned} &4\phi\pi^2 a_{\mu_1\nu_1}\left(\mu_1^2+\nu_1^2\right) \\ &+\sum_{\mu_2,\nu_2=-q_m}^{q_m} a_{\mu_2\nu_2}\sum_{j=1}^{n_0} f_{\mu_1}(x_j)f_{\nu_1}(y_j)f_{\mu_2}(x_j)f_{\nu_2}(y_j) \\ &=\sum_{j=1}^{n_0} z_j f_{\mu_1}(x_j)f_{\nu_1}(y_j) \end{aligned} \tag{13}$$

Here we redefine the original form of equation (13) as equation (15), in order to incorporate systems where the cell dimensions $L_x \neq L_y$ to account for anisotropy in L_x and L_y, since not all simulation methodologies will be able to have strict control over the cell size parameters. This is achieved through the use of the step function $\Psi(\mu,\nu)$ (equation (14)) and the inclusion of the cross-section ratio L_x/L_y into the diagonal terms of matrix $\mathbf{A}$. We have also introduced the weighting function $\Pi(\mu_1,\nu_1,\mu_2,\nu_2)$, the rationale behind which shall be explained later in section IV. The derivation of $\Psi(\mu,\nu)$, $\Pi(\mu_1,\nu_1,\mu_2,\nu_2)$ and equation (15) can be found in the supplementary material. Equation (15) will produce equivalent solutions to (13) if $\Pi(\mu_1,\nu_1,\mu_2,\nu_2)=1$ and $L_x = L_y$.

$$\Psi(\mu,\nu) = \begin{cases} 4 & \mu = 0 \;\; and \;\; \nu = 0 \\ 2 & \mu = 0 \;\; or \;\; \nu = 0 \\ 1 & else \end{cases} \tag{14}$$

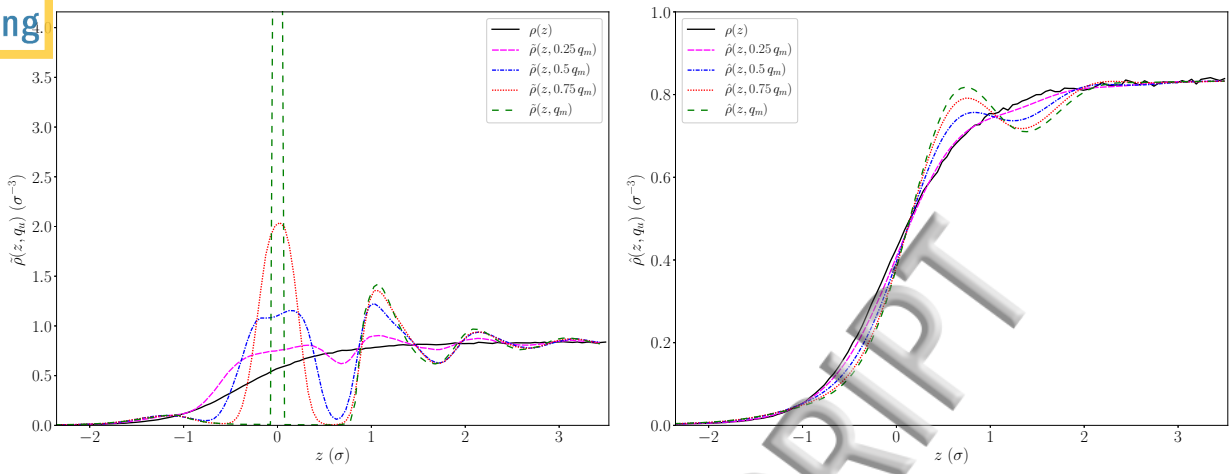


FIG. 1. Intrinsic density distributions (left) $\tilde{\rho}(z, q_u)$ (equation (16)) of increasing resolution q_u for ISM surfaces of an LJ air-liquid interface at $T = 0.711\, k_B\epsilon^{-1}$ (see the supplementary material for simulation details). A Dirac delta peak can be seen at $z = 0$ for high resolutions as $q_u \to q_m$. Corresponding convoluted density distributions (right) $\hat{\rho}(z, q_u)$ (equation (18) for each resolution, showing an decrease in the ability to resolve the correct form of the laboratory frame density distribution $\rho(z)$ as $q_u \to q_m$. The liquid phase density is measured as $\rho_l = 0.832\,\sigma^{-3}$

$$\begin{aligned}
&\Psi(\mu_1, \nu_1)\phi\pi^2 a_{\mu_1\nu_1}\left(\frac{\mu_1^2 L_y}{L_x} + \frac{\nu_1^2 L_x}{L_y}\right) \\
&\quad + \sum_{\mu_2,\nu_2=-q_m}^{q_m} a_{\mu_2\nu_2}\Pi(\mu_1, \nu_1, \mu_2, \nu_2) \\
&\quad \times \sum_{j=1}^{n_0} f_{\mu_1}(x_j) f_{\nu_1}(y_j) f_{\mu_2}(x_j) f_{\nu_2}(y_j) \\
&\quad = \sum_{j=1}^{n_0} z_j f_{\mu_1}(x_j) f_{\nu_1}(y_j)
\end{aligned} \tag{15}$$

D. Intrinsic Density $\tilde{\rho}(z, q_u)$

The ISM can be used to generate an intrinsic density distribution $\tilde{\rho}(z, q_u)$ from an interfacial simulation of N particles as a function of particle position $\mathbf{r}(x_k, y_k, z_k)$ relative to the solution of $\xi(x_k, y_k, q_u)$ at a given resolution q_u (equation (16)). For periodic simulations containing two interfaces, both surfaces are mapped independently from each other and an

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

average plot of $\tilde{\rho}(z, q_u)$ is generated.

$$\tilde{\rho}(z, q_u) = \frac{1}{A} \left\langle \sum_{k=1}^{N} \delta \left(z - z_k + \xi(x_k, y_k, q_u) \right) \right\rangle \tag{16}$$

A comparison of the intrinsic density $\tilde{\rho}(z, q_u)$ across various resolutions (including the density as viewed from the laboratory frame $\rho(z) \equiv \tilde{\rho}(z, 0)$ for an interfacial LJ fluid simulation at $T = 0.711\, k_B \epsilon^{-1}$ is illustrated in figure 1. Equation (8) dictates an important measure of internal consistency for the ISM regarding the mapping back of $\tilde{\rho}(z, q_u)$ to $\rho(z)$ across all resolutions. This can be achieved via equation (17), using the probability distribution $P(\delta\xi, q_u)$ of height deviations $\delta\xi$ from the surface plane.

$$\rho(z) \equiv \int_{-\infty}^{\infty} \tilde{\rho}(z - \delta\xi, q_u)\, P(\delta\xi, q_u)\, d\delta\xi \tag{17}$$

In this work we will write the convolution performed in equation (17) as $[\tilde{\rho}(q_u) * P(q_u)]\,(\delta\xi)$. As previously described, CWT assumes a normal distribution[5] of $\delta\xi$, therefore $P(\delta\xi, q_u)$ can be represented by a Gaussian function $\Omega(\delta\xi, q_u)$ with a variance $\Delta\xi(q_u)$.

$$\hat{\rho}(z, q_u) = [\tilde{\rho}(q_u) * \Omega(q_u)]\,(\delta\xi) \tag{18}$$

$$\Omega(\delta\xi, q_u) = \frac{1}{2\pi\Delta\xi(q_u)} \exp\left(-\frac{\delta\xi^2}{2\Delta\xi(q_u)} \right) \tag{19}$$

The value of $\Delta\xi(q_u)$ can be estimated by a summation of the average squared coefficients $\langle a_{\mu\nu}^2 \rangle$, in equivalence to that of the components in a Fourier series that describe a 2D wave[5].

$$\Delta\xi(q_u) = \left[\sum_{\mu,\nu=-q_u}^{q_u} \langle a_{\mu\nu}^2 \rangle \right] - \langle a_{00} \rangle^2 \tag{20}$$

Including our function $\Psi(\mu, \nu)$ into this summation to allow for anisotropy of L_x and L_y results in equation (21) (see the supplementary material).

$$\Delta\xi(q_u) = \left[\frac{1}{4} \sum_{\mu,\nu=-q_u}^{q_u} \Psi(\mu, \nu) \langle a_{\mu\nu}^2 \rangle \right] - \langle a_{00} \rangle^2 \tag{21}$$

The resultant global density $\hat{\rho}(z, q_u)$ is assumed to be uniform across all resolutions, since its effect on $\tilde{\rho}(z, q_u)$ is accounted for by the value of $\Delta\xi(q_u)$. Therefore the ISM can be considered internally consistent if $\rho(z)$ can be recovered for resolutions $q_u \leq q_m$. However, as shown in figure 1, previous investigations have highlighted that the current ISM is unable to demonstrate this[5,8] . We show next that a possible cause of this inability comes from an inaccurate distribution of surface curvature, caused by over-fitting of $\xi(\mathbf{r}_{\parallel}, q_m)$ to the set of pivot points in equation (15).

E. Surface Curvature $H(\mathbf{r}_\parallel, q_u)$

Extended CWT includes higher derivatives of $\xi(\mathbf{r}_\parallel)$ to account for the curvature of the intrinsic surface, specifically the second derivative $\xi''(\mathbf{r}_\parallel)$, which can be approximated[43] from the mean surface curvature $H(\mathbf{r}_\parallel)$.

$$\xi''(\mathbf{r}_\parallel) \approx H(\mathbf{r}_\parallel) = \frac{1}{2}\left[\frac{1}{R_1(\mathbf{r}_\parallel)} + \frac{1}{R_2(\mathbf{r}_\parallel)}\right] \tag{22}$$

The principal radii of curvatures $R_1(\mathbf{r}_\parallel)$ and $R_2(\mathbf{r}_\parallel)$ are defined as the maximum and minimum radii of curvature at position $\mathbf{r}_\parallel$. It has been previously suggested[5] that a bias on the fitting of $\xi(\mathbf{r}_\parallel, q_m)$ to the set of pivot particles at maximum resolution q_m, enforces the Fourier series to fit tightly to each pivot at $\mathbf{r}_\parallel(x_j, y_j)$, so that $\xi(x_j, y_j, q_m)$ tends to go through a point of inflection. This means that the (18) also becomes greatly dependent on the curvature of the surface around the interface region, as the density is no longer homogeneous for each solution to $\xi(\mathbf{r}_\parallel, q_u)$. We analyse this effect by calculating the mean curvature $H(\mathbf{r}_\parallel, q_u)$ for each resolution q_u and including this as a variable in a coupled intrinsic density distribution $\tilde{\rho}(z, H, q_u)$. For systems possessing relatively smooth curvature, $H(\mathbf{r}_\parallel, q_u)$ can be estimated in terms of secondary partial derivatives only[45] from equation (22).

$$H(\mathbf{r}_\parallel, q_u) \approx \frac{\partial^2 \xi(\mathbf{r}_\parallel, q_u)}{\partial x^2} + \frac{\partial^2 \xi(\mathbf{r}_\parallel, q_u)}{\partial y^2} \tag{23}$$

The partial second derivatives of $\xi(\mathbf{r}_\parallel, q_u)$, as defined by equation (11), are straight forward to solve.

$$\frac{\partial^2 \xi(\mathbf{r}_\parallel, q_u)}{\partial x^2} = \sum_{\mu,\nu=-q_u}^{q_u} a_{\mu\nu} f''_\mu(x) f_\nu(y) \tag{24a}$$

$$\frac{\partial^2 \xi(\mathbf{r}_\parallel, q_u)}{\partial y^2} = \sum_{\mu,\nu=-q_u}^{q_u} a_{\mu\nu} f_\mu(x) f''_\nu(y) \tag{24b}$$

Where the functions $f''_\mu(x)$ are defined as:

$$f''_\mu(x) = -\frac{4\pi^2\mu^2}{L_x^2} \cos\left(\frac{2\pi\mu x}{L_x}\right) \tag{25a}$$

$$f''_{-\mu}(x) = -\frac{4\pi^2\mu^2}{L_x^2} \sin\left(\frac{2\pi\mu x}{L_x}\right) \tag{25b}$$

We can clearly see $f''_\mu(x) \equiv -\frac{4\pi^2\mu^2}{L_x^2} f_\mu(x)$ and consequently $H(\mathbf{r}_\parallel, q_u)$ becomes:

$$H(\mathbf{r}_\parallel, q_u) = -4\pi^2 \sum_{\mu,\nu=-q_u}^{q_u} a_{\mu\nu} \left(\frac{\mu^2}{L_x^2} + \frac{\nu^2}{L_y^2}\right) f_\mu(x) f_\nu(y) \tag{26}$$

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

CWT predicts a normal distribution[45] of $H(\mathbf{r}_\parallel)$, and integrating equation (26) analytically across all space provides the expectation value $E\left[H(\mathbf{r}_\parallel, q_u)\right] = 0$ (see the supplementary material), thereby preserving the equality $\tilde{\rho}(z) = \langle \rho(\mathbf{r}) \rangle_{\mathbf{r}_\parallel}$ from equation (3). Additionally, according to equation (3), we would expect the correction term $\tilde{\rho}_H(z)$ to be independent of $\xi''(\mathbf{r}_\parallel)$ and therefore $\tilde{\rho}(z, H, q_u)$ should not be sensitive to $H(\mathbf{r}_\parallel, q_u)$.

We shall now investigate the air-liquid interface of a Lennard-Jones (LJ) fluid model using the ISM. All properties are reported in reduced units with respect to the LJ energy ϵ and hard-sphere radius σ, resulting in density units of σ^{-3}, surface tension units of $\epsilon\sigma^{-2}$, temperature units of $k_B\epsilon^{-1}$ and curvature units of σ^{-1}. A description of our the simulation procedure can be found in the supplementary material.

III. ORIGINAL ISM SURFACES

Here we explore both the distribution of surface curvature $P(H, q_u)$ and the coupled intrinsic density distribution $\tilde{\rho}(z, H, q_u)$ of LJ intrinsic surfaces simulated at $T = 0.711\, k_B\epsilon^{-1}$, as resolved by the ISM.

A. Curvature Distribution $P(H, q_m)$

Figure 2 illustrates the probability distribution of mean curvature $P(H, q_m)$ at the maximum resolution q_m, sampled at pivot particle sites $\mathbf{r}_\parallel(x_j, y_j)$. It appears that sampling the surface curvature at the pivot sites yields a distribution that includes a bias on larger values of $H(x_j, y_j, q_m)$ than would be expected from a randomly distributed sample. Larger values of mean curvature may indicate positions where the surface is undergoing a point of inflection. This may therefore signify over-fitting of $\xi(\mathbf{r}_\parallel, q_m)$ to $\mathbf{r}_\parallel(x_j, y_j)$, as the extrema are tightly bound to the positions of the pivot particles. Additionally, the distribution of $H(x_j, y_j, q_m)$ is slightly skewed, suggesting that there is an imbalance in the size of curvature around $\xi(x_j, y_j, q_m)$. Considering that pivot sites form the first particle layer of the intrinsic density distribution, it is expected that any sampling biases at $\mathbf{r}_\parallel(x_j, y_j)$ will be carried through to $\tilde{\rho}(z, H, q_u)$, which we now go on to explore.

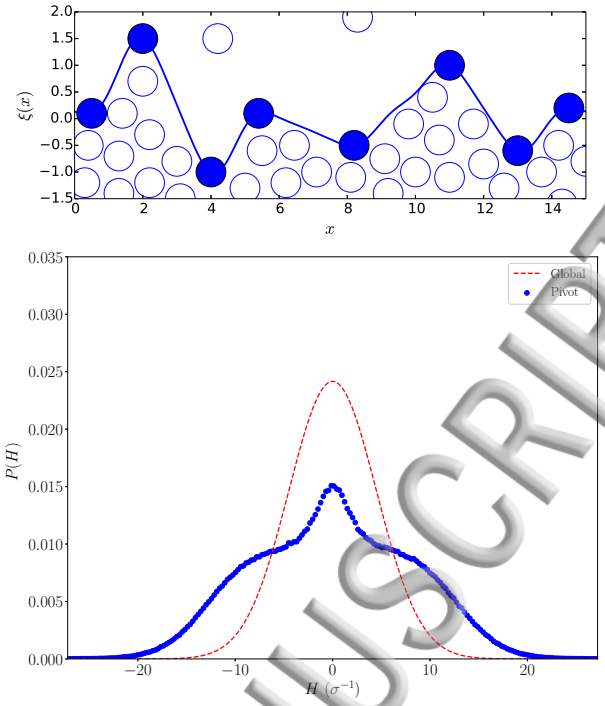


FIG. 2. Demonstration of the curvature bias at pivot particle sites of intrinsic surfaces defined using the ISM. A 1D illustration (top) of the ISM fitting routine, where the intrinsic surface $\xi(x)$ (blue line) is fit to pivot sites x_j (blue dots). Note that positions of extrema in $\xi(x)$ are likely to be heavily biased around x_j. Distribution taken from MD simulations of the LJ air-liquid interface at $T = 0.711\,k_B\epsilon^{-1}$ (bottom) of the mean surface curvature $H(\mathbf{r}_\parallel, q_m)$ sampled at surface pivot particle sites $\mathbf{r}_\parallel(x_j, y_j)$ (blue) for a regular ISM surface. The expected global (normal) distribution of H using a variance given by (27) is shown by a dashed red line.

B. Coupled Intrinsic Density $\tilde{\rho}(z, H, q_u)$

Figure 3 demonstrates the impact of this bias on the coupled intrinsic density distribution $\tilde{\rho}(z, H, q_u)$ for a range of surface resolutions q_u, where we use the absolute value of $|H|$ for visualisation purposes only. We can see that $\tilde{\rho}(z, H, q_u)$ is highly dependent on H at regions close to the intrinsic surface $z = 0$ for high resolutions of $q_u \geq 0.75\,q_m$. This dependency becomes less significant as z increases towards the bulk liquid, which would be expected if the bias only largely affects particles in the interface region. It also appears that at maximum resolution q_m, the distribution $\tilde{\rho}(0, H, q_m)$ is significantly decreased at regions of low curvature, and enhanced at those with a greater curvature. This coincides with our

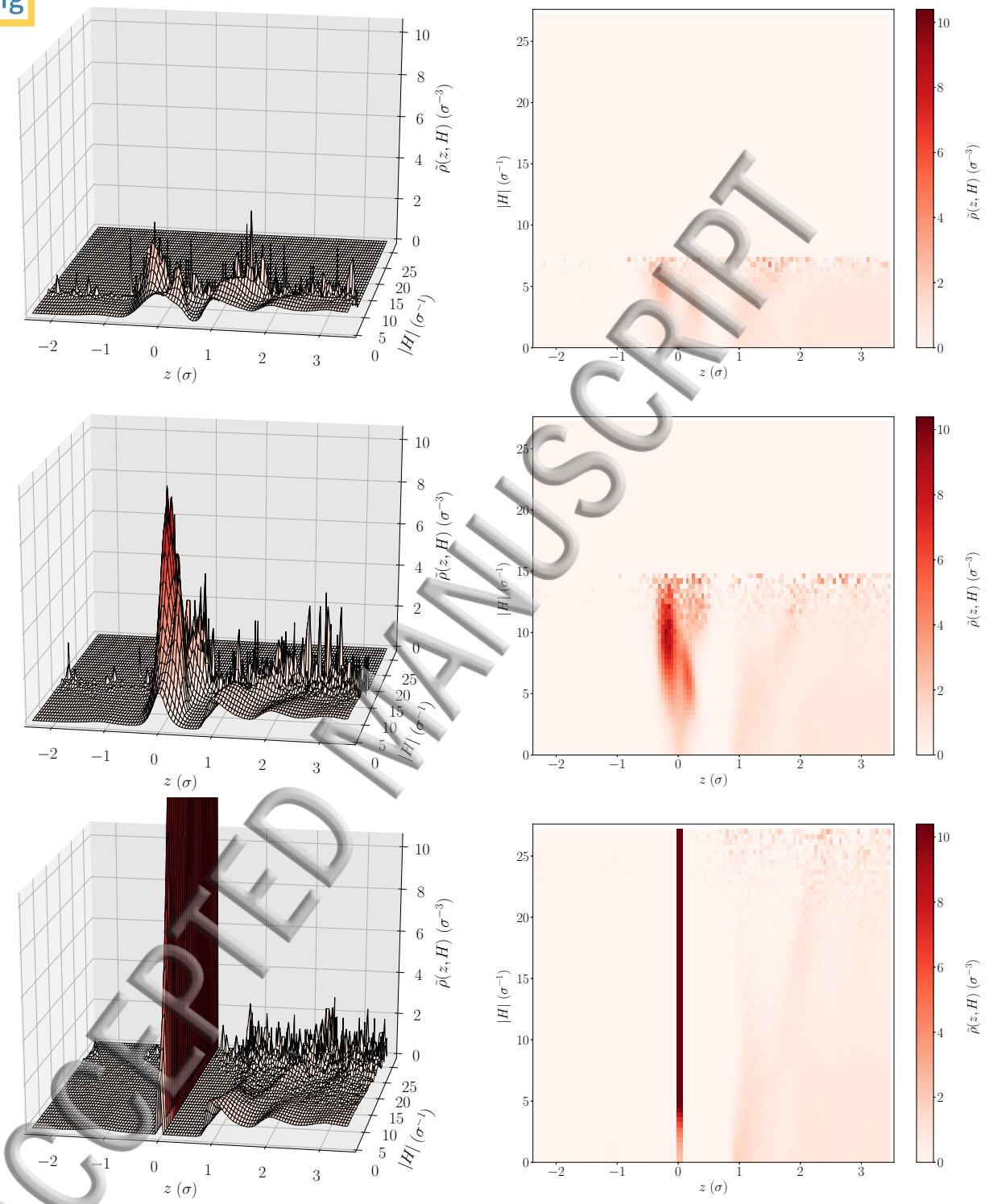


FIG. 3. Illustrations of the coupled intrinsic density distribution $\tilde{\rho}(z, H, q_u)$ of a LJ air-liquid interface at $T = 0.711\, k_B\epsilon^{-1}$ at resolutions $q_u = 0.5\, q_m$ (top), $q_u = 0.75\, q_m$ (middle), and $q_u = q_m$ (bottom) (see supplementary material for simulation details). Left plot includes $\tilde{\rho}(z, H, q_u)$ as a 3D projection on the z axis, whereas the right side reduces this axis to a 2D colour map.

conclusion that over-fitting the intrinsic surface to the pivot sites $\mathbf{r}(x_j, y_j, z_j)$ results in a bias on larger values of $H(x_j, y_j, q_m)$. Additionally, there appears to be a splitting of $\tilde{\rho}(z, H, q_u)$ close to $z = 0$ for values of $|H| \approx 7\,\sigma^{-1}$ at resolution $q_u \leq 0.75\,q_m$. This splitting at lower resolutions may also suggest that the intrinsic surface is over-fit at q_m, since removing waves in the Fourier sum appears to artificially deform the surface, causing it to no longer go through each pivot site. Consequently it is expected that the higher derivative term in equation $\tilde{\rho}_1(z)$ is not being evaluated correctly, violating the equality $\tilde{\rho}(z) = \langle \rho(\mathbf{r}) \rangle_{\mathbf{r}_\parallel}$ and preventing successful convolution of the intrinsic density back to the laboratory frame (figure 1) using equation (18).

A way to avoid over-fitting or under-fitting in parametrisation of $\xi(\mathbf{r}_\parallel, q_u)$ could involve optimisation of ϕ in equation (10). However, this is not a trivial task for the ISM, considering that choice of pivot particles typically need to be determined in a self-consistent way, and will therefore be dependent on ϕ. In statistical terms, our training dataset changes with regularisation parameter, and therefore we cannot be sure that a minima of ϕ can be found. The only other way to address this issue previously without altering the definition of $\xi(\mathbf{r}_\parallel, q_u)$ has been to use lower resolutions of q_u, effectively decreasing the range of H so that the fitting bias becomes less significant[8]. As a result, the estimate of the laboratory density $\rho(z)$ in (18) also remains dependent on the resolution q_u and therefore lacks internal consistency. Clearly this is a huge hindrance for ISM analysis, as the inability to transform confidently between the local intrinsic surface and laboratory frames across all resolutions restricts the prediction of experimentally observable properties. Consequently we propose a method to reduce the bias of H by recovering the correct global distribution of mean curvature at the pivot particle sites during the ISM fitting procedure.

Our methodology uses the rationale that usually any sample bias can be addressed by reselecting data to reflect the global distribution. However, since our pivot coordinates $\mathbf{r}(x_j, y_j, z_j)$ must remain fixed, we propose to perform an equivalent routine by altering the description of the function $\xi(\mathbf{r}_\parallel, q_m)$ to ensure the distribution of $H(x_j, y_j, q_m)$ across all pivot sites matches that of the global distribution of $H(\mathbf{r}_\parallel, q_m)$. We now outline an amendment to the minimisation scheme in equation (15) that resolves the coefficients $a_{\mu\nu}$ in equation (11) to optimise the variance of $H(x_j, y_j, q_m)$ and recover the expected mean curvature distribution $P(H)$.

IV. SURFACE RECONSTRUCTION METHOD

Considering that $E\left[H(\mathbf{r}_{\parallel}, q_u)\right] = 0$, the variance in mean curvature $\mathrm{Var}[H(q_u)] = E\left[H(\mathbf{r}_{\parallel}, q_u)^2\right] - E\left[H(\mathbf{r}_{\parallel}, q_u)\right]^2$ can be expressed by the single expectation value $E\left[H(\mathbf{r}_{\parallel}, q_u)^2\right]$. Performing an integration of $H(\mathbf{r}_{\parallel}, q_u)^2$ over the entire surface then yields an estimation for $\mathrm{Var}[H(q_u)]$ (see the supplementary material).

$$\mathrm{Var}[H(q_u)] = 4\pi^4 \sum_{\mu,\nu=-q_u}^{q_u} a_{\mu\nu}^2 \Psi(\mu,\nu) \left(\frac{\mu^4}{L_x^4} + \frac{\nu^4}{L_y^4} + 2\frac{\mu^2\nu^2}{L_x^2 L_y^2} \right) \tag{27}$$

During the ISM fitting procedure at resolution q_m, we make the presumption that the variance in $H(x_j, y_j, q_m)$ sampled at n_0 pivot sites should reproduce the similar variance as the global distribution of $H(\mathbf{r}_{\parallel}, q_m)$. Therefore, using equations (26) and (27) we also enforce the following condition in our minimisation scheme.

$$\begin{aligned} &\sum_{\mu,\nu=-q_m}^{q_m} \Psi(\mu,\nu) a_{\mu\nu}^2 \left(\frac{\mu^4}{L_x^4} + \frac{\nu^4}{L_y^4} + 2\frac{\mu^2\nu^2}{L_x^2 L_y^2} \right) \\ =&\frac{4}{n_0} \sum_{j=1}^{n_0} \left[\sum_{\mu,\nu=-q_m}^{q_m} a_{\mu\nu} \left(\frac{\mu^2}{L_x^2} + \frac{\nu^2}{L_y^2} \right) f_\mu(x_j) f_\nu(y_j) \right]^2 \end{aligned} \tag{28}$$

In practice we do so by iteratively re-optimising the coefficients $a_{\mu\nu}$ at each cycle in the self-consistent pivot selection (hence our term "surface reconstruction"). This is achieved by including the right hand side of equation (28) into (15), resulting in a new expression of the weighting term $\Pi(\mu_1, \nu_1, \mu_2, \nu_2)$, given by equation (29).

$$\begin{aligned} &\Pi(\mu_1, \nu_1, \mu_2, \nu_2) \\ &= \left[1 + \psi \frac{4}{n_0} \left(\frac{\mu_1^2\mu_2^2}{L_x^4} + \frac{\nu_1^2\nu_2^2}{L_y^4} + \frac{\mu_1^2\nu_2^2 + \mu_2^2\nu_1^2}{L_x^2 L_y^2} \right) \right] \end{aligned} \tag{29}$$

We can then optimise the weighting scalar ψ using a Newton-Raphson method to obtain the criteria in equation (28) within a threshold value. The full derivations of equations (27), (28) and (29) can be found in the supplementary material.

We analyse the structural effects of this new optimisation by comparing the distribution of mean curvature along the normal axis $H(z, q_u)$, the coupled intrinsic position-surface norm distribution $\tilde{\rho}(z, H, q_u)$ and the effective density distribution $\hat{\rho}(z, q_u)$. We now explore the effect that applying our surface reconstruction has on the previously identified intrinsic surfaces of simulated LJ air-liquid interfaces in section III.

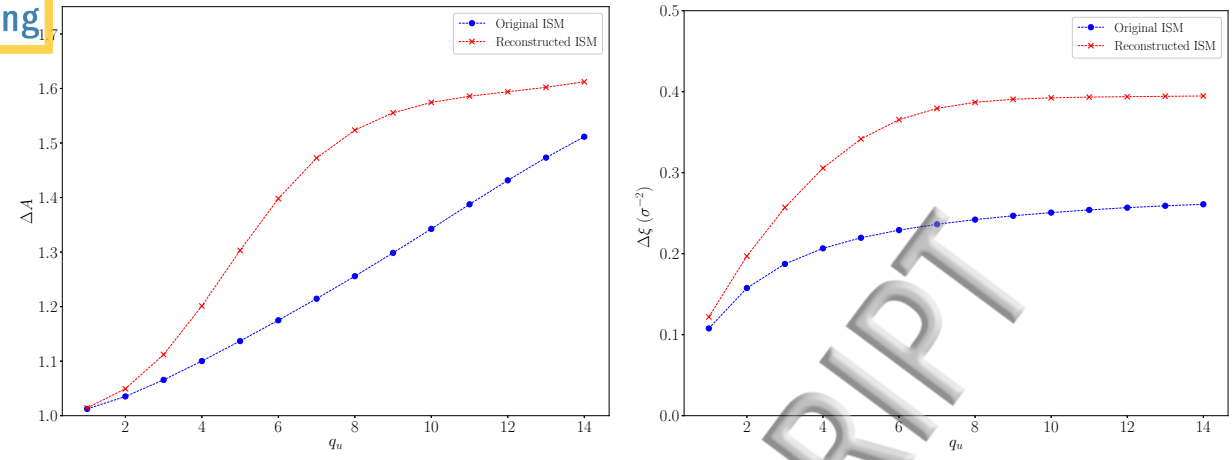


FIG. 4. Surface excess area $\Delta A(q_u)$ (left) and surface variance $\Delta\xi(q_u)$ (right) as a function of resolution q_u for intrinsic surfaces of an LJ fluid simulation at $T = 0.711\,k_B\epsilon^{-1}$ fitted via the original and reconstructed ISM protocol (see supplementary material for simulation details).

	$\sqrt{A}\ (\sigma)$	ϕ	q_m	$n_s\ (\sigma^{-2})$	n_0	ΔA	$\Delta\xi\ (\sigma^{-2})$
Original ISM	14.86	5×10^{-8}	14	0.946	209	1.51	0.261
Reconstructed ISM	14.86	5×10^{-8}	14	0.846	187	1.61	0.395

TABLE I. ISM parameters for intrinsic surfaces of LJ air-liquid interface at $T = 0.711\,k_B\epsilon^{-1}$: simulation cell unit length $\sqrt{A}$, regularisation parameter ϕ, maximum resolution parameter q_m, surface pivot density n_s, surface pivot number n_0, surface excess area ΔA and surface variance $\Delta\xi$ calculated at the maximum resolution q_m.

V. RECONSTRUCTED ISM SURFACES

Once again we explore both the distribution of surface curvature $P(H)$ and the coupled intrinsic density distribution $\tilde{\rho}(z, H, q_u)$ to explore the impact that surface reconstruction has on air-liquid interfaces of a LJ fluid, resolved by the ISM. Additionally, by modelling the behaviour of $\gamma(\mathbf{q})$ as $|\mathbf{q}| \to 0$, we can explore the impact that surface reconstruction has on an experimentally observable property γ_0 (equations (6) and (7)), as well as whether it continues to support traditional CWT theories that suggest a minima at $|\mathbf{q}| = 0$ or other recent theories that predict a minima for $|\mathbf{q}| > 0$. For the dataset of trajectories analysed in section III, we have re-run the ISM fitting procedure whilst including a surface reconstruction at each iteration in the pivot selection cycle. We report a 96% overlap in surface particle

"pivots" selected by the original and reconstructed ISM. However, the number of pivot particles n_0 has also been re-optimised so as to minimise their diffusion rate[10], resulting in a decrease of 10% in the optimal surface particle density n_s. Consequently, it is not advisable to simply use the original set of pivots obtained by the ISM, despite the high overlap, as it is not always possible to find appropriate solutions to equation (15) for an over-saturated layer. We also report a 10% increase in the surface excess area $\Delta A(q_m)$ (see supplementary material for derivation) and a and 50% increase in the surface variance $\Delta\xi(q_m)$ at the maximum resolution q_m (table I). Interestingly, the reconstructed surfaces appear to show both a converging $\Delta A(q_u)$ and $\Delta\xi(q_u)$ as $q_u \to q_m$ (figure 4), whereas other sampling methods have been previously required to converge the surface excess area[35,36].

A. Curvature Distribution $P(H, q_m)$

We can see that the reconstruction algorithm effectively redistributes the intrinsic surface in order to replicate the global distribution of mean curvature for the pivot particle sites $H(x_j, y_j, q_m)$ (figure 2). This includes a reduction in skewness and over-representation of larger curvature, so that it now yields a normal distribution of $H(\mathbf{r}_\parallel, q_m)$, predicted to be the correct description by CWT[45]. It is important to note that this redistribution arises naturally from our routine, since we only optimise the variance of H, indicating that including the curvature in the intrinsic surface fitting is crucial for obtaining the correct parametrisation of $\xi(\mathbf{r}_\parallel, q_m)$.

B. Coupled Intrinsic Density $\tilde{\rho}(z, H, q_u)$

Figure 6 illustrates the effect that this redistribution of surface curvature has on the coupled intrinsic density $\tilde{\rho}(z, H, q_u)$. It is clear that our correction to the surface curvature leads to an $\tilde{\rho}(z, H, q_u)$ distribution that is considerably more independent of H across all resolutions, including the removal of extreme density values in $\tilde{\rho}(0, H, q_m)$. The splitting of $\tilde{\rho}(z, H, q_u)$ close to $z = 0$ for values of $|H| \approx 7\,\sigma^{-1}$ at resolutions $q_u \leq 0.75\,q_m$ is also effectively removed upon surface reconstruction, recovering a more robust description of the intrinsic surface across all resolutions. There is still, however, a slight decrease in $\tilde{\rho}(z, H, q_u)$ for small values of $|H| \leq 3.4\,\sigma^{-1}$, demonstrating the limits of the current routine.

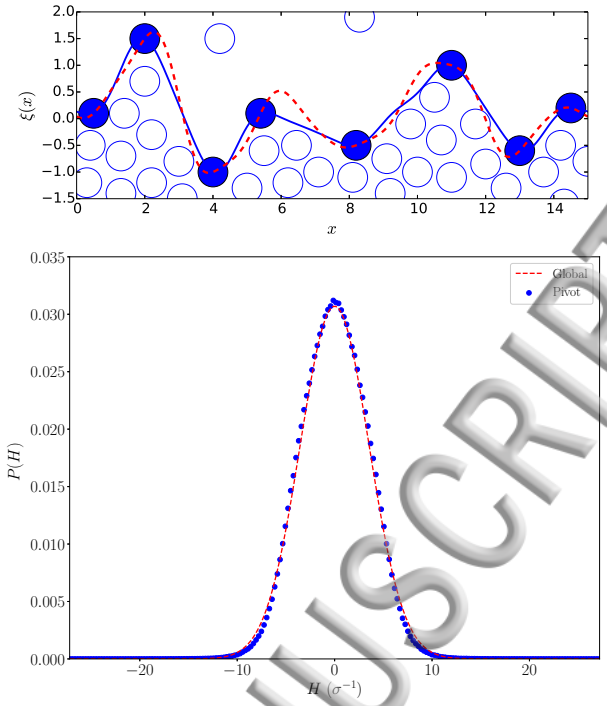


FIG. 5. Demonstration of the surface reconstruction method to reduce curvature bias at particle pivot sites. A 1D illustration (top) of an intrinsic surface $\xi(x)$ (blue line) fit to pivot sites x_j (filled blue circles) using original ISM procedure with a reconstructed surface (red dashed line). Reconstruction enforces the variance of curvature at pivots sites x_j to be the same as variance across the range of the function $\xi(x)$. Distribution taken from MD simulations of the LJ air-liquid interface at $T = 0.711\, k_B\epsilon^{-1}$ (bottom) of the mean surface curvature $H(\mathbf{r}_\parallel, q_m)$ at surface pivot particle sites $\mathbf{r}_\parallel(x_j, y_j)$ (blue) for a reconstructed ISM surface. The global (normal) distribution of H using a variance given by (27) is shown by a dashed red line.

Any remaining dependence of $\tilde{\rho}(z, H, q_u)$ on H between the first and second surface layers can be attributed to the inaccuracy between the radial distance of a particle from the surface and the depth along the z axis only[5]. The "true" value would therefore be the minimal distance between a position in space $\mathbf{r}_\parallel(x_j, y_j)$ and all possible solutions to the function $\xi(\mathbf{r}_\parallel, q_u)$. Clearly this discrepancy would increase for larger curvatures, as seen in figures 3 and 6, but becomes relatively insignificant for particle coordinates far away from the position of the intrinsic surface. Although an approximation to the correction between both radial and depth frames was suggested, this was not considered suitable for convolution back to the global axis[5]. Consequently we are unable to continue further at this moment in

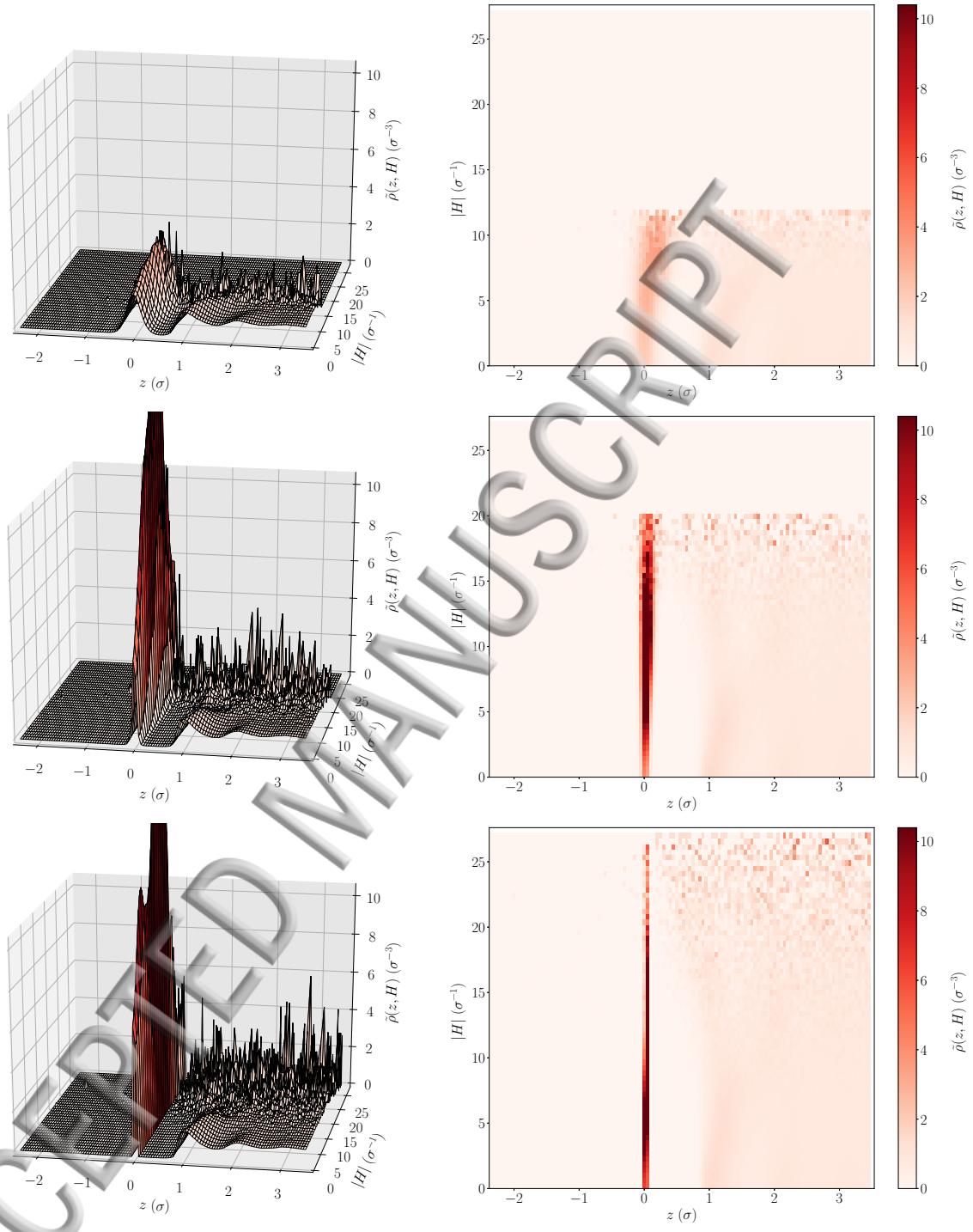


FIG. 6. Illustrations of the coupled intrinsic density distribution $\tilde{\rho}(z, H, q_u)$ of a LJ air-liquid interface at $T = 0.711\, k_B\epsilon^{-1}$ for resolutions $q_u = 0.5\, q_m$ (top), $q_u = 0.75\, q_m$ (middle), and $q_u = q_m$ (bottom) after having undergone surface reconstruction (see supplementary material for simulation details). Left plot includes $\tilde{\rho}(z, H, q_u)$ as a 3D projection on the z axis, whereas the right side reduces this axis to a 2D colour map.

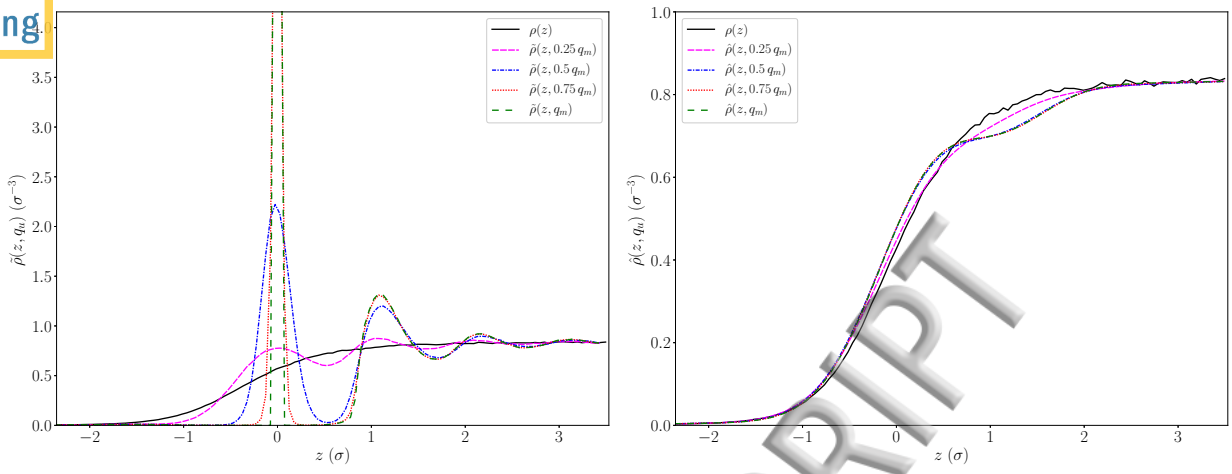


FIG. 7. Intrinsic density distributions (left) $\tilde{\rho}(z, q_u)$ (equation (16)) of increasing resolution q_u for reconstructed ISM surfaces of the LJ air-liquid interface at $T = 0.711\,k_B\epsilon^{-1}$ (see supplementary material for simulation details). The development of a Dirac delta peak at lower resolutions q_u and significant reduction of density in the vapour region $z < 0$, compared to figure 1, illustrates the effect of our surface reconstruction routine. Consequently, the corresponding convoluted density distributions (right) $\hat{\rho}(z, q_u)$ (equation (18)) for each resolution resemble the correct form of the laboratory frame density distribution $\rho(z)$ across all resolutions as $q_u \to q_m$. The liquid phase density is measured as $\rho_l = 0.832\,\sigma^{-3}$

time to reduce this additional dependency of $\tilde{\rho}(z, H, q_u)$ on H.

Additionally, we can see that surface reconstruction extends the range of possible curvature values, indicated by decreased noise in the recorded spectra of $\tilde{\rho}(z, H, q_u)$, especially for low values of q_u. We also note an unexpected effect of surface reconstruction is that the delta peak at $\tilde{\rho}(0, H, q_u)$ develops at lower resolutions, suggesting once again that we are able to reproduce a theoretically expected description of the intrinsic surface.

C. Convolution to Laboratory Frame

Figure 7 shows the effect that this new distribution of $\tilde{\rho}(z, H, q_u)$ has on the intrinsic density distribution $\tilde{\rho}(z, q_u)$. The Dirac delta peak at $z = 0$ appears to develop at much lower resolutions than the original ISM method shown in figure 1. Additionally, we find a significant reduction of density just inside the vapour region $z \approx -\sigma$, which has also been considered an artefact of over-fitting[5]. Upon performing the convolution $[\tilde{\rho}(q_u) * P(q_u)]\,(\delta\xi)$,

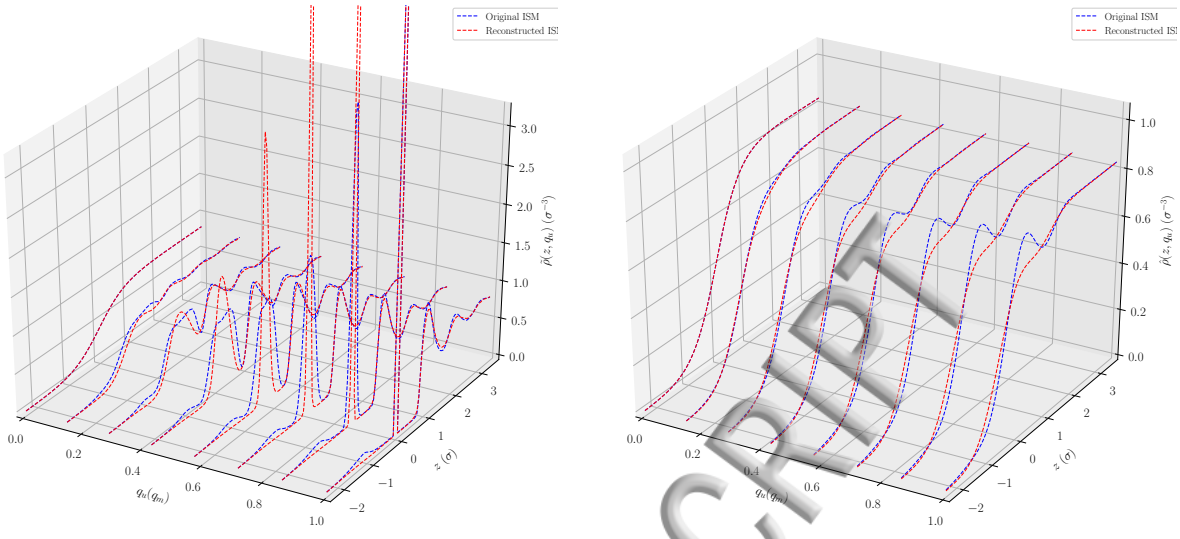


FIG. 8. The intrinsic particle density (left) for a LJ fluid as defined by (16) for the original fitted ISM (blue) and one having undergone surface reconstruction (red). The convoluted particle density (right) for a LJ fluid as defined by (18) for the original fitted ISM (blue) and one having undergone surface reconstruction (red).

we report a significant reduction the size of the artificial behaviour of the effective density distribution (figure 7), leading to a more converged form of $\hat{\rho}(z, q_u)$ at the interface region, and allowing for a more accurate recovery of the laboratory frame density $\rho(z)$ as the resolution $q_u \to q_m$. The side-by side comparison between original and reconstructed ISM density profiles, illustrated in figure 8, reveals large differences between $\tilde{\rho}(z, q_u)$ around the interface region for resolutions $0.2\, q_m \leq q_u \leq 0.8\, q_m$, although the particle distribution beyond the Dirac delta peak is relatively unaffected. The significant increase in surface variance $\Delta\xi(q_u)$ at each resolution (figure 4) also plays a large part in smoothing the reconstructed distributions of $\hat{\rho}(z, q_u)$. However, we ascertain that it is not enough on its own to yield the level of convergence in the effective density distribution, and consequently a combination of alterations to both $\tilde{\rho}(z, q_u)$ and $\Delta\xi(q_u)$ is crucial for minimising the artefacts in $\hat{\rho}(z, q_u)$. Nevertheless, we do not see these artefacts disappearing completely, suggesting that additional factors may also be contributing.

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

D. Power Spectrum and Surface Tension

Additional insight into the effects of our surface reconstruction routine on the CWT description of the ISM is provided by the power spectrum $\langle|\xi(\mathbf{q})^2|\rangle$. Mapping each coefficient $a_{\mu\nu}$ to a Fourier amplitude $\xi(\mathbf{q})$ using $\mathbf{q}\left(\frac{2\pi\mu}{L_x}, \frac{2\pi\nu}{L_y}\right)$ results in equation (30).

$$\left\langle|\xi(\mathbf{q})^2|\right\rangle = \frac{L_x L_y}{4}\Psi(\mu,\nu)\left\langle a_{\mu\nu}^2\right\rangle \tag{30}$$

Figure 9 demonstrates the effect of surface reconstruction on the the power spectrum of $\xi(\mathbf{r}_{\|}, q_m)$. We observe that surface reconstruction causes a slight increase of $\left\langle|\hat{\xi}(\mathbf{q})^2|\right\rangle$ for small values of $|\mathbf{q}| < 5\sigma^{-1}$ compensated by a sharp reduction $\left\langle|\hat{\xi}(\mathbf{q})^2|\right\rangle$ for larger values of $|\mathbf{q}| < 7.5\sigma^{-1}$. This is to be expected since the large structural features dominated by low frequency waves remain tightly constrained by the pivot coordinates so that possible solutions to (28) are easier to find by forcing changes to small structural features dominated by high frequency waves. As a consequence we now see non-monotonic behaviour as the wavelength decreases towards the atomic radius σ, which may be considered an artefact of our routine since this feature has not been previously reported upon for the ISM. Further research is therefore required to clarify whether we are obtaining the optimal solution of equation (10) through our methodology.

Interestingly there is also deviation of the power spectrum from exponential decay for smaller values of $0 < |\mathbf{q}| < 4\,\sigma^{-1}$. This region is the current focus of on-going investigations into the form of the frequency dependent surface tension $\gamma(\mathbf{q})$ in CWT as discussed earlier. Combining equations (30) and (5) allows us to report on $\gamma(\mathbf{q})$.

$$\gamma(\mathbf{q}) = \frac{k_B T}{\pi^2\Psi(\mu,\nu)\left\langle a_{\mu\nu}^2\right\rangle\left(\frac{\mu^2 L_y}{L_x} + \frac{\nu^2 L_x}{L_y}\right)} \tag{31}$$

Figure 9 shows CWT-SR harmonic behaviour of $\gamma(\mathbf{q})$ from intrinsic surfaces using the original ISM method, as previously reported[24]. We are therefore able to fit this $\gamma(\mathbf{q})$ spectrum to equation (6) and derive an estimation of the macroscopic surface tension $\gamma_0 = 1.03 \pm 0.01\,\epsilon\sigma^{-2}$. Features predicted by CWT-LR have previously been previously identified in the low frequency region ($|\mathbf{q}| < \sigma^{-1}$) of surface tension spectra produced by the ISM[33]. However, we have been unable to confidently reproduce these subtle effects due to the significant amount of noise still present in our spectra, even after averaging over 6000 independent samples (see inlay figure 9). Nevertheless, we do observe CWT-LR behaviour

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

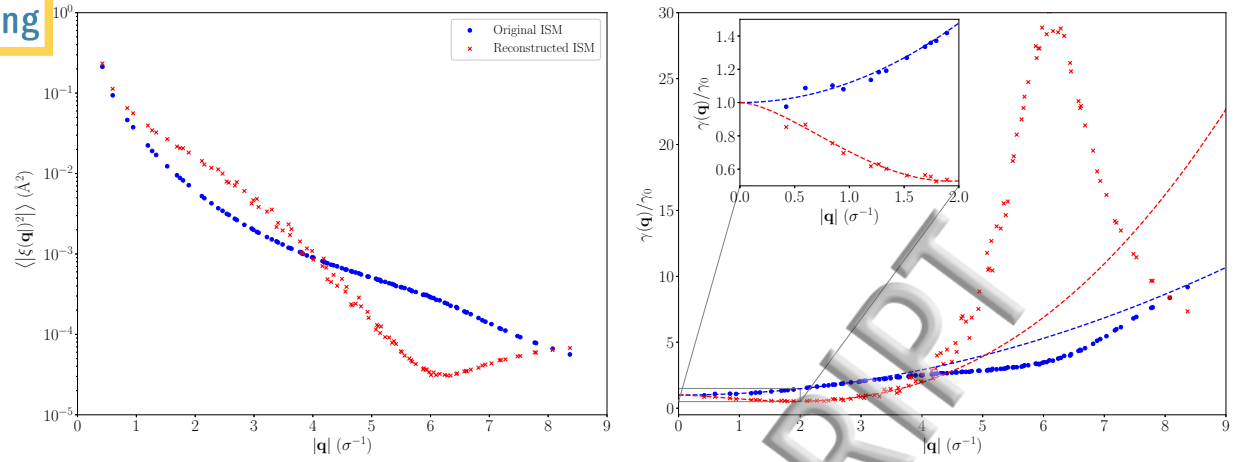


FIG. 9. Power spectrum of Fourier series modes $\langle|\xi(\mathbf{q})^2|\rangle$ (left) and frequency dependent surface tension spectrum $\gamma(\mathbf{q})$ (right) for original fitted intrinsic surfaces (blue) and those having undergone surface reconstruction (red) of the LJ air-liquid interface at $T = 0.711\, k_B\epsilon^{-1}$. Solid trend lines indicate data fitted to equation (6), whereas dashed lines are fitted to equation (7).

in our reconstructed spectrum, which also differs markedly from any prior study. In particular, it is a much stronger effect, with a minimum appearing in a much higher region of the spectrum $|\mathbf{q}| \approx 3.5\,\sigma^{-1}$. Consequently we are able to produce estimations of the macroscopic surface tension $\gamma_0 = 1.07 \pm 0.04\,\epsilon\sigma^{-2}$ by fitting our surface reconstructed $\gamma(\mathbf{q})$ spectrum to equation (7). Our reported values are all comparable to each other and the statistical mechanic Kirkwood-Buff (KB) estimation of $\gamma_0 = 1.12 \pm 0.16\,\epsilon\sigma^{-2}$. via the diagonal components of the global pressure tensor $\mathbf{P}$ (equation (32)), which is most commonly employed to calculate surface tensions in MD simulations[55].

$$\gamma = \frac{L_z}{2}\left(P_{zz} - \frac{P_{xx} + P_{yy}}{2}\right) \tag{32}$$

The pressure tensor components are calculated at each time step using the mass m_i and velocity $\mathbf{v}_i$ of each particle i and the pairwise Cartesian distance $\mathbf{r}_{ij}$ and force $\mathbf{F}_{ij}$ vectors between particles i and j (equation (33)).

$$P_{\alpha\beta}V = \frac{1}{2}\sum_{i}^{N} m_i v_{i_\alpha} v_{i_\beta} + \sum_{i}^{N-1}\sum_{j>i}^{N} r_{ij_\alpha} F_{ij_\beta} \tag{33}$$

It has been previously reported[27,56] that the $\gamma(\mathbf{q})$ of simple fluids and Ising models show anomalous behaviour for large values of $|\mathbf{q}|$, commonly attributed to an additional frequency-dependent component which weights the bulk gas and liquid contributions. As a result, we

This manuscript was accepted by J. Chem. Phys. Click here to see the version of record.

ing only able to report on the behaviour of $\gamma(\mathbf{q})$ within the region $0 < |\mathbf{q}| < 4\,\sigma^{-1}$, since the LJ fluid forcefield is unlikely to replicate behaviour predicted by any CWT models outside this range. Nevertheless, the sharp increase in $\gamma(\mathbf{q})$ for $4 < |\mathbf{q}| < 9\,\sigma^{-1}$ has been reported experimentally for other liquids, including LJ fluids, using X-ray reflectivity techniques[50,51], although these are unable to probe accurately beyond a resolution limit of $|\mathbf{q}| < 0.5$ nm$^{-1} \approx 6\,\sigma^{-1}$.

Therefore, whether alterations to the power spectrum and ISM surface structure present simulated evidence of the predictions of extended CWT or is itself an unintended artefact is not immediately clear, and so conclusions drawn from this analysis should be treated with care.

VI. DISCUSSION

Our surface reconstruction routine is able to successfully redistribute the surface curvature H sampled over the pivot particles to reflect that of the global population, which had also been pushed toward the theoretically expected normal distribution. This has led to a significant reduction in the artificial density enhancement in $\hat{\rho}(z, q_u)$ at the interface region, through it has not been fully removed. Additionally, our surface reconstruction routine has redistributed the amplitudes of each wave in the Fourier series describing the structure of the interface to reflect the predictions of CWT-LR in a way that has not been previously reported upon.

Whilst testing our routine we found an optimal starting value for the weighting scalar in (29) was $\psi = \phi A$. This relation links the impact of the surface area minimisation term to the surface curvature distribution and allows for the same number of initial variable parameters as the original method. Therefore our procedure does not increase the level of complexity of the ISM, which is important to reduce the likelihood of over-fitting. The algorithm is also highly efficient when performed using vector handling libraries written in compiled code such as Numerical Python (NumPy: http://www.numpy.org), since we can perform nearly all operations via matrix multiplication. It typically adds an additional 25% computational time to the intrinsic surface fitting routine. Generally, the section of code that requires the largest amount of computational resources is the iterative pivot search, which is difficult to improve upon without including additional artefacts, but remains relatively independent of

ing reconstruction algorithm.

It is worth investigating the reduction in optimal surface density n_s, obtained via minimisation of the pivot diffusion rate[10], upon reconstruction. Additionally, assessment of particle numbers in each surface layer, as well as the intrinsic profiles of excess density, temperature, surface tension, energy and free energy for a range of chemically relevant species would be highly advantageous to explore further effects of our surface reconstruction routine. This analysis has been performed previously using differed intrinsic sampling methodologies[17,19,57,58], and so is likely to provide insight into the transferability of predicted properties obtained from surfaces with an explicitly enforced Gaussian distribution of mean curvature.

Additionally, previous investigations have shown that a more robust estimation of the surface tension can be gathered through analysis of the behaviour of the interface via hydrodynamic theory[15]. In the structural approach, equations (6) and (7) focus on the ensemble average of wave amplitudes, considering the dynamics of each wave, or mode, to be distributed in thermal equilibrium. However, the dynamics of each mode in our Fourier series is predicted by hydrodynamic theory of CWT to crossover from oscillatory to over-damped behaviour as the wavelength decreases[59]. Analysis of the temporal correlation of each wave amplitude throughout a simulation can provide further insight into the true dynamics of the system. Analysing the modes described by the ISM for hydrodynamic behaviour has not only been able to predict macroscopic surface tension in agreement with a structural approach, but that also remain valid up to much smaller wavelengths[15]. Considering recent experimental reports of the complex surface tension of water, measured using similar techniques[60,61] it would be highly interesting to investigate the temporal correlations of surface reconstructed intrinsic surfaces for further comparison with the original ISM routine.

The distribution of the surface reconstructed power spectrum $\langle|\xi(\mathbf{q})^2|\rangle$ and frequency dependent surface tension spectrum $\gamma(\mathbf{q})$ shown in 9 raises many unexpected questions. It may be that the ISM approach does support the enhanced surface tension predictions of extended CWT as long as the global distribution of surface curvature is correctly assigned. However, we cannot be sure that this is not an artefact of our reconstruction routine, since it is desired that the original form of the surface should be maintained, resulting in an effective change of sample points without altering the particle coordinates of the configuration. To achieve this we do not restrict the values of the coefficients $a_{\mu\nu}$, only the variance of

surface curvature. Nevertheless, it would be highly coincidental if this effect is unrelated to the distinction between traditional and extended CWT, considering the extremely close correlation of reconstructed $\gamma(\mathbf{q})$ to equation (7). The consequences of which are therefore worth investigating further with care.

VII. CONCLUSION

We have developed a "surface reconstruction" method that successfully improves the internal consistency of the ISM. It does so by reducing the curvature bias on surface particle pivot sites during the intrinsic surface fitting routine. The methodology is designed to be integrated very efficiently into the original computational procedure with no additional parameters requiring optimisation, resulting in only a minor addition in calculation time. We present analysis of this routine, proving its ability to enforce the correct distribution surface curvature at particle sites, resulting in a sharp reduction in sampling bias for the intrinsic density distributions close to the interface. We also show that addressing this bias produces a significant reduction in artefacts upon convolution of the intrinsic density back to the laboratory frame, an important measure of internal consistency and ability to predict experimental properties. Furthermore, the surface reconstruction routine appears to significantly alter the power spectrum of surface modes, aligning the distribution more with the predictions of extended capillary wave theories, rather than traditional CWT, which had supported by the original ISM routine. It is hoped that future research into the true nature of the power-spectrum will allow further refinement of the method.

VIII. ACKNOWLEDGEMENTS

This work was supported by the EPSRC DTC grant EP/G03690X/1. Simulations were performed using the IRIDIS 4 supercomputing facilities at the University of Southampton, UK.

We cite the developer group[62] of AMBER 12:

We acknowledge the use of the MDTraj python library[63] for analysis of simulation data.

A Python library containing source code used to perform the intrinsic sampling analysis in this investigation can be found at https://github.com/franklongford/ALIAS.

REFERENCES

[1]F. P. Buff, R. A. Lovett, and F. H. Stillinger, Phys. Rev. Lett. **15**, 621 (1965).

[2]J. D. Weeks, J. Chem. Phys. **67**, 3106 (1977).

[3]P. Linse, J. Chem. Phys. **86**, 4177 (1987).

[4]I. Benjamin, J. Chem. Phys. **97**, 1432 (1992).

[5]E. Chacón and P. Tarazona, Phys. Rev. B. **70**, 235407 (2004).

[6]E. Chacón and P. Tarazona, J. Phys.: Condens. Matter **17**, S3493S3498 (2005).

[7]E. Chacón, P. Tarazona, and J. Alejandre, J. Chem. Phys. **125**, 014709 (2006).

[8]F. Bresme, E. Chacón, and P. Tarazona, Phys. Chem. Chem. Phys. **10**, 4704 (2008).

[9]F. Bresme, E. Chacón, P. Tarazona, and K. Tay, Phys. Rev. Lett. **101**, 056102 (2008).

[10]E. Chacón, E. M. Fernández, D. Duque, R. Delgado-Buscalioni, and P. Tarazona, Phys. Rev. B **80**, 195403 (2009).

[11]M. Darvas, L. B. Pártay, P. Jedlovszky, and G. Horvai, J. Mol. Liq. **153**, 88 (2010).

[12]Y. Yang and B. B. Laird, J. Phys. Chem. B **118**, 8373 (2014).

[13]J. Hernández-Muñoz, E. Chacón, and P. Tarazona, J. Chem. Phys. **148**, 084702 (2018).

[14]D. Duque, P. Tarazona, and E. Chacón, J. Chem. Phys. **128**, 134704 (2008).

[15]R. Delgado-Buscalioni, E. Chacón, and P. Tarazona, J. Phys.: Condens. Matter **20**, 494229 (2008).

[16]B. Fábián, M. V. Senćanski, I. N. Cvijetić, P. Jedlovszky, and G. Horvai, J. Phys. Chem. C **120**, 8578 (2016).

[17]J. Muscatello, E. Chacón, P. Tarazona, and F. Bresme, Phys. Rev. Lett. **119**, 045901 (2017).

[18]A. Idrissi, G. Hantal, and P. Jedlovszky, Phys. Chem. Chem. Phys. **17**, 8913 (2015).

[19]M. Sega, B. Fábián, G. Horvai, and P. Jedlovszky, J. Phys. Chem. C **120**, 27468 (2016).

[20]M. Sega, B. Fábián, and P. Jedlovszky, J. Phys. Chem. Letts. **8**, 2608 (2017).

[21]D. Beaglehole, J. Phys. Chem. **91**, 5091 (1987).

[22]R. Greef and J. G. Frey, Phys. Stat. Sol. **5**, 1184 (2008).

[23]A. Stocco and K. Tauer, The European Physical Journal E **30**, 431438 (2009).

[24]P. Tarazona, R. Checa, and E. Chacón, Phys. Rev. Lett. **99**, 196101 (2007).

[25]F. Sedlmeier, D. Horinek, and R. R. Netz, Phys. Rev. Lett. **103**, 136102 (2009).

[26]L. G. MacDowell, J. Benet, N. A. Katcho, and J. M. Palanco, Adv. Colloid Interface Sci.

206, 150 (2014).

[27]A. O. Parry, C. Rascón, and R. Evans, Phys. Rev. E **91**, 030401 (2015).

[28]J. Chowdhary and B. M. Ladanyi, J. Phys. Chem. B **110**, 15442 (2006).

[29]L. B. Pártay, G. Hantal, P. Jedlovszky, Á. Vincze, and G. Horvai, J. Comp. Chem. **29**, 945956 (2007).

[30]M. Jorge and M. D. Cordeiro, J. Phys. Chem. C **111**, 17612 (2007).

[31]M. Jorge, P. Jedlovszky, and M. N. D. Cordeiro, J. Phys. Chem. C **114**, 11169 (2010).

[32]M. Jorge, P. Jedlovszky, and M. N. D. Cordeiro, J. Phys. Chem. C **114**, 1865618663 (2010).

[33]E. Chacón, E. M. Fernández, and P. Tarazona, Phys. Rev. E **89**, 042406 (2014).

[34]E. Chacón and P. Tarazona, Phys. Rev. Letts. **91**, 166103 (2003).

[35]P. Tarazona, E. Chacón, and F. Bresme, J. Chem. Phys. **139**, 094902 (2013).

[36]E. Chacón, P. Tarazona, and F. Bresme, J. Chem. Phys. **143**, 034706 (2015).

[37]E. Lindahl and O. Edholm, Biophys. J. **79**, 426 (2000).

[38]S. J. Marrink and A. E. Mark, J. Phys. Chem. B **105**, 6122 (2001).

[39]K. R. Mecke, J. Phys.: Condens. Matter **13**, 4615 (2001).

[40]R. L. C. Vink, J. Horbach, and K. Binder, J. Chem. Phys. **122**, 134905 (2005).

[41]E. Chacón and P. Tarazona, J. Phys.: Condens. Matter **28**, 244014 (2016).

[42]E. M. Blokhuis, J. Chem. Phys **130**, 014706 (2009).

[43]M. Napiorkówski and S. Dietrich, Phys. Rev. E **47**, 1836 (1993).

[44]K. R. Mecke and S. Dietrich, Phys. Rev. E **59**, 6766 (1999).

[45]M. Longuet-Higgins, Mathematical Proceedings of the Cambridge Philosophical Society **54**, 439 (1958).

[46]J. Meunier, Journal de Physique **48**, 1819 (1987).

[47]A. O. Parry and C. J. Boulter, J. Phys.: Condens. Matter **6**, 7199 (1994).

[48]E. M. Blokhuis, J. Groenewald, and D. Bedeaux, Mol. Phys. **96**, 397 (1999).

[49]C. Fradin, A. Braslau, D. Luzet, D. Smilgies, M. Alba, N. Boudet, K. Mecke, and J. Daillant, Nature **403**, 871 (2000).

[50]S. Mora, J. Daillant, K. Mecke, D. Luzet, A. Braslau, M. Alba, and B. Struth, Phys. Rev. Lett. **90**, 216101 (2003).

[51]D. Li, B. Yang, B. Lin, M. Meron, J. Gebhardt, T. Graber, and S. A. Rice, Phys. Rev. Lett. **92**, 136102 (2004).

[52]P. Pershan, Colloids Surf., A **171**, 149 (2000).

[53]O. Shpyrko, P. Huber, A. Grigoriev, P. Pershan, B. Ocko, H. Tostmann, and M. Deutsch, Phys. Rev. B. **69**, 245423 (2004).

[54]J. Meunier, Physica A **230**, 27 (1996).

[55]J. G. Kirkwood and F. P. Buff, J. Chem. Phys. **19**, 774 (1951).

[56]F. Höfling and S. Dietrich, EPL **109**, 46002 (2015).

[57]M. Sega, B. Fábián, and P. Jedlovszky, The Journal of Chemical Physics **143**, 114709 (2015).

[58]C. Braga, J. Muscatello, G. Lau, E. A. Müller, and G. Jackson, The Journal of Chemical Physics **144**, 044703 (2016).

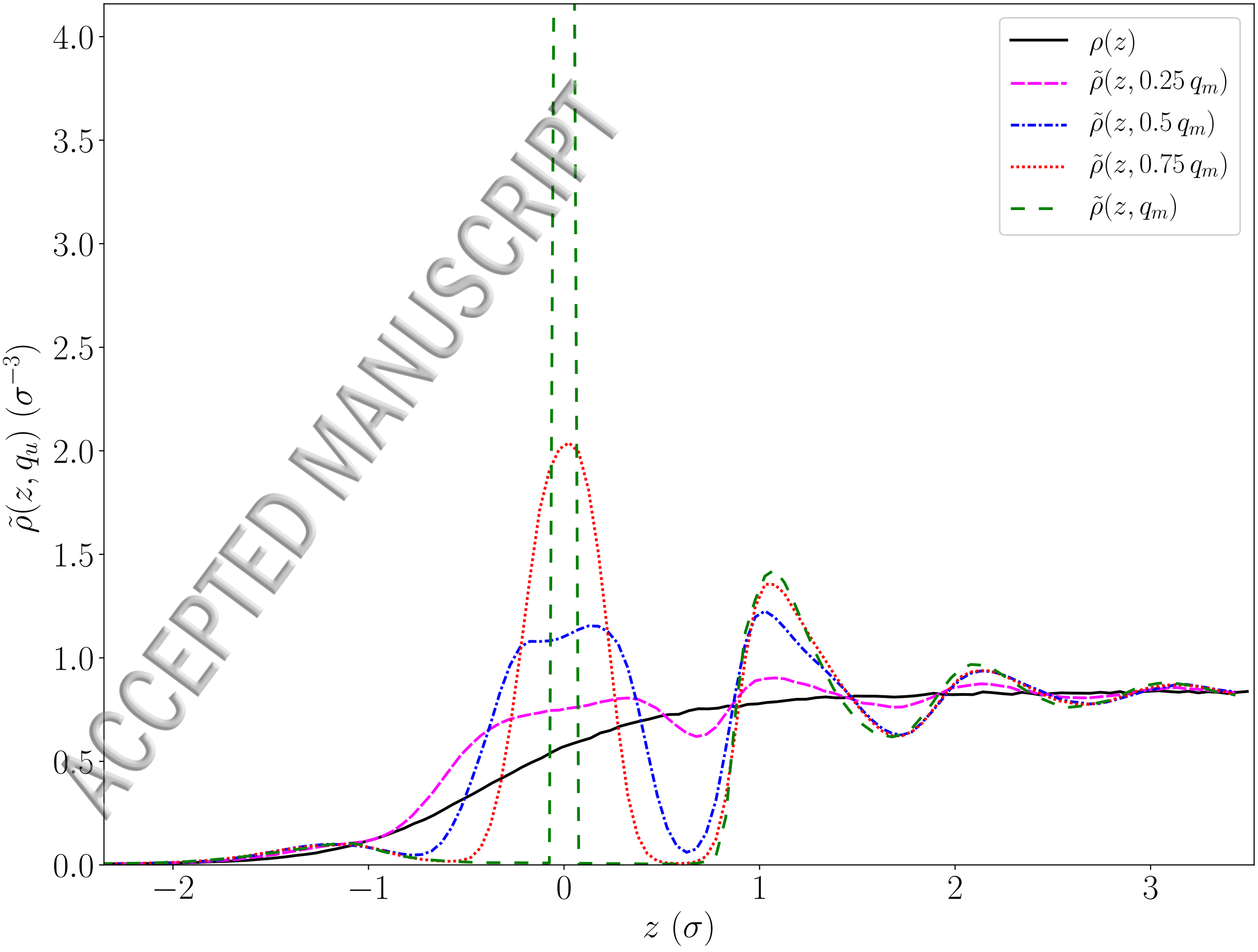
[59]U.-S. Jeng, L. Esibov, L. Crow, and A. Steyerl, J. Phys.: Condens. Matter **10**, 4955 (1998).

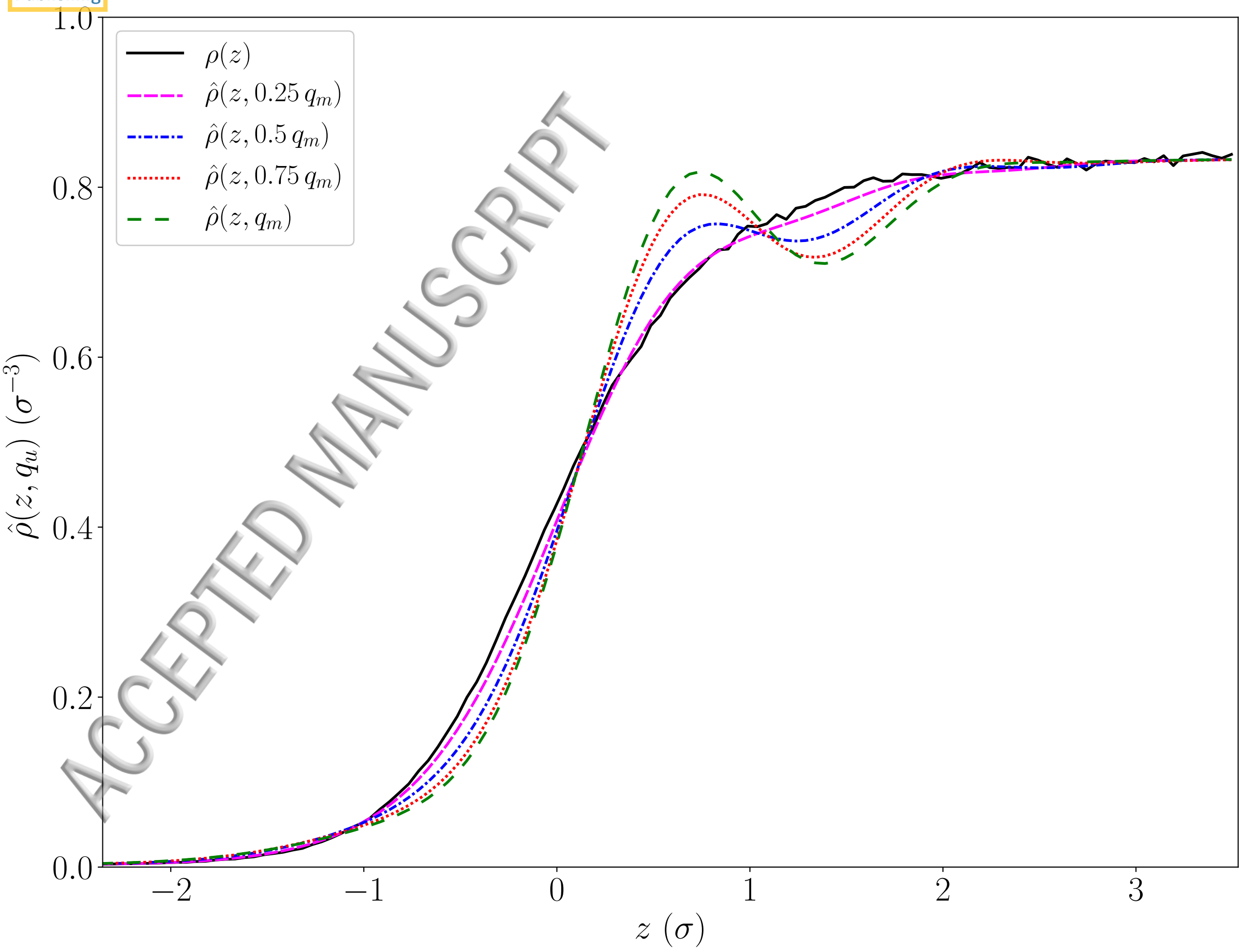
[60]X. Xiao-Min, C. Lan, Z. Wen-Long, L. Long-Fei, Y. Yue-Bin, P. Zhi-Yong, and Z. Jin-Xiu, Chin. Phys. Letts. **31**, 076801 (2014).

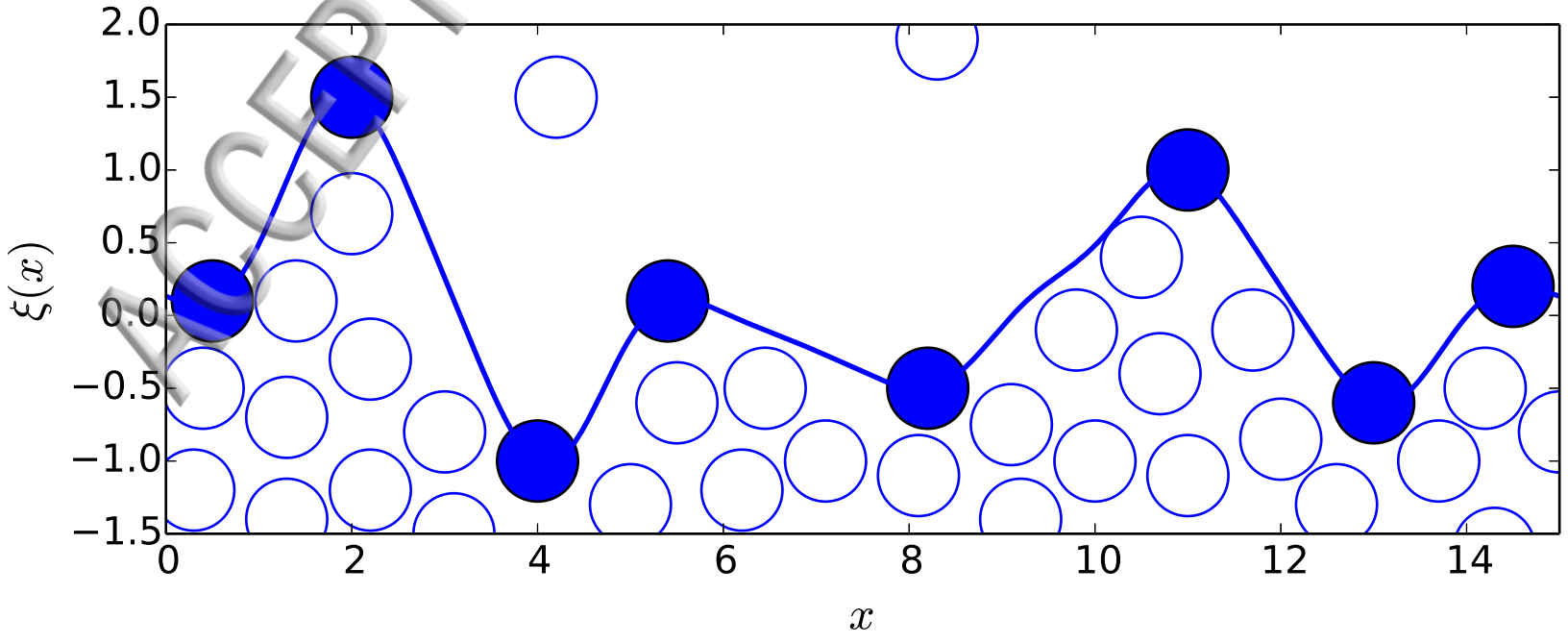
[61]L. Chen, W. Zuo, L. Li, Y. Yang, J. Zhang, and X. Xiong, J. Phys. Chem. C **118**, 26877 (2014).

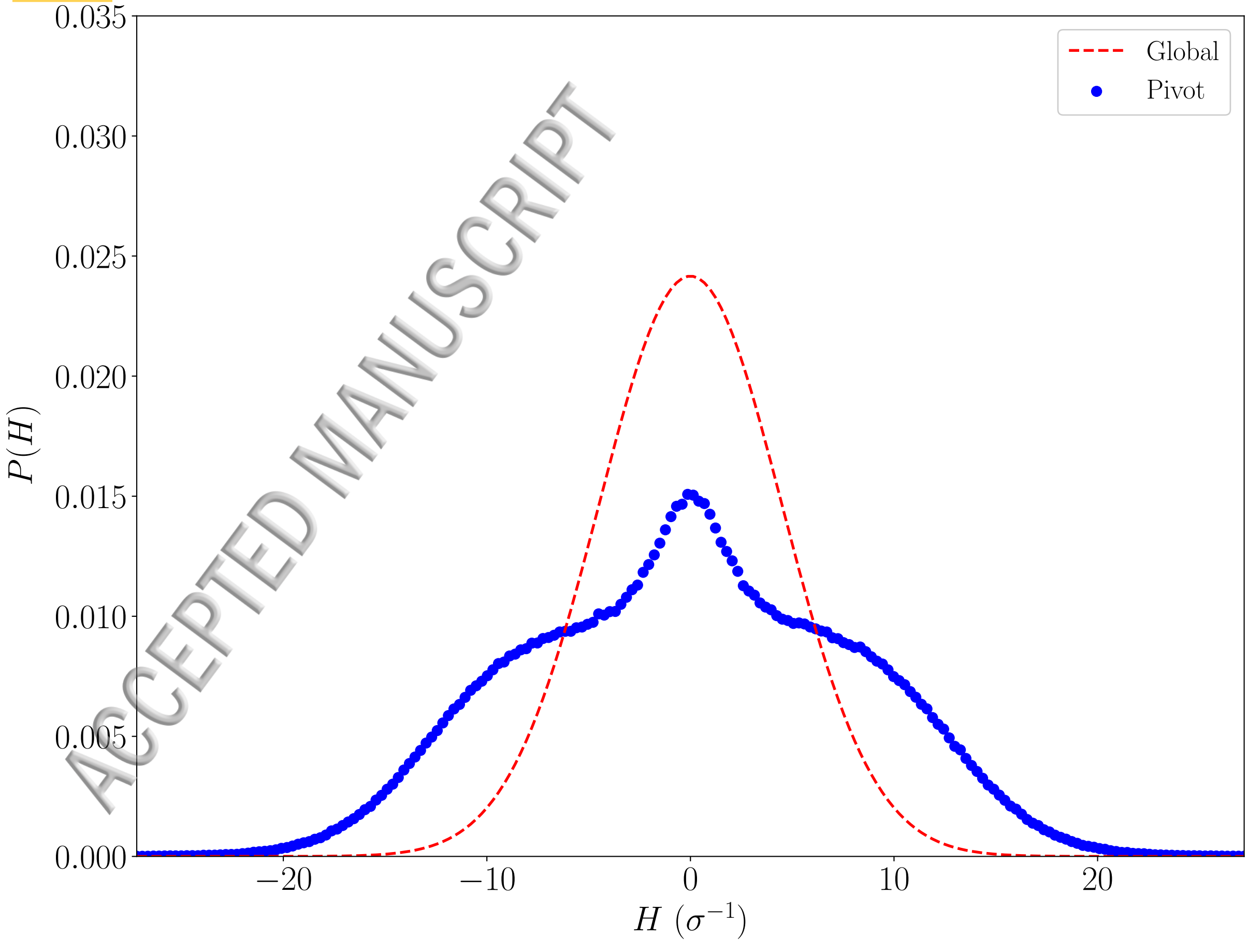
[62]D. Case, T. Dardena, T. C. III, C. Simmerling, J. Wang, R. Duke, R. Luo, R. Walker, W. Zhang, K. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. Götz, I. Kolossváry, K. Wong, F. Paesani, J. Vanicek, R. Wolf, J. Liu, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. Roe, D. Mathews, M. Seetin, R. Salomon-Ferrer, C. Sagui, T. L. V. Babin, S. Gusarov, A. Kovalenko, and P. Kollman, AMBER 12, University of California, San Francisco (2012).

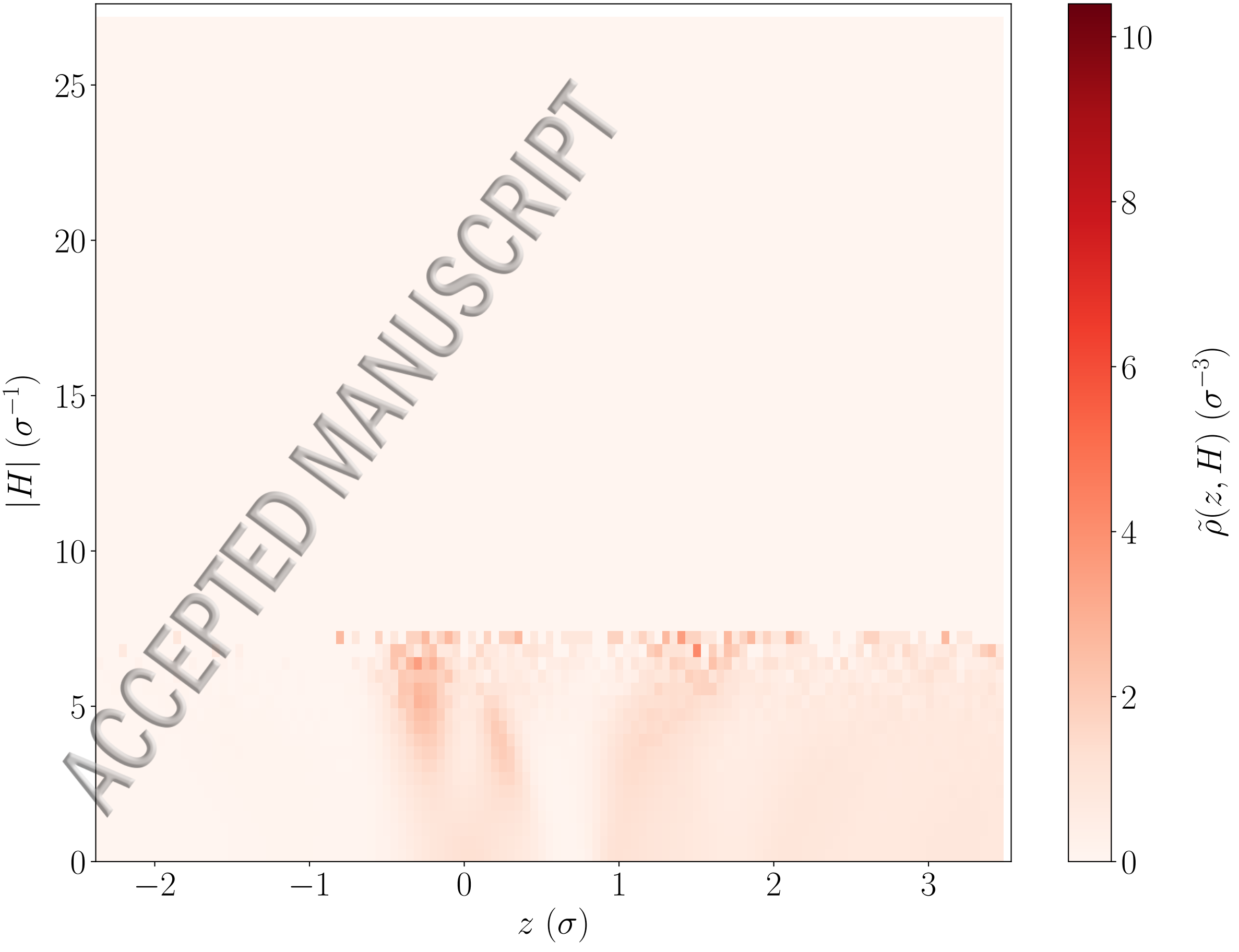
[63]R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, and V. S. Pande, Biophys. J. **109**, 1528 (2015).

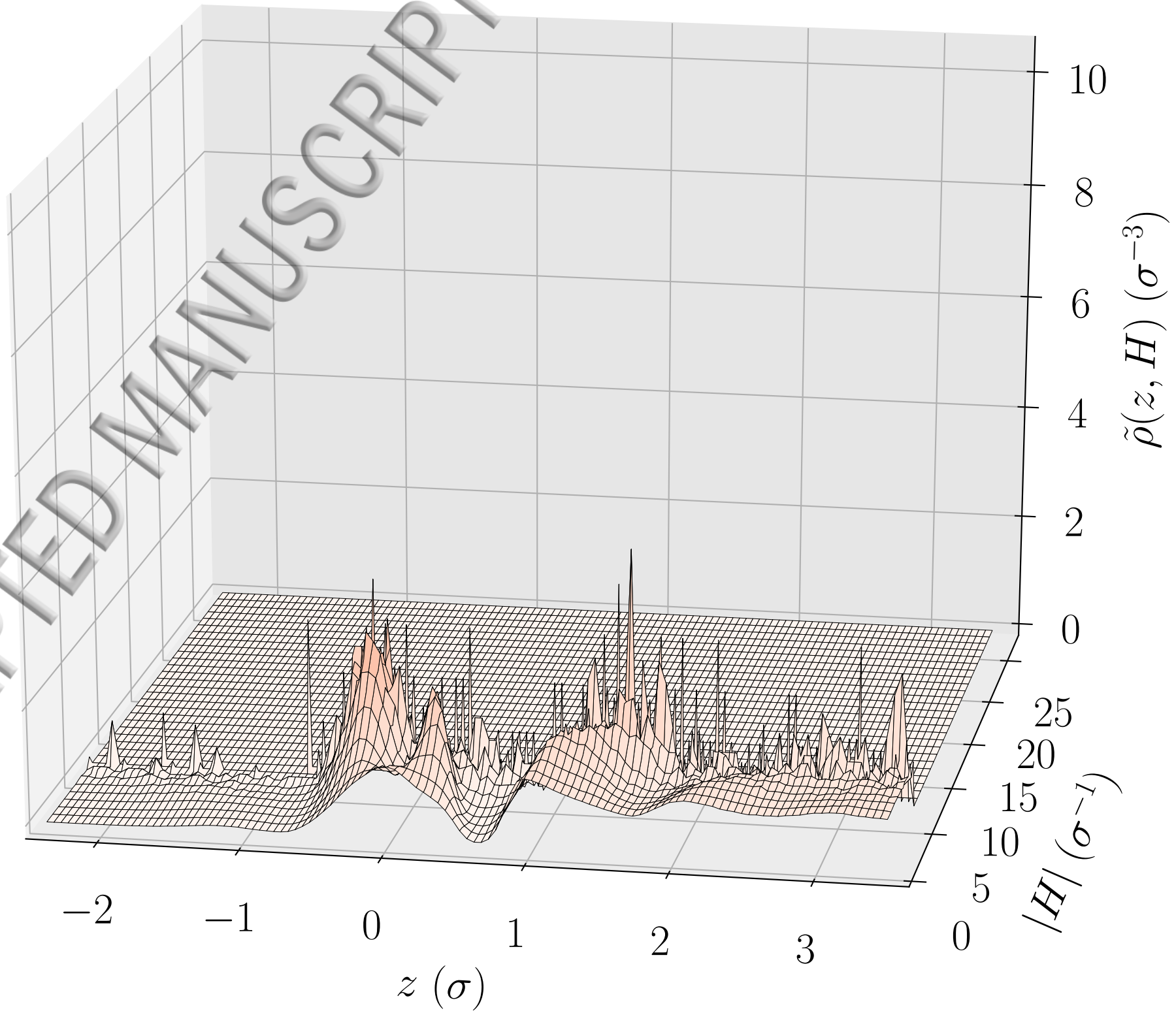


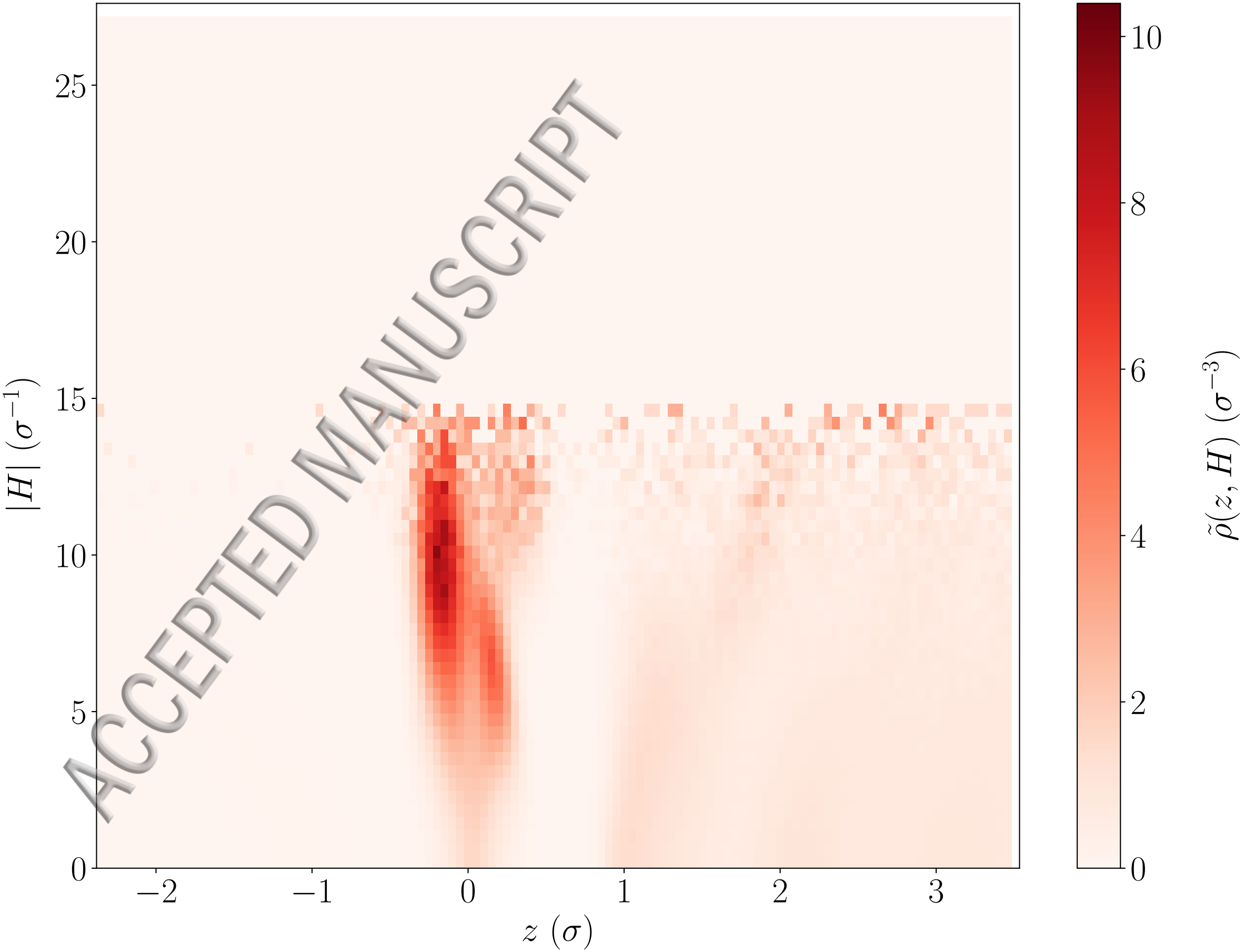


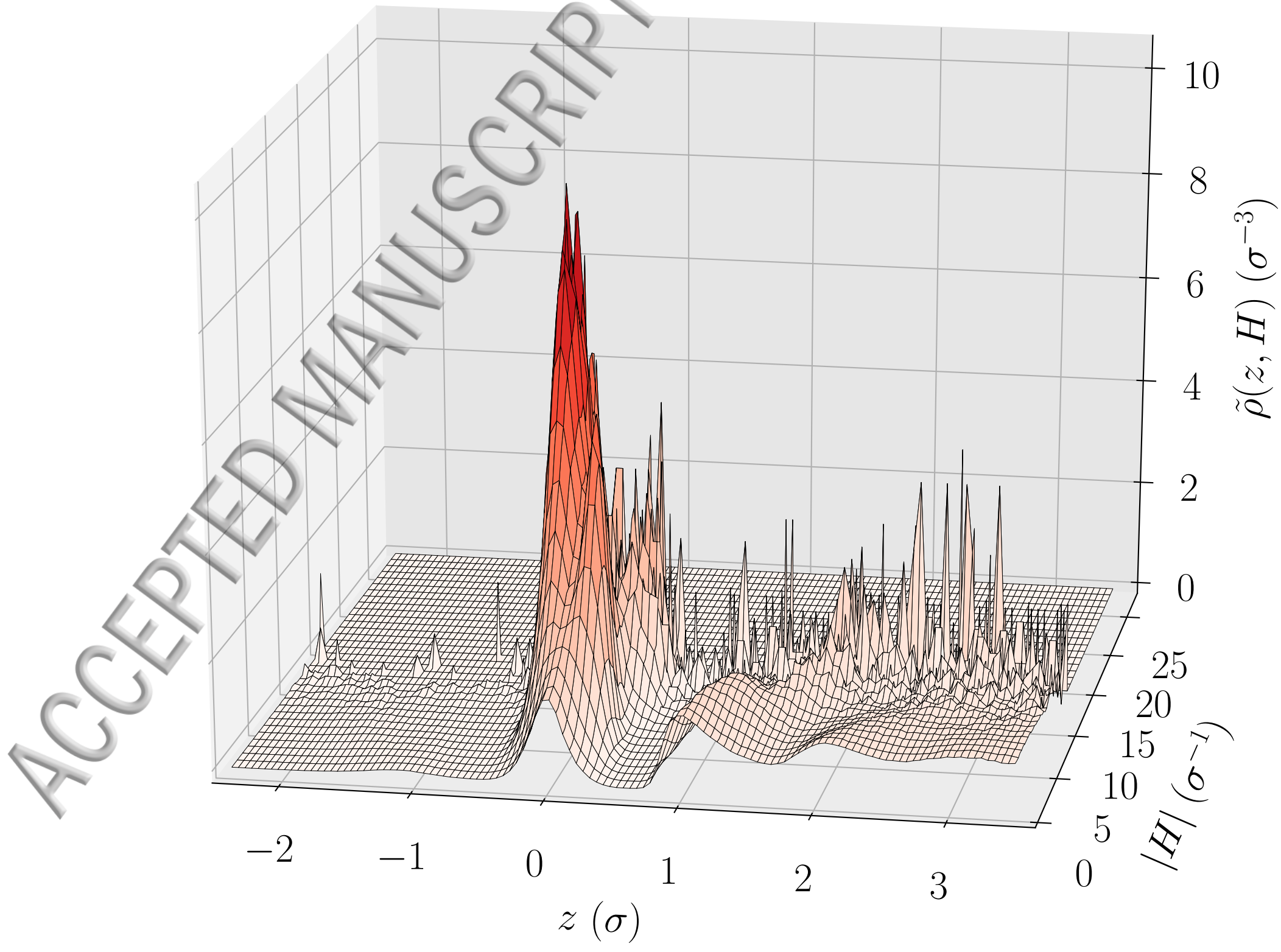


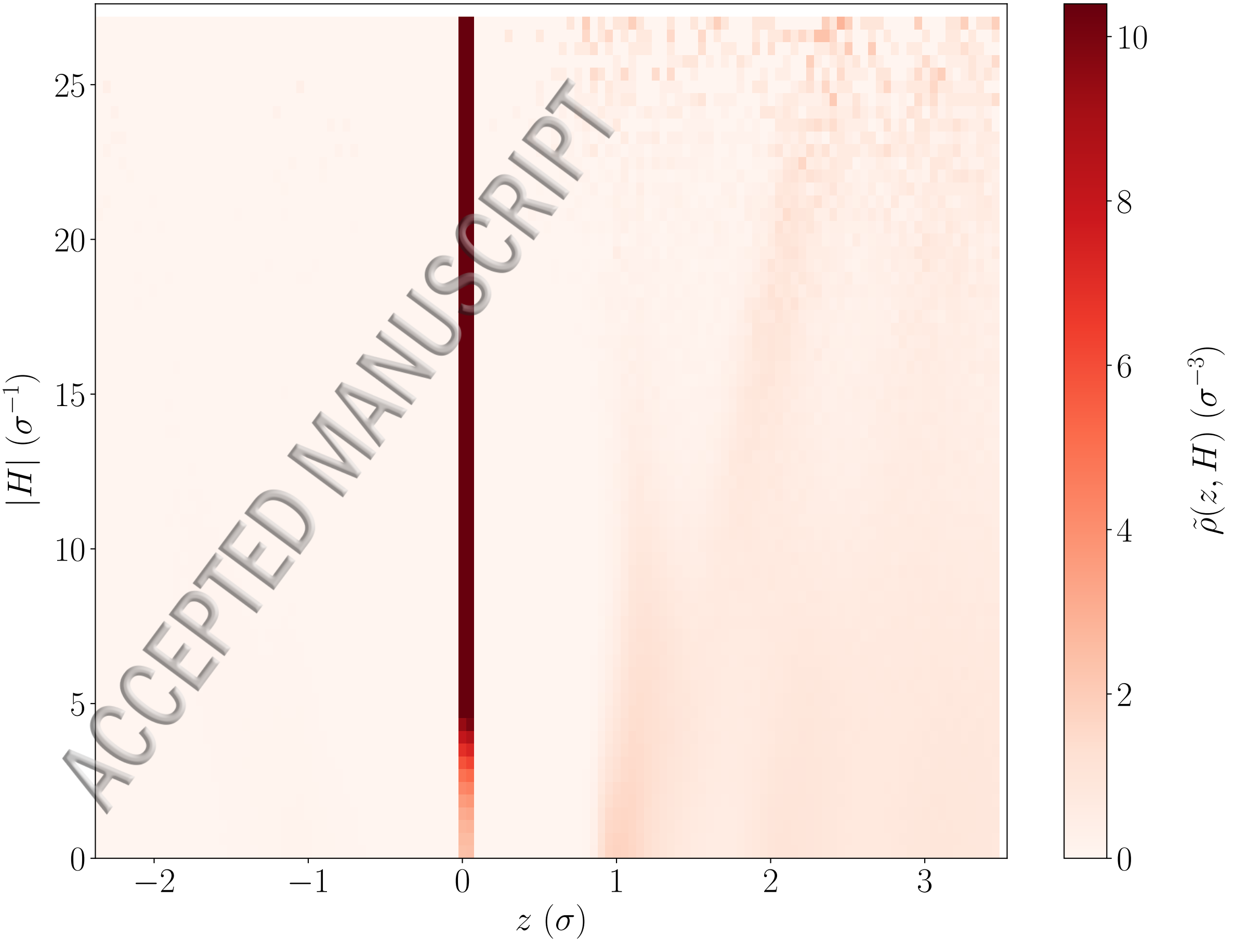


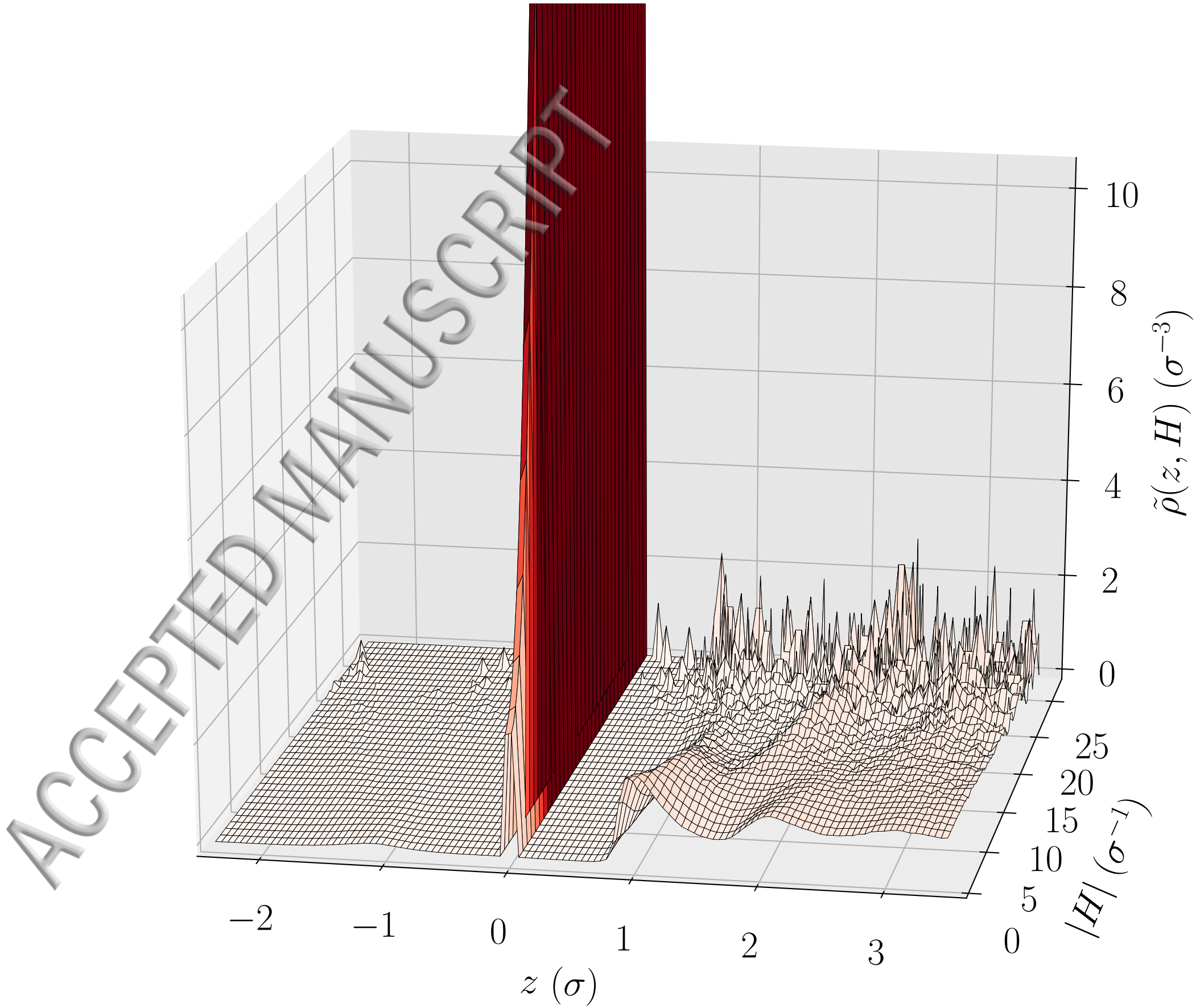

ACCEPTED MANUSCRIPT
$|H|\ (\sigma^{-1})$
0
5
10
15
20
25
$z\ (\sigma)$
−2
−1
0
1
2
3
$\tilde{\rho}(z, H)\ (\sigma^{-3})$
0
2
4
6
8
10

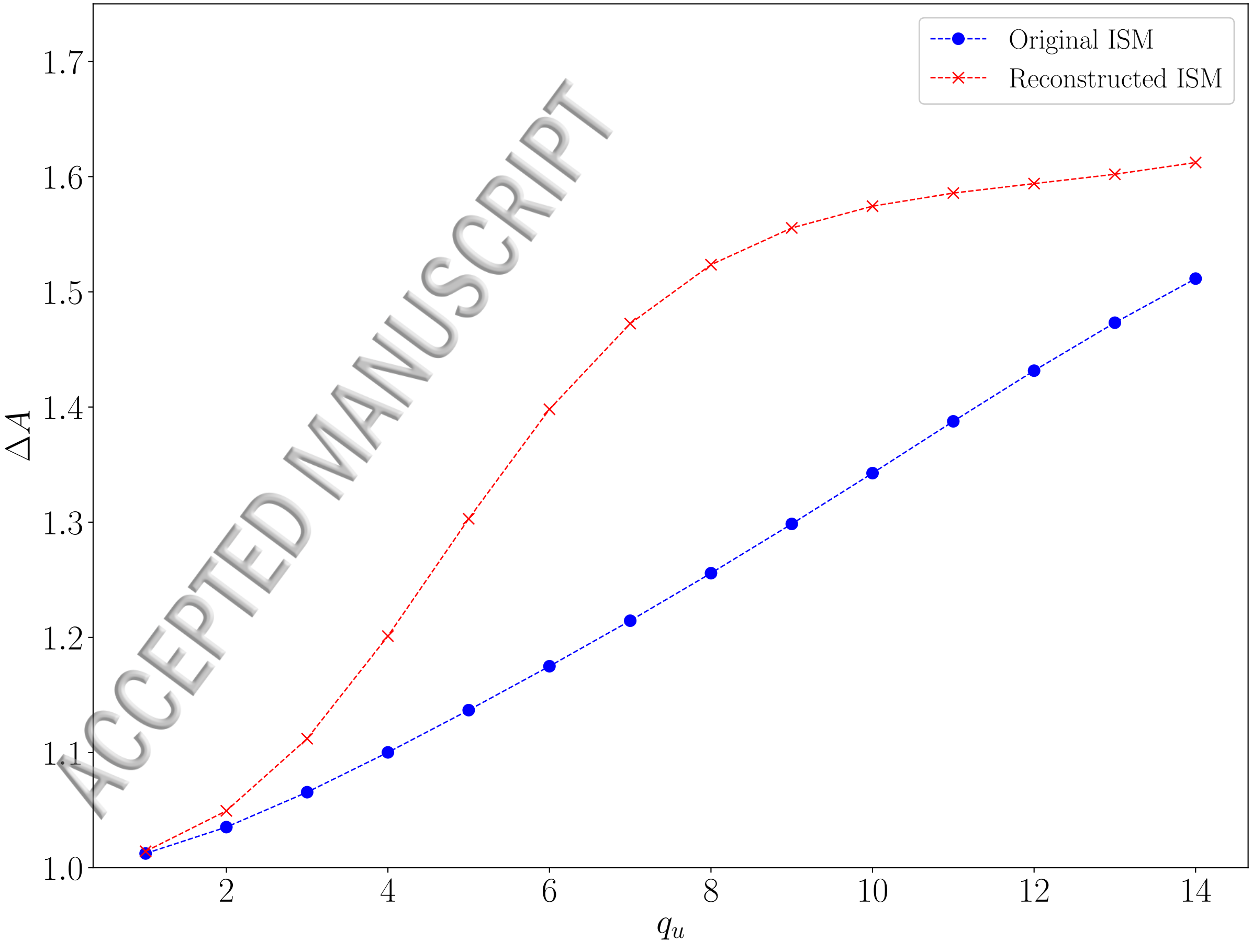


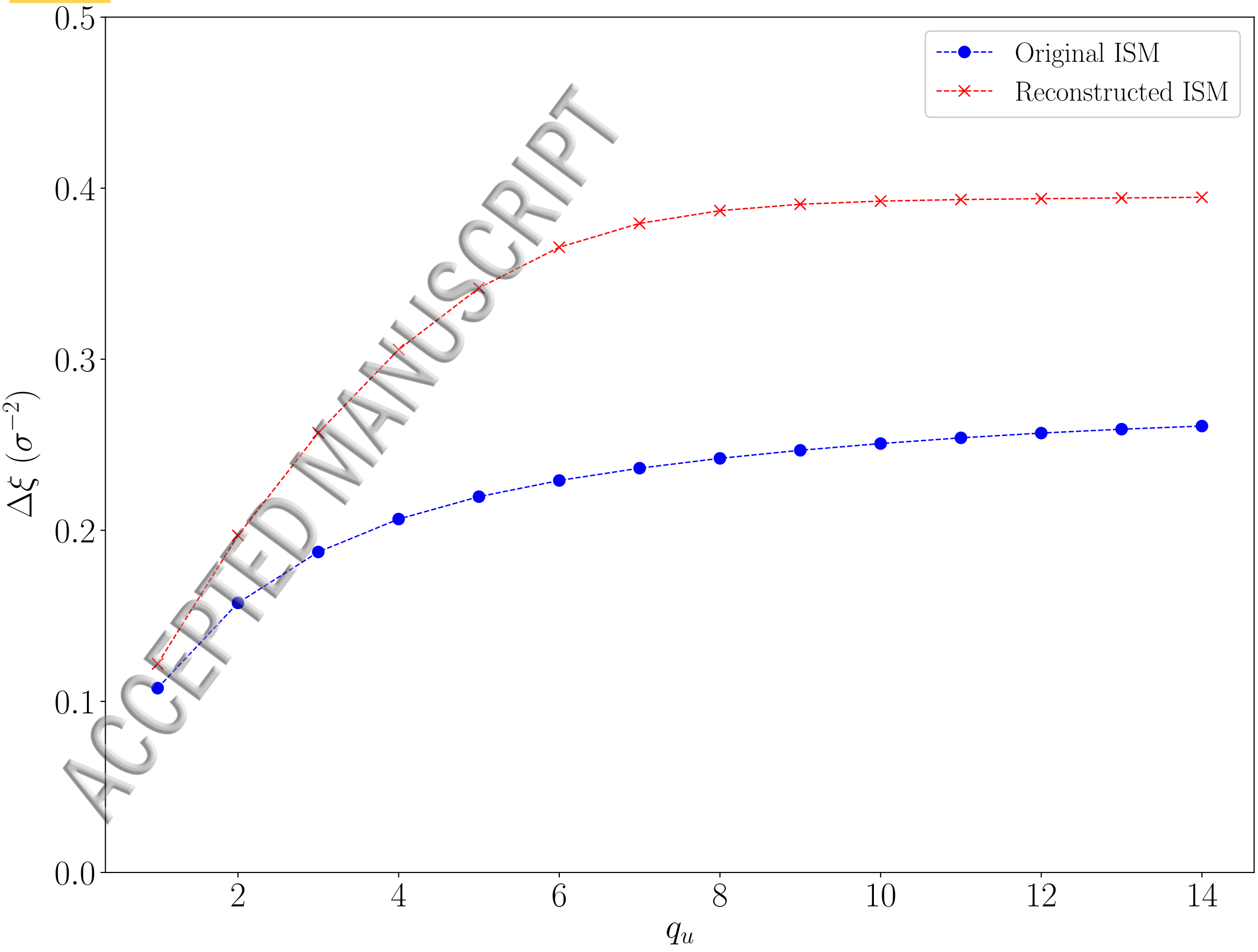


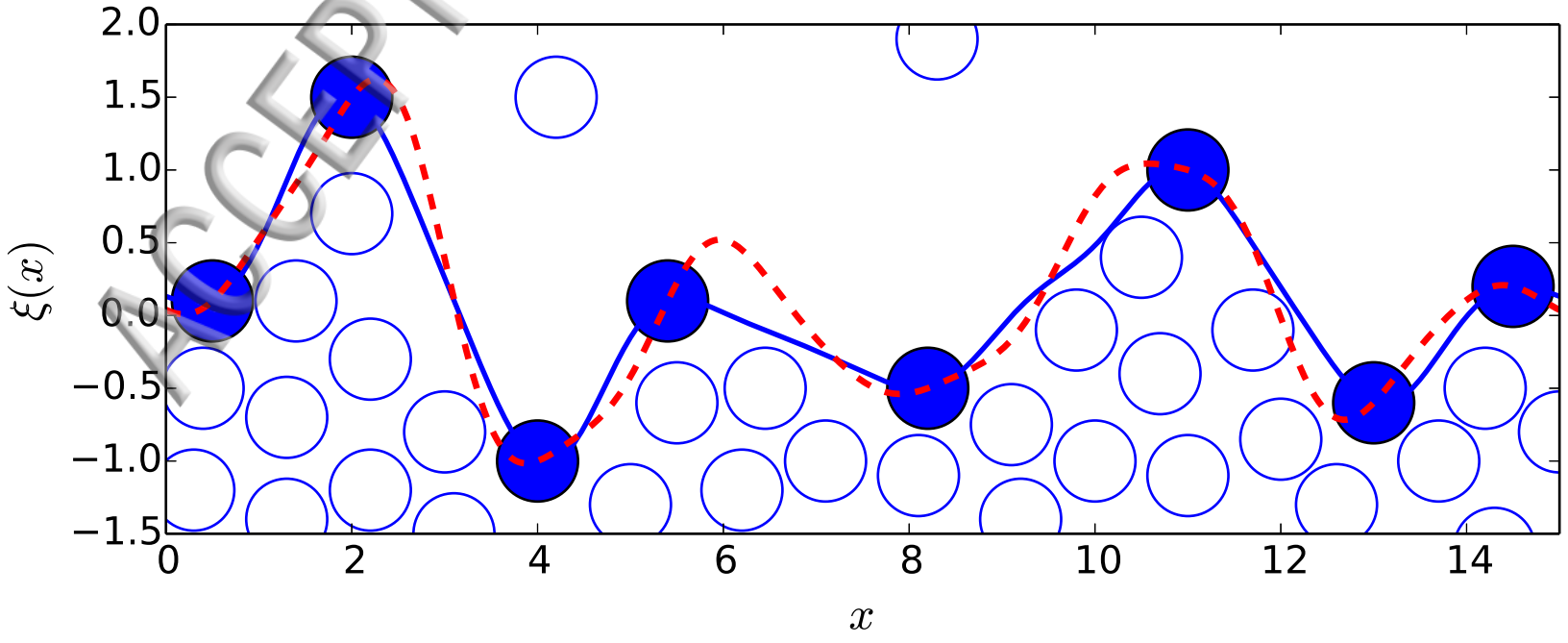


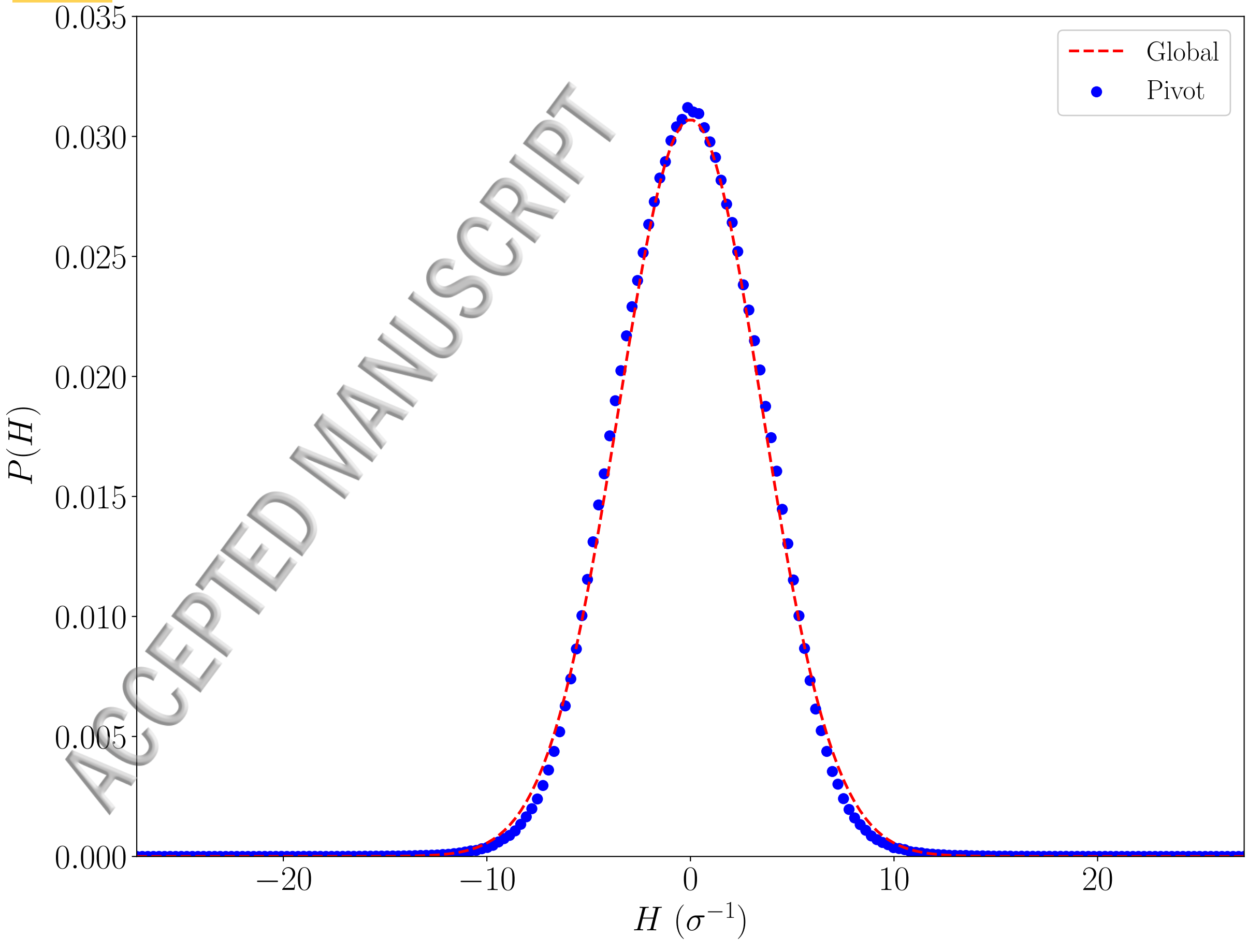


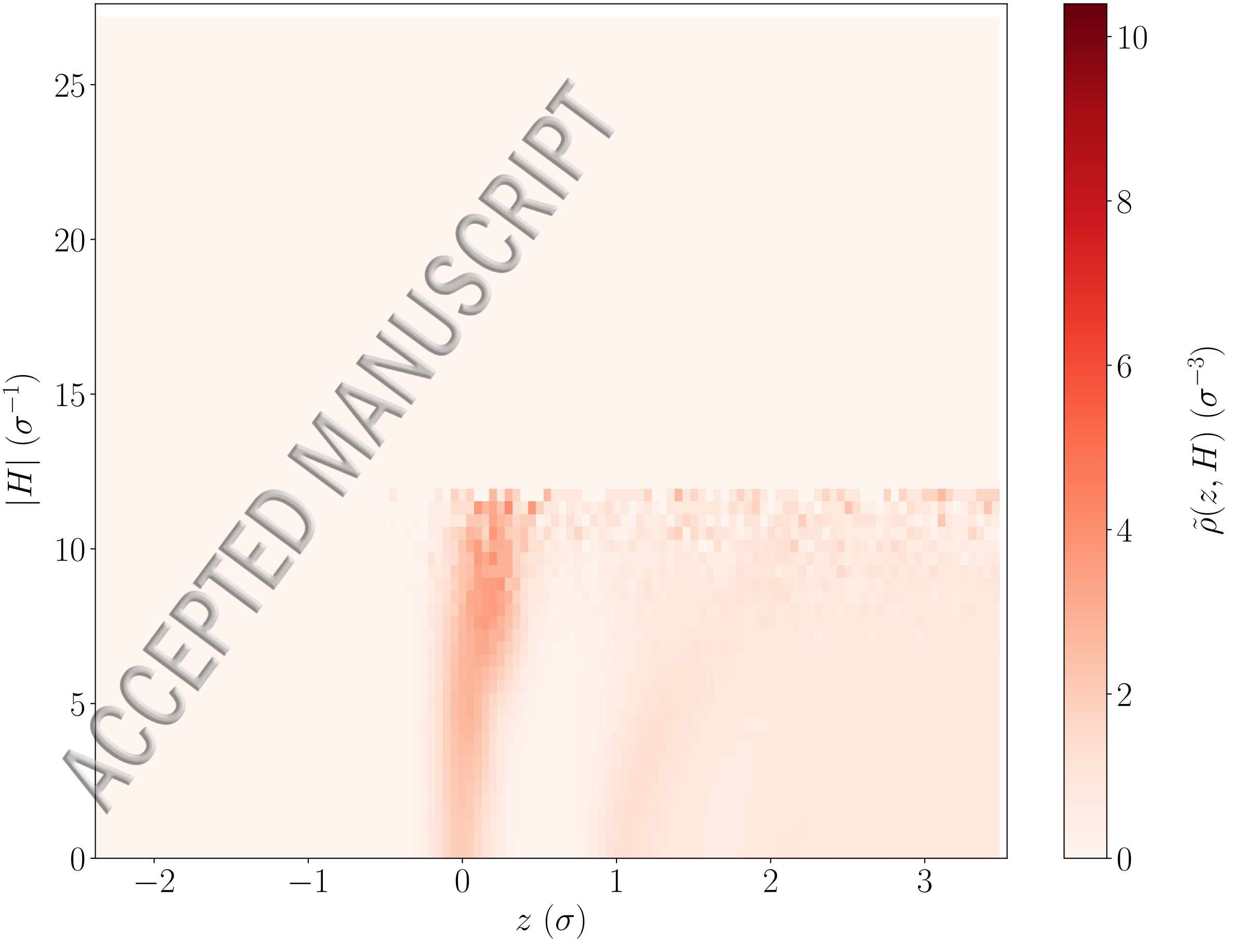


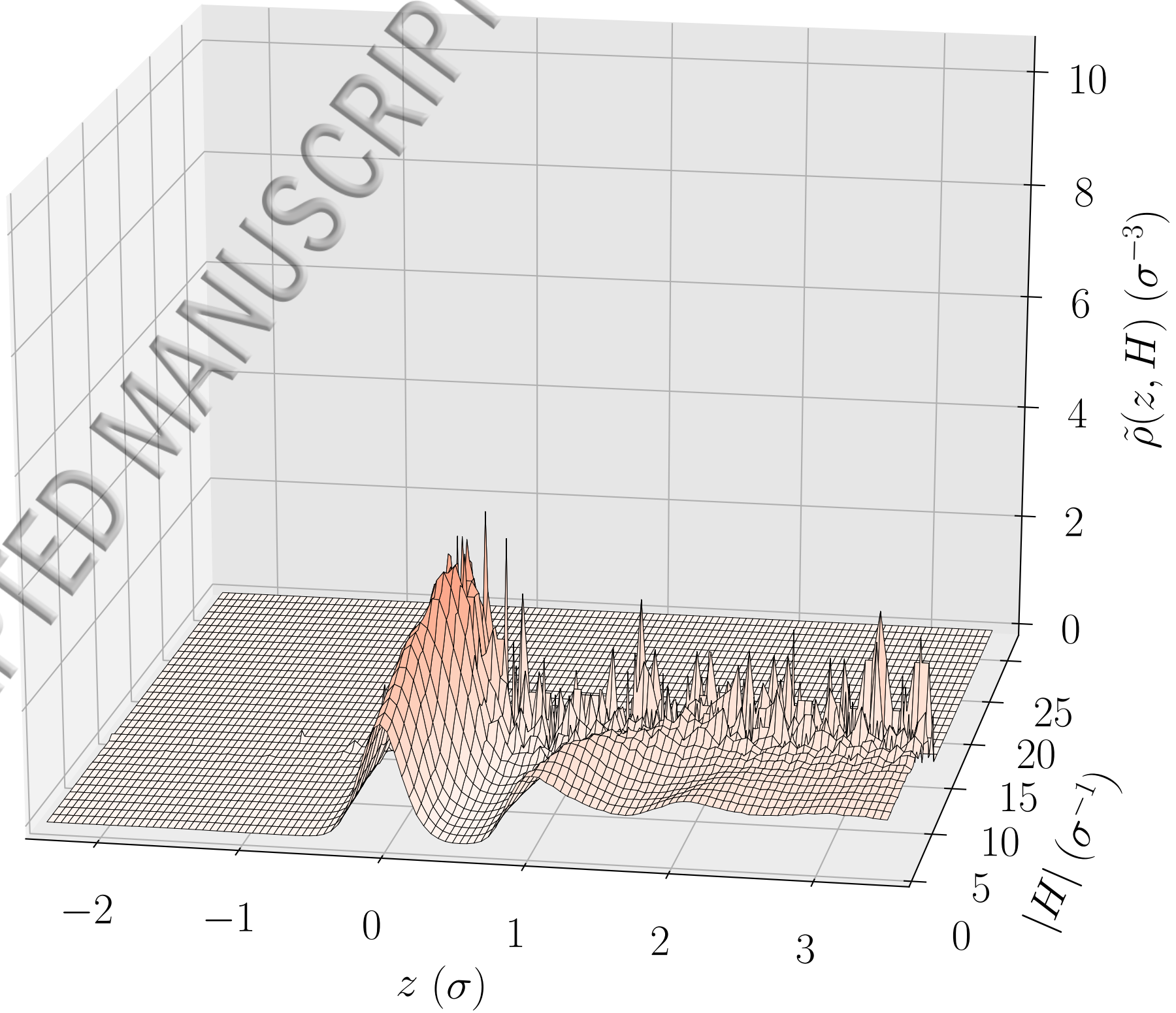


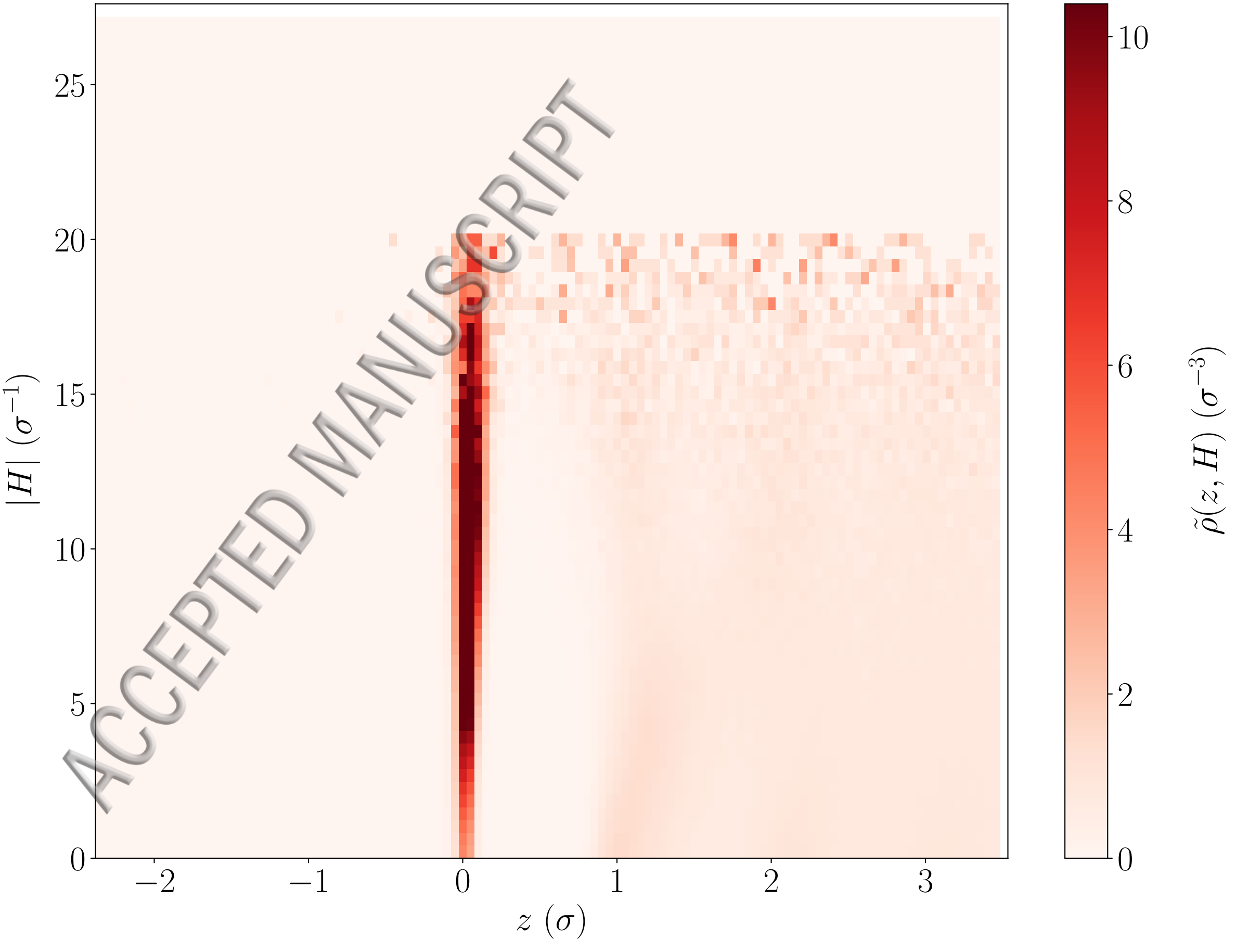



2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
$\xi(x)$
0
2
4
6
8
10
12
14
x




ACCEPTED MANUSCRIPT
$|H|\ (\sigma^{-1})$
0
5
10
15
20
25
$z\ (\sigma)$
−2
−1
0
1
2
3
$\tilde{\rho}(z, H)\ (\sigma^{-3})$
0
2
4
6
8
10




ACCEPTED MANUSCRIPT
$|H|$ (σ^{-1})
25
20
15
10
5
0
−2
−1
0
1
2
3
z (σ)
10
8
6
4
2
0
$\tilde{\rho}(z, H)$ (σ^{-3})