A 4-bank multileaf collimator provides a decomposition advantage for delivering intensity-modulated beams by step-and-shoot

S. Webb

Joint Department of Physics, Institute of Cancer Research and Royal Marsden NHS Foundation Trust, Downs Road, Sutton, Surrey, SM2 5PT, UK

Received 12 May 2011; received in revised form 10 August 2011; accepted 24 August 2011
Available online 9 September 2011

Abstract During the delivery of intensity-modulated radiation therapy (IMRT) by the multileaf collimator (MLC) step-and-shoot technique, it is required to sequence the two-dimensional intensity-modulated beams (2D IMBs) into a series of components that can be shaped with the MLC. In this paper it is shown that the use of an MLC that has four banks of leaves (in two orthogonal pairs) instead of two banks only is advantageous for this purpose.

Introduction

Most commercially-available multileaf collimators (MLCs) have two banks of leaves, each bank having an equal number of leaves so that they may be viewed as a set of paired “left” with “right” leaves [1]. This is interesting given that the very first patent for an MLC actually had four banks of leaves, a “left-right” set and a “top-bottom” set [2] (see also [3]). We might discount this latter as something of a historical artefact since, apart from a crude implementation by Toshiba in which the “leaves” were more the size of “blocks”, nothing came of it and the radiotherapy community had to wait for the first commercial MLC from Brahma [4] and Scanditronix in the mid 1980s, since when, with just the one exception now to be mentioned, all MLCs have had just 2 banks of leaves. Indeed it is these that have been pressed into service in many ways to deliver IMRT [5–8].

A commercial MLC has been developed by the Initia/Azimutherapy Company [9] that has four banks of MLC leaves, two operating “left-right” and the other pair operating at 90° “top-bottom”. Leaf sets can operate independently. It is emphasised that the two banks are one beneath the other and so the operation of one bank in no way interferes with the other. No orthogonal leaf clashes can occur. For the purposes of this preliminary study all the leaf widths are regarded as the same and equal to the bixel size of the 2D IMB to be decomposed. The advertised advantage relates to two main features: (1) when there is...
a requirement to collimate a shape that has a concave outline then the "top-bottom" leaves can be operated to reduce the unwanted irradiation that would arise with the collimation from just a "left-right" set; (2) when the goal is to track a moving target the use of the "top-bottom" leaves can eliminate the need for the outermost "left-right" leaves to exceed their maximum physical speed. This would happen when the target outline moved "up-down" such as to invoke the opening up or the closure of those outer "left-right" leaves [10]. Both these advantages were explained graphically in a promotional video released by the Company in 2006 but it is not known how much advantage was taken of these possibilities in practice.

In this paper a possible third advantage for the 4-bank MLC is explained, namely that it can reduce both the number of components and the delivered monitor units (MU) when IMRT is delivered in the step-and-shoot mode i.e. when the task is set to "sequence" the delivery of a two-dimensional (2D) intensity-modulated beam (IMB) the task can be performed more advantageously with a 4-bank than with a 2-bank MLC. It might be (correctly) regarded that this is self-evident and does not need proving. However, the purpose of this paper is to give some quantitative data on the magnitude of the improvement and also to comment more on 2D IMB sequencing.

Methods

Let us state that the computer-modelling task is the following: given a 2D IMB of square dimensions \(N \times N\) decompose this into a "best set" of components that in turn sum to give the original 2D IMB. This task is often described as "sequencing" or "component stripping", the latter recognising that the process generally takes place by sequential stripping of some "best-chosen component" from "configuration options" to create at each "strip event" a "residual" and then this residual is sequenced in turn until the final residual is the empty matrix. It is assumed that the elements of the original 2D IMB are integers and the sequencing is into integer-valued components. We might write that each element \(V(i, j)\) in the 2D IMB is given by \(V(i, j) = \sum_{k=1}^{K} M_{Uk} \times \delta(i, j)\) where there are \(K\) components each with unique value \(M_{Uk}\) and \(\delta(i, j)\) is unity if the element in that component is open (i.e. irradiated) and zero if, conversely, it is closed (i.e. shielded by leaves). The MU-efficiency of this process is given by \(\varepsilon = \max[V(i, j)]/\sum_{k=1}^{K} M_{Uk}\) where \(\varepsilon\) is the largest element in the 2D IMB. \(\varepsilon\) will always be less than or equal to unity. It might be stated that a "good" step-and-shoot delivery would be one with as few components \(K\) and as high MU-efficiency \(\varepsilon\) as possible because the total irradiation would be minimised and the speed of delivery would be maximised.

Now let us proceed to how to compute the decomposition. Immediately we encounter a computational difficulty of the kind sometimes referred to as a computational explosion. Consider sequencing just one row of the \(N \times N\) matrix using a 2-bank MLC. To do this the left and right leaves of a pair are set (in turn) at each of the locations they may occupy, the right leaf clearly having to be ahead of the left leaf. Then the elements that are uncovered are inspected. If there are no exposed zeroes then the configuration is a physical possibility. We can compute the minimum value of those exposed elements and that becomes a candidate component at this cycle of the component stripping. Even so a choice between candidate components (configuration options) has to be invoked and a common one is to strip the component that has the largest product of number of exposed elements and minimum value of those exposed elements (largest area of strip). Already this has introduced an element of subjectivity into the sequencing because other choices could have been made (e.g. to strip the largest number of elements or to strip just the largest-valued element and so on.) The computational explosion arises as follows. There are \(C(N) = \sum_{i=1}^{N} i + 1 = [(N(N + 1)/2) + 1]\) ways to set the leaf pair and that require the above analysis (the \(+1\) is to allow a completely shielded row). However, in stripping a \(N \times N\) matrix all the rows are considered as an ensemble. Hence there are \(C^N(N)\) options for leaf-component setting when using just a 2-bank MLC. If the 2-bank MLC is also considered to be rotatable by 90° then the whole process can be repeated over column stripping leading to the requirement to investigate \(2C^N(N)\) configuration options. This would also be the number of options if the 4-bank MLC were to be used but with the 2 orthogonal sets of banks operating separately, first one set then the other set. Now imagine we sequence with the 4-bank MLC using the four banks simultaneously and independently to define components. Then the number of possible component geometries is \(C^N(N)\).

Ideally the study we wish to perform here would like to have been able to investigate all of these options, something we refer to as a "complete" search. Apart from the parameters for the choice of component—and maximum "area stripped" makes sense and was used throughout here—the computer search would then guarantee the least number of components and the highest MU-efficiency.

We should like to make a comparison between the component decomposition of any given 2D IMB via the three methods: (i) 2-banks used "left-right" only; (ii) 2-banks used separately in "left-right" and "top-bottom" mode (equivalent to 4-banks being used with the pairs independently) and (iii) 4-banks used simultaneously. For this to be a fair comparison the same conditions must hold for each. In this study we sequenced a large number \(N_M\) of 2D IMBs of maximum elemental value \(M\) and with the elements randomly set as integers in the range 1 to \(M\). This was done for \(M = 1, 2, 3, \ldots 50\) in turn. Then it was possible to report the mean number \(\langle K(M)\rangle\) of components and the mean number \(\langle \sum_{k=1}^{K} M_{Uk}\rangle\) of MUs required and the mean efficiency \(\langle\varepsilon(M)\rangle\) as a function of \(M\).

When \(N = 3\) we see that \(C^N(N)=7^6\) which is easily manageable. So we can report the results of a "complete search" for each of the three collimation methods using \(N_M=10^4\). Indeed gathering such results took seconds. However when \(N = 6\) we see that \(C^N(N)=22^{12}\) and to execute this complete search is totally impossible. Furthermore imagine \(N = 10\) then \(C^N(10)=56^{20}\)! Recall that this number of candidate configurations has to be examined at each cyclical component strip stage. This is the "computational explosion". So, how can this be overcome? We want to decompose \(N \times N = 6 \times 6\) matrices. It was necessary, when \(N = 6\), to resort to a kind of Monte Carlo limited sampling as follows:
For collimation methods (i) and (ii) above (the use of just a 2-bank MLC with or without rotation through 90°) a candidate component was created by setting the left and the right leaves at randomly selected locations (right leaves ahead of left leaves of course) and $10^6$ candidates were considered and the “best” component was selected as that with the largest stripped area. Typically it now took about 10 s to find this component. However, repeating a preliminary study (just one $M$ value and just one such example) with, instead, $10^7$ randomly selected candidates taking now 100 s per cycle, then a “slightly better” component emerged i.e. one with a slightly larger area. Repeating this preliminary study with, instead, $10^8$ randomly selected candidates taking now 1000 s per cycle, then a marginally improved “slightly better” component emerged i.e. one with a marginally slightly larger area. However, one cannot make computer strips with $10^7$ or $10^8$ candidate configurations (or more) because the total computer time over large ensembles of 2D IMBs would stretch to months and years. Some “trick” has to be invoked as follows. Maintaining the number of candidate searches as $10^6$ the best was selected. Then this best component was challenged by retracting the left leaves either one or two elements and advancing the right leaves the same. A retraction or advance of just one leaf was made each time. If this generated a better component (one with larger total stripped area) then this became the running best component. This requires just a further $4N$ configuration options to be investigated (including the “top-bottom” equivalent adjustments) and, whilst the code looks considerably complex, the execution time is very rapid, indeed negligible, compared with the main search over the initial $10^6$ candidates.

For collimation method (iii) (the use of 4-banks simultaneously) firstly a candidate component was created by setting the left, right, top and bottom leaves at randomly selected locations (right leaves ahead of left leaves of course; bottom leaves ahead of top leaves) and $10^6$ candidates were considered and the “best” component was selected as that with the largest stripped area. A different “trick” was then required for the subsequent refining stage. It is now not possible to consider sequential retractions of left and top leaves and advances of right and bottom leaves because of the complex interaction between the collimation provided by the orthogonal banks. Instead the following was performed. A subset $Q$ ($Q = 4$ was found best) of the leaves were randomly selected to be retracted or advanced by any random integer number of bixels between 1 and the number that would take that leaf to the matrix boundary. Again the newly defined candidate was tested against the best-so-far from the initial search over $10^6$ candidates. This second-stage “tweaking” was performed $10^6$ times.

It is clearly acknowledged that the steps outlined above are computer recipes and in that sense somewhat arbitrary. It has already been explained that a complete search cannot be sensibly made beyond about $N = 3$. Do these tricks work? The way it was established that they do was by tedious and time consuming manual computer runs in which the result of each stage was visualised by running the codes with multiple screen displays and pauses. One could see the best components from the initial searches being improved by the second-stage processes. Finally, this satisfied, automatic batch runs were set running on a Dell Optiplex 755 for $M = 1, 2, \ldots 50$ and $N_M = 600$. Typically each decomposition took of the order 10 days.

Results

Figures 1 and 2 show the average outcomes for methods (i), (ii) and (iii) of collimation when $N = 3$. Figures 3 and 4 show the average outcomes for methods (i), (ii) and (iii) of collimation when $N = 6$.

These Figures largely speak for themselves. We can see that whatever the size $N$ and whatever the decomposition technique ((i), (ii) or (iii)) both the mean number of components and the mean number of MUs increase as the peak 2D IMB matrix value $M$ increases. That is entirely as we should expect. It is also apparent that for any given peak matrix value $M$ and for any specific decomposition method both the mean number of components and the mean number of MUs increase as the dimensionality $N$ of the matrix size increases. Again that is as we should expect. All the lines showing the increase of the mean number of MUs as the peak value $M$ increases are almost exactly straight lines whatever are the matrix dimensions $N$ and for all decomposition methods. Again that is what we should expect. Conversely, the lines depicting the mean number of components are more gentle shallow curves with the curvature flattening out as the peak value $M$ increases.

It can be immediately seen that the use of 4 banks independently is better than the use of 2 banks alone. Moreover the use of 4 banks simultaneously is better than
the use of 4 banks independently. The advantage is greater at \( N = 6 \) than at \( N = 3 \). Some specific values are extracted to Table 1.

Discussion

Traditionally 2D IMBs have been sequenced by somewhat heuristic methods. For example one way is to perform a decomposition in which all leaf pairs of a 2-bank MLC move from left to right. An alternative is the so-called “close-in” method. Others have attempted simple power-of-two stripping. These choices were forced in the 1990s because it was recognised that the number of component options was huge and could not be completely searched (see extensive analyses by Webb [11,12]) in realistic computer times. Conversely component stripping using jaws-only is relatively straightforward to code because this is as if the “leaves” all shape up in concert creating rectangular shapes.

In this paper we have explicitly explained why complete computer decompositions are generally impossible for realistically-sized matrices and have shown how carefully adjusted stripping menus can create components. It has been shown that the use of 4 banks simultaneously is better than the use of 4 banks independently and 2-banks alone.

The comparison has been deliberately applied to matrices of random integers. This is a well-known and accepted benchmark test. Of course it will generate the largest difference between the three methods and so represent a measure of maximum gain. Conversely, if the 2D IMBs were all uniform with value \( M \) then every method would require just one component; efficiency would be unity and the mean number of MUs would exactly equal the peak value, i.e. the line of \( \sum_{k=1}^{K} MU_k(M) \) versus peak \( M \) would be at 45° to the axes. Further very lengthy “runs” (not reported here) showed that as the range of values decreased so this situation was gradually approached.

The Initia/Azimuththerapy 4-bank MLC has variable sized leaf widths but this has not been considered in this preliminary study of the decomposition of small 2D IMBs where the central same-size leaves would be used.

We may enquire how much the use of the 4-bank MLC in fully coupled mode (decomposition method (iii)) improves the treatment delivery. The answer is that there is zero change to the patient dosimetry. If we imagine a series of 2D IMBs irradiating a patient then these exact same 2D IMBs are delivered to the patient whatever method of delivery decomposition is used. That is at the heart of the study. What changes is the MU-efficiency and the number of decomposition components. These two quantities do depend on the decomposition method as shown in Figs. 1–4 and Table 1. The quantitative magnitude of the improvement depends on the precise form and number of the 2D IMBs being delivered. Hypothetically let us suppose that say a number \( N_{IMB} \) of 2D IMBs are to be delivered, each being on a matrix size of \( N \times N = 6 \times 6 \) and with a peak value in each of 50 MU. Then changing from decomposition method (i) (the worst) to decomposition method (iii) (the best) would lead to an improved MU-efficiency of 19.7% irrespective of the number \( N_{IMB} \) (see also Table 1). At the same time the

![Figure 2](image)

**Figure 2** The mean number \( \langle K(M) \rangle \) of components and the mean number \( \langle \sum_{k=1}^{K} MU_k(M) \rangle \) of MUs required and the mean efficiency \( \langle e(M) \rangle \) as a function of \( M \) via decomposition method (iii) [blue, black and red lines respectively] when \( N = 3 \) and \( N_M = 10^4 \) (complete search).

![Figure 3](image)

**Figure 3** The mean number \( \langle K(M) \rangle \) of components and the mean number \( \langle \sum_{k=1}^{K} MU_k(M) \rangle \) of MUs required and the mean efficiency \( e(M) \) as a function of \( M \) via decomposition methods (i) [blue, black and red lines respectively] and (ii) [maroon, green and yellow lines respectively] when \( N = 6 \) and \( N_M = 600 \) (Monte Carlo search and second-stage tweaking).
improvement in the number of field components (and thus improvement in delivery time) is 33.8%. Of course in practice the number of fields will vary, the size of those fields will vary and the 2D IMBs will not be random. For any given clinical situation the exact quantitative improvement could only be determined by decomposing the 2D IMBs appropriate for that clinical situation. It is considered to always be clinically relevant to decrease the number of MUs delivered (increase the MU-efficiency) and to reduce the number of components because that leads to a lower leakage bath and a faster treatment.

The issue of characterising fluence-map decomposition in terms of the mean number of MUs and mean number of components is a very old one that has been extensively studied. Let us put the current study in this context. Firstly there is the famous multiple-static-field decomposition method of Bortfeld et al. [13]. This is well-known to generate the solution with the minimum mean number of MUs. Ahuja and Hamacher [14] used a network-flow algorithm to prove that the Bortfeld et al sweep algorithm is the most MU efficient. Indeed an expression for this was derived in detail by Webb [15] and used as a benchmark [16]. The mean number \( \langle MU_{B-B} \rangle \) of MUs is re-quoted here as

\[
\langle MU_{B-B} \rangle = \frac{(N-1)}{2} \times \left( \frac{M^2 - 1}{3M} \right) + \left( \frac{M}{2} \right)
\]  

From this we may deduce that (e.g.) when \( N = 6 \) and \( M = 50 \), \( \langle MU_{B-B} \rangle = 66.6 \) which is somewhat less than the value given by method (i) in Table 1 or Fig. 3. It is also less than the value from method (iii). However the first two terms in equation (1) give approximately the number of field components and come out at 41, somewhat greater than that of all the methods studied here. This illustrates the well-known conclusions that whilst the Bortfeld method gives the optimum (minimum) mean number of MUs it does not give the optimum (minimum) number of field components. This was known as long ago as 1998 when Xia and Verhey [17] made plots showing a comparison of decomposition algorithms (including that of Galvin et al. [18]) and by Que [19] for eight methods of field decomposition. They applied these to both random and clinical cases. The Bortfeld method often came out with a much larger number of components than its rivals. This is widely discussed again in a review by Webb [20]. We should also note that when comparing field-decomposition techniques a concern is whether machine constraints (e.g. lack of the possibility to interdigitate and also the tongue-and-groove effect) are considered or overlooked. In the study in this paper it is assumed that interdigitation can occur and the TG effect is ignored as it is now known that this problem was somewhat over-exaggerated in the 1990s. The proof of Ahuja and Hamacher [14] ignored the MLC machine constraints and also ignored the optimisation of the time taken to reset components (as does our present study).

Very recently Caner Taskin et al. [21] have given an exact algorithm for minimising both the mean number of MUs and the mean number of components and applied it to the case of a 2-bank-only MLC with and without 90\(^\circ\) rotation. It may be that such a method could be extended to the 4-bank MLC considered here but that is for the future. It is likely that the conclusions of the current paper are not strictly dependent on the choice of sequencing algorithm and the current study bears out the intuitive hypothesis that a 4-bank MLC gives some decomposition advantage over a 2-bank MLC.

![Figure 4](image-url)

**Figure 4** The mean number \( \langle K(M) \rangle \) of components and the mean number \( \langle \sum_{i=1}^{K} MU_{i}(M) \rangle \) of MUs required and the mean efficiency \( \langle \epsilon(M) \rangle \) as a function of \( M \) via decomposition method (iii) [blue, black and red lines respectively] when \( N = 6 \) and \( N_M = 600 \) (Monte Carlo search and second-stage tweaking).

<table>
<thead>
<tr>
<th>Matrix size</th>
<th>Peak value M in random matrix</th>
<th>Decomposition method</th>
<th>Mean number of components</th>
<th>Mean number of MU</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>50</td>
<td>(i)</td>
<td>6.484</td>
<td>65.058</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>(ii)</td>
<td>6.010</td>
<td>57.550</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>(iii)</td>
<td>5.879</td>
<td>56.058</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>(i)</td>
<td>16.835</td>
<td>107.970</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>(ii)</td>
<td>14.387</td>
<td>95.435</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>(iii)</td>
<td>11.152</td>
<td>86.738</td>
</tr>
</tbody>
</table>
Conclusion

The 4-bank MLC outperforms the 2-bank MLC when used for component determination in IMRT.

Acknowledgements

The ICR and RMH are supported by the NIHR. I am grateful to Dr Gavin Poludniowski for interest in this work and for suggesting the Monte Carlo approach. I should like to thank an anonymous referee for bringing some less-mainstream radiotherapy-physics literature to my attention and for comments that, in responding, have improved the quality of this paper, specifically contextually.

References