# *AB INITIO* MAXIMUM LIKELIHOOD RECONSTRUCTION OF HELICAL MACROMOLECULES USING ELECTRON MICROSCOPY

*Min Woo Kim, and Jong Chul Ye*

Bio-Imaging & Signal Processing Lab., Dept. Bio& Brain Engineering
Korea Advanced Institute of Science & Technology (KAIST)
373-1 Guseong-dong Yuseong-gu, Daejon 305-701, Korea
Email: jong.ye@kaist.ac.kr

## ABSTRACT

Recent advance in cryo-electron microscopy allows structural determination of helical macromolecules without forming crystals. A conventional reconstruction method for helical structure is the Fourier-Bessel transform approach. However, this approach often fails in some instances such as flexible and disordered specimens. In order to overcome the drawbacks, a single particle reconstruction approach has been recently proposed. However, it often requires a good initial guess to guarantee the convergence. This paper describes an accurate *model free* maximum likelihood helical reconstruction algorithm by exploiting the symmetry and sparsity, which overcomes the drawbacks of the conventional methods.

***Index Terms—*** helical macromolecule, electron microscopy, symmetry, sparsity, diffraction pattern

## 1. INTRODUCTION

Many macromolecules such as microtubule, actin and filamentous viruses are naturally assembled into helical polymers. Recent advance in cryo-electron microscopy allows structural determination of helical macromolecules without x-ray crystallography. The classical reconstruction method for helical structures in electron microscopy is the Fourier-Bessel transform approach that exploits the diffraction pattern. However, the Fourier-Bessel transform approach often fails in some instances such as flexible and disordered specimens. For these type of specimens, a single particle reconstruction technique called iterative helical real space reconstruction (IHRSR) has been investigated as an alternative [1]. The main drawback of IHRSR is, however, that it does not often guarantee the correct reconstruction unless a good initial estimate of unknown particle structure is available.

This work was supported by the Basic Research Program of Korea Research Foundation Grant funded by the Korean Government (MOEHRD) KRF-2007-313-D00593, and also supported by the Korea Science and Engineering Foundation under Grant R01-2005-000-10035-0. The authors would like to thank Prof. Esther Bullitt at Boston university school of medicine for providing the P-pili micrograph sample.

This paper proposes a new *ab initio* maximum likelihood helical reconstruction algorithm that encompasses the conventional Fourier-Bessel approach and the single particle reconstruction methods. The main idea is to exploit the symmetry and the sparsity of the helical structure in real and reciprocal spaces, respectively. Helical symmetry provides specific number of symmetric views determined by the number of subunits per repeat. Even though the number of available view from symmetry is not sufficient, the angular aliasing artifacts due to the limited views can be significantly reduced by exploiting the sparsity of the helical structure in real space using the compressed sensing [2]. Hence, as long as the helix parameters are correctly estimated, a reasonable 3-D structure can be obtained even from a single projection. If additional projections for other segments are available, accurate alignment between segments may provide the sufficient number of views for the best resolution 3-D reconstruction.

Therefore, the most important step in the proposed algorithm is the estimation of the correct helix parameters. This paper presents the maximum-likelihood approach, in which the helical volume as well as the symmetry parameters are estimated by maximizing the likelihood. We can show the convergence of our estimation algorithm using alternative minimization technique.

## 2. CONVENTIONAL HELICAL RECONSTRUCTION METHODS

The molecules making up the helix have periodic electron density in the long direction. The symmetry of a helical structure can be determined by subunit axial translation $z_s$ and azimuthal rotation angle $\phi_s$. Under the lattice model, the helical symmetric density model is defined by $\rho(r, \phi, z) = \rho(r, \phi + m\phi_s, z + nz_s)$, where $m, n$ denotes any integer.

The Fourier-Bessel transform approach has been widely used for reconstruction of helical structures. This approach exploit the diffraction pattern in reciprocal space. However, it is often difficult to apply the Fourier-Bessel algorithm to some specimen such as flexible and disordered specimens.

A real space approach based on single particle reconstruction techniques may allow many advantages over helical approaches in these aspects. One popular single particle reconstruction method for helical structures is the iterative helical real space reconstruction (IHRSR) [1]. To deal with the flexibility of filament, the IHRSR first cuts the projection image into short segments. Then, the reference projections are generated by rotating the 3-D model reference about the filament axis. These reference projections are used with the raw images to determine several parameters that are applied to each short segment images for alignment; then these aligned images are used with the known azimuthal orientation to generate a 3D reconstruction. In order to impose the helical symmetry, a least square search is implemented to first find the $(z_s, \phi_s)$ using the reconstructed volume, and then these parameters are then imposed to generate a new helically symmetric reference volume for next iteration. The process is iterated until convergence. In IHRSR, the initial estimate of the helical structure is important since it is used to generate the initial reference image. However, no formal proof has been made and some authors have observed that the final reconstruction is dependent on the initialization [3].

## 3. *AB INITIO* MAXIMUM LIKELIHOOD APPROACH

### 3.1. Diffraction Pattern Analysis

Beside $(\phi_s, z_s)$ screw displacement, the helix can also have pure rotational symmetry. For example, $N$-fold rotational symmetry about its axis ($C_N$ point group), and sometimes dihedral symmetry ($D_N$ point group). Therefore, the complete description of symmetry in helical macromolecules should account for the screw displacement plus point-group symmetry. For the case of complete helix, each layer line has nonzero intensity on each side of meridian, and the distance between the layer line should be given by $\Delta Z = N/P_s$, where $P_s$ denotes the pitch. The meridian intensities occurs at every $1/z_s$ intervals, where $z_s$ is the subunit axial translation. Hence, the remaining symmetry parameter to be estimated is the point group symmetry $N$, and the twist angle $\phi_s$:

$$\phi_s = \pm\frac{2\pi z_s \Delta Z}{N}, \quad N \in \mathbb{N}^+, \tag{1}$$

owing to the relationship $\phi_s = 2\pi z_s/P_s$ and $\Delta Z = N/P_s$. Here, the positive $\phi_s$ implies the righthanded helix.

If the length of the helical segment are shorter than $2P_s/N$, we cannot expect the first layer lines to appear at $N/P_s$ since the corresponding frequency resolution of the short segment should be larger than $N/P_s$. Even in this case, the subunit axial translation $z_s$ can be often calculated as the reciprocal of the distance between the origin and the nearest meridian reflection. Hence, the twist angle $\phi_s$ can be estimated as

$$\phi_s = \pm\frac{2\pi t}{u}, \quad t \in \mathbb{N}^+, u \in \mathbb{N}^+ \tag{2}$$

owing to the relationship $z_s/P_s = t/u$, where $t$ and $u$ denote the number of helical twist per repeat and the number of subunits per repeat, respectively. Hence, the unknown helix parameters to be estimated are two integer numbers $(t, u)$ and the rotational group symmetry $N$.

Finally, in some measurement situation, only a distance between adjacent layer line, $\Delta Z = N/P_s$, can be measured. In this case, we need to estimate both screw symmetry parameters $(\phi_s, z_s)$ as well as the number of point group symmetry $N$. Interestingly, even in this scenario, all the unknown parameter can be estimated using three integer-valued parameters thanks to the following relationship:

$$\phi_s = \pm\frac{2\pi t}{u}, \quad z_s = \frac{1}{\Delta Z}\frac{tN}{u}, \quad t \in \mathbb{N}^+, u \in \mathbb{N}^+ , \tag{3}$$

since $z_s = P_s t/u$ and $P_s = N/\Delta Z$.

### 3.2. Helix Parameter Estimation and Reconstruction

Then, how can we estimate the integer number $(N, t, u)$ ? To estimate these parameters accurately, we employ the parameter searching criteria based on maximum-likelihood approaches. Our model can be represented in discrete form by

$$\mathbf{Y} = \mathbf{A}\mathbf{R}_{(N,t,u)} + \mathbf{V} \tag{4}$$

where $\mathbf{Y} = [\mathbf{y}_0, \cdots, \mathbf{y}_{K-1}]$,$\mathbf{R} = \left[\boldsymbol{\rho}_0, \cdots, \boldsymbol{\rho}_{K-1}\right]$,$\mathbf{V} = [\mathbf{v}_0, \cdots, \mathbf{v}_{K-1}]$. $\mathbf{A}$ is projection operator and $\mathbf{v}_k$ is zero mean Gaussian distribution, respectively. Here, $\boldsymbol{\rho}_k$ and $\mathbf{y}_k$ denote a helical structure at the $k$-th rotation/translation and its corresponding projection image, respectively. Each $\boldsymbol{\rho}_k$ can be represented as an affine transform from a reference helix $\boldsymbol{\rho}_0$:

$$\boldsymbol{\rho}_k = \mathbf{P}_{(\Delta\phi_k, \Delta z_k)}\boldsymbol{\rho}_0, \tag{5}$$

where $\mathbf{P}_{(\Delta\phi_k, \Delta z_k)}$ describes an affine transformation with rotation of $\Delta\phi_k$ and axial shifting $\Delta z_k$. We now define the unknown parameter vectors $\boldsymbol{\Xi}$ and $\boldsymbol{\Theta}$:

$$\begin{aligned} \boldsymbol{\Xi} &= \left[\ \boldsymbol{\rho}_0 \quad \Delta\phi_1 \quad \cdots \quad \Delta\phi_{N-1} \quad \Delta z_1 \quad \cdots \quad \Delta z_{N-1}\ \right] \\ \boldsymbol{\Theta} &= \left[\ \boldsymbol{\Xi} \quad N \quad t \quad u\ \right] \end{aligned} \tag{6}$$

Assuming that noise for each projection measurement is independent, the log-likelihood function is then given by

$$\begin{aligned} L(\boldsymbol{\Theta}) &= \log \prod_{k=0}^{K-1} P\left(\mathbf{y}_k | S, t, u, \Delta\phi_k, \Delta z_k, \boldsymbol{\rho}_0\right) \\ &= -\frac{1}{2\sigma^2}\sum_{k=0}^{K-1} ||\mathbf{y}_k - \mathbf{A}\mathbf{P}_{(\Delta\phi_k, \Delta z_k)}\boldsymbol{\rho}_0||^2 + \alpha. \end{aligned} \tag{7}$$

Therefore, the optimal solution to maximum-likelihood approach corresponds to the parameter set $\hat{\boldsymbol{\Theta}}$ that maximize the

log-likelihood Eq. (7):

$$\hat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}} \log L(\boldsymbol{\Theta}) \quad (8)$$

$$= \arg\min_{\boldsymbol{\Theta}} \sum_{k=0}^{K-1} ||\mathbf{y}_k - \mathbf{A}\mathbf{P}_{(\Delta\phi_k, \Delta z_k)}\boldsymbol{\rho}_0||^2 \quad (9)$$

The cost function in Eq. (9) can be further elaborated by noting that helical symmetry and point group symmetry provides symmetric projection views. Therefore, we have the following cost function $C(\boldsymbol{\Theta})$:

$$\frac{1}{uN}\sum_{n=0}^{N-1}\sum_{m=0}^{u-1}\sum_{k=0}^{K-1} ||\mathbf{y}_k - \mathbf{A}\mathbf{P}_{(m\phi_s+\Delta\phi_k+\frac{2\pi}{N}n, mz_s+\Delta z_k)}\boldsymbol{\rho}_0||^2 \quad (10)$$

where the helical parameters $(\phi_s, z_s)$ can be calculated for a each $(N, t, u)$. The direct minimization of $C(\boldsymbol{\Theta})$ is quite complicated due to the coupling between the parameters. Rather, we find that the decoupling of the cost function allows computational efficient minimization. The minimum cost $C(\hat{\boldsymbol{\Xi}}, N, t, u)$ should be calculated first for all feasible helix parameter $(N, t, u)$, and we chose the optimal helix parameter by comparing the individual cost function.

**Step1**. Initial Guess of Reference : For a given integer parameter $(N, t, u)$, the unknown parameter to be estimated is the reference helix $\boldsymbol{\rho}_0$ and alignment parameter $(\Delta\phi_k, \Delta z_k)_{k=1}^{K-1}$. However, the unknown parameters are coupled to each other except at $k = 0$. For this type of problem, alternating minimization provides an computationally feasible problem. For alternative minimization, a good initial guess is required to accelerate the convergence, and we find that the initial guess of $\boldsymbol{\rho}_0$ can be readily obtained from a reasonable criterion. Specifically, consider the following maximum-likelihood of one projection measurement $\mathbf{y}_0$:

$$\hat{\boldsymbol{\rho}}_0 = \arg\max_{\boldsymbol{\rho}_0} \log P(\mathbf{y}_0|\boldsymbol{\rho}_0) = \arg\min_{\boldsymbol{\rho}_0} ||\mathbf{y}_0 - \mathbf{A}\boldsymbol{\rho}_0||^2. \quad (11)$$

However, since only single projection is available to estimate the three dimensional structure of $\boldsymbol{\rho}_0$, the estimation problem is severely ill-posed. Of course, the helical symmetry provides additional projection view at each $(m\phi_s, mz_s)_{m=1}^{u-1}$, and the minimization problem turns into

$$\hat{\boldsymbol{\rho}}_0 = \arg\min_{\boldsymbol{\rho}_0} \frac{1}{uN}\sum_{n=0}^{N-1}\sum_{m=0}^{u-1} ||\mathbf{y}_0 - \mathbf{A}\mathbf{P}_{(m\phi_s+\frac{2\pi n}{N}, mz_s)}\boldsymbol{\rho}_0||^2 \ . \quad (12)$$

Although additional views can be generated from a single projection due to the helical symmetry, it is not enough for reasonable quality image reconstruction. However, for sparse objects, the recent theory of so-called "compressed sensing" [2] shows that perfect reconstruction is possible even from samples dramatically smaller than the Nyquist sampling limit by $l_1$ minimization. Therefore, we employ the compressed sensing technique for reconstruction of structure $\boldsymbol{\rho}_0$.

**Step2**. Parameter Refinement : After obtaining the estimate of $\hat{\boldsymbol{\rho}}_0$, we need to minimize Eq. (10) with respect to alignment parameters. More specifically, we generate the library of projections for each sampling points. Then, by minimizing the distance between the $\mathbf{y}_k$ and the projection images within the library, we can readily find the corresponding alignment parameters. After alignment parameters are estimated, the initial model $\boldsymbol{\rho}_0$ can be further refined. Then, using the refined estimate of $\boldsymbol{\rho}_0$, the alignment parameter can be further refined. This procedure continues until convergence.

**Step3**. Estimation of Helix Parameter : Recall that for a fixed $(N, t, u)$, Step 1 and Step 2 correspond to the following:

$$\hat{\boldsymbol{\Xi}} = \arg\min_{\boldsymbol{\Xi}} C(\boldsymbol{\Xi}, N, t, u) \quad (13)$$

Therefore, the helix parameters and additional parameters for discretization problem can be estimated by

$$\left(\hat{N}, \hat{t}, \hat{u}\right) = \arg\min_{(N,t,u)} \hat{C}(\hat{\boldsymbol{\Xi}}, N, t, u) \ . \quad (14)$$

## 4. EXPERIMENTAL RESULTS

### 4.1. Synthetic Helices

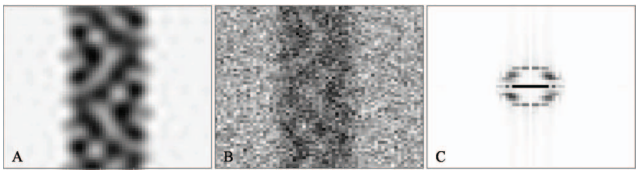


**Fig. 1**. A : A projection image. B : A projection image with 0db noise. C : The corresponding diffraction pattern.

We first tested our new algorithm using synthetic data. The synthetic micrograph images were made by projecting the synthetic model along several rotation angles and axial transitions. To be sure that our algorithm perform under noise data, we also added Gaussian noise to micrograph so that the overall SNR equal 0dB. One of the synthetic projections and the projections with noise are illustrated in Figs. 1(a) and (b), respectively. The corresponding diffraction pattern is illustrated in Figs. 1(c). The layer pattern of this sample corresponds to Type B and the unknown helix parameter is $(t, u)$ since only the subunit axial translation $z_s$ can be directly obtained from the layer line pattern in Fig. 1(b). The estimation provides that $t$ is one, and $u$ is nine. Figs. 2(a) and (b) denote the three dimensional structure of the original synthetic model and the final reconstruction structure when the 500 noiseless helix projections with unknown view angles are available, respectively. Figs. 2(c) illustrates the final reconstruction structure from 500 noisy helix projections. We can show that accurate structures are obtained by our new method.

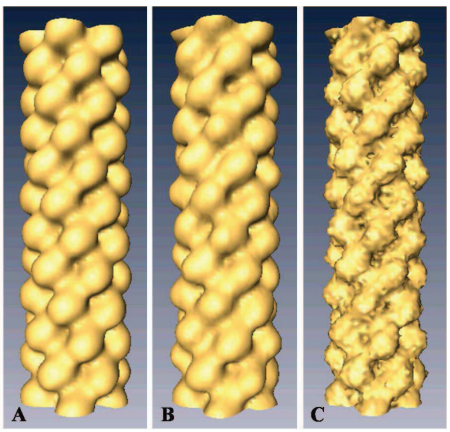


**Fig. 2**. A : The synthetic microtubule structure. B : Final reconstruction structure using noiseless 500 segments projection images. C : Final reconstruction structure using 500 segments projection image with 0db noise

### 4.2. Real P-pili data

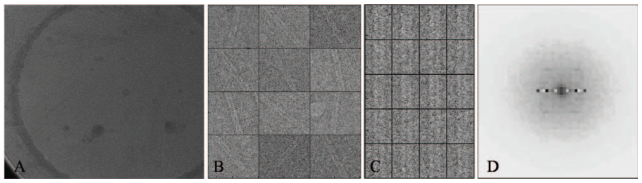


**Fig. 3**. A : P-pili Micrograph. B : Boxed segments of P-pili. C : Aligned segments. D : The average diffraction pattern

The bacterial attachment is a crucial initiating step in infection. Organelles called P-pili mediate microbial adhesion by binding to receptors in kidney cells. Therefore, high resolution reconstruction of P-pili structure contributes to clear the binding mechanism as well as design the therapeutics for preventing bacterial infections [4]. Figs. 3(a) is real P-pili micrograph taken by cryo-EM using negative staining. Figs. 3 (b)(c) are boxed segments of filaments before and after alignment, respectively. Rotational angle and shift value can be estimated efficiently by using radon transform. Among aligned particles, best 9,000 particles are finally selected for reconstruction steps, and the average diffraction pattern is illustrated in Figs. 3 (d). The diffraction pattern corresponds to Type C and the unknown parameters are $(N, t, u)$. The estimation provides that the number of strands is one, the number of turns per repeat is ten, and the number of subunit per repeat is thirty three. From these estimated helix parameters, we can obtain the rotation angle $109.1^\circ$ and axial translation 7.54 Å. After the symmetry parameter estimation, more elaborated three dimensional structure can be generated from iterative reconstruction. Fig. 4 illustrates the final converged three dimensional reconstruction result.

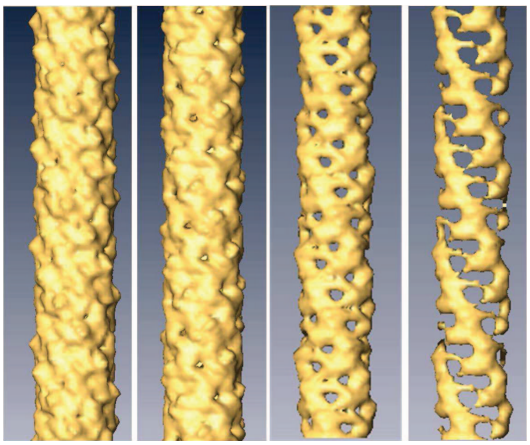

**Fig. 4**. Reconstructed P-pili structures using our algorithm with respect to several threshold values.

## 5. CONCLUSION

This paper described a novel model free iterative reconstruction algorithm for helical structure by exploiting the helical symmetry and the sparsity of the sample in the real space. We showed that the helix parameters can be estimated by criterion based on maximum-likelihood approach. Optimal values can be easily obtained by comparing likelihood for all feasible parameters. Thanks to the diffraction analysis, we showed that the parameter search space can be reduced to a space of 1, 2, or 3 dimensional *integers*, which drastically reduces the estimation complexity compared to the conventional single particle approach using *real* unknown parameters. We described the compressed sensing approach to improve the accuracy of the parameter estimation as well as the final reconstruction. Experiment results confirmed that our algorithm provides superior reconstruction of 3D helical structure.

## 6. REFERENCES

[1] E. H. Egelman, "A robust algorithm for the reconstruction of helical filaments using single-particle methods," *Ultramicroscopy*, vol. 85, no. 4, pp. 225–234, 2000.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[3] X.Q. Mu, S.J. Savarino, and E. Bullitt, "The three-dimensional structure of CFA/I adhesion pili: travelers diarrhea bacteria hang on by a spring," *Journal of Molecular Biology*, vol. 376, no. 3, pp. 614–620, 2008.

[4] X.Q. Mu and E. Bullitt, "Structure and assembly of P-pili: A protruding hinge region used for assembly of a bacterial adhesion filament," *Proceedings of the National Academy of Sciences*, vol. 103, no. 26, pp. 9861, 2006.