

# UNIVERSITÀ DEGLI STUDI DI PADOVA

**Dipartimento di Fisica e Astronomia "Galileo Galilei"**

**Master Degree in Physics**


**Final Dissertation**

# Tensor Network Methods for Radiotherapy

# Dose Optimization


**Thesis supervisor**

**Prof. Simone Montangero**

**Thesis co-supervisor**

**Dr. Marta Paiusco**

**Candidate**

**Samuele Cavinato**


**Academic Year 2018/2019**

*Dedicated to all people fighting*
*their own battle against cancer.*

*Luigi, Loredana, Silvia,*
*Il Branco, Sid & Snoopy*

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# Abstract

The very first successful application of tensor network methods (TNMs) to the solution of the dose optimization problem in radiotherapy is presented. The particular case of the *inverse planning* optimization in I*ntensity Modulated Radiation Therapy* (IMRT) is considered. This technique provides a method to modulate the local beam intensities, dividing the beam into smaller beamlets. This allows to reduce the radiation toxicity for healthy organs and deal with irregular and inhomogeneous tumors. Plan's goals are encoded as mathematical constraints into a cost function expressing the distance between the prescribed and delivered dose. Aim of the optimization is to minimize the function associating to each beamlet the optimal weight $x_j$. In this thesis a classical quadratic cost function is mapped into an Ising-like Hamiltonian, where the beamlets weights are described by a system of long-range interacting qubits. The aim of the work is to solve the dose optimization problem using a binary tree tensor network (bTTN) to find the Hamiltonian's ground-state and show the applicability of TNM to the IMRT dose optimization problem.

# 1 INTRODUCTION

At the end of 2018 the Internation Agency for Reasearch on Cancer (IARC) released the latest estimates on the global burden of cancer [38], providing results about incidence and mortality of 36 types of cancer in 185 different countries. The data are baffling: one in five men and one in six women worldwide develop cancer during their lifetime; one in eight men and one eleven women die from the desease. The general increasing trend may be due to different factors from population growth to social and economic developement. On the other side, an incredible effort has been made in the past years, expecially in developed countries, in order to provide new tools to detect cancers in their early stages and to treat it effectively .

In this thesis, we'll focus our attention on a specific kind of cancer treatment: the radiotherpy, which aims at killing cancer cells by means of ionizing radiation. Radiotherapy plays a key role in many treatments and it has become fundamental in about the 50% of cancer treaments wordwide [10].

The most difficult task of treating patients with ionizing radiations, which is the focus of most of nowadays studies in this particular field, is to reduce as much as possible the radiation toxicity for healthy tissues. Critical organs placed in proximity of the tumor to treat could get damaged if exposed to too high radiation doses. This issue becomes particularly crictial when the radioactive source is placed outside the patient body, a technique called *external radiotherapy*.

The particle beam (electrons, photons or adrons) produced by means of a particles accelerator has to pass through the patient's body, potentially inducing a damage on the healthy tissues' cells it finds on its path. This calls for to the necessity of performing an accurate planning procedure in order to maximize

the damage to cancer cells, while sparing healthy tissues.

Among the different kind of radiotherapy treatments, the most promising from this point of view is the so-called *Intensity Modulated Radiation Therapy* (IMRT) [45]. It is an evolution of the standard 3D conformal radiotherapy, providing control also on the local beam intensity (beam fluence). Thanks to its high precision, it's very suited to treat tumors with very irregular shapes and very close to organs whose radiosensitivity is a critical parameter to account for in the treatment.

In IMRT, each beam is subdivided into grid of pencil-beams, called *beamlets*. Each beamlet's intensity can be tuned independently from all the others' in order to produce the optimal *fluence map*, a process called *fluence map optimization* (FMO). Unfortunately due to the high number of beamlets typically involved in a treatment optimization, this cannot be done by hand. For this reason, the other main feature distinghuishing IMRT from other techniques is that the former is *inverse planned*: the treatment goals are encoded as mathematical constraints into a cost function expressing the distance between the prescribed and the delivered dose. The optimization of a treatment plan becomes the problem of optimizing a cost function, being the beamlets intensity $x_j$ the variables.

In nowadays cilinal applications, this is done by means of classical algoriithms based on different techniques [49, 39, 17, 7, 50, 11, 25]. In 2015, D.P. Nazareth and J.D. Spaans [28] tried to reformulate the problem and solve it exploiting quantum annealing on a quantum annealer with 512 qubits produced by D-Wave System. They compared the result to those obtained using simulated annealing and Tabu search, pointing out that quantum annealing could outperform classical techniques in terms of time performances, even though non definitive results were provided because of the limited number of beamelts used.

In this thesis, we attack this problem using *Tensor Network Methods* (TNMs), a widespread class of algorithms used to simulate quantum many-body systems on classical computers. TNMs give systematic methods to *compress* the information contained in a quantum state in order to transform, when possible,

exponentially scaling problems into polinomial ones.
We started from a classical quadratic cost function describing the difference between the prescribed and the delivered dose and implemented *ex-novo* a method to adapt it to the TNM algorithm architecture. In particular, the cost function was mapped into an Ising-like Hamiltonian describing a system of long-range interacting qubits. Each beamlet's weight is represented by $N_Q$ qubits in the lattice and the classical dose-optimization problem becomes the problem of finding the Hamiltonian's ground state, a task which typically can be efficiently solved with tensor networks.
The final aim of the work is to investigate the applicability of TNM to the solution of the dose optimization problem in IMRT in order to probe if it can be considered as a potential alternative to standard optimization methods.
In the following lines, a brief outline of the chapters is reported, in order to help the reader to follow the thread of the discussion through the pages.

**CHAPTER 1 – RADIOTHERAPY.** A general introduction to is presented, in order to provide all necessary information to follow the following chapters. Particular attention is put on the treatment planning process, in order to underline which are the difficulties it hides. In the last part, a detailed description of the IMRT and the inverse planning technique is provided, focusing on the particular cost function used in this work.

**CHAPTER 2 – QUANTUM MANY-BODY SYSTEMS.** It presents a brief overview on the essential concepts of many-body theory related to tensor network methods. In the first section quantum-lattice systems are described, while in the second an introduction to the description of entanglement in many-body quantum systems we'll be provided. In particular, we'll focus on key role it plays in the description of many-body states in the tensor networks framework.

**CHAPTER 3 – TENSOR NETWORK METHODS.** Tensor network methods are described from a very general point of view, with the aim of highlighting how

the can be used to efficiently represent operations among tensors. The process of the Hamiltonian's ground state search via tensor networks is described. In the last part, an introduction to the binary tree tensor networks anstaz is presented, since it is the one implemented in the algorithm used in this work.

**CHAPTER 4 – STUDY AND RESULTS.** Starting from the mapping of the classical quadratic function into an Ising-like Hamiltonian we'll pass through all the several intermediate steps of the study. The final aim of the chapter is to accurately describe all the results which come out from this study and show that TNMs can be used to solve the dose optimization problem in IMRT.

# 2 RADIOTHERAPY

## 2.1 THE COMPLEXITY BEHIND RADIOTHERAPY

Radiotherapy is a technique used to cure or control solid tumor by means of different kind of ionizing radiation. The dose released into the cancer tissue by the interaction of the ionizing radiation with the matter damages the DNA of the tumor cells leading them to death or slowing down their growth [4, 47, 3]. Radiotherapy is often used in combination with other therapy like surgery, chemiotherapy or immunotherapy in order to improve the global efficacy of the therapy.

- **Internal radiotherapy.** A radioactive source is placed directly inside the patient's body. The source can be either solid (*brachytherapy*) or liquid (*systemic therapy*). The former consists in the placement of seeds, capsules or ribbons containing the radioactive source inside or very close to the tumor. This placement can be either temporary or permanent. Brachytherapy is part of the so-called *local methods*, which are all the methods involving only the specific part of the body where the tumor is situated. On the other end, with systemic therapy a liquid radioactive source travels in the blood or tissues, seeking out and killing cancer cells. This method is not local.

- **External beam radiotherapy (EBT).** External particles beam are produced by a linear accelerator (LINAC) and directed towards the region where the tumor is. The particles used may be either electrons or high energy X-rays (photons). In general, the patient is irradiated from multiple direc-

tions, in order to increase as much as possible the damage to the tumor, while sparing the sorrounding healty tissues. This is a local method too.

The choice of the most suitable radiotherapy treatment depends on many different factors: the type of cancer, the size of the tumor, the tumor's location in the body, how close the tumor is to normal tissues whose radiosensitivity cannot be neglected, the health and medical history of the patient, the concomintance with other therapy treatments and other factors, such as the age and the medical conditions.
It follows that scheduling a radiotherapy treatment is not a trivial task: different figures are involved, from the oncologist to the medical physicist to the radiology technician. The number of beams, their intensity and spatial geometry, the number of part the treatment is subdivided are only some of the several factors that one has to accounts for a correct execution of the treatment. In the following pages, an introductive discussion to the typical workflow of a radiotherapy treatment we'll be presented, focusing on the LINAC treatment with photons beams.

## 2.2 EXTERNAL BEAM RADIOTHERAPY (EBT)

EBT is a branch of radiotherapy treating patients by the use of a radiation beam external to the body. Althought a LINAC generates both electrons than photon beams in the following we'll focus on radiotherapy with photons, which is the most used technique nowadays. EBT is a well structured process including: patient immobilization, dose prescription , treatment planning and treatment delivery. The treatment planning, object of this work, plays a crucial role in the radiotherapy worflow. It can be divided into five main steps: images acquisition, volumes contouring, dose prescription, plan optimization and evaluation of the quality of the plan. They are all discussed in this section; a brief discussion on the LINAC treatment delivery is also included.

### 2.2.1 Image acquisition

Patient 3D images must be acquired both for the identification of the target volumes and for carrying out an accurate dosimetric calculation. Three different acquisition modalities may be applied: Computed Tomography (CT), Magnetic Resonance (RM) and Positiron Emission Tomography (PET). Morphological information of the target volume and surrounding healthy organs are gathered from CT and MR while functional and biochemical ones from RM and PET.
Today CT is the scan used in treatment planning as it is the only one measuring the electronic tissue density needed to characterize the interaction of radiation with matter and to accurately calculate the dose.
MR measures the orientation of protons in a magnetic field, produces images at very high resolution and enhances the soft-tissue tumor contrast.
PET measures the concentration of a radiopharmaceutical (e.g. FGD) inside the tissues given information about the cellular activity , the main limit of the PET is its low spatial resolution. PET and RM are often combined with CT to characterize area of different aggressiveness inside the tumor as well as hypoxic area. Indeed the identification, by means of PET and RM, of the so called Biological Target Volume (BTV) is of great importance in defining the treatment dose.
Each 3D medical image is made up of hundreds or thousands of elementary constituents, called *voxels*, which are the equivalent of pixels in 2D images. We introduce this concept since it will be used in the following. Medical images are stored in a specific format named DICOM, which is more than a standard image compression algorithm: DICOM files includes also several about the patient and the scan ifself.

### 2.2.2 Volumes contouring

Contouring is the act of defining the volume of the tumor and healthy organs around it, called Organs At Risk (OAR). This a very crucial part of the plan-

ning procedure: determining tumor location, its dimension and phenotype is crucial for the dose prescription. In order to guarantee the safety of the patient during the radiotherapy treatment, as well as guarantee the repetibility of the treatments, different region need to be identified around the tumor itself. The main two are the followings:

- **Gross Tumor Volume (GTV)**: this is the volume which can be seen by eye (or feel by palpation) either on the patient itself or with the help of imaging techniques.

- **Clinical Target Volume (CTV)**: it's the volume containing the GTV and/or other subclinical malignant disease which has to receive a curative dose. This volume has to be treated adequatelly in order to achieve the aim of the therapy, either cure or palliation. The definition of the CTV is not unique and depends mainly on the ability of the oncoligist.

In Figure 1 an example of the difference between the GTV and the CTV for a lung cancer is shown. We see that it's not difficult to identify the GTV by eye, while the detection of the CTV is not trivial; combination of different images modality could help in defining the CTV. In order to account for all possible

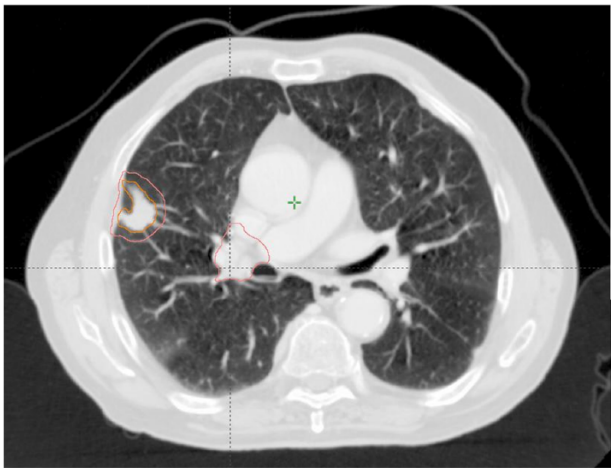


**Figure 1:** Example of the difference between the GTV (orange line) and the CTV (pink line) for a lung cancer.

source of uncertainty during the radiotherapy treatment, like organ motions or error in the patient positioning, two other volumes are defined:

- **Internal Target Volume (ITV)**: it is obtained by adding to the CTV the so-called *internal margin* (IM), which accounts for the motion of the CTV (e.g. motion of the lung during the respiration cycle).
- **Planning Target Volume (PTV)**: this volumes accounts for errors in the patient positioning, mechanical accuracy of the equipment and human error. It's obtained by adding to the ITV the so-called *set-up margin* (SM). The PTV is used to choose the appropriate beam size and arrangements to ensure that the prescribed dose is actually and correctly delivered to the *whole* CTV.

In Figure 2 an example of the contouring of the ITV and PTV is shown.
Once the tumor's volumes have been defined, particular care has to be given the *organs at risk* (OARs), whose radiation sensitivity is such that the dose received during the treatment should be less than their tolerance.
In Figure 3 a schematical summary of the different volumes discussed above and their reciprocal relation is shown.

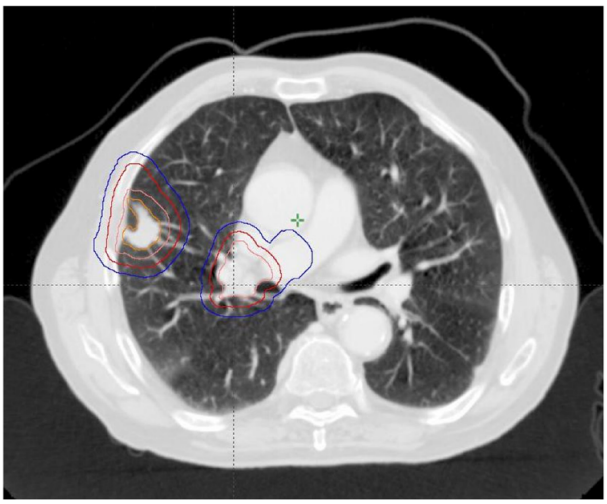


**Figure 2:** Example of the ITV (red line) and the PTV(blue line) for a lung cancer.

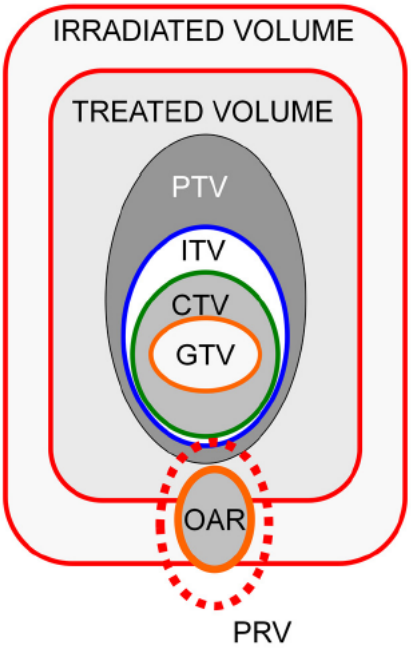



Figure 3: Representation of the different volumes which can be countoured around the GTV.

### 2.2.3 Dose prescription and plan optimization

The main goal of the radiotherapy is to treat the tumor sparing the surrounding healthy tissue and OAR as much as possible. To this aim it is mandatory to optimize the dosimetric plan. A dosimetric plan is computed on a patient CT by a medical physicist and represent the distribution of the dose inside the patients as delivered by different photon beam that affect patients from different angles or entry points This opens to the necessity to *optimize* the plan, in order to cause the largest damage to the tumor, while sparing healty tissues. The goodness of a dosimetric plan depends on several factors: accuracy of the imaging and contouring, reliability of the dose calculation, delivery technique, number and geometry of the beams. The optimization process concerns the number and geometry of the beam and specifically the fluence of the beams. There are two different optimization procedure which are used in practice: *forward planning* and *inverse planning*. The forward planning procedure is a sort of

trial and error approach, where the planner chooses a particular beam geometry to deliver the prescribed dose and evalutes the outcomes. If the latter are not satisfied, some parameter is changed and the new plane evaluated, until a good result is reached. This procedure is particular apted in case of tumor with simple shapes and far from critical organs.
For all the other cases, an inverse planning procedure is required. This procedure is much more complex than the previous one. Instead of starting from a configuration and evaluating the results, the desired treatment outcomes are codified in terms of mathematical constraints into an *objective function.* The optimization process , based on a cost function, modified the fluences of the planned beams to reach the prescribed dose and satisfy the clinical constraints. The results of the process is a delivery technique called intensity modulated radiation therapy (IMRT) .In section 2.3 all the details concerning the mathematical formulation of the problem inMRT will be discussed.
An optimizer is used to minimize the cost function and extrapolate the correct values of the parameters. This is in general an expensive procedure from the computational point of view, but allows to deal with tumors with very complex shapes and very close to critical organs.

### 2.2.4 Plan's quality evaluation

Provided a adiotherapy plan, it's necessary to have some criteria to evaluate its quality in terms of agreement with the clinical needs and contraints. The *dose-volume histogram* (DVH) is the tool used in the plan evaluation activity. Defined for each volume it gives a global information about the dose distribution inside it. We have two version of the DVH, both used in practice:

- **Differential DVH:** it is an histogram relating the radiation dose to tissues' volume; it represents the percetage volume which received a given dose value, in a range defined by the minimum and maximum dose.

- **Cumulative DVH:** it's the cumulative of the previous one, namely the percentage volume which receives at least a defined dose.

Analysing these histograms, it is possible to verified whether the constraints imposed on the different organs are satisfied. In relation to the different types of organs, the constraint can be expressed as an average value , Dose max, Dose min or ,more commonly, as the maximum volume receiving a defined dose or the maximum dose delivered to certain fraction of volume. For the tumor constrains generally describe the coverage and the dose uniformity.

### 2.2.5 Treatment delivering: the LINear ACcelerator (LINAC)

Once the optimal plan has been obtained and approved, the next step is to deliver it to the patient using a *LINear ACcelerator* (LINAC) [21], which is a device that uses high Radio-Frequency (RF) electromagnetic waves to accelerate charged particles (i.e. electrons) to high energy in a linear path, through a tube-like structure called accelerator waveguide, which is a resonating cavity whose typycal frequency is around 3 GHz. Electrons energies range from 6 to 22 MeV, depending either on the machinary and on the clinical needs. This is the other bottelneck of radiotherapy treatments based on X-rays: LINAC is responsible of the accurate production, monitoring, control and conformation of the radiation beam to the plan target . The success of a radiotherapy treatment depends on the ability of the linear accelerator to deliver the prescribed dose to the tumor, while ensuring minimum radiation to normal tissues. It follows that, and we'll see it later in the discussion about inverse planning technique with IMRT, any software aimed at improving the quality of a radiotherapy treatment, must be able to interface with this very powefull machinary. In Figure 4 a schematic representation of the outer and inner components of a LINAC is shown. The two main external components are the *stand*, which is a large rectangular cabinet secured to the treatment floor, and the *gantry*, which rotates on the horizontal axis bearings located inside the stand, allowing to irradiate the patient on the *treatment coach* from different angles. In particular, it can rotate

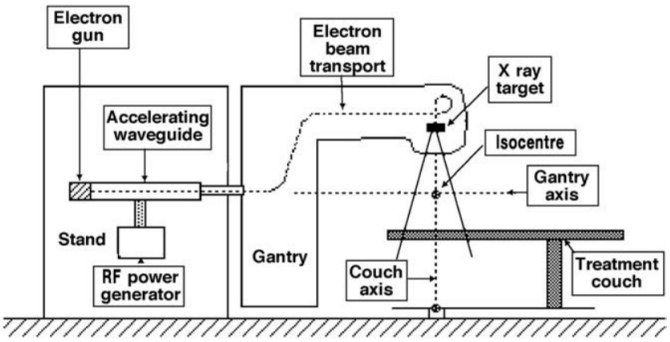



**Figure 4:** Schematic representation of the main components of a LINAC used for radiotherapy.

360° around a point called *isocenter*. The stand also contains a radiofrequency generator, a RF waveguide and an electron gun. Other geometries are possible, with the electron gun placed inside the gantry. Radiofrequency waves are pulsed into the waveguide by the RF generator. This process is syncronized with the injection of electrons into the waveguide by the electron gun. They are accelerated by the radiofrequency waves into the accelerator waveguide, at a speed which is more than 99% of speed of light, and transported along the waveguide toward a tungsten target situated at the exit of the gantry's *treatment head*. In the interaction between the electrons and the target, an X-rays beam is produced via Bremmstrahlung. The energy of the final X-rays is determined by the power and frequency of the radiofrequency waves, which are controlled by the RF generator, and they typically range from 1 MeV to 25 MeV. It's interesting to notice that about the 94% of the electrons energy goes into heat in the interaction with the target. The number of electron injected into the waveguide, and thus the final intensity of the X-rays beams produced via Bremstrahlung, is determined by the electron gun. The latter consists in a tungsten filament crossed by a current where the electrons are boiled out by thermionic emission. The tipical temperatures of the filaments are over 1000°C. A pulsed DC voltage between the gun's electrod accelerates the electrons toward the acceleration wave guide to a speed of about 1/4 the speed of light and a convergent electric field allows to form a pencil beam. The higher the temperature of the

filment is, more electrons are produced. The accelerator waveguide is made up of series of *copper cells*, each one connected to its neighbour cells by holes which allow the electrons to travel through them. The presence of the holes also helps to focus the beam along the waveguide axis. In addition to that, the correct beam focus is guardanteed by quadrupole magnets placed along the waveguide. In order to increase the mean free path of the electrons, a vacuum is created. Furthermore, the whole system is cooled by water in order to allow the components to work at a proper constant temperature. In case of electron energies greater than 6 MeV, a bending magnet, whose function is to focus the beam, is placed the end of the accelerator waveguide. The key role of this component is to focus electrons with slightly different energies onto the same point on the target plan, a property called *achromatic behaviour.*

The X-ray beam produced is typically forward peaking shaped in the direction of the patient's tumor, meaning that the distrubution of the photons across the beam section is not uniform. In order to solve this issue, a *flattening filter* is placed after the target: it is a conical shaped metal absorber which absorbes more photons in the central region of the beam than in the edges, allowing to produce a uniform photons distribution. The effect of the flattening filter is shown Figure 5. Another important contribution to the beam shaping is given

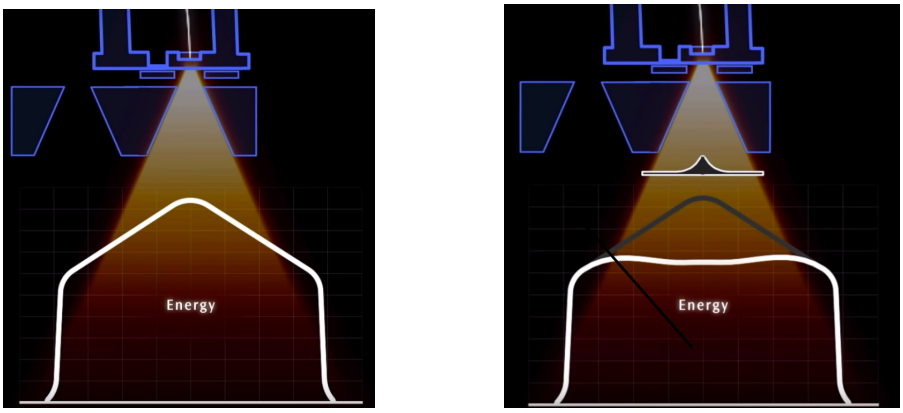
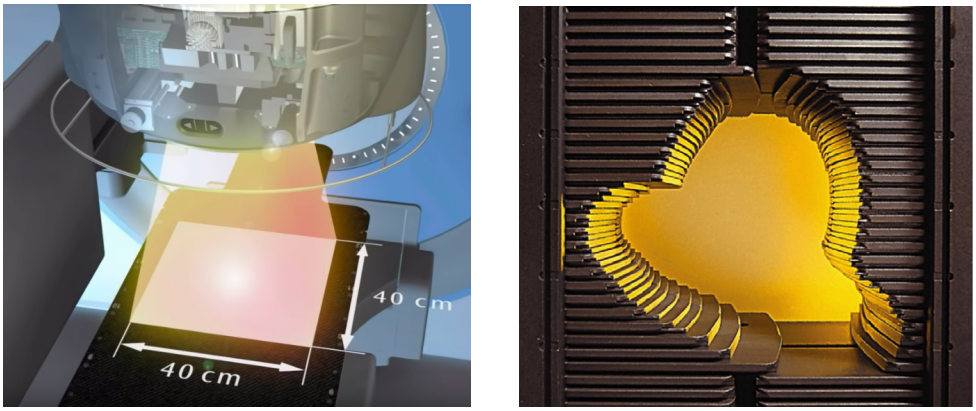



**Figure 5:** Effect of the flattening filter on the beam's photons distribution. *(Left)* Before; *(Right)* after.

by the *collimators*, which are jaws made up of high Z number elements, such

as tungsten or lead. Typically there are two collimators, called *primary* and *secondary* collimator, respectively. The primary collimator allows only the forward travelling X-rays to pass through, producing a cone shaped beam. In this way, it minimizes leakages by absorbing scattered photons travelling in the lateral directions. The combination of the primary and secondary collimators allows to define the maximum area of the resulting clinical radiation beam, which is typically a $(40 \times 40)\ cm^2$ rectangle, as shown in Figure 6a.




**(a)** **(b)**

**Figure 6:** **(a)** Resulting clinical radiation beam produced by the combined action of the primary and secondary collimators. **(b)** Example of multi-leaf collimator used in IMRT treatments.

In order to measure the dose delivered to the patient as well as the beam quality, a system made up of two ionization chambers is placed at the exit of the gantry's treatment head. They are called primary and secondary dosimeters, respectively, they are sealed together but power supplied and read independently. In Figure 7 a representation of how the two chambers work is shown. The first chamber (the upper one) has the role of monitoring the dose delivered to the patient and stopping the beam once the desired dose has been delivered. The lower one works as a backup, controlling the correct functioning of the first one and stopping the beam in case of a failure of the latter. At the end of all this chain, a further mechanical collimator can be placed between the grantry exit and the patient in order to conform the beam shape to that of the

tumor. A typical choice in case of IMRT treatments is the *multi-leaf collimator* (MLC), which is a device made up of individual *leaves* of a high atomic numbered material, usually tungsten, that can move independently in and out of the path of a radiotherapy beam in order to shape it and vary its intensity. An example is shown in Figure 6b. We'll see in next section the way an optimizer can interface with the MLC in order to produce the desired dose distribution.

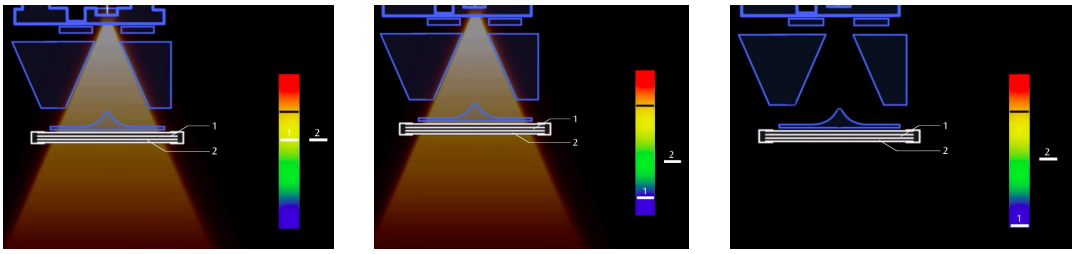


**Figure 7:** Representation of the primary and secondary ion chambers of a LINAC. *(Left)* In case of correct functioning, the two chamber measure the same dose; (*Center)* The first chamber may fail, therefore the two dose measurements become different; *(Right)* The difference in the two measurements triggers the beam stop.

## 2.3 INTENSITY MODULATED RADIATION THERAPY (IMRT)

External beam radiotherapy treatments can be subdivided into 2D and 3D [29], according to the planning procedure. Thanks to the evolution of CT in the last years, it has been possible to move from conventional 2D treatments to the so-called *3D conformal radiotherapy*. This because unless x-ray images, those obtained with CT are 3D and allow a well definition of the target volume. The use of the MLC placed at the exit of the gantry's treatment head allow to conform the beam to the tumor's shape in order not to irradiate the sorrounding tissues. In Figure 8 a comaparison between the planning on a 2D and 3D image is shown. The introduction of the 3D conformal therapy has lead to the possibility both to increase the dose delivered to the tumor and reduce the tox-

icity for the sorrounding OARs at the same time [13]. An advanced form of 3D

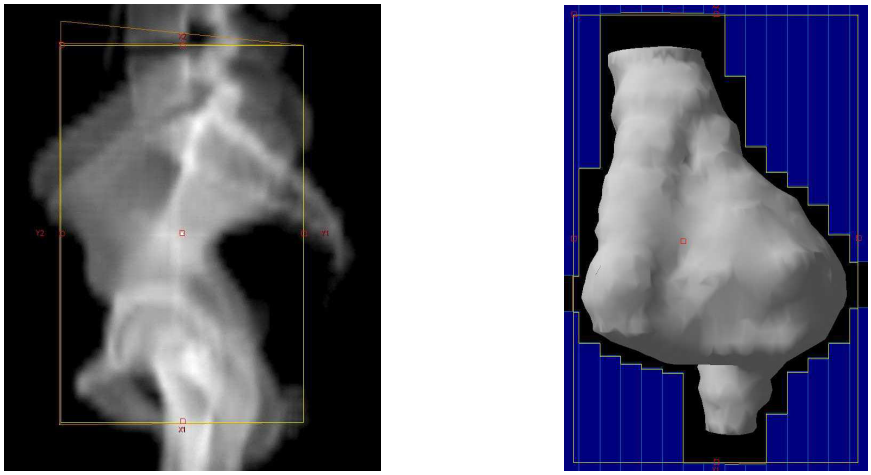


**Figure 8:** Comparison between 2D *(left)* and 3D conformal *(right)* beam shaping. The 3D planning technique allow to conform the beam to the tumor shape.

conformal radiotherapy is the so-called *Intensity Modulated Radiation Therapy* (IMRT), one of the most promising treatments methods used nowadays. This technique is very suitable in case of tumors having very irreguar or concave shapes, strong inhomogeineities or situated very close to vital organ (OARs). The great improvement provided by IMRT relies in the possibility to modulate the *local beam fluence* by means of the MLCs, in contrast with the uniform beam intensity obtained with other techniques. From the practical point of view, the beam is divided into a number of smaller pencil-beams, called *beamlets* and a real weight $x_j \geqslant 0$ is associated to each of them.The main advantages of using this techniques are of dosimetric nature: improvement of target conformity, production of intentional dose inhomogeneities (*dose-painting*), increase of normal tissue sparing, dose escalation. On the other end, the main disadvantages are related to the increase of the planning and delivering time: clinician time for target and organ outlining, use of extensive quality assurance programme, machine treatment time. Furthermore, even the problem of an increased total body irradiation dose is not negligible.

IMRT differentiates from all other techniques also in the planning procedure: conformal therapy is *forward planned*, while IMRT is *inverse planned*: the desired

dose map is encoded in terms of *constraints* and *priorities* into a *cost-function* of the beamlet weights $x_j$, expressing the distance between the delivered and the prescribed dose. The aim of the planning procedure is to exploit the power of an optimization algorithm in order to minimize the cost function with respect to the $x_j$s, a process which should provide the *optimal* configuration of the beamlets weights, i.e. the one which optimizes the fluence according to the prescribed dose. In Figure 9 a schematic representation of the main steps of an IMRT treatment planning is shown. In general, the choice of the cost function is neither simple nor unique and its definition is strictly rrelated to the kind of constraints introduced.

Let's suppose to have a 3D image with R OARs and a PTV. Let's consider a number $N_{beams}$ of radiation fields, each one divided into a $N_B^{(b)}$ beamlets, with $b$ running over the radiation fields. If we refer with $a_{i(r)j(b)}$ to the *elementary dose contribution* per unit intensity of the $j^{th}$ beamlet of the $b^{th}$ field on the $i^{th}$ voxel of the $r^{th}$ OAR (or PTV), the total dose delivered to the same voxel can be obtained as:

$$D_{i(r)} = \sum_{b=1}^{N_{beams}} \sum_{j(b)=1}^{N_B^{(b)}} a_{i(r)j(b)} x_{j(b)} \tag{1}$$

where $r = 0$ refers to the PTV.

In order to simplify the notation, we can define the *total number of beamlets* $N_B$ as:

$$N_B = \sum_{b=1}^{N_{beams}} N_B^{(b)} \tag{2}$$

so that Equation (1) becomes:

$$D_{i(r)} = \sum_{j=1}^{N_B} a_{i(r)j} x_j \tag{3}$$

which states that the total dose delivered to the $i^{th}$ voxel of the $r^{th}$ object (OAR or PTV) is given by the sum of the elementary dose contributions from all the beamlets acting on it, each one rescaled by the corresponding weighting fac-

tor $x_j$. The set of all the $a_{i(r)j}$s form a matrix called *influence matrix*, with the row and columns indices running over the voxels and the beamlets, respectively. This matrix can be be either very sparse or very dense, depending on the interaction between different beamlets. The entries of the influnce matrix are calculated by dose calculation algorithm, generally based on montecarlo methods. In general, what can make the density matrix more or less dense is the scattering of the particles inside the body.
The $x_j$s are the set of variables to optimize and their final configuration form the so-called optimized fluence map. This particular approach to the dose optimization thrugh the optimization of the fluence is called fluence map optimization (FMO). Just for completeness, it's important to tell thath this is not the only optimization scheme used in IMRT: optimization of the beam angles as well as *direct aperture optimization* (DAO) are used in practice [8].
Focusing on the FMO problem, the next step is the definition of an objective function, whose general form is given by:

$$F(x_1, \ldots, x_{N_B}) = \sum_{r=0}^{R} \sum_{i \in R} \gamma_{i(r)} \times \mathcal{D}\left[D_{i(r)}(x_1, \ldots x_n), D^P_{i(r)}\right] \tag{4}$$

where $D^P_{i(r)}$ is the prescribed dose to the $i^{th}$ voxel of the $r^{th}$ object and $\mathcal{D}[A, B]$ indicates an arbitrary distance-function between the two arguments; $\gamma_{i(r)}$ quantifies the *priority* of a given objective, namely the importance of satisfying a given prescription for a given voxel with respect to all the others. Some of the most used distance functions are the following:

- **(A - B)**$^2$: the square of the difference between the arguments; it's the simplest one from the mathematical point of view since it's convex and it can be optimized even through simple grandient descent methods. However, since a penalty is added for any violation of a given constraint, either in excess or in defect, working with this function could be not trivial.

- **[max(**0**, A - B)]**$^2$: it adds a cost only if the delivered dose is higher than the prediction. This kind of constraint is of fundamental importance in

case of OARs which could get damaged if irradiated with a too much high dose.

- **[min(0, A - B)]$^2$**: it works in the opposite way than the maximum; it's usually important to deliver at least a given dose inside the PTV it in order to achieve the treatment aims, thus a positive cost is added if this condition is violated.

- **DVH-based constraints:** they are usually refered as *dose-volume constraints* and they are obtained from the DVH, either differential and cumulative. They can be encoded in terms of maximum and minimum dose, as well as in terms of mean dose or median dose. All optimization function used in nowdays commercial systems are based on this kind of constraints.

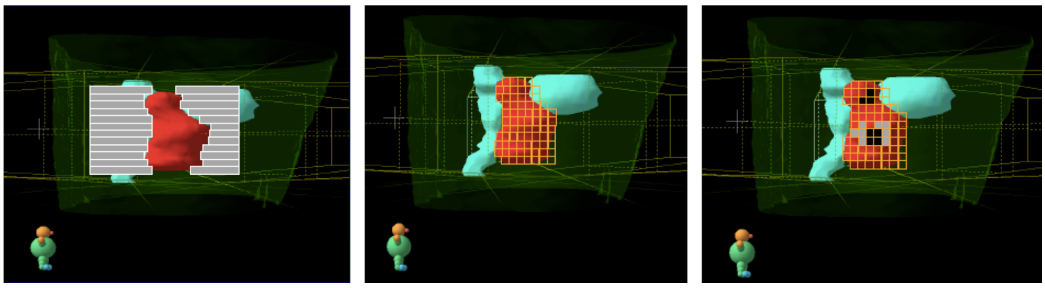


**Figure 9:** Main steps of the IMRT treatment planning. First the beam is conformed to the tumor's shape *(left)*; then the area is divided into sufficient number of beamlets (*(center)*. In the end, after the optimization procedure, a weight is associated to each beamlet (*(Right)*.

In this work a quadratic cost function was used whose final expression given by:

$$F(x_1, \ldots, x_n) = \sum_{r=0}^{R} \sum_{i \in R} \gamma_{i(r)} \times \left( \sum_{j=1}^{N_B} a_{i(r)j} x_j - D^P_{i(r)} \right)^2 \tag{5}$$

We can further simply the notation by absorbing the obejct index $r$ into the voxel index $i$. This can alwat be done by creating a single influence matrix $A$ for all the obejcts. In the end, Equation (5) becomes:

$$F(x_1, \ldots, x_n) = \sum_i \gamma_i \times \left( \sum_{j=1}^{N_B} a_{ij} x_j - D_i^P \right)^2 \tag{6}$$

The reasons behind this choice of the cost function are manily two: being the first study of this kind we've decided to start from the simplest possible scenario from the mathematical point of view; secondly, we'll see in the following that such a kind of function is very suitable to be converted into a quadratic Ising-like Hamiltonian, without introducing any approximation.

# 3 MANY-BODY QUANTUM SYSTEMS

Many-body systems are ubiquitos in the world we live, from big planets and stars populating the universe [32, 22, 6] to atoms and molecules in materials [24, 19]. The complexity of providing a physical description of such systems relies on the fact that they are made up of several elementary constituents interacting non-trivially.

We can consider as a many-body system each system composed by a *large* number of such elements, even though the definition is strongly related to the particular scenario one is dealing with.[1] The study of this class of systems in quantum mechanics has a strong impact on various disciplines such as condensed matter [9], computational chemistry [43], material science [26] and quantum information [30].

The aim of this chapter is to provide an introduction about a particular class of many-body quantum systems, the so-called *quantum lattice systems*. After having defined what they are, we'll go into a description of the entanglement for this systems, showing how it can be quantified starting from the description of the wave-funciton in the *density matrix* formalism. At the end of the chapter we'll see how entanglement is at the base of the tensor network methods' representation of a many-body wave function.

1 Usually one distinguishes between *one-body*, *few-body* and *many-body*. The boundary between the last two is somehow blurry.

## 3.1 QUANTUM LATTICE SYSTEMS

Quantum lattice systems are many-body quantum systems whose underlying structure is a given graph $\mathcal{G} = (V, E)$, with $V$ representing the degrees of freedom of the system and $E$ the interaction among them. Even though any underlying structure could be used to represent a quantum lattice systems, regular graphs are mostly used. In particular, many lattice systems are described using a *cubic lattice* $V = L^D$, being $D$ the dimension of the problem. The distance between two sites $i, j$ is the *graph distance* $d(i, j)$, i.e. the minimum number of sites to walk through in order to reach $i$ from $j$, and viceversa. A *local* Hilbert space $\mathcal{H}_k$ is associated to each lattice site $k$ and the global Hilbert space $\mathcal{H}$ for the many-body quantum lattice is given by the product of single-particle spaces:

$$\mathcal{H} = \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_N \tag{7}$$

One of the easiest examples of such systems is the 1D Ising Model, a spin model whose underlying structure is a one-dimensional chain where the quantum degree of freedom associated to each site $k$ is the spin of the occupying particle. The local Hilbert space has dimension $\dim(\mathcal{H}_k) = 2$, which corresponds to the spin configurations (up and down). According to Equation (7) the global Hilbert space is given by the composition $(\mathbb{C}^2)^{\otimes N}$, whose dimension is $\dim(\mathcal{H}) = 2^N$, with $N$ the number of sites. It exhibits an exponential scaling in the number of constituents which is a general behaviour of any quantum many-body system and the main challenge for simulating them on classical computers.

The physics of the lattice is described by the system's Hamiltonian, whose general form is:

$$H = \sum_{i \in V} h_i \tag{8}$$

where each $h_i$ is the Hamiltonian associated to the $i^{th}$ elementary constituent which may be in general composed by local and interaction terms. For in-

stance, in the particular case of the 1D Ising model in transverse field, the local Hamiltonian $h_i$ takes the following form:

$$h_i = -\lambda \cdot \sigma_i^z + \frac{1}{2} \sum_{j \in \langle i,j \rangle} \sigma_i^x \sigma_j^x \tag{9}$$

where the $\sigma_i^{x(z)}$s are the Pauli's matrices, $\lambda$ is the magnetic field and the symbol $\langle i, j \rangle$ means that the two-body interaction is limited to nearest-neighbor sites only. The exponentially growing dimension of the associated global Hilbert space can be extract from Equation (7) as follow. For a system of $N$ sites the formal representation of the nearest-neighbor interaction between sites $i$ and $i+1$ is given by:

$$\sigma_i^x \sigma_{i+1}^x \equiv \mathbb{I}_1 \otimes \mathbb{I}_2 \otimes \ldots \otimes \mathbb{I}_{i-1} \otimes \sigma_i^x \otimes \sigma_{i+1}^x \otimes \mathbb{I}_{i+2} \ldots \otimes \mathbb{I}_N \tag{10}$$

which is a square matrix of dimension $2^N$, with $\mathbb{I}$ the identity matrix. The idea behind Equation (10) remains valid also for the local term at site $i$, simply performing the substitutions $\sigma_{i+1}^x \to \mathbb{I}_{i+1}$ and $\sigma_i^x \to \sigma_i^z$.

Let's now suppose we are interested in finding the *ground state* of such a system: we would need to explore $2^N$ configurations in order find the one which fits our request. If we want to approach the problem numerically, we need an algorithm which can do that efficiently. In quantum mechanics the situtation is particularly complex because of the presence of matrix and vectors of order $d^N$, with $d$ the generic single-particle space dimension, which are typically difficult to manipulate and store as $N$ increases.

This has lead to the developement of several algorithms such as Density Matrix Renormalization Group (DMRG) [48] and Tensor Network Methods (TNMs) [33, 40, 42] in the past years in order to perform numerical simulation of many-body quantum systems. The main reason why these algorithms succeded in solving different quantum many-body problems is that they provide a systematic way to represent a many-body wave function and to *compress* the information contained in a quantum state, simplifying its representation but keeping a

sufficiently high information content at the same time.
Given a many-body quantum state its complete representation contains all the information available and is described by $d^N$ parameters. On the other side, a description of the state in terms of *mean-field* approximation allows to reduce the number of parameters to $d$, which is the lowest-dimensional possible representation of a quantum state. TNMs interpolate between the two cases, which is the reason why they have been so intensely studied during the last years.
In next section we'll go deeper into the discussion, explaining the key role played by *entanglement* in the representatation of a many- body state.

## 3.2 ENTANGLEMENT IN QUANTUM SYSTEMS

From quantum mechanics we know that a generic single-particle state is described by its wave function $|\psi\rangle \in \mathcal{H}$, with the normalization condition $||\psi||^2 = \langle\psi|\psi\rangle = 1$. If we consider an orthonormal basis $|\alpha\rangle_1, \ldots, |\alpha\rangle_n$ of $\mathcal{H}$, $|\psi\rangle$ is given by a *coherent superposition* of such basis states, namely:

$$|\psi\rangle = \alpha_1 |\alpha\rangle_1 + \ldots + \alpha_n |\alpha\rangle_n \tag{11}$$

where $\{\alpha_1, \ldots, \alpha_n\} \in \mathbb{C}$ are the so-called *probability amplitudes*. It reads that $|\alpha_i|^2$ is the probability that, after a measurement, the state $|\psi\rangle$ collapses onto the state $|\alpha\rangle_i$. We can extend this argument to many-body state, saying that all states whose representation can be performed in terms of *bra-* and *-ket* are usually refered as *vector states*, or more in general as *pure states*, in order to differentiate them from *mixed states* which are statistical mixtures of pure states. We can represent a generic quantum state using the *density matrix* formalism, a description which allows also to embody and extract information about the

entanglement of the system.
Taking a generic pure state $|\psi\rangle$, the associated density matrix is defined as[2]:

$$\rho = |\psi\rangle \langle\psi| \tag{12}$$

with the properties:

1. $\rho = \rho^\dagger$:

2. $\mathrm{Tr}(\rho) = 1$

3. $\mathrm{Tr}(\rho^2) \leqslant 1$

The connection of density matrix formalism to the system's entanglement is all contained in property 3. It can be shown that the equality holds for pure states only [5].
Let's now consider a system of two particles, A and B. Suppose now to have them in a pure state $|\psi\rangle_{AB}$ whose density matrix is given by $\rho = |\psi\rangle_{AB} \langle\psi|_{AB}$ according to Equation (12). We can introduce the *reduced density matrix* for one of the two dipartitions, A for example, which is:

$$\rho_A = \mathrm{Tr}_B(\rho_{AB}) = \sum_k \langle B_k| \rho_{AB} |B_k\rangle \tag{13}$$

which corresponds to the *partial trace* computed with respect to particle B., with the same properties defined before. If the system is entangled, we'll find the single particle in a statistical mixture of states, namely $\mathrm{Tr}[\rho_A^2] < 1$.
All the previous description allows to introduce one most spread measures of entanglement for pure quantum states, the so-called *Von Neumann entropy*

2 More formally, this is called *density operator* and the density matrix can be obtained from the latter by choosing a basis in the underlying space. However, in practice the two terms are often used interchangeably.

or *entanglement entropy*. Let's consider a generic state described by its density matrix $\rho$, the entanglement entropy is defined as:

$$S(\rho) = -\mathrm{Tr}[\rho \log(\rho)] = -\sum_j \lambda_j \log(\lambda_j) \tag{14}$$

with $\lambda_j$ the $j^{th}$ eigenvalue of $\rho$. If $\rho$ describes a pure state, the entropy is vanishing. In order to quantify the entanglement in a pure system we always need to consider a bipartion of the system into two subsystems A and B, compute the reduced density matrix of one them and evaluate its entanglement entropy in order to quanfity whether the system is separable ($S = 0$) or entangled ($S > 0$). This concept can be further extended if we consider the so-called *Schmidt decomposition* of the state. Given a many-body state $|\psi\rangle$ describing a system of N particles, we can always apply a *Liouville transformation* and write it in a matricial form L with respect to a given bipartition $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ of the Hilbert space. The Schmidt decomposition of the state consists in performing the Singular Value Decomposition (SVD) of L, as:

$$L = S \times V \times D \tag{15}$$

where S(D) is a unitary matrix whose columns (rows) are the left (right) singular vectors $\{|\alpha_A\rangle\}(\{\langle\alpha_B|\})$, which form an orthonormal basis for the space A (B) and V is a diagonal matrix whose entries are the *singular values* $\Lambda_\alpha \geqslant 0$. The Schmidt decomposition of the state $|\psi\rangle$ thus becomes:

$$|\psi\rangle = \sum_\alpha \Lambda_\alpha |\alpha_A\rangle |\alpha_B\rangle \tag{16}$$

where the index $\alpha$ runs over all non-zero coefficients $\Lambda_\alpha$. The number of non-zero Schmidt coefficients is also refered to as *Schmidt rank*.
The connection with the density matrix formalism is that the complete set of $\{\Lambda_\alpha^2\}$ corresponds to the eigenvalues of the reduced density matrix of the bipartition (Equation 13) and takes the name of *entanglement spectrum*.

In general, TNM representations of quantum states exploits Equation (16) and introduce a *cutoff* $\chi$ in the representation of the state to approximate it in a lower-dimensional subspace, which is:

$$|\psi\rangle = \frac{1}{Z}\sum_{\alpha=1}^{\chi} \Lambda_\alpha |\alpha_A\rangle |\alpha_B\rangle \tag{17}$$

with $Z = \sqrt{\sum_{i=1}^{\chi} \Lambda_i^2}$ a normalization constant and $\chi$ the so-called *bond dimension*. This is done by cutting off the $n$ lowest singular values $\Lambda_\alpha$, which has been proven [20] to be the best possible lower-rank approximation in term of the Frobenius norm $||.||_F$:

$$||M - M_\chi||_F^2 = \sum_{\alpha=\chi+1}^{S_R} \Lambda_\alpha^2 \tag{18}$$

with $S_R$ being the Schmidt rank.

The introduction of a bond dimension allows to compress the representation of the state, interpolating between a mean-field approximation and the complete description, as it was already discussed in the previous section. For example, the *Matrix Product State* (MPS) [34], one of the most frequently used tensor networks astantze, can be obtained by applying iteratively Liouville transformations followed by Schmidt decompositions to a given state [27].

In the next chapter we'll go on with the discussion, introducing the basics about TNM and showing how they can be exploited to look for the ground states of many-body quantum lattice systems.

# 4 TENSOR NETWORK METHODS

In this chapter, we'll provide a general introduction to TNM for the numerical simulations of quantum lattice systems. First, we'll briefly describe how they can be used to represent tensors and their operations.
In the second part, we'll discuss in more detail how the problem of ground-state search through TNM can be approached. In particular, we'll first describe mean-field approximation for 1D systems, showing how the latter can be thought of as a special case of TNM, with the bond dimension $\chi = 1$.
Finally, we'll conclude with a general description of the *tree tensor network* ansatz, which is the one used in this work.

## 4.1 TENSORS DEFINITION AND REPRESENTATION

Tensor calculus is a technique that can be regarded as an extension of linear algebra. In linear algebra we usually deal with *vector* and *matrices*: tensors have to be thought of as a generalization of such concepts, owning all their properties and adding new very powerfull features at the same time. A tensor can be defined as follow:

**Definition 1 (Tensor)** *Given a finite set $\{\mathcal{H}_1, \ldots, \mathcal{H}_N\}$ of vector spaces over a common field $\mathcal{F}$ one can form a new vector space $\mathcal{H}$ as the result of the tensor product operation, as follow:*

$$\mathcal{H} = \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_N \tag{19}$$

*The generic element of $\mathcal{H}$ this space is called tensor.*

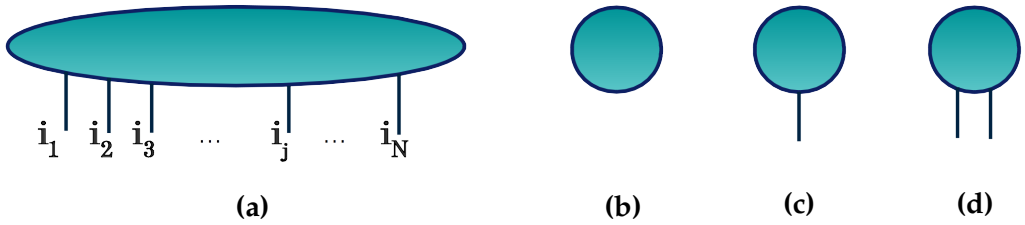
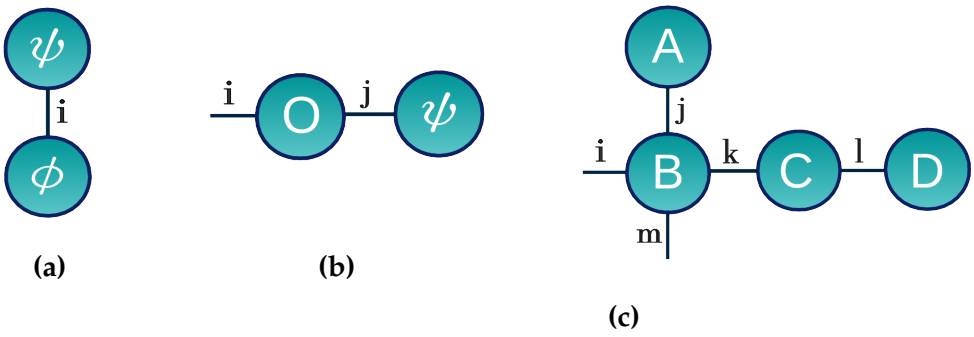



**Figure 10:** **(a)** Generic tensor representation; **(b)** representation of a scalar; **(c)** representation of a vector; **(d)** representation of a matrix.

From a very general point of view, a tensor is a container of numbers, where each of the latter can be addressed by a set of indexes. The total number of indexes equals the number N of vector spaces $\mathcal{H}$ is made up of. The total number of indexes is called *rank* of the tensor and their dimension is equal to the dimension of the vector space they originated from. A generic tensor can be represented as follow:

$$T_{\alpha_1,\ldots,\alpha_N} \tag{20}$$

For a detailed description of all mathematical properties of tensor see [14].
For our purposes, we can think of the vector spaces $\{\mathcal{H}_1,\ldots,\mathcal{H}_N\}$ as the single-particle Hilbert spaces associated with each lattice site of a quantum lattice system.
Using a graphical notation, an N-rank tensor can be thought of as a round box representing all numbers, with a number of *legs* equal to the rank, as shown in Figure 10a. It follows that a scalar, which is a 0-rank tensor, will be represented as a circle with no indexes, as shown in Figure 10b. In the same way a vector (1-rank tensor) and a matrix (2-rank tensor) will be represented as boxes with one and two legs respectively, as it is shown in Figure 10c and 10d.
Manipulating tensors with many indexes is not a trivial task in general, but it can become quite simple if we imagine to represent each operation among them as a *tensor network*, where *connections indicate contractions*.




(a) (b) (c)

**Figure 11:** Representation of some basic operation among tensor. **(a)** Scalar product between two vectors; **(b)** Matrix-vector multiplication; **(c)** Contraction of four tensors leading to a rank 2 tensor as final result.

For example, in Figure 11a the following scalar product is represented:

$$\langle \phi | \psi \rangle = \sum_i \phi_i^* \psi_i \tag{21}$$

where the resulting tensor has no free links, meaning that it is a scalar, as it was expected. In the same way, in Figure 11b a representation of the following contraction for a matrix-vector multiplication is given:

$$(\mathcal{O}\psi)_i = \sum_j \mathcal{O}_{ij} \psi_j \tag{22}$$

where the final tensor has rank 1, meaning that it is a vector. One can generalize the previous arguments to any number of tensors with arbitrary ranks, obtaining a real network of tensors, where the number of free indexes determines the rank of the resulting tensor after the indicated contractions. An example is shown in Figure 11c, where it's quite easy to see that the complex contractions involving four tensors lead to a rank 2 tensor (i.e. a matrix) as final result.

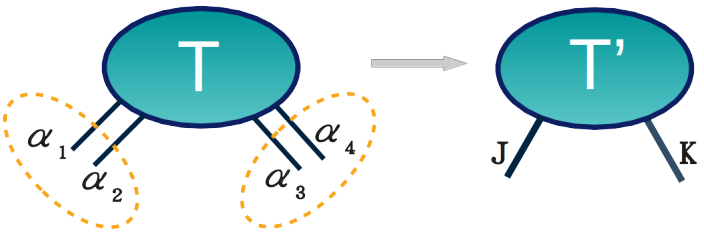



**Figure 12:** Representation of the *index fusion* procedure.

## 4.2 TENSORS MANIPULATION

Let's now show how it is possible to reshape the representation of a tensor to a lower (higher) rank tensor by *fusing* (*splitting*) its indexes. Suppose we have a 4-rank tensor $T_{\alpha_1,\alpha_2,\alpha_3,\alpha_4}$. We can group the first two indexes into one single index as $\{\alpha_1, \alpha_2\} \succ \mathbf{J}$ and the other two as $\{\alpha_3, \alpha_4\} \succ \mathbf{K}$. The meaning of this *index fusion* process can be easily understood if we turn to the graphical notation.

Looking at Figure 12 we see that the initial 4-rank tensor gets transformed into a 2-rank one, which is a matrix. What we are practically doing is to change the way information is stored. In fact, the dimension of the new index is given by the product of the dimension of the fused indexes, namely $\dim(\alpha_{1(3)}) \times \dim(\alpha_{2(4)}) = \dim(\mathbf{J}(\mathbf{K}))$.

However, the immediate consequence of this procedure is that we can apply all the powerfull tools provided by linear algebra even to tensors, simply reshaping them into matrices and going back to the initial representation at the end of the computation by splitting the indexes if needed. All previous manipulations have a computational cost, but it's in general very convenient to face the problem from this point of view [42].

If we now recall the discussion of Section 3.2 about the Liouville transformation of an N-particle wave function, we see that it's exactly the process of splitting the indexes with respect to a given bipartition of the system in order to reshape the vector into a matrix.

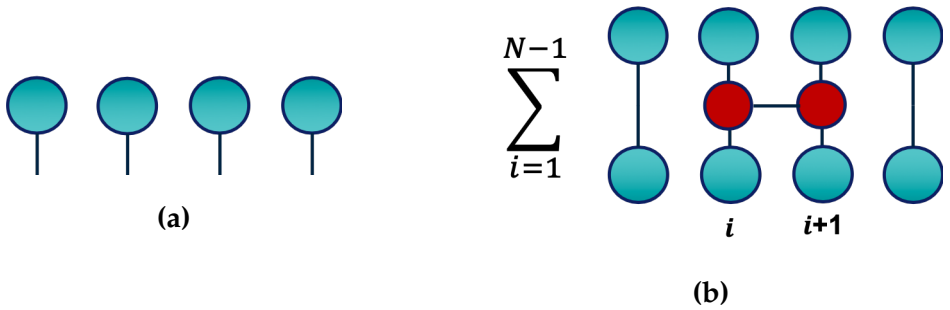



**Figure 13:** **(a)** Representation of a many-body state using the mean-field ansatz. **(b)** Energy expectation value in the mean-field approximation for the Ising in transverse field with $\lambda = 0$.

## 4.3 GROUND STATES VIA TENSORS NETWORKS

When TNMs are used for the ground-state search, they work as *variational algorithms*, meaning that they aim at solving the following variational problem:

$$E_{GS} = \min_{\psi} \langle \psi | H | \psi \rangle \tag{23}$$

where $H$ is the system Hamiltonian. One of the main features related to the use of tensor network methods is that they provide different possible *ansatze* through which the wave function $\psi$ can be represented. The method we'll show in a moment is very general and can be adapted to any tensor network.
Let's suppose we want to use tensor network method to find the ground state of the 1D Ising model in transverse field in the *mean-field* approximation. To be more general, we can relax the translational invariance and consider a generic product-state. This is the simplest tensor network ansatz, where the bond dimension is $\chi = 1$. Working in the graphical notation, the state can be represented as shown in Figure 13a where all tensor are equal in case of mean-field approximation.
Without loss of generality, refering to Equation (9), we can set $h_i = 0$ for sim-

plicity, obtaining the following Hamiltonian, where only the two-site nearest-neighbour interactions are present:

$$H = \sum_{i=1}^{N-1} J_{ii+1} \sigma_i^x \sigma_{i+1}^x \tag{24}$$

The minimization problem in Equation (23) correspond to minimize the energy rpresented in Figure 13b, where the blue tensors represent the particles, while the red ones the $\sigma^x$ matrices.

If we now move to a more general description, any two-site operator can be described as a tensor $H^{\alpha_i' \alpha_j'}_{\alpha_i, \alpha_j}$ acting between particles at sites $i$ and $j$. The minimization problem in case of nearest neighbour interactions can be now expressed using the method of the *Lagrange multipliers* to impose the normalization of the state as follows:

$$\mathcal{L}(\psi_1, \psi_2, \ldots, \psi_N, \psi_1^*, \psi_2^*, \ldots, \psi_N^*) = \langle \psi | H | \psi \rangle - \lambda \left( \langle \psi | \psi \rangle - 1 \right) \equiv E - \lambda(\mathcal{N} - 1) \tag{25}$$

with:

$$\langle \psi | H | \psi \rangle = \sum_{i=1}^{N-1} H^{\alpha_i' \alpha_{i+1}'}_{\alpha_i \alpha_i + 1} \psi^*_{\alpha_i} \psi^*_{\alpha_{i+1}} \psi_{\alpha_i'} \psi_{\alpha_{i+1}'} \tag{26}$$

where the sum over the tensors' components is implied. Equation (25) has to be minimized with respect to the components of each tensor in the network, as follow:

$$\frac{\partial \mathcal{L}}{\partial \psi_i^*} = H^{\alpha_{i-1}' \alpha_i'}_{\alpha_{i-1}, \alpha_i} \psi^*_{\alpha_{i-1}} \psi_{\alpha_{i-1}'} \psi_{\alpha_i'} + H^{\alpha_i' \alpha_{i+1}'}_{\alpha_i \alpha_{i+1}} \psi^*_{\alpha_{i+1}} \psi_{\alpha_i'} \psi_{\alpha_{i+1}'} - \lambda \psi_{\alpha_i'} = 0 \tag{27}$$

The effect of the differentiation of the network with respect to one of its tensor is to remove that tensor from the network. If we now define an *effective Hamiltonian* $\tilde{H}$ as:

$$\tilde{H}^{\alpha_i'}_{\alpha_i} \equiv H^{\alpha_{i-1}' \alpha_i'}_{\alpha_{i-1}, \alpha_i} \psi^*_{\alpha_{i-1}} \psi_{\alpha_{i-1}'} + H^{\alpha_i' \alpha_{i+1}'}_{\alpha_i \alpha_{i+1}} \psi^*_{\alpha_{i+1}} \psi_{\alpha_{i+1}'} \tag{28}$$

we see that Equation (27) can be expressed as:

$$\tilde{H}^{\alpha_i'}_{\alpha_i}\psi_{\alpha_i'} = \lambda\psi_{\alpha_i'} \tag{29}$$

which is a *local* eigenvalue problem for the effective Hamiltonian $\tilde{H}^{\alpha_i'}_{\alpha_i}$ and the tensor $\psi_{\alpha_i'}$. This procedure can be applied iteratively to all tensors in the network, updating their entries with the result from the minimization at each step, until convergence is reached.
As it's already been said at the beginning of the discussion, this procedure is very general and can be extended to any operator, provided its expression. TNM provides two very powerfull ansazte for an efficient representation of many-body operators. They are the *Matrix Product Operator* (MPO) [37] and the *Tensor Product Operator* (TPO) [42]. The former is very suited in case of MPS representation of the wave function, while the second can be regarded as a generalization of the first to different network structures.
Starting from this very general introduction to tensor network methods in the next section we'll present the essentials about the binary tree tensor network ansatz, the one used in this work.

## 4.4 BINARY TREE TENSOR NETWORKS (bTTN)

### 4.4.1 Introduction to bTNN representation

Let's consider a one dimensional lattice with $N = 2^L$ sites, where each site has an associated local Hilbert space of dimension $d$ and $L$ is the number of layers. A *binary tree tensor network* (bTTN) [16, 42, 27] is a loopless structure built on the top of a one dimensional lattice and it's fully composed by tensors with three links. We see from Figure 14a, the $n^{th}$ tensor of the $l^{th}$ layers is refered to as $\Lambda^{[l,n]}$; layers are enumerated from the top to the bottom. The links in the last layer (the lowest one) are the so-called *physical links*, they represent the sites of

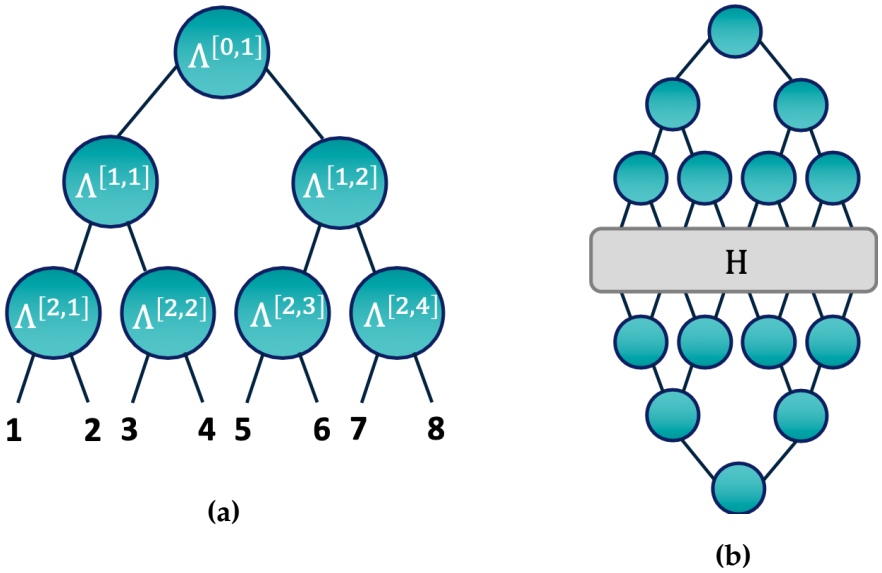



Figure 14: **(a)** Example of binary tree tensor network for a system of eight sites. **(b)** Graphical representation of the energy expectation value computation with bTTN.

the lattice and their dimension equals the local dimension of the associated single-particle Hilbert space. All other links are called *virtual links*.
The tree structure has the property of mapping two sites of the $(l+1)^{th}$ layer into one coarse-grained site of the $l^{th}$ layer. The full Hilbert spaces of the sites at layer $l$ have dimension $M(l) = d^{2^{L-l}}$, which increases exponentialy with the number of physical sites which are blocked together in layer $l$. An efficient numerical representation of the many-body state requires to introduce a cutoff in the dimension of the virtual links. It' can be done by setting a *maximum bond dimension* $\chi_{max}$ in way that all virtual links must have a dimension $\chi_l \leqslant \min(\chi_{max}, d^{2^{L-l}})$. The number of variational parameters in a bTTN scales as $\mathcal{O}(N\chi^3)$.

### 4.4.2 Ground state search with bTTN

It's now interesting to see how bTTN can be used to find the ground state of the system. The procedure explained below can be thought of as a more

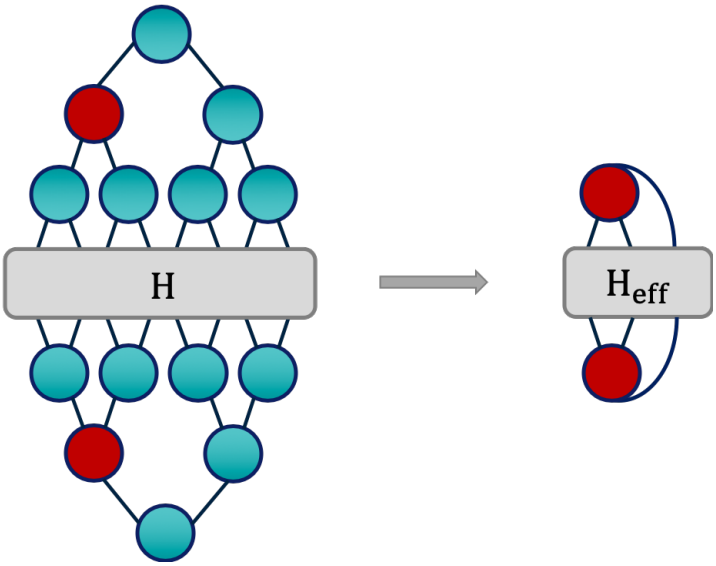



**Figure 15:** Introduction of the effective Hamiltonian for the local optimization of the tensors in the network.

efficient version of the one explained in Section 4.3 and which is also the one implemented in the algorithm. The complete description of this procedure can be found in [42]; here we'll provide only general details to give to the reader the essential knowledge.

For the bTTN, the variational problem in Equation (23) aims at minimizing the energy in Figure 14b with respect to each tensor. Even in this case, the underlying idea is to solve local optimization problems for each tensor in the network, defining an effective Hamiltonian. Let's look at Figure 15 and suppose we want to solve the variational problem for the red tensor. It means that all other tensor entires are kept fixed. In this way we can contract the Hamiltonian with all other tensors in the network, obtaining the new tensor network on the right hand side of the figure, which represents an eigenvalue problem for the red tensor and the effective Hamiltonian $H_{eff}$. Once it has been solve, the red tensor gets replaced by the eigenvector corresponding to the lowest energy eigenvalue. This is done for all tensors in the network and a single optimization of all tensors is called a *sweep*. This procedure can be applied iteratively until convergence to the global ground state is reached.

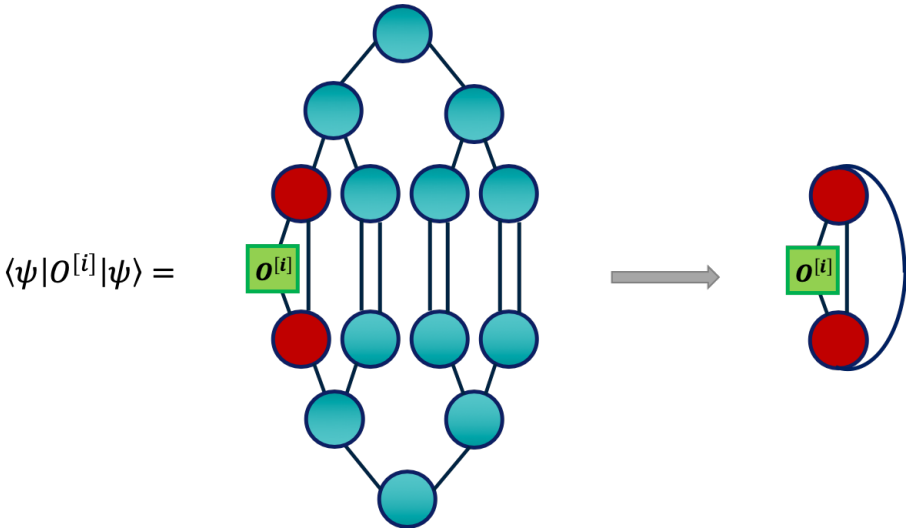



**Figure 16:** Computation of the expectation value of a local observable for the generic site $i$ of the lattice. The introduction of the isometries in the network allows to reduce the number of relevant contractions.

As already mentioned in the remarks at the end of previous section, all this procedure can be performed very efficiently if the Hamiltonian $H$ is represented using the TPO anstatz.

### 4.4.3 Computation of expectation values

Once all tensors in the bTTN have been optimized, we can use the ground state in order to compute the expectation values of observables. As we'll see in next chapter a case of particular interest for the purposes of this work is the computation of the expectation value of a local observable with support only on the $i^{th}$ single-particle space, as shown in Figure 16.
An efficient way to perform this operations is to isometrise (gauge) the network with respect to the red tensor, a concept in a way that the large part of the contractions lead to identities, resulting that only those represented one the right hand side of Figure 16 are actually relevant [42].
A bTTN can be isometrised with respect to any of its tensor and this results

in the fact that the number of contractions to perfom in order to compute the expectation value of local observables is $\mathcal{O}(1)$.

# 5 STUDY AND RESULTS

In this chepter the whole process done in order to apply TNM to the solution of the dose optimization problem in IMRT will be discussed. In the first part the mathematical procedure implemented to convert the classical objective function into an Ising-like hamitonian we'll be presented. This procedure we'll be validated using an algorithm to perform classical iterative search along the diagonal of the Hamiltonian. The toy-model we built to this aim will be described. Finally, in the second part the results obtained with the bTTN algorithm both on the toy-model and on two more realistic cases will be presented.

## 5.1 MATHEMATICAL DESCRIPTION OF THE PROBLEM

In Section 2.3 the mathematical details about the formulation of the IMRT treatment's goals in terms of mathematical constraints have been discussed. In this section discuss how the quadrati cost function in Equation 6 can be converted into an Ising-like model Hamiltonian. We know from discussion in Section 2.3 that the beamlets weights are real valued and such that $x_j \geqslant 0$. In this study they will be always normalized in order to have $x_j \in [0; 1]$.
A description in terms of real values doesn't fit the one in terms of descrete spin orientations, thus we need to convert the $x_j$ into a discrete representation. The procedure we implemented is based on the *decimal-binary* conversion:

$$x_j \approx \frac{1}{(B/2)} \sum_{l=0}^{L-1} 2^l b_l^{(j)} \tag{30}$$

with L the *bit-depth*, i.e. the maximum number of bits used to represent the number, $b_l^j = \{0, 1\}$ and $\frac{B}{2} = \sum_{l=0}^{L-1} 2^l$ a normalization constant needed to keep $x_j$ in the interval $[0; 1]$. The number L is very critical in this context, since it represents the resolution in the representation of final weights. $2^L$ corresponds to the number of discrete levels the interval $[0; 1]$ si subdivided into.
In real clinical applications, after any optimization procedure performed with classical algorithm, the final real-valued beamlets weights are converted into descrete numbers too. Therefore, L is always a critical number, independently from the optimizer used. Typical values range from $L = 3,\ 4$ or 5.
In order to describe the model as a spin system, each $b_l^{(j)}$ has to be mapped into $s_l^{(j)} = \{-1, +1\}$, as:

$$s_l^{(j)} = 2b_l^{(j)} - 1 \iff b_l^{(j)} = \frac{s_l^{(j)} + 1}{2} \tag{31}$$

giving the correspondence $b_l^{(j)} = 0 \longleftrightarrow s_l^{(j)} = -1$ and $b_l^{(j)} = 1 \longleftrightarrow s_l^{(j)} = +1$.
Let's now go back to Equation (6) and expand the square:

$$F(\vec{x}) = \sum_i \gamma_i \left[ \underbrace{\left( \sum_j a_{ij} x_j \right)^2}_{H_1} + \underbrace{D_i^{2(P)} - 2D_i^{(P)} \sum_j a_{ij} x_j}_{H_2} \right] \tag{32}$$

We'll consider $H_1$ and $H_2$ separately. For compactness of notation, we'll negletc the sum over the number of voxels i in the calculations. It'll be considered only in the final results.

**COMPUTATION OF $H_1$.** This term can be rewritten as:

$$H_1 = \left( \sum_j a_{ij} x_j \right)^2 = \underbrace{\sum_j \left( a_{ij} x_j \right)^2}_{H_1^A} + \underbrace{2 \sum_{j<k} a_{ij} a_{ik} x_j x_k}_{H_1^B} \tag{33}$$

If we insert Equation (30) and (31) into the expression for $H_1$, we obtain:

$$\begin{aligned} H_1 = \sum_j \left( a_{ij} \frac{\sum_{n=0}^{N-1} 2^n \left( s_n^{(j)} + 1 \right)}{\sum_{n=0}^{N-1} 2^{n+1}} \right)^2 + \\ + 2 \sum_{j<k} a_{ij} a_{ik} \left( \frac{\sum_{n=0}^{N-1} 2^n \left( s_n^{(j)} + 1 \right)}{\sum_{n=0}^{N-1} 2^{n+1}} \right) \left( \frac{\sum_{m=0}^{M-1} 2^m \left( s_m^{(k)} + 1 \right)}{\sum_{m=0}^{M-1} 2^{m+1}} \right) \end{aligned} \tag{34}$$

Let's simplify the notation by setting:

$$\sum_{m=0}^{M-1} 2^{m+1} = \sum_{n=0}^{N-1} 2^{n+1} \equiv B \tag{35}$$

In this way, we have that:

$$H_1^A = \sum_j \frac{a_{ij}^2}{B^2} \left[ \sum_{n=0}^{N-1} \left( 2^n s_n^{(j)} \right)^2 + 2 \sum_{n<m} 2^n 2^m s_n^{(j)} s_m^{(j)} + \left( \sum_{n=0}^{N-1} 2^n \right)^2 + B \left( \sum_{n=0}^{N-1} 2^n s_n^{(j)} \right) \right] \tag{36}$$

where two constant terms are present:

$$\sum_i \gamma_i \left[ \sum_j \frac{a_{ij}^2}{B^2} \sum_{n=0}^{N-1} \left( 2^n s_n^{(j)} \right)^2 \right] \equiv C_1 \tag{37}$$

$$\sum_i \gamma_i \left[ \sum_j \frac{a_{ij}^2}{B^2} \left( \sum_{n=0}^{N-1} 2^n \right)^2 \right] \equiv C_2 \tag{38}$$

Since in the ground-state search the constants don't play any role, they can be neglected in the final expression of each term. For this reason, the final expression for $H_1^A$ becomes:

$$\boxed{H_1^A = \sum_i \gamma_i \left[ \sum_j \frac{a_{ij}^2}{B^2} \left( 2 \sum_{n<m} 2^n 2^m s_n^{(j)} s_m^{(j)} + B \sum_{n=0}^{N-1} 2^n s_n^{(j)} \right) \right]} \tag{39}$$

Using the procedure, $H_1^B$ is given by:

$$H_1^B = 2 \sum_{j<k} \frac{a_{ij} a_{ik}}{B^2} \left[ \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} 2^n 2^m s_n^{(j)} s_m^{(k)} + \frac{B}{2} \sum_{n=0}^{N-1} 2^n s_n^{(j)} + \frac{B}{2} \sum_{m=0}^{M-1} 2^m s_m^{(k)} + \frac{B^2}{4} \right] \tag{40}$$

Even in this case, we can rename the constant as:

$$\sum_i \gamma_i \left[ \frac{1}{2} \sum_{j<k} a_{ij} a_{ik} \right] \equiv C_3 \tag{41}$$

obtaining that the final expression for $H_1^B$ is:

$$\boxed{H_1^B = \sum_i \gamma_i \left[ 2 \sum_{j<k} \frac{a_{ij} a_{ik}}{B^2} \left( \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} 2^n 2^m s_n^{(j)} s_m^{(k)} + \frac{B}{2} \sum_{n=0}^{N-1} 2^n s_n^{(j)} + \frac{B}{2} \sum_{m=0}^{M-1} 2^m s_m^{(k)} \right) \right]} \tag{42}$$

**COMPUTATION OF $H_2$.** The first term of $H_2$ is a constant and it can be redefined as:

$$\sum_i \gamma_i \left[ D_i^{2(\text{pred})} \right] \equiv C_4 \tag{43}$$

Therefore, we obtaind the following expression for $H_2$:

$$H_2 = -\frac{2D_i^{(pred)}}{B} \sum_j a_{ij} \left[ \sum_{n=0}^{N-1} 2^n s_n^{(j)} + \sum_{n=0}^{N-1} 2^n \right] \tag{44}$$

If we now introduce the constant term:

$$\sum_i \gamma_i \left[ -D_i^{(pred)} \sum_j a_{ij} \right] \equiv C_5 \tag{45}$$

the final expression for $H_2$ becomes:

$$\boxed{H_2 = -\sum_i \gamma_i \left[ \frac{2D_i^{(pred)}}{B} \sum_j a_{ij} \left( \sum_{n=0}^{N-1} 2^n s_n^{(j)} \right) \right]} \tag{46}$$

Finally, the resulting Hamiltonian is:

$$H = H_1^A + H_1^B + H_2 + c \tag{47}$$

with $c$ the sum of all constant terms previously defined. $H$ describes a two dimensional system, with $N_B$ columns and $N_Q$ rows. $N_B$ corresponds to the number of beamlets, while $N_Q$ to the number of qubits[1]. In other words, we are representing each beamlet's weight $x_j$ with the spin configuration of $N_Q$ qubits, and the two are connected by a binary-decimal conversion. Interactions are typically *long-range*, but the number of terms is determined by the way the beamlets physically interact inside the tissues, as we'll see in the following.

1 It is the analogue of the bit in classical information theory and identifies a particle with two possible states.

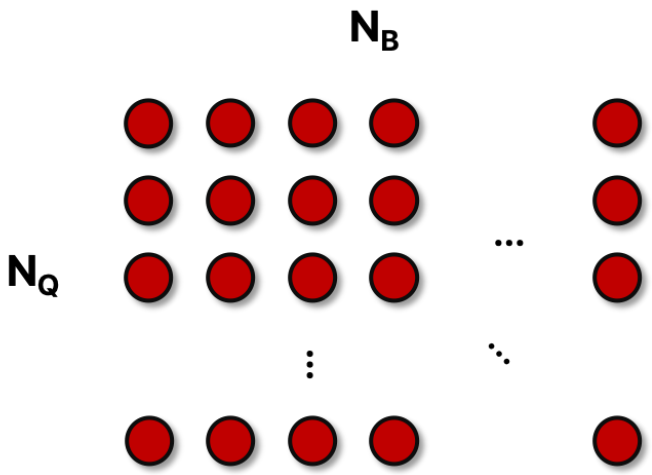



**Figure 17:** Representation of the two dimensional lattice described by the Ising-like Hamiltonian obtained from the conversion of the quadratic objective function.

We now need to rearrange the terms in such a way the final Hamiltonian takes an Ising-like form, such as:

$$H(\{h_{n(j)}\},\{J_{n(j),m(k)}\}) = \sum_j \sum_n h_{n(j)} s_n^{(j)} + \sum_{j,k} \sum_{n,m} J_{n(j),m(k)} s_n^{(j)} s_m^{(k)} \tag{48}$$

where $j$ and $k$ run over the $N_B$ beamlets, while $n$ and $m$ over the $N_Q$ qubits. The notation $n(j)$ and $m(k)$ stands for the $n^{th}(m^{th})$ qubit of the $j^{th}(k^{th})$ beamlet. In other words, we need to find an expression for the cuplings $(\{h_{n(j)}\},\{J_{n(j),m(k)}\})$. As far as the single-particle terms are concerned, they can be rearranged in the following way:

$$H_{SP} = \sum_j \sum_n \underbrace{\left[\sum_i \frac{\gamma_i}{B}\left(a_{ij}\sum_k a_{ik} - 2D_i^{(P)} a_{ij}\right) 2^n\right]}_{h_{n(j)}} s_n^{(j)} \tag{49}$$

On the other end, the two-body interaction term becomes:

$$H_{2B} = \underbrace{\sum_{j\neq k}\sum_{n,m}\underbrace{\left[\sum_i \gamma_i \frac{a_{ij}a_{ik}}{B^2}2^n 2^m\right]}_{J_{n(j)m(k)}} s_n^{(j)} s_m^{(k)}}_{\text{Qubits of different beamlets}} + \underbrace{\sum_j \sum_{m\neq n}\underbrace{\left[\sum_i \gamma_i \frac{a_{ij}^2}{B^2}2^n 2^m\right]}_{J_{n(j)m(j)}} s_n^{(j)} s_m^{(j)}}_{\text{Qubits of the same beamlet}} \tag{50}$$

where we have split $H_{2B}$ into two terms representing interactions of qubits belonging to different beamlets and the same beamlet, respectively. The coupling coefficients are represented by the terms in the square brakets.
The final procedure one needs to apply in order to solve the optimization problem once the classical objective function has been provided, is the following:

1. Fix the bit-depth $N_Q$;
2. Find the expression of the couplings $(\{h_{n(j)}\},\{J_{n(j),m(k)}\})$;
3. Search the ground state of the system with the desired algorithm (e.g. TNM);
4. Read the final state of the qubits in the system, which will have the general form:
$$\begin{bmatrix} s_1^{(1)} & s_1^{(2)} & \dots & s_1^{(N_B)} \\ s_2^{(1)} & s_2^{(2)} & \dots & s_2^{(N_B)} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N_Q}^{(1)} & s_{N_Q}^{(2)} & \dots & s_{N_Q}^{(N_B)} \end{bmatrix} \tag{51}$$
5. Decodify the weights $x_j$ using Equation (30) and (31).

We see that the problem has been formulated in a very general algorithm-independent form, where the optimization algorithm enters at step 3 only. In the following we'll first discuss an entry-level approach based on the iterative search of the lowest energy eigenvalue through the iteration over the diagonal(ized) Hamiltonian. It will be usefull to validate the procedure illustrated so far as well as to show the scaling of the computational time for inceasing

numbers of qubits and the necessity to change approach. Finally, in the second part we'll discuss the application of TNM.

## 5.2 CLASSICAL ITERATIVE SEARCH APPROACH

Let's consider Equation (48) and try to write it in a matricial form. This can be done by making the substitution $s_j^{(n)} \rightarrow \sigma^z$, with the latter the Pauli's matrix in the $z$ direction, obtaining the following expression:

$$H \equiv H(\{h_{n(j)}\}, \{J_{n(j),m(k)}\}) = \sum_j \sum_n h_{n(j)} \sigma_n^{z(j)} + \sum_{j,k} \sum_{n,m} J_{n(j),m(k)} \sigma_n^{z(j)} \sigma_m^{z(k)} \quad (52)$$

where $\sigma^z$ was chosen because it's both diagonal and real valued, but in general the same procedure could be applied also using $\sigma^{x(y)}$. The only important fact is that the same matrix needs to be used to represent each spin, since we are dealing with a classical problem, which means that operators have to commute one each other. Written in this form, the optimization problem becomes the following:

$$H|\psi\rangle = E|\psi\rangle \quad (53)$$

which is an eigenvalues problem. However, since the Hamiltonian depends only on $\sigma_z$, it's already diagonal. Therefore, finding the ground state reduces to iterate the diagonal of $H$ and take the smallest energy eigenvalue, whose corresponding eigenvector should represent the solution to the initial problem. Working in the so-called *computational basis*, the latter has all entries equal to $0$, except for the one in the position corresponding to the position of its eigenvalue in the diagonal of $H$. For, istance if we consider the eigenvalue in the $4^{th}$ position, the corresponding eigenvector will be:

$$\psi_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix} \tag{54}$$

with the length of the vector equal to the dimension of the many-body Hilbert space. Some remarks about this procedure:

- Equation 54 is a good representation of the generic ground state because we're solving a classical problem and we expect to end up with a product state, meaning the the state of the qubits in the lattice at the end of the optimization is well determined and no quantum uncertainty is present;
- It may happens that more than one configuration fit the constraints, leading to a degenerate ground state: we keep one since they have all the same energy[2]. From the practical point of view, this degeneration arises from the physical interaction between different beamlets;
- The solution is *exact*, meaning that no approximation is introduced in the ground state research.

From this discussion, it looks like that this very basic approach is enough to solve efficiently the problem. However if we consider again Equation (52), we already know from our discussion in Chapter 3 that its dimension increases exponentially with the number of qubits in the system. We know that the final Hamiltonian is the result of the iterative application of the tensor product, hence we our ability to solve the problems depends on the efficiency in the representation of the final Hamiltonian. Furthermore as a consequence of the exponential scaling of the dimension of the many-body Hilbert space, the time the algorithm takes to iterate all the entries of the diagonal increases exponentially too. Actually, as far as the first problem is concerned, we're particularly

2 Within the numerical precision.

lucky, since we're using only $\sigma^z$ to represent the spins, thus typically a recursive formula can be found to compute a given entry of the diagonal without the necessity of computing the tensor product and storing large matrices. We implemented an algorithm which can do that. The real problem is related to the second question: if the length of the diagonal increases exponentially, also the time to iterate all its elements does. In Section 5.4 we'll provide a quantitative analysis of this fact, definitely pointing out the reason why it's necessary to change approach.

## 5.3 A TOY-MODEL: THE 3D BOX

In order to validate the procedure discussed in Section 5.1 we built a toy-model. It was realized with the aim of reproducing a simplified version of a 3D medical image of the body, where the tumor and other regions of interest can be identified. It was realized with the idea of simplifying as much as possible the description of a real system, owning its basic properties at the same time. We'll see during the discussion that it is a useful tool to investigate on the properties of the different algorithms used. The model consists in a 3D box of arbitrary dimensions where the following parameters can be tuned:

- Number of voxels on the $x$, $y$ and $z$ directions;
- The number of region of interests in the box;
- The dose prescription $D_i^P$ for each voxel in the box;
- The number of beams and their directions;
- The number of beamlets for each beam.
- The beamlets dose profile, i.e. the dose released by each beamlet $j$ to each voxel $i$ as a function of the depth;
- The priority $\gamma_i$ for each voxel.

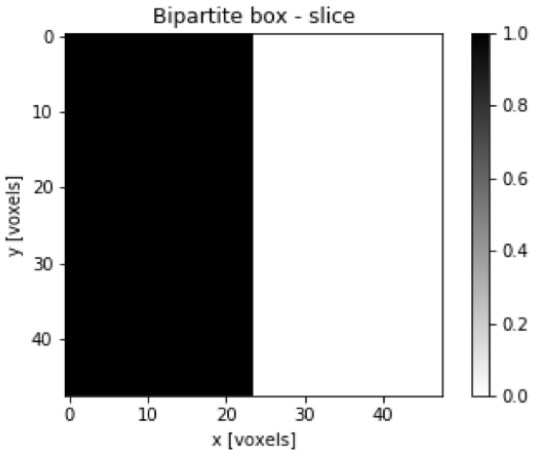
The following simplifications were introduced:

- The possible beam directions were limited to $0°$ and $180°$. This was done in order to simplify the geometrical description of the beam propagation in the box;
- The dose profile can be any analytical function of the distance, but we always worked with an *ideal beam* whose dose deposit is the same for each voxel;
- The number of voxel in a particular direction is always chosen as a multiple of the number of beamlets, in order to avoid the case of beamlets hitting only part of voxels;
- The dose prescription were chosen in order to end up with a system where the ground state is at $E_{GS} = 0$. This can be done by choosing dose values which are compatible with the number of discretization levels $2^{N_Q}$ used. In practice, this condition is not fullfiled in general, but in this first part of the study it was needed in order to check if everything was working properly.

In the end, the information we can extract from this model are the following:

- The influence matrix $a_{ij}$;
- The list of priorities $\gamma_i$;
- The list of prescriptions $D_i^P$.

They are exactly the same information a real therapy planning system needs in order to optimize the plan. These information were stored on separate .txt files, in addition to a file where the bit-depth clould be specified, in order to make them easily accessible to the optimization algorithm. In Figure 18a an example of a slice on the $(x, y)$ plane of a 3D dimensional made up of 48x48x48 voxels where a bipartition into two regions of interest was created is shown. The numbers (0 and 1 in this case), identify uniquely each region of interest.




**(a)**

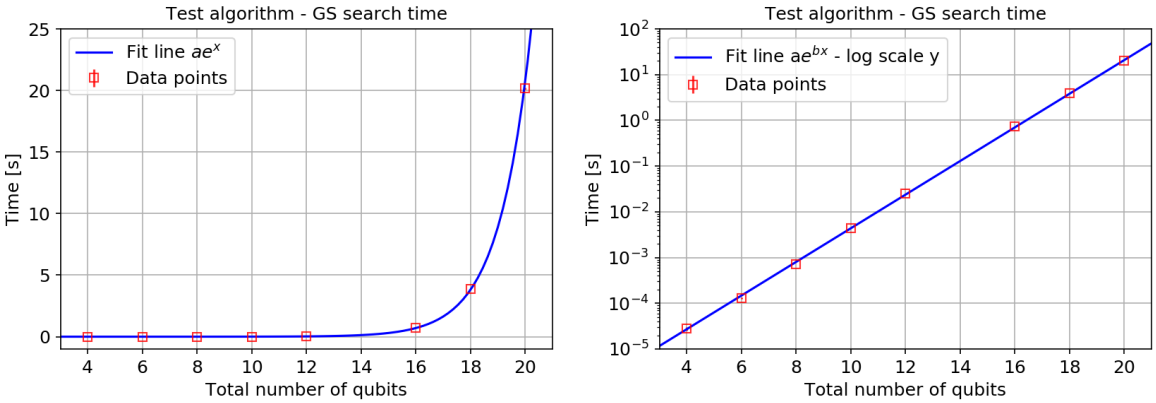
**Figure 18:** Example of a slice of the 48x48x48 voxels 3D box subdivided into two regions of interest.

## 5.4 3D BOX OPTIMIZATION: ITERATIVE SEARCH

In this section we'll show the main results obtained from the optimization of the 3D box with the entry-level algorithm for iterative search discussed in section 5.2. First, the results about the scaling of time with respect to the total number of qubits in the system will be discussed. Then, a first optimization result will be shown, in order to validate all the procedure explained so far.

### 5.4.1 Scaling of the computational time

We've already discussed about the pros and cons of the iterative-search of the ground state, pointing out that the main limit of the procedure is that the computational time is expected to increase exponentially as the total number of qubits in the system increases. In order to investigate on this particular behaviour, we collected samples ($N_{qubit}$, Time) for $N_{qubit}$ between 4 and 20. In particular, all the samples were collected on the bipartite box shown in Figure 18a, by choosing different number of beamlets (1 or 2) and different




**(a)** **(b)**

**Figure 19:** Computational time vs total number of qubits in the system for the iterative-search algorithm. **(a)** Natural scale; **(b)** Logarithmic scale on the y-axis.

values for the bit depth (3,4 or 5), depending on the total number of qubits available. Two different doses were prescribed to the two sides the bipartition, varying the values in order to fullfill the condition $E = 0$ for the ground state. We perfomed time averages for each sample in order to obtained an extimate of the computational time for each of them. After that we performed a fit with an exponential function $t(N_{qubits}) = ae^{bN_{qubit}}$. The results is shown in Figure 19, either in natural and logarithmic scale. The fitting parameters are $a = (9.2 \pm 0.8) \times 10^{-8}$ and $b = (0.847 \pm 0.005)$. The optimizations were run on ASUS X750J equiped with Intel i7 $4^{th}$ gen., 8 GB RAM.

In Table 1 the values used for the fit are reported. The function obtained was used to make a prediction of the computational time for higher system's dimension. We see from the table that simulating a system of 32 qubits would take about 148 hours, while for 64 qubits about $10^{13}$ hours would be required. As explained before when the algorithm was described, this is the main limit of this entry-level approach and the main reason why a more efficient algorithm needs to be used in order to deal with higher system dimensions.

| $N_{qubits}$ | Time [s] |
|---|---|
| 4 | $(2.80 \pm 0.30) \times 10^{-5}$ |
| 6 | $(1.29 \pm 0.04) \times 10^{-4}$ |
| 8 | $(7.22 \pm 0.05) \times 10^{-4}$ |
| 10 | $(4.41 \pm 0.04) \times 10^{-3}$ |
| 12 | $(2.52 \pm 0.01) \times 10^{-2}$ |
| 16 | $(7.37 \pm 0.03) \times 10^{-1}$ |
| 18 | $(3.88 \pm 0.02) \times 10^{0}$ |
| 20 | $(2.02 \pm 0.06) \times 10^{1}$ |
| 32 | $\approx 148$ h |
| 64 | $\approx 10^{13}$ h |

**Table 1:** Data points in Figure 19 and extrapolation of the computational time for systems of 32 and 64 qubits.

### 5.4.2 First optimization result

We show here the result of the optimization of the 3D bipartite box. In this particular case two opposite beams of two beamlets each one were used. The bit depth was set to $N_Q = 4$, resulting in a system of 16 qubits, a number which can be easily approached with the algorith for iterative-search. The dose for the left hand side was set to $D^P_{left} = 6$, while it was set to $D^P_{right} = 14$ for the right hand one. Doses are expressed in arbitary units. The maximum dose delivered by each beamlets was set to $D_{max} = 15$ In Figure 20a and 20c the final configuration for each beamlet is shown, together with the final intensity map, which results in $x^{tot}_{left} = 0.4$ and $x^{tot}_{right} = 0.933$. If we multiply the two values by $D_{max}$ we obtain precisely the dose prescription imposed at the beginning.
This simple example has a central importance, since it's a proof of the fact that the procedure we implemented for the convertion of the classical function into an Ising-like Hamiltonian is correct. Now we're ready to go on with the discussion, focusing on the core of the work: the application of tensor network methods to the solution of the problem.

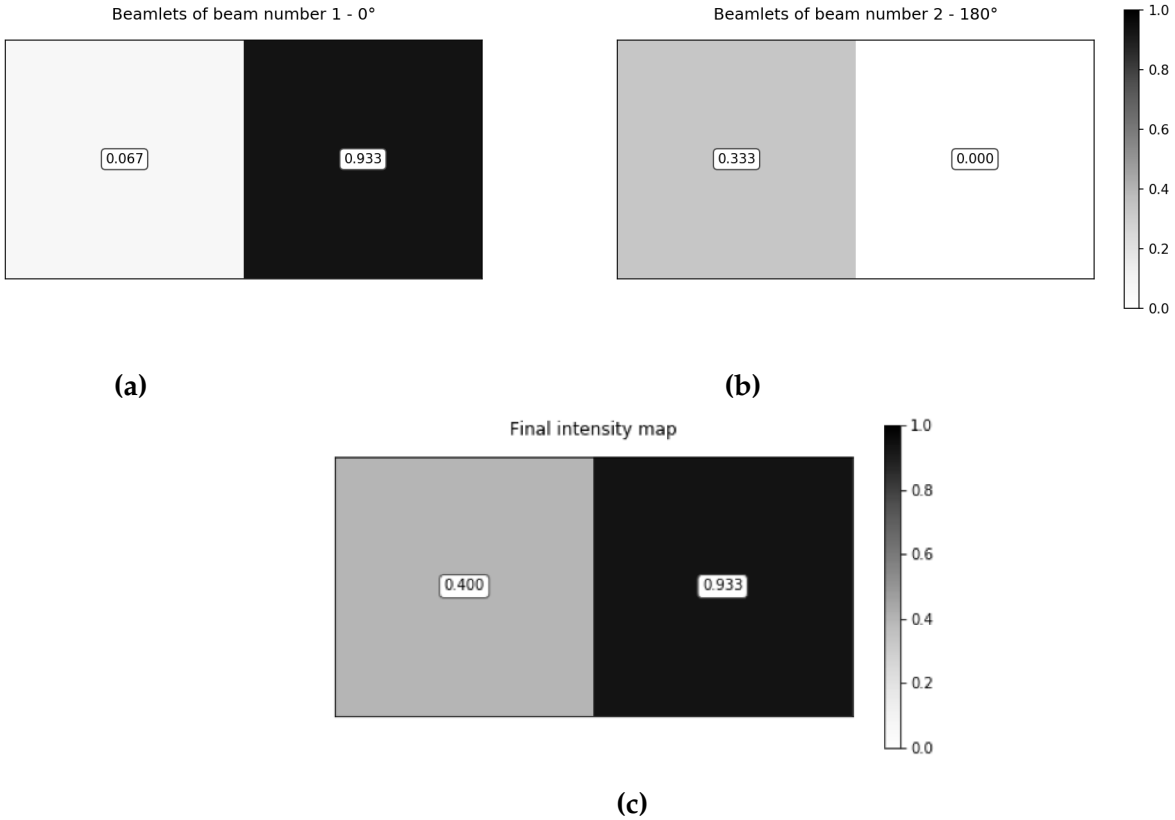



**Figure 20:** Result of the optimization of the 3D bipartite box in with two opposite beams of two beamlets each one. **(a)** Final beamlets' values for the first beam; **(b)** final beamlets' values for the second beam; **(c)** final intensity map as result of the sum of the other two.

## 5.5 bTTN ALGORITHM: AN OVERVIEW

The algorithm used to find the GS of the Hamiltonian is very sophisticated and its complete study and description goes beyond the aim of this thesis. However, we want to give an idea about its functioning. The algorithm implements a bTTN and it was coded in order to work either with 1D or 2D system.
The main elements needed to the algorithm to perform the optimization are the following:

1. The dimensionality of the problem, namely the number of qubits in the $x$ and $y$ directions in our case;

2. The coupling coefficients, both for the local and interaction terms, as well as the operators involved ($\sigma_z$ in our case);

3. The maximum bond dimension $\chi_{max}$, which determines the maximum dimension of the virtual links;

4. A rule to stop the algorithm when convergence has been reached.

5. The observables to measure at the end of the computation (spin orientation for each qubit in our case).

As far as point 3 is concerned, the convergence can be set both in terms of maximum number of sweeps and in terms of relative error on the enery values over a given number of sweeps. In other words, the algorithm is stopped for having reached the convergence if the error on the last $n^{conv}_{sweeps}$ sweeps is $\epsilon_{rel} \leqslant \epsilon^{thr}_{rel}$. The choice of both $n^{conv}_{swees}$ and $\epsilon^{thr}_{rel}$ depends on the specific problem. In particular, it turns out that the minimum $\epsilon^{thr}_{rel}$ depends on the bond dimension: too small values can lead to errors in the optimization routines.
There's also a more practical consideration to be made for $\epsilon^{thr}_{rel}$. Its values determines the accuracy on the final solution. Typically, dose prescriptions are given with tolerances of the order of *some percent*, meaning that it might be useless to perform optimization with a threshold set to a very small value, since

higher threshold could lead to equally good results, from the clinical point of view. In other words, it might happen that a solution found by the algorithm is sufficient even though it's not the absolute minimum. This consideration is very general, the mapping between the errors on the energy and on the dose is not so straight forward, but it's something which needs to be taken into account.
Another important feature provided by the algorithm, is the possibililty to be run on multiple threads, in order to reduce the computational time. This features becomes particularly usefull when high dimensional systems are considered, with a big number of interaction terms.

## 5.6 3D BOX OPTIMIZATION: BTTN ALGORITHM

In this section the application of the bTTN algorithm to the toy-model of the 3D box presented before will be discussed. The description will follow the same structure as in Section 5.4. First, we'll describe the scaling properties of the algorithm as a function of the number of qubits in the system. The role of the bond dimension will be then discussed, being it the central parameter in any tensor network algorithm. In the last part, we'll show an optimization result on the bipartite box described before.

### 5.6.1 Scaling of the computational time

From the information provided in Section 4.4 about the solution of the variational problem in Equation (23) with bTTN, we know that the ground state search is an iterative process. Tensors get optimized one after the other by solving local eigenvalues problem for the effective Hamiltonian at each position in the tree. A single complete optimization of all tensors in the three is called *sweep*, and the number of sweeps needed to reach the result depends on the specific problem.

| $N_Q$ | 16 | 64 | 256 |
|---|---|---|---|
| **Time [s]** | $6.7 \pm 0.6$ | $33 \pm 1$ | $175 \pm 2$ |

**Table 2:** Values of the points in Figure 21

The numerical complexity for the tree tensor network scale as $\mathcal{O}(NS\chi^4)$, with N, S and $\chi$ the number of qubits in the physical sites, the number of sweeps and the bond dimension, respectively. In other words, given a value for the bond dimension, we expect the see a linear scaling of the computational time as the number of physical sites increases. Furthermore, we can focus on the single-sweep time, since the number of iterations acts as a multiplication factor on it, leaving the general trend unthouched.
This property was investigated using the toy-model of the bipartite 3D box. We collected samples of 100 independent sweeps for three different system dimensions[3], 16, 64, 256, and $\chi_{max} = 6$. After the random initialization of the tree we did just one sweep. This operation was repeated 100 times for each value of N. In the end, the average and the standard error were computed for each data point1. We set $N_Q = 4$ and used two opposing beams. The dose delivered by $j^{th}$ beamlet to the voxels on its path was set to $D_j = 15$. The dose prescriptions were set to $D^P_{right} = 14$ and $D^{left}_P = 6$. The beamlets grid for each beam becomes thicker as the number of qubits in the system increases, as shown in Figure 22. It means that each voxel is always hit by two beamlets. From the result shown in Figure 21 we observe that in first approximation the expected linear trend is fulfilled, up to the limited number of points available.

In Table 2 the values of the points in the plot are reported. The simulation were run on an Asus X750J, Inteli/ $4^{th}$ gen., 8 GB RAM. Recalling the results obtained for the iterative-search algorithm, it's easy to notice the great improvement introduced by bTTN algorithm. For example, the time required to simulate a system of 64 qubits is $N_{sweeps} \times \mathcal{O}(10^1\ s)$, order of magnitued smaller than the $10^{13}$ found before, on the same Asus x750j, Intel i7, 8 GB RAM.

3 We were forced to work with $N = 2^K$ for computational reasons.

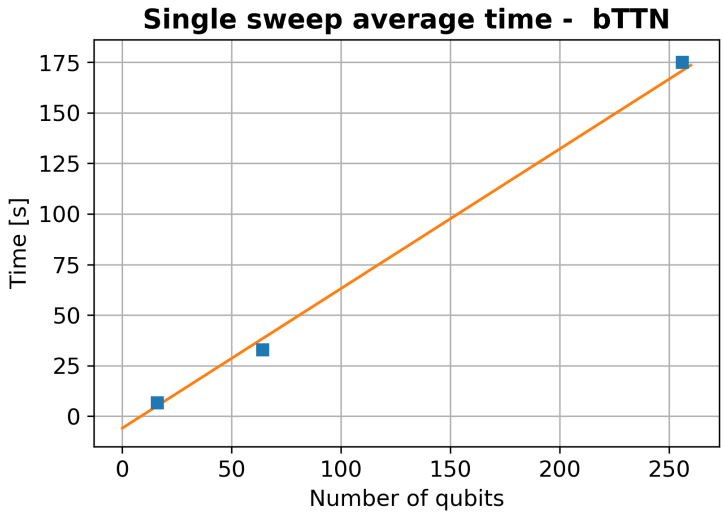



Figure 21: Trend of the scaling of the single-sweep computational time vs the number of qubits, bTTN algorithm

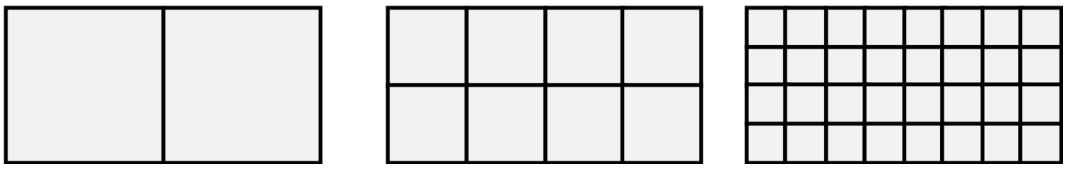


Figure 22: Beamlets grids for each beam for different number of qubits in the system. *(Left)* 16 qubits, 2 beamlets per beam; *(Center)* 64 qubits, 8 beamlets per beam; *(Right)* 256 qubits, 32 beamlets per beam.

In the end, bTTN looks like it is an effective choice in order to simulate high dimensional system. However, we'll see that the price to pay in order to keep the computational time low is that the convergence to the optimal solution is not always guaranteed. From this point of view ground state search with bTTN is a statistical process determined by the random intialization of the tensors.The probability to get to the minimum increases with the bond dimension.

### 5.6.2 A zero-th order analysis

We want now to give an idea about the statistical nature of the ground state search for increasing systems sizes. We collected samples of 60 independent

| $N_{qubits}$ | 16 | 64 | 256 |
|---|---|---|---|
| $f(E_{GS} = 0)$ | 60/60 | 2/60 | 0/60 |

**Table 3:** Frequency of $E_{GS} =$ for three system dimensions on samples of 60 independent optimizations. $\chi_{max} = 6$.

and randomly initialized optimizations for each of the three system sizes and counted the number of times the algorithm reaches the $E_{GS} = 0$ ground state. We set the threshold for the convergence to $\epsilon_{rel}^{thr} = 10^{-5}$, requiring it to be satisfyed for $n_{sweeps}^{conv} = 10$ sweeps. From Table Table 3 we see that the frequence of $E_{GS} = 0$ in the samples descreases as the the system size increases. The values are listed in Table 3. There are two important fact to mention about this procedure:

- 60 points in a sample is a very small number considering that we're working with random initialization of the tree. The result simply gives an idea of the complexity of reaching the GS when the system size increases
- The value found for 256 qubits, which is 0/60, doesn't mean that it's *impossible* to reach the GS for a system of that size. Even in this case, the number of points in the sample plays a crucial role: it may mean that the probability is very low, thus more points are needed in order to see at least once the $E_{GS} = 0$ ground state.

### 5.6.3 The role of the bond dimension

In general, we solve a complex quantum spin system with long-range interactions. Even though we expect a "classical solution" (product state), the optimization may require to pass through complex quantum states with high correlations. In the end, the product state is obtained by truncating all virtual links in the tree to $\chi_{max} = 1$. The role of the bond dimension in the ground state search was investigated by considering a fixed system size (64 qubits, the same used before) and collecting samples of optimizations for different values of $\chi_{max}$.

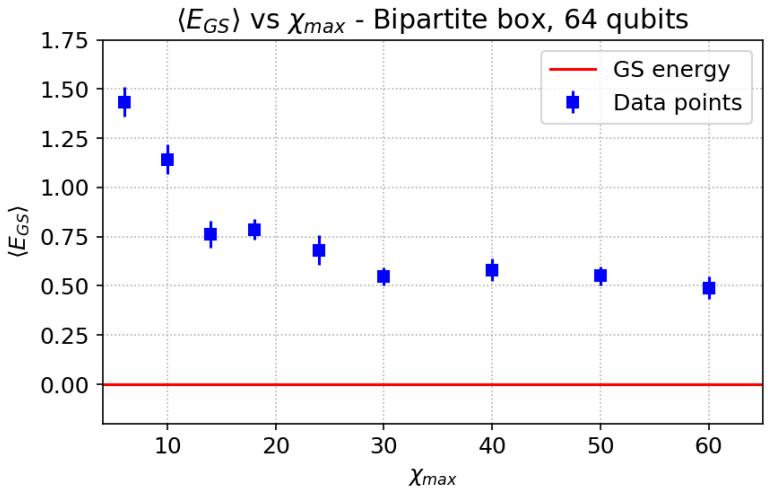
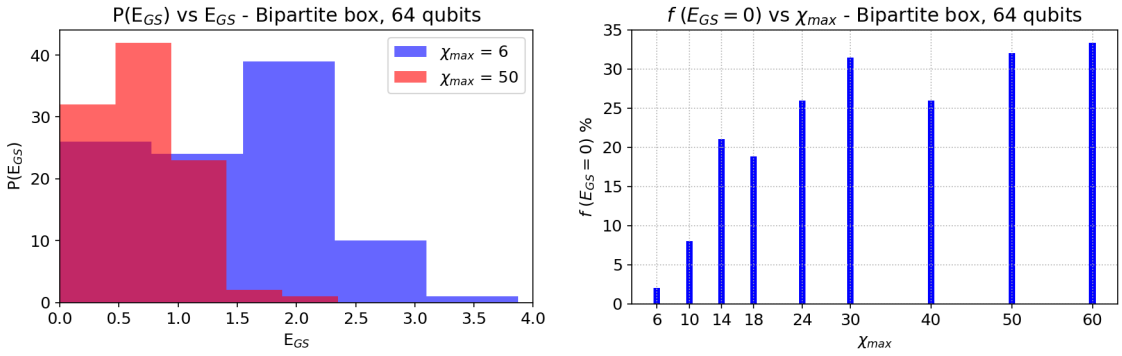



**Figure 23:** Trend of the average GS energy returned by the bTTN algorithm for different values of the maximum bond dimension $\chi_{max}$.

In Figure 23 the average energy $\langle E_{GS} \rangle$ for different values of the bond dimension is shown. The global decreasing trend and it seems to be compatible with a polynomial function. In Figure 24a a comparison between the distribution of $E_{GS}$ for $\chi_{max} = 6$ and $\chi_{max} = 50$ is shown and we see that the latter is visibly shifted towards smaller energy values. The statistical difference among the two distributions was probed using the Mann-Whitney non-parametric statistical test for indepedent samples[4]. The null hypothesis is rejected with a $p_{value} = 2.7 \times 10^{-13}$, which confirms that the two distributions are significantly different for both $\alpha = 0.05$ and $\alpha = 0.01$, being $\alpha$ the threshold such that if $p_{value} < \alpha$ the null hypothesis is rejected. This means that the higher bond dimension guarantees a smaller average error on the ground state $E_{GS}$.
On the other hand, in the region form $\chi_{max} = 30$ to $\chi_{max} = 60$. the situation is different. The results of the MW test between the distributions of $E_{GS}$ for $\chi_{max} = 60$ and $\chi_{max} = 30$, 40 and 50 are reported in Table 4. No significant differences arise from the test, showing that working with $\chi_{max} = 60$ doesn't give an siginificat improvement with respect to $\chi_{max} = 30$.

4 The null hypothesis $H_0$ is that the two samples are not statistically different, while the alternative hypothesis $H_1$ is that they are.




**(a)** **(b)**

**Figure 24:** **(a)** Comparison of the distributions of $E_{GS}$ for $\chi_{max} = 6$ and $\chi_{max} = 50$; **(b)** Frequency of the $E_{GS} = 0$ for the different values of the bond dimension $\chi_{max}$.

| $\chi_{max}$ | **30-60** | **40-60** | **50-60** |
|---|---|---|---|
| $p_{value}$ | 0.25 | 0.25 | 0.42 |

**Table 4:** Result of the Mann-Whitney test between the $E_{GS}$ distribution for $\chi_{max} = 60$ and $\chi_{max} = 30,\ 40$ and 50.

In Figure 24b a plot showing the number of time $E_{GS} = 0$ was found for each sample is presented. We see the global trend increasing trend.
It is intuitively expected to observe better results for higher $\chi_{max}$. On the other hand, we know that the numerical complexity in a bTTN scales as $\mathcal{O}(NS\chi^4)$, thus we should work with the lowest possible value of the bond dimension to keep the time as smaller as possible. It's also true that higher bond dimension could lead to the desired result in a lower number of iterations, as a significantly higher subspace of the complete Hilber space is explored. It comes out that specific consideration have to be made in each specific case. We'll see for example in Section 5.7.1 that a very good result can be obtained on the apparently very complicated sphere, even with $\chi_{max} = 2$ and in a few iterations. In next section we'll go deeper into the problem showing that, in general, even configurations with a slightly higher energy than $E_{GS} = 0$ could lead to a good result as far as the beamlets configuration is concerned.

### 5.6.4 Optimization example

In Figure 25 the final beamlets configuration for the system of 64 qubits is shown in the case of $E_{GS} = 0$. As in the case of the iterative search, the results show that the prescriptions are fullfilled. It is interesting to see also which are

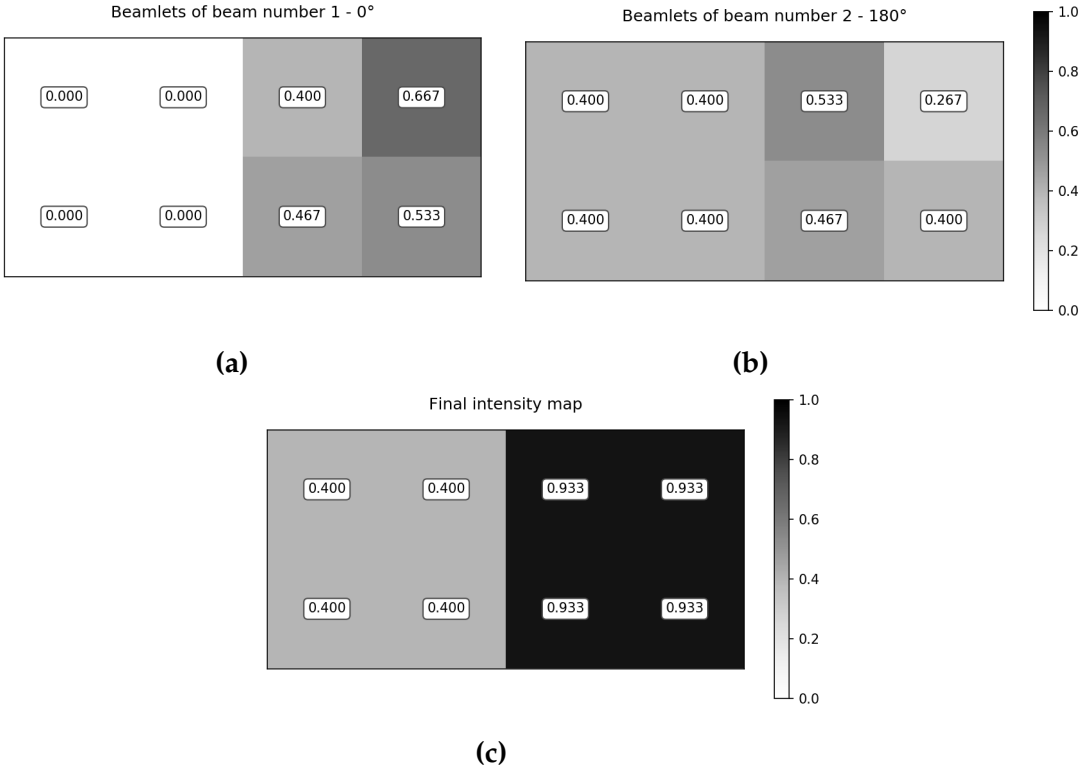



**Figure 25:** Resulting configuration of the beamlets for the first-lowest-energy $E_{GS} = 0$ optimization result. **(a)** Beamlet at 0°; **(b)** Beamlet at 180° (looking from 0°); Final beamlets configuration.

the configurations corresponding to higher energy values. In Figure 26 and 27 example of configurations corresponding to $E_{2nd} = 0.13824$ and $E_{3rd} = 0.55296$ are shown, respectively. They correspond to the second-lowest and the third-lowest energies in the sample. We observe for the former the configuration is exactly the same as for $E_{GS} = 0$, which is the prescribed one. The reason of this fact is the following. There are cases where the algorithm doesn't end up in a product-state, even after cutting the bond dimension to $\chi_{max} = 1$. The state is

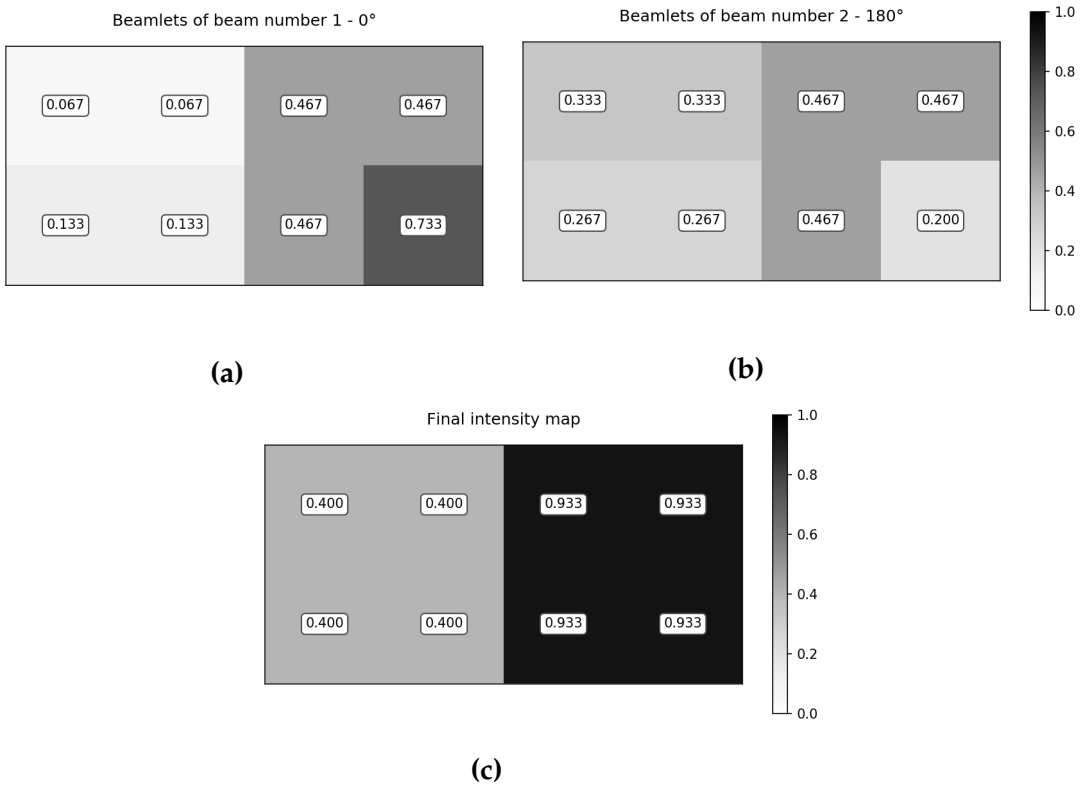



**Figure 26:** Resulting configuration of the beamlets for the second-lowest-energy $E_{GS} = 0.13824$ optimization result. **(a)** Beamlets at 0°; **(b)** Beamlet at 180° (looking from 0°); Final beamlets configuration.

in a superposition and the uncertainty is found also in the expectation value of the local observables, which takes values in $[-1, 1]$ and not exactly $\pm 1$. The interesting fact is that if we choose the state with the highest probability by rounding positive and negative values to $+1$ and $-1$ respectively, the solution we find is the one corresponding to the closest product-state.

In the second case, we see that the prescription is not satisfied for one weight only. In particular, the weights is $x = 0.533$ instead of $x = 0.400$. We already know rom Section 5.1 that beamlets are discretized with $N_Q = 4$. Refering to Equation 51, the beamlets are decodifies using binary-decimal convertion proceding from the top to the bottom. The difference between the two values is $\Delta_x = 0.133 = 2/(2^4 - 1)$, meaning that they differ only for the flip of one

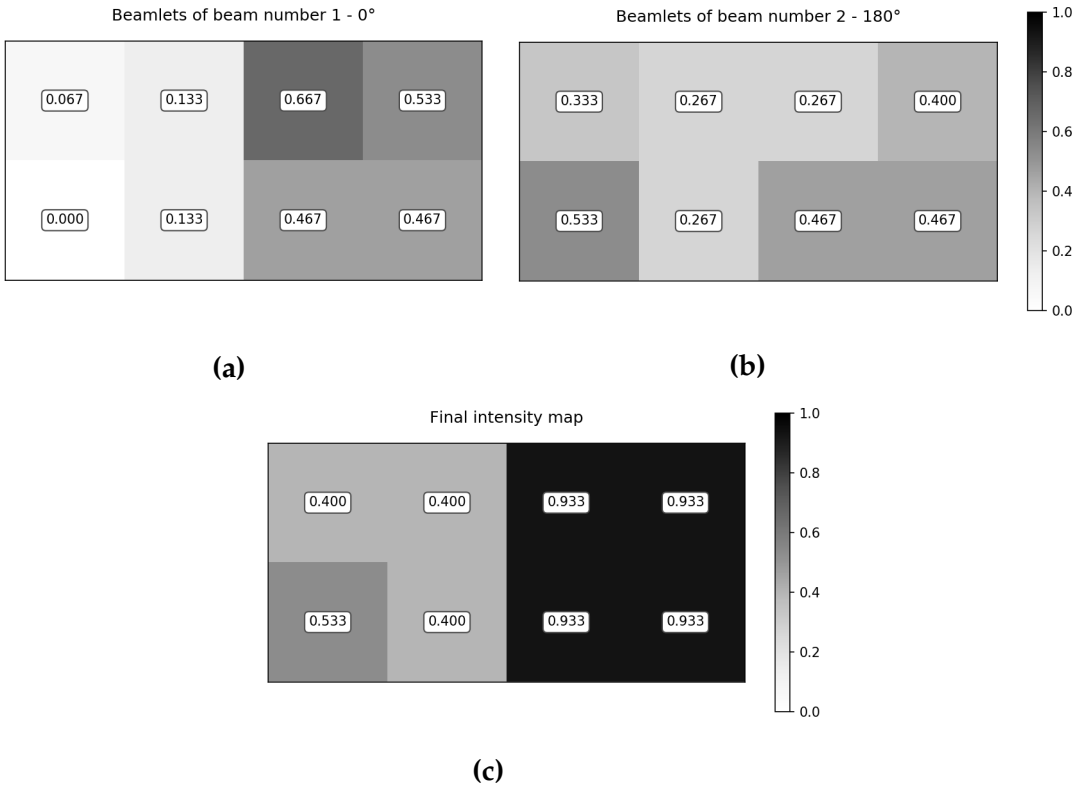



**Figure 27:** Final configuration of the beamlets for the third-lowest-energy $E_{GS} = 0.55296$ optimization result. **(a)** Beamlet at 0°; **(b)** Beamlet at 180° (looking from 0°;) Final beamlets configuration.

qubit in the second position. This fact is very interesting, since it show that the algorithm got extremely close to the best configuration. The are two important remarks to do:

1. In general, we don't know which is the value of $E_{GS}$, and thus how much far we are from the best configuration. Furthermore, we don't know if this trend is confirmed in general for more complex cases. If it were, once the optimization has finished we could iterate the lattice and flip the qubits in order to see whether the energy decreases. If only a few qubits need to be flipped, this procedure will not be computational expensive and we can get to the global minimum. The drawback is that the presence

of local minima for more complex functions could make things more difficult.

2. What we need is an optimal radiotherapy plan. The result we obtain can be evaluated with different tools, and in general a very good plan (where *good* means *compatible with the clinical requests*) can be obtain even if the global minimum has not been reached. This is one of the factors to take into account in real cases, where typically the *best* plan is always a compromise between accuracy and computational time.

In next section we'll show two applications of bTTN algorithm to more realistic cases.

## 5.7 APPLICATION OF bTTN TO REAL DATA

In this section we want to present and discuss two applications of the bTTN algorithm to more realistic optimization problems. In 5.7.1 we'll consider a sphere made up of an homogeneous medium, while in 5.7.2 we'll consider a real medical image of a prostate cancer, where two organs at risk, the urinary bladder and the rectum, have been countoured. There are two main features of the procedure adpoted which make this study very different from that on the toy-model:

1. The use of images in DICOM format, which is the standard format used to store medical images;

2. The use of a therapy planning system.

| | DICOM | Beam | Geometry | Scattering |
|---|---|---|---|---|
| **Toy-model** | No | Ideal | 0°, 180° | Yes |
| **Real case** | Yes | Real | Arbitrary | Yes |

**Table 5:** Comparison between the main features of the toy-model and the real case.

The second element has to do mainly with the creation of the *influence matrix* . We used CERR software [2], a free and open-source therapy planning system based on Matlab. Even if its clinical use is forbidden, it works in the same way as commercial therapy planning systems. For our purposes, the main advantage provided by the use of this software are the following:

- It works with real medical images in DICOM format;
- It allows to create IMRT plans with different beam geometries, allowing to set all typical parameters of interest.
- It works with *real beams*. The dose calculation can be performed using different algorithms, which can also account for the scattering of the particles in the body.
- The optimal number of beamlets is computed by the software itself. What can be set is the dimension of each beamlet in the $x$ and $y$ direction.
- It provides several tools for the visualization of the results, as well as for the comparisons among different plans. Some of these tools wll be used in the following.

In Table 5 a summary of the difference between the study on the toy-model and the real case is reported.

### 5.7.1 The homogeneous sphere

The homogeneous sphere was realized with the aim of creating a connection between the toy-model described before and real medical images acquired by means of CT. For this reason, we implemented a procedure to store the sphere

in DICOM format, as it were a real CT scan. This process was performed with the help of three different software, whose role is described in the following paragraphs.

**MATLAB FOR VOLUME CREATION.** We created a 3D array of dimensions (100, 100,100). In real unit it will correspond to a (10,10,10) cm box. The array was filled with all ones, and this is what makes the medium homogeneous. Using Matlab built-in functions, we converted it into a DICOM structure, associating to it all metastructures information needed to make it most possibly close to a real image. This square box plays the role of the body. The next step is to create contours for the regions of interest, the sphere in our case.

**BLENDER FOR 3D SPHERICAL REGION CREATION.** Using Blender, a free and open source 3D creation suite, we created a spherical surface of radius 2 cm and stored it in a separate file. The last needed step is to connect this structure to the image of the box, in order to make it ready for radiotherapy planning.

**3D SLICER FOR SEGMENTATIONS CREATION.** 3D Slicer is an open source software platform for medical image informatics, image processing, and three-dimensional visualization. One of the tools provided by this software is the possibility to import 3D structures in a proper format, such as those create with Blender, and use them as segmentations[5] for DICOM images. In the end, they can be saved into DICOM format, ready to be used by a real therapy planning system.

| $D^P$ | **# Beams** | **# Beamlets** | **Beamlets dim** | **Scattering method** |
|---|---|---|---|---|
| 1.0 Gy | 2 opposing | 64 | $(0.4 \times 0.4)$cm | QIB G+P+S |

**Table 6:** List of the main properties of the IMRT plan for the homogeneous sphere.

5 It's a synonim of contours

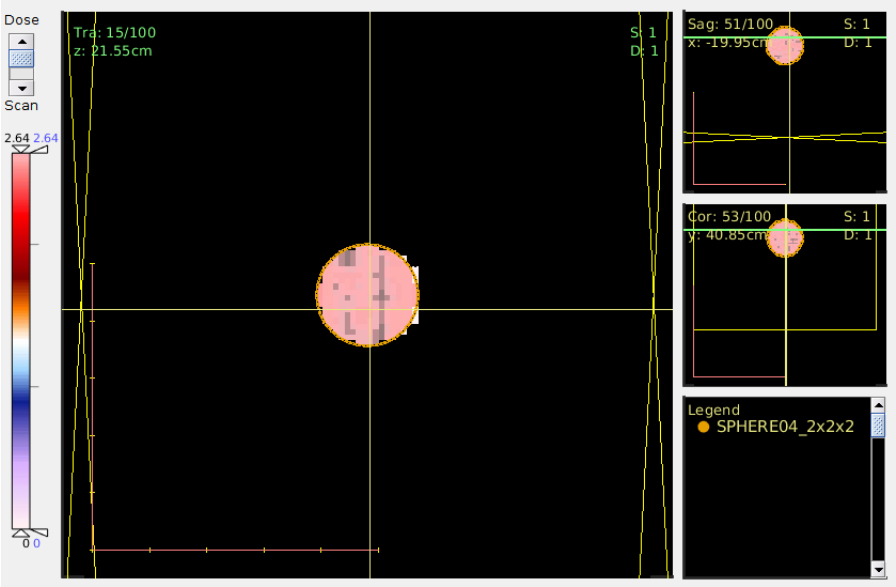



**Figure 28:** Dose distribution in the sphere before the optimization

We created an IMRT plan with the features listed in Table 6. The computation of the dose in each voxels per unitary values of the beamlets, i.e. the entries of the influence matrix, is computed using the CERR's build-in QIB algorithm [18], which accounts for the contribution of the primary beam as well as for the scattering. In Figure 28 an image of a slice of the sphere on the three planes before the optimization is show.
The dose prescription for each voxel was set to $D_i^P = 1.0\text{Gy}$. We'll see also in the following that in general it's convenient to normalize the prescriptions with respect to maximum prescription, keeping all values in the interval $[0; 1]$. This because we're interested in finding the relative weights of each beamlet, which is independent to the *dose scale*. Then, the dose can be rescaled to the desired value, by keeping the same beamlets weights. Another things which can help in general is to rescale the entries of the dose matrix so that $x_j \in [0; 1]$. As in the previous cases, starting from these quantities, the coupling coefficients of the Hamiltonian can be computed. The final Hamiltonian describes a system of 256 interacting qubits, where the number of interaction terms amounts to $N_{\text{inter}} = 32638$, a huge number if compared with the $\approx 700$ terms

for the toy-model with 256 qubits. The reason of this big number can be completely found in the more dense structure of the influence matrix, due to the presence of scattering. The scattering makes spreads particles in the volume and makes them hitting also voxels which are not on their path.

In Table 7 a summary of the main parameters of the simulation with bTTN

| **# qubits** | **# interactions** | $\epsilon_{rel}^{thr}$ | $n_{sweeps}^{conv}$ |
|---|---|---|---|
| 256 | 32638 | $10^{-5}$ | 4 |

**Table 7:** Main parameters of the bTTN algorithm for the optimization of the dose in the sphere and computational time result.

algorithm are reported. In this case, the value of the ground state energy is not known, thus other tools were needed in order to validate the quality of the result.

The same optimization was performed using the Matlab built-in function *quadprog*. The convergence of this function is guaranteed in case of convex optimization problems as in this case, thus it was taken as a reference. In particular, the beamlets configuration returned by *quadprog* was used in order to compute a reference energy to evaluate the performances of the bTTN algorithm. In Figure 29 the convergence of the algorithm of $\chi_{max} = 2,\ 5$ and 10 is shown. The green dotted line represent the reference value computed with *quadprog*, and we see that the bTTN solution converges to the reference as the $\chi_{max}$ increases.

The final distribution of the $x_j$ for the *quadprog*function and the bTTN algorithm are shown in Figure 30a and 30b, respectively. The two distribution are very different, there are only a few beamlets having the same values.

However, if we look at the cumulative DVHs in Figure 32, we see that the resulting dose distributions are very close each other, even though the two

| | $D_{min}$ | $D_{max}$ | $D_{mean}$ |
|---|---|---|---|
| **quadprog** | 0.95 | 1.05 | $0.99 \pm 0.01$ |
| **bTTN** | 0.95 | 1.05 | $1.01 \pm 0.01$ |

**Table 8:** Comparison between $D_{max}$, $D_{min}$ and $D_{mean}$ for the *quadprog* and the bTTN algorithm.

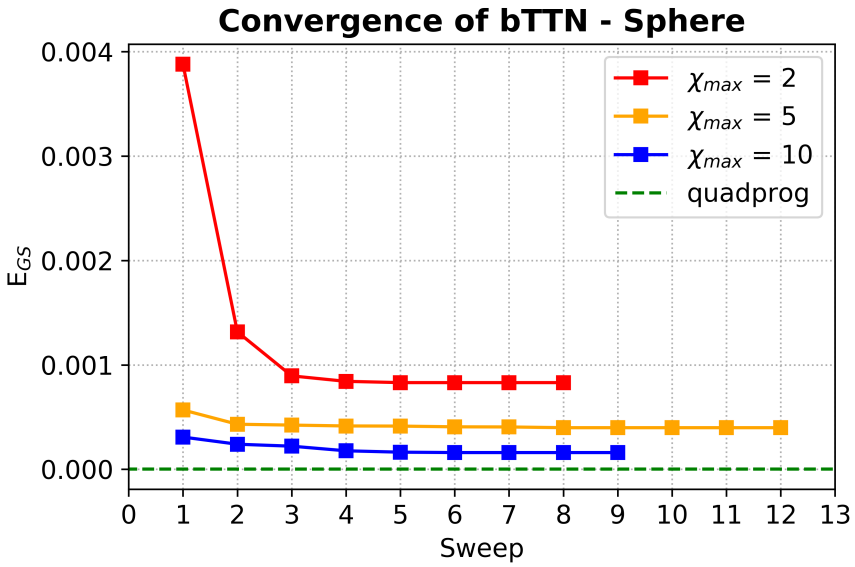
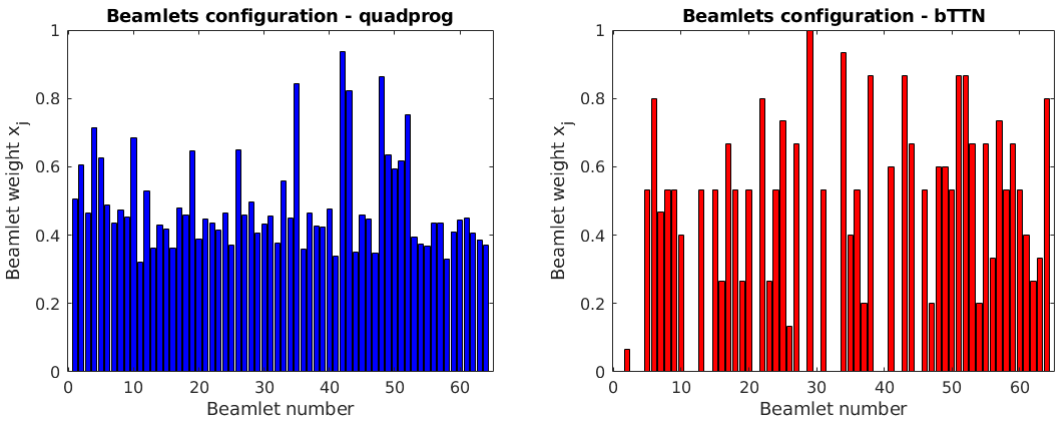



**Figure 29:** Convergence of the bTTN algorithm for different values of $\chi_{max}$ compared to the expected ground state energy computed with *quadprog*.

beamlets distributions are quite different. It's important to remember that the beamlets weights for the bTTN are descretized using $N_Q = 4$, which means that the interval $[0, 1]$ is subdivided into steps $\Delta x_j = 0.0666..7$. The fact that the bTTN algorithm works with discrete weights could be one of the reasons of this difference. At the moment, the only drawback of using bTTN seems to be the computational time. For the case of $\chi_{max} = 2$ the simulation took $\approx 2.5$h on the Cloudveneto cluster [35], while the one with *quadprog* only a few seconds on an ASUS X750J personal computer with Intel i7, $4^{th}$ generation, GB RAM.
In Table 8 the values of $D_{max}$, $D_{min}$ and $D_{mean}$ in the two cases is reported. The values confirm that the results obtained with the two algorithms are consistent. The reason why the prescribed dose is not respected for all the voxels but only on average is probably due to the particular geometry of the beams used.
To conclude, in Figure 31 an example of the dose distribution on a slice of the sphere after the optimization is shown. We observe that compared to the ini-




**(a)** **(b)**

**Figure 30:** Final beamlets values for **(a)** the *quadprog* function and **(b)** the bTTN algorithm.

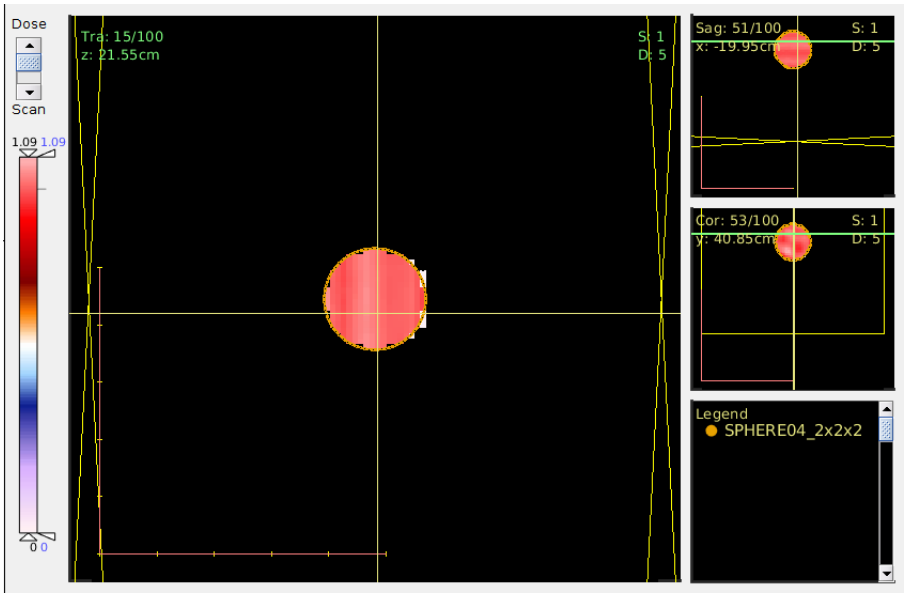



**Figure 31:** Example of the dose distribution on one slice of the sphere after the optimization with bTTN.

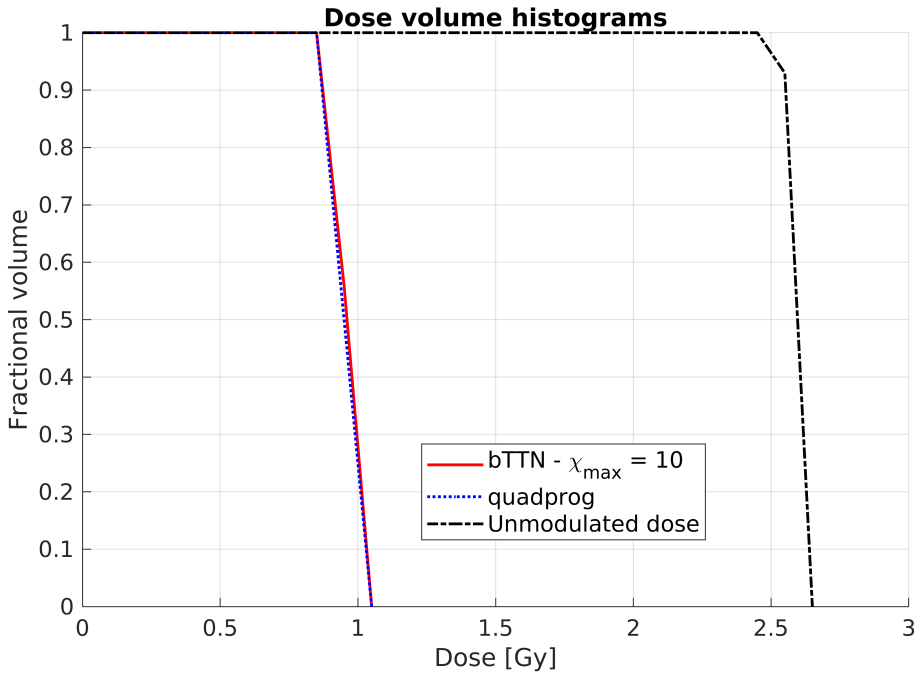



**Figure 32:** Comparison of the cumulative DVHs obtained with quadprog (blue line) and the bTTN (red line).

tial dose in Figure 28 the maximum is lower and the dose distribution is more homogeneous.

### 5.7.2 Prostate with two OARs

After having probed the functioning of the algorithm with the homogenoues sphere described in 5.7.1, we moved to the real case of a prostate cancer. The data used were taken from the database of [46].

The prostate was consider as the target, while the urinary bladder and the rectum as OARs. In Table 9 the main parameters of the treatment plan are re-

| # Beams | # Beamlets | Beamlets dim | Scattering method |
|---|---|---|---|
| $\theta_1 = 90°, \theta_2 = -90°$ | 64 | $(0.9 \times 0.9)$cm | QIB - G+P+S |

**Table 9:** List of the main properties of the IMRT plan for the prostate with two organs at risk.

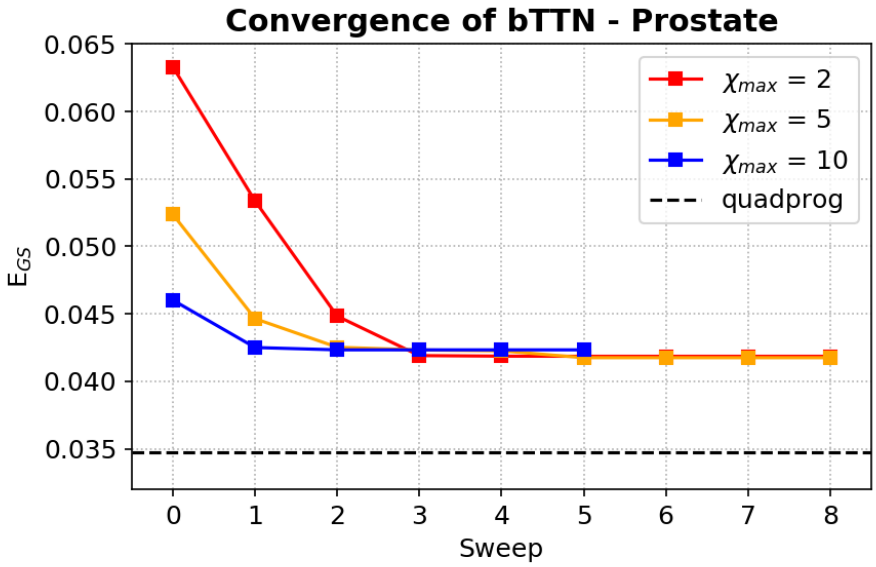


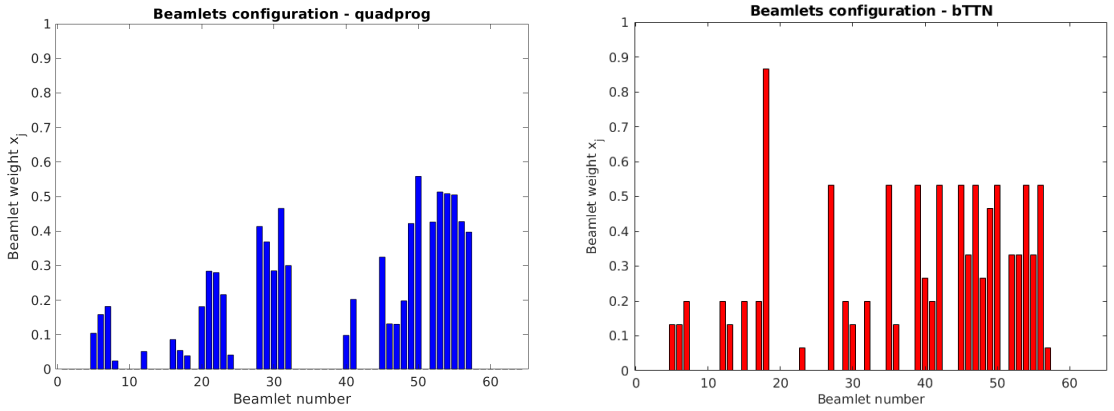

**Figure 33:** Trend of the ground state energy $E_{GS}$ obtained with bTTN on the prostate cancer for different values of $\chi_{max}$

| # qubits | # interactions | $\epsilon_{rel}^{thr}$ | $n_{sweeps}^{conv}$ |
|---|---|---|---|
| 256 | 32638 | $10^{-5}$ | 4 |

**Table 10:** Main parameters of the bTTN algorithm for the optimization of the dose in the prostate and computational time result.

ported. The dose for the each voxel of the prostate was set to $D^{p}_{prostate} = 1.0$ Gy; it was set to $D_{OARs} = 0.0$ Gy for the two OARs. Even in this case, the QIB algorithm to compute the influence matrix was used. In Table 10 the parameters of the simulation are repored. The interesting fact is that nothing has changed as far as the lattice is concerned with respect to the case of the sphere.

In Figure 33 the trend of $E_{GS}$ for varying $\chi_{max}$ is shown. The three results are very close each other and in this case a general descresing trend is not observed. It's probable that it could be observed by increasing $\chi_{max}$. In Figure 34 the resulting beamlets distributions the quadprog and bTTN algorithm are shown. Even though the distribution aren't equal, we observe that there are group of beamlets which are activated in both the two cases. This happens, for example, between 44 and 57, around 30 and before 10. The discretization




**(a)** **(b)**

**Figure 34:** Final beamlets values for **(a)** the *quadprog* function and **(b)** the bTTN algorithm.

could play some role even in this case.

In Figure 35 the resulting cumulative histograms are shown. The unmodulated dose is compared to the results from *quadprog* and bTTN. For the urinary bladder and the rectum the distribution coincide. Only small differences are observed for the prostate. The two results are consistent, showing the well functioning of the bTTN algorithm. Even in this case, the only drawback of using bTTN has to do with the computational time. For $\chi_{max} = 2$ the optimization took $\approx$ 3h on the Cloudveneto cluster. The optimization with *quadprog* took a few second on an ASUS X750J, Intel i7 $4^{th}$ gen., 8 GB RAM.

The reason why the initial prescriptions are not exactly satisfied for both *quadprog* and bTTN is a limit of using this very simple function. The priorities for the three organs were set all to $\gamma = 1$, meaning the all of them have the same importance. This lead to a conflict between the different objectives and the resulting dose is a compromise. In general, the use of more sophisticated cost functions including $D_{max}$, $D_{min}$, $D_{mean}$ or other different constraints, can help to give to the distribution the desired shape. Furthermore, we are using two opposing beam at $\theta_1 = -90°$ and $\theta_2 = 90°$. Different geometries as well as using an higher number of beams could improve the quality of the result. In

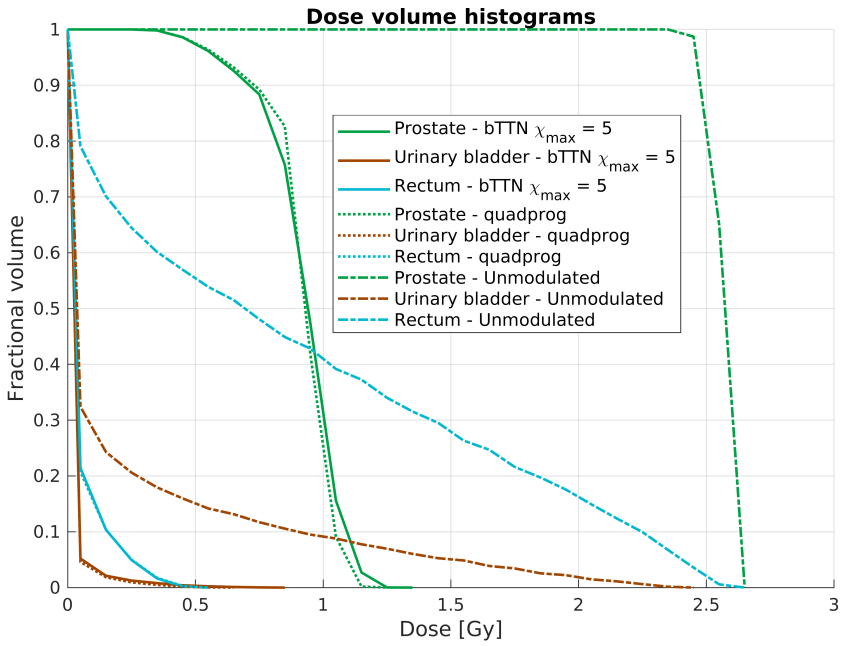



**Figure 35:** A comparison between the unmodulated dose distribution and those obtained with quadprog and bTTN for the prostate cancer.

Figure 36 we see a comparison of the dose distribution on one slice before and after the optimization. From the top to the bottom the urinary bladder, the prostate and the rectum are shown.

The initial dose uniformity in the prostate is loss after the opimization, since the unitary bladder and the rectum have to be spared. As a consequence, we see that the dose inside both the rectum and the urinary bladder takes lower values after the optimization. Furthermore, we observe a reduction of the high dose contribution on the interface between the two OARs and the prostate. This is a practical proof of the benefit introduced by using IMRT.

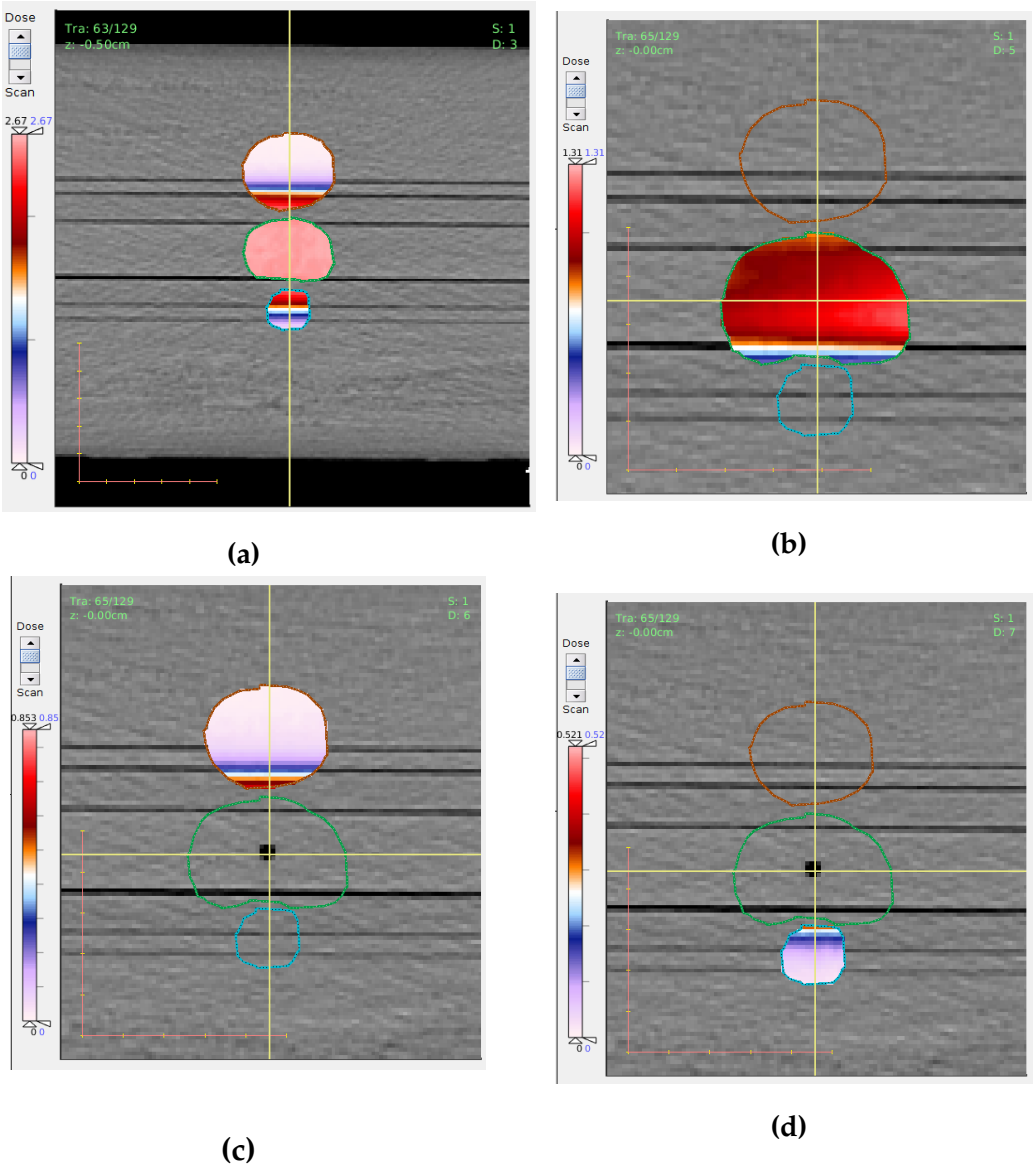



**Figure 36:** Dose distribution on one slice of prostate and the two OARS before and after the optimization.

# 6 CONCLUSION

This thesis aimed to investigate for the very first time the applicability of TNM to the dose optimization problem in IMRT. We focused on simplest case, the optimization of a quadratic function expressing the difference between the delivered and the prescribed dose.
A procedure to map the classical quadratic function into an Ising-like Hamiltonian was implemented, in order to adapt it to the architecture of the bTTN algorithm. As shown, it was done by discretizing the beamlets weights $x_j$ and exploiting the binary-decimal convertion.
In the first part of the study, we wrote an algorithm to perform iterative-search of the lowest energy eigenvalue along the diagonal of the Hamiltonian, with the aim of validating the mapping procedure from the classical to the quantum framework. In particular, it was applied to the toy-model of a bipartite box and the first successful results were obtained. However, the drawback of the exponential scaling of the computational time for inceasing system dimensions arose, showing the limited range of applicability of this method. It has been found that a simulation on a system of 64 qubits would take $\mathcal{O}(10^{13}\ h)$.
For this reason, a more sophisticated algorithm implementing a bTTN was used. The choice was validated by showing the great improvement in terms of computational time introduced by the latter. For the same system of 64 qubits it was shown to be $\mathcal{O}(10^{1}\ s)$. However, from the first results, the statistical nature of the ground state search with bTTN came out.
In addition to that, the role of the maximum bond dimension $\chi_{max}$ was investigated. From the study on the bipartite box for the system of 64 qubits, it came out that the $\langle E_{GS} \rangle$ has a decreasing trend up to a given value of $\chi_{max}$, reaching a plateau after that. The results were valitated using the MW test.

From a more detailed analysis on the final configurations, it resulted that in general the first configurations in energy differs only for the filp of a few qubits. This fact was considered as an hint of the fact that solutions slightly above the ground state could also deal to acceptable configurations from the cilinical point of view.

Finally, in the third part bTTN was applied to more realistic cases: the homogeneous sphere and the prostate cancer with the urinary bladder and the rectum as OARs. The results were validated by comparing them to those obtained with the Matab build-in function *quadprog*.

In both the two cases, the results have been found to be very close to the one computed by *quadprog*, even with low values of $\chi_{max}$. Furthermore, the role of the bond dimension was pointed out in this case too, showing how the bTTN's solution converges to the optimum solution as $\chi_{max}$ increases. This was a further hint that configurations which are close to the minimum could lead to acceptable solutions and in principle those solution can be found even with small $\chi_{max}$.

The only drawback of TNM with respect to the quadprog function is the computational time. Both for the sphere and for the prostate cancer, the computation with quadprog takes only a few seconds, while bTTN requires some hours.

However, the aim of the work was to probe the applicability of TNM to IMRT dose optimization, and the results gave a positive response.

There are some interesting aspects which could be further investigated. First of all, it would be very interesting to work with more realistic and sophisticated cost functions, in order insert all the constraints typically used in cilincs. The use of more complex cost functions would be also interesting in order to check if TNM can outperform standard algorithms in the search of the global minimum when many local minima are preent, a task that simple gradient descent-based methods usually fails to solve.

Another possible direction to explore could be to test different ways to codify the beamlets weights. One possibility might be to replace the $N_Q$ qubits repre-

senting each beamlets with a single *qudit* with $2^{N_Q}$ possibile spin orientations. This would reduce the dimension of the system, creating a 1-to-1 correspondence between the number of sites in the lattice and the number of beamlets in the beam.

To conclude, there's still a lot of work do in order to apply all this machinary to real IMRT treatments, but the results obtained are promising and gives a glimmer of hope that one day TNM might be used in the fight against cancer.

# REFERENCES


[1] INTERNATIONAL ATOMIC ENERGY AGENCY. *Radiation dose in radiotherapy from prescription to delivery*. IAEA, Vienna, 1996. URL: https://www-pub.iaea.org/MTCD/Publications/PDF/te_896_prn.pdf.

[2] A. Apte et al. *CERR*. 2010. URL: https://github.com/cerr/CERR.

[3] R. Baskar et al. "Biological response of cancer cells to radiation treatment". In: *Frontiers in molecular biosciences* (2014). DOI: 10.3389/fmolb.2014.00024.

[4] R. Baskar et al. "Cancer and radiation therapy: current advances and future directions". In: *International Journal of Medical Science* (Feb. 2012). DOI: 10.7150/ijms.3635.

[5] K. Blum. *Density Matrix Theory and Applications*. 3rd ed. Springer-Verlag Berlin Heidelberg, 2012. DOI: 10.1007/978-3-642-20561-3.

[6] Tjarda Boekholt and Simon Portegies Zwart. *On the Reliability of N-body Simulations*. 2014. URL: https://arxiv.org/abs/1411.6671.

[7] Sebastiaan Breedveld et al. "Fast, multiple optimization of quadratic dose objective function in IMRT". In: *Physics in Medicine and Biology* 51.14 (July 2006). DOI: 10.1088/0031-9155/51/14/019.

[8] M. Broderick, M. Leech, and M. Coffey. "Direct aperture optimization as a mean of reducing the complexity of intensity modulaated radiation therapy plans". In: *Radiation Oncology* (Feb. 2009). DOI: 10.1186/1748-717X-4-8.

[9] P.M. Chaikin and T.C. Chaikin. *Principles of Condensed Matter Physics*. Cambridge University Press, 22 1995. DOI: 10.1017/CBO9780511813467.

[10] A. Codignola. *La radioterapia non è una cura vecchia ma vive di nuova giovinezza*. Oct. 2017. URL: https://www.airc.it/news/la-radioterapia-non-e-una-cura-vecchia-ma-vive-una-nuova-giovinezza.

[11] David Craft et al. "An Approach for Practical Multiobjective IMRT Treatment Planning". In: *International Journal of Radiation Oncology* 69.5 (Dec. 2007), pp. 1600–1607. DOI: 10.1016/j.ijrobp.2007.08.019.

[12] David Craft et al. "Shared datasets for IMRT, beam angle optimization, and VMAT research". In: (). URL: http://gray.mgh.harvard.edu/attachments/article/226/c5.pdf.

[13] D P Dearnaley, V S Khoo, and Norman A R. "Comparison of radiation side-effects of conformal and conventional radiotherapy in prostate cancer: a randomised trial". In: *Lancet* (Jan. 1999). DOI: "10.1016/S0140-6736(98)05180-0".

[14] K. Dullemond and K. Peeters. *Introduction to Tensor Calculus*. 2010. URL: http://www.ita.uni-heidelberg.de/~dullemond/lectures/tensor/tensor.pdf.

[15] J. Eisert. *Entanglement and tensor networks states*. Sept. 2013. URL: https://arxiv.org/abs/1308.3318.

[16] M. Gerster et al. "Unconstrained tree tensor network: An adaptive gauge picture for enhanced performance". In: *Physical Review B* (Sept. 2014). DOI: 10.1103/physrevb.90.125154.

[17] Tarek Halabi, David Craft, and Thomas Bortfeld. "Application of constrained optimization to radiotherapy planning". In: *Physics in Medicine and Biology* 51.15 (July 2006). DOI: 10.1088/0031-9155/51/15/014.

[18] Y Hardiyanti et al. "The Comparison Study of Quadratic Infinite Beam Program on Optimization Instensity Modulated Radiation Therapy Treatment Planning (IMRTP) between Threshold and Exponential Scatter Method with CERR® In The Case of Lung Cancer". In: *Phys.: Conf. Ser.* **739** *012144* (2016). DOI: doi:10.1088/1742-6596/739/1/012144.

[19] A. Hospital et al. "Molecular dynamics simulations: advances and applications". In: *Advances and applications in bioinformatics and chemistry* (2015). DOI: 10.2147/AABC.S70333.

[20] R. M. Johnson. "On a theorem stated by eckart and young". In: *Psychometrika* (Sept. 1963). DOI: 10.1007/BF02289573.

[21] C.J. Karzmark, C.S. Nunan, and E. Tanabe. *Medical Electron Accelerators*. McGraw-Hill, 1993. URL: https://epdf.pub/medical-electron-accelerators.html.

[22] Andrei P. Kirilyuk. *Complex-Dynamic Cosmology and Emergent World Structure*. 2004. URL: https://arxiv.org/abs/physics/0408027.

[23] R F Laitano. *Fondamenti di dosimetria delle radiazioni ionizzanti*. 4th ed. Frascati: ENEA, Oct. 2015. URL: http://www.enea.it/it/seguici/pubblicazioni/pdf-volumi/FondamentidosimetriaradiazioniionizzantiIV.pdf.

[24] P. Mangin and R. Kahn. *Superconductivity*. Springer International Publishing, 2017. DOI: 10.1007/978-3-319-50527-5.

[25] Ivaho B Mihaylov and Eduardo G Moros. "Integral dose based inverse optimization objective function promises lower toxicity in head-and-neck". In: *Physica Medica: European Journal of Medical Physics* 54 (Oct. 2018), pp. 77–83. DOI: 10.1016/j.ejmp.2018.06.635.

[26] E. J. Mittemeijer. *Fundamentals of Materials Science*. Springer-Verlag Berlin Heidelberg, 2011. URL: https://www.springer.com/gp/book/9783642104992.

[27] S. Montangero. *Introduction to tensor network methods*. 1st ed. Springer, 2018. URL: https://doi.org/10.1007/978-3-030-01409-4.

[28] D. P. Nazareth and J. D. Spaans. "First application of quantum annealing to IMRT beamlet intensity optimization". In: *Physics in Medicine and Biology* 60.10 (). DOI: 10.1088/0031-9155/60/10/4137.

[29] M. Nesreen Ahamd et al. "Conventional (2D) Versus Conformal (3D) Techniques in Radiotherapy for Malignant Pediatric Tumors: Dosimetric Prospectives". In: *Journal of the Egyptian Nat. Cancer Inst.* (Dec. 2009). URL: https://www.ncbi.nlm.nih.gov/pubmed/21415867.

[30] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Oct. 2000. URL: http://mmrc.amss.cas.cn/tlb/201702/W020170224608149940643.pdf.

[31] U Oelfke, S Nill, and J J Wilkens. *Image-Guided IMRT*. Ed. by T Bortfled et al. Berlin, Heidelberg: Springer, 2006. Chap. 4, pp. 31–45. DOI: 10.1007/3-540-30356-1_4.

[32] Daniele Oriti. *Spacetime as a quantum many-body system*. 2017. URL: https://arxiv.org/abs/1710.02807.

[33] Román Orús. "A practical introduction to tensor networks: Matrix product states and projected entangled pair states". In: *Annals of Physics* (Oct. 2014). DOI: 10.1016/j.aop.2014.06.013.

[34] Stellan Östlund and Stefan Rommer. "Thermodynamic Limit of Density Matrix Renormalization". In: *Phys. Rev. Lett.* 75 (Nov. 1995), pp. 3537–3540. DOI: 10.1103/PhysRevLett.75.3537.

[35] University of Padua and INFN. *Cloudveneto*. URL: http://cloudveneto.it/.

[36] Laura Panizzi. "Sviluppo di un sistema per l"ottimizzazione della dose in radioterapia tramite 'dose-painting' basato sui voxels". Tesi di Laurea. Università degli Studi di Bologna, 2013.

[37] B. Pirvu et al. "Matrix product operator representations". In: *New Journal of Physics* 12.2 (Feb. 2010). DOI: 10.1088/1367-2630/12/2/025012.

[38] International Agency for Research on Cancer. *Latest global cancer data: Cancer burden rises to 18.1 milion new cases and 9.6 milion cancer deaths in 2018*. Sept. 2018. URL: https://www.who.int/cancer/PRGlobocanFinal.pdf.

[39] Otto A. Sauer, David M. Shepard, and T. Rock Mackie. "Application of constrained optimization to radiotherapy planning". In: *Medical Physics* 26.11 (1999), pp. 2359–2366. DOI: 10.1118/1.598750.

[40] Ulrich Schollwöck. "The density-matrix renormalization group in the age of matrix product states". In: *Annals of Physics* (Jan. 2011). DOI: 10.1016/j.aop.2010.09.012.

[41] Y.-Y. Shi, L.-M. Duan, and G. Vidal. "Classical simulation of quantum many-body systems with a tree tensor network". In: *Physical Review A* 74 (Aug. 2006). DOI: 10.1103/physreva.74.022320.

[42] P. Silvi et al. "The Tensor Networks Anthology: Simulation techniques for many-body quantum lattice systems". In: *SciPost Physics Lecture Notes* (Mar. 2019). DOI: 10.21468/scipostphyslectnotes.8.

[43] A. Szabo and N. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Ed. by McGraw-Hill. May 1989. URL: https://chemistlibrary.files.wordpress.com/2015/02/modern-quantum-chemistry.pdf.

[44] L. Tagliacozzo, G. Evenbly, and G. Vidal. "Simulation of two-dimensional quantum systems using a tree tensor network that exploits the entropic area law". In: *Physical Review B* 23 (Dec. 2009). DOI: 10.1103/physrevb.80.235127.

[45] A Taylor and M E B Powell. "Intensity-modulated radiotherapy–what is?" In: *Cancer Imaging: the official publication of the International Cancer Imaging* 4.2 (Mar. 2004), pp. 68–73. DOI: 10.1102/1470-7330.2004.0003.

[46] E. Villaggi et al. "Plan quality improvement by DVH sharing and planner's experience: Results of a SBRT multicentric planning study on prostate". In: *Physica Medica* (June 2019). URL: https://www.sciencedirect.com/science/article/abs/pii/S1120179719301103.

[47] J. S. Wang, H. J. Wang, and H. L. Qian. "Biological effects of radiation on cancer cells". In: *Military Medical Research* (2018). DOI: 10.1186/s40779-018-0167-4.

[48] Steven R. White. "Density matrix formulation for quantum renormalization groups". In: *Phys. Rev. Lett.* 69 (Nov. 1992), pp. 2863–2866. DOI: 10.1103/PhysRevLett.69.2863.

[49] Pangfei Zhang, Neng Fan, and Jie Shan. "Mixed integer programming with dose-volume constraints in intensity-modulated proton therapy". In: (Aug. 2017). DOI: "DOI:10.1002/acm2.12130".

[50] Yin Zhang and Michael Merritt. "Dose-volume-based IMRT fluence optimization: A fast least-squares approach with differentiability". In: *Linear Algebra and its Applications* 428.5 (2008), pp. 1365–1387. DOI: 10.1016/j.laa.2007.09.037.