



JOURNAL OF  
APPLIED  
CRYSTALLOGRAPHY

**Volume 50 (2017)**

**Supporting information for article:**

**A novel fast Fourier transform accelerated off-grid exhaustive search method for cryo-electron microscopy fitting**

**Alexandre Hoffmann, Valérie Perrier, and Sergei Grudinin**

# Supplementary Information for “A novel FFT-accelerated off-grid exhaustive search method for cryo-EM fitting”

Alexandre Hoffmann<sup>a,b,c</sup>, Valérie Perrier<sup>a,b,c</sup>, and Sergei Grudinin<sup>\*a,b,c</sup>

<sup>a</sup>University of Grenoble Alpes, LJK, F-38000 Grenoble, France

<sup>b</sup>CNRS, LJK, F-38000 Grenoble, France

<sup>c</sup>Inria, F-38000 Grenoble, France

## 1 Stability of the method

Here we study the stability of the method. In order to do so, we will put ourselves in a very simplistic framework. More precisely, we will consider a one-dimensional problem, with a molecule consisting of a single atom placed at the origin. Without loss of generality, we will assume that our experimental density is band-limited with the highest frequency written as a cosine function with an arbitrary phase shift  $\phi$ ,

$$d_a(x) = \cos(2\pi\omega(x - \phi)) \quad \omega \in (0, 1). \quad (1)$$

In this case, one can show that the CCF can be written as follows,

$$CCF_{\cos}(\tau, y) = CCF_{\cos}(\tau - y) = \cos(2\pi\omega(\tau + \phi - y)) \exp(-\sigma^2 2\pi^2 \omega^2) \quad \omega \in (0, 1), \quad (2)$$

where  $y$  is the off-grid displacement. Suppose now that we apply the quadratic approximation method starting from a given value of  $\tau$ . By using (??), one can see that the quadratic approximation of the CCF can be written as follows,

$$C\tilde{C}F_{\cos}(\tau, y) = CCF_{\cos}(\tau) + 2\pi\omega CCF_{\sin}(\tau)y - 2\pi^2\omega^2 CCF_{\cos}(\tau)y^2. \quad (3)$$

The derivation of  $CCF_{\cos}$  and  $CCF_{\sin}$  can be found in Appendix.

Now, we will find an interval in which our method can effectively reduce the fitting error of an in-grid exhaustive search. Let us first introduce the in-grid

---

\*Sergei.Grudinin@inria.fr

pose  $\tau^{IG}$  and the best off-grid pose  $\tau^{OF}$ , which can be written as follows,

$$\tau_j^{OF} = \frac{j}{\omega} - \phi \quad \forall j \in \mathbb{Z}, \quad \omega \in (0, 1). \quad (4)$$

Our method seeks for the off-grid displacement  $y$  that reduces the distance between  $\tau^{IG} - y$  and  $\tau^{OF}$ . As we mentioned in the main text, we observed that it is preferable to have a concave cost function  $CCF_{\cos}(\tau, \cdot)$ . By using (??) and (??), one can show that this is the case *if and only if*  $CCF_{\cos}(\tau, 0) > 0$ , meaning that

$$\tau^{IG} \in \left( \frac{4j-1}{4\omega} - \phi, \frac{4j+1}{4\omega} - \phi \right) \quad \forall j \in \mathbb{Z}, \quad \omega \in (0, 1). \quad (5)$$

Let now express the fitting error reduction of our method. In order to do so, we will use a classical proof of the Newton's method convergence rate. The idea is to use a first order Taylor's series of the first derivative of (??) around  $\tau^{OF} = \tau^{IG} - y^*$ ,

$$\partial_y CCF_{\cos}(\tau^{IG} - y^*) = \partial_y CCF_{\cos}(\tau^{IG}) - y^* \partial_y^2 CCF_{\cos}(\tau^{IG}) + \frac{(y^*)^2}{2} \partial_y^3 CCF_{\cos}(\tau^{IG} - \xi), \quad \xi \in (0, y_j^*). \quad (6)$$

Since  $\tau^{OF}$  is the optimal pose, the derivative of the CCF is null at  $\tau^{OF}$ . Then we divide (??) by the second derivative of the CCF to obtain the following equality,

$$y^* - \frac{\partial_y CCF_{\cos}(\tau^{IG})}{\partial_y^2 CCF_{\cos}(\tau^{IG})} = \frac{(y^*)^2}{2} \frac{\partial_y^3 CCF_{\cos}(\tau^{IG} - \xi)}{\partial_y^2 CCF_{\cos}(\tau^{IG})}, \quad \xi \in (0, y_j^*). \quad (7)$$

Let us recall that  $\frac{\partial_y CCF_{\cos}(\tau^{IG})}{\partial_y^2 CCF_{\cos}(\tau^{IG})}$  is the solution given by our off-grid search method, thus the fitting error produced by our method can be written as follows,

$$E(\tau^{IG}) = |y^* - y| = \left| \frac{\partial_y^3 CCF_{\cos}(\tau^{IG} - \xi)}{2\partial_y^2 CCF_{\cos}(\tau^{IG})} \right| (y^*)^2, \quad \xi \in (0, y_j^*). \quad (8)$$

The second and the third derivatives can be computed by integration by parts (see Appendix for more detail), and then we find the following equality,

$$E(\tau^{IG}) = |y^* - y| = \left| \frac{2\pi\omega \sin(2\pi\omega(\tau^{IG} + \phi - \xi))}{\cos(2\pi\omega(\tau^{IG} + \phi))} \right| (y^*)^2, \quad \xi \in (0, y_j^*). \quad (9)$$

We now aim at bounding  $E(\tau^{IG})$ . First, note that since  $\tau^{OF} = \tau^{IG} - y^*$  and using (??), we can write the following inequalities,

$$\frac{4j-1}{4\omega} < \frac{j}{\omega} = \tau^{IG} + \phi - y < \tau^{IG} + \phi - \xi < \tau^{IG} + \phi \leq \frac{4j+1}{4\omega} \quad (10)$$

We know that on such an interval,  $\sin(2\pi\omega \cdot)$  is a strictly increasing function, therefore, we have the following identity,

$$\sin(2\pi\omega(\tau^{IG} + \phi - \xi)) < \sin(2\pi\omega(\tau^{IG} + \phi)). \quad (11)$$

Finally, our fitting error can be bounded as follows,

$$E(\tau^{IG}) = |y^* - y| < \left| \frac{2\pi\omega \sin(2\pi\omega(\tau^{IG} + \phi))}{\cos(2\pi\omega(\tau^{IG} + \phi))} \right| (y^*)^2 < 2\pi |\tan(2\pi\omega(\tau^{IG} + \phi))| \omega h^2, \quad (12)$$

where  $h$  is the grid spacing. This is a very conservative estimation, provided that we can bound the tangent function by shrinking the allowed interval of  $\tau^{IG} + \phi$ . Nonetheless, this estimation clearly demonstrates that the error of the method is linearly proportional to the highest frequency of the map (or inversely proportional to its resolution), and quadratically proportional to the grid spacing. Following this result, we can provide the following asymptotic approximation of the fitting error,

$$E(\tau^{IG}) \approx O(\omega h^2). \quad (13)$$

## Appendix. Analytical formulae for the CCF of a sinusoidal function

Here we focus on computing the following integrals,

$$\begin{aligned} CCF_{\cos}(\tau, 0) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} \cos(2\pi\omega(x - \phi)) \exp\left(-\frac{(x + \tau)^2}{2\sigma^2}\right) dx \\ CCF_{\sin}(\tau, 0) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} \sin(2\pi\omega(x - \phi)) \exp\left(-\frac{(x + \tau)^2}{2\sigma^2}\right) dx \end{aligned}, \quad (14)$$

which are the real and the imaginary parts of the following integral,

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} \exp(2i\pi\omega(x - \phi)) \exp\left(-\frac{(x + \tau)^2}{2\sigma^2}\right) dx. \quad (15)$$

This integral is the Fourier transform of a Gaussian evaluated at  $-\omega$ , or equivalently at  $\omega$ , since the Fourier transform of a real valued function is Hermitian. Thus by using the translation property of the Fourier Transform and by recalling that the Fourier transform of a Gaussian is a scaled Gaussian, we get the following identity,

$$\begin{aligned} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} \exp(2i\pi\omega(x - \phi)) \exp\left(-\frac{(x + \tau)^2}{2\sigma^2}\right) dx = \\ \exp(-2i\pi\omega(\tau + \phi)) \exp(-\sigma^2 2\pi^2 \omega^2). \end{aligned} \quad (16)$$

Finally, we get the two following identities,

$$\begin{aligned} CCF_{\cos}(\tau, 0) &= \cos(2\pi\omega(\tau + \phi)) \exp(-\sigma^2 2\pi^2 \omega^2) \\ CCF_{\sin}(\tau, 0) &= -\sin(2\pi\omega(\tau + \phi)) \exp(-\sigma^2 2\pi^2 \omega^2). \end{aligned} \quad (17)$$

Figure S1: RMSD as a function of the angular sampling distance for fitting a subunit of the GroEL complex to simulated EDMs of different resolutions. (a) : Resolution 5 Å. (b): Resolution 11 Å. (c) : Resolution 15 Å. (d): Resolution 21 Å. Red : Fitting with off-grid optimization. Blue: Fitting without off-grid optimization.

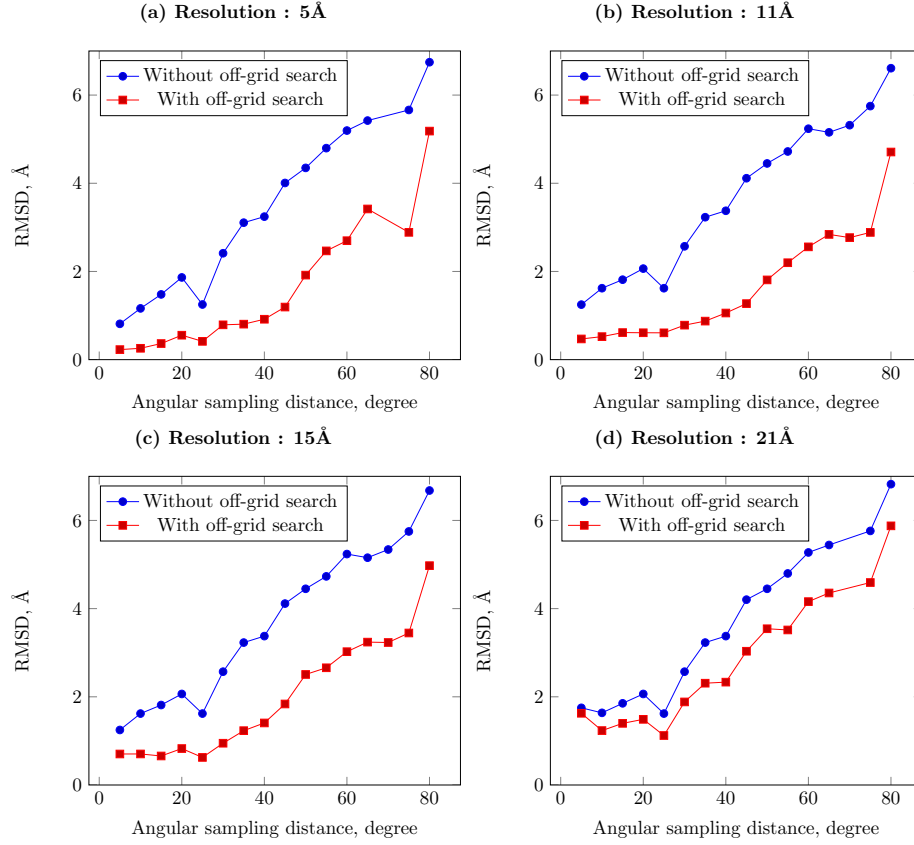


Figure S2: RMSD as a function of the angular sampling distance for fitting a subunit of the GroEL complex at different model resolutions to a simulated EDM of 11 Å resolution. (a) : Model resolution 1 Å. (b): Model resolution 2 Å. (c) : Model resolution 3 Å. (d): Model resolution 4 Å. Red : Fitting with off-grid optimization. Blue: Fitting without off-grid optimization.

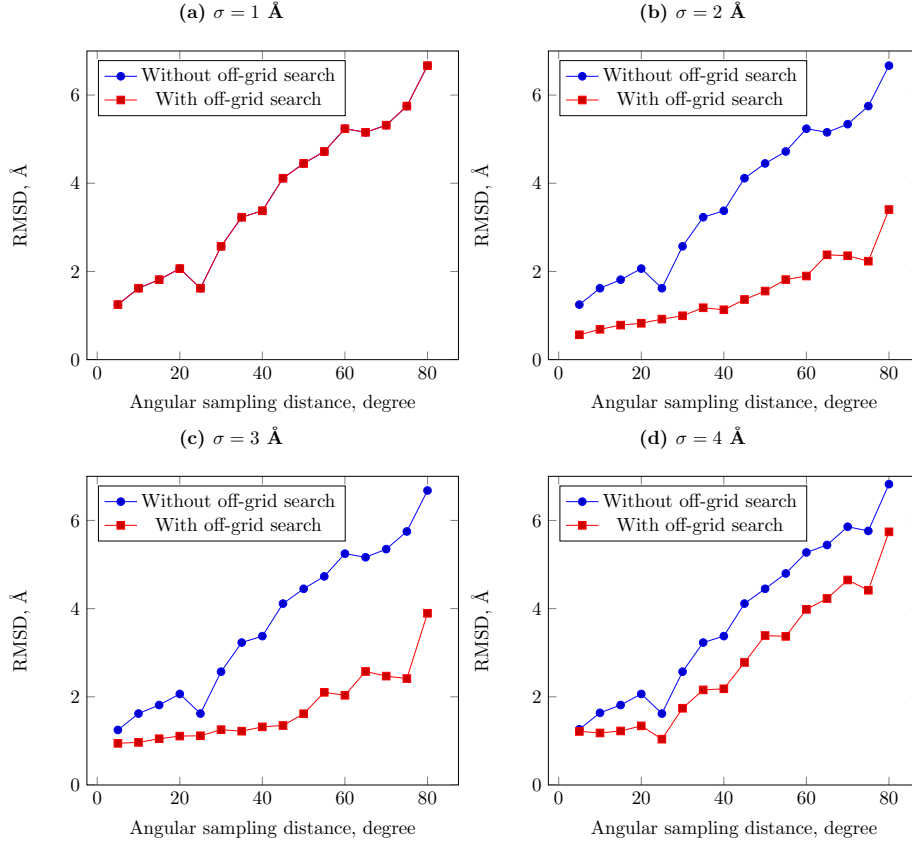


Figure S3: RMSD as a function of the total number of sampled translations for different model resolutions. (a) : Model resolution 1 Å. (b): Model resolution 2 Å. (c) : Model resolution 3 Å. (d): Model resolution 4 Å. Blue : Fitting without off-grid optimization. Red: Fitting with off-grid optimization.

