A New Approximation Algorithm for Matrix Partitioning in Presence of Strongly Heterogeneous Processors

Olivier Beaumont, Lionel Eyraud-Dubois, Thomas Lambert

A New Approximation Algorithm for Matrix Partitioning in Presence of Strongly Heterogeneous Processors

Olivier Beaumont[*], Lionel Eyraud-Dubois[†] and Thomas Lambert[‡]

Inria
University of Bordeaux
Talence, France
[*]olivier.beaumont@inria.fr, [†]lionel.eyraud-dubois@inria.fr, [‡]thomas.lambert@inria.fr

Abstract—In this paper, we consider the problem of partitioning a square into a set of zones of prescribed areas, while minimizing the overall size of their projections onto horizontal and vertical axes. This problem typically arises when considering the amount of communications induced when partitioning matrices for dense linear algebra kernels onto a set of heterogeneous processors. It has been first introduced for matrix multiplication in the 2000's, with a best known approximation ratio was 1.75. Since then, two main new ingredients have been introduced. First, Lastovetsky et al. proposed a special partitioning in the case of 2 or 3 strongly heterogeneous processors, as in the case of a platform made of CPUs and GPUs, relaxing the constraint of a rectangular based partitioning. Second, Nagamochi et al. have introduced clever recursive partitioning techniques and proved, thanks to a careful analysis, that their algorithm achieves a 1.25 approximation ratio. In this paper, we combine both ingredients in order to obtain a non-rectangular recursive partitioning (NRRP), whose approximation ratio is $\frac{2}{\sqrt{3}} \simeq 1.15$. Moreover, we observe on a large set of realistic platforms built from CPUs and GPUs that this proposed NRRP algorithm allows to achieve very efficient partitionings on all considered cases.

Index Terms—

I. INTRODUCTION

The problem of partitioning a matrix into a set of sub matrices has received a lot of attention in the last few years. This operation is indeed crucial when considering dense linear algebra kernels on heterogeneous platforms. Let us for instance consider dense matrix multiplication based on Canon's-like algorithm, restricted for the sake of simplicity to the multiplication $C = AB$ of two square $n \times n$ matrices A and B. Let us further assume that the matrices are partitioned into blocks, whose size is chosen so as to be well adapted to all types of resources (typically CPUs and GPUs). Then, at step k of the algorithm, the outer product of the k-th column of blocks of A and the k-th row of blocks of B is computed. Let us assume that processor P holds a set of s blocks whose projections along the different axis have respective size h and w. Then, the volume of computations P needs to perform is proportional to s and the volume of communications is proportional to $h+w$. In order to balance the computing load, each processor should receive a number of blocks proportional to its relative speed. In turn, the overall volume of communications is proportional to the sum of the projections of the areas owned by the different processors along the axes. Therefore, in order to minimize the processing time while minimizing the overall volume

of communication, the optimization problem is amenable to the problem of partitioning a square into a set of zones of prescribed area (in order to balance the load) such that the sum of the projections along the two axes is minimized (in order to minimize the communications).

Related Works

This optimization problem has been first introduced by Lastovetsky and Kalinov in [1]. In [2], it has been proven that the problem is NP-Complete, and a first approximation algorithm with bounded ratio (1.75) has been proposed. This algorithm has been improved along two directions. On the one hand, Lastovetsky et al. have proposed to relax the assumption stating that the zones allocated to the processors should consist in a single rectangle and have proposed optimal algorithms, but limited to 2 processors [3] and more recently to 3 processors [4]. On the other hand, recursive partitioning algorithms have recently been proposed, in which at each step, the set of processors is split into two parts. Sophisticated proof techniques enabled Nagamochi and Abe [5] to improve the approximation ratio down to 1.25. Recently, Fügenschuh et al. [6] improved this result to 1.15, but under the assumption that if we consider processors in decreasing order of their processing speeds, there is no abrupt change in the performance between 2 successive processors. Unfortunately, such an abrupt decrease typically happens when considering nodes consisting of CPUs and GPUs, such that Fügenschuh's algorithm is limited to the case of relatively homogeneous platforms. In this paper, our goal is to keep the best of both worlds. More specifically, we adapt the idea of non rectangular partitioning proposed by Lastovetsky and we extend it to any number of processors by adapting the recursive partitioning algorithm proposed by Nagamochi, which facilitates approximation ratio proofs. These two ingredients lead to an improvement of the approximation ratio down to $\frac{2}{\sqrt{3}} \simeq 1.15$ that does not require any specific assumption on the relative speed of resources and is therefore applicable to nodes consisting of both regular cores and accelerators.

This partitioning problem can be used as a building block for many dense linear algebra kernels. For instance, it has been extended to LU factorization and other dense linear algebra kernels in [7], [8]. In this case, block cyclic principle is combined to the initial partitioning in order to obtain 2D-

cyclic ScaLAPACK solutions [9], where the load is balanced throughout the whole computation. These partitionings have also been adapted to distributed hierarchical and highly heterogeneous platforms in [10], where the partitioning is applied at two levels (intra-node and inter-node), based on sophisticated performance models. The same partitioning has also been extended to finite-difference time-domain (FDTD) method to obtain numerical solutions of Maxwell's equations in [11]. The extension to more dynamic settings has also been considered in [12]. In this case, the partitioning problem can be used in order to provide an initial static partitioning algorithm that can be modified in order to dynamically maintaining load balancing. Recently, in order to cope with resource heterogeneity and the difficulty to build optimal schedules, the use of dynamic runtime schedulers have been proposed, such as StarPU [13], StarSs [14], QUARK [15] or PaRSEC [16]. At runtime, the scheduler takes the scheduling and allocation decisions based on the set of ready tasks (tasks whose all data and control dependences have been solved), on the availability of the resources (estimated using expecting processing and communication times), and on the actual location of input data. The comparison between static scheduling strategies (such as the one proposed in this paper) and runtime scheduling strategies has been recently considered in [17], where the analysis of the behavior of static, dynamic, and hybrid strategies highlights the benefits of introducing more static knowledge and allocation decisions in runtime libraries.

All these papers are based on the partitioning problem considered in this paper and can therefore directly benefit from an improvement in the performance and approximation ratio.

Paper Outline

The paper is organized as follows. In Section II, we formally present the partitioning problem PERI-SUM and the notations that will be used throughout the paper. The algorithm we propose for solving PERI-SUM is presented in Section III and its $\frac{2}{\sqrt{3}} \simeq 1.15$ approximation ratio is proved in Section IV (and in Appendix VII due to lack of space). At last, we finish with a set of simulations in order to assess the efficiency of proposed algorithm on realistic instances in Section V and we provide conclusions and perspectives in Section VI.

II. PROBLEM STATEMENT AND NOTATIONS

In this section, we define the notations that will be used in the rest of this paper and we present the formal version of the optimization problem that corresponds to enforcing a perfect load balancing while minimizing the amount of communications.

Let us consider the unitary square $S = [0,1] \times [0,1]$. Let Z denote a zone (a set of points) included in the unit square. We denote by $s(Z)$ its area and by $R(Z)$ its covering rectangle, *i.e.* the Cartesian product of the projections of Z along both dimensions. If $R(Z) = [x_1, x_2] \times [y_1, y_2]$, then let us define the height of Z by $h(Z) = x_2 - x_1$ and the width of Z by $w(Z) = y_2 - y_1$. Finally, let us define $p(Z) = h(Z) + w(Z)$,

the half-perimeter of $R(Z)$ and $\rho(Z) = \frac{max(h(Z), w(Z))}{min((h(Z), w(Z))}$, its aspect ratio.

We consider the following problem :

Problem 1 (PERI-SUM)**.** *Given a set of p rational numbers $\{s_1, \ldots, s_p\}$ such that $\sum s_k = 1$, and the square $S = [0,1] \times [0,1]$, find for each s_k an area $Z_k \in S$ such that the area of Z_k is s_k, $\bigcup Z_k = S$, and such that $\sum p(Z_k)$ is minimized.*

The decision problem associated to this optimization has been proved to be in NP-Complete in [2]. In the following, we denote $\sum p(Z_k)$ as $c(Z_1, \ldots, Z_p)$ and its optimal value as c_{opt}. A lower bound has been proposed by Ballard et al. in[18], that comes from an application of the Loomis-Whitney inequality. This lower bound simply states that the perimeter of an zone Z_k of given area $s(Z_k)$ is minimal when the zone is shaped as a square.

$$c(Z_k) \geq 2\sqrt{s(Z_k)} \qquad (1)$$

Of course, in general, it is not possible to partition the unit square into a set of squares (consider for instance the case of two identical zones of area $\frac{1}{2}$), so that this lower bound is in general too optimistic. On the other hand, it is always possible to partition it into a set of rectangles (consider a 1D partition with vertical lines), and the 1.25 approximation ratio of Nagamochi et al. [5] is indeed based on rectangle-based partition. In this paper, we do not enforce the zones allocated to the different processors to be rectangle-based, what enables to obtain better results for small heterogeneous settings, as advocated by Lastovetsky in [3], [4] (consider for instance the case of a $(1-\epsilon)$ area and an ϵ area, where the optimal solution consist in having a small square for the ϵ area that punches a hole into the unit square.

III. ALGORITHM

In this section, we describe NRRP, the approximation algorithm we propose to solve PERI-SUM. This algorithm is based on a divide and conquer paradigm. At each step, it tries to split the actual rectangle into two parts (three in a few cases), and is applied recursively on each part. In the following, we use the terms **simple** and **composed** zones. Simple zones are terminal and are allocated to a single processor. In what follows, they will be denoted using letter Z. Composed zones are the union of simple zones that are encountered during the algorithm. In what follows, they will be denoted using letter R.

The description of NRRP is presented in Algorithm 5. It relies on basic subroutines that are described in Section III-A, III-B, III-B, III-D and III-E.

The proof that NRRP is indeed a $\frac{2}{\sqrt{3}}$-approximation algorithm is decomposed into two parts. We have observed with Equation (1) that zones allocated to processors should be as close to squares as possible but we have noticed in Section II that it is unfortunately not possible to partition the unit square into a set of squares of prescribed area. Nevertheless, enforcing that all rectangles should have an aspect ratio smaller than $5/2$ would be enough to prove the claimed approximation ratio (actually a slightly better one, as given by Lemma 9).

In order to keep the proofs relatively simple, we enforce that NRRP can only be applied to a (simple or composed) zone R that fulfills the following properties: R must be (i) a rectangle (ii) whose aspect ratio is less than $5/2$. Proving that these two properties hold true in Algorithm 5 is done in Section III-F.

Unfortunately, it is not always possible to partition the unit square into rectangles whose aspect ratios are all smaller than $5/2$, and NRRP may create such zones under the following two conditions: (i) these zones have to be terminal (ii) each time NRRP creates such a set of zones, the $\frac{2}{\sqrt{3}}$-approximation ratio must be fulfilled for the whole set of zones (maybe not for each individual zone, but globally). Proving that these two properties hold true in Algorithm 5 is done in Section IV.

In both cases (Section III-F and Section IV), proofs are rather technical and involve many subcases, but all these subcases are required to enforce claimed approximation ratio.

A. Guillotine

The first routine of NRRP is the Guillotine routine, depicted in Figure 1. It is the main ingredient of Nagamochi's algorithm [5]. Given a composed zone R and a rational number $\alpha \in [0, 1]$, $Guillotine(R, \alpha)$ splits R along the largest dimension into two rectangles of respective areas $\alpha(R)$ and $(1 - \alpha)s(R)$. In some cases, we may need to perform two Guillotine calls in sequence, and we denote it with an additional input parameter. In order to reduce the space required by Guillotine , we allow it to have one input parameter.More specifically, if $R_1, R_2 = Guillotine(R, \alpha)$, then $Guillotine(R, \alpha, \beta) = Guillotine(R_1, \beta), R_2$.
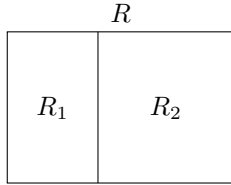

Figure 1. An illustration of the Guillotine routine.

Algorithm 1: Guillotine(R, α)

Input: A rectangle $R = [x_1, x_2] \times [y_1, y_2]$, $\alpha \in [0, 1]$
Output: Two zones of areas $\alpha S(R)$, $(1 - \alpha)S(R)$
$h = h(R)$; $w = w(R)$;
if $h \geq w$ **then**
$\quad | \quad R_1 = [x_1, x_1 + \alpha h] \times [y_1, y_2]$
else
$\quad \lfloor \quad R_1 = [x_1, x_2] \times [y_1, y_1 + \alpha w]$
$R_2 = R \setminus R_1$;
return R_1, R_2

B. Square

The second routine is the Square routine, depicted in Figure 2. Given a rectangle R and a rational number $\alpha \in [0, 1]$, $Square(R, \alpha)$ returns a square R_1 of area $\alpha s(R)$ and a zone

Z_2 which corresponds to the initial rectangle R punched by square R_1. The covering rectangle of Z_2 is R and Z_2 will always be used to host a simple zone.
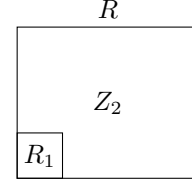

Figure 2. An illustration of the Square routine.

Algorithm 2: Square(R, α)

Input: A rectangle $R = [x_1, x_2] \times [y_1, y_2]$, $\alpha \in [0, 1]$
Output: Two zones of area $\alpha S(R)$, $(1 - \alpha)S(R)$
$s = s(R)$;
$R_1 = [x_1, x_1 + \sqrt{\alpha s}] \times [y_1, y_1 + \sqrt{\alpha s}]$;
$Z_2 = R \setminus R_1$;
return R_1, Z_2

C. Tripartition

The third routine is the Tripartition routine, depicted in Figure 3. In some (rare) cases, neither Guillotine nor Square routines are able to provide either simple zones or composed zones consisting of a rectangle whose aspect ratio is smaller than $5/2$. In this case, we use $Tripartition(R, \alpha, \beta)$, that returns three rectangles R_1, Z_2 and Z_3 of respective areas $\alpha s(R)$, $\beta s(R)$ and $(1 - \alpha - \beta)s(R)$. The difference with the result given by $Guillotine(R, \alpha + \beta, \alpha/(\alpha + \beta)$ is that we do not perform the second split along the largest dimension of R'. In practice, the covering rectangle Z_2 and Z_3 will always be used to host simple zones. Hence, only the aspect ratio $\rho(R_1)$ needs to be smaller than $5/2$.
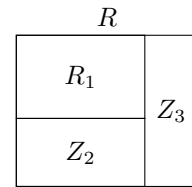

Figure 3. An illustration of the Tripartition routine.

D. Superposition

The fourth and last routine is the Superposition routine. $Superposition(R, \alpha, \epsilon)$ returns three zones, R_1, Z_2 and Z_3 of respective areas $\epsilon s(R)$, $(\alpha - \epsilon)s(R)$ and $(1 - \alpha)s(R)$. R_1 is a square that can be placed in the upper left corner, Z_2 a rectangle which is placed under R_1 in the bottom left corner and Z_3 is the remaining zone, *i.e.* R punched by both R_1 and Z_2. In practice, Z_2 and Z_3 will always be used to host simple zones. Hence, only the aspect ratio $\rho(R_1)$ needs to be smaller than $5/2$, what is always the case since R_1 is a square.

Algorithm 3: Tripartition(R, α, β)

Input: A rectangle $R = [x_1, x_2] \times [y_1, y_2]$, $\alpha, \beta \in [0, 1]$
Output: Three zones of respective areas $\alpha S(R)$, $\beta S(R)$,
    $(1 - \alpha - \beta)S(R)$
$h = h(R)$; $w = w(R)$;
if $h \geq w$ **then**
  $h_1 = (\alpha + \beta)h$; $w_1 = \frac{\alpha}{\alpha + \beta}w$;
  $R_1 = [x_1, x_1 + h_1] \times [y_1, y_1 + w_1]$;
  $Z_2 = [x_1, x_1 + h_1] \times [y_1 + w_1, y_2]$
else
  $h_1 = \frac{\alpha}{\alpha + \beta}h$; $w_1 = (\alpha + \beta)w$;
  $R_1 = [x_1, x_1 + h_1] \times [y_1, y_1 + w_1]$;
  $Z_2 = [x_1 + h_1, x_2] \times [y_1, y_1 + w_1]$
$Z_3 = R \setminus (R_1 \cup R_2)$;
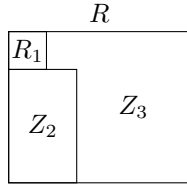**return** R_1, Z_2, Z_3

Figure 4.   An illustration of the Superposition routine.

E. Packing

Given a list $\{s_1, \ldots s_k\}$ sorted in increasing order, two rational values s and s' and a rectangle R such that $s(R) = \sum s_i$, $Packing(\{s_1, \ldots, s_k\}, s, s', R)$ returns a list of couples (R_j, S_j) where R_j is a rectangle and S_j is a subset of $\{s_1, \ldots, s_k\}$ such that $\bigcup R_j = R$, $s(R_j) = \sum_{i \in S_j} s_i$ and $s \leq s(R_j) \leq s'$. For conciness, we allow s_k to be larger than s' and in this case $Packing(\{s_1, \ldots, s_k\}, s, s', R) = Packing(\{s_1, \ldots, s_{k-1}\}, s, s', R_1) + (R_2, s_k)$, where $R_1, R_2 = Guillotine(R, (s(R) - s_k)/s(R))$. We will explicitly discuss the existence of such a function each time we use it. We will denote by Map the function that allocates a specific zone to the R_js.

Algorithm 4: Superposition(R, α, ϵ)

Input: A rectangle $R = [x_1, x_2] \times [y_1, y_2]$, $\alpha, \epsilon \in [0, 1]$
Output: Three zones of respective areas $\epsilon S(R)$,
    $(\alpha - \epsilon)S(R)$, $(1 - \alpha)S(R)$
$h = h(R)$; $w = w(R)$; $s = s(R)$;
$l = \sqrt{\epsilon s}$; $R_1 = [x_1, x_1 + l] \times [y_1, y_1 + l]$;
if $h \geq w$ **then**
  $h_1 = \frac{(\alpha - \epsilon)hw}{w - l}$; $Z_2 = [x_1, x_1 + h_1] \times [y_1 + l, y_2]$
else
  $w_1 = \frac{(\alpha - \epsilon)hw}{h - l}$; $Z_2 = [x_1 + l, x_2] \times [y_1, y_1 + w_1]$
$Z_3 = R \setminus (R_1 \cup Z_2)$;
return R_1, Z_2, Z_3

F. Correction Proof

Algorithm $NRRP(R, \{s_1, \ldots, s_n\})$ is depicted in Algorithm 5. In order to prove its correctness, we need to prove the following theorem, that states that all the composed rectangles (on which the algorithm is recursively applied) have an aspect ratio lower than $5/2$, since this property is crucial in order to establish the approximation ratio proved in Section IV. In all the following, we consider that the list of s_i values is sorted in increasing order, $s_1 \leq s_2 \leq \ldots \leq s_n$. Furthermore, we assume that rectangles R are composed zones, *i.e.* $n > 1$.

Theorem 1 (Correctness)**.** *When executing* $NRRP(R, \{s_1, \ldots, s_n\})$ *with* $\rho(R) < 5/2$, *all the recursive calls to* $NRRP(R', \{s_1', \ldots, s_k'\})$ *are performed on a rectangle area* R' *such that* $\rho(R') < 5/2$.

Proof: The first step of the algorithm finds k, the smallest index such that $s' = \sum_{i=1}^{k} s_i \geq \frac{2s}{5\rho}$, where $s = s(R)$ and $\rho = \rho(R)$ (Line 5). Depending on the value of k, there are two cases (Line 7).

Case (A) **(lines 8-13)** corresponds to the case $k < n$, that is split into two subcases.

Case (A1) (lines 9-10) corresponds to the case where $s - s'$ is also larger than $\frac{2s}{5\rho(R)}$. Then, with $\alpha = s'/s$, Lemma 2 applies.

Lemma 2. *Let R be a rectangle with $\rho(R) \leq 5/2$, $\alpha \in [0, 1]$ and $R_1, R_2 = Guillotine(R, \alpha)$.*

- *If $\alpha \geq \frac{2}{5\rho(R)}$ then $\rho(R_1) \leq 5/2$.*
- *If $(1 - \alpha) \geq \frac{2}{5\rho(R)}$ then $\rho(R_2) \leq 5/2$.*

Proof: Let us assume without loss of generality that $h = h(R) \geq w(R) = w$, and denote $\rho = \rho(R) = \frac{h}{w}$. Then $\rho(R_1) = \min(\frac{\alpha h}{w}, \frac{w}{\alpha h})$. We have $\frac{\alpha h}{w} \leq \frac{h}{w} = \rho \leq 5/2$ and $\frac{w}{\alpha h} = \frac{1}{\alpha \rho} \leq \frac{5\rho}{2\rho} \leq 5/2$ (under the assumption $\alpha \geq \frac{2}{5\rho}$). Therefore, $\alpha \geq \frac{2}{5\rho}$ implies $\rho(R_1) \leq 5/2$ and for the same reason, $(1 - \alpha) \geq \frac{2}{5\rho}$ implies $\rho(R_2) \leq 5/2$. ∎

Therefore, in case (A1), $Guillotine(R, \alpha)$ returns two rectangles R_1, R_2 whose aspect ratios are smaller than $5/2$ and we can apply NRRP on each of them.

Case (A2) (lines 12-13) corresponds to the case where $s' = \sum_{i=1}^{k} s_i \geq \frac{2s}{5\rho}$, $s - s' \leq \frac{2s}{5\rho}$ and $k < n$. In this case, we rely on the Tripartition routine and the following lemma states that if $R_1, Z_2, Z_3 = Tripartition(R, (s' - s_{n-1})/s, s_{n-1}/s)$, then $\rho(R_1) \leq 5/2$ and Z_2 and Z_3 are simple.

Lemma 3. *If $s' = \sum_{i=1}^{k} s_i \geq \frac{2s}{5\rho(R)}$, $k < n$ and $s - s' < \frac{2s}{5\rho(R)}$, then $k = n - 1$ and $\rho(R) < 6/5$. In addition, if $R_1, R_2, R_3 = Tripartition(R, (s' - s_{n-1})/s, s_{n-1}/s)$, then $\rho(R_1) \leq 5/2$.*

Proof: Let us assume without loss of generality that $h = h(R) \geq w(R) = w$ and define $\rho = \rho(R)$. By definition of k, we know that $s'' = \sum_{i=1}^{k-1} s_i < \frac{2s}{5\rho}$. Therefore
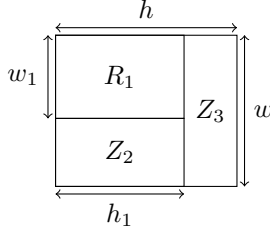
$$s_k = s' - s''$$
$$s_k = s - (s - s') - s''$$

Figure 5. **Case (A2)**

$$s_k > s - \frac{2s}{5\rho} - \frac{2s}{5\rho}$$

$$s_k > s(1 - \frac{4}{5\rho}) = s\frac{4\rho - 4}{5\rho}$$

Since the s_i values are sorted, $s - s' = \sum_{i=k+1}^{n} s_i \geq (n - k)s_{k+1} \geq (n-k)s_k$. Hence $\frac{2s}{5\rho} > (n-k)s_k > (n-k)s\frac{5\rho-4}{5\rho}$. This implies that $n - k < \frac{2}{5\rho-4} \leq 2$ since $\rho \geq 1$, and then $k \geq n - 1$. Thus, the only possible value for k is $n - 1$ (remember that $k < n$). Therefore, $s = s'' + s_{n-1} + s_n < 3\frac{2s}{5\rho}$ and $\rho < 6/5$.

Let us denote $\alpha = s''/s$, $\beta = s_{n-1}/s$ and $\gamma = s_n/s$, $h_1 = h(R_1)$ and $w_1 = w(R_1)$ (see Figure 5). We want to prove that both h_1/w_1 and w_1/h_1 are smaller than $5/2$, which is equivalent to proving $2/5 \leq h_1/w_1 \leq 5/2$. First, recall that $h_1 = (\alpha + \beta)h$ and $w_1 = \frac{\alpha}{\alpha+\beta}w$.

Let us now establish lower bounds on α, β and γ (the following upper bounds holds true $\alpha < \frac{2}{5\rho}$ and $\beta \leq \gamma < \frac{2}{5\rho}$). Let us notice that $1 - \alpha = \beta + \gamma < \frac{4}{5\rho}$. Therefore $\alpha > \frac{5\rho-4}{5\rho}$, and similarly $\beta > \frac{5\rho-4}{5\rho}$. For γ, noticing that $2\gamma \geq \beta + \gamma = 1 - \alpha > \frac{5\rho-2}{5\rho}$, we obtain $\gamma > \frac{5\rho-2}{10\rho}$.

Since $\alpha + \beta = 1 - \gamma$, then

$$\frac{5\rho - 2}{5\rho} < \alpha + \beta < \frac{5\rho + 2}{10\rho}.$$

Moreover, since $h_1/w_1 = \rho\frac{(\alpha+\beta)^2}{\alpha}$, then

$$\rho\frac{(5\rho - 2)^2}{25\rho^2} \times \frac{5\rho}{2} < \frac{h_1}{w_1} < \rho\frac{(5\rho + 2)^2}{100\rho^2} \times \frac{5\rho}{5\rho - 4}$$

and

$$\frac{(5\rho - 2)^2}{10} < \frac{h_1}{w_1} < \frac{(5\rho + 2)^2}{20(5\rho - 4)}.$$

Trivially, $\frac{(5\rho-2)^2}{10} \geq 9/10 > 2/5$. Moreover, $\frac{(5\rho+2)^2}{20(5\rho-4)} \leq 49/20 < 5/2$ since $x \mapsto \frac{(5x+2)^2}{20(5x-4)}$ is a decreasing function on $[1, 6/5]$. Hence, we prove $2/5 \leq h_1/w_1 \leq 5/2$ and therefore, $\rho(R_1) \leq 5/2$. ∎

Above Lemmas prove the correctness of the calls to NRRP in cases (A1) and (A2) (lines (10) and (13) of Algorithm 5). Case (B) (lines 15-48) corresponds to the case where $k = n$, which happens when s_n is significantly larger than the other values. Let us denote $s' = s - s_n$ (Line 15). Depending on the value of s', several subcases can occur.

Case (B1) (lines 17 and 18) corresponds to a small s', *i.e.* $s' \leq s(1 - \frac{3(\rho+1)^2}{16\rho})$. In this case, Square is called and

generates R_1 that is a square (with aspect ratio is $1 < 5/2$) and the simple zone Z_2.

Before continuing, let us note that in all remaining cases, $\rho - \frac{3(\rho+1)^2}{16} < \frac{\rho s'}{s} < 2/5$ (where $\rho = \rho(R)$). In addition, $13/64 \leq \rho - \frac{3(\rho+1)^2}{16}$ for $\rho \in [0, 5/2]$. Thus,

$$13/64 \leq \rho - \frac{3(\rho + 1)^2}{16} < \rho\frac{s'}{s} < 2/5 \qquad (2)$$
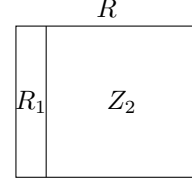

Figure 6. **Case (B2)**

The situation is depicted in Figure 6, where Z_2 is simple (and such that $\rho(Z_2) \leq 5/2$). Unfortunately, $\rho(R_1) > 5/2$ so that NRRP cannot directly be called on R_1, and needs to be further split into several rectangles with acceptable aspect ratio. Lemma 4 details (if we set $\alpha = \rho\frac{s'}{s}$) the conditions such that this holds true.

Lemma 4. *Let R be a rectangle such that $\rho(R) = 1/\alpha$ and $R', R'' = Guillotine(R, x)$. If $13/64 \leq \alpha \leq 2/5$ and $\frac{2\alpha}{5} \leq x \leq \frac{5\alpha}{2}$, then $\rho(R') \leq 5/2$.*

Proof: Let us suppose without loss of generality that $w = w(R) \leq h(R) = h$. In this case, $w(R') = w$ and $h(R') = xh$ with $x \in [0, 1]$. $s(R') = xh \times w = xS(R)$. Therefore $\frac{2\alpha}{5} \leq x \leq \frac{5\alpha}{2}$. Since $\rho(R') \leq 5/2 \Leftrightarrow 2/5 \leq h(R')/w(R') \leq 5/2$ is equivalent to $2/5 \leq x/\alpha \leq 5/2$ and $\frac{2\alpha}{5} \leq x \leq \frac{5\alpha}{2}$, which is true by construction. ∎

Let us denote $s'' = s' - s_{n-1}$.

Case (B2-a) (Lines 21-40) corresponds to the case $\frac{2\rho s'^2}{5s} \leq s''$.

Case (B2-a1) (Lines 23-24) correspond to the case $\frac{2\rho s'^2}{5s} \leq s'' \leq \rho\frac{5s'^2}{2s}$. Lemma 4 proves that if $R_1, Z_2, Z_3 = Guillotine(R_1, s'/s, s''/s')$ (see Figure 7), then $\rho(R_1) \leq 5/2$ and that Z_2 and Z_3 are simple so that the call $NRRP$ on R_1 at line 24 is valid.
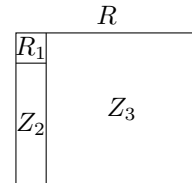

Figure 7. **Case (B2-a1)**

Case (B2-a2) (lines 26-40) corresponds to the case where $s'' > \rho\frac{5s'^2}{2s}$. It is again split into several subcases depending on the value of $s''' = s'' - s_{n-2}$. Let us first note that $s''' > 0$. Indeed, thanks to equation (2), we know that $s''/s' > \rho\frac{5s'}{2s} >$

$65/128 > 1/2$. Therefore $s_{n-2} \leq s_{n-1} < s'(1 - 1/2) < s'/2 < s''$.

Case (B2-a2') (Lines 28-30) corresponds to the case $s''' \geq \rho \frac{5s'^2}{2s}$. Then, the conditions of Lemma 5 hold true (with $\alpha = \rho \frac{s'}{2}$) and we can build $Packing(\{s_1, \ldots, s_{n-1}\}, \frac{2\rho s'^2}{5s}, \frac{5\rho s'^2}{2s}, R_1)$ and call NRRP on each element of the list, since all the rectangles of the list, except possibly one simple rectangle, have an aspect ratio smaller than $5/2$ (see Figure 8).
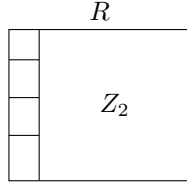

Figure 8. **Case (B2-a2')**

Lemma 5. *Let R be a rectangle such that $\rho(R) = 1/\alpha$ and $\{s_1, \ldots, s_n\}$ an ordered list such that $\sum s_i = s(R)$. Then, if $n \geq 3$, $\sum_{i=1}^{n-2} s_i \geq \frac{2\alpha}{5}$, $\sum_{i=1}^{n-1} s_i > \frac{5\alpha}{2}$ and $13/64 < \alpha < 2/5$, then we can build $Packing(R, \{s_1, \ldots, s_n\}, \frac{2\alpha}{5}, \frac{5\alpha}{2})$ (in linear time).*

Proof: Let us denote $s = s(R)$, $s' = s - s_n$ and $s'' = s' - s_{n-1}$ (we know that $s'' \geq \frac{2\alpha}{5}$). Note that $\frac{5\alpha}{2} > 1/2$.

- Let us first assume that $s_n + s_{n-1} > \frac{5\alpha}{2}$. In this case, $2s_n \geq s_n + s_{n-1} > \frac{5\alpha}{2}$ and therefore $s_n > \frac{5\alpha}{4} > \frac{2\alpha}{5}$. In addition, $s'' \leq \frac{5\alpha}{2}$ (otherwise $s = s'' + s_{n-1} + s_n > s/2 + s/2$). Therefore, let us consider R_1, where $R_1, R_2 = Guillotine(R, s'/s)$. Let j be such that $\sum_{i=1}^{j} s_i \leq s' - \frac{2\alpha}{5}$ and $\sum_{i=1}^{j+1} s_i > s' - \frac{2\alpha}{5}$. We have two cases:
  (i) If $s_{n-1} \geq \frac{2\alpha}{5}$, then $j = n - 2$. Therefore, $\frac{2\alpha}{5} \leq s'' \leq \frac{5\alpha}{2}$ and $\frac{2\alpha}{5} \leq s_{n-1} \leq \frac{5\alpha}{2}$ ($s_{n-1} > \frac{5\alpha}{2}$ implies $s_n > \frac{5\alpha}{2}$ and for the same reason, since $s'' \leq \frac{5\alpha}{2}$, this is impossible). Therefore, if $R_1, R_3 = Guillotine(R_1, s''/s')$, $[(R_1, \{s_1, \ldots, s_{n-2}\}), (R_3, \{s_{n-1}\}), (R_2, \{s_n\})]$ is a valid return of $Packing(R, \{s_1, \ldots, s_n\}, \frac{2\alpha}{5}, \frac{5\alpha}{2})$.
  (ii) If $s_{n-1} < \frac{2\alpha}{5}$, then $\forall i \leq n-1, s_i < \frac{2\alpha}{5}$ and $j < n-1$. Let us denote $s_{bis} = \sum_{i=1}^{j} s_i$ and $s_{ter} = \sum_{i=j+1}^{n-1} s_i$. $s_{bis} < s''' \leq \frac{5\alpha}{2}$ and

$$s_{bis} + s_{j+1} > s' - \frac{2\alpha}{5} > \frac{5\alpha}{2} - \frac{2\alpha}{5}$$

Hence

$$s_{bis} > \frac{5\alpha}{2} - \frac{2\alpha}{5} - \frac{2\alpha}{5} = (5/2 - 4/5)\alpha \geq \frac{2\alpha}{5}.$$

Moreover, $s' - s_{bis} = s_{ter}$ implies that $s_{ter} \geq s' - (s' - \frac{2\alpha}{5})$ and therefore $s_{ter} \geq \frac{2\alpha}{5}$, and $s_{ter} \leq s' - (s' - \frac{4\alpha}{5}) < \frac{5\alpha}{2}$. Therefore, $R_1, R_3 = Guillotine(R_1, (s_{bis} + s_{ter})/s', s_{ter}/(s_{bis} + s_{ter}))$, $[(R_1, \{s_1, \ldots, s_J\}), (R_3, \{s_{j+1}, \ldots, s_{n-1}\}), (R_2, \{s_n\})]$ is a valid return of $Packing(R, \{s_1, \ldots, s_n\}, \frac{2\alpha}{5}, \frac{5\alpha}{2})$.

- Otherwise, $s_n + s_{n-1} \leq \frac{5\alpha}{2}$. Since $s_{n-1} \leq s_n$, we know that $s_{n-1} \leq \frac{5\alpha}{4}$ and therefore $\forall i \leq n - 1, s_i \leq \frac{5\alpha}{4}$. Let us apply the following procedure:
  - If $s_i \geq \frac{2\alpha}{5}$, then, since $s_i \leq s_n \leq \frac{5\alpha}{2}$, we can leave s_i alone since it fulfills the condition.
  - Else, let j_i be such that $\sum_{j=j_i}^{i} s_j \geq \frac{2\alpha}{5}$ and $\sum_{j=j_i+1}^{i} s_j < \frac{2\alpha}{5}$. Note that

$$\sum_{j=j_i}^{i} s_j = s_{j_i} + \sum_{j=j_i+1}^{i} s_j < s_i + \sum_{j=j_i+1}^{i} s_j < \frac{4\alpha}{5} < \frac{5\alpha}{2}$$

Then, the set $\{s_{j_i}, \ldots, s_i\}$ fulfills the condition.

After the execution of the above algorithm, there may exist a k such that $\sum_{j=1}^{k} s_j < \frac{2\alpha}{5}$. If $s_{k+1} \geq \frac{2\alpha}{5}$, then

$$\sum_{j=1}^{k+1} s_j \leq \frac{2\alpha}{5} + s_{n-1} \leq \frac{2\alpha}{5} + \frac{5\alpha}{4} = \frac{33\alpha}{20} < \frac{5\alpha}{2}$$

and trivially $\sum_{j=1}^{k+1} s_j \geq \frac{2\alpha}{5}$. Therefore the set $\{s_1, \ldots, s_{k+1}\}$ fulfills the condition. Otherwise, there exists i such that $k = j_i$. Then

$$\sum_{j=1}^{i} s_j \leq \frac{2\alpha}{5} + \frac{4\alpha}{5} = \frac{6\alpha}{5} < \frac{5\alpha}{2}$$

and trivially $\sum_{j=1}^{i} s_j \geq \frac{2\alpha}{5}$. Therefore the set $\{s_1, \ldots, s_i\}$ fulfills the condition.

Then, in any possible case, we have built a valid result for $Packing(R, \{s_1, \ldots, s_n\}, \frac{2\alpha}{5}, \frac{5\alpha}{2})$ (in linear time). ∎

Case (B2-a2") (Lines 33-40) corresponds to the case $s''' < \frac{2\rho s'^2}{5s}$. Note that $s''' < \frac{2\rho s'^2}{5s}$ and $s'' > \frac{5\rho s'^2}{2s}$ is a possible situation, for example with $s_{n-1} = s_{n-2} = \frac{21\rho s'^2}{10s} + \epsilon$). In this case, either we successively apply Square and Guillotine on R_1 or we use Superposition on R (the choice will be discussed in Section IV). Both cases are depicted in Figure 9 and in all case, at most one rectangle (R_1) is not simple, and since it is shaped as a square, then its aspect ratio is less than $5/2$, and the calls to NRRP at lines 36 and 40 are both valid.
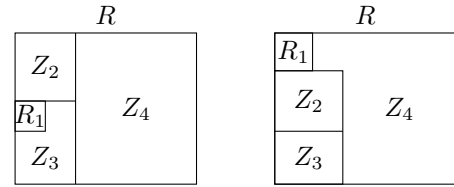

Figure 9. **Case (B2-a2")**

Case (B2-b) (lines 42 to 48) corresponds to the case $s' < \frac{2\rho s'^2}{5s}$. In this case, we rely on the technique described for Case (B2-a2"). We successively apply Square and Guillotine on R_1 or we use Superposition on R (the choice will be discussed in Section IV) and are depicted in Figure 10.

This ends the proof of Theorem 1 ∎

Algorithm 5: $\text{NRRP}(R, \{s_1, \ldots, s_n\})$

Input: A rectangle R, a set of values $\{s_1, \ldots, s_n\}$ such that $\sum s_i = s(R)$ and $s_1 \leq s_2 \leq \ldots \leq s_n$

Output: For each $1 \leq i \leq n$, a zone A_i such that $s(A_i) = s_i$ and $\bigcup A_i = R$

1 **if** $n = 1$ **then**
2   **return** R
3 **else**
4     $\rho = \rho(R)$;
5     $k = $ the smallest k such that $\sum_{i=1}^{k} s_i \geq \frac{2s}{5\rho}$;
6     $s' = \sum_{i=1}^{k} s_i$;
7   **if** $k < n$ **then**
8   **if** $s - s' \geq \frac{2s}{5\rho}$ **then**
9           $R_1, R_2 = \text{Guillotine}(R, s'/s)$;
10   **return** $\text{NRRP}(R_1, \{s_1, \ldots, s_k\}) + \text{NRRP}(R_2, \{s_{k+1}, \ldots, s_n\})$
11   **else**
12           $R_1, R_2, R_3 = \text{Tripartition}(R, (s' - s_{n-1})/s, (s_{n-1})/s)$;
13   **return** $\text{NRRP}(R_1, \{s_1, \ldots, s_k\}) + R_2 + R_3$
14   **else**
15        $s' = \sum_{i=1}^{k-1} s_i$;
16   **if** $s'/s \leq 1 - \frac{3(\rho+1)^2}{16\rho}$ **then**
17           $R_1, A_2 = \text{Square}(R, s'/s)$;
18   **return** $\text{NRRP}(R_1, \{s_1, \ldots, s_k\}) + A_2$
19   **else**
20           $s'' = s' - s_{n-1}$;
21   **if** $s'' \geq \frac{2\rho s'^2}{5s}$ **then**
22   **if** $s'' \leq \frac{5\rho s'^2}{2s}$ **then**
23                 $R_1, R_2, R_3 = \text{Guillotine}(R, s'/s, s''/s')$;
24   **return** $\text{NRRP}(R_1, \{s_1, \ldots, s_{n-2}\}) + R_2 + R_3$
25   **else**
26                 $s''' = s'' - s_{n-2}$;
27   **if** $s''' \geq \frac{2\rho s'^2}{5s}$ **then**
28                     $R_1, R_2 = \text{Guillotine}(R, s'/s)$;
29                     $L = \text{Packing}(\{s_1, \ldots, s_{n-1}\}, \frac{2\rho s'^2}{5s}, \frac{5\rho s'^2}{2s}, R_1)$;
30   **return** $\text{Map}(MonAlgo,) + R_2$
31   **else**
32   **if** $s'''/s \leq \frac{(1 - \sqrt{1 - \rho * s'/s})^2}{\rho}$ **then**
33                     $R_1, R_2, R_4 = \text{Guillotine}(R, s'/s, (s''' + s_{n-1})/s)$;
34                     $R_1, A_3 = \text{Square}(R_1, s'''/(s''' + s_{n-1}))$;
35   **return** $\text{NRRP}(R_1, \{s_1, \ldots, s_{n-3}\}) + R_2 + A_3 + R_4$
36   **else**
37                     $R_1, R_4 = \text{Guillotine}(R, s'/s)$;
38                     $R_1, R_2, A_4 = \text{Superposition}(R, s'/s, s'''/s)$;
39                     $R_2, R_3 = \text{Guillotine}(R_2, s_{n-2}/(s' - s'''))$;
40   **return** $\text{NRRP}(R_1, \{s_1, \ldots, s_{n-3}\}) + R_2 + R_3 + A_4$

41   **else**
42   **if** $s''/s \leq \frac{(1 - \sqrt{1 - \rho * s'/s})^2}{\rho}$ **then**
43                 $R_1, R_3 = \text{Guillotine}(R, s'/s)$;
44                 $R_1, A_2 = \text{Square}(R_1, s''/s')$;
45   **return** $\text{NRRP}(R_1, \{s_1, \ldots, s_{n-2}\}) + A_2 + R_3$
46   **else**
47                 $R_1, R_2, A_3 = \text{Superposition}(R, s'/s, s''/s)$;
48   **return** $\text{NRRP}(R_1, \{s_1, \ldots, s_{n-2}\}) + R_2 + A_3$
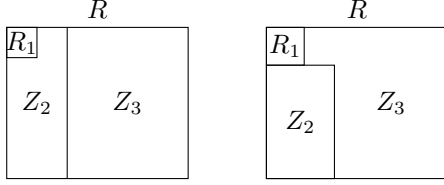
Figure 10. **Case (B2-b)**

IV. APPROXIMATION PROOF

In this section we prove our claim that NRRP is a $\frac{2}{\sqrt{3}}$-approximation for PERI-SUM(Theorem 6).

Theorem 6. *NRRP is a $\frac{2}{\sqrt{3}}$-approximation for PERI-SUM.*

Proof:

We will extensively rely on the following lemmas in order to prove Theorem 6.

Lemma 7. *Given a set of p rational numbers $\{s_1, \ldots, s_p\}$ such that $\sum s_k = 1$, and the square $S = [0,1] \times [0,1]$, find for each s_k an area $A_k \in S$ such that the area of A_k is $_k$, $\bigcup A_k = S$, then $\sum p(A_k) \geq 2 \sum \sqrt{s_k}$*

Proof: This result is a direct consequence of Equation (1), which states that the property is valid term by term. This lower bound is used in what follows, both to prove the approximation ratio in Theorem 6 or to present the performance of the different algorithms in Section V ∎

Lemma 8. *Let A, B, C and D denote 4 rational numbers, then if $\frac{A}{B} \leq \alpha$ and $\frac{C}{D} \leq \alpha$, then $\frac{A+C}{B+D} \leq \alpha$.*

We omit the proof of this textbook lemma. It plays a crucial role in the following proof. Indeed let us partition the s_ks into I subsets $S_1, \ldots, S_I$. Then, by Lemma 8, if

$$\forall i, \frac{\sum_{k \in S_i} p(A_k)}{2 \sum_{k \in S_i} \sqrt{s_k}} \leq \frac{2}{\sqrt{3}}, \text{ then } \frac{\sum p(A_k)}{2 \sum \sqrt{s_k}} \leq \frac{2}{\sqrt{3}}.$$

Based on this result, the rationale of the proof is as follows. For all the cases of Algorithm NRRP that are depicted in Figures 5, 6, 7, 8, 9, 10, we can distinguish between composed rectangles (denoted by R letter) and the simple (terminal) zones (denoted by Z letter).

We have proved in Section V that all composed rectangles have an aspect ratio less than $5/2$. This is enough to prove our result since a such a rectangle satisfies, by Lemma 9

$$\frac{p(A_k)}{2\sqrt{s_k}} \leq \frac{7\sqrt{2}}{4\sqrt{5}} < \frac{2}{\sqrt{3}}.$$

To establish the approximation ratio, using Lemma 8, we can therefore consider those rectangles independently.

Unfortunately, there are composed Z zones for which the $5/2$ aspect ratio does not hold true, and it is not true in all cases that $\frac{p(Z_k)}{2\sqrt{s(Z_k)}} \leq \frac{2}{\sqrt{3}}$. Nevertheless, for each case depicted in in Figures 5, 6, 7, 8, 9, 10, if we group all the terminal simple zones Z_1, Z_2 (and possibly Z_3), then we can prove that $\frac{\sum p(Z_k)}{2 \sum \sqrt{s(Z_k)}} \leq \frac{2}{\sqrt{3}}$, so that the bound holds globally if it is

not the case for individual zones. Then, we can conclude with Lemma 8 that the $\frac{2}{\sqrt{3}}$ bound holds true since it is enough to exhibit one partitioning of the s_ks such that the bounds holds for each individual group of the partition.

The rest of the proof is rather technical and simply proves that the above bound holds true for all possible subcases of Algorithm 5.

The following lemma will be used at several places in the proof.

Lemma 9. *Let R be a rectangle. Then*

$$\frac{c(R)}{2\sqrt{s(R)}} = \frac{\rho(R)+1}{2\sqrt{\rho(R)}}$$

Proof: Let us assume without loss of generality that $h = h(R) \geq w(R) = w$. Therefore $\rho = \rho(R) = h/w$. $c(R) = h + w = (\rho + 1)w$ and $s(R) = hw = \rho w^2$. Hence $\frac{c(R)}{2\sqrt{s(R)}} = \frac{\rho(R)+1}{2\sqrt{\rho(R)}}$. ∎

Let us now check the different possible results of NRRP.

First if $n = 1$, then a single rectangle R is returned, and by construction (see Section III) (Line 2) $\rho(R) \leq 5/2$, what ends the proof.

Case (A1) (lines 9-10): NRRP returns no simple areas.

Case (A2) (lines 12-13) corresponds to the case described in Figure 11 with 2 simple zones Z_2 and Z_3. Lemma 11 proves that $c(Z_2, Z_3) \leq \frac{2}{\sqrt{3}}(2(\sqrt{s_{n-1}} + \sqrt{s_n}))$, what ends the proof of this case.
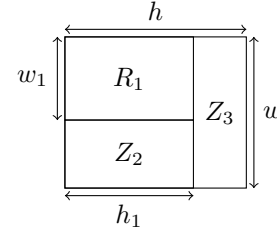

Figure 11. **Case (A2)**

Lemma 10. *Let $x \in [1, 6/5]$, $z \in [\frac{5x-2}{10x}, \frac{2}{5x}]$ and $y \in [\frac{5x-4}{5x}, z]$ and $f(x,y,z) = \frac{x+1+\frac{y}{1-z}}{2\sqrt{x}(\sqrt{y}+\sqrt{z})}$. If $\alpha = 1 - z - y \geq 1/5$ and $xz \leq 2/5$, then $f(x,y,z) \leq f(1, 1/5, 3/10) = \frac{16\sqrt{5}}{7(2+\sqrt{6})} < \frac{2}{\sqrt{3}}$*

Proof: Due to lack of space, the proof is presented in Section VII ∎

Lemma 11. *Let us suppose that $s' = \sum_{i=1}^{k} s_i \geq \frac{2s}{5\rho(R)}$, $k = n - 1$ and $s - s' < \frac{2s}{5\rho(R)}$. Then, if $R_1, Z_2, Z_3 = Tripartition(R, (s' - s_{n-1})/s, s_{n-1}/s)$, $\frac{c(Z_2, Z_3)}{2(\sqrt{s(Z_2)} + \sqrt{s(Z_3)})} \leq \frac{2}{\sqrt{3}}$*

Proof: We use the same notations as for Lemma 3: without loss of generality, $h = h(R) \geq w(R) = w$ and denote $\rho = \rho(R)$, $\alpha = s''/s$, $\beta = s_{n-1}/s$ and $\gamma = s_n/s$. Let us also denote $h_1 = h(R_1)$ and $w_1 = w(R_1)$ (see Figure

11). Remember that we have proved in the proof of Lemma 3 that

$$\frac{5\rho - 4}{5\rho} \leq \alpha \leq \frac{2}{5\rho}$$

$$\frac{5\rho - 4}{5\rho} \leq \beta \leq \gamma$$

$$\frac{5\rho - 2}{10\rho} \leq \gamma \leq \frac{2}{5\rho}$$

$$1 \leq \rho \leq 6/5$$

and that $h_1 = (\alpha + \beta)h$ and $w_1 = \frac{\alpha}{\alpha+\beta}w$. With these notations, $p(Z_2) = h_1 + w - w_1$ and $p(Z_3) = h - h_1 + w$. Then $c(Z_2, Z_3) = p(Z_2) + p(Z_3) = h + 2w - w_1 = \rho w + 2w - \frac{\alpha}{\alpha+\beta}w = w(\rho + 2 - \frac{\alpha}{\alpha+\beta})$. Since $\alpha + \beta = 1 - \gamma$, $\frac{\alpha}{\alpha+\beta} = \frac{1-\gamma-\beta}{1-\gamma} = 1 - \frac{\beta}{1-\gamma}$. Then, $c(Z_2, Z_3) = w(\rho + 1 + \frac{\beta}{1-\gamma})$. Moreover, $s(Z_2) = hw\beta = \beta\rho w^2$ and $s(Z_3) = hw\gamma = \gamma\rho w^2$. Thus, with f defined as in Lemma 10:

$$\frac{c(Z_2, Z_3)}{2(\sqrt{s(Z_2)} + \sqrt{s(Z_3)})} = \frac{\rho + 1 + \frac{\beta}{1-\gamma}}{2\sqrt{\rho}(\sqrt{\beta} + \sqrt{\gamma})} = f(\rho, \beta, \gamma)$$

Since $1 - \beta - \gamma = \alpha > 1/5$ and $\gamma\rho = \rho\frac{s-s'}{s} \leq 2/5$, we can apply Lemma 10 and obtain our result. ∎

Let us now move to Case (B), *i.e.* $k = n$.

Case (B1) (lines 17 and 18): NRRP returns a single simple zone Z_2 defined by $R_1, Z_2 = Square(R, s'/s)$. In this case, Lemma 12 proves that $c(Z_2) \leq \frac{2}{\sqrt{3}} \times 2\sqrt{s(Z_2)}$.

Lemma 12. *Let R be a rectangle. If $\alpha \leq 1 - \frac{3(\rho(R)+1)^2}{16\rho(R)}$ and $R_1, Z_2 = Square(R, \alpha)$, then $\frac{c(Z_2)}{2\sqrt{Z_2}} \leq \frac{2}{\sqrt{3}}$.*

Proof: We suppose, without loss of generality that $h = h(R) \geq w(R) = w$ and then $\rho = \rho(R) = h/w$. If $R_1, Z_2 = Square(R, \alpha)$, then $R(Z_2) = R$ and $c(Z_2) = h + w = (\rho + 1)w$. In the same time $s(Z_2) = (1 - \alpha)s(R) = \rho(1 - \alpha)w^2$. Then

$$\frac{c(Z_2)}{2\sqrt{Z_2}} = \frac{\rho + 1}{2\sqrt{\rho(1-\alpha)}}$$

As $x \mapsto \frac{1}{\sqrt{1-x}}$ is a increasing function on $[0, 1[$ and $\alpha \leq 1 - \frac{3(\rho(R)+1)^2}{16\rho(R)}$, we have:

$$\frac{c(Z_2)}{2\sqrt{Z_2}} = \frac{\rho + 1}{2\sqrt{\rho(1-\alpha)}} \leq \frac{\rho + 1}{2\sqrt{\rho(1 - 1 + \frac{3(\rho+1)^2}{16\rho})}}$$

Therefore:

$$\frac{c(Z_2)}{2\sqrt{Z_2}} \leq \frac{\rho + 1}{2\sqrt{\frac{3(\rho+1)^2}{16}}} = \frac{2}{\sqrt{3}}$$

∎

In the rest of the cases, equation (2) applies and

$$13/64 \leq \rho - \frac{3(\rho + 1)^2}{16} < \rho\frac{s'}{s} < 2/5$$

Let us first start with **Case (B2-b) (lines 42 to 48)**, and we will show later that other cases are dominated by these ones.

Lemma 13 justifies the choice between performing successively Guillotine and Square (Lines 43 to 45) or Superposition (Lines 47 and 48). In both cases, there are exactly two simple zones: Z_2 and Z_3 in the first case, Z_2' and Z_3' in the second one (see Figure 12).

Lemma 13. *Let R be a rectangle, $\rho = \rho(R)$, α such that $1 - \frac{3(\rho+1)^2}{16\rho} < \alpha < \frac{2}{5\rho}$ and ϵ be such that $0 \leq \epsilon \leq \frac{2\rho\alpha^2}{5}$. Let $R', Z_3 = Guillotine(R, \alpha)$, $R_1, Z_2 = Square(R', \epsilon/\alpha)$ and $R_1', Z_2, Z_3' = Superposition(R, \alpha, \epsilon)$. Then,*

- *If $\epsilon \leq \frac{(1-\sqrt{1-\rho\alpha})^2}{\rho}$ then $c(Z_2, Z_3) \leq c(Z_2', Z_3')$.*
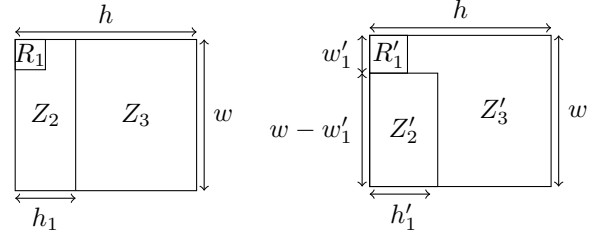- *Otherwise $c(Z_2, Z_3) \geq c(Z_2', Z_3')$.*

Figure 12.   **Case (B2-b)**

Proof: Let us suppose, without loss of generality, that $h = h(R) \geq w(R) = w$. We denote $h_1 = h(Z_2)$, $h_1' = h(Z_2')$ and $w_1' = w(R_1') = h(R_1')$ and, by construction, $w(Z_2) = w(Z_3) = w(Z_3') = w$, $w(Z_2') = w - w_1'$, $h(Z_3) = h - h_1$ and $h(Z_3') = h - w_1'$ (see Figure 12). Trivially,

$$c(Z_2, Z_3) = p(Z_2) + p(Z_3) = h_1 + w + h - h_1 + w = (\rho+2)w$$

and

$$
\begin{aligned}
c(Z_2', Z_3') &= p(Z_2') + p(Z_3') = h_1' + w - w_1' + h - w_1' + w \\
&= (\rho + 2)w + h_1' - 2w_1' \\
&= c(Z_2, Z_3) + h_1' - 2w_1'
\end{aligned}
$$

Thus, $c(Z_2, Z_3) \leq c(Z_2', Z_3') \iff h_1' - 2w_1' \geq 0$. Yet, by definition, $w_1'^2 = s(R_1') = \epsilon hw$, then $w_1' = \sqrt{\rho\epsilon}w$. Moreover, $h_1'(w - w_1') = s(Z_2') = (\alpha - \epsilon)s(R)$. Thus, $h_1' = \frac{\rho(\alpha-\epsilon)}{1-\sqrt{\rho\epsilon}}w$ and finally

$$h_1' - 2w_1' = \left(\frac{\rho(\alpha-\epsilon)}{1-\sqrt{\rho\epsilon}} - 2\sqrt{\rho\epsilon}\right)w.$$

Hence,

$$
\begin{aligned}
c(Z_2, Z_3) \leq c(Z_2', Z_3') &\iff \frac{\rho(\alpha-\epsilon)}{1-\sqrt{\rho\epsilon}} - 2\sqrt{\rho\epsilon}\, gleq 0 \\
&\iff 2\sqrt{\rho\epsilon} \leq \frac{\rho(\alpha-\epsilon)}{1-\sqrt{\rho\epsilon}} \\
&\iff 2\sqrt{\rho\epsilon} - 2\rho\epsilon \leq (\alpha-\epsilon)\rho \\
&\iff 2\sqrt{\rho\epsilon} \leq (\alpha+\epsilon)\rho \\
&\iff 2\sqrt{\epsilon} \leq (\alpha+\epsilon)\sqrt{\rho} \\
&\iff -\epsilon\sqrt{\rho} + 2\sqrt{\epsilon} - \alpha\sqrt{\rho} \leq 0
\end{aligned}
$$

Moreover, $x \mapsto -\sqrt{\rho}x^2 + 2x - \alpha\sqrt{\rho}$ is positive over $[\frac{1-\sqrt{1-\alpha\rho}}{\sqrt{\rho}}, \frac{1+\sqrt{1-\alpha\rho}}{\sqrt{\rho}}]$. Since $\frac{1+\sqrt{1-\alpha\rho}}{\sqrt{\rho}} \geq \frac{1}{\sqrt{\rho}} \geq \sqrt{\frac{2}{5}}$ and $\sqrt{\epsilon} \leq \alpha \leq 2/5$ then, in any case $\sqrt{\epsilon} \leq \frac{1+\sqrt{1-\alpha\rho}}{\sqrt{\rho}}$. So:

$$c(Z_2, Z_3) \leq c(Z_2', Z_3') \iff \sqrt{\epsilon} \leq \frac{1-\sqrt{1-\alpha\rho}}{\sqrt{\rho}}$$
$$\iff \epsilon \leq \frac{(1-\sqrt{1-\rho(R)\alpha})^2}{\rho}$$

what achieves the proof. ∎

In the first case (lines 43 to 45),

$$0 \leq \rho\frac{s''}{s} \leq (1-\sqrt{1-\rho(R)\alpha})^2$$

Thus, Lemma 15 applies (with $\alpha = s'/s$ and $\epsilon = s''/s$, where the bounds on α comes from equation (2)) and proves claimed result.

Lemma 14. *Let* $f(x,y,z) = \frac{x+2}{2(\sqrt{y-z}+\sqrt{x-y})}$. *Then, for* $x \in [1, 5/2]$, $y \in [x - \frac{3(x+1)^2}{16}, 2/5]$ *and* $z \in [0, (1-\sqrt{1-y})^2]$:

$$f(x,y,z) \leq \frac{2}{\sqrt{3}}$$

Proof: Due to lack of space, the proof is presented in Section VII. ∎

Lemma 15. *Let* R *be a rectangle,* $\rho = \rho(R)$ *and* α *such that* $1 - \frac{3(\rho+1)^2}{16\rho} < \alpha < \frac{2}{5\rho}$ *and let* ϵ *be such that* $0 \leq \rho\epsilon \leq (1-\sqrt{1-\rho\alpha})^2$. *Let us denote* $R', Z_3 = Guillotine(R, \alpha)$ *and* $R_1, Z_2 = Square(R', \epsilon/\alpha)$. *Then,*

$$\frac{c(Z_2, Z_3)}{2(\sqrt{s(Z_2)}+\sqrt{s(Z_3)})} \leq \frac{2}{\sqrt{3}}$$

Proof: With the same notations as in the proof of Lemma 13, $c(Z_2, Z_3) = (\rho + 2)w$. Moreover, $s(Z_2) = (\alpha - \epsilon)s(R) = \rho(\alpha-\epsilon)w^2$ and $s(Z_3) = (1-\alpha)s(R) = \rho(1-\alpha)w^2$. Then,

$$\frac{c(Z_2, Z_3)}{2(\sqrt{s(Z_2)}+\sqrt{s(Z_3)})} = \frac{\rho+2}{2(\sqrt{\rho(\alpha-\epsilon)}+\sqrt{\rho(1-\alpha)})}.$$

We can thus use Lemma 14 with $x = \rho$, $y = \alpha\rho$ and $z = \epsilon\rho$, and we obtain the desired result. ∎

In the other case (lines 47 and 48), since

$$(1-\sqrt{1-\rho(R)\alpha})^2 \leq \rho\frac{s''}{s} < \frac{2\rho s'^2}{5s^2},$$

Lemma 17 applies and achieves the proof.

Lemma 16. *For any fixed* x, y *such that* $1 \leq x \leq \frac{5}{2}$ *and* $\frac{1}{5} \leq y \leq \frac{2}{5}$, *consider* $f(z) = \frac{x+2-B(y,z)}{\sqrt{x-y}+\sqrt{y-z}}$, *where* $B(y,z) = 2\sqrt{z} - \frac{y-z}{1-\sqrt{z}}$, *and denote* $z_0 = (1-\sqrt{1-y})^2$ *such that* $B(y, z_0) = 0$. *Then for all* z *such that* $z_0 \leq z \leq \frac{2y^2}{5}$, $f(z) \leq f(z_0)$.

Proof: Due to lack of space, the proof is presented in Section VII. ∎

Lemma 17. *Let* R *be a rectangle with* $\rho = \rho(R)$, *and*
- α *such that* $1 - \frac{3(\rho+1)^2}{16\rho} < \alpha < \frac{2}{5\rho}$
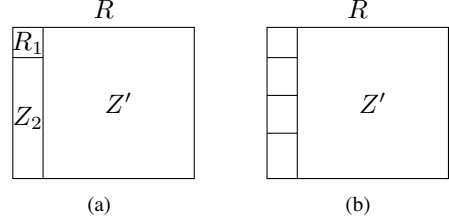

Figure 13. **Case (B2-a1)** and **Case (B2-a2')**

- ϵ *such that* $(1-\sqrt{1-\rho\alpha})^2 \leq \rho\epsilon \leq \frac{2\rho\alpha^2}{5}$
- $R_1, Z_2, Z_3 = Superposition(R, \alpha, \epsilon)$.

Then,

$$\frac{c(Z_2, Z_3)}{2(\sqrt{s(Z_2)}+\sqrt{s(Z_3)})} \leq \frac{2}{\sqrt{3}}$$

Proof: With the same notations as in the proof of Lemma 13 (see on the right of Figure 12),

$$c(Z_2, Z_3) = (\rho+2)w + h_1' - 2w_1'$$
$$= (\rho+2)w + (\frac{\rho(\alpha-\epsilon)}{1-\sqrt{\rho\epsilon}} - 2\sqrt{\rho\epsilon})w$$
$$= (\rho+2+\frac{\rho(\alpha-\epsilon)}{1-\sqrt{\rho\epsilon}} - 2\sqrt{\rho\epsilon})w.$$

Moreover, $s(Z_2) = (\alpha-\epsilon)s(R) = \rho(\alpha-\epsilon)w^2$ and $s(Z_3) = (1-\alpha)s(R) = \rho(1-\alpha)w^2$, so that

$$\frac{c(Z_2, Z_3)}{2(\sqrt{s(Z_2)}+\sqrt{s(Z_3)})} = \frac{\rho+2+\frac{\rho(\alpha-\epsilon)}{1-\sqrt{\rho\epsilon}}-2\sqrt{\rho\epsilon}}{2(\sqrt{\rho(\alpha-\epsilon)}+\sqrt{\rho(1-\alpha)})}.$$

If we set $x = \rho$, $y = \alpha\rho$ and $z = \epsilon\rho$ and $B(y,z) = 2\sqrt{z} - \frac{y-z}{1-\sqrt{z}}$, then we can use Lemma 16 to show that for all x, y the worst case happens when $\epsilon\rho = (1-\sqrt{1-\alpha\rho})^2$, which implies $\frac{\rho(\alpha-\epsilon)}{1-\sqrt{\rho\epsilon}} = 2\sqrt{\rho\epsilon}$. We can thus use Lemma 14 to conclude the proof. ∎

Case (B2-a1) (Lines 23-24) and **Case (B2-a2') (Lines 28-30)** are presented in Figure 13 and **Case (B2-a2'') (Lines 33-36 and 38-40)** are presented in Figure 14.

First, let us note that all four cases contain a large zone, denoted Z', and a set of smaller zones on the side. Among these smaller zones, there is at most one simple zone included in a rectangle whose aspect ratio is larger than $5/2$. Therefore, the other zones are either composed (and thus not considered here) or their aspect ratio is smaller than $5/2$ (which thanks to Lemma 9 yields the desired bound on the cost). The idea of the rest of the proof is to group this small zone with the large zone Z' and to show that their combined cost satisfy the bound.

A general proof can be obtained thanks to the remark that since we do not need to take into account the small zones which are correct, all four cases can be reduced to one case, which is described in Figure 15, where Z_1 is the rectangle whose aspect ratio is too large. Then, a general result expressed in Lemma 18 shows that all of these cases have a better cost ratio than the case (B2b) shown on Figure 12, and this concludes the proof that $\frac{c(Z_1, Z')}{2(\sqrt{s(Z_1)}+\sqrt{s(Z')})} \leq \frac{2}{\sqrt{3}}$.
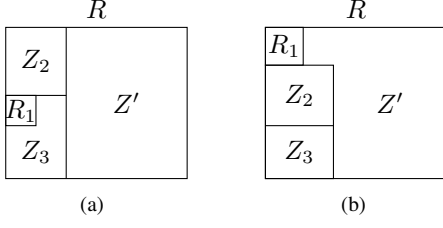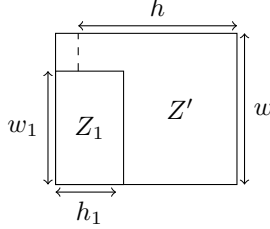
Figure 14. **Case (B2-a2")**

Figure 15. Subproblem studied in Lemma 18.

It only remains to show that all cases satisfy the assumptions of Lemma 18. To show $\rho(Z_1) \geq \frac{4}{3}$, we can actually show $\rho(Z_1) > 21/10$. In cases (B2-a1) and (B2-a2'), the question is not raised anyway unless $\rho(Z_1) > 5/2$. In case (B2-a2"), $\rho(Z_1) \leq 21/10$ would imply that $s_{n-2} \leq s_{n-1} \leq \frac{21s'}{s}$ and then $s''' \geq \frac{2s'}{5s}$, what is a contradiction with the assumption $s''' < \frac{2s'}{5s}$. In addition $s(Z') = s - s' \geq s - s(1 - \frac{3(\rho(R)+1)^2}{16\rho(R)})$ and $s = \rho w^2$. Hence we can indeed use Lemma 18 to solve these cases (we can notice than the case in Figure 14(a) does not exactly reduce like the others, but Lemma 18 and its proof can be adapted to solve it).

Lemma 18. *Let Z_1 and Z_2 be as depicted in Figure 15. If $\rho(Z_1) = \frac{w_1}{h_1} \geq 4/3$ and $\frac{s(Z_2)}{w^2} \geq \frac{3(\rho(R)+1)^2}{16}$, then $\frac{c(Z_1,Z_2)}{2(\sqrt{s(Z_1)}+\sqrt{s(Z_2)})}$ is maximum when $w_1 = w$.*

Proof: Let us denote $x = \rho(Z_1)$. Trivially, $c(Z_1, Z_2) = h_1(\frac{h+w}{h_1} + 1 + x)$ and $\sqrt{s(Z_1)} + \sqrt{s(Z')} = h_1(\frac{\sqrt{s(Z')}}{h_1} + \sqrt{x})$. Therefore, $\frac{c(Z_1,Z')}{2(\sqrt{s(Z_1)}+\sqrt{s(Z')})} = \frac{\frac{h+w}{h_1}+1+x}{2(\frac{\sqrt{s(Z')}}{h_1}+\sqrt{x})} = f(x)$.

$f'(x) = \frac{2\frac{\sqrt{s(Z')}}{h_1}\sqrt{x}+x-\frac{h+w}{h_1}+1}{4\sqrt{x}(\frac{\sqrt{s(Z')}}{h_1}+\sqrt{x})^2}$ and therefore $f'(x) \geq 0$ is equivalent to $2\frac{\sqrt{s(Z')}}{h_1}\sqrt{x} + x - (\frac{h+w}{h_1} + 1) \geq 0$. Since $x \geq 1$, we have to prove that $2\sqrt{s(Z')}\sqrt{x} - (h + w) \geq 0$. Let us denote $\rho = \rho(R)$, so that $h + w = (\rho+1)w$. Let us prove that $2\sqrt{s(Z')}\sqrt{x} - (\rho + 1)w \geq 0$.

$$2\sqrt{s(Z')}\sqrt{x} - (\rho + 1)w \geq 0 \Longleftrightarrow 2\sqrt{s(Z')}\sqrt{x} \geq (\rho + 1)w$$
$$\Longleftrightarrow 4s(Z')x \geq (\rho + 1)w^2$$
$$\Longleftrightarrow x\frac{s(Z')}{w^2} \geq \frac{(\rho + 1)^2}{4}$$

Since we assumed that $x \geq 4/3$ and $\frac{s(Z')}{w^2} \geq \frac{3(\rho)+1)^2}{16}$, we have $x\frac{s(Z')}{w^2} \geq \frac{(\rho+1)^2}{4}$ and we obtain claimed result.

Therefore, f is increasing with x, and, when h_1 is fixed, the cost ratio is maximum when $w_1 = w$. ∎

This achieves the proof of all possible cases of NRRP and therefore the proof of Theorem 6. ∎

Note that our bound is tight since there are cases where the ratio between the sum of perimeters and the lower bound based on Equation 1 is indeed $\frac{2}{\sqrt{3}}$. Let us define for $1 < k \leq n$, $s_k = 3/4^{n-k+1}$ and $s_1 = 1/4^{n-1}$. One can notice that $s_k = \frac{3}{4}\sum_{i=1}^{k} s_i$. Hence each step of $NRRP([0,1]^2, \{s_1, \ldots, s_n\})$ corresponds to **Case (B1)** and therefore the Square routine is called. In addition, this corresponds the extremal position of **Case (B1)** and we can notice in the proof of Lemma 12 that in this case the bound is tight. Hence, if $Z_1, \ldots, Z_n = NRRP([0,1]^2, \{s_1, \ldots, s_n\})$, $c(Z_1) = 2\sqrt{s(Z_1)}$ and for $k > 1$, $c(Z_k) = \frac{4}{\sqrt{3}}\sqrt{s(Z_k)}$. Thus,

$$\lim_{n \to \infty} \frac{c(Z_1, \ldots, Z_n)}{2\sum s(Z_i)} = \frac{2}{\sqrt{3}}$$

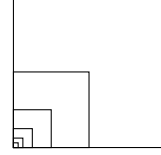An illustration of this case is depicted in Figure 16.

Figure 16. An illustration of the worst possible case.

V. EXPERIMENTAL EVALUATION

We present in this section some experimental results in order to assess the practical performance of our algorithm. In contrast to previous work, we consider strongly heterogeneous platforms made of different kinds of processing units (CPU cores, GPUs and Xeon Phi), which is a common situation in today's computers. According to recent benchmarks from the literature, as far as matrix multiplication is concerned, GPUs such as the Nvidia Tesla are roughly 30 times faster than a standard core, and Xeon Phi accelerators are about 20 times faster than a core. In our experiments, we thus generate instances with n cores, m accelerators and p GPUs, where the speed of accelerators is randomly chosen between 15 and 25, and the speed of GPUs is chosen between 25 and 35, so as to cover a wide range of values. We use values of n in $\{1, 2, 4, 8, 12, 16, 24, 32, 64\}$, and m and p vary from 0 to 8, and we generate 10 random instances for all values of n, m and p. On these instances, we analyze the results of our NRRP algorithm together with two previous solutions from the literature, namely Nagamochi's algorithm [5] and the Column-based approach from [8], and we compute the ratio of the partition returned by each algorithm to the lower bound from equation (1).

The results we obtain show that the Column-Based algorithm returns very efficient partitionings for large enough platforms (namely, when the number of CPUs is larger than 16), with ratios consistently below 1.05, and with a median

value below 1.02. However, the performance is much lower for smaller platforms, and ratios as high as 1.5 are observed on platforms with few CPUs. The Nagamochi and NRRP algorithms achieve better results than Column Based on small platforms, but are unable to compete for larger platform sizes. The most difficult cases correspond to very small platforms, with a few processing units. In this case, NRRP always achieves an approximation ratio below 1.1, whereas the ratios obtained by Nagamochi can be as large as 1.3. Note that this is higher than the approximation ratio proved in [5], but it is not a contradiction because in the setting of the proof of Nagamochi et al., a refined lower bound is used for smaller instances. Except for these very small cases, the behaviors of Nagamochi and NRRP algorithms are fairly comparable, but can be quite different depending on the instance. In summary, worst case values are 1.106 for NRRP, 1.3 for Nagamochi, and 1.526 for Column Based.

This complementary behavior of all three algorithms is an incentive to compute all three partitionings on a given instance, and then return the one with the lowest communication cost. As expected this simple heuristic provides very efficient solutions in all cases. Indeed, on average, the solutions are within 2% of the lower bound over all ranges of values, and at most within 8% of the (optimistic) lower bound. For comparison, the same heuristic without NRRP has a maximum ratio of 30%.

VI. Conclusion

In this paper, we propose a new algorithm NRRP for partitioning a square into a set of zones of prescribed area while minimizing the perimeter of the zones. This optimization problem is of crucial importance when a 2D computational domain has to be split into parts that will be allocated to heterogeneous processing resources, while minimizing the overall amount of communications. This work therefore applies naturally to Matrix Multiplication and other linear algebra kernels on heterogeneous (CPU, GPU and other accelerators) platforms. Such static partitionings are more flexible than block-cyclic schemes, and can be used as a basis for more dynamic strategies to obtain very good performance even in noisy environments where execution times cannot be perfectly estimated. Due to its practical importance, it has been the subject of a vast literature. Combining ideas proposed by Lastovetsky et al. on the one hand and Nagamochi et al. on the other hand, we have been able to improve the approximation ratio from 1.25 to 1.15. Even if the algorithm and the proof are non trivial, it is worth noticing that the overall complexity of the algorithm is low, $i.e.$ $p \log p$. Experimental analysis on realistic instances show that NRRP is very efficient on difficult instances and that combining several heuristics together yields to very good practical solutions, $i.e.$ on average 2% larger and at worst 8% larger than the optimal solution, on a large set of realistic platforms consisting of both CPUs and accelerators. This work opens several perspectives. First, we conjecture that a lower approximation ratio could be proved using a different partitioning technique, but this would require to rely on different proof techniques, given the complexity of the actual proof. Second, the extension to 3D computational domain is relevant and open, and would encompass the 2.5D algorithms that have recently been proposed.

References

[1] A. Kalinov and A. Lastovetsky, "Heterogeneous distribution of computations solving linear algebra problems on networks of heterogeneous computers," *Journal of Parallel and Distributed Computing*, vol. 61, no. 4, pp. 520–535, 2001.

[2] O. Beaumont, V. Boudet, F. Rastello, Y. Robert *et al.*, "Partitioning a square into rectangles: Np-completeness and approximation algorithms," *Algorithmica*, vol. 34, no. 3, pp. 217–239, 2002.

[3] B. Becker, A. Lastovetsky *et al.*, "Towards data partitioning for parallel computing on three interconnected clusters," in *Parallel and Distributed Computing, 2007. ISPDC'07. Sixth International Symposium on*. IEEE, 2007, pp. 39–39.

[4] A. DeFlumere and A. Lastovetsky, "Optimal data partitioning shape for matrix multiplication on three fully connected heterogeneous processors," in *Euro-Par 2014: Parallel Processing Workshops*. Springer, 2014, pp. 201–214.

[5] H. Nagamochi and Y. Abe, "An approximation algorithm for dissecting a rectangle into rectangles with specified areas," *Discrete Applied Mathematics*, vol. 155, no. 4, pp. 523 – 537, 2007.

[6] A. Fügenschuh, K. Junosza-Szaniawski, and Z. Lonc, "Exact and approximation algorithms for a soft rectangle packing problem," *Optimization*, vol. 63, no. 11, pp. 1637–1663, 2014.

[7] O. Beaumont, A. Legrand, F. Rastello, and Y. Robert, "Static lu decomposition on heterogeneous platforms," *International Journal of High Performance Computing Applications*, vol. 15, no. 3, pp. 310–323, 2001.

[8] O. Beaumont, V. Boudet, A. Petitet, F. Rastello, and Y. Robert, "A proposal for a heterogeneous cluster scalapack (dense linear solvers)," *Computers, IEEE Transactions on*, vol. 50, no. 10, pp. 1052–1070, 2001.

[9] L. S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet *et al.*, *ScaLAPACK users' guide*. siam, 1997, vol. 4.

[10] D. Clarke, A. Ilic, A. Lastovetsky, and L. Sousa, "Hierarchical partitioning algorithm for scientific computing on highly heterogeneous cpu+ gpu clusters," in *Euro-Par 2012 Parallel Processing*. Springer, 2012, pp. 489–501.

[11] R. Shams and P. Sadeghi, "On optimization of finite-difference time-domain (fdtd) computation on heterogeneous and gpu clusters," *Journal of Parallel and Distributed Computing*, vol. 71, no. 4, pp. 584–593, 2011.

[12] N. Mohamed, J. Al-Jaroodi, and H. Jiang, "Ddops: dual-direction operations for load balancing on non-dedicated heterogeneous distributed systems," *Cluster Computing*, vol. 17, no. 2, pp. 503–528, 2014.

[13] C. Augonnet, S. Thibault, R. Namyst, and P.-A. Wacrenier, "StarPU: A Unified Platform for Task Scheduling on Heterogeneous Multicore Architectures," *Concurrency and Computation: Practice and Experience, Special Issue: Euro-Par 2009*, vol. 23, pp. 187–198, Feb. 2011. [Online]. Available: http://hal.inria.fr/inria-00550877

[14] J. Planas, R. M. Badia, E. Ayguadé, and J. Labarta, "Hierarchical task-based programming with StarSs," *International Journal of High Performance Computing Applications*, vol. 23, no. 3, pp. 284–299, 2009.

[15] A. YarKhan, J. Kurzak, and J. Dongarra, *QUARK Users' Guide: QUeueing And Runtime for Kernels*, UTK ICL, 2011.

[16] G. Bosilca, A. Bouteiller, A. Danalis, M. Faverge, T. Hérault, and J. Dongarra, "PaRSEC: A programming paradigm exploiting heterogeneity for enhancing scalability," *Computing in Science and Engineering*, vol. 15, no. 6, pp. 36–45, Nov. 2013.

[17] O. Beaumont, L. Eyraud-Dubois, A. Guermouche, and T. Lambert, "Hierarchical partitioning algorithm for scientific computing on highly heterogeneous cpu+ gpu clusters," in *Proceedings of the 26th IEEE International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*. IEEE, 2015, pp. 1–10.

[18] G. Ballard, J. Demmel, O. Holtz, and O. Schwartz, "Minimizing communication in linear algebra," *SIAM Journal on Matrix Analysis and Applications*, vol. 32, no. 3, pp. 866–901, Jul. 2011, arXiv: 0905.2485. [Online]. Available: http://arxiv.org/abs/0905.2485

VII. Annexe

A. Proof of Lemma 10

First let us prove that f is decreasing in y and z.

$$\frac{\partial f}{\partial y}(x,y,z) = \frac{2\sqrt{y}(\sqrt{y}+\sqrt{z}) - (1-z)(x+1) - y}{4\sqrt{xy}(1-z)(\sqrt{(y)}+\sqrt{z})^2}$$

Therefore, $\frac{\partial f}{\partial y}(x,y,z) \le 0$ is equivalent to $2\sqrt{y}(\sqrt{y}+\sqrt{z}) - (1-z)(x+1) - y \le 0$.

$$
\begin{aligned}
2\sqrt{y}(\sqrt{y}+\sqrt{z}) - (1-z)(x+1) - y &= 2y + 2\sqrt{yz} - x - y - 1 + xz + z \\
&= y + z + 2\sqrt{yz} + zx - x - 1 \\
&= 2\sqrt{yz} - (1-\gamma)x - \alpha
\end{aligned}
$$

As $x \ge 1$ and $y \le z$, we have:

$$
\begin{aligned}
2\sqrt{y}(\sqrt{y}+\sqrt{z}) - (1-z)(x+1) - y &\le 2z - +z - \alpha \\
&\le 3z - 1 - \alpha
\end{aligned}
$$

Yet $z \le \frac{2}{5\rho} \le 6/5$ and $\alpha \ge 1/5$, hence:

$$2\sqrt{y}(\sqrt{y}+\sqrt{z}) - (1-z)(x+1) - y \le 6/5 - 1 - 1/5 = 0$$

and thus $\frac{\partial f}{\partial y}(x,y,z) \le 0$.

$$\frac{\partial f}{\partial z}(x,y,z) = \frac{2z^{3/2}(\sqrt{y}+\sqrt{z}) - (1-z)^2(x+1) - (1-z)y}{2\sqrt{xz}(1-z)^2(\sqrt{(y)}+\sqrt{z})^2}$$

So $\frac{\partial f}{\partial z}(x,y,z) \le 0$ is equivalent to $2z^{3/2}(\sqrt{y}+\sqrt{z}) - (1-z)^2(x+1) - (1-z)y \le 0$.

$$
\begin{aligned}
2z^{3/2}(\sqrt{y}+\sqrt{z}) - (1-z)^2(x+1) - (1-z)y &= 2z^2 + 2z\sqrt{yz} - x - 1 + 2zx + 2z - xz^2 - z^2 - y + xy \\
&\le 4z^2 - x - 1 + 2xz + 2z - xz^2 - z^2 - y - z^2 \\
&\le 4z^2 - x - 1 + 2xz + 2z - xz^2 - y \\
&\le 3z^2 - 2 + 2xz + 2z - y
\end{aligned}
$$

As we have $xz \le 2/5$:

$$
\begin{aligned}
2z^{3/2}(\sqrt{y}+\sqrt{z}) - (1-z)^2(x+1) - (1-z)y &\le 3z^2 - 2 + 4/5 + 2z - y \\
&\le 12/25 + 4/5 - 1/5 - 6/5 \\
&\le -3/25
\end{aligned}
$$

Thus $\frac{\partial f}{\partial z}(x,y,z) \le 0$.

Then $f(x,y,z) \le f(x, \frac{5x-4}{5x}, \frac{5\rho-2}{10\rho})$. We denote $g(x) = f(x, \frac{5x-4}{5x}, \frac{5\rho-2}{10\rho})$ and one can prove that

$$g(x) = \frac{\sqrt{5}(x(5x+17) - 6)}{(5x+2)(2\sqrt{5x-4} + \sqrt{10x-4})}$$

We are interested in proving that $g(x)$ is decreasing when x varies in $[1, 6/5]$.
First,

$$
\begin{aligned}
g'(x) = {}& \frac{\sqrt{(5)}(10x+17)}{(5x+2)(2\sqrt{(5x-4)} + \sqrt{(10x-4)})} \\
& - \frac{5\sqrt{(5)}(x(5x+17) - 6)}{(5x+2)^2(2\sqrt{(5x-4)} + \sqrt{(10x-4)})} \\
& - \frac{\sqrt{(5)}(x(5x+17) - 6)(\frac{5}{\sqrt{(10x-4)}} + \frac{5}{\sqrt{(5x-4)}})}{(5x+2)(2\sqrt{(5x-4)} + \sqrt{(10x-4)})^2}
\end{aligned}
$$

so that $g'(x) \leq 0 \iff$

$$(25x^2 + 20x + 64) - \frac{5(x(5x+17)-6)(5x+2)(\frac{1}{\sqrt{(10x-4)}} + \frac{1}{\sqrt{(5x-4)}}))}{(2\sqrt{(5x-4)} + \sqrt{(10x-4)})} \leq 0$$

Moreover $5x + 2$ is increasing with x in the interval $[1, 6/5]$, so that $5x + 2 \geq 7$. To finish the proof, let us denote $A(x) = \frac{(\frac{1}{\sqrt{(10x-4)}} + \frac{1}{\sqrt{(5x-4)}})}{(2\sqrt{(5x-4)} + \sqrt{(10x-4)})}$ Clearly, A(x) is i decreasing with x in the interval $[1, 6/5]$, so that in particular in the interval $[1, 11/10]$, $A(x) \geq A(11/10)$ and in the interval $[11/10, 6/5]$, $A(x) \geq A(6/5)$.

Therefore, in the interval $[1, 11/10]$,

$$B(x) = (25x^2 + 20x + 64) - \frac{(5(x(5x+17)-6)(5+2)(\frac{1}{\sqrt{(10*11/10-4)}} + \frac{1}{\sqrt{(5*11/10-4)}}))}{(2\sqrt{(5*11/10-4)} + \sqrt{(10*11/10-4)}))} \leq 0 \implies g'(x) \leq 0$$

. Moreover, B is a polynomial of degree 2 that tends to $-\infty$ when x tends to ∞, that is equal to $40\sqrt{(42)} - 146 > 0$ in 0 and to $669 - 320\sqrt{(14/3)} < 0$ in 1 so that $B(x)$ is negative in the interval $[1, 11/10]$

Similarly, in the interval $[11/10, 6/5]$,

$$C(x) = (25x^2 + 20x + 64) - \frac{(5(x(5x+17)-6)(5+2)(\frac{1}{\sqrt{(10*6/5-4)}} + \frac{1}{\sqrt{(5*6/5-4)}}))}{(2\sqrt{(5*6/5-4)} + \sqrt{(10*6/5-4)})} \leq 0 \implies g'(x) \leq 0$$

. As previously, C is a polynomial of degree 2 that tends to $-\infty$ when x tends to ∞, that is equal to $827/8 > 0$ in 0 and to $-435/64 < 0$ in $11/10$ so that $B(x)$ is negative in the interval $[11/10, 6/5]$

Therefore $f(x, y, z) \leq g(x) \leq g(1) = \frac{16\sqrt{5}}{7(2+\sqrt{6})} < \frac{2}{\sqrt{3}}$ and we have our result.

$\square$

B. Proof of Lemma 14

First it is easy to see that f is increasing in z. Therefore, for all x, y:

$$f(x, y, z) \leq f(x, y, (1 - \sqrt{1-y})^2)$$
$$\leq \frac{x+2}{2(\sqrt{y - (1 - \sqrt{1-y})^2} + \sqrt{x - y})}$$
$$\leq \frac{x+2}{2(\sqrt{y - 1 + 2\sqrt{1-y} - (1 - y)} + \sqrt{x - y})}$$
$$\leq \frac{x+2}{2(\sqrt{2(\sqrt{1-y} - (1 - y))} + \sqrt{x - y})}$$

In the following we denote $z_{max} = (1 - \sqrt{1-y})^2$.

Let us consider $g_x(y) = \sqrt{2(\sqrt{1-y} - (1-y))} + \sqrt{x - y}$. Successive derivations give us

$$g_x''(y) = -\frac{1}{4(x-y)^{3/2}} - \frac{(1 - \frac{1}{2\sqrt{1-y}})^2}{2\sqrt{2}(\sqrt{1-y} - (1-y))^{3/2}} - \frac{1}{4\sqrt{2}(1-y)^{3/2}\sqrt{\sqrt{1-y} - (1-y)}}$$

Therefore, on $y \in [x - \frac{3(x+1)^2}{16}, 2/5]$, $g_x(y) \leq 0$ and g_x is concave.

Then for all $y \in [x - \frac{3(x+1)^2}{16}, 2/5]$, $g_x(y) \geq min(g_x(x - \frac{3(x+1)^2}{16}), g_x(2/5))$.

Let us first suppose $min(g_x(x - \frac{3(x+1)^2}{16}), g_x(2/5)) = g_x(2/5)$. Then:

$$f(x, y, z) \leq \frac{x+2}{2(\sqrt{2(\sqrt{1-y} - (1-y))} + \sqrt{x - y})}$$
$$\leq \frac{x+2}{2g_x(y)}$$
$$\leq \frac{x+2}{2g_x(2/5)}$$
$$\leq \frac{x+2}{2(\sqrt{x - 2/5} + \sqrt{2}\sqrt{\sqrt{3/5} - 3/5}})$$

Let us denote $A = \sqrt{2}\sqrt{\sqrt{3/5} - 3/5}$. Let us prove that $G_1(x) = \frac{x+2}{2(\sqrt{x-2/5}+A)}$ is inferior or equal to $\frac{2}{\sqrt{3}}$ for $x \in [1, 5/2]$. One can prove that

$$G_1'(x) = \frac{x - 14/5 + 2A\sqrt{x - 2/5}}{4\sqrt{x - 2/5}(\sqrt{x - 2/5} + A)^2}$$

Then $G_1'(x) \geq 0$ is equivalent to $x - 14/5 + 2A\sqrt{x - 2/5} \geq 0$. As $\rho \leq 14/5$ it is equivalent to $2A\sqrt{x - 2/5} \geq 14/5 - x$ which is equivalent to $4A^2(x - 2/5) \geq 196/25 - 28/5x + x^2$. By replacing A^2 by $2(\sqrt{3/5} - 3/5)$ we obtain:

$$G_1'(x) \geq 0 \iff -x^2 + (8\sqrt{\frac{3}{5}} + \frac{4}{5})x - (\frac{16\sqrt{3}}{5\sqrt{5}} + \frac{148}{25}) \geq 0$$

One can prove that in our case this is equivalent to $x \geq \frac{2(1+2\sqrt{15})}{5} - \frac{4\sqrt{6}}{5} > 1$ and therefore G_1 is decreasing on $[1, \frac{2(1+2\sqrt{15})}{5} - \frac{4\sqrt{6}}{5}]$ and increasing on $[\frac{2(1+2\sqrt{15})}{5} - \frac{4\sqrt{6}}{5}, 5/2]$. Therefore $G_1(x) \leq max(G_1(1), G_1(5/2))$. As we have

$$G_1(1) = \frac{3\sqrt{5}}{2(\sqrt{3} + \sqrt{2(\sqrt{15} - 3)})} < \frac{2}{\sqrt{3}}$$

$$G_1(5/2) = \frac{9\sqrt{5}}{2\sqrt{2}(\sqrt{21} + 2\sqrt{(\sqrt{15} - 3)})} < \frac{2}{\sqrt{3}}$$

We can deduce that $G_1(x) \leq \frac{2}{\sqrt{3}}$ and, if $min(g_x(x - \frac{3(x+1)^2}{16}), g_x(2/5)) = g_x(2/5)$, then $f(x, y, z) \leq \frac{2}{\sqrt{3}}$.

Else, we now suppose $min(g_x(x - \frac{3(x+1)^2}{16}), g_x(2/5)) = g_x(x - \frac{3(x+1)^2}{16})$. Then:

$$f(x, y, z) \leq \frac{x + 2}{2(\sqrt{2(\sqrt{1-y} - (1-y))} + \sqrt{x - y})}$$

$$\leq \frac{x + 2}{2g_x(y)}$$

$$\leq \frac{x + 2}{2g_x(x - \frac{3(x+1)^2}{16})}$$

$$\leq \frac{x + 2}{2(\sqrt{\frac{3(x+1)^2}{16}} + \sqrt{2(\sqrt{1-y} - (1-y))})}$$

$$\leq \frac{2}{\sqrt{3}} \times \frac{x + 2}{x + 1 + \frac{4\sqrt{2}}{\sqrt{3}}\sqrt{\sqrt{1-y} - (1-y)}} = G_2(x)$$

Hence :

$$G_2(x) \leq \frac{2}{\sqrt{3}} \iff x + 2 \leq x + 1 + \frac{4\sqrt{2}}{\sqrt{3}}\sqrt{\sqrt{1-y} - (1-y)}$$

$$\iff 1 \leq \frac{4\sqrt{2}}{\sqrt{3}}\sqrt{\sqrt{1-y} - (1-y)}$$

$$\iff 1 \leq \frac{32}{3}(\sqrt{1-y} - (1-y))$$

If we denote $X = \sqrt{1-y}$, then:

$$G_2(x) \leq \frac{2}{\sqrt{3}} \iff 1 \leq \frac{32}{3}(X - X^2)$$

$$\iff -\frac{32}{3}X^2 + \frac{32}{3}X - 1 \geq 0$$

One can show that this is equivalent to

$$\frac{1}{2} - \frac{\sqrt{10}}{8} \leq X \leq \frac{1}{2} + \frac{\sqrt{10}}{8}$$

and furthermore, this is equivalent to:

$$1 - (\frac{1}{2} + \frac{\sqrt{10}}{8})^2 \leq y \leq 1 - (\frac{1}{2} - \frac{\sqrt{10}}{8})^2$$

As $1 - (\frac{1}{2} + \frac{\sqrt{10}}{8})^2 \leq 13/64 \leq y$ and $1 - (\frac{1}{2} - \frac{\sqrt{10}}{8})^2 \geq 2/5 \geq y$, then for all x, $G_2(x) \leq \frac{2}{\sqrt{3}}$ and we have our result.

So, in any case, we prove that $f(x, y, z) \leq \frac{2}{\sqrt{3}}$.

$\square$

C. Proof of Lemma 16

Computing $f(z_0) - f(z)$ yields

$$f(z_0) - f(z) = \frac{(x+2)(\sqrt{x-y} + \sqrt{y-z}) - (x+2-B(z))(\sqrt{x-y} + \sqrt{y-z_0})}{(\sqrt{x-y} + \sqrt{y-z})(\sqrt{x-y} + \sqrt{y-z_0})}$$

Hence $f(z_0) - f(z)$ has same sign as $g(z) = B(z)(\sqrt{x-y} + \sqrt{y-z_0}) - (x+2)(\sqrt{y-z_0} - \sqrt{y-z})$. We note $C(x, y) = \sqrt{x-y} + \sqrt{y-z_0}$, so that we can write the derivative $g'(z) = B'(z)C(x, y) - \frac{x+2}{2\sqrt{y-z}}$.

We first note that, given the bounds on x and y, $C(x, y) \geq \sqrt{x-y} \geq \sqrt{\frac{3}{5}}$, and $x + 2 \leq \frac{7}{2}$. Since we have $z \leq \frac{2y^2}{5}$, we get $y - z \geq y(1 - \frac{2y}{5}) \geq \frac{1}{5}(1 - \frac{4}{25}) \geq \frac{1}{6}$, which yields $\frac{1}{2\sqrt{y-z}} \leq \frac{\sqrt{6}}{2}$.

Similarly, the bounds on y provide the following bounds on $\sqrt{z}$:

- $\sqrt{z} \geq \sqrt{z_0} = 1 - \sqrt{1-y} \geq 1 - \sqrt{\frac{4}{5}}$ because $y \geq \frac{1}{5}$
- $\sqrt{z} \leq y\sqrt{\frac{2}{5}} \leq \zeta = \frac{2}{5}\sqrt{\frac{2}{5}}$ because $y \leq \frac{2}{5}$.

Computing $B'(z)$ gives $B'(z) = \frac{2-y-\sqrt{z}(2-\sqrt{z})}{\sqrt{z}(1-\sqrt{z})^2} = \frac{2-y}{\sqrt{z}(1-\sqrt{z})^2} - \frac{2-\sqrt{z}}{(1-\sqrt{z})^2}$. Simple analysis shows that both $u \mapsto \frac{1}{u(1-u)^2}$ and $u \mapsto -\frac{2-u}{(1-u)^2}$ are decreasing on $I = [1 - \sqrt{\frac{4}{5}}, \zeta]$. This implies that $B'(z) \geq B'(\zeta^2)$, which together with $y \leq \frac{2}{5}$ yields $B'(z) \geq \frac{2 - \frac{2}{5} + \zeta(2-\zeta)}{\zeta(1-\zeta)^2} \geq 8$.

Putting all together provides $g'(z) \geq 8.\sqrt{\frac{3}{5}} - \frac{9}{2}.\frac{\sqrt{6}}{2} \geq 0$. Hence g is an increasing function of z, and since $g(z_0) = 0$ by construction, we get $f(z_0) - f(z) \geq 0$ for all $z \geq z_0$.

$\square$