Notes: A Water Model with an Explicit Electron Density: from X-ray and DFT to ML and Surface Potential

Zhao Yihao
zhaoyihao@protonmail.com

July 22, 2024

1 Theoretical Basis of X-ray Scattering and Atomic Form Factors

The X-ray scattering amplitude reflects the overall distribution of electrons within the system; that is,

$$I(\mathbf{k}) = \frac{1}{M}\langle|\hat{\rho}(\mathbf{k},\bar{\mathbf{R}})|^2\rangle, \tag{1}$$

where M is the number of molecules, $\bar{\mathbf{R}}$ denotes the set of nuclear coordinates, and $\rho_e(\mathbf{r})$ represents the total electron density distribution of the system. It can therefore be expanded as:

$$\begin{aligned} I(\mathbf{k}) =& \frac{1}{M}\left\langle\left|\int_{-\infty}^{\infty} d\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}}\rho_e(\mathbf{r})\right|^2\right\rangle \\ =& \frac{1}{M}\left\langle\left|\int_{-\infty}^{\infty} d\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}}\sum_{i=1}^{N}\delta(\mathbf{r}-\mathbf{R}_i)\star\rho_e^i(\mathbf{r})\right|^2\right\rangle \end{aligned} \tag{2}$$

Applying the convolution theorem,

$$I(\mathbf{k}) = \frac{1}{M}\left\langle\left|\sum_{i=1}^{N} e^{i\mathbf{k}\cdot\mathbf{R}_i}\mathscr{F}[\rho_e^i](\mathbf{k})\right|^2\right\rangle \tag{3}$$

Although the electron cloud distribution of identical atoms in the same molecule may vary slightly due to their local environments, a key approximation can be made: these distributions are assumed to be identical. This allows the Fourier transforms of the electron density to be factored out of the ensemble average. For a pure water system, this leads to:

$$\begin{aligned} I(\mathbf{k}) =& S_{\mathrm{OO}}(\mathbf{k})\cdot|\mathscr{F}[\rho_{\mathrm{O}}]|^2 \\ &+2S_{\mathrm{OH}}(\mathbf{k})\cdot\mathscr{F}[\rho_{\mathrm{O}}]\mathscr{F}[\rho_{\mathrm{H}}] + S_{\mathrm{HH}}(\mathbf{k})\cdot|\mathscr{F}[\rho_{\mathrm{H}}]|^2, \end{aligned} \tag{4}$$

where the structure factors are defined as:

$$S_{\mathrm{OO}}(\mathbf{k}) = \frac{1}{M}\left\langle \left| \sum_{i\in \mathrm{O}} e^{i\mathbf{k}\cdot\mathbf{R}_i} \right|^2 \right\rangle, \tag{5}$$

$$S_{\mathrm{OH}}(\mathbf{k}) = \frac{1}{2M}\left\langle \sum_{i\in \mathrm{O}} \sum_{j\in \mathrm{H}} (e^{i\mathbf{k}\cdot(\mathbf{R}_i-\mathbf{R}_j)} + e^{i\mathbf{k}\cdot(\mathbf{R}_j-\mathbf{R}_i)}) \right\rangle, \tag{6}$$

$$S_{\mathrm{HH}}(\mathbf{k}) = \frac{1}{M}\left\langle \left| \sum_{i\in \mathrm{H}} e^{i\mathbf{k}\cdot\mathbf{R}_i} \right|^2 \right\rangle. \tag{7}$$

The Fourier transform of a single atom's electron density is termed its atomic form factor:

$$f_{\mathrm{O}}(\mathbf{k}) = \mathscr{F}[\rho_{\mathrm{O}}], \quad f_{\mathrm{H}}(\mathbf{k}) = \mathscr{F}[\rho_{\mathrm{H}}]. \tag{8}$$

It is commonly assumed that the electron cloud is spherically symmetric, implying that the form factor depends only on the magnitude of the wave vector: $f(\mathbf{k}) = f(k)$. Consequently, the X-ray scattering intensity for water can be compactly expressed as a function of these radially symmetric form factors and the partial structure factors:

$$\begin{aligned} I(\mathbf{k}) = {} & S_{\mathrm{OO}}(\mathbf{k}) \cdot f_{\mathrm{O}}^2(k) \\ & + 2S_{\mathrm{OH}}(\mathbf{k}) \cdot f_{\mathrm{O}}(k) f_{\mathrm{H}}(k) + S_{\mathrm{HH}}(\mathbf{k}) \cdot f_{\mathrm{H}}^2(k). \end{aligned} \tag{9}$$

This formula encounters different problems depending on the context in which it is applied.

First, the problem of attributing electron density in *ab initio* calculations. In molecular dynamics-based quantum chemical calculations, the atomic configurations of the system (i.e., $\bar{\mathbf{R}}$) are known. Therefore, the partial structure factors $S_{\mathrm{OO}}(k)$, $S_{\mathrm{OH}}(k)$, and $S_{\mathrm{HH}}(k)$ can be directly calculated from the trajectories. At this point, the **only two unknowns** in the scattering intensity formula are the atomic form factors $f_{\mathrm{O}}(k)$ and $f_{\mathrm{H}}(k)$. Although a single equation contains two unknowns, the fact that the amplitude of the hydrogen form factor is much smaller than that of oxygen ($|f_{\mathrm{H}}(k)| \ll |f_{\mathrm{O}}(k)|$) allows for the simultaneous high-precision determination of both $f_{\mathrm{O}}(k)$ and $f_{\mathrm{H}}(k)$ under reasonable approximations. The resulting bulk water electron density distribution $\rho_e(\mathbf{r})$ not only accurately describes the electron distribution but also possesses a clear atomic attribution (i.e., it can be decomposed into contributions from f_{O} and f_{H}). This attributed, precise electron density distribution serves two main purposes: on one hand, it can be used to correct the surface potential of classical water models; on the other hand, it provides a foundation for the rigorous separation of long– and short-range electrostatic interactions in machine learning *ab initio* datasets.

Second, the problem of resolving microscopic structure based on experimental measurement data. In experiments, $I(k)$ is the directly measurable quantity. To extract microscopic structural information about water

(e.g., the oxygen-oxygen radial distribution function $g_{\mathrm{OO}}(r)$) from $I(k)$, one must first obtain the partial structure factors $S_{\alpha\beta}(k)$. This necessitates introducing form factors into the formula. The closer the form factor is to the true physical situation, the more accurate the inversely solved $S_{\mathrm{OO}}(k)$ will be, and consequently, the more precise the $g_{\mathrm{OO}}(r)$ obtained via inverse Fourier transform. This is the standard practice for experimentally determining the RDF of water.

2 Obtaining Atom-Attributed Electron Density from Ab Initio Calculations

The simplest and most direct source for the form factor is the electron density of an isolated, ground-state, spherically symmetric atom. This is known as the Independent Atom Model (IAM). We denote the IAM form factors for the water molecule as $f_{\mathrm{H},0}$ and $f_{\mathrm{O},0}$, respectively. In particular, for an isolated hydrogen atom, we can directly utilize the analytic solution to its Schrödinger equation:

$$\rho_{\mathrm{H},0}(\mathbf{r}) = \frac{1}{\pi a_0^3} e^{-\frac{2r}{a_0}}, \tag{10}$$

where $a_0 \approx 0.529\,\text{Å}$ is the Bohr radius. Its corresponding form factor is obtained via the Fourier transform:

$$f_{\mathrm{H},0}(k) = \mathscr{F}[\rho_{\mathrm{H},0}](k) = \frac{16}{(k^2 a_0^2 + 4)^2}. \tag{11}$$

For atoms with higher atomic numbers, although an analytical expression may not be available, the atomic form factors can be accurately represented by a six-Gaussian function fit to *ab initio* data. As proposed by Zhengwei Su and Philip Coppens in 1997 [**?**], the functional form is:

$$f_{\mathrm{X},0}(k) = \sum_{j=1}^{6} a_j e^{-b_j k^2/(8\pi)^2}, \tag{12}$$

where the index X denotes the atomic species. For oxygen and hydrogen, the fitted parameters a_j and b_j are listed in Table 1.

2.1 Limitations of IAM and the MAFF Model

However, the IAM does not account for effects such as chemical bonding, dielectric environment, and the resulting distortion of electron cloud distributions. To address these limitations, Sorenson *et al.* proposed the Modified Atomic Form Factor (MAFF). We denote the MAFF for hydrogen and oxygen as $f_{\mathrm{H},1}$ and $f_{\mathrm{O},1}$, respectively.

Table 1: Parameters for the six-Gaussian fit of the e where $Z_{\mathrm{O}} = 8$ for oxygen and $Z_{\mathrm{H}} = 1$ for hydrogen are the atomic numbers. a_j and b_j are the coefficients and variance of the Gaussians for an isolated X-atom [**?**].

X	a_1 b_1	a_2 b_2	a_3 b_3	a_4 b_4	a_5 b_5	a_6 b_6
H	0.43028	0.28537	0.17134	0.09451	0.01725	0.00114
	23.02312	10.20138	51.25444	4.13511	1.35427	0.24269
O	2.34752	1.83006	1.61528	1.52402	0.41423	0.26867
	9.6971	18.59876	5.19879	0.32408	39.79099	0.0115

These modified form factors are obtained by transforming the IAM form factors with a charge transfer parameter q, which quantifies the amount of electron density transferred from each hydrogen atom to oxygen during bond formation. The specific transformations are given by:

$$f_{\mathrm{H},1}(k) = f_{\mathrm{H},0}(k)\left[1 - \frac{q}{Z_{\mathrm{H}}} e^{-\frac{k^2\sigma^2}{4}}\right], \tag{13}$$

$$f_{\mathrm{O},1}(k) = f_{\mathrm{O},0}(k)\left[1 + \frac{2q}{Z_{\mathrm{O}}} e^{-\frac{k^2\sigma^2}{4}}\right], \tag{14}$$

where $Z_{\mathrm{H}} = 1$ and $Z_{\mathrm{O}} = 8$ are the atomic numbers of hydrogen and oxygen, respectively, and σ is a width parameter characterizing the spatial extent of the charge transfer.

The parameter q directly determines the dipole moment μ of the water molecule in the MAFF model:

$$\mu = |2q(\mathbf{r}_{\mathrm{OH}_1} + \mathbf{r}_{\mathrm{OH}_2})|, \tag{15}$$

where $\mathbf{r}_{\mathrm{OH}_1}$ and $\mathbf{r}_{\mathrm{OH}_2}$ are the displacement vectors from the oxygen atom to the first and second hydrogen atoms, respectively.

Since the form factor corresponds to a real-space electron density distribution with definite physical meaning, the density must be non-negative everywhere. The MAFF model, however, does not guarantee this fundamental requirement. We can demonstrate this explicitly by deriving the real-space density $\rho_{\mathrm{H},1}(r)$ from the MAFF form factor:

$$\begin{aligned}\rho_{\mathrm{H},1}(r) =& \mathscr{F}^{-1}[f_{\mathrm{H},0}] - \mathscr{F}^{-1}\left[f_{\mathrm{H},0}\frac{q}{Z_{\mathrm{H}}} e^{-\frac{k^2\sigma^2}{4}}\right] \\ =& \rho_{\mathrm{H},0} - \frac{q}{Z_{\mathrm{H}} a_0^4}\frac{16}{(2\pi)^3} I\end{aligned} \tag{16}$$

with

$$\begin{aligned} I =& \int_{-\infty}^{\infty} d\mathbf{k}\, \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{(k^2+(2/a_0)^2)^2} e^{-k^2\sigma^2/4} \\ =& 2\pi \int_0^{\infty} dk\, \frac{k^2}{(k^2+(2/a_0)^2)^2} e^{-k^2\sigma^2/4} \int_0^{\pi} d\theta\, \sin(\theta) e^{ikr\cos(\theta)} \\ =& \frac{4\pi}{r} G, \end{aligned} \tag{17}$$

defining $\beta = 2/a_0$ and $\alpha = \sigma^2/4$

$$G = \int_0^{\infty} dk\, \frac{k\sin(kr)}{(k^2+\beta^2)^2} e^{-\alpha k^2} \tag{18}$$

The integral for G can be evaluated starting from a fundamental integral proven during my derivation of the 2D Ewald summation (see Appendix of my thesis):

$$\begin{aligned} F =& \int_0^{\infty} dk\, \frac{\cos(kr)}{k^2+\beta^2} e^{-\alpha k^2} \\ =& \frac{\pi}{4\beta} e^{\beta^2\alpha} \left[e^{\beta r} \mathrm{erfc}\left(\beta\sqrt{\alpha} + \frac{r}{2\sqrt{\alpha}} \right) + e^{-\beta r} \mathrm{erfc}\left(\beta\sqrt{\alpha} - \frac{r}{2\sqrt{\alpha}} \right) \right] \end{aligned} \tag{19}$$

Observe that:

$$\frac{\partial F}{\partial \beta} = -2\beta \int_0^{\infty} dk\, \frac{\cos(kr)}{(k^2+\beta^2)^2} e^{-\alpha k^2}. \tag{20}$$

Furthermore,

$$\frac{\partial}{\partial r}\left[(\frac{-1}{2\beta}) \frac{\partial F}{\partial \beta} \right] = -\int_0^{\infty} dk\, \frac{k\sin(kr)}{(k^2+\beta^2)^2} e^{-\alpha k^2}. \tag{21}$$

Consequently,

$$G = \frac{1}{2\beta} \frac{\partial^2 F}{\partial r \partial \beta}, \tag{22}$$

Substitute the error function form of F into the above equation :

$$\begin{aligned} \frac{\partial F}{\partial \beta} =& \frac{\pi}{4} e^{\beta^2\alpha} \left[\left(2\alpha - \frac{1}{\beta^2} + \frac{r}{\beta} \right) e^{\beta r} \mathrm{erfc}\left(\beta\sqrt{\alpha} + \frac{r}{2\sqrt{\alpha}} \right) \right. \\ & \left. + \left(2\alpha - \frac{1}{\beta^2} - \frac{r}{\beta} \right) e^{-\beta r} \mathrm{erfc}\left(\beta\sqrt{\alpha} - \frac{r}{2\sqrt{\alpha}} \right) - \frac{4}{\beta} \frac{\sqrt{\alpha}}{\sqrt{\pi}} e^{-\beta^2\alpha - r^2/(4\alpha)} \right] \end{aligned} \tag{23}$$

Subsequently,

$$\begin{aligned} \frac{\partial^2 F}{\partial\beta\partial r} = \frac{\pi}{4} e^{\beta^2\alpha} \left[(2\alpha\beta + r) e^{\beta r} \mathrm{erfc}\left(\beta\sqrt{\alpha} + \frac{r}{2\sqrt{\alpha}} \right) \right. \\ \left. - (2\alpha\beta - r) e^{-\beta r} \mathrm{erfc}\left(\beta\sqrt{\alpha} - \frac{r}{2\sqrt{\alpha}} \right) \right]. \end{aligned} \tag{24}$$

Thus, we obtain G by substituting $\beta = 2/a_0$ and $\alpha = \sigma^2/4$ back into the expression:

$$G = \frac{\pi a_0}{16} e^{\frac{\sigma^2}{a_0^2}} \left[\left(\frac{\sigma^2}{a_0} + r \right) e^{\frac{2r}{a_0}} \operatorname{erfc}\left(\frac{\sigma}{a_0} + \frac{r}{\sigma} \right) - \left(\frac{\sigma^2}{a_0} - r \right) e^{-\frac{2r}{a_0}} \operatorname{erfc}\left(\frac{\sigma}{a_0} - \frac{r}{\sigma} \right) \right]. \tag{25}$$

Finally,

$$\rho_{\mathrm{H},1}(r) = \rho_{\mathrm{H},0} - \frac{q}{Z_{\mathrm{H}}} \frac{1}{\pi a_0^3} \frac{1}{2r} e^{\frac{\sigma^2}{a_0^2}} \left[\left(\frac{\sigma^2}{a_0} + r \right) e^{\frac{2r}{a_0}} \operatorname{erfc}\left(\frac{\sigma}{a_0} + \frac{r}{\sigma} \right) - \left(\frac{\sigma^2}{a_0} - r \right) e^{-\frac{2r}{a_0}} \operatorname{erfc}\left(\frac{\sigma}{a_0} - \frac{r}{\sigma} \right) \right]. \tag{26}$$

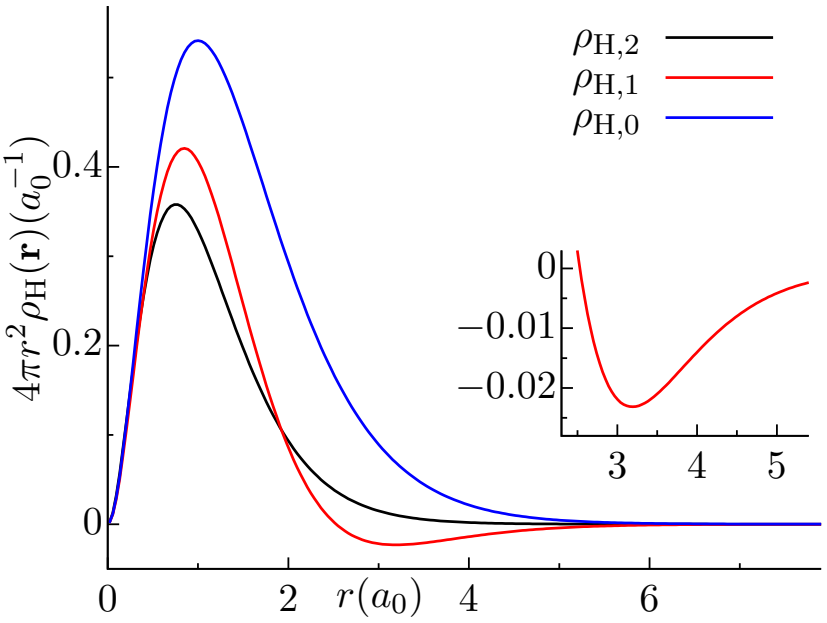

Figure 1: Radial electron density distribution $\rho_{\mathrm{H},1}(r)$ from the MAFF model. The shaded region highlights the unphysical negative density values.

Figure 1 shows the radial electron density distribution $\rho_{\mathrm{H},1}(r)$ computed using the MAFF parameters quoted in the original literature: $q = 0.5$ and $\sigma = \sqrt{2}/2$ Å. As clearly demonstrated, $\rho_{\mathrm{H},1}(r)$ attains negative values for $r \gtrsim 2.3\,a_0$ (shaded region). Such an electron density distribution is unphysical. To overcome this fundamental flaw, a modified electron density model that guarantees positivity is required.

2.2 A New Form Factor Model

To address this flaw, we propose a new form factor model that modifies only the hydrogen electron density distribution:

$$\rho_{\mathrm{H},2}(r) = \frac{1-q}{Z_{\mathrm{H}}} \frac{\alpha^3}{8\pi} e^{-\alpha r}, \tag{27}$$

where q remains the transferred charge, and the remaining charge is distributed as $e^{-\alpha r}$, which guarantees non-negativity. The factor $\alpha^3/(8\pi)$ is a normalization constant. The parameter α regulates the distribution profile. When $\alpha = 5.0$, the distribution is shown as the black line in Fig. 1. The corresponding form factor $f_{\mathrm{H},2}(k)$ is obtained via Fourier transform:

$$f_{\mathrm{H},2}(k) = \frac{1-q}{Z_{\mathrm{H}}} \frac{\alpha^4}{(k^2+\alpha^2)^2} \tag{28}$$

2.3 Parameter Determination and Physical Constraints

Both the MAFF model and our new model introduce human-chosen parameters, which is necessary to correct the deviation of the IAM's physical picture from reality. Among these parameters, the charge transfer q has clear physical constraints and can be obtained from the average molecular dipole moment and geometry:

$$q = \frac{\langle \mu \rangle}{2\sqrt{\langle d_{\mathrm{OH}} \rangle^2 - \langle d_{\mathrm{HH}} \rangle^2/4}}. \tag{29}$$

In contrast, the physical constraints on σ and α are ambiguous. However, in *ab initio* calculations, the spectral data and the structure factors $S_{XX}(k)$ are known. Therefore, these parameters can be determined by minimizing the difference between the model X-ray spectrum and the *ab initio* X-ray spectrum:

$$\mathrm{Error}(\theta) = \int dk \, \left[I_{\mathrm{model}}(k;\theta) - I_{ab\ initio}(k) \right]^2, \quad \theta = \theta \Big|_{\partial \mathrm{Error}/\partial\theta = 0}. \tag{30}$$

In addition to optimizing parameters via scattering spectra, the spatial distribution characteristics of the molecular electron cloud provide an independent means of physical validation. The spatial extent of a molecule's electron cloud is quantified by its total variance, $\langle \rho_e^{tot}(\mathbf{r}) r^2 \rangle$. For models where the electron density is partitioned into spherical atomic contributions, this total variance can be rigorously decomposed into two physically distinct components: the intra-atomic spread variance $\mathcal{R}_{\mathrm{s}}^2$ and the geometric center variance $\mathcal{R}_{\mathrm{c}}^2$.

The geometric variance, $\mathcal{R}_{\mathrm{c}}^2$, originates from the spatial arrangement of the atomic centers. It is computed by treating the entire electron cloud associated with each atom as a point charge located at its nucleus (with weight $Z_\alpha + q_\alpha$) and evaluating the variance of these point charges relative to the molecular center of

negative charge. Consequently, $\mathcal{R}_c^2$ depends solely on the molecular geometry—specifically the bond lengths d_{OH} and d_{HH}—and the amount of charge transfer q:

$$\mathcal{R}_c^2 = \frac{(4-3q-q^2)(4d_{OH}^2 - d_{HH}^2)}{10} + (1-q)\frac{d_{HH}^2}{2}. \tag{31}$$

In contrast, the intra-atomic spread variance, $\mathcal{R}_s^2$, captures the innate diffuseness of the electron cloud *around each atomic nucleus*. It is defined as the weighted sum of the variances of the three spherical atomic densities relative to their own centers, irrespective of atomic positions. This term reflects how chemical bonding modifies the compactness of the atomic orbitals. The total variance of the molecular electron cloud is then the sum of these two independent contributions:

$$\langle \rho_e^{\text{tot}}(\mathbf{r}) r^2 \rangle = \mathcal{R}_s^2 + \mathcal{R}_c^2. \tag{32}$$

The connection between the form factor $f_X(k)$ and the atomic variance $\mathcal{R}_X^2$ follows from the properties of the Fourier transform. For a spherically symmetric electron density $\rho_e(r)$:

$$\frac{\partial^2 f_X(\mathbf{k})}{\partial k^2} = -\int_{-\infty}^{\infty} d\mathbf{r}\, r^2 \rho_e(r) e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{33}$$

At $k = 0$, this gives two fundamental relations:

$$-\frac{\partial^2 f_X(\mathbf{k})}{\partial k^2}\bigg|_{k=0} = \mathcal{R}_X^2, \quad f_X(\mathbf{k})\bigg|_{k=0} = Z_X, \tag{34}$$

where $\mathcal{R}_X^2$ is the variance of atom X's electron distribution, and Z_X is its atomic number.

In our new model ($f_{H,2} + f_{O,1}$), the intra-atomic variance takes the form

$$\mathcal{R}_s^2 = (1 + \frac{2q}{Z_O})R_{O,0}^2 + q\sigma^2 + 2\frac{1-q}{Z_H}\frac{4}{\alpha^2}, \tag{35}$$

where the parameter α explicitly controls the contraction of the hydrogen electron cloud. The MAFF model, however, yields

$$\mathcal{R}_s^2 = (1 + \frac{2q}{Z_O})\mathcal{R}_{O,0}^2 + 2(1 - \frac{q}{Z_H})\mathcal{R}_{H,0}^2, \tag{36}$$

which lacks an adjustable parameter to describe orbital contraction, thereby directly coupling the spread variance to the charge transfer q.

2.3.1 Water in Gas-Phase

First, we consider an isolated water molecule, which represents the gas-phase limit. The benchmark X-ray scattering signal $I_{ab\ initio}(k)$ is computed via direct Fourier transform of the molecular electron density obtained from quantum chemistry calculations, as defined in eq(1). Our computation is performed at

the PBE0/def2-TZVPD level using the BDF software [1] for a monomer with experimental bond lengths $d_{\mathrm{OH}} = 0.9753\,\text{Å}$ and $d_{\mathrm{HH}} = 1.559\,\text{Å}$ [2, 3, 4]. The calculated dipole moment of 1.8486 Debye closely matches the experimental value of 1.8546 Debye [5].

Alternatively, assuming that the electron density of the rigid molecule is a sum of spherical atomic densities, the scattering intensity for an isolated molecule can be computed via the Debye formula [**?**, 4]:

$$I(k) = f_{\mathrm{O}}^2(k) + 2f_{\mathrm{H}}^2(k) + 4f_{\mathrm{O}}(k)f_{\mathrm{H}}(k)\frac{\sin(kd_{\mathrm{OH}})}{kd_{\mathrm{OH}}} + 2f_{\mathrm{H}}^2(k)\frac{\sin(kd_{\mathrm{HH}})}{kd_{\mathrm{HH}}}, \quad (37)$$

where $f_{\mathrm{O}}(k)$ and $f_{\mathrm{H}}(k)$ are the k-dependent form factors defined in any of the models discussed above.

Following the gradient-based optimization procedure described in Section 2.3, we obtain the optimal parameters for each model by minimizing the difference between the model prediction and the benchmark $I_{ab\ initio}(k)$.

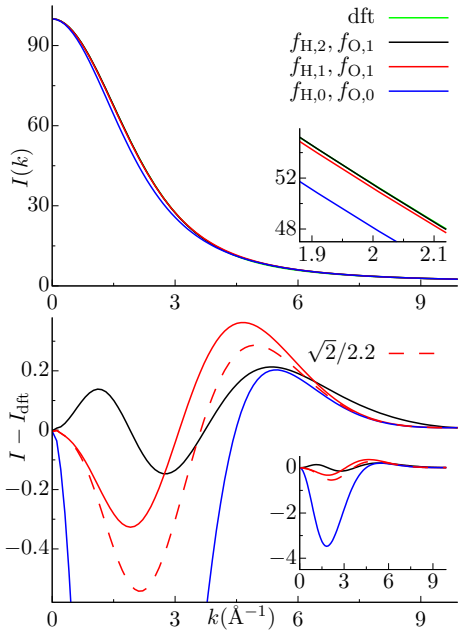

Figure 2: Comparison of X-ray scattering intensities for gas-phase water: benchmark DFT result (black solid line), IAM (gray dashed), MAFF with suggested σ (orange dashed), MAFF with optimized σ (green dashed), and the new model with optimized parameters (blue solid).

Figure 2 compares the scattering intensities computed from different models with their respective optimized parameters. Notable differences around $k \approx 2\,\text{Å}^{-1}$ are observed between the benchmark $I_{ab\ initio}(k)$ and the result from the parameter-free Independent Atom Model ($f_{\mathrm{H},0}$ and $f_{\mathrm{O},0}$). Significant

improvement is achieved when using modified form factors. With $q = 0.32737$ (consistent with the dipole moment) and $\sigma = \sqrt{2}/2.2\,\text{Å}$ [6], the discrepancy is reduced to less than 0.5. Further reduction is obtained by least-squares optimization, yielding $\sigma = 0.5870\,\text{Å}$ for the MAFF model.

As expected, the new model ($f_{\text{H},2} + f_{\text{O},1}$) with optimized parameters $\sigma = 0.6346\,\text{Å}$ and $\alpha = 5.1746\,\text{Å}^{-1}$ achieves the best agreement with $I_{ab\ initio}(k)$, demonstrating superior spectral fitting capability.

Table 2: Optimized parameters and physical validation for gas-phase water models. The charge transfer $q = 0.32737$ is fixed by the dipole moment for all modified models.

Exp.	$d_{\text{OH}} = 0.9763$Å;	$d_{\text{HH}} = 1.559$Å;	1.8546 (Debye)	
Model	Parameters	Dipole (Debye)	$\mathcal{R}_{\text{s}}^2$ (Å^2)	$\mathcal{R}_{\text{c}}^2$ (Å^2)
DFT	PBE0 def2-TZVPD	1.8486	1.4642	0.4066
$f_{\text{H},2}$ $f_{\text{O},1}$	$q = 0.32737$ $\sigma = 0.6346$ $\alpha = 5.1746$	1.8486	1.4606	0.4066
$f_{\text{H},1}$ $f_{\text{O},1}$	$q = 0.32737$ $\sigma = 0.5870$	1.8486	1.5045	0.4066
$f_{\text{H},1}$ $f_{\text{O},1}$	$q = 0.32737$ $\sigma = \sqrt{2}/2.2$	1.8486	1.5045	0.4066
IAM	$q = 0$	0	1.6025	0.5894

Table 2 lists the optimized parameters and their corresponding physical validations. While all modified models reproduce the correct dipole moment (by construction), the new model shows significant advantages in physical consistency. The intra-atomic spread variance $\mathcal{R}_{\text{s}}^2$, which measures the compactness of atomic orbitals after chemical bonding, is reproduced with only 0.025% error by the new model, an order of magnitude more accurate than the MAFF models. This indicates that the parameter α in the new model not only achieves better spectral fitting (Fig. 2) but also provides a more physically reasonable description of orbital contraction upon bond formation.

The geometric variance $\mathcal{R}_{\text{c}}^2$, which depends only on molecular geometry and charge transfer q, is correctly reproduced by all models as expected. The superior performance of the new model in both spectral fitting and physical validation demonstrates that it offers a more precise and physically grounded distribution form and parameters for describing electron density distribution in water molecules.

2.3.2 Water in Bulk-Phase

Next, we examine liquid water under ambient conditions. We perform *ab initio* molecular dynamics simulations using the revPBE0/TZV2P-GTH level for an NVE ensemble of 64 water molecules at the experimental mass density of $0.997\,\mathrm{g\,cm^{-3}}$ and $T = 298.15\,\mathrm{K}$. A total of 150 independent trajectories are generated from equilibrated SPC/E configurations, each run for 2 ps after 12 ps of equilibration, with a time step of 0.5 fs and configurations sampled every 25 fs.

Following the same gradient-based optimization procedure as for the gas phase, we obtain the optimal parameters for bulk water by minimizing the difference between the model prediction and the benchmark $I_{ab\ initio}(k)$.

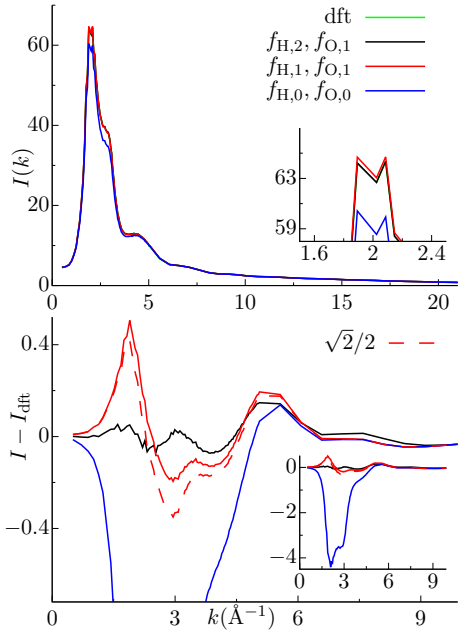

Figure 3: Comparison of X-ray scattering intensities for bulk water: benchmark DFT result (black solid line), IAM (gray dashed), MAFF with $\sigma = \sqrt{2}/2$ (orange dashed), MAFF with optimized σ (green dashed), and the new model with optimized parameters (blue solid).

Figure 3 compares the scattering intensities computed from different models with their respective optimized parameters for bulk water. From the average molecular dipole moment $\langle\mu\rangle = 2.9244\,\mathrm{Debye}$ and geometry $\langle d_{OH}\rangle = 0.9859\,\mathrm{Å}$, $\langle d_{HH}\rangle = 1.563\,\mathrm{Å}$, we obtain $q = 0.506702$.

As shown in Figure 3, the new model ($f_{H,2}+f_{O,1}$) with optimized parameters $\sigma = 0.6720\,\mathrm{Å}$ and $\alpha = 5.6307\,\mathrm{Å}^{-1}$ achieves the best agreement with $I_{ab\ initio}(k)$ for bulk water, demonstrating consistent superiority across both gas and con-

densed phases.

Table 3: Optimized parameters for bulk water models compared with the original MAFF model ($\sigma = \sqrt{2}/2$) suggested by Skinner *et al.*

Exp.	$\langle d_{\mathrm{OH}} \rangle = 0.9859$Å;	$\langle d_{\mathrm{HH}} \rangle = 1.563$Å;	$\langle \mu \rangle = 2.9244$ (Debye)	
Model	Parameters	Dipole (Debye)	$\mathcal{R}_{\mathrm{s}}^2$ (Å^2)	$\mathcal{R}_{\mathrm{c}}^2$ (Å^2)
DFT	revPBE0 TZV2P-GTH	2.9244	–	0.3080
$f_{\mathrm{H},2}$ $f_{\mathrm{O},1}$	$q = 0.506702$ $\sigma = 0.6720$ $\alpha = 5.6307$	2.9244	1.5278	0.3080
$f_{\mathrm{H},1}$ $f_{\mathrm{O},1}$	$q = 0.506702$ $\sigma = 0.6912$	2.9244	1.4508	0.3080
$f_{\mathrm{H},1}$ $f_{\mathrm{O},1}$	$q = 0.506702$ $\sigma = \sqrt{2}/2$	2.9244	1.4508	0.3080
IAM	$q = 0$	0	1.6025	0.5999

Table 3 lists the optimized parameters and their corresponding variances. While all modified models reproduce the correct average dipole moment (by construction from q), the new model shows distinct values for the intra-atomic spread variance $\mathcal{R}_{\mathrm{s}}^2$. The geometric variance $\mathcal{R}_{\mathrm{c}}^2$, which depends only on molecular geometry and charge transfer q, is correctly reproduced by all models as expected from Eq. (31).

However, in bulk water, electrons cannot be unambiguously attributed to individual molecules, which prevents direct calculation of $\mathcal{R}_{\mathrm{s}}^2$ from the total electron density. An independent validation can be obtained from the system's surface potential. We have previously established a quantitative relationship linking the surface potential contribution from charge delocalization, χ_{deloc}, to the intramolecular charge distribution:

$$\chi_{\mathrm{deloc}} = \frac{|e|}{2\varepsilon_0} \rho_M (\mathcal{R}_-^2 - \mathcal{R}_+^2), \tag{38}$$

where $\mathcal{R}_-^2 = \mathcal{R}_{\mathrm{s}}^2 + \mathcal{R}_{\mathrm{c}}^2$ represents the total delocalization extent of negative charges within the molecule, $\mathcal{R}_+^2$ is that of positive charges, and ρ_M is the molecular number density in the bulk phase.

Using the same computational level as for the bulk simulations, we directly computed the surface potential difference of water to be +4.18 V, with the dipole moment contribution being +0.295 V. The remainder arises from charge delocalization, i.e., the second moment contribution. We employ this result to assess the physical reliability of the optimized parameters in Table 3.

Table 4: Surface potential contributions χ_{deloc} computed via Eq. (38) for models and parameters in Table 3.

Model	DFT	$f_{\mathrm{H},2}$ $f_{\mathrm{O},1}$	$f_{\mathrm{H},1}$ $f_{\mathrm{O},1}$	IAM
χ_{deloc} (V)	3.885	3.727	3.495	4.832

The χ_{deloc} values in Table 4 provide an independent physical validation. Among the form factor models, the new model ($f_{\mathrm{H},2} + f_{\mathrm{O},1}$) yields $\chi_{\mathrm{deloc}} = 3.727\,\mathrm{V}$, closest to the DFT benchmark of $3.885\,\mathrm{V}$. This indicates that the optimized parameters ($q = 0.506702$, $\sigma = 0.6720\,\text{Å}$, $\alpha = 5.6307\,\text{Å}^{-1}$) not only achieve superior spectral fitting (Fig. 3) but also provide a physically accurate description of electron delocalization, as validated by the surface potential.

3 Structure Resolution from Experimental Scattering Spectra

References

[1] Yong Zhang, Bingbing Suo, Zikuan Wang, Ning Zhang, Zhendong Li, Yibo Lei, Wenli Zou, Jun Gao, Daoling Peng, Zhichen Pu, Yunlong Xiao, Qiming Sun, Fan Wang, Yongtao Ma, Xiaopeng Wang, Yang Guo, and Wenjian Liu. BDF: A relativistic electronic structure program package. *The Journal of Chemical Physics*, 152(6):064113, 02 2020.

[2] W. S. Benedict, N. Gailar, and Earle K. Plyler. Rotation-Vibration Spectra of Deuterated Water Vapor. *The Journal of Chemical Physics*, 24(6):1139–1165, 1956.

[3] Shuzo Shibata and L. S. Bartell. Electron-Diffraction Study of Water and Heavy Water. *The Journal of Chemical Physics*, 42(4):1147–1151, 1965.

[4] A. H. Narten and H. A. Levy. Liquid Water: Molecular Correlation Functions from X-Ray Diffraction. *The Journal of Chemical Physics*, 55(5):2263–2269, 1971.

[5] Shepard A. Clough, Yardley Beers, Gerald P. Klein, and Laurence S. Rothman. Dipole moment of water from Stark measurements of H2O, HDO, and D2O. *The Journal of Chemical Physics*, 59(5):2254–2259, 1973.

[6] Jon M. Sorenson, Greg Hura, Robert M. Glaeser, and Teresa Head-Gordon. What can x-ray scattering tell us about the radial distribution functions of water? *The Journal of Chemical Physics*, 113(20):9149–9161, 2000.