Fast Exact/Quasi-Exact FBP Algorithms for Triple-Source Helical Cone-Beam CT

Yang Lu*, Alexander Katsevich, Jun Zhao, Member, IEEE, Hengyong Yu, Senior Member, IEEE, and Ge Wang, Fellow, IEEE

Abstract—Cardiac computed tomography (CT) has been improved over past years, but it still needs improvement for higher temporal resolution in the cases of high or irregular cardiac rates. Given successful applications of dual-source cardiac CT scanners, triple-source cone-beam CT seems a promising mode for cardiac CT. In this paper, we propose two filtered-backprojection algorithms for triple-source helical cone-beam CT. The first algorithm utilizes two families of filtering lines. These lines are parallel to the tangent of the scanning trajectory and the so-called L lines. The second algorithm utilizes two families of filtering lines tangent to the boundaries of the Zhao window and L lines, respectively, but it eliminates the filtering paths along the tangent of the scanning trajectory, thus reducing the required detector size greatly. The first algorithm is theoretically exact for $\tau < 0.265R$ and quasi-exact for $0.265R \leq \tau < 0.495R$, and the second algorithm is quasi-exact for $\tau < 0.495R$, where $\tau$ and $R$ denote the object radius and the trajectory radius, respectively. Both algorithms are computationally efficient. Numerical results are presented to verify and showcase the proposed algorithms.

Index Terms—Cardiac CT, computed tomography (CT), cone-beam CT (CBCT), filtered backprojection (FBP), triple-source.

I. INTRODUCTION

THANKS to the modern detection technique, projection data can now be collected on a 2-D array of detectors [1], which is much faster than with the conventional single- or multi-row detectors. The more detector rows we use, the larger the longitudinal coverage we have, and the faster the data acquisition is for volumetric scanning. However, the detector array cannot be too large in a cost and dose efficient way, and the mechanical rotation of an X-ray source has already approached its limit. One alternative solution is to use multi-source computed tomography (CT) architecture, which led to the emergence of Siemens’ dual-source CT [2]. Given successful applications of dual-source cardiac CT scanners, triple-source cone-beam CT (CBCT) seems a promising mode for cardiac CT [3].

For the dual-source scanner, the minimum rotation interval is $90^\circ$ plus the fan angle $\alpha$ compared to $180^\circ + \alpha$ in a third-generation geometry. Since the fan angle $\alpha$ is relatively small, there is > 40% reduction in acquisition time. As a result, the dual-source scanner reduces the data acquisition time to 0.35 s, yielding 83 ms temporal resolution with cardiac gating.

There are major problems with current cardiac CT including dual-source CT [2]. Actually, even in the favorable cases, retrospectively reconstructed cardiac images still suffer from substantial motion blurring for patients who have high and irregular heartbeats because in practice each projection sector covers a projection angular range of a substantial length. Within such an angular range, the heart moves appreciably, especially when it is not in a relative stationary phase. As a benchmark, we routinely achieve ~0.3 mm spatial resolution in spiral CT of the temporal bone where the motion magnitude is much less than that of the heart [4], [5]. On the other hand, the spatial resolution with cardiac CT is at best in the millimeter domain.

As a natural extension to dual-source CT, triple-source CT promises to bring temporal resolution down to 0.2 s and hopefully to be as fast as electron-beam CT. Interestingly, for helical cone-beam reconstruction [6], [7], the trinity (triple-source architecture) is better than the duality (dual-source architecture) because the triple-source helical scan allows a perfect mosaic of longitudinally truncated cone-beam data to satisfy the Orlov condition and yields better noise performance than the dual-source counterpart [8].

Since the 1-D Hilbert transform can be obtained on a generalized PI-line/chord by backprojecting the weighted differential projection data, which was first proposed by Gel’fand and Graev [9] and generalized in the 2D SPECT case [10] and rediscovered in the 2D and 3D CT cases [11], we previously formulated a backprojection-filtration (BPF) approach for triple-source helical CBCT [12] and a corresponding “slow” filtered-backprojection (FBP) algorithm [13]. In this paper, we present two “fast” exact/quasi-exact FBP algorithms in this important case. The algorithms are not specifically designed for cardiac gating which is widely used in cardiac CT. However, the triple-source scanner achieves a data acquisition speed three times faster than the conventional CT, thus the temporal resolution...
is significantly increased and may be applied without gating in multiple dynamic cases such as perfusion studies in which gating is generally not possible.

This paper is organized as follows. In Section II, we summarize the general reconstruction formulae and the notations of the inter-PI lines and inter-PI arcs. In Section III, we analyze the intersections of the plane through a reconstruction point and the inter-PI arcs associated with the point. In Section IV, we derive the filtering directions and backprojection coefficients for our first algorithm. In Section V, we present the second algorithm after defining a different weight function. In Section VI, we analyze the percentage of the incorrectly weighted Radon planes. In Section VII, we describe the numerical results to validate our algorithms. In Section VIII, we discuss the relevant issues and conclude the paper.
where $D_f(y, \beta)$ is the cone-beam transform of $f, \theta$ the polar angle in the plane perpendicular to $[\beta(s, x), \alpha(s, x, \theta)]$ a unit vector perpendicular to $[\beta(s, x), \theta_m$, a point where $\phi(x, s, \theta)$ is discontinuous, $n(s, x, \alpha)$ a weight function, $C$ a finite union of $C^\infty$ curves in $\mathbb{R}^3$, $-\infty < a_1 < b_1 < \infty$, and $\gamma(s) := dy/\text{ds}$.

The aforementioned general inversion formula can be applied to any trajectory that satisfies Tuy’s condition, but only when the weight function $n(s, x, \alpha)$ is well designed can the inversion formula have a shift-invariant filtering structure. To derive fast exact FBP algorithms for triple-source helical CBCT, our general approach involves the following key concepts of and analyses on the inflection line, $A_\tau, T_\tau, L_\tau$, and $B_\tau$-curves.

**B. Infection Line**

On the detector plane, the boundaries of the Zhao window are expressed as

$$u(s) = \frac{D}{1 - \cos \theta}, \quad v(s) = \frac{Dh}{2\pi R (1 - \cos \theta)^3}, \quad (8)$$

where $D$ is the distance between the detector and the source, $s$ is the angular parameter relative to the corresponding source position, $\Delta s = -2/3\pi$ and $\Delta s = -4/3\pi$ are for the top and bottom boundaries respectively. Then

$$\dot{u}(s) = \frac{D}{1 - \cos \theta}, \quad (9)$$

$$\dot{v}(s) = \frac{D}{1 - \cos \theta} (1 - \cos \theta)^2, \quad (10)$$

$$\ddot{u}(s) = \frac{Dh}{2\pi R} \left[ \frac{1 - \cos \theta - (s + \Delta s) \sin \theta}{(1 - \cos \theta)^3} \right], \quad (11)$$

$$\ddot{v}(s) = \frac{Dh}{2\pi R} \left[ \frac{(1 - \cos \theta)(s + \Delta s) \cos s + 2 \sin s \cos s + s \sin s + \Delta s \sin s - 1)}{(1 - \cos \theta)^3} \right], \quad (12)$$

$$\frac{d^2 v}{dv^2} = \frac{\ddot{u}(s) - \ddot{u}(s)}{[\dot{u}(s)]^3} = \frac{h}{2\pi R} (s + \Delta s - \sin s), \quad (13)$$

The inflection point exists when $d^2 v/dv^2 = 0$. Thus, we obtain $s(t) = 2.60553$ and $s(t) = 3.66791$ when $\Delta s = -2/3\pi$ and $-4/3\pi$. The slope of the tangent line at $s = 0$ is computed as

$$\frac{dv}{du} = \frac{\ddot{u}(s)}{[\dot{u}(s)]^3} = \frac{h[1 - \cos \theta - (s + \Delta s) \sin \theta]}{2\pi R (1 - \cos \theta)}$$

$$= \frac{h}{2\pi R} \cos s \quad (14)$$

Because $\cos s(t) \approx \cos s(t) = -0.8506$, the slope is the same $(-0.13689/R)$ at both inflection points. For practical medical applications, it is common that $R_{FOV} \leq 0.5R$, and we can include a boundary limitation $x_1^2 + x_3^2 \leq r^2, r = 0.495R$, which is shown as the vertical lines $\Gamma_1$ and $\Gamma_2$ in Fig. 4. Now, the inflection lines (the tangent lines at $s(t)$ and $s(t)$, where $s(t)$ and $s(t)$ are the projection of $y_j(s(t))$ on the detector plane) and the boundary lines split the
Zhao window into the following three regions: \( G_1, G_2 \) and \( G_3 \). Only the points in \( G_1 \) and \( G_2 \) can have tangent lines with \( \Gamma^{\pm 1} \).

C. \( A \)-Curve and \( T \)-Curve

To construct an appropriate weight function, at first we must know how Radon planes intersect with the trajectories. Recall that the number of intersection points only changes when a Radon plane is tangent to the trajectory or contains one PI line/inter-PI line. Hence, if we find all such Radon planes, we can determine the distribution of the intersection points. Since each plane is uniquely determined by its normal vector, in the following sections we use unit vectors instead of the Radon planes. Let us define several key notions introduced in [17]. An \( A \)-curve consists of all unit vectors orthogonal to an inter-PI line. A \( T \)-curve consists of all unit vectors

\[
\alpha(s) = \pm \frac{(x - y(s)) \times \dot{y}(s)}{[(x - y(s)) \times \dot{y}(s)]^2}
\]

where \( s \) belongs to an inter-PI arc. Actually, the \( A \)-curve represents all Radon planes containing one inter-PI line, and the \( T \)-curve represents all Radon planes tangent to the trajectory. Since there are three inter-PI lines and three inter-PI arcs for a fixed \( x \), there are accordingly three \( A \)-curves and three \( T \)-curves. Similar to [17], let us use spherical coordinates \((\theta_1, \theta_2)\) to describe these curves on the unit sphere

\[
\alpha = (\cos \theta_1 \sin \theta_2, \sin \theta_1 \sin \theta_2, \cos \theta_2),
\]

\[-\pi \leq \theta_1 \leq 0, 0 \leq \theta_2 \leq \pi.
\]

With the identification \((\theta_1, \theta_2) \cong (\theta_1 + \pi, \theta_2) \mod 2\pi, \pi - \theta_2\), each \( \alpha \) corresponds to a unique plane through \( x \) with the normal vector \( \alpha \).

As an example, the \( A \)-curves and \( T \)-curves of point \( x = (1, 0, 0, 0) \) are illustrated in Fig. 5, where \( R = 1, h = 2\pi, T_1, T_2, \) and \( T_3 \) stand for the \( T \)-curves corresponding to the inter-PI arcs \( s_1^1 s_1^2, s_1^2 s_1^3, \) and \( s_1^3 s_1^4 \), respectively. Similarly, \( A_1, A_2, \) and \( A_3 \) are for the \( A \)-curves corresponding to the inter-PI arcs \( s_2^1 s_2^2, s_2^2 s_2^3, \) and \( s_2^3 s_2^4 \), respectively.

The \( A \)-curves and \( T \)-curves divide the surface of the unit sphere into several connected domains, in each of which all the planes through \( x \) have the same number of intersection points (IPs) with the inter-PI arcs of \( x \). Given an object point and one source trajectory, the number of IPs changes only when a Radon plane is tangent to the trajectory or contains an endpoint of the trajectory. The \( A \)-curve represents all planes containing the end-points of the trajectory, and the \( T \)-curve represents all planes tangent to the trajectory. If we pick any Radon plane and rotate it around one direction, the normal vector of this plane forms a curve on the unit sphere. Clearly, only when this curve intersects with the \( A \)-curve or \( T \)-curve does the number of IPs change. Thus, the \( A \)-curve and \( T \)-curve define the boundaries of different domains in which the number of IPs is constant. The distribution of IPs over the inter-PI arcs is listed in Table I.

To determine the distribution of IPs, we first pick a vector \( \alpha(\theta_1, \theta_2) \) in each domain, and then generate the plane through \( x \) and perpendicular to \( \alpha(\theta_1, \theta_2) \), and compute numerically the number of IPs.

By construction, a \( T \)-curve always starts from an \( A \)-curve and ends on another \( A \)-curve. It can be seen from Fig. 5 that a \( T \)-curve is possibly not smooth at some point \( a_c \), but the limits of unit tangent vectors at \( a_c^- \) and \( a_c^+ \) are equal. Such a point \( a_c \) is called a “cusp.” The cusp indicates that the two vectors determine the same plane, and \( a_c \) is the normal vector to that plane. It has been proved in [17] that the cusp is equivalent to the osculating plane \( \Pi(x) \) which goes through \( y_k(s_k(x)) \), \( i \in \{1, 2, 3\} \), parallel to \( \dot{y}_k(s_k(x)) \), and contains \( x \) (see Fig. 6). On the detector plane, this corresponds to a point where the projected boundary has a point of zero curvature, i.e., the point of inflection [18], [19].

The diagram we plot in spherical coordinates deforms smoothly as a function of \( x \). The new diagram is equivalent to the old one if the distribution of IPs remains the same. An essential change could happen when three boundaries intersect each other at one point, which is called “critical event.” A “critical event” may happen in the following seven cases.

1) Three \( A \)-curves intersect at one point.
2) Three \( T \)-curves intersect at one point.
TABLE I

<table>
<thead>
<tr>
<th></th>
<th>$s_1^1s_1^1$</th>
<th>$s_2^2s_2^2$</th>
<th>$s_3^3s_3^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1$</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$D_2$</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>$D_3$</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>$D_4$</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$D_5$</td>
<td>2</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>$D_6$</td>
<td>2</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>$D_7$</td>
<td>1</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>$D_8$</td>
<td>-</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>$D_9$</td>
<td>1</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>$D_{10}$</td>
<td>-</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$D_{11}$</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$D_{12}$</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>$D_{13}$</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

3) Two $A$-curves and one $T$-curve intersect at one point.
4) One $A$-curve and two $T$-curves intersect at one point.
5) A $T$-curve becomes tangent to an $A$-curve at a point of nonsmoothness (i.e., cusp).
6) When the order of tangency (i.e., the zero derivatives of this order) at the beginning of the $T$-curve is increased, the $T$-curve reemerges on the other side of the $A$-curve.
7) $T$-curve develops a smooth dent and becomes tangent to an $A$-curve.

From Lemmas 2–3, we know Case 1, 2, 4, 6, 7 do not occur for $r < 0.495R$ and Case 5 does not occur for $r < 0.265R$. Case 3 is possible. If Case 3 takes place, the tangency of $T$-curve and $A$-curve will move across another $A$-curve, then one domain disappears. For example, when $x = (0, -0.15, 0)$ gradually changes to $x = (0, 0, 0)$, the tangency of $T_3$ and $A_2$ will move across $A_3$, and domain $D_{10}$ disappears [see Fig. 7(a)]. Case 5 is possible for $r \geq 0.265R$. If it takes place, a $T$-curve will only intersect $A$-curves at the endpoints. That is, the cusp of that $T$-curve and one domain disappear [see Fig. 7(b)].

D. $L$-Curve

The $L$-curve is used to split the domain $D_4$ into subdomains, making the weight function $n$ continuous across all the $A$-curves. This is the key requirement, which will allow us to develop efficient reconstruction algorithms (see Section IV below). Thus, the $L$-curve should not go across an $A$-curve. Let us fix $x$ and run $s$ over the three inter-PI arcs, $\hat{x}$ forms a trajectory on the detector plane. Because at the endpoint of the inter-PI arc the line connecting $y_k(s)$ and $x$ happens to be an inter-PI line, $\hat{x}$ always starts from one endpoint of the inter-PI arc on the boundary of the Zhao window, and ends at the other one. Hence, whatever the trajectory of $\hat{x}$ is, part of the trajectory is in $G_2$. In other words, $\hat{x}$ will run across one inflection line, then move in $G_2$, and finally cross the other inflection line. Note that $\hat{x}$ on the inflection line indicates a plane containing the inflection line, i.e., a cusp in one $T$-curve. From Lemma 3, the cusps always belong to the boundary of domain $D_4$. Thus, they can be used as the endpoints of the $L$-curve. A family of $L$-curves is formed as follows. Run $s$ over the three inter-PI arcs of $x$. If $\hat{x}$ is in $G_2$ and above $\hat{s}_4$, where $\Gamma^{a-1}$ intersects $L_{ab}$ (Fig. 4), find the plane through $\hat{x}$ and $\hat{s}_4$. If $\hat{x}$ is in $G_2$ and below $\hat{s}_4$, where $\Gamma^{b-1}$ intersects $L_{ab}$, find the plane through $\hat{x}$ and $\hat{s}_4$. If $\hat{x}$ is in $G_2$ and between $\hat{s}_4$ and $\hat{s}_u$, find the plane through $\hat{x}$ and parallel to the $v$-axis. Then, we can plot all the normal unit vectors of these planes in the spherical coordinates $(\theta_1, \theta_2)$. This gives us three $L$-curves. The corresponding lines on the detector plane are called $L$-lines. Fig. 8 shows the $L$-curves on the diagram in spherical coordinates $(\theta_1, \theta_2)$, where $L_1, L_2, L_3$ denote the $L$-curves corresponding to the inter-PI arcs $s_1^1s_1^1, s_3^3s_3^3$, and $s_5^5s_5^5$, respectively. As is seen from the above construction, the $L$-curve always starts and ends on the cusps, and not defined for those parameter values when $\hat{x}$ is not in $G_2$.

For $r \geq 0.265R$, one or more cusps will disappear if the critical event Case 5 occurs, then the $L$-curve may start from the intersection of $T$-curve and $A$-curve, and end at one $A$-curve. Also, the $L$-curve may start from one cusp and end at one $A$-curve or start from the intersection of $T$-curve and $A$-curve, and end at one $A$-curve. For example, see Fig. 9. $L_1$ starts from the intersection of $T_3$ and $A_2$, and ends at the intersection of $T_2$ and $A_2$;
$L_2$ starts from one point on $A_2$, and ends at the cusp of $T_1$; $L_3$ starts from the cusp of $T_1$, and ends at one point on $A_2$.

Whatever the endpoints of $L$-curves are, the $L$-curves intersects at one point, i.e., $\theta_2 = \pi$ or 0 in the spherical coordinates (which corresponds to the plane containing $x$ and parallel to the $x_1 = x_2$ plane). Then $D_4$ is split into several sub-domains. If the endpoints of $L$-curves are cusps, by Lemma 5, each sub-domain contains only one $A$-curve. If not, small “line segments” on $A$-curves may appear and the sub-domains may contain more than one $A$-curve [see Fig. 9(b)].

**E. $B_3$-Curve**

A $B_3$-curve consists of all unit vectors perpendicular to $x = y_0(s), i \in \{1, 2, 3\}$. Each intersection of $B_3$- and $A$-curves corresponds to a plane containing an inter-PI line and $y_i(s)$. Each intersection of $B_3$ and $T$-curves corresponds to a plane tangent to an inter-PI arc and containing $y_i(s)$. For example, pick $x \in \Omega$ with $\hat{x} \in G_2$, $\hat{x}$ is above $L_0$, where $L_0$ is the projection of the helical tangent at the current source position. Denote $L(\theta) := DP(s) \cap \Pi(x, \alpha(s, x, \theta)), \text{ where DP}(s)$ is the detector plane corresponding to the source position $s$, and $L(\theta)$ is the projection of the plane through $x$ with the normal vector $\alpha(s, x, \theta)$ (Fig. 10). As $\theta$ increases, $\alpha(s, x, \theta) \in \beta(s, x)$ rotates clockwise on DP($s$), and the following sequence of events takes places. First, $\Pi(x, \alpha(s, x, \theta))$ intersects $s_1 \neq s_2$, and a pair of IPs is born. On the unit sphere, this is seen as an intersection of $B_3$ and $A_1$, after which $B_3$ enters $D_5$ [Fig. 11(a)]. Second, $\Pi(x, \alpha(s, x, \theta))$ intersects $s_2 \neq s_1$ and another pair of IPs is born. On the unit sphere, this is seen as an intersection of $B_3$ and $A_3$, after which $B_3$ enters $D_{11}$. Third, a swap of two IPs takes place.

Fig. 10. Domains on the detector plane.

On DP($s$) this happens when $\theta = \theta_0$, $L(\theta_0)$ is parallel to the helical tangent. On the unit sphere, this means that $B_3$ is tangent to $T_1$ at $\alpha_0 = \alpha(s, x, \theta_0)$. Fourth, $B_3$ exits $D_{11}$ by intersecting $T_1$. On DP($s$) this takes place when $L(\theta_0)$ is tangent to $\Gamma^{+2}$. Finally, $\Pi(x, \alpha(s, x, \theta))$ intersects the $L$-line. This will not change the number of IPs but it will be useful for construction of the weight function. On the unit sphere, this is seen as an intersection of $B_3$ and $L_1$.

The jumps across an $A$-curve can only be of two types: from a 1-IP domain to a 3-IP domain and from a 3-IP domain to a 5-IP domain. Note that the $B_3$-curve is tangent from the inside to $T_1$, which means a swap of two intersection points at $\alpha = \alpha_0$ where $\text{sgn}(\alpha_0 \cdot y(s)) = 0$ (see [17]). For a fixed $s$, if $x$ is allowed to change slightly inside the Zhao window, the tangency point will move from $D_{11}$ to $D_5$ across $A_3$ (or from $D_{11}$ to $D_0$ across $A_1$) [Fig. 11(a) and (b)]. If $x$ projects into $G_1$ or $G_3$, the $B_3$-curve will pass not only through $D_5$ (or $D_0$) and $D_{11}$ but also through $D_7$ and $D_{12}$ [Fig. 11(c)]. The similar results can be obtained if the source is on $y_2(s)$ or $y_3(s)$.

**IV. FIRST FAST FBP ALGORITHM**

Now, we are ready to design our first fast FBP algorithm for triple-source helical CBCT. We will start with specifying the weight function $n(s, x, \alpha)$. Then, we will determine the filtering directions by finding the discontinuities of $\phi(s, x, \theta) := \text{sgn}(\alpha \cdot y(s)) n(s, x, \alpha)$. Then, the backprojection coefficients can be calculated according to (6). Once the filtering lines and the backprojection coefficients are determined, we can use (3) to reconstruct the object.
To have an efficient FBP structure, the weight function $n(s,x;\alpha)$ should be continuous across all $A_i$-curves. Thus, the weight function can be defined as shown in Table II. The values in Table II are the weights assigned to IPs. For example, in the $D_1$ domain the Radon plane has only one IP on the inter-PI segment $s_1^1 s_2^1$. Accordingly, we assign weight 1 to this IP and use “…” to indicate that there is no IP on the inter-PI segments $s_1^1 s_2^2$, $s_1^2 s_2^3$, and $s_3^2 s_3^3$. In the $D_{11}$ domain the Radon plane has three IPs on $s_1^1 s_1^2$, one IP on $s_2^2 s_3^2$, and one IP on $s_3^2 s_3^3$. Thus, we assign weight 1 to one IP on $s_1^1 s_1^2$ and weight 0 to all other IPs.

### Table II

<table>
<thead>
<tr>
<th>$A_i$</th>
<th>$s_1^i s_2^i$</th>
<th>$s_2^i s_3^i$</th>
<th>$s_3^i s_4^i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1$</td>
<td>+1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$D_2$</td>
<td>-</td>
<td>+1</td>
<td>-</td>
</tr>
<tr>
<td>$D_3$</td>
<td>-</td>
<td>-</td>
<td>+1</td>
</tr>
<tr>
<td>$D_4$</td>
<td>+1,0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>$D_5$</td>
<td>0</td>
<td>+1,0</td>
<td>-</td>
</tr>
<tr>
<td>$D_6$</td>
<td>-</td>
<td>+1,0</td>
<td>0</td>
</tr>
<tr>
<td>$D_7$</td>
<td>0</td>
<td>-</td>
<td>+1,0</td>
</tr>
<tr>
<td>$D_8$</td>
<td>-</td>
<td>0</td>
<td>+1,0</td>
</tr>
<tr>
<td>$D_9$</td>
<td>0</td>
<td>0</td>
<td>+1,0,0</td>
</tr>
<tr>
<td>$D_{10}$</td>
<td>0</td>
<td>+1,0,0</td>
<td>0</td>
</tr>
<tr>
<td>$D_{11}$</td>
<td>0</td>
<td>0</td>
<td>+1,0,0</td>
</tr>
<tr>
<td>$D_{12}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$D_{13}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$D_{14}$</td>
<td>0</td>
<td>0</td>
<td>+1</td>
</tr>
<tr>
<td>$D_{15}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### A. Construction of the Weight Function $n(s,x;\alpha)$

To find the backprojection coefficients, let us pick a representative point in each area, determine the discontinuities of $\phi(s,x;\alpha) := \text{sgn}(\alpha \cdot \dot{y}(s)) n(s,x;\alpha)$, and extend the results by continuity to the entire area. A discontinuity of $\text{sgn}$ occurs only when a $B_x$-curve is tangent to a $T$-curve from the inside. In other words, the $B_x$-curve does not go across the $T$-curve, but stays on one side in a neighborhood of the point of tangency. On the detector plane, $L(\hat{x})$ is parallel to the helical tangent for all $\hat{x} \in G_1 \cup G_2 \cup G_3$. Hence, this gives a family of filtering lines parallel to $L_0$, where $L_0$ is the projection of the helical tangent at $y(s)$. The swap of two IPs changes the weight from $\phi^+ = 0$ to $\phi^- = 1$. The backprojection coefficient is computed as $c_0 = \phi^+ - \phi^- = (+1)(0) - (-1)(1) = 1$ [Fig. 12(a)].

#### C. Discontinuity of $n(s,x;\alpha)$

By construction, the weight function $n$ is continuous across all inter-PI lines. A discontinuity of $n$ occurs only when a $B_x$-curve intersects a $T$-curve or an $L$-curve. Without loss of generality, pick $y_1(s_0)$ on $s_1^1 s_1^2$. For $\hat{x} \in G_2$, after the swap mentioned in the above paragraph the weight at the current position is zero. Hence, when the $B_x$-curve passes through a $T$-curve, i.e., from $D_{11}$ to $D_{14}$, $n$ is continuous. Possible jumps of $n$ may only occur when a $B_x$-curve passes through an $L$-curve, i.e., from $D_{13}$ to $D_{14}$ or from $D_{12}$ to $D_{14}$ in Fig. 8. On the detector plane, this occurs when $L(\hat{x})$ overlaps the $L$-line of $\hat{x}$. Then, the backprojection coefficients are computed as $c_1 = \phi^+ - \phi^- = (+1)(1) - (+1)(0) = 1$ [Fig. 12(b)]. For $\hat{x} \in G_1 \cup G_3$, the $B_x$-curve will not enter $D_{14}$. Instead, it passes through a second $T$-curve twice, i.e., from $D_{43}$ to $D_{12}$ and from $D_7$ to $D_1$. From Table II, the jumps of $n$ may only occur in the latter intersection. On the detector plane, this happens when $L(\hat{x})$ overlaps the line tangent to $\Gamma^{\pm 1}$. Then, the backprojection coefficients are computed as $c_1 = \phi^+ - \phi^- = (+1)(1) - (+1)(0) = 1$. 

---

**Fig. 12.** (a) Determination of $c_0$. (b) Determination of $c_L$. 

---

**Fig. 11.** (a) $B_x$-curve is tangent to a $T$-curve in $D_{11}$ for the source on $y_1(s)$. (b) $B_x$-curve is tangent to a $T$-curve in $D_5$ for the source on $y_1(s)$. (c) $B_x$-curve passes through the second $T$-curve when the source on $y_1(s)$. 

---

Authorized licensed use limited to: to IEEEXplore provided by Virginia Tech Libraries. Downloaded on June 16, 2010 at 14:59:35 UTC from IEEE Xplore. Restrictions apply.
D. Required Detector Area

Fig. 13(a) and (b) summarizes the filtering lines and the backprojection coefficients discussed above. In these figures, \( L_0' \) is the line parallel to \( L_0 \) and \( L_q \) denotes the \( L \)-line. To implement the proposed algorithm, the filtering lines cannot be truncated. Thus, the detector size should be large enough to cover the area bounded by \( \Gamma_1, \Gamma_2, L_{\text{max}}, \) and \( L_{\text{min}} \), where \( L_{\text{max}} \) and \( L_{\text{min}} \) are the lines across the intersections of (1) \( \Gamma_1 \) and \( \Gamma_1' \) and (2) \( \Gamma_2 \) and \( \Gamma_2' \), respectively, and parallel to \( L_0 \) (Fig. 14). The required detector area is determined by two factors: 1) the ratio of the pitch \( h \) and the scanning radius \( R \) and 2) the ratio of the object support radius \( r \) and the scanning radius \( R \). If \( R \) is fixed, the required detector area grows as \( h \) or \( r \) increases.

V. SECOND FAST FBP ALGORITHM

Again, the design of our second fast FBP algorithm starts with specifying new weights (Table III). By construction, \( n(s, x, \alpha) \) is continuous across all inter-\( P \) lines. More importantly, a swap of two IPs takes place when a \( B_s \)-curve becomes tangent to a \( T \)-curve [17], and \( n(s, x, \alpha) \) changes from +1 to -1. Recall that the discontinuity of \( \text{sgn}(\alpha \cdot \hat{y}(s)) \) appears only when a \( B_s \)-curve is tangent to a \( T \)-curve from inside.

Since both \( n(s, x, \alpha) \) and \( \text{sgn}(\alpha \cdot \hat{y}(s)) \) are discontinuous at that point, the function \( \hat{y}(s, x, \theta) := \text{sgn}(\alpha \cdot \hat{y}(s)) n(s, x, \alpha) \) is continuous. Thus, the filtering operation along the tangent of the scanning trajectory is eliminated.

A discontinuity of \( n \) can occur only when a \( B_s \)-curve intersects a \( T \)-curve or an \( L \)-curve. Following the discussion in Section IV, jumps of \( n \) may occur when 1) a \( B_s \)-curve passes through a \( T \)-curve, i.e., from \( D_{11} \) to \( D_1 \) or from \( D_1 \) to \( D_{12} \) in Fig. 11(a) and (c), and 2) a \( B_s \)-curve passes through an \( L \)-curve, i.e., from \( D_{12} \) to \( D_{12} \) or from \( D_{12} \) to \( D_{12} \) in Fig. 8. On the detector plane, this gives two families of filtering lines: the lines tangent to \( \Gamma_1 \) or \( \Gamma_2 \) and the \( L \)-lines. Note that the filtering lines tangent to \( \Gamma_1 \) are different from those for our first fast FBP algorithm (Fig. 15), because the discontinuity of \( n(s, x, \alpha) \) occurs on the different side of the cusp. Then, the backprojection

\[
\begin{array}{cccc}
D_1 & s_1 s_2 & s_1' s_2' & s_1' s_1' \\
D_2 & -1 & 1 & - \\
D_3 & - & - & 1 \\
D_4 & +1, -1 & 1 & - \\
D_5 & +1, -1 & 1 & - \\
D_6 & +1, -1 & 1 & - \\
D_7 & - & +1, -1 & 1 \\
D_8 & +1 & - & +1, -1 \\
D_9 & -1, 1, -1 & 1 & +1 \\
D_{10} & 0 & +1 & +1 \\
D_{11} & 0 & 0 & 1 \\
\end{array}
\]
coefficients can be calculated as $c_i = \phi^+ - \phi^- = (+1)(1) - (+1)(0) = 1$ and $c_i = \phi^+ - \phi^- = (+1)(1) - (+1)(0) = 1$.

The reconstruction formula for the second algorithm is very similar to that for the first algorithm. The only difference lies in the selection of the filtering lines. For clarity, our second fast FBP algorithm is illustrated in Fig. 16. Because the filtering paths along the tangent of the scanning trajectory are eliminated, the required detector area is reduced by at least 30% (Fig. 14).

**VI. ANALYSIS ON THE QUASI-EXACTNESS**

By Lemma 3, there are two types of “line segments” according to different critical events. First, let us consider the “line segment” related to a critical event in Case 3. Recall that before entering $D_4$ the $B_4$-curve will be tangent to a $T$-curve. For the first algorithm, at the tangency the weight $n_1$ changes from 1 to 0, then it does not change whether the $B_4$-curve enters $D_4$ across an $A$-curve or a $T$-curve. For the second algorithm, if the weight $n$ changes from 1 to $-1$ at the tangency, then it will jump from $-1$ to 0 when the $B_4$-curve enters $D_4$. If the $B_4$-curve enters $D_4$ across an $A$-curve [i.e., the “line segment”; see Fig. 7(a)], the FBP structure is ruined. Thus, the critical event in Case 3 will only affect the second algorithm, without damaging the FBP structure of the first algorithm. Then, let us consider the “line segment” related to a critical event in Case 5, which only occurs when $r \geq 0.265R$. From the discussion in Section III.D such “a line segment” is the boundary of the one-IP and three-IP domains. Hence, for both algorithms the weight $n_1$ will jump from 1 to 0 if the $B_4$-curve enters $D_4$ across the “line segment,” and the FBP structure is ruined (if the exactness is to be maintained. By sacrificing the exactness we retain the FBP structure). Consequently, the first algorithm is theoretically exact for $r < 0.265R$ and not exact for $0.405R \geq r \geq 0.265R$, and the second algorithm is not exact for $r < 0.405R$.

Since the algorithms are not always exact, we are interested in estimating what percentage of the Radon planes is incorrectly calculated. If the Radon planes with approximate weighting only have a small percentage, the algorithms can be considered quasi-exact, and we can still reconstruct with high image quality.

First, let us consider the incorrectly weighted planes caused by the critical event in Case 3. It appears in the area $r < 0.405R$. Let us fix $x$ for $r < 0.405R$; denote the intersection of the Radon plane and the detector plane as $I(\theta, \theta \in [0, 2\pi])$, run $s$ over the three inter-PI arcs, and see what happens with $\hat{x}$ and $I(\theta)$. Based on the discussion in Section III-E, for $\hat{x}$ in $G_2$, if the critical event occurs, the $B_4$-curve will first intersect a $T$-curve, and then go across an $A$-curve. For example, in Fig. 11(a) the $B_4$-curve will first intersect $T_2$, then enter $D_4$ across $A_3$. On the detector plane, this corresponds to that $\hat{x}$ intersects the tangent of $\Gamma^{2+}$ before the inter-PI line $\overline{\Gamma_3 \Gamma_2}$ while $I(\theta)$ is rotated clockwise. Therefore, the Radon planes between the tangent of $\Gamma^{2+}$ and $\overline{\Gamma_3 \Gamma_2}$ are not exactly weighted. Because the case of $\overline{\Gamma_3 \Gamma_2}$ is positive and the slope of the tangent of $\Gamma^{2+}$ is less than $\frac{\pi}{2}R$, the percentage of the incorrectly weighted Radon planes is less than $\rho = \beta/\pi$, where $\alpha < \beta < 2 \arctan(2)/\sqrt{3} \tan(\alpha/(2)) = \alpha = \arctan(h/(2\pi R))$ (Appendix). It is common that $h/R < 0.2$ in practical applications, hence $\rho < 1.17\%$. For $\hat{x}$ in $G_1$ ($G_3$), if the critical event occurs, the $B_4$-curve will first go across an $A$-curve, and then over a $T$-curve. On the detector plane, this corresponds to the case when $I(\theta)$ intersects the tangent to $\Gamma^{2+}(\Gamma^{-1})$ before the inter-PI line $\overline{\Gamma_3 \Gamma_2}$ while $I(\theta)$ is rotated clockwise. Hence, the Radon planes between the tangent of $\Gamma^{2+}(\Gamma^{-1})$ and $\overline{\Gamma_3 \Gamma_2}$ are not exactly weighted. Because the slope of $\overline{\Gamma_3 \Gamma_2}$ is negative and the slope of the tangent of $\Gamma^{2+}(\Gamma^{-1})$ is more than $-\alpha\pi R$, the percentage of the incorrectly weighted Radon planes is less than $\rho = 2.33\%$ for $h/R = 0.2$.

Then, let us consider the incorrectly weighted planes caused by the critical event in Case 5. Recall that the $L_2$-curves are used to split the domain $D_4$ into sub-domains, making the weight function $n$ continuous across all the $A$-curves, and the cusps are
the starting and ending points of the $L$-curves. If the endpoints of the $L$-curves are not the cusps, there will be small fractions (or “line segments”) on the $A$-curves, making the weight function $n$ discontinuous across them and ruining the FBP structure of our algorithms. It possibly occurs for $r \geq 0.265R$. As discussed above, the $B_{0L}$-curve will first enter a 1-IP domain from a 3-IP domain across the “line segment,” and then pass through an $L$-curve. On the detector plane, this corresponds to that $L(\theta)$ intersects the inter-PI line $\frac{\theta_s}{\theta} \overrightarrow{\theta} \overrightarrow{\theta}$ before the $L$-line while $L(\theta)$ is rotated clockwise. Recall that if the cusp is not in $D_{1L}, \theta_s$ is possibly to the left of $\hat{s}_u$ or $\hat{s}_r$ is to the right of $\hat{s}_r$. Thus, the slope of $\frac{\theta_s}{\theta} \overrightarrow{\theta} \overrightarrow{\theta}$ is always more than $-0.35\frac{h}{R}$. Because the slope of the $L$-curve is never positive, the percentage of the incorrectly weighted Radon planes is less than $\rho = 2.57\%$ for $h/R = 0.2$. On the other hand, based on the discussion on Lemma 3, one or more cusps may possibly remain even when $r \geq 0.265R$, which means that less “line segments” related to critical events will appear in Case 5, and in fact more Radon planes may be correctly weighted.

VII. NUMERICAL RESULTS

To verify and showcase our proposed fast FBP algorithms, numerical tests were performed using the Clock phantom. This phantom consists of ellipses, as parameterized in Tables IV. In the simulations, the origin of the reconstruction coordinate system was set to the center of each phantom. The spherical phantom support was of 375 mm for the experiment. Three sources were arranged uniformly along a circle with their corresponding detectors on the opposite side. The source-detector distance was 1000 mm. Projections were generated from 1000 view angles while the sources and the detectors were constantly moved along three helices in one turn. The helix was of 750 mm distance was 1000 mm. Projections were generated from 1000 responding detectors on the opposite side. The source-detector system was set to the center of each phantom. The spherical phantom consists of ellipses, as parameterized in Tables IV.

In the simulations, the origin of the reconstruction coordinate system was set to the center of each phantom. The spherical phantom support was of 375 mm for the experiment. Three sources were arranged uniformly along a circle with their corresponding detectors on the opposite side. The source-detector distance was 1000 mm. Projections were generated from 1000 view angles while the sources and the detectors were constantly moved along three helices in one turn. The helix was of 750 mm distance was 1000 mm. Projections were generated from 1000 responding detectors on the opposite side. The source-detector system was set to the center of each phantom. The spherical phantom consists of ellipses, as parameterized in Tables IV.

The implementation of our algorithms consists of the following steps. 

Step 1) Differentiate each projection with respect to variable $s$.

Step 2) For each $y_i(s), i \in \{1, 2, 3\}$ perform the Hilbert transform of derivative data along the given filtering directions on the corresponding detector plane.

Step 3) Backproject the filtered data on the inter-PI segments to reconstruct the object point.

A more detailed description of the proposed algorithm is similar to [20] except for the following differences: 1) in triple-helices geometry the filtered data are backprojected on inter-PI segments; 2) there are two families of filtering lines for each algorithm so that each point on the detector plane will be filtered twice. Since our algorithms allow shift-invariant filtration, all results are in Cartesian coordinates directly, and there is no coordinate transform like what we have used in the slow-FBP algorithm [13] or BPF algorithm [12].

The algorithms were coded in MATLAB and executed on a regular PC (Intel Core2 Duo CPU 3.06 GHz, 4 GB RAM). Reconstructed images are shown in Fig. 17. Our numerical results show that in the case of $r = 0.495R$ both two algorithms produced high quality images.

VIII. DISCUSSIONS AND CONCLUSION

Although our previously published BPF algorithm for triple-source helical CBCT can indeed produce excellent image quality, FBP algorithms (either “slow” or “fast”) are computationally desirable for several reasons, such as being amenable for parallel processing as discussed in [13]. In particular, while the computational structures of our BPF algorithm and FBP algorithms are quite similar, the FBP algorithms avoid densely sampled intermediate reconstruction in the PI-line-based coordinate system, and more importantly they can reconstruct a region of interest (ROI) or volume of interest (VOI) much more efficiently than the BPF counterpart. Note that ROI/VOI reconstruction is very common in medical imaging. A related technology called “interior tomography” is being actively developed to target this type of problems [21], [22]. Then, an interesting possibility would be to develop tripe-source interior CBCT.

The proposed two fast exact/quasi-exact FBP algorithms for triple-source helical CBCT have their advantages and disadvantages. From the perspective of exact reconstruction, the first algorithm is more desirable than the second algorithm because it is not affected by critical events in Case 3. However, in terms of efficient data acquisition, it requires a larger detector area than the second algorithm. In the medical CT field, the rectangular detector shape is most popular, and the helical pitch may be varied case by case. Therefore, it is practically possible to have projection data for reconstruction using either or both of our two fast

<table>
<thead>
<tr>
<th>Table IV</th>
<th>PARAMETERS OF THE CLOCK PHANTOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>$x_c$</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>0.4</td>
</tr>
<tr>
<td>4</td>
<td>0.69</td>
</tr>
<tr>
<td>5</td>
<td>0.8</td>
</tr>
<tr>
<td>6</td>
<td>0.69</td>
</tr>
<tr>
<td>7</td>
<td>0.4</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>-0.4</td>
</tr>
<tr>
<td>10</td>
<td>-0.69</td>
</tr>
<tr>
<td>11</td>
<td>0.5</td>
</tr>
<tr>
<td>12</td>
<td>-0.69</td>
</tr>
<tr>
<td>13</td>
<td>0.4</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0.25</td>
</tr>
<tr>
<td>16</td>
<td>0.43</td>
</tr>
<tr>
<td>17</td>
<td>0.5</td>
</tr>
<tr>
<td>18</td>
<td>0.43</td>
</tr>
<tr>
<td>19</td>
<td>0.25</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>-0.25</td>
</tr>
<tr>
<td>22</td>
<td>-0.43</td>
</tr>
<tr>
<td>23</td>
<td>-0.5</td>
</tr>
<tr>
<td>24</td>
<td>-0.43</td>
</tr>
<tr>
<td>25</td>
<td>-0.25</td>
</tr>
</tbody>
</table>
Fig. 17. Reconstructed images of the Clock phantom with $r = 3.75$ mm using (a) the first fast FBP algorithm and (b) the second fast FBP algorithm. Images were reconstructed at $x_3 = 0$ mm and displayed in [0.95, 1.05]. Additionally, (c) and (d) show the differences between the reconstructed images and the ground truth respectively in the display window [-0.05, 0.05].

FBP algorithms. A detailed comparison of these two algorithms should be done in terms of image quality and radiation dosage. Other designs for fast FBP in the triple-source helical CBCT are also possible.

One concern with triple-source CT is the increment in scattering artifacts, which was addressed in our earlier paper for the BPF formulation [13]. Although the height of the Zhao window for triple-source helical CT is one third of the Tam-Danielsson window, our proposed fast FBP formulation requires additional data beyond the Zhao window. In this case, the total dose delivered to the patient can be controlled to be comparable to that in the single-source case, and the total amount of scattering would be similar as well. Another way is to utilize source multiplexing so that not all the sources are on simultaneously. Other scattering correction means, such as interior tomography [21], [22], may be applied as well.

In conclusion, based on our previous triple-source helical CBCT work and guided by the Katsevich general reconstruction scheme, we have proposed two fast FBP algorithms for triple-source helical CBCT. The proposed algorithms utilize the inter-PI line and inter-PI arcs, and have a shift-invariant filtering structure. Unlike our slow-FBP algorithm performing filtration spatial-variantly line by line, the proposed fast-FBP algorithms filter projection data spatial-invariantly view by view, representing a significant computational benefit. Since triple-source helical CBCT may triple temporal resolution, it seems a promising mode for cardiac CT and other applications, and our proposed algorithms may find applications in this context.

APPENDIX

A. Auxiliary Lemmas

Let us fix a point $x \in \Omega$ and find its three associated inter-PI lines as shown in Fig. 3. Then, we can select a source position $s \in (s^j_1, s^j_3), j \in \{1, 2, 3\}$ and determine how the inter-PI lines project onto the corresponding detector plane. For simplicity, here and below the projection of $y_j(s), j \in \{1, 2, 3\}$ on a detector plane is denoted by $\hat{s}_j$.

Lemma 1: On a detector plane, the slopes of the projected inter-PI lines $\frac{s^j_1}{s^j_3}$ and $\frac{s^j_2}{s^j_3}$ are always positive, and that of the inter-PI line $\frac{s^j_1}{s^j_3}$ is always negative.

Proof: Without loss of generality, suppose the source position we select is $y_j(s_0), s_0 \in (s^j_1, s^j_3)$. By construction, $s^j_2 - s^j_1 < 2\pi, s^j_3 - s^j_1 < 2\pi, s^j_3 - s^j_2 > 0$. Hence, the projections of $s^j_1, s^j_2$, and $s^j_3$ are always to the left of those of $s^j_2, s^j_3$, and $s^j_1$ respectively (Fig. 18). When $s \in (s^j_1, s^j_3)$ changes, the point $\hat{x}$, i.e., the projection of $x$ onto the detector plane, moves inside the region $G := G_1 \cup G_2 \cup G_3$. Clearly, $\hat{x}$ reaches its highest (resp., lowest) position in the vertical direction when $\hat{x}$ is at the intersection of $\Gamma_1$ and $\Gamma_1^2$ (resp., $\Gamma_2$ and $\Gamma_2^2$). Also, the vertical coordinates at these points are $\nu_{\text{max}} = 1.3794(Dh)/(2\pi R)$ and $\nu_{\text{min}} = -1.3794(Dh)/(2\pi R)$, respectively. Moreover, the lowest point on $\Gamma_1 + 2$ and the highest point on $\Gamma_2$ are $\nu_{\text{min}} = 1.3801(Dh)/(2\pi R)$ and $\nu_{\text{max}} = -1.3801(Dh)/(2\pi R)$, respectively. Evidently, $\nu_{\text{max}} \leq \nu_{\text{min}}$ and $\nu_{\text{max}} \leq \nu_{\text{max}}$. Since $s^j_1$ and $s^j_3$ are to the left of $s^j_2$ and $s^j_3$, the slopes of the inter-PI lines $\frac{s^j_1}{s^j_3}$ and $\frac{s^j_2}{s^j_3}$ are positive for all $\hat{x} \in G$.

As follows from the definition in [14], the inter-PI line $\frac{s^j_1}{s^j_3}$ satisfies:

$$\begin{align} x_1 &= \frac{Rt \cos s^j_2 + R(1 - t) \cos s^j_3}{t} \\
 x_2 &= \frac{Rt \sin s^j_2 + R(1 - t) \sin s^j_3}{t} \\
 x_3 &= \frac{R \sin (s^j_2 - \frac{2\pi}{3}) + \frac{R}{2} (1 - t) (s^j_3 - \frac{4\pi}{3})}{t} \end{align}$$

(17)

where $G_2$ and $s^j_2$, $s^j_3 \in (s_0, s_0 + 2\pi)$. Let

$$x_1 = \nu_0 \cos \mu_0, \quad x_2 = \nu_0 \sin \mu_0$$

(18)
where $r_0 \in [0, 0.495R]$ and $\mu_0 \in [0, 2\pi]$. We have
\[
\cos \frac{\Theta_s - \Theta_0}{2} = \frac{r_0 \sin (\mu_0 - \Theta_s)}{\sqrt{R^2 + r_0^2 - 2R_0 \cos (\mu_0 - \Theta_s)}}. \tag{19}
\]

Then, (19) can be rewritten as
\[
\cos \frac{\Theta_s - \Theta_0}{2} = \frac{\sin (\mu_0 - \Theta_s)}{\sqrt{\left(\frac{B}{r_0} - \cos (\mu_0 - \Theta_s)\right)^2 + \sin^2 (\mu_0 - \Theta_s)}}. \tag{20}
\]

When $\mu_0 - \Theta_s$ is fixed and $r_0$ is reduced, $[\cos(1/2)(\Theta_s - \Theta_0)]$ decreases. Therefore, the right side of (20) reaches its maximum or minimum when $r_0$ is maximized, i.e., at $r_0 = 0.495R$. Those maximum and minimum values can be numerically calculated [Fig. 19(a)], and we have
\[
-0.4949 \leq \cos \frac{1}{2} (\Theta_s - \Theta_0) \leq 0.4949
\]
or
\[
2.1062 \leq \Theta_s - \Theta_0 \leq 4.1773. \tag{21}
\]

Next we prove that $0 \leq \Theta_s - \Theta_0 \leq 4.1773$ implies
\[
v_1(\Theta_s - \Theta_0) - v_2(\Theta_0 - \Theta_s) > 0 \tag{22}
\]

where $v_1(s) = (Dh)/(2\pi R)(s - 4\pi/3)/(1 - \cos s)$ and $v_2(s) = (Dh)/(2\pi R)(s - 4\pi/3)/(1 - \cos s)$. Fig. 19(b) shows the function $\Phi(s) = \cos s - (s - 4/3\pi) \sin s$ in the range $s \in (0, 2\pi)$, demonstrating that $v_1(s)$ is always positive. Hence $v_1(\Theta_s - \Theta_0)$ is monotonically increasing. Let us fix $\Theta_s = \Theta_0 + 2.1062$ and plot the function $\Psi(\Theta_0) = v_1(\Theta_s - \Theta_0) - v_2(\Theta_0 - \Theta_s)$ [Fig. 19(c)]. Clearly, this function is always positive in the range $0 \leq \Theta_0 - \Theta_s \leq 4.1773$. Note that $s = 4.1773 + \Theta_0$ is the intersection of $\Gamma_1$ and $\Gamma_1^+$. $\Theta_s$ cannot be to the left of this point (otherwise, $\hat{x}$ is outside $G$).

Equation (22) indicates that $\Theta_s$ is always lower than $\Theta_0$ in the vertical direction. Since $\Theta_0$ is to the right of $\Theta_s$, the slope of the inter-PI line $\frac{\Theta_s - \Theta_0}{2}$ is always negative. Due to symmetry, the other two cases $s \in (\Theta_s, \Theta_0)$ and $s \in (\Theta_0, \Theta_s)$ can be handled similarly. This finishes the proof.

**Lemma 2:** A $T$-curve cannot be tangent to an $A$-curve at an interior point of $T$.

**Proof:** The interior point of $T$-curve is any point of the $T$-curve except an endpoint. It has been proved in [17] that a $T$-curve is smooth everywhere, except possibly at a cusp. Let $s_i$, $i \in \{1, 2, 3\}$ be the point where the cusp occurs. Suppose a $T$-curve is tangent to an $A$-curve at $\alpha(s)$. If $s = s_i \in (s_i, s_i')$, where $s_i$ and $s_i'$ are the endpoints of the $T$-curve, then the osculating plane $\Pi(x)$ intersects the helix $y_i(s)$ at only one point and it contains one inter-PI line. By construction, $\Pi(x)$ intersects the detector plane at the asymptote of the Tam-Danielson window boundary and $\hat{x}$ belongs to the asymptote. Connecting $\hat{x}$ and $s_i$, $\hat{x}$ and $s_i'$ we get two inter-PI lines. Clearly, $\Pi(x)$ will not contain any of them. By Lemma 1, the third inter-PI line has a negative slope, thus it will not overlap the asymptote. Hence, $\Pi(x)$ will not contain it. Consequently, $s \neq s_i$ and $T$-curve is smooth in a neighborhood of $\alpha(s)$. Let $\theta$ be the polar angle for the great circle $(x - y_i(s'))^2 + s_i'^2 \in (s_i, s_i')$, $i \in \{1, 2, 3\}$. Then, the $A$-curve consists of all the unit vectors $\alpha \theta(x) \in (x - y_i(s'))^2$. Clearly, $\alpha \theta(\theta)$ is perpendicular to $\alpha \theta(\theta)$ and $(x - y_i(s'))$. By construction, the $T$-curve is tangent to the $A$-curve at $\alpha(s)$. Hence, $\alpha \theta(\theta)$ is parallel to $\alpha \theta(\theta)$. That is, $\alpha \theta(s)$ is parallel to $\alpha \theta(\theta)$ and $(x - y_i(s'))$. Because $\alpha \theta(s)$ is also parallel to $(x - y_i(s'))^2$, $s_i \in (s_i, s_i')$, $(x - y_i(s'))^2$ is parallel to $(x - y_i(s'))$ and $s_i' \leq s_i$, which contradicts the assumption that $s_i$ is an inner point of the $T$-curve. This finishes the proof.

**Lemma 3:** Case 1, 2, 4, 6, 7 do not occur for $r < 0.495R$ and Case 5 does not occur for $r < 0.265R$.

**Proof:** Cases 1 and 2 are impossible because they mean that there will be a plane containing three inter-PI lines or tangent to three inter-PI arcs.

In Case 4, there will be one plane $\Pi$ containing one inter-PI line and tangent to two inter-PI arcs. Assume this inter-PI line is $s_i/s_i'$, let us pick a point $s = s_i$ on $y_i(s)$ and denote $L = \Pi \cap D(s_i)$. By construction, on the detector plane $\hat{x}$ is on $\Gamma^+$ and it overlaps $s_i$. Then, $L$ may be tangent to $\Gamma^+$ and $\Gamma^+$ or $\Gamma^+$ and $\Gamma^-$, see Fig. 20. Because the points of tangency are on the inter-PI arcs, the endpoint $s_i'$ is to the left of the tangency for case A and $s_i$ is to the right of the tangency for case B. Connecting $s_i$ and $\hat{x}$ (or $s_i'$ and $\hat{x}$) we find that the slope of the inter-PI line $\frac{s_i - s_i'}{2}$ is negative. By Lemma 1, these two cases are impossible.

In Case 5, there will be one plane containing one inter-PI line and tangent to one inter-PI arc at the inflection point. Thus, the inter-PI line on the detector overlaps the inflection line, i.e., $\overline{s_i/s_i'}$ overlaps $L_{s_i}$ when $s_i = s_{it}$ and $s_i' = 2s_{it}$, where $s_{it}$ is difference.
between $s_2^3$ and $s_2^5$. Let us look at (20). When $\mu_0 - s_2^3$ is fixed, the absolute value of $\cos(s_3^2 - s_2^3)/(2)$ is monotonically decreasing when $r_0$ is reduced. Then, the range of $s_2^3 - s_2^5$ is narrowed. In other words, the difference between $s_2^3$ and $s_2^5$ becomes closer to $\pi$. Case 5 occurs when $s_2^3 - s_2^5 = s_{u}$. If we want to exclude Case 5, the range of $s_2^3 - s_2^5$ cannot cover the value $s_{u} = 2.0053$. Hence, the minimum range of $s_2^3 - s_2^5$ is $2.0053 < s_2^3 - s_2^5 < 2\pi - 2.0053$. That is, $-0.2649 < \cos(s_2^3 - s_2^5)/(2) < 0.2649$ and $\cos(s_3^2 - s_2^5)/(2)$ reaches its extreme when $r_0 = 0.265R$. From (20) and (21) we have $s_2^3 - s_2^5 = s_{u}$ only for $r \geq 0.265R$, which contradicts our condition. Hence, Case 5 is impossible for $r < 0.265R$.

In Case 6, a $T$-curve will intersect one $A$-curve twice before meeting a cusp. Suppose this takes place at inter-PI arc $s_1^3s_2^5$. Let us pick a point $s = s_1^3$ on $y_l(s)$ and see what happens on the detector plane when $s$ moves. By construction, the plane $\Pi$ containing $y_l(s)$ and $x$ intersects the detector plane at the line $L$ which is parallel to the helix tangent across $\hat{x}$. At $s = s_1^3$, $\hat{x}$ is on $\Gamma^{+1}$ and $\Pi$ contains inter-PI line $s_1^3s_2^5$. As $s$ moves along $y_l(s)$, $\hat{x}$ moves downwards. Notice that $L$ is parallel to the asymptote of the Tam–Dannielson window, so it will not intersect $\Gamma^{-2}$ provided that $\hat{x}$ moves across the asymptote, at which the cusp occurs. Hence, $n$ will not contain the inter-PI line $s_1^3s_2^5$ and Case 6 is impossible.

By Lemma 2, Case 7 is impossible. This finishes the proof.

Let $G_{21}$ be the area bounded by $L_{4\hat{x}}$, $\Gamma'_{l}$ and the line $v = Dh(s_{u} + (4)/(3\pi))/(2\pi R(1 - \cos s_{u}))$ (i.e., the line across $s_{u}$ and parallel to the $u$-axis) and $G_{22}$ the area bounded by $L_{4\hat{x}}$, $\Gamma'_{r}$ and the line $v = Dh(s_{u} - (4)/(3\pi))/(2\pi R(1 - \cos s_{u}))$ (i.e., the line across $s_{u}$ and parallel to the $u$-axis, see Fig. 21). $\Gamma'_{l}$ and $\Gamma'_{r}$ are the vertical lines across $\hat{x}$ and $s_{u}$. They correspond to the boundary condition $r_0 = 0.265R$.

Lemma 4: The inflection point $s_{u}(\hat{x})$ is inside the inter-PI arc when $\hat{x}$ is in $G_{22}(G_{22})$.

Proof: By Lemma 3, any point in the area $r < 0.265R$ has three cusps in the diagram. Note that there is one IP in each inter-PI arc within $D_{\hat{x}}$. Since all three cusps are in $D_{\hat{x}}$, an oscillating plane of one inter-PI arc intersects two other inter-PI arcs exactly once at one point. Assume that this oscillating plane $\Pi_{c}$ contains $x$. Consider $\Pi_{c}$ of the second inter-PI arc (i.e., of $y_2(s)$). Let $s_1^0$ be the point where it intersects the first inter-PI arc (i.e., on $y_1(s)$). Let us move $s$ along the first inter-PI arc and observe what happens with $\hat{x}$ on the detector. When $s = s_1^0$, $\hat{x}$ enters the Zhao window through $\Gamma_{+1}$. When $s = s_0^0$, $\hat{x}$ belongs to $L_{4\hat{x}}$. As follows from the diagram, the point $s_{u}$ must be inside the second inter-PI arc, i.e., between $s_2^3$ and $s_2^5$. As the point $s$ moves further, the difference $s_2^3 - s$ becomes smaller, and the point $s_2^5$ moves to the right of $s_{u}$ along $\Gamma_{+1}$. The inter-PI line $s_1^3s_2^5$ has a positive slope. Thus, as long as $\hat{x}$ is inside $G_{21}$, the point $s_2^5$ is always to the left of $s_{u}$. The case where $\hat{x}$ is in $G_{22}$ can be similarly treated. This proves Lemma 4.

Lemma 5: An $L$-curve never intersects an $A$-curve for $r < 0.265R$.

Proof: If an $L$- and $A$-curves intersect, an $L$-line through $\hat{x}$ overlaps the inter-PI line. Let us pick one point on the inter-PI arc $s_3^2s_5^3$ and consider the slope of the inter-PI line on the corresponding detector plane. By construction, an $L$-curve always starts from a cusp of a $T$-curve and ends on a cusp of another $T$-curve. For the oscillating plane $\Pi_{c}$, its intersection with the detector plane is the line tangent to $\Gamma_{+1}^{1}/(\Gamma_{-1}^{1})$ at $s_{d}(\hat{x}_{d})$. By Lemma 4, the endpoints of the inter-PI arc on $\Gamma_{+1}^{1}/(\Gamma_{-1}^{1})$ are on different sides of $s_{d}(\hat{x}_{d})$, and one of them on the right (left) side is also an endpoint of the inter-PI line for helices $y_2(s)$ and $y_3(s)$, and denoted as $s_1^0(s_1^0)$ in Fig. 22.

If $\hat{x}$ is in $G_{21}$ and above $s_{d}$, we can have the $L$-line by connecting $\hat{x}$, $s_{d}$, and the inter-PI line by connecting $\hat{x}$, $s_1^0$ and the $L$-line by connecting $\hat{x}$, $s_2^5$. Clearly, in any case the slope of $L$-line is between zero and the slope of the inter-PI line. That is, the $L$-line will not overlap the inter-PI line. For other $\hat{x}$, the $L$-line is parallel to the $u$ axis. By Lemma 1, it is always between two inter-PI lines and will not overlap with any of them.

For the point on other inter-PI arcs, the situation is the same. This finishes the proof.

B. Derivation of Formulae in Section VI

The angles in the detector plane are not equivalent to that in the spherical coordinate system because the detector plane is parallel to the $\varphi_{3}$-axis while the plane in spherical coordinates
When and reaches its and inflection line and above between lines and its minimum 0 one point in denote the source position, and is perpendicular to the line at.

By construction

\[
\begin{align*}
(P_1 - A) \cdot (S - A) &= 0 \\
(P_1 - S) &= c_1(P_1 - B) \\
(P_2 - A) \cdot (S - A) &= 0 \\
(P_2 - S) &= c_2(P_2 - C)
\end{align*}
\]  

where \( B = (d + d\cos\alpha, d\sin\alpha, 0), C = (d + d\cos(\alpha + \theta), d\sin(\alpha + \theta), 0) \), \( c_1 \) and \( c_2 \) are two nonzero constants.

Solving (23) and (24), we have

\[
P_1 = \frac{d(R^2 + d^2)(1 + \cos\alpha)}{R^2 + d^2(1 + \cos\alpha)}
\]

\[
P_2 = \frac{d(R^2 + d^2)(1 + \cos(\alpha + \theta))}{R^2 + d^2(1 + \cos(\alpha + \theta))}
\]

\[
\cos\varphi = \frac{(P_1 - A) \cdot (P_2 - A)}{\sqrt{(P_1 - A) \cdot (P_2 - A)}}
\]

\[
= \frac{\cos\alpha \cos(\alpha + \theta) + (1 + k^2)\sin\alpha\sin(\alpha + \theta)}{\sqrt{1 + k^2\sin^2\alpha} \sqrt{1 + k^2\sin^2(\alpha + \theta)}}
\]

where \( k = \frac{d}{R} \).

Computing the derivative of (27) we find that \( \varphi \) reaches its maximum at \( \alpha = -\theta/2 \) or \( \alpha = -\theta/2 + \pi \). In this case, (27) becomes

\[
\cos\varphi = \frac{\cos^2\frac{\theta}{2} - (1 + k^2)\sin^2\frac{\theta}{2}}{\cos^2\frac{\theta}{2} + (1 + k^2)\sin^2\frac{\theta}{2}}
\]

For a fixed \( \theta \), \( \cos\varphi \) is monotonically decreasing while \( k \) is increased. Since \( \tau < 0.495R/k \), \( \varphi \) reaches its maximum \( 1/\sqrt{3} \) when the line \( OX \) is perpendicular to the line \( SA \) and its minimum 0 when the line \( OX \) is parallel to the line \( OS \). Hence,

\[
\theta < \varphi < \arccos \left( \frac{3 - 4\tan^2\frac{\theta}{2}}{3 + 4\tan^2\frac{\theta}{2}} \right)
\]

or

\[
\theta < \varphi < \arctan \left( \frac{2}{\sqrt{3}} \tan \frac{\theta}{2} \right).
\]

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their comments and suggestions.

REFERENCES


