Università degli Studi di Milano-Bicocca

Facoltà di Scienze MM.FF.NN.
Dottorato di Ricerca in Informatica
XXIV Ciclo

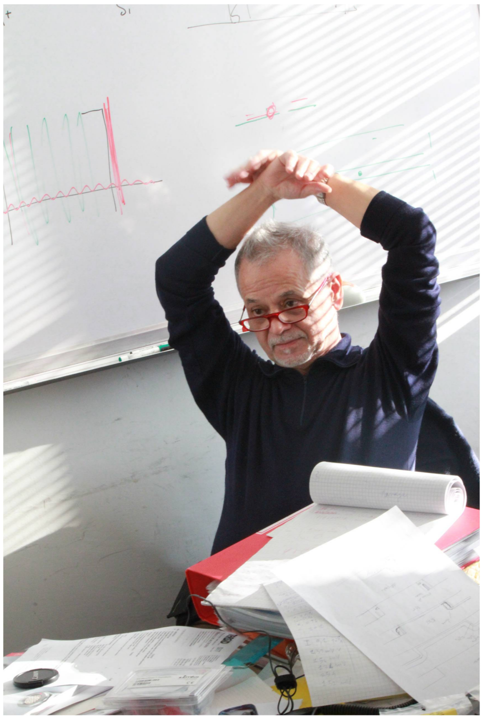
Swarm-Intelligence Strategy for Diagnosis of Endogenous Diseases by Nanobots

Paolo Amato
October 2012

Supervisor Prof. Giancarlo Mauri
Tutor Prof. Stefania Bandini

To Lalla and our marvelous kids Aurora and Gabriele

To the memory of Gianfranco (“G.F.”) Cerofolini



Gianfranco Cerofolini
June 14, 1946 – April 3, 2012

Contents

List of Figures

List of Tables

Chapter 1

Introduction

An ambitious long-term goal of medicine is to make analyses and deliver drugs selectively at cell level. In particular such “ideal” drugs should be able to travel through the vasculature, reach the intended target at full concentration, and there act selectively on diseased cells and tissues only, without creating undesired side effects.

The state of the art in pursuing this goal is represented by *nanoparticles* (Ferrari, 2010), which can be roughly defined as the combination of a drug molecule with a suitable vector of nanoscale dimensions. One of the main limitations of nanoparticles is that they do not have an onboard control system. Their “program” has to be defined in the drug design phase. A different program means the development of a new nanoparticle. Moreover the lack of onboard control means also that the complexity of the actions to be performed by nanoparticles is quite restricted.

The basic idea sustaining this goal is inspired by the immune system. That the immune system is a fantastic machine for surveilling and fighting against exogenous diseases is too well known to deserve discussion. However, it is not perfect.

A drawback of the immune system system is the poor recognition of endogenous pathological cells like those responsible for cancer or self-immune diseases. In this context it should be interesting to develop artificial devices (*nanobots*) working as blood white cells but addressed to the recognition and eventually the destruction of endogenous pathological states.

A nanobot (Requicha, 2003) can be defined as any artificial machine with overall size of the order of a few microns or less ($\simeq$ red cells), constituted by nanoscopic components with individual dimensions in the interval $1 - 10^2$ nm, and able to perform sophisticated functions like navigation, cell recognition, data collection and transmission. To the best of my knowledge no attempt toward the definition of an auxiliary immune system is however known. Of course, this goal is extremely ambitious and is expected to require decades. Nonetheless, sketching a scenario is not a mere speculation, but rather is useful to identify the nature of problems posed by the development of nanobots for real applications.

One of the possible reasons for the ineffectiveness of the immune system in detecting malignant cells is that the common features associated with cancer genesis and growth (hyperthermia, hypoxia and acidification) are the same as those characteristic of muscle under physical exercise. My program is thus that of *endowing the immune system with an artificial surveillance system devoted to detecting the simultaneous occurrence of hyperthermia, hypoxia and excessive acidity due to localized cancerous states without confusing it with the similar conditions produced under physiological conditions*. Considering the diagnosis of cancer as a crime detection, the search of the perpetrator is not addressed to the identification of the smoking gun but rather to a frame evidence resulting from the simultaneous occurrence of three events (hyperthermia, hypoxia and excessive acidity) and their persistence in time.

The aim of this thesis is twofold. On one side I want to describe an architecture of nanobots which is not only able to embed sophisticated functions, but also suitable for being manufactured by processes compatible with today and likely tomorrow semiconductor industries. On the other side, I want also to analyze and develop the algorithms needed for successfully tackling the nanobot tasks. In particular, since the nanobots would be devices with very limited computational resources and the interaction between them could only happen locally, the algorithms to control them could be based on swarm intelligence; i.e. it could be inspired by the collective behavior of social-insect colonies and other animal societies (Bonabeau et al., 1999). The thesis is articulated in the following 4 activities, whose relationships are shown in Figure 1.1:

- Identification of nanobot tasks for early diagnosis of diseases (Chapter 5)
- High-level description of nanobots architecture (Chapters 5 and 8)
- Development of a model of the environment in which nanobots operate: the circulatory system (Chapter 6)
- Development of swarm-intelligence algorithms for clustering nanobots around diseased cells (Chapter 7)

Figure 1.2 shows the structure of the thesis chapters.

Chapters 2–4 review some of the background relevant for this thesis. In particular Chapter 2 gives a very brief introduction to cancer, the disease which could be the major target of the nanobots. Then Chapter 3 describes the current state of nanoparticles. Finally Chapter 4 shows some results and challenges of of micro- and nano-robotics.

Chapter 5 describes the task and the architectures of the nanobots. The investigated auxiliary immune system is envisioned as constituted by two parts: a central unit, externally accessible but permanently resident in the organism (e.g., as an earring), and a swarm of nanobots. Nanobots can be considered as circulating nano-laboratories for blood analysis in situ. An ellipsoidal disk ($50 \times 2.5 \times 1\mu\mathrm{m}$) can move quite freely through the entire circulatory system. Hence the idea of surveying the state of the organism with swarms of such devices in the blood. By using next generation silicon planar technology, such a disk could host 10^5 gates to be used for implementing all the needed circuitry.

Once injected in the circulatory system, the first task of the nanobots is to cluster around the diseased cells in order to help the identification of their positions. To accomplish this task, the nanobots need to autonomously disperse in the capillary bed, take chemical sensor reading, mark the region where a positive signal is detected, and form a cluster in that region. The most complex subtask is the last one. Since in the microscopic world collisions are not an issue, communication between nanobots can happen through direct physical clashes. This approach, at best of my knowledge, has not been previously investigated. The chapter describes also how a nanobot with suitable appendages can anchor on the surface of the capillaries, and bond to other nanobots. The main results presented in this chapter have been published in (Cerofolini et al., 2010a) and (Cerofolini and Amato, 2012).

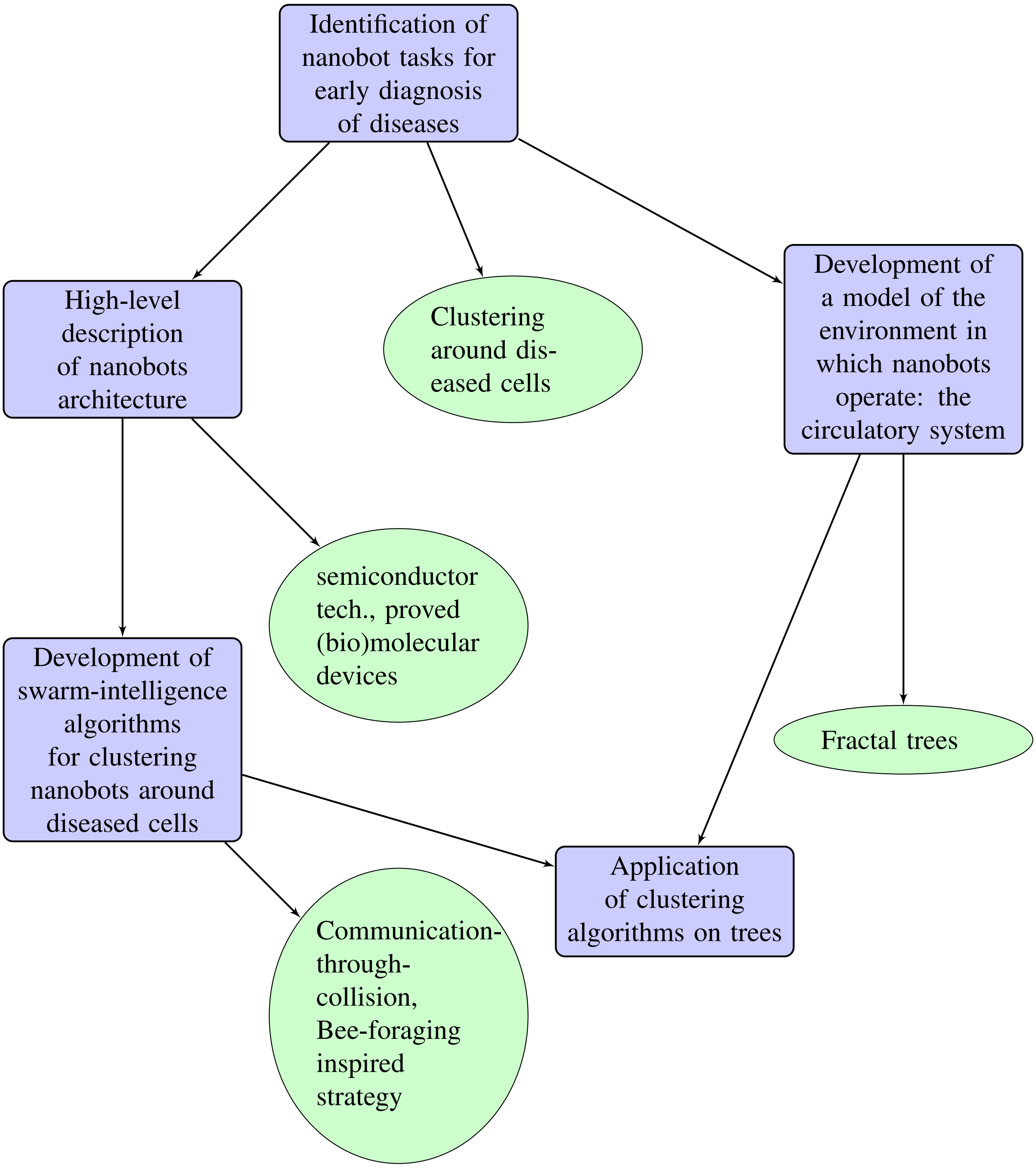


Figure 1.1: Thesis work flow (original contribution)

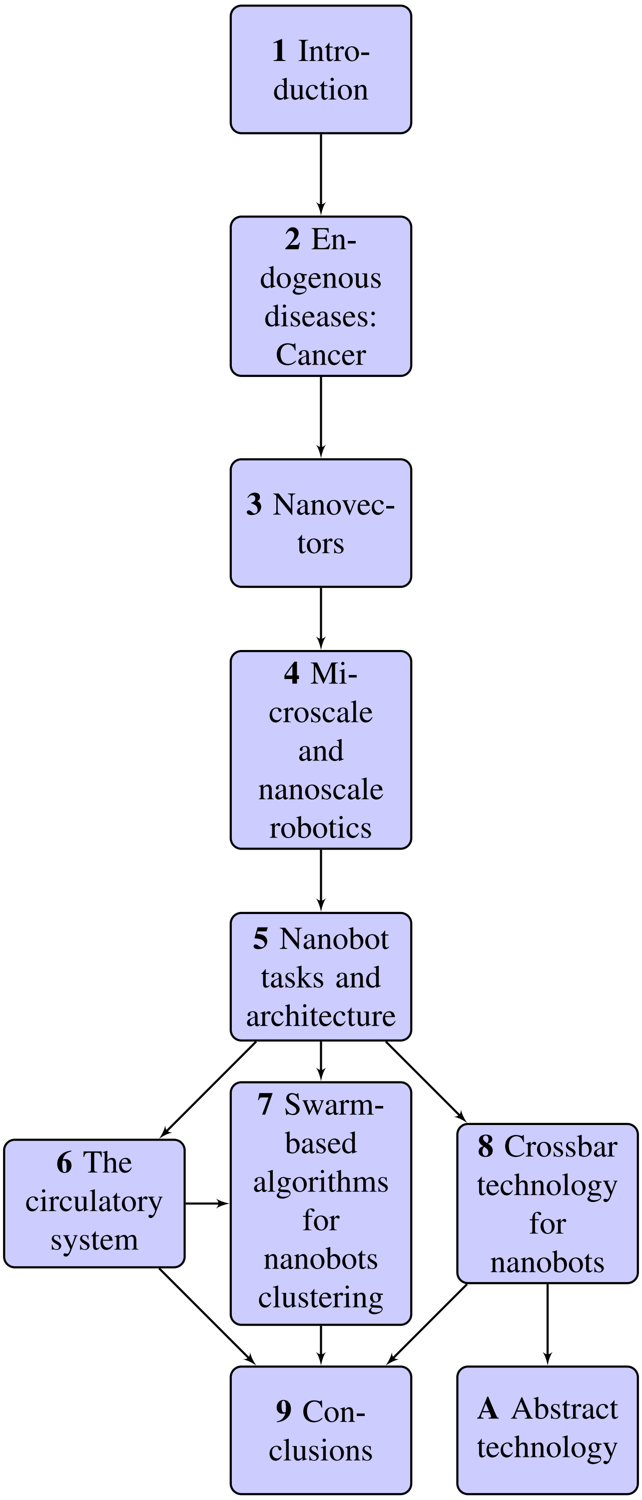


Figure 1.2: Chapters structure

The design of an efficient algorithm to control the swarm of nanobots injected in vivo requires an adequate model of the circulatory system. Chapter 6 describes the development of such a model. Since the work of Mandelbrot, the fractal geometry of biological structures has suggested that fractal methods could be used for modeling the human circulatory system. In fact, in the recent years the structure of blood vessel in any part of human body has been described in fractal terms; in particular by using *fractal trees*. The main result of this chapter is the fractal tree model of the systemic circulatory system described in section 6.5, that has been published in (Cerofolini and Amato, 2012). This model has been developed from scratch by introducing a system of equations able to link the physiological parameters of the circulatory system with the parameters of a symmetrical fractal trees. Although it is an extreme simplification to consider the circulatory system as a binary symmetric fractal tree, by optimizing the model parameters the error on physiological parameter is less than 3%.

Chapter 7 describes, simulates and compares three different strategies I developed for the clustering of nanobots around diseased cells: only target (OT), self-docking (SD) and division of labor (DL). These strategies can be located in the framework of swarm intelligence. In particular DL is inspired by how honeybee colonies organize the work during foraging. To obtain foraging success in a rapidly changing flower market, the colony must receive both updates on old food sources and reports on promising new ones. This is achieved by keeping two distinct types of employed foragers (Seeley, 1996), and a similar approach can be used for nanobots looking for diseased cells. Simulation of the 3 strategies has been performed by considering two different scenarios: a square region and a tree. The square region mimics a static scenario (e.g., blood analysis inside a test tube), while the tree scenario mimics the circulatory system (as described in Chapter 6). In both scenarios SD turns out to be the best strategy when dealing with just one target, while DL is the best strategy dealing with many targets. At the end of the chapter, the results obtained in the double-tree scenario are analyzed in the perspective of the application for *in vivo* monitoring. A preliminary version of the results of this chapter have been published in (Amato et al., 2011).

Chapter 8 describes the *crossbar architecture*, one of the key technological elements of the envisioned nanobots. Nanobots are far from being manufacturable objects. Considering as a framework the technology developed by today (and likely tomorrow) semiconductor industry, a key point to enable nanobots feasibility is the availability of low-cost high-density (in the order of terascale bit per cm^2) integrated circuits. In fact, to implement the on-board control system for all the nanobot functions hundreds thousands (or millions) logic gates will be needed on an area of around 100 μm^2. The most promising approach for reaching such densities is the crossbar architecture. In particular the 3-dimensional crossbar architecture seems to be a solution compatible with semiconductor industry and able to increase the gates density by an order of magnitude at least. With this approach the number of gates available to implement the nanobot control and storage system will raise up at least to 10^6. The main results presented in this chapter have been published in (Cerofolini et al., 2011a).

Chapter 9 briefly summarizes the achieved results.

At last Appendix A introduces *abstract technology (AT)*, a mathematical theory for the formal description of the complicated sequence of processes involved in the fabrication of integrated circuits (ICs). This theory has been used to develop the analysis made in Chapter 8. The main results presented in this chapter have been published in (Amato, 2011)

Part I

State-of-art

Chapter 2

Endogenous diseases: Cancer

The main application area discussed in this thesis is the early detection of endogenous diseases, and in particular of cancer. This chapter introduces basic facts about cancer. The aim is to give the basic medical terminology and some notions used in the later chapters. A quick and recent introduction to the topic can be found in (James, 2011).

2.1 What is cancer

Cancer is a term used for diseases in which abnormal cells divide without control and are able to invade other tissues. Moreover cancer cells can spread to other parts of the body through the blood and lymph systems (National Cancer Institute, 2010). Both the diseases themselves and the treatments are major causes of pain and distress for worldwide population. In fact in 2008, around 12.7 millions of people were diagnosed with cancer, of whom 7.9 million died; this accounting for roughly 13% of all deaths. See Fig. 2.1.

Cancer is not just one disease but many diseases. There are more than 100 different types of cancer, and most of them are named from the organ or type of cell in which they start. Just to give 2 examples example, cancers that begin in the colon are called colon cancer, while cancers that begin in basal cells of the skin are called basal cell carcinomas.

Cancer types can be grouped into categories:

- *Carcinomas*, the most common types of cancer, arise from the cells that cover external and internal body surfaces. Lung, breast, and colon are the most frequent cancers of this type in the United States.

- *Sarcomas* are cancers arising from cells found in the supporting tissues of the body such as bone, cartilage, fat, connective tissue, and muscle.

- *Lymphomas* are cancers that arise in the lymph nodes and tissues of the body's immune system.

- *Leukemias* are cancers of the immature blood cells that grow in the bone marrow and tend to accumulate in large numbers in the bloodstream.

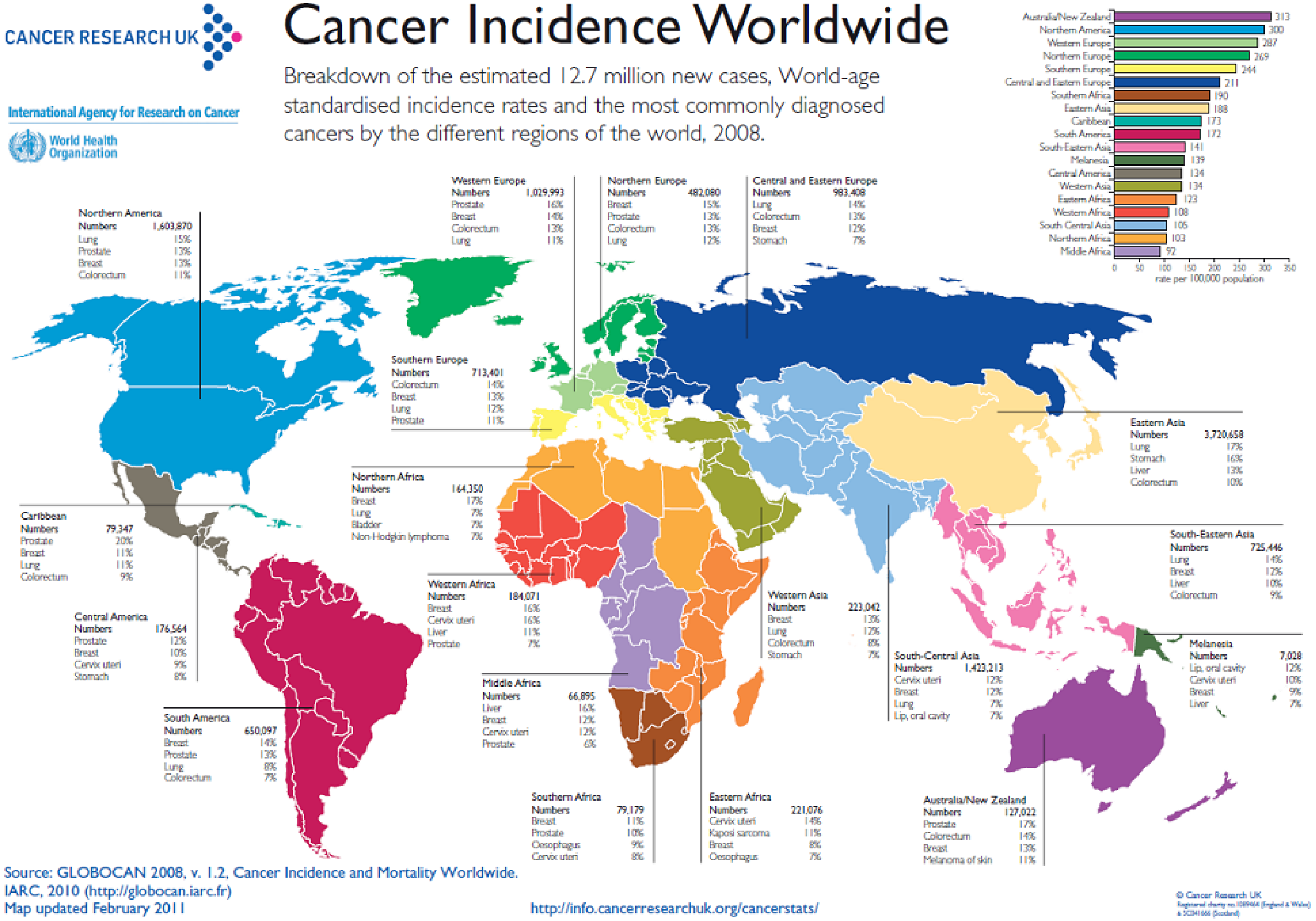


Figure 2.1: Worldwide cancer incidence (Cancer Research UK, 2011)

2.2 Origins of cancer

All cancers begin in cells, the body's basic unit of life. To understand cancer, it's helpful to know what happens when normal cells become cancer cells.

The cells are the basic building blocks that make up all living things. The human body is made up of many types of cells. All cells in the organism have their own carefully controlled life cycle. They grow and divide in a controlled way to produce more cells as they are needed to keep the body healthy. When cells become old or damaged, they die and are replaced with new cells.

However, sometimes this orderly process goes wrong. The genetic material (DNA) of a cell can become damaged or changed, producing mutations that affect normal cell growth and division. When this happens, cells do not die when they should and new cells form when the body does not need them. The extra cells may form a mass of tissue called a *tumor* (or *neoplasm*). See Figure 2.2.

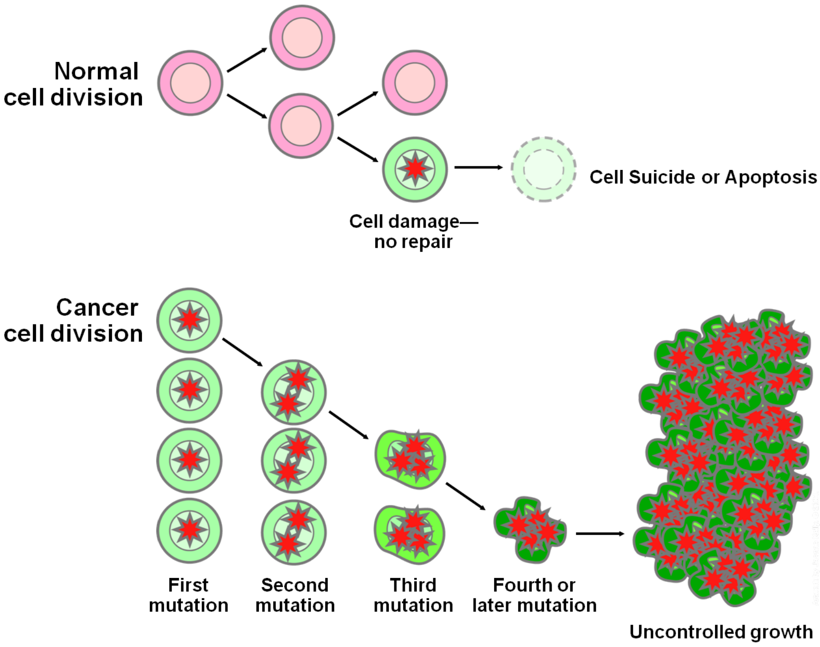


Figure 2.2: Cancer arises from a loss of normal growth control. In normal tissues, the rates of new cell growth and old cell death are kept in balance. In cancer, this balance is disrupted. This disruption can result from uncontrolled cell growth or loss of a cell's ability to undergo cell suicide by a process called *apoptosis*. Apoptosis, or "cell suicide", is the mechanism by which old or damaged cells normally self-destruct. Taken from (National Cancer Institute, 2010)

Anything which replicates (living cells) will probabilistically suffer from errors (mutations). Unless error correction and prevention is properly carried out, the errors will survive, and might be passed along to daughter cells. Normally, the body safeguards against cancer via numerous methods, such as: apoptosis, helper molecules (some DNA polymerases), possibly senescence, etc. However sometimes these error-correction methods fail to correct all the errors, especially in environments that make errors more likely to arise and propagate. For example, such environments can include the presence of disruptive substances called carcinogens, or periodic injury (physical, heat, etc.), or environments that cells did not evolve to withstand, such as hypoxia[1].

All agents that damage DNA therefore are potential carcinogens — agents that cause cancer. The reverse is not true: not all agents that can cause cancer directly damage DNA (although this always lies at the end of the process). Examples of cancer-causing substances that do not directly damage DNA include alcohol and the sex hormones involved in causing breast and prostate cancer. There are many sorts of carcinogens; cigarette smoke and ionizing radiation are well know examples. Considering cigarette smoking, it is known that usually it is necessary to smoke many cigarettes for many years for cancer to develop. This suggests that cancer is a progressive disease, and these progressive errors slowly accumulate until a cell begins to act contrary to its function in the organism.

To illustrate what is meant by normal growth control, consider the skin. The thin outermost layer of normal skin, called the epidermis, is roughly a dozen cells thick. Cells in the bottom row of this layer, called the basal layer, divide just fast enough to replenish cells that are continually being shed from the surface of the skin. Each time one of these basal cells divides, it produces two cells. One remains in the basal layer and retains the capacity to divide. The other migrates out of the basal layer and loses the capacity to divide. The number of dividing cells in the basal layer, therefore, stays the same. See Figure 2.3.

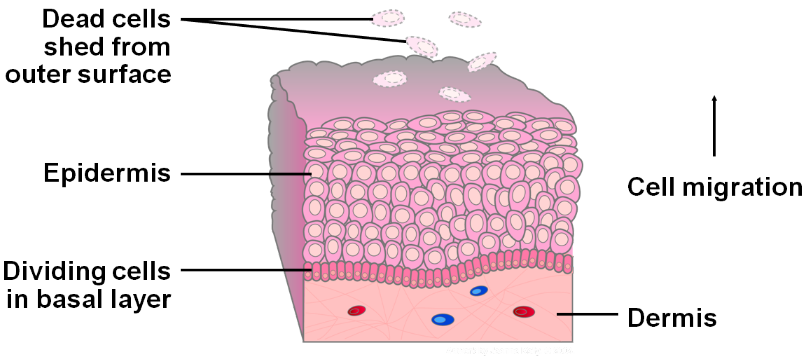


Figure 2.3: Example of normal growth. Taken from (National Cancer Institute, 2010)

During the development of skin cancer, the normal balance between cell division and cell loss is disrupted. The basal cells now divide faster than is needed to replenish the cells being shed from the surface of the skin. Each time one of these basal cells divides, the two newly formed cells will often retain the capacity to divide, thereby leading to an increase in the total number of dividing cells. See Figure 2.4.

This gradual increase in the number of dividing cells creates a growing mass of tissue (the tumor). If the rate of cell division is relatively rapid, and no "suicide" signals are in place to trigger cell death, the tumor will grow quickly in size; if the cells divide more slowly, tumor growth will be slower. But regardless of the growth rate, tumors ultimately increase in size because new cells are being produced in greater numbers than needed. As more and more of these dividing cells accumulate, the normal organization of the tissue gradually becomes disrupted. See Figure 2.5.

---

[1] Hypoxia, or hypoxiation, is a pathological condition in which the body as a whole (generalized hypoxia) or a region of the body (tissue hypoxia) is deprived of adequate oxygen supply

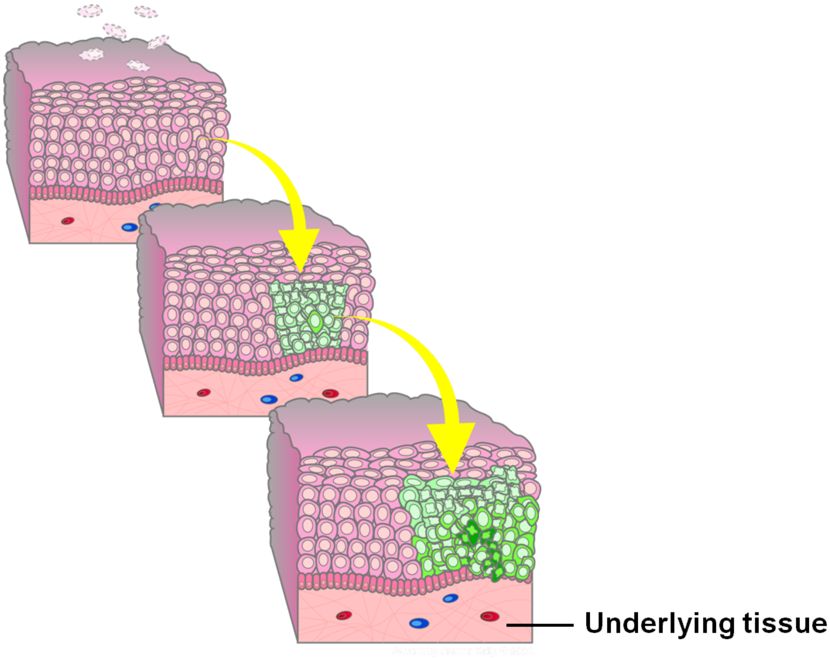


Figure 2.4: The beginning of cancerous growth. Taken from (National Cancer Institute, 2010)

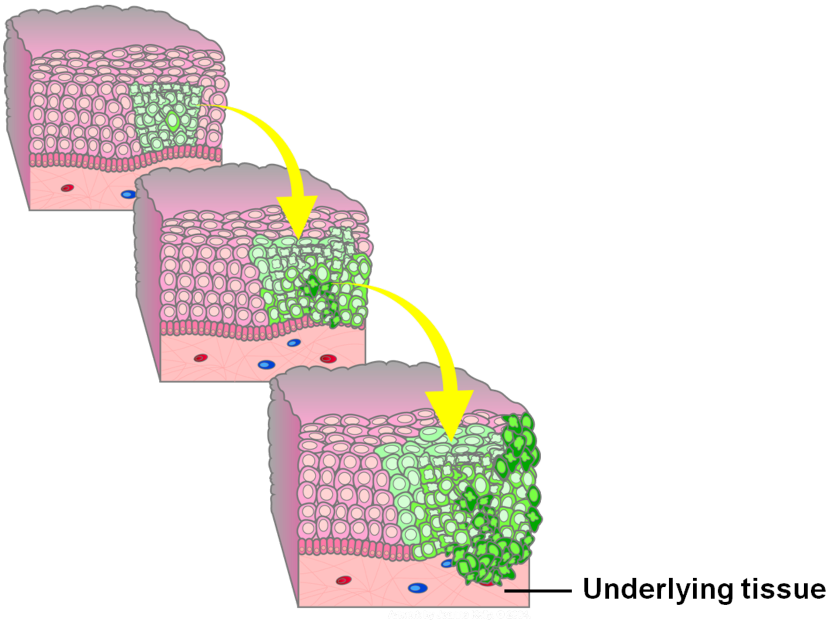


Figure 2.5: Tumors (Neoplasms). Taken from (National Cancer Institute, 2010)

Cancers are capable of spreading throughout the body by two mechanisms: *invasion* and *metastasis*. Invasion refers to the direct migration and penetration by cancer cells into neighboring tissues. Metastasis refers to the ability of cancer cells to penetrate into lymphatic and blood vessels, circulate through the bloodstream, and then invade normal tissues elsewhere in the body. See Figure 2.6.

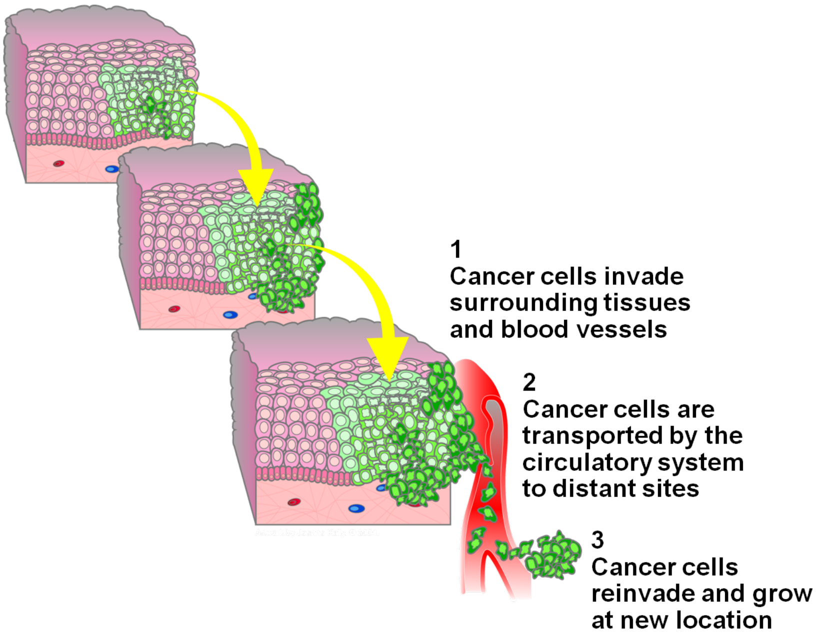


Figure 2.6: Invasion and metastasis. Taken from (National Cancer Institute, 2010)

Not all tumors are cancerous; tumors can be benign or malignant (Figure 2.7).

- Benign tumors aren't cancerous. They can often be removed, and, in most cases, they do not come back. Cells in benign tumors do not spread to other parts of the body
- Malignant tumors are cancerous. Cells in these tumors can invade nearby tissues and spread to other parts of the body. The spread of cancer from one part of the body to another is called *metastasis*.

Some cancers do not form tumors. For example, leukemia is a cancer of the bone marrow and blood.

2.3 The hallmarks of cancer

A cancer cell differs from normal cells in that it divides in an unregulated fashion. In addition, cancer cells have the ability to spread and to invade other parts of the body.

In (Hanahan and Weinberg, 2000) and (Hanahan and Weinberg, 2011) it is suggested that most if not all cancers have acquired the same set of functional capabilities during their development, albeit through various mechanistic strategies. It seems that the vast catalog of cancer cell genotypes is a manifestation of six essential alterations in cell physiology that collectively

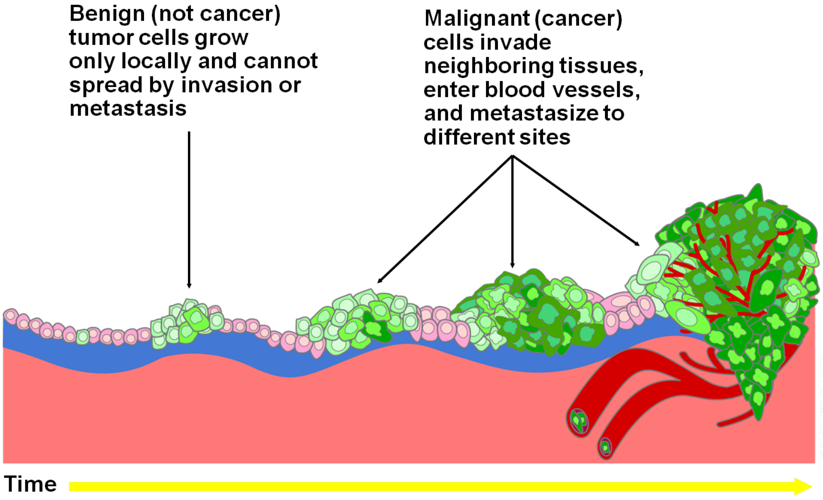


Figure 2.7: Benign tumors are tumors that cannot spread by invasion or metastasis; hence, they only grow locally. Malignant tumors are tumors that are capable of spreading by invasion and metastasis. By definition, the term ”cancer” applies only to malignant tumors. Taken from (National Cancer Institute, 2010)

dictate malignant growth. These acquired functional capabilities that allow cancer cells to survive, proliferate, and disseminate are called *hallmarks of cancer* (Hanahan and Weinberg, 2000):

1. Sustaining proliferative signals
2. Evading growth suppressors
3. Resisting cell death
4. Enabling replicative immortality
5. Inducing angiogenesis
6. Activating invasion and metastasis

These six hallmarks are shown in Fig. 2.8 and are briefly described in the rest of the section.

The first three hallmarks — Sustaining proliferative signals, evading growth suppressors, and resisting cell death — all lead to an uncoupling of a cell growth program from signals in its environment. However, research performed over the past 30 years indicates that this acquired disruption of cell-to-cell signaling, on its own, does not ensure expansive tumor growth. This is the reason why the fourth hallmark is required.

The fifth hallmark of cancer is the ability to grow a new blood supply. The last hallmark is related to a key aspect of cancer growth: the acquisition of the ability to grow in the wrong place, and this is a feature that distinguishes a malignant tumor from a benign one, which can grow but not spread or invade.

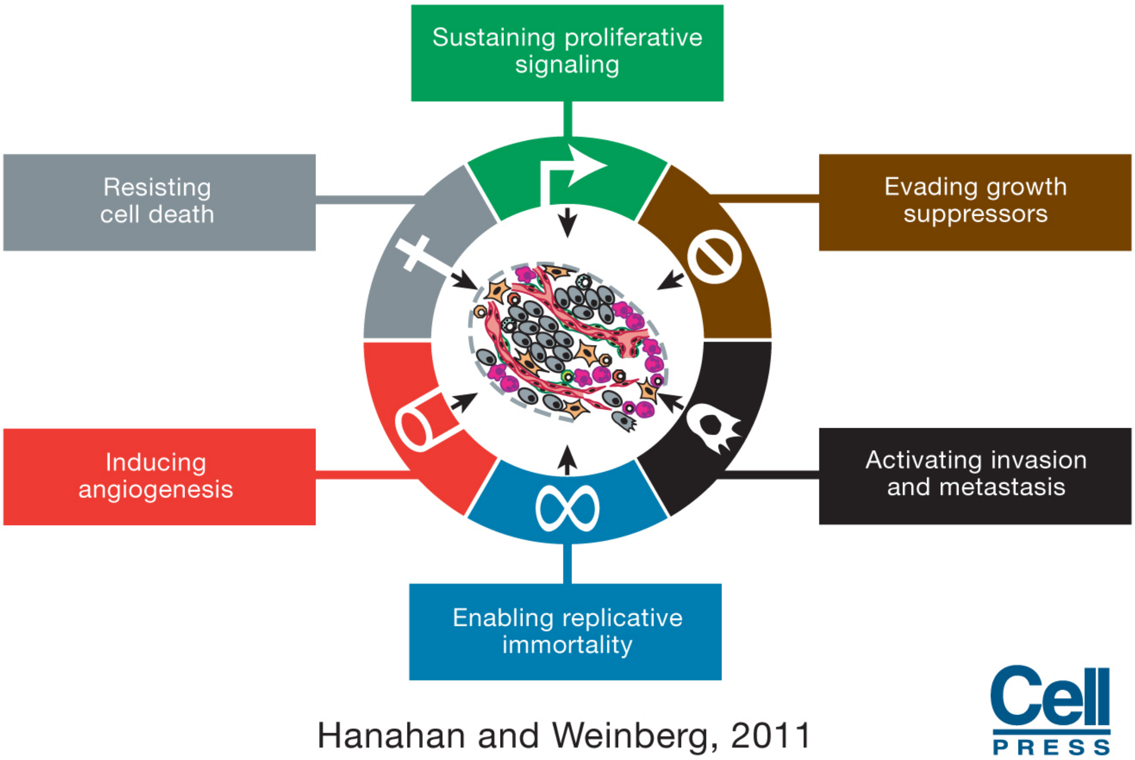


Figure 2.8: The Hallmarks of Cancer. This illustration encompasses the six hallmark capabilities originally proposed in (Hanahan and Weinberg, 2000). The past decade has witnessed remarkable progress toward understanding the mechanistic underpinnings of each hallmark. Taken from (Hanahan and Weinberg, 2011)

Sustaining Proliferative Signaling Arguably the most fundamental trait of cancer cells involves their ability to sustain chronic proliferation. Normal tissues carefully control the production and release of growth-promoting signals that instruct entry into and progression through the cell growth-and-division cycle, thereby ensuring a homeostasis of cell number and thus maintenance of normal tissue architecture and function. Cancer cells, by deregulating these signals, become masters of their own destinies.

Evading Growth Suppressors In addition to the hallmark capability of inducing and sustaining positively acting growth-stimulatory signals, cancer cells must also circumvent powerful programs that negatively regulate cell proliferation; many of these programs depend on the actions of tumor suppressor genes. Dozens of tumor suppressors that operate in various ways to limit cell growth and proliferation have been discovered through their characteristic inactivation in one or another form of animal or human cancer; many of these genes have been validated as bona fide tumor suppressors through gain- or loss-of-function experiments in mice.

Resisting cell death The concept that programmed cell death by apoptosis serves as a natural barrier to cancer development has been established by compelling functional studies conducted over the last two decades (Adams and Cory, 2007). Elucidation of the signaling circuitry governing the apoptotic program has revealed how apoptosis is triggered in response to various physiologic stresses that cancer cells experience during the course of tumorigenesis or as a result of anticancer therapy.

The structure of the apoptotic machinery and program, and the strategies used by cancer cells to evade its actions, were widely appreciated by the beginning of the last decade. The most notable conceptual advances since then have involved other forms of cell death that broaden the scope of "programmed cell death" as a barrier to cancer.

- *Autophagy* represents an important cell-physiologic response that, like apoptosis, normally operates at low, basal levels in cells but can be strongly induced in certain states of cellular stress, the most obvious of which is nutrient deficiency

- In contrast to apoptosis, in which a dying cell contracts into an almost-invisible corpse that is soon consumed by neighbors, *necrotic cells* become bloated and explode, releasing their contents into the local tissue microenvironment. Necrotic cell death releases proinflammatory signals into the surrounding tissue microenvironment, in contrast to apoptosis and autophagy, which do not. As a consequence, necrotic cells can recruit inflammatory cells of the immune system

Enabling Replicative Immortality By 2000, it was widely accepted that cancer cells require unlimited replicative potential in order to generate macroscopic tumors.This capability stands in marked contrast to the behavior of the cells in most normal cell lineages in the body, which are able to pass through only a limited number of successive cell growth-and-division cycles. This limitation has been associated with two distinct barriers to proliferation: *senescence*, a typically irreversible entrance into a nonproliferative but viable state, and *crisis*, which involves cell death. These two barriers to proliferation ave been rationalized as crucial anticancer defenses that are hard-wired into our cells, being deployed to impede the outgrowth of clones of preneoplastic and frankly neoplastic cells.

Inducing Angiogenesis Like normal tissues, tumors require sustenance in the form of nutrients and oxygen as well as an ability to evacuate metabolic wastes and carbon dioxide. In the adult, as part of physiologic processes such as wound healing and female reproductive cycling, angiogenesis is turned on, but only transiently. In contrast, during tumor progression, an *angiogenic switch* is almost always activated and remains on, causing normally quiescent vasculature to continually sprout new vessels that help sustain expanding neoplastic growth.

The blood vessels produced within tumors by chronically activated angiogenesis and an unbalanced mix of proangiogenic signals are typically aberrant: tumor neovasculature is marked by precocious capillary sprouting, convoluted and excessive vessel branching, distorted and enlarged vessels, erratic blood flow, microhemorrhaging, leakiness, and abnormal levels of endothelial cell proliferation and apoptosis.

Activating Invasion and Metastasis In 2000, the mechanisms underlying invasion and metastasis were largely an enigma. It was clear that as carcinomas arising from epithelial tissues progressed to higher pathological grades of malignancy, reflected in local invasion and distant metastasis, the associated cancer cells typically developed alterations in their shape as well as in their attachment to other cells and to the extracellular matrix (ECM).

The multistep process of invasion and metastasis has been described as a sequence of discrete steps, often termed the invasion-metastasis cascade (Talmadge and Fidler, 2010). See Fig. 2.9.

2.3.1 Growth of cancer

A single cancerous cell surrounded by healthy tissue will replicate at a rate higher than the other cells, placing a strain on the nutrient supply and elimination of metabolic waste products. Once a small tumor mass has formed, the healthy tissue will not be able to compete with the cancer cells for the inadequate supply of nutrients from the blood stream. Tumor cells will displace healthy cells until the tumor reaches a diffusion-limited maximal size. While tumor cells will typically not initiate apoptosis in a low nutrient environment, they do require the normal building blocks of cell function like oxygen, glucose and amino acids. The vasculature was designed to supply the now extinct healthy tissue that did not place as high a demand for nutrients due to its slower growth rate.

Tumor cells will therefore continue dividing because they do so without regard to nutrient supply but also many tumor cells will perish because the amount of nutrients is insufficient. The tumor cells at the outer edge of a mass have the best access to nutrients while cells on the inside die creating a necrotic core within tumors that rely on diffusion to deliver nutrients and eliminate waste products. In essence, a steady state tumor size forms, as the rate of proliferation is equal to the rate of cell death until a better connection with the circulatory system is created. This diffusion-limited maximal size of most tumors is around 2 mm^3. To grow beyond this size, the tumor must recruit the formation of blood vessels to provide the nutrients necessary to fuel its continued expansion. An illustration of tumor development from a single cell to a diffusion-limited tumor is shown in Fig. 2.10. It is thought that there could be numerous tumors at this diffusion-limited maximal size throughout the body. Until the tumor can gain that access to the circulation it will remain at this size and the process can take years. The exact molecular mechanisms that initiate angiogenesis at a tumor site are not known and could be unique to site of origin but more information about what factors play a role in this process is being

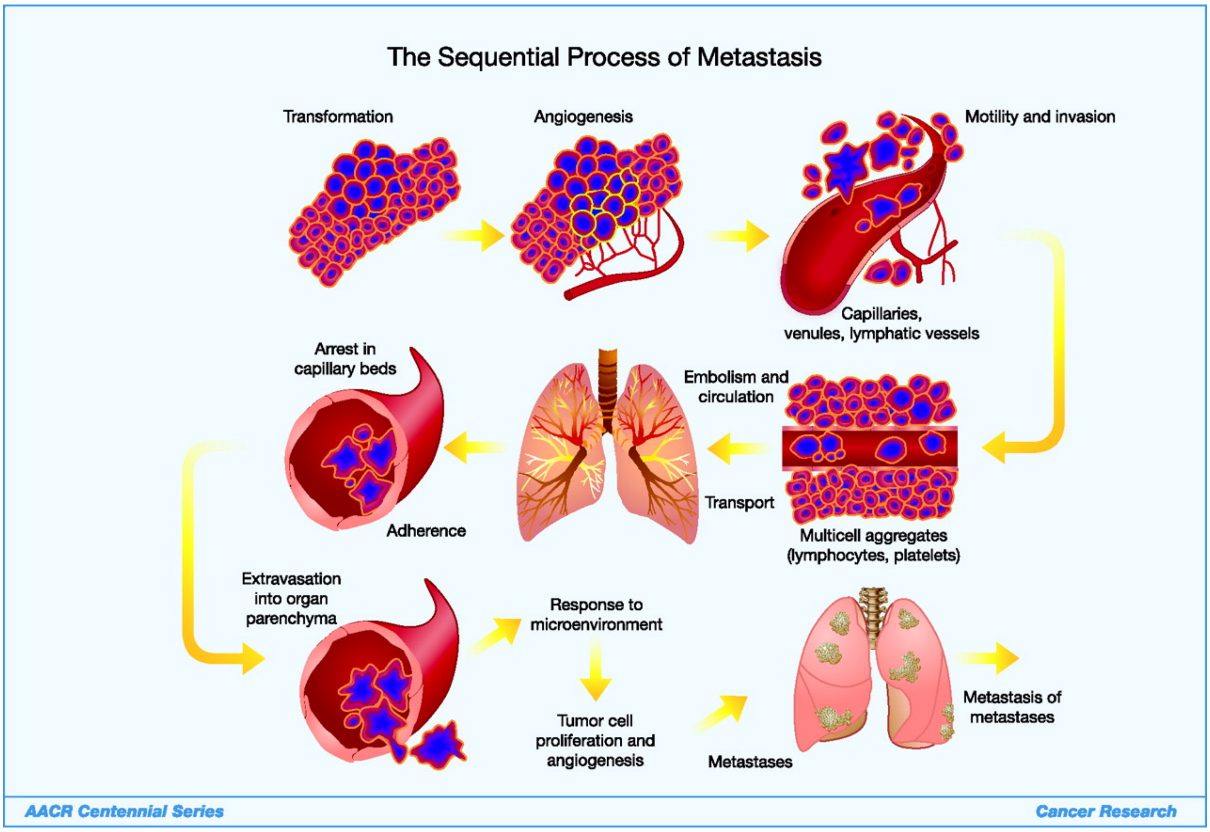


Figure 2.9: The process of cancer metastasis consists of sequential, interlinked, and selective steps with some stochastic elements. The outcome of each step is influenced by the interaction of metastatic cellular subpopulations with homeostatic factors. Each step of the metastatic cascade is potentially rate limiting such that failure of a tumor cell to complete any step effectively impedes that portion of the process. Therefore, the formation of clinically relevant metastases represents the survival and growth of selected subpopulations of cells that preexist in primary tumors. Taken from (Talmadge and Fidler, 2010)

discovered. As more is known about the molecular mechanisms that stimulate angiogenesis, the factors involved present new therapeutic targets to prevent tumor development.

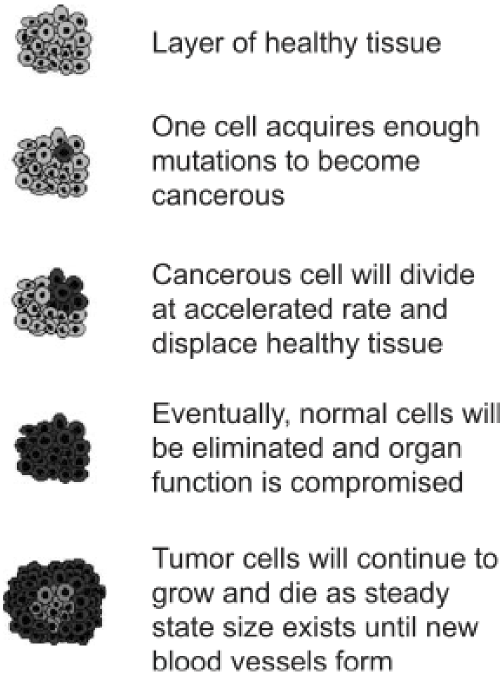


Figure 2.10: Tumor development from initial carcinogenesis to diffusion-limited maximal size. Taken from (Brannon-Peppas and Blanchette, 2004)

2.3.2 Angiogenesis

Angiogenesis is a process vital to the continued development of a tumor mass. This process has been the subject of intense research due to its role in cancer development and has proven to be the result of numerous interactions between regulators, mediators and stimulatory molecules.

Tumor cells are capable of secreting molecules that initiate the angiogenic process. The new vessels will allow the tumor to grow beyond the diffusion-limited maximal size. Some tumor masses never grow beyond this point, as they are incapable of recruiting new vessels. Acquisition of the angiostimulatory phenotype, also called the angiogenic switch, is thought to result from a local imbalance between positive and negative regulators of angiogenesis.

The inability of the body to halt tumor-induced angiogenesis can have a number of explanations. Tumor cells in some cases no longer express angiogenesis inhibitors that would stop the process. Tumor cells and the surrounding stromal cells can be induced to express angiogenesis promoters at accelerated rates. As the blood vessels begin to form, immune cells that can secrete angiogenesis stimulators gain access to the tumor cells to continue to promote neovascularization. As the tumor cells are in closer proximity to blood vessels, tumor cells may disseminate from the tumor into the circulation. Upon finding a suitable environment, like a distant capillary bed or nearby lymph node, these cells can become metastatic foci of the primary tumor. The end result of a tumor that makes the angiogenic switch is a tumor capable of increased growth, fueled by both paracrine and autocrine factors, with access to the blood stream to create additional tumors in other organs. The continued development of a tumor beyond the diffusion-limited maximal size is shown in Fig. 2.11.

Because a vascularized tumor is capable of increased growth and is more readily able to metastasize, increasing amounts of research are focused on developing treatments to slow angiogenesis and limit tumor growth and dissemination. Because so many different molecules are involved in angiogenesis there are many potential targets for therapy. Some examples of therapeutic strategies include limiting endothelial proliferation and motility, increased expression of angiogenesis inhibitors and use of molecules such as soluble VEGF receptor to try and decrease the amount of angiogenesis stimulatory factors at the tumor site. A comprehensive list of anti-angiogenic therapies in clinical trials can be found in the website of US National Cancer Institute [http://cancernet.nci.nih.gov].

2.3.3 Impact on the system

In addition to all the molecular changes that occur within a cancer cell, the environment around the tumor changes dramatically as well. The cancer cell loses receptors that would normally respond to neighboring cells that call for growth to stop. Instead, tumors amplify their own supply of growth signals. They also flood their neighbors with other signals called cytokines and enzymes called proteases. This action destroys both the basement membrane and surrounding matrix, which lies between the tumor and its path to metastasis–a blood vessel or duct of the lymphatic system. See Figure 2.12.

2.3.4 Enabling Characteristics and Emerging Hallmarks

In (Hanahan and Weinberg, 2000) the hallmarks of cancer have been defined as acquired functional capabilities that allow cancer cells to survive, proliferate, and disseminate; these functions are acquired in different tumor types via distinct mechanisms and at various times during the course of multistep tumorigenesis. Their acquisition is made possible by two enabling characteristics. Most prominent is the development of genomic instability in cancer cells, which generates random mutations including chromosomal rearrangements; among these are the rare genetic changes that can orchestrate hallmark capabilities. A second enabling characteristic involves the inflammatory state of premalignant and frankly malignant lesions that is driven by cells of the immune system, some of which serve to promote tumor progression through various means.

Yet other distinct attributes of cancer cells have been proposed to be functionally important for the development of cancer and might therefore be added to the list of core hallmarks. Two such attributes are particularly compelling:

- Genome Instability and Mutation
- Tumor-Promoting Inflammation

The first involves major reprogramming of cellular energy metabolism in order to support continuous cell growth and proliferation, replacing the metabolic program that operates in normal tissues and fuels the physiological operations of the associated cells. The second involves active evasion by cancer cells from attack and elimination by immune cells; this capability highlights

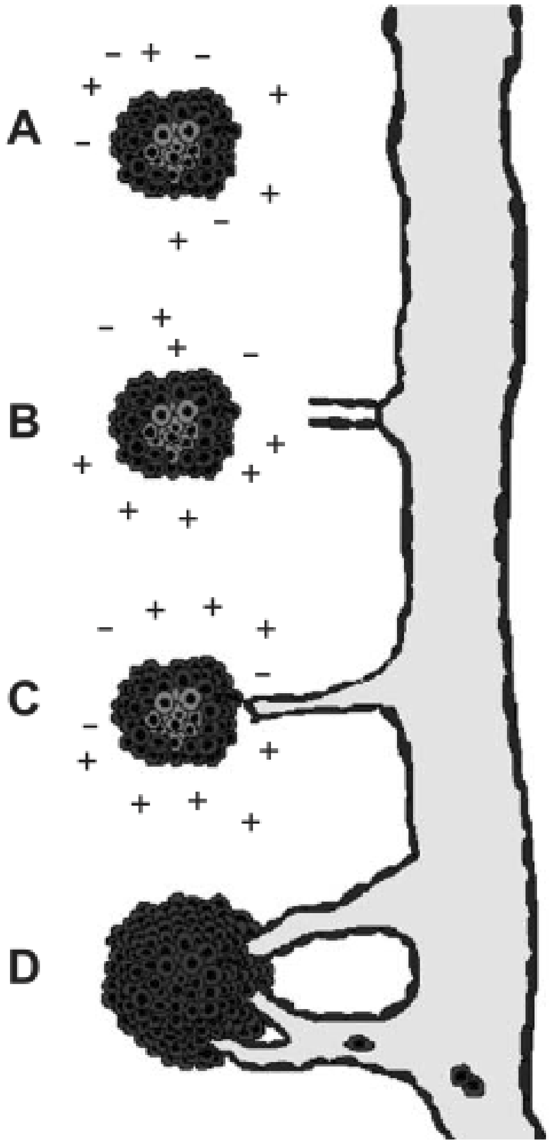


Figure 2.11: Continued tumor development beyond diffusion-limited maximal size (A). In (B), the angiogenic switch has occured creating an imbalance of positive to negative regulators causing endothelial cell proliferation and migration. These endothelial cells form a vessel which extends towards the tumor and provides nutrients to sustain cell proliferation (C). A fully vascularized tumor (D) is capable of continued growth with metastatic potential due to the proximity to the blood stream. Taken from (Brannon-Peppas and Blanchette, 2004)

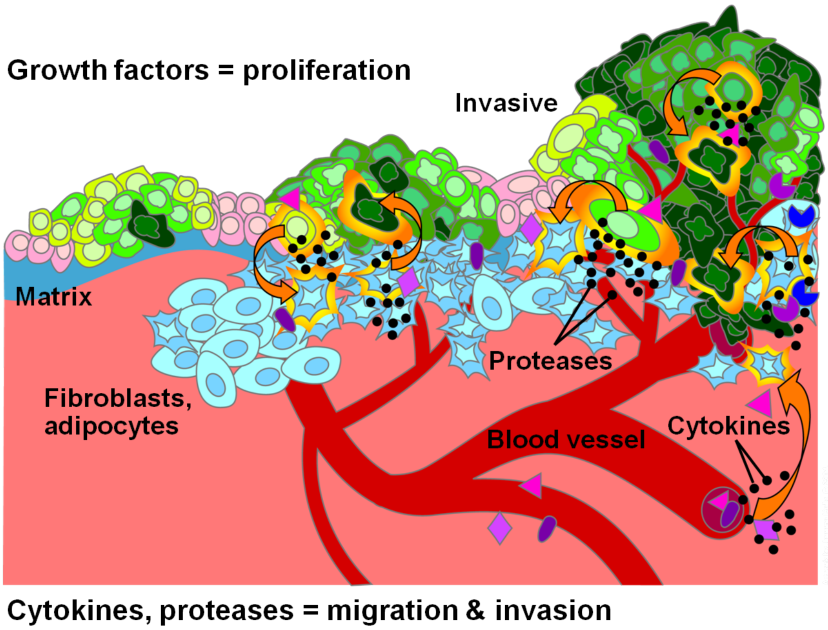


Figure 2.12: Cancer tends to corrupt surrounding environment. Taken from (National Cancer Institute, 2010)

the dichotomous roles of an immune system that both antagonizes and enhances tumor development and progression.

These enabling characteristics and emerging hallmarks are depicted in Fig.2.13

Reprogramming Energy Metabolism The chronic and often uncontrolled cell proliferation that represents the essence of neoplastic disease involves not only deregulated control of cell proliferation but also corresponding adjustments of energy metabolism in order to fuel cell growth and division.

Evading Immune Destruction A second, still-unresolved issue surrounding tumor formation involves the role that the immune system plays in resisting or eradicating formation and progression of incipient neoplasias, late-stage tumors, and micrometastases. The long-standing theory of immune surveillance proposes that cells and tissues are constantly monitored by an ever-alert immune system, and that such immune surveillance is responsible for recognizing and eliminating the vast majority of incipient cancer cells and thus nascent tumors. According to this logic, solid tumors that do appear have somehow managed to avoid detection by the various arms of the immune system or have been able to limit the extent of immunological killing, thereby evading eradication.

2.4 Current practice for cancer therapy

Cancer treatment is complex and typically involves inputs from a number of different groups: doctors, surgeons, oncologists, pathologists, radiologists,... In any case, current cancer treatments can be divided in a few main classes (James, 2011):

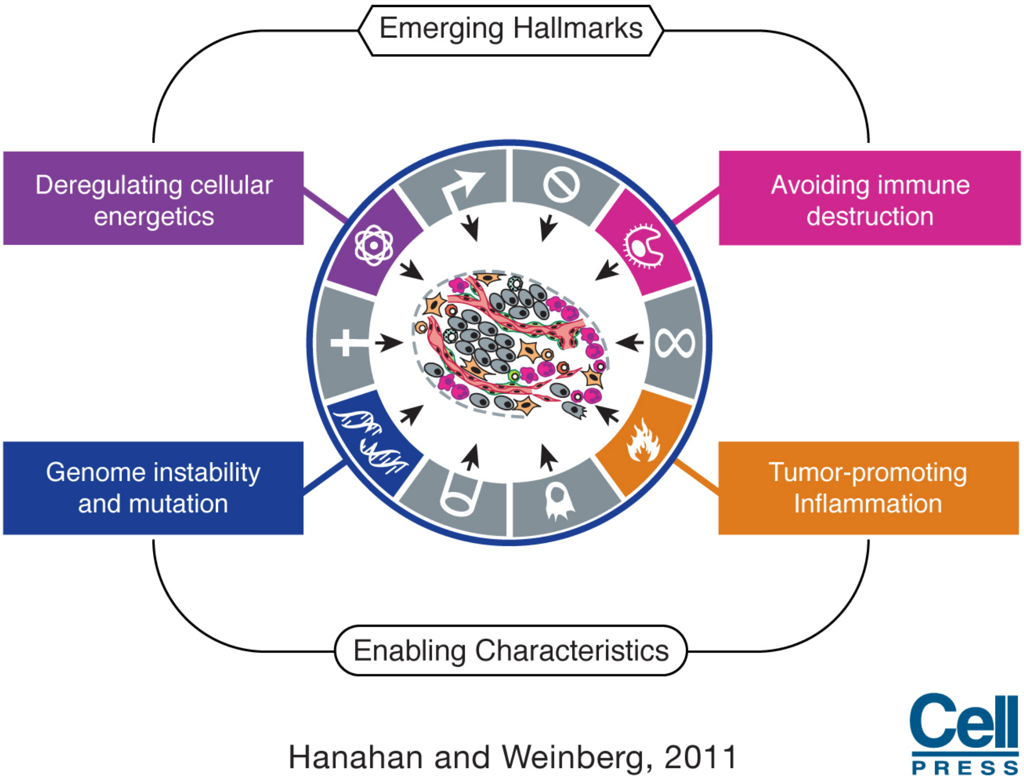


Figure 2.13: Enabling characteristics and emerging hallmarks. Taken from (Hanahan and Weinberg, 2011)

- Surgery
- Radiotherapy
- Hormone therapy
- Chemotherapy
- Monoclonal antibodies
- Targeted molecular therapies

The choice of therapy depends upon the location and grade of the tumor and the stage of the disease, as well as the general state of the patient. In all cases, the effectiveness of the treatment is directly related to the treatment's ability to target and to kill the cancer cells while affecting as few healthy cells as possible. The degree of change in the patient's quality of life and eventual life expectancy is directly related to this targeting ability of the treatment. However, some side effects will always occur and sometimes are so intense that the patient must discontinue therapy before the drugs have a chance to eradicate the cancer. Unfortunately, not all treatments, even if carried through to the oncologists specifications, are effective in killing the cancer before the cancer kills the patient. The advances in treatment of cancer are progressing quickly both in terms of new agents against cancer and new ways of delivering both old and new agents. Hopefully this progress can move us away from near-toxic doses of non-specific agents (Brannon-Peppas and Blanchette, 2004).

The introduction of mechanism-based targeted therapies to treat human cancers has been heralded as one of the fruits of three decades of remarkable progress of research into the mechanisms of cancer pathogenesis. In (Hanahan and Weinberg, 2011) it is argued that the description of hallmark principles is beginning to inform therapeutic development at present and may increasingly do so in the future. In fact, the rapidly growing armamentarium of targeted therapeutics can be categorized according to their respective effects on one or more hallmark capabilities, as illustrated in the examples presented in Fig. 2.14.

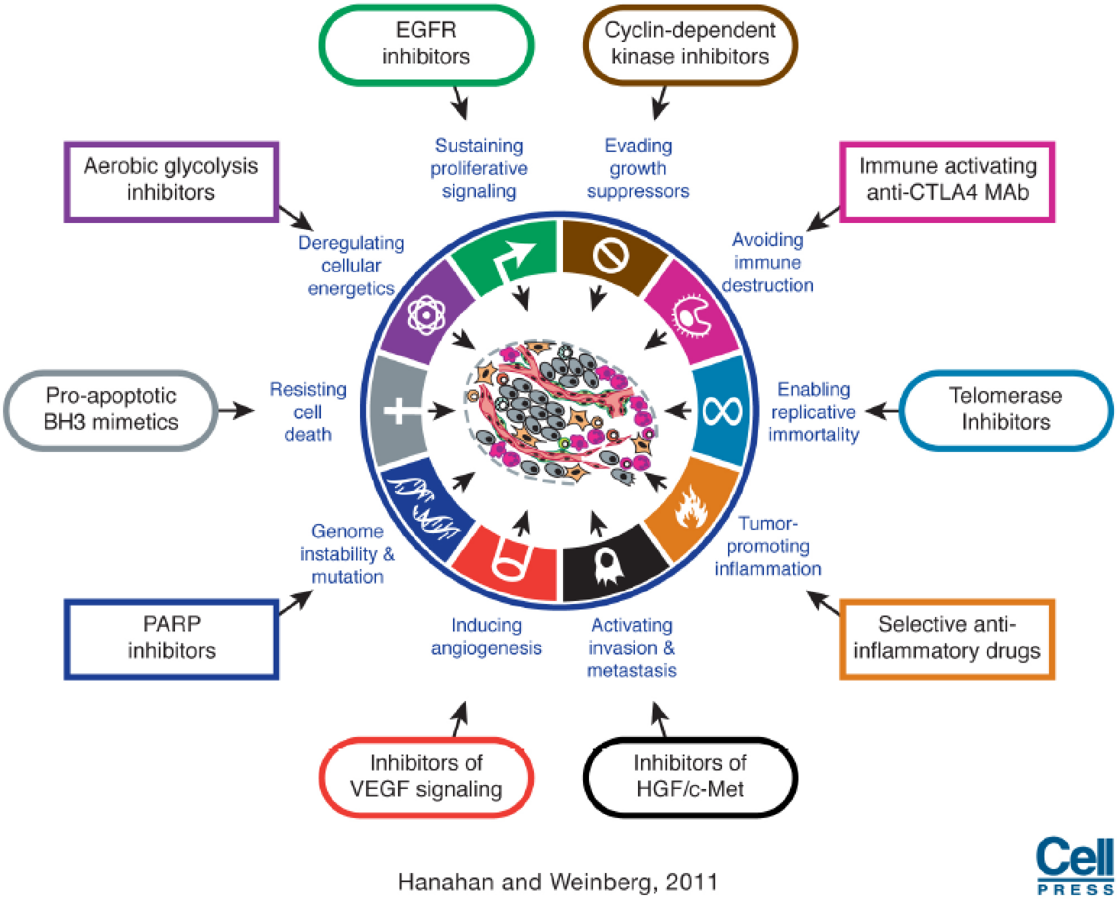


Figure 2.14: Emerging Hallmarks and Enabling Characteristics. An increasing body of research suggests that two additional hallmarks of cancer are involved in the pathogenesis of some and perhaps all cancers. One involves the capability to modify, or reprogram, cellular metabolism in order to most effectively support neoplastic proliferation. The second allows cancer cells to evade immunological destruction, in particular by T and B lymphocytes, macrophages, and natural killer cells. Because neither capability is yet generalized and fully validated, they are labeled as emerging hallmarks. Additionally, two consequential characteristics of neoplasia facilitate acquisition of both core and emerging hallmarks. Genomic instability and thus mutability endow cancer cells with genetic alterations that drive tumor progression. Inflammation by innate immune cells designed to fight infections and heal wounds can instead result in their inadvertent support of multiple hallmark capabilities, thereby manifesting the now widely appreciated tumor-promoting consequences of inflammatory responses. Taken from (Hanahan and Weinberg, 2011)

Chapter 3

Nanovectors

A *Nanovector* is a hollow or solid structure, with diameter in the $1 - 10^3$ nm range, which can be filled with anticancer drugs and detection agents. Targeting moieties can also be attached to the surface (Ferrari, 2005a). A *Nanoparticle* is a solid nanovector, typically made of a single material (e.g. gold).

Intravascularly injectable nanovectors are a major class of nanotechnological devices of interest for use in cancer. For them the following uses are envisioned (Ferrari, 2005a):

- in vivo, non-invasive visualization of molecular markers of early stages of disease
- the targeted delivery of therapeutic agents, with a concurrent, substantial reduction of deleterious side effects
- and, by a combination of the first 2, the interception and containment of lesions before they reach the lethal or even the malignant phenotype, with minimal or no concurrent loss of quality of life.

Nanovectors delivering drugs and contrast agents have the potential to provide breakthrough solutions for the cure of diseases, as cancer and cardiovascular, and for in-vivo high-resolution imaging. Differently from freely administrated molecules, a nanovector can carry simultaneously several and different therapeutic and imaging agents supporting both complex treatments through the co-localized release of different drugs and multi-modal imaging. It has been proved that targeted nanoparticles administered at the systemic level can recognize and adhere to diseased cells (biological target) expressing specific biological molecules (receptors), thus increasing the specificity and efficiency of the therapeutic action and the imaging resolution, whilst reducing possible side effects (Decuzzi and Ferrari, 2008).

The rationale of using nanoparticles for tumor targeting is based on the following facts:

1. nanoparticles will be able to deliver a concentrate dose of drug in the vicinity of the tumor targets via the enhanced permeability and retention effect or active targeting by ligands on the surface of nanoparticles;
2. nanoparticles will reduce the drug exposure of health tissues by limiting drug distribution to target organ.

3.1 Archetipal nanovectors: Liposomes

Liposomes are the archetypal, simplest form of a nanovector. A liposome can be thought of as a hollow sphere whose size ranges from 30 nm to 10 μm. Liposomes are made from phospholipid molecules which are the same molecules that comprise cell membranes. Phospholipids are *amphipathic*, that is, part of their structure is water-soluble (hydrophilic or water-loving) and the other part is oil-like (hydrophobic or water-fearing or fat-soluble). Therefore, when added to water, the water-soluble part of the phospholipid interacts with the water and the oil-like part of the molecule avoids the water[1].

In order to accomplish this, the phospholipids align themselves side-by-side with their oil-like portions orienting themselves towards each other as shown in cross-section in the middle of Fig. 3.1. This structure is known as a phospholipid bilayer. This bilayer extends itself in water to form a sheet which then curls into a liposome. Liposomes smaller than about 200 nm usually only consist of one bilayer (*unilamellar liposomes*) but larger liposomes can contain concentric layers of lipid (like an onion) or several smaller liposomes can be formed inside large liposomes. These larger multicompartmented liposomes are known as *multilamellar liposomes*.

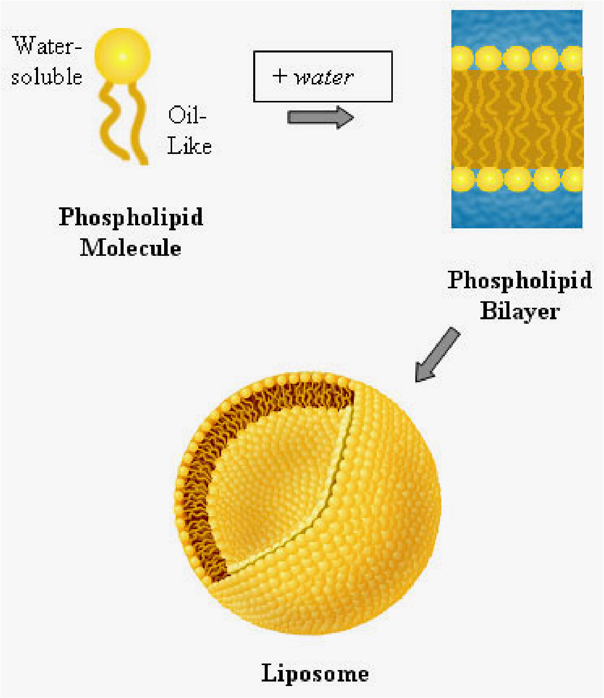


Figure 3.1: Liposome structure. Taken from (Encapsula Nano Sciences, 2011)

The interior of liposomes is filled with water and therefore, molecules which are soluble in water can be encapsulated in the interior of the liposome. A less recognized but very important property of liposomes is that molecules which are not water soluble, or oil-like, can be entrapped in the oil-like portion of the phospholipid bilayer. Consequently, liposomes can serve as carriers for all types of molecules including both water-soluble and water-insoluble

---

[1] Common use amphipathic substances are soaps and detergents.

compounds. In fact, a single liposome can carry both types of molecules or combinations of each type of molecule.

Liposomes use the overexpression of fenestrations in cancer neovasculature to increase drug concentration at tumor sites; a phenomenon termed Enhanced Penetration and Retention (*EPR*), see Fig. 3.2. They concentrate preferentially in tumors because the new network of blood vessels supplying the tumors presents gaps (fenestrations) through which the liposomes can permeate, if they are of the right size (Ferrari, 2008). Thus, size has been an agent of therapeutic specificity from the very beginning of cancer nanotechnology, but in a “passive sense”, i.e., without the use of biomolecular recognition molecules that can provide greater degrees of specificity. Further attempts at adding specificity by decorating the surface of nanoparticles with biological targeting moieties such as antibodies have had very modest success.

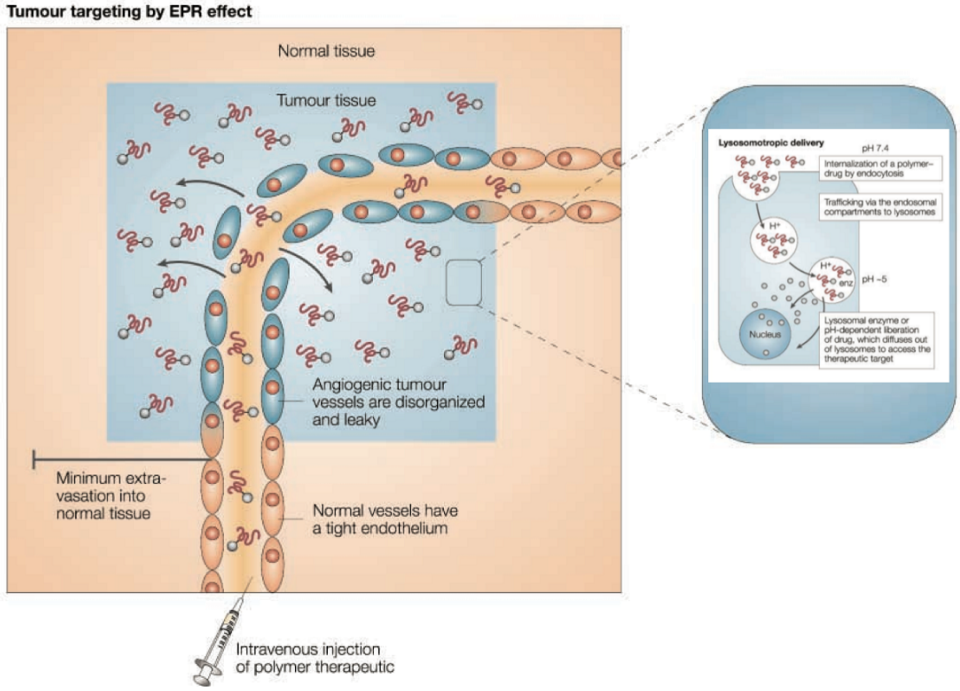


Figure 3.2: Tumor targeting through EPR effect. Hyperpermeable angiogenic tumor vasculature allows preferential extravasation of nanovectors. The therpeutics act once present in the tumor interstitium, Taken from (Duncan, 2003)

Liposome-encapsulated formulations of doxorubicin were approved in the 90s for the treatment of Kaposis sarcoma[2], and are now used against breast cancer and refractory ovarian cancer. Liposomes continue to be refined and applied to more cancer indications. They are only the first in an ever-growing number of nanovectors under development for novel, more efficacious drug-delivery modalities. Liposomes were the first drug carrying nanoparticles to reach cancer clinics.

Several types of nanoparticle for the enhancement of magnetic resonance imaging (MRI) contrast have been used clinically and in research protocols. These include gadolinium-based,

---

[2] Kaposis sarcoma is a cancer that causes patches of abnormal tissue to grow under the skin, in the lining of the mouth, nose and throat or in other organs.

ironoxide-based nanoparticles and multiple-mode imaging contrast nano-agents that combine magnetic resonance with biological targeting and optical detection. Low-density lipid nanoparticles have been used to enhance ultrasound imaging. For each current clinical modality it is actually possible to develop nanoparticles that can provide signal enhancement, combined with biomolecular targeting capabilities.

3.2 Nanovectors

Nanovectors in general have at least a tripartite constitution which enhances the biodistribution and tumor targeting of the nanoparticle dispersion (see Figure3.3). The three parts are:

- a core constituent material
- a therapeutic and/or imaging payload
- and biological surface modifiers,

Nanovector formulations are designed to reduce the clearance time of small peptide drugs, provide protection of active agents from enzymatic or environmental degradation, and avoid obstacles to the targeting of the active moiety. Examples of such obstacles include the protective exclusion by the bloodbrain barrier or the vascular endothelium; the augmented osmotic pressure states in cancer lesions, resulting in outward convection of the therapeutic moiety; and nanoparticle sequestration by the *Reticulo-endothelial system* (RES)[3].

Silicon and silica are emerging as interesting candidate materials for injectable nanovectors (Ferrari, 2005a). Porosified silicon is biodegradable, with kinetics that are much more rapid (minutes to hours) than those of biodegradable polymers (weeks to months), and therefore release drugs with previously unattainable time profiles.

Metal-based nanovectors include *nanoshells*, i.e. nanoparticles composed of a gold shell surrounding a semiconductor. When nanoshells reach their target they can be irradiated to make the nanoshell hot; the heat kills the cancer cell.

The thickness of the gold layer can be precisely tuned, so that the nanoshell can be selectively activated through tissue irradiation with near-infrared light to perform localized therapeutic thermal ablation. The approach was recently used to eradicate transmissible venereal tumors in mice. Beyond its specific merits, this approach introduces the concept that nanovectors can be used as highly selective, externally activated therapeutic agents.

It is estimated that several thousand different nanovector types have been reported in the literature. Just a minute fraction of their potential uses against cancer have been explored, yet these offer technological foundations for meeting the fundamental cancer nanotechnology challenges.

3.2.1 Breakthrough opportunities

A major clinical advantage sought by the use of nanovectors over simple immunotargeted drugs is the specific delivery of large amounts of therapeutic or imaging agents per targeting biorecognition event.

---

[3] A system composed of monocytes and macrophages that is located in reticular connective tissue (for example, in the spleen). These cells are responsible for phagocytosing and removing cellular debris, pathogens and foreign substances from the bloodstream.

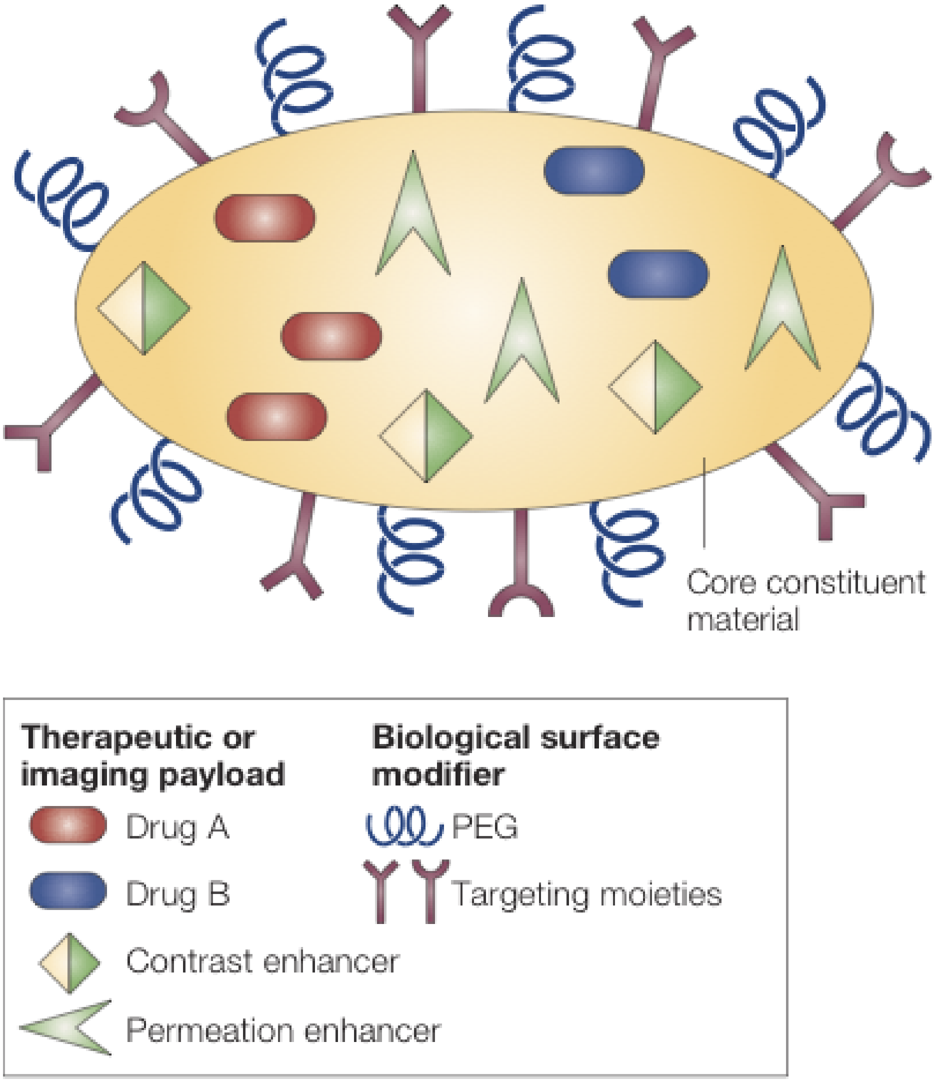


Figure 3.3: Multifunctional nanoparticle. The following are illustrated: the ability to carry one or more therapeutic agents; biomolecular targeting through one or more conjugated antibodies or other recognition agents; imaging signal amplification, by way of co-encapsulated contrast agents; and biobarrier avoidance, exemplified by an endothelial tight-junction opening permeation enhancer, and by polyethylene glycol (PEG) for the avoidance of macrophage uptake by macrophages. Taken from (Ferrari, 2005a)

This advantage can be summarized in three, closely interrelated main aspects (Ferrari, 2005b):

1. the recognition of target cells and tissues
2. the ability to reach the diseased sites where the target cells and tissues are located
3. third, the ability to deliver multiple therapeutic agents

The first two aspects comprise the notion of achieving preferred, substantially higher concentration of therapeutic action at lesion sites. This phenomenon is called *localization* in (Ferrari, 2005b), as opposed to the term *targeting* that in literature is usually employed to identify drugs that provide specific action against a target biological pathway.

Recognition of target cells and tissues There is a general taxonomy that can be applied to nanovectors, which divides them into the three main subclasses or generations (Sakamoto et al., 2010) as schematically shown in Fig. 3.4.

The first generation of nanovectors are nanoparticles that penetrate the disease site by using passive mechanisms. The main subclass in this category comprises the liposomes (previuosly discussed).

Significant strides in the fields of chemistry and materials science have yielded several other nanosized carriers with immense potential for drug delivery, including polymer-drug conjugates, polymer micelles, and dendrimers. While first generation of nanovector comprises carriers with no active mechanisms of disease site location and therapy, the second generation of nanovectors encompasses delivery systems with further functionalities. These functionalities can have two origins:

1. specific molecular recognition moieties on the nanovector to receptors overexpressed on the tumor cells or adjacent blood vessels (e.g. mAb-conjugated liposomes) or
2. a possibility for active/triggered release of the payload at the diseased location (e.g. magnetic liposomes).

Although second-generation nanovectors are superior to their precursor (by employing additional complexities such as targeting moieties, remote activation, and environmentally sensitive components), represent simply a progressive evolution of the first-generation nanovectors.

The fundamental problem of various obstacles on the way of therapeutics to reach their target, has given rise to a paradigm shift in the design of nanoparticles with the emergence of third-generation particles — *logic embedded vectors* (LEVs) (Sakamoto et al., 2010). LEVs are therapeutic multi-component constructs specifically engineered to avoid biological barriers, in which the functions of bio-recognition, cytotoxicity and biobarrier avoidance are decoupled, yet act in efficacious operational harmony. The ideal injected chemotherapeutic strategy is envisioned to be capable of navigating through the vasculature after intravenous administration, to reach the desired tumor site at full concentration, and to selectively kill cancer cells with a cocktail of agents with minimal harmful side effects.

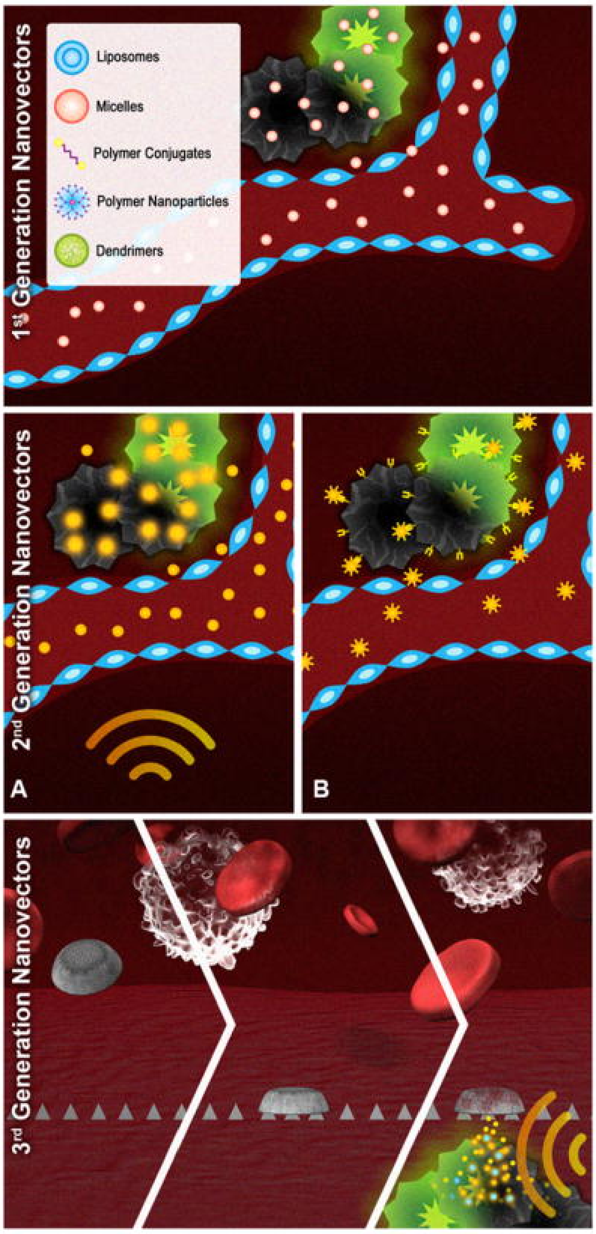


Figure 3.4: Schematic presentation of three generations of therapeutic nanovectors. First generation: nanoparticles localizing in tumor through the EPR passive mechanism; second generation: nanovectors possessing additional level of complexity such as (a) remote activation by means of radiofrequency (RF) or near-infrared (NIR) energy or (b) active targeting through specific ligands overexpressed on tumor cells; third generation: logic embedded vectors, LEV comprised of different nano-components which act through a time-sequence of synergistic and logic-driven events. Taken from (Sakamoto et al., 2010)

Reaching the target lesions Intrinsic to the body defense systems are several extremely effective obstacles (collectively termed *biobarriers*) that largely prevent injected chemicals, biomolecules, nanoparticles and any other foreign agent of therapeutic action from reaching their intended destinations. Any efficient delivery method must be provided with tools that allow it to overcome all of these barriers. Some of the leading developments (Ferrari, 2005b) arte tackling the following types of biobarriers:

- Endothelial and epithelial barriers
- Sequestration by the reticulo-endothelial system
- Adverse osmotic pressure

Delivering multiple agents of therapy Contemporary cancer therapeutics generally involve the simultaneous use of multiple drugs, based on the need to intervene on a spectrum of cancer-overactive mechanisms possibly acting in a parallel or redundant fashion.

3.2.2 Design

The major goals in designing nanoparticles as a delivery system are to control particle size, surface properties and release of pharmacologically active agents in order to achieve the site-specific action of the drug at the therapeutically optimal rate and dose regimen (Mohanraj and Chen, 2006).

The advantages of using nanoparticles as a drug delivery system include the following:

- Particle size and surface characteristics of nanoparticles can be easily manipulated to achieve both passive and active drug targeting after parenteral administration.
- They control and sustain release of the drug during the transportation and at the site of localization, altering organ distribution of the drug and subsequent clearance of the drug so as to achieve increase in drug therapeutic efficacy and reduction in side effects.
- Site-specific targeting can be achieved by attaching targeting ligands to surface of particles or use of magnetic guidance.

In spite of these advantages, nanoparticles do have limitations. For example, their small size and large surface area can lead to particle-particle aggregation, making physical handling of nanoparticles difficult in liquid and dry forms. In addition, small particles size and large surface area readily result in limited drug loading and burst release. These practical problems have to be overcome before nanoparticles can be used clinically or made commercially available.

Chapter 4

Microscale and nanoscale robotics

4.1 Introduction

Due to the developments in micro- and nanoscale science and technology and biotechnology in last two decades, handling µm- and nm-scale objects and accessing tiny spaces down to µm/nm sizes have become critical issues. Progress in robotics over the past years has dramatically extended our ability to explore the world from perception, cognition and manipulation perspectives at a variety of scales (see Figure 4.1). In particular many research groups have proposed various micro- and nanoscale robotic manipulation systems and miniature mobile robots since the beginning of the 1990s (Sitti, 2007). Some of these robots have enabled many new applications, such as prototyping micro/nanodevices and characterizing micro/nano- materials using micro/nanorobotic manipulation systems; minimally invasive diagnosis and treatment in-

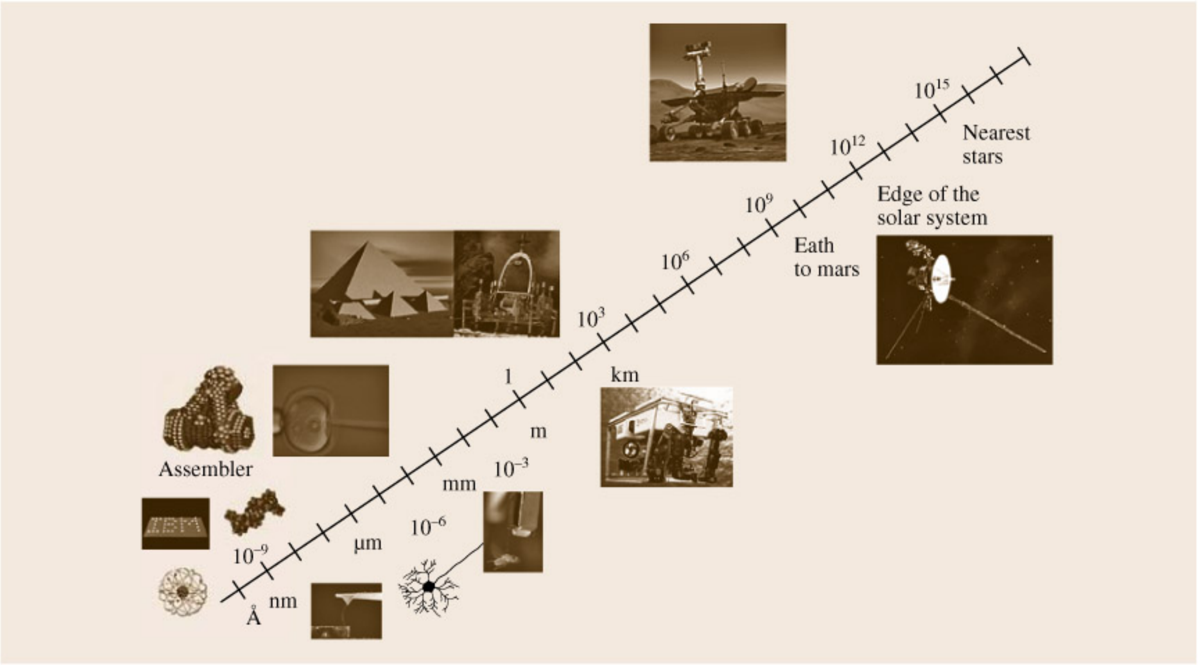


Figure 4.1: Robotics exploration. Taken from (Nelson and Dong, 2010)

side or outside the human body; and mobile sensor networks for environmental monitoring, search and rescue, and homeland security using miniature mobile robots.

Depending on their overall size, sensing and actuation precision, part or tool size, task space, and physics being $\mu\mathrm{m}$ or nm scale, robotic systems can be classified as microrobotics or nanorobotics, respectively. Micro/nanorobotics represents these two different scale robotics areas jointly while keeping their clear scale differences in mind.

Two major research areas can be identified:

- new methods for programmable manipulation and assembly of micro- and nano-scale entities (*micromanipulation*)
- miniaturization of mobile robots down to $\mu\mathrm{m}$ overall sizes (*microrobotics*)

The first area deals with the manipulation and/or assembly of nanoscale components with macroscale instruments or robots (i.e., nanomanipulators). Due to the advances in nanotechnology and its rapidly growing number of potential applications, it is evident that practical technologies for the manipulation and assembly of nanoscale structures into functional nanodevices need to be developed. Nanomanipulation and nanoassembly may also play a crucial role in the development of artificial nanorobots themselves. Manipulation at the nanoscale is still in its infancy and the physical and chemical phenomena at this scale are not completely understood (Sitti, 2001; Sitti, 2007). At last, it is important to stress that, here, the overall micro/nanorobotic system size can be very large, while only the manipulation tools, manipulated objects, and sensing, actuation, and manipulation precision are required to be at the micro/nanoscale

The second area deals with the design, simulation, control, and coordination of robots with nanoscale dimensions (i.e., nanorobots). Nanorobots, nanomachines, and other nanosystems are objects with overall dimensions at or below the micrometer range and are made of assemblies of nanoscale components with individual dimensions ranging approximately between 1 to 100 nm. Much of the research conducted in this area remains highly theoretical at the present, primarily due to the difficulties in fabricating such devices. Although artificial nanorobots do not yet exist, in nature biological nanorobotic systems do exist and provide evidence that such systems are at least possible (Requicha, 2003). As a result, nanorobots have for the most part been explored in the biological context of nanomedicine. In these mobile robotic systems, overall system size is very limited, which induces severe constraints in utilized actuators, sensors, motion mechanisms, power sources, computing power, and wireless communication capability (Sitti, 2007).

Since this thesis is about nanorobots, here the main interest is on the miniaturization of mobile robots. Before discussing the issues miniaturization has to deal with and the current state of the art, swarm robotics will be briefly introduced. In fact swarm robotics is the approach that seems most promising to tackle the problem of controlling a large number of micro- and nano-robots. A recent introductory book on robotics, that describes also swarm robotics and the challenges of nanorobotics, is (Winfield, 2012).

4.2 Swarm Robotics

In a *multi-robot system* multiple robots, instead of a single robots, are used together to achieve a task (Winfield, 2012). The idea of multi-robot systems is not new and, recently, even commercial systems became available; like the Kiva warehouse robot system (Wurman et al., 2007). The standard approach in multi-robot systems is to have centralized control and coordination. *Swarm robotics* offers an alternative approach in which there is no centralized control.

As stated in (Dorigo and Şahin, 2004), swarm robotics can be loosely defined as the study of how collectively intelligent behavior can emerge from local interactions of a large number of relatively simple physically embodied agents. In a swarm of robots, the control is decentralized and distributed between the robots themselves.

Swarm robotics can be characterized by the following 4 features (Dorigo and Şahin, 2004):

1. The swarm consists of a large number of robots, and their control and coordination is decentralized

2. The swarm should consist of a few homogeneous groups of robots — each group containing a large number of robots

3. The robots in the swarm should have difficulties in carrying out or completing the considered task on their own, and cooperation among robots should improve their performance

4. The robots in the swarm should only have local and limited sensing and communication abilities

In swarm robot systems there is no hierarchy, no predetermined leader, and control is completely distributed between the robots of the swarm.

The use of the term “swarm” (instead of “fleet”, for example) is related to the fact that swarm robotics embodies the principles of swarm intelligence. The next section will give a definition of swarm intelligence. Then, to better understand how the principle of swarm robotics can be implemented in practice and how swarm of robots can tackle real world-applications, 3 recent European projects will be described: SWARM-BOT, SWARMANOID and I-SWARM.

4.2.1 Swarm intelligence

The expression swarm intelligence is used to describe any attempt to design algorithms or distributed problem-solving devices inspired by the collective behavior of social insect colonies and other similar societies (Bonabeau et al., 1999). Since the early 1990s, a significant amount of work has been done using social insect-inspired algorithms to solve both ‘toy’ and ‘real’ problems.

There are many examples of the impressive capabilities of social insects (Bonabeau et al., 1999). Weaver ant workers form chains of their own bodies, allowing them to cross wide gaps and pull stiff leaf edges together to form a nest. Honey bees have sophisticated foraging strategies to exploit food sources according to their quality, number and distance from the hive. Nest construction in the wasp *Polybia occidentalis* involves 3 groups of workers, pulp foragers, and builders. Some species of termites build really complex mounds, comprised of roughly cone-shaped outer walls; such nests have humidity and temperature control, fungus gardens, nursery chambers, and peripheral galleries constructed both above and below ground which connect the mound to its foraging sites.

The puzzle is how insects that seem to be so incapable at the individual level, can create such complex and sophisticated structures. Moreover the same insects are also able to repair and maintain their productions.

We know that the collective behavior of the insects is not governed by an architect, a king or a 'brain' individual. It seems that, on one side, some of the mechanics underlying cooperation are genetically determined (e.g., anatomical differences between individuals). On the other side the activities are self-organized. Each individual is characterized by a set of simple behaviors, each triggered by some condition — interaction with another individual, physical contact with a specific part of the environment, chemicals (like pheromones),... No individual has knowledge of the global aim (or plan), but collectively all the individuals are able to achieve remarkable results.

Swarm intelligence encompasses a wide range of families of algorithms. One of the most popular swarm intelligence algorithm is ant colony optimization (Dorigo et al., 1996). But there are many others: Particle swarm optimization (Kennedy and Eberhart, 1995), artificial immune systems (Dasgupta, 2006),...

4.2.2 SWARM-BOT

The aim of SWARM-BOT project (Dorigo, 2005) is to design small mobile robots capable of autonomous physical aggregation into specific shapes so as to perform specific functions. To this aim small robots, called *s-bots*, with a number of sensors and motors, basic communication devices, and limited computational capabilities have been built. The s-bots have an additional key feature: an appendage that allows them to form collective physical structures and disband at will. These collective physical structures are called *swarm-bots*. A prototype s-bot is shown if Figure 4.2 (top), it has a gripper that can be used to grab objects or to connect to other s-bots. The s-bot has a diameter of about 12 cm, and weights 660 g.

A swarm-bot is an aggregate of s-bots that has the potential to exhibit capabilities that go beyond those of a single s-bot. Once assembled, swarm-bots are not limited to a single configuration, but can change its shape while moving, according to their needs. Figure 4.2 (middle) shows a swarm-bot with a linear shape composed of 4 s-bots and moving on rough terrain.

The experimental setting scenario settled in the project required a swarm of s-bots must transport a heavy object from an initial to a target location. There were several possible paths between these two locations; paths with different lengths and requiring to avoid obstacles (e.g., walls and holes). The weight of the object was such that its transportation requires the coordinated effort of many s-bots (Dorigo, 2005). Figure 4.2 (bottom) shows the experimental arena built in the project with 9 s-bots and some obstacles; see (O'Grady et al., 2011) for details.

The swarm control system developed in the project, based on artificial neural networks shaped by evolutionary algorithms (Dorigo et al., 2004), can find simple and effective solutions that allow the s-bots to move in a coordinated way independently of the topology of the swarm-bot. Moreover these controllers proved robust with respect to variations in the size and shape of the object they had to transport (Groß and Dorigo, 2004).

The presence of many autonomous entities, which can assemble into a single body and disassemble back into disparate elements as required, ensures hardware versatility and robustness. Smaller, simpler robots are usually each more reliable than the equivalent same-task-achieving single robot

Figure 4.2: The s-bot (*top*), a swarm-bot with a linear shape composed of 4 s-bots and moving on rough terrain (*middle*), and the experimental arena (*bottom*). Taken from (Dorigo, 2005)

Figure 4.3: The swarmanoid robots: foot (*top*), hand (*middle*) and eye with ceiling attachment system (*bottom*). Taken from (Dorigo et al., 2012)

4.2.3 SWARMANOID

The SWARMANOID project (Dorigo et al., 2012) is a successor of SWARM-BOT. The latter is an example of homogeneous system, because each s-bot is identical to the others. The former, instead, is an example of heterogeneous system. In fact the swarm is composed of 3 distinguished types of robots: *foot-bots*, *hand-bots* and *eye-bots*. This heterogeneity gives the overall system more flexibility to undertake a wider range of tasks, while retaining the robustness of the swarm.

The foot-bot (Figure 4.3, top) is an evolution of the s-bot platform. The robot is conceptually modular at all levels (mechanics, electronics and software). The modules can be stacked one on top of another and are the following: base, gripper, range and bearing, distance scanner, and upper. The foot-bot is 28 cm high and has a diameter of 13 cm (i.e., it is slightly bigger than the s-bot).

The hand bot (Figure 4.3, middle) has no autonomous mobility on the ground, but is able to climb standard office furniture, grasp small objects such as books or letters, and bring such objects to the ground. For the swarmanoid to transport an object, the hand-bot can grasp the object while itself being transported by the foot-bots. The hand-bot can thus interact physically with other robots of the swarmanoid. The hand-bot combines a grasping hand with a climbing assistance device based on a rope launcher and a magnetic attachment system. At the electronic level, the robot has an architecture identical to the foot-bot, and shares most of the basic components. The hand-bot is 29 cm high, 41 cm wide in its widest configuration (with its arms fully retracted) and 47 cm long in its longest configuration (with its arms fully extended).

The eye-bot (Figure 4.3, bottom) is an autonomous flying robot designed to operate in indoor environments. The eye-bots work in synergy with the rest of the swarmanoid: they provide an aerial view to detect the objects of interest and to direct the actions of other robot types. The eye-bot features also a ceiling attachment system that enables an energy saving operation mode in which the eye-bot can power down its flight systems, but continue to scan the environment and communicate with the rest of the swarmanoid. Finally, the eye-bot has 4 carbon fiber legs for support. In total, the carbon fiber structure weighs only 270 g. The outer diameter is 50 cm and the total height including the legs and ceiling attachment is 54 cm.

The task to test the swarmanoid concept was to search and retrieve objects in a complex 3D environment. Objects were placed on one or more shelves in a human habitable space (such as an office building). The swarmanoid robots (assumed to start from a single deployment area) had to find the shelves containing relevant objects, and then to transport the objects from the shelves back to the deployment area.

In the project this task has been successfully tested in a real-world experimental setting, in an environment containing a single shelf and a book (the object to be retrieved). Moreover a new open source multi-robot simulator (targeted at the real time simulation of large heterogeneous swarm of robots) has been developed (Pinciroli et al., 2011).

4.2.4 I-SWARM

The vision of the I-SWARM project (Seyfried et al., 2005) was to take a leap forward in robotics research by combining expertise in micro-robotics and in self-organizing biological swarm systems. The project aim was to facilitate the mass-production of micro-robots and to employ them as a “real” swarm consisting of up to 1000 members. Although the large-scale swarm of micro-robots has not been built, I-SWARM went a long way in this direction.

The prototype robot, shown in Figure 4.4, measured between 3 and 4 mm in each direction, with a total volume of 23 mm^2 and total mass of 65 mg (Edqvist et al., 2009). The robot consists of an Flexible printed circuit boards acting as the robot backbone connecting the modules: a solar cell (on the top), capacitors for energy storage, an infra red communication module, an application specified integrated circuit, and a locomotion module. The locomotion module consists of three legs and includes a vibrating contact sensor.

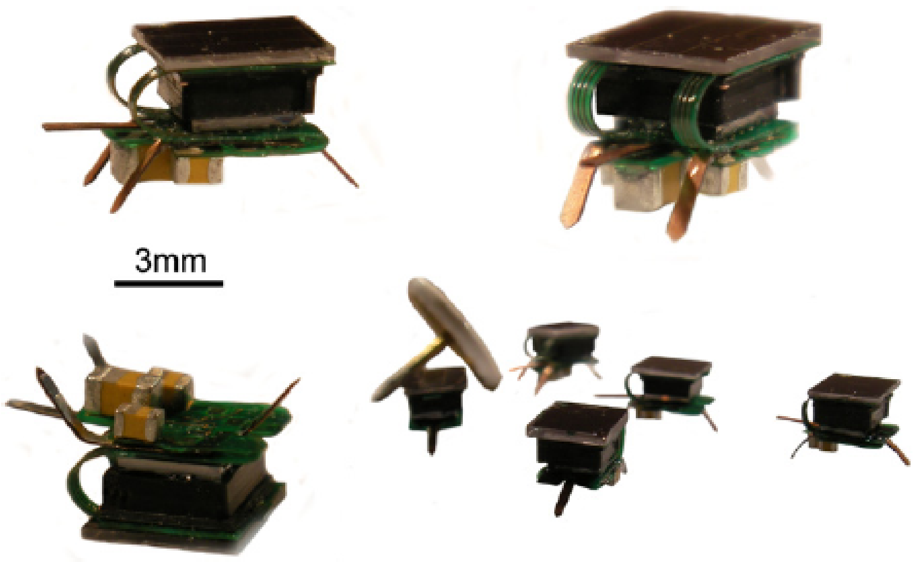


Figure 4.4: The fully assembled and folded I-SWARM robot in three different views. The bottom right image illustrates a small swarm of folded robots and a drawing pin. Taken from (Edqvist et al., 2009)

Robots communicate with each other by using a microfabricated optical module containing 4 pair of LEDs and photodetectors (Corradi et al., 2009). In the communication strategy proposed in (Corradi et al., 2009), robots walk randomly and frequently stop in order to emit some light pulses from the LED. During communication, each LED belonging to one robot is identified with a particular bit string; for the 4 LEDs case, 2-bit strings are enough. Surrounding robots can detect not only the presence and position of one or more robots, but also understand if it/they is on a direct collision course and reacts accordingly: the specific LED codes received inform the robots about the walking direction of the transmitting robot.

4.3 Life at miscoscopic scale

Micro/nanorobots are different than macroscale robots in many aspects. First, they can access smaller spaces or scales directly due to their small size and grippers. Next, one of the major differences is in their physics.

Swimming or flying in fluids seems more attractive than walking or crawling on a surface, since most objects likely to be encountered on a surface are large and difficult to overcome by a nanoscale walking or crawling machine. Bacteria are good models for nanorobots because they have sizes on the order of a few micrometers, which are likely to be comparable to those of future nanorobots, and move in fluids.

In the movie *Fantastic Voyage* (20th Century Fox) a team of scientists and their submarine are miniaturized using a top-secret technology, enabling them to navigate through the human blood stream to save the life of a dying man trying to destroy the blood clot that threatens the patients life.

This is an imaginative vision of noninvasive surgery, but the problem is that even if this amazing shrinking technology existed, the miniature submarine and its crew would be totally ineffective (Abbott et al., 2007):

- The submarine propellers would not function at that scale
- The crew would not be able to walk around inside the submarine because of the adhesive forces between their feet and the floor
- Even if the crew could make it out of the submarine, the tiny scuba divers could kick their legs all day and go nowhere

This is due to the fact that objects of a few microns (like a bacteria, the movie submarine and a nanorobot) move in the Stokes (or low-Reynolds number) regime, which can be highly counterintuitive (Purcell, 1977; Berg, 1993): inertia is negligible, motion is controlled entirely by friction, sliding is impossible, propulsion cannot be achieved by symmetric motions, jet propulsion does not work,…Bacteria move in this regime typically by using cilia or rotating flagella.

What is changing moving from the macroscopic to the microscopic world is the relative importance of the physical laws. Lets describe the scale of a system by its linear length L (also called *linear scale*). It is usually assumed that all directions scale proportionally to L. This means that the surface area scales as $\simeq L^2$ and the volume as $\simeq L^3$. Friction, heat transfer and surface forces are associated to surface area, while inertia, weight, heat capacity and body forces are associated to volume.

4.3.1 Adhesion force

An interesting example of the effects of scaling is given by adhesion forces. In fact they are always present, bur their relevance with respect to other forces increases with the decreasing of the linear scale.

Let's consider the vand der Waals force F_W between a sphere with radius r and an infinite halfspace (Abbott et al., 2007):

$$F_W = \frac{Hr}{8\pi d^2},$$

where d is the distance between the sphere and the halfspace, and H is the material-dependent Hamaker constant (typically $\simeq 10^{-20}$ J). Scaling only the radius r as $\simeq L$ results in $F_W \simeq L$. If also d is scaled as $\simeq L$, the force will scale as $\simeq L^{-1}$. This means that the importance of F_W increases when L decreases.

Moreover, in the case that a liquid water film can form between the sphere and the halfspace, the associated tension force F_t has to be considered as well (Abbott et al., 2007). This force can be approximated as:

$$F_t = 2\pi\gamma r,$$

where γ is the material-dependent surface tension expressed in N/m. Since $r \simeq L$ also $F_t \simeq L$.

4.3.2 Reynolds number and fluid mechanics

The *Reynolds number* is a dimensionless parameter in the equations of motion of a fluid that indicates the relative size of terms that describe inertial forces (forces required to accelerate masses) and viscous forces (forces due to viscous shear).

Definition 1 (Reynold number) *The* Reynolds number R *is defined as*

$$R = \frac{vL\rho}{\eta},$$

where v is the velocity of the fluid (the velocity of a particle moving through the fluid), L is the linear scale of the motion (the size of the particle), ρ is the specific gravity (the density) of the fluid, and η is its viscosity.

Example 1 (Fish) *For a fish $v = 1$ m/s, $L = 0.1$ cm, $\rho \simeq 10^3$ kg/m^3, $\eta = 10^{-3}$ kg/(m s). Then $R = 10^5$.*

Example 2 (Bacterium) *For a bacterium $v = 10^{-5}$ m/s, $L = 10^{-6}$ m, $\rho \simeq 10^3$ kg/m^3, $\eta = 10^{-3}$ kg/(m s). Then $R = 10^{-5}$.*

Example 3 (Nano-submarine) *For a nano-submarine (like that in Fantastic voyage movie) moving in the capillaries $v = 5 \cdot 10^{-4}$ cm/s (Stücker et al., 1996), $L = 5 \cdot 10^{-4}$ cm, $\rho \simeq 10^3$ kg/m^3, $\eta = 10^{-3}$ kg/(m s). Then $R = 0.25$.*

The Reynold number of the fish is very large, and those of bacterium and nano-submarine are very small. This means that the fish propels itself by acceleration, while bacterium and submarine by using *viscous shear*. The fish knows a lot about inertia, the bacterium and nano-submarine know nothing; at low Reynolds number inertia is totally irrelevant. In other words, the hydrodynamics worlds at large and small Reynold numbers are very different.

To stress this point, let compute the distance that a bacterium can coast when it stops swimming (Berg, 1993). To do so, the bacterium is approximated as a sphere of radius r and density ρ_s. According to Newton's second law, $F = mv'(t)$, the sphere decelerates according to the equation:

$$-mv'(t) = 6\pi\eta r v(t),$$

where $m = \frac{4}{3}\pi r^3 \rho_s$ is the mass of the sphere. Rearranging the previous equation, we get the differential equation:

$$\frac{v'(t)}{v} = -\frac{9\eta}{2r^2\rho_s},$$

which has as solution:

$$v(t) = v(0)e^{-\frac{t}{\tau}},$$

where $v(0)$ is the initial velocity of the sphere, and

$$\tau = \frac{2r^2\rho_s}{9\eta}.$$

For $r = 10^{-4}$ cm, and $\rho_s = 1$ g/cm^3, $\tau = 2 \cdot 10^{-7}$. This mean that the bacterium stops in less than 1 μs. meanwhile, it coasts a distance d given by

$$d = \int_0^\infty v(t)dt = v(0)\tau.$$

Consequently, if the initial velocity was 10^{-5} m/s, it coasts a distance $d = 2 \cdot 10^{-12}$ m, which is small even compared with the diameter of a hydrogen atom[1]. If you are at very low Reynolds number, what you are doing at the moment is entirely determined by the forces that are exerted on you at that moment, and by nothing in the past (as in Aristotle's mechanics)

Due to these properties, the specificity of tasks increases exponentially at the micro/nanoscale

4.4 Programmable manipulation and assembly of micro- and nanoscale entities

Although the focus of this work is on nanorobots, it is worthwhile to give a brief description of the research area devoted to programmable micro/nanomanipulation.

By the early 1980s, scanning tunneling microscopes (STMs) radically changed the ways in which we interacted with and even regarded single atoms and molecules. The very nature of proximal probe methods encourages exploration of the nanoworld beyond conventional microscopic imaging. Scanned probes now allow us to perform engineering operations on single molecules, atoms, and bonds, thereby providing a tool that operates at the ultimate limits of fabrication. They have also enabled exploration of molecular properties on an individual non statistical basis (Nelson and Dong, 2010).

STMs and other nanomanipulators are nonmolecular machines but use bottom-up strategies. Although performing only one molecular reaction at a time is obviously impractical for making large amounts of a product, it is a promising way to provide the next generation of nanomanipulators. Most importantly, it is possible to realize the directed assembly of molecules or supermolecules to build larger nanostructures through nanomanipulation. The products produced by nanomanipulation could be the first step of a bottom-up strategy in which these assembled products are used to self-assemble into nanomachines.

Nanomanipulation (positional or force control at the nanometer scale) is a key enabling technology for nanotechnology by filling the gap between top-down and bottom-up strategies, and may lead to the appearance of replication-based molecular assembler. These types of assemblers have been proposed as general-purpose manufacturing devices for building a wide range of useful products as well as copies of themselves (self-replication).

---

[1] The bacterium is still subject to Brownian motion. This means that it does not actually stop. The drift goes to zero, not the diffusion

Nanomanipulation was enabled by the inventions of the STM, of the atomic force microscope (AFM), and other types of scanning probe microscope (SPM). Nanorobotic manipulators (NRMs) are characterized by the capability of 3-D positioning, orientation control, independently actuated multiple end-effectors, and independent real-time observation systems, and can be integrated with scanning probe microscopes. NRMs largely extend the complexity of nanomanipulation. STM, AFM, and NRM technology are compared in Figure 4.5

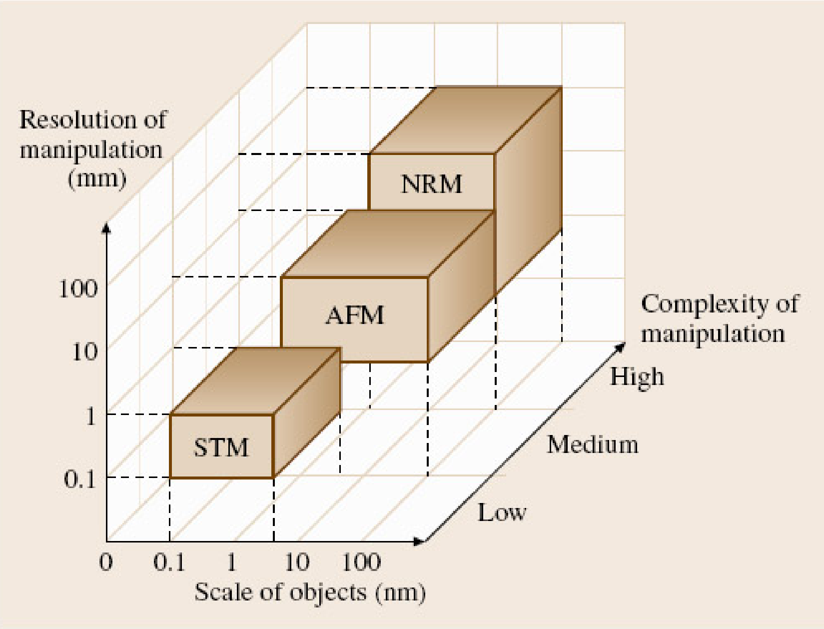


Figure 4.5: Comparison of nanomanipulators. Taken from (Nelson and Dong, 2010)

The first nanomanipulation experiment was performed back to 1990 (Eigler and Schweizer, 1990). They used an STM and materials at almost absolute zero temperature to position individual xenon atoms on a single-crystal nickel surface with atomic precision. The manipulation enabled them to fabricate rudimentary structures of their own design, atom by atom. The result is the famous set of images showing how 35 atoms were moved to form the three-letter logo IBM, demonstrating that matter could indeed be maneuvered atom by atom as Feynman suggested (Feynman, 1992).

After the mid of 1990s, AFM has become one of the major micro/nanorobotic systems to push, pull, cut, and indent any mate- rial and surface in any environment (air, liquid, or vacuum), as illustrated in Figure 4.6. The main advantage of using the AFM as a micro/nanomanipulator is its use of sense force and image three-dimensional (3-D) topography.

All the precision micro/nanomanipulation systems have the following main technical challenges to be solved (Sitti, 2007): there are no micro/nanophysical simulation and design tools for any random micro/nanogripper and entity geometries and materials in air, liquid, or vacuum; autonomous and intelligent control of the micro/ nanomanipulation systems has not been established; 3-D programmable assembly of micro/ nanoscale building blocks has not yet been demonstrated; manipulation of entities down to molecular sizes and subnanometer precision is very challenging; and massively parallel and distributed active control of an array of micro/nanomanipulators has not been established.

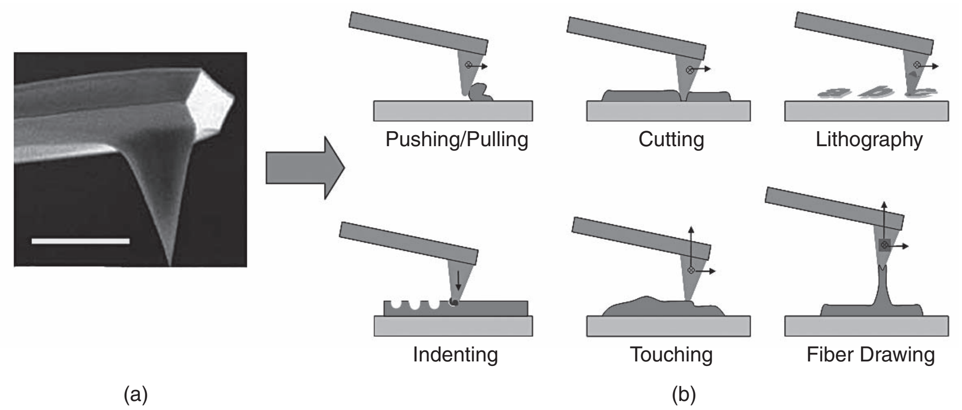


Figure 4.6: a) An SEM image of a silicon AFM probe with about a 10-nm tip radius (scale bar: 10 μm) and (b) example 1-D and 2-D mechanical micro/nanomanipulation tasks using such an AFM probe. Taken from (Sitti, 2007)

4.5 Miniature mobile robots

Micro/nanomanipulation robots have the grippers, traps, imaging, and positioning precision at the micro/nanoscale, but their overall size is not necessarily on the order of a few millimeters or centimeters. Therefore, the miniature micro/nanorobot research area focuses on miniaturization of robots down to submillimeter sizes as an entire system. Such a mobile robot mainly consists of miniature actuators, mechanisms, sensors, computing, communication, and other electronics, and the power source to drive them.

4.5.1 An untethered silicon microrobot (Dartmouth College)

An example of microrobot, whose overall size is about $60 \times 250 \times 10 \mu\mathrm{m}^3$, has been developed at Dartmouth College (Donald et al., 2006). Figure 4.7 shows the structure of this device. It consists of an untethered scratch drive actuator (A), with a cantilevered steering arm (B) that protrudes from one side. The untethered scratch drive is used for locomotion, while the steering arm is used to raise or lower a stylus into contact with the substrate. When this stylus is in contact with the substrate, it provides enough friction to cause the device as a whole to turn.

The device receives its electrical power and control signals through an array of insulated interdigitated electrodes that cover the device's operating environment. Since the control signal and electrical power are both available to the device anywhere within this environment, the device can move freely, untethered by the wires and rails that power most electrostatic MEMS devices.

Since the scratch drive and the steering arm are fabricated monolithically from the same sheet of conductive polysilicon, the independent actuation of these two components is obtained by exploiting the electro-mechanical hysteresis of the components trough the application of sequences of voltages in a control waveform. The scratch drive can be either flexed or relaxed, and the steering arm can either be in contact with the substrate, or out of contact. Hence, the state of the device can be discretized into four possible values, according to whether or not each

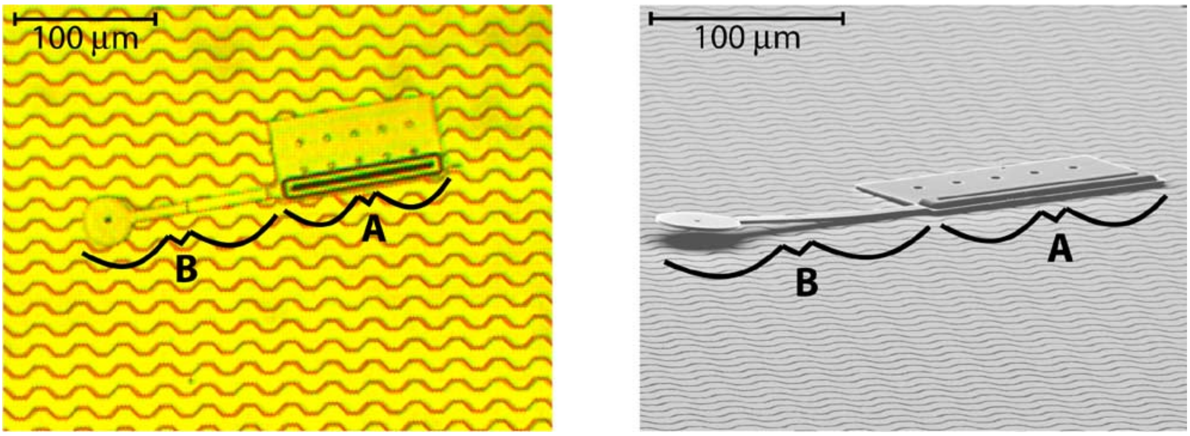


Figure 4.7: Optical (left) and electron (right) micrographs of an electrostatic micro-robot. The device consists of an untethered scratch drive actuator (A), with a cantilevered steering arm (B) that protrudes from one side. The untethered scratch drive is used for propulsion, while the steering arm can be raised or lowered to turn. An array of insulated interdigitated electrodes (lighter-colored background) provides electrical power and control signals to the device. Taken from (Donald et al., 2006)

of the two components has crossed its transition point. The control system for the micro-robot can then be treated as a finite state machine, as shown in Figure 4.8. Although simple, this finite state machine enables complex motions of the microrobot.

4.5.2 Nanorobotics

Nanorobots are but one example of nanoelectromechanical systems (NEMS), which represent a new frontier in miniaturization, looming beyond the microelectromechanical systems (MEMS) that today constitute a multibillion-dollar industry (e.g., in mobile devices and videogame controllers).

Nanorobotics is an emerging interdisciplinary technology area raising new scientific challenges and promising revolutionary advancements in applications such as medicine, biology and industrial manufacturing (Mavroidis and Ferreira, 2009). It is concerned with (Requicha, 2003):

- design and fabrication of nanorobots with overall dimensions at or below the micrometer range and made of nanoscopic components with individual dimensions ranging between 1 and 100 nm
- programming and coordination of large numbers (swarms) of such nanorobots
- programmable assembly of nanometer-scale components either by manipulation with macro or micro devices, or by self-assembly on programmed templates or scaffolds.

Interest in nanorobotics is growing rapidly. This growth of interest reflects the enormous potential of the technology, and also recent technical advances (for instance, in nanomachine

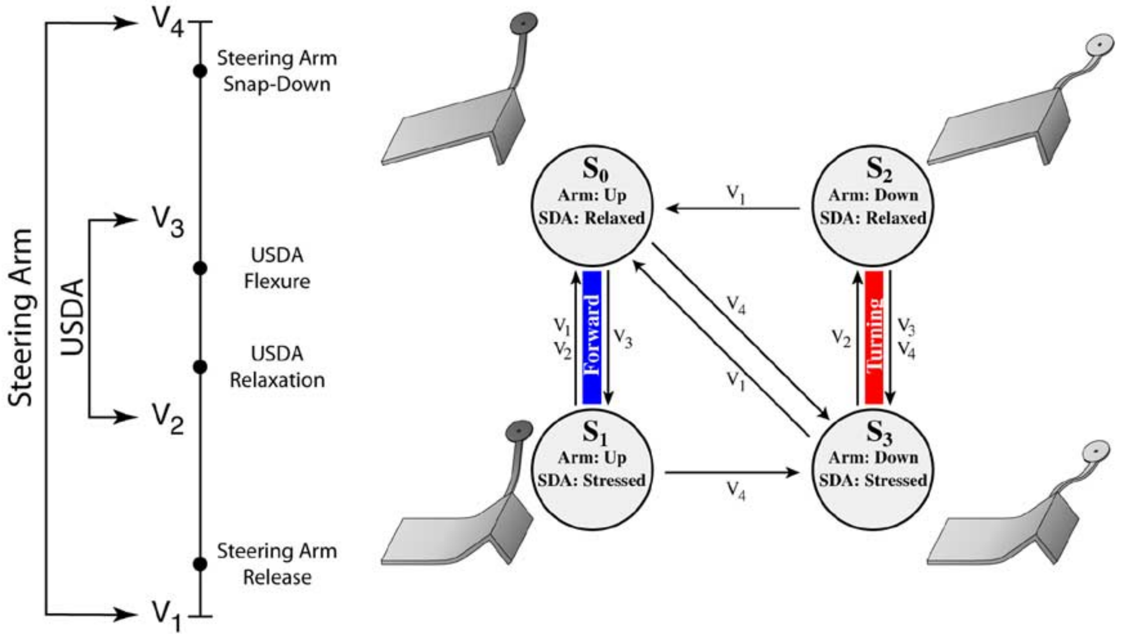


Figure 4.8: The state transition diagram of the micro-robot. Four voltages, $V_1 < \ldots < V_4$, are used in constructing control waveforms. For the untethered scratch drive actuator (USDA), any voltage below the relaxation voltage will cause it to relax, and any voltage above the flexure voltage will cause it to flex. For the steering arm, any voltage below the release voltage will raise the arm, and any voltage above the snap-down voltage will lower it. The four system states ($S_0, \ldots, S_3$) correspond to the possible combinations of the states of the steering arm and the USDA. When the voltage is changed, the system will transition to a new state. Since the control voltages of the USDA nest within the control voltages of the steering arm, there is a sequence of voltages that allows the system to transition between any pair of states. The device moves forward by repeatedly transitioning between states S_0 and S_1 , and turns by transitioning between states S_2 and S_3. Taken from (Donald et al., 2006)

synthesis) that suggest that nanorobots will not remain in the realm of science fiction much longer.

Nanorobots could be defined as intelligent systems with overall dimensions at or below the micrometer range (i.e., overall dimensions comparable to those of biological cells and organelles) that are made of assemblies of nanoscale components with individual dimensions ranging between 1 and 100 nm. Nanorobots would be able to perform at least one of the following actions: actuation, sensing, signaling, information processing, intelligence, swarm behavior at the nano scale.

In general, nanorobotics carries a variety of definitions throughout the literature. Consequently, the field of nanorobotics can be generally divided into two main focus areas (Requicha, 2003).

Order-of-magnitude feasibility calculations (Requicha, 2003) indicate that nanorobots are not physically impossible. They would be extremely useful not only in the medical field but also in applications such as:

- monitoring and interacting with harmful microorganisms in the air or in water
- building intelligent surfaces with a controllable (programmable) structure, for example, with variable rough- ness and friction.

Nanorobotics and, more generally, NEMS research involves design (which often is biologically inspired), prototyping, fabrication, programming, and applications such as biomedical nanotechnology.

4.5.3 Locomotion

When mobile robots are scaled down, robot motion dynamics change significantly. At first, the robot motion characteristic time scale gets shorter, i.e., the structural resonant frequencies increase. Then, drag and friction forces dominate the inertial forces (as described in Sec. 4.3), and surface-area-based power sources such as solar cells become more feasible than bulk dominated power sources such as batteries (Sitti, 2007).

To see these scaling effects on robot locomotion, different locomotion types can be considered. For flying, Reynolds number becomes very small. Thus, the aerodynamics of a centimeter-size robot or a fly is unsteady and gets very complicated to analyze and control.

For climbing, miniaturization of robots is favorable since the attachment forces mostly depend on the foot pad area while the robot weight depends on the volume. Small animals such as geckos and many insects are light and can climb on surfaces very quickly and robustly.

For tiny robots, even unique locomotion principles become possible. For example, repulsive surface tension forces can dominate buoyancy for a water strider insect to lift its body weight so that a similar insect robot can stay and walk on water very quickly and efficiently by never breaking the water surface, similar to its biological counterpart.

As example of such a robot, in (Song and Sitti, 2007) it has been proposed a reprogrammable and controllable microrobot capable of walking on water inspired by a water strider (see Figure 4.9). Water strider are tiny insects, weighing only 0.01 g, able to move on top of the surface of water at the speed of 1.5 m/s without breaking the surface. This insect maintains itself on the surface of water mainly utilizing the surface tension force. Similar to the biological system, this robot demonstrates surface tension dominated locomotion on water. The robot weight is

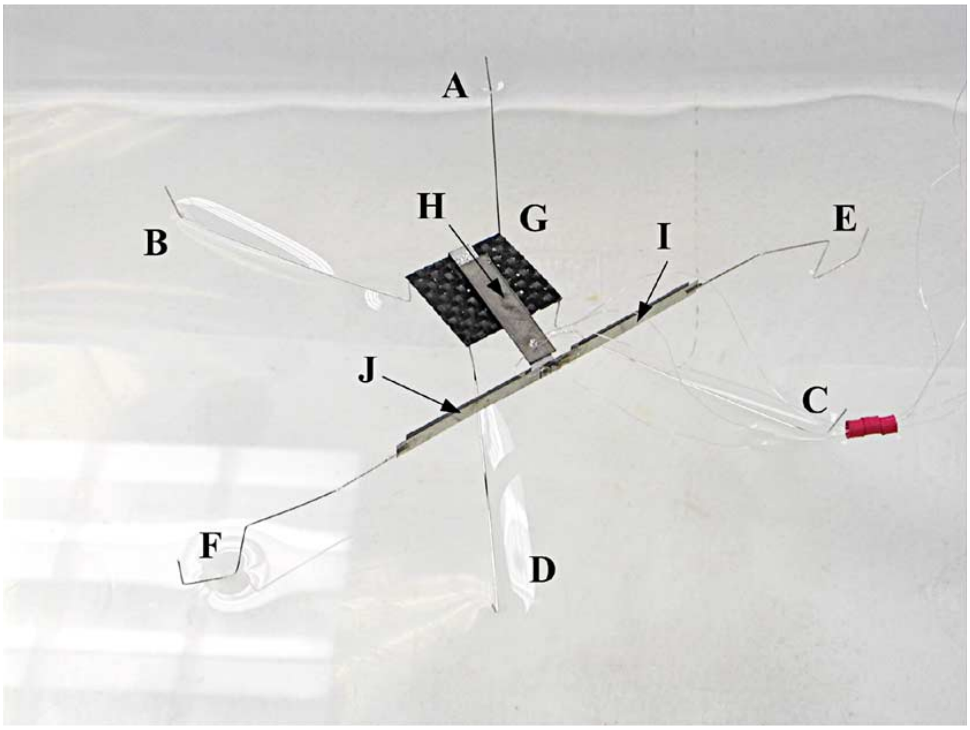


Figure 4.9: Photo of the prototype robot on the surface of water. Taken from (Song and Sitti, 2007)

about 1 g. Four 5-cm-long Teflon coated legs are fabricated and glued to the body which is made of three interlacing layers of carbon fiber sheet (2×2 cm^2). The robot also successfully demonstrated controlled motion in every direction, including the backward and left turn. The forward speed of the robot measured in the experiment is about 3 cm/s (Song and Sitti, 2007).

4.5.4 Power sources

Miniature on-board power sources are one of the most significant bottlenecks of miniature robots. Current miniature batteries continue to get smaller, more flexible, and lighter, but due to the scaling laws they deliver low power at small sizes. Micro fuel cells, MEMS-based silicon solar cells, supercapacitors, microbatteries, and radioactive thin films have been proposed as possible on-board power sources. But, still there is no feasible on-board power source for mm-scale or very light (less than 1 g) robots.

As a potential solution to this critical issue, a miniature robot moving in a limited work space such as inside the human body can be externally powered depending on the locomotion principle or application.

The next possible solution is to scavenge power (Roundy et al., 2004) in the miniature robot environment using sources such as:

- Photovoltaics (Solar Cells)
- temperature gradients
- human power

- air, fluid and wind flows
- mechanical vibrations
- electrical charge gradients

Unlike power sources that are fundamentally energy reservoirs, power scavenging sources are usually characterized by their power density rather than energy density. Energy reservoirs have a characteristic energy density, and how much average power they can provide is then dependent on the lifetime over which they are operating. On the contrary, the energy provided by a power scavenging source depends on how long the source is in operation.

As a new approach, hybrid (biotic and abiotic) miniature actuation and powering methods can be developed by integrating biological motors or microorganisms with inorganic micro/nanostructures. In (Behkam and Sitti, 2007) to demonstrate the feasibility of using bacterial flagella as bioactuators, 10 μm polystyrene (PS) beads are propelled by several *S. marcescens* bacteria attached to them.

The motion of the PS beads was observed with a 60× oil immersion phase objective. Figure 4.10 depicts a PS bead at $t = 0$ and the same bead at $t = 6$ s. he total displacement of the bead was measured to be approximately 90 μm. The average velocity of the bead shown in Figure 4.10 is calculated to be 15 μm/s and there were 10 bacteria attached to the bead.

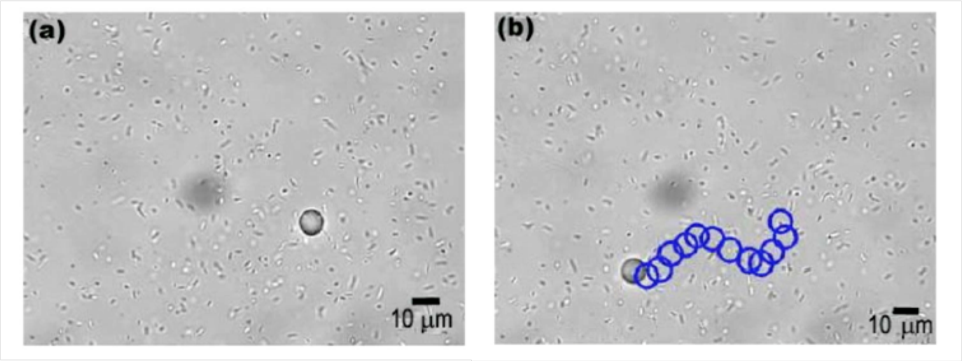


Figure 4.10: Phase-contrast optical microscope images of a mobile 10 μm PS bead with several *S. marcescens* bacteria attached to it at (a) $t = 0$ and (b) $t = 6$ s. PS beads path is shown with rings. Taken from (Behkam and Sitti, 2007)

The power and actuation for this bacterial flagella-based solution are harvested from the bacteria while they are provided with the required chemical energy source and preferred environmental condition.

Current challenges of power sources are the lack of high-power-density, lightweight, long-lifetime, and miniature on-board power sources, and low harvested power magnitudes (mostly on the order of mW).

4.5.5 Major challenges

Current major challenges of miniature mobile robots can be summarized as follows (Sitti, 2007):

1. miniaturization of on-board power sources and actuators, which limit the overall size, operation time, and performance of current miniature robots

2. current miniature robot design and fabrication methods are very case specific, and there are no unified methodology, analysis, and simulation tools or rapid prototyping methods for designing and fabricating such robots

3. when these miniature robots are developed and manufactured reliably, low- and high-level control of these robots is very challenging due to unsteady and nonlinear dynamics, with many uncertain and time-varying physical parameters and limited computing, actuation, sensing, communication, and power source.

4.6 Nanomedicine

Nanomedicine is the application of nanotechnology (the engineering of tiny machines) to the prevention and treatment of disease in the human body. This evolving discipline has the potential to dramatically change medical science.

Established and near-future nanomedicine applications include activity monitors, chemotherapy, pacemakers, biochips, OTC tests, insulin pumps, nebulizers, needleless injectors, hearing aids, medical flow sensors and blood pressure, glucose monitoring and drug delivery systems.

Here are a few examples of how nanomedicine could transform common medical procedures:

- Diagnostic nanomachines could be employed to monitor the internal chemistry of the body. Mobile nanorobots, equipped with wireless transmitters, could circulate in the blood and lymph systems, and send out warnings when chemical imbalances occur or worsen.

- Similar fixed nanomachines could be planted in the nervous system to monitor pulse, brain-wave activity, and other functions.

- Implanted nanotechnology devices could dispense drugs or hormones as needed in people with chronic imbalance or deficiency states.

- In heart defibrillators and pacemakers, nanomachines could affect the behavior of individual cells.

- Artificial antibodies, artificial white and red blood cells, and antiviral nanorobots might be devised.

The most advanced nanomedicine involves the use of nanorobots as miniature surgeons. Such machines might repair damaged cells, or get inside cells and replace or assist damaged intracellular structures. At the extreme, nanomachines might replicate themselves, or correct genetic deficiencies by altering or replacing DNA (deoxyribonucleic acid) molecules.

4.6.1 Potential impact of medical microrobots

The world of a medical microrobot consists of fluid-filled lumens (i.e., tubes) and cavities, as well as soft tissues. Medical microrobots must be designed specifically to work in these environments, but the relative changes in the size, geometry, and material properties of the environment within a given procedure present design challenges. Consider, for example, a microrobot designed to work in the urinary system. The microrobot would be inserted into the urethra and would travel along this lumen to reach the bladder. If the goal is to reach a kidney, the microrobot must first navigate the bladder, which appears as a relatively large open cavity to the microrobot. To reach the kidney, the microrobot must enter the ureter, which is a lumen that enters the bladder at an oblique angle, and then navigate to the kidney. Designing a single microrobot that is able to negotiate these changing environments is not trivial. Fluid flow in the microrobots environment also presents a significant design challenge. Consider a microrobot designed to work in the circulatory system. In addition to dealing with varying blood-vessel diameter, the microrobot must compete against the pulsating flow of blood, which is significant to a small, untethered device (Nelson et al., 2010).

Although designing functional medical microrobots is challenging from an engineering perspective, the potential rewards are vast. In (Nelson et al., 2010) it is given a tentative classification of simple medical microrobot tasks (see Fig 4.11):

(a) Targeted drug delivery will enable increased effectiveness of therapies while minimizing side effects. This microrobot, for example, is docked to a blood vessel as a proposed therapy for retinal vein occlusions

(b) Microrobots can aid in brachytherapy by carrying radioactive seeds to destroy tumors

(c) Time-varying magnetic fields induce heating in a microrobot, which can be used for hyperthermia therapy

(d) In vitro stem-cell research is progressing rapidly, and microrobots can provide a method that applies these results in vivo.

(e) Microrobots can perform ablation to remove unwanted tissue

(f) Microrobots can obtain tissue samples for biopsy, using mechanisms such as this device designed for a miniature endoscopic robot

(g) Microrobots can act as support scaffolding for tissue-engineering applications. This image shows a construct that consists of mineralized ingrowth into a polymer scaffolding

(h) Microrobots can act as stents that can be used to keep passages open.

(i) Microrobots can act as occlusions in cases where a passage needs to be blocked

(j) Microrobots can carry or act as electrode implants in the brain

(k) Microrobots can perform wireless remote sensing, such as this intraocular oxygen sensor whose luminescence is a function of oxygen concentration

(l) A microrobot transmitting its own location is useful for localization, as well as for marking pathologies of interest.

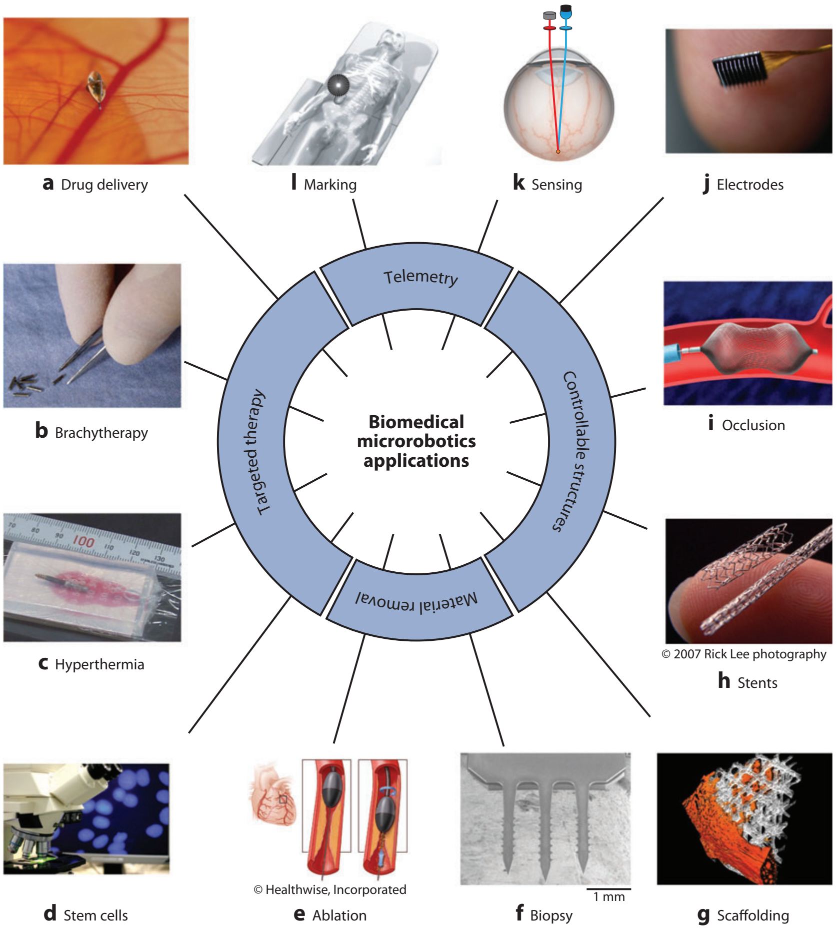


Figure 4.11: Medical tasks for microrobots, including targeted therapy, material removal, controllable structures, and telemetry. Taken from (Nelson et al., 2010)

4.6.2 Application areas

Once equipped with some basic functionalities and miniaturized to a size of a few millimeters or less, microrobots have the potential to perform medical procedures throughout the body. Potential application areas are (Nelson et al., 2010):

1. The circulatory system
2. The central nervous system
3. The eye
4. The ear
5. The fetus

For its relevance to this thesis, here I briefly describe the circulatory system application area.

The circulatory system consists of the heart and the vessels used to carry blood around the body. Blood vessels consist mainly of four types of tissue: endothelium (internal wall lining), elastic tissue (for pulsation damping), smooth muscle, and connective tissue. The relative composition of a given vessel depends on its function and distance from the heart. As the distance between the heart and a given blood vessel increases, the pressure pulsation resulting from the beating heart becomes attenuated, reaching a limiting behavior of constant flow in the capillaries.

Figure 4.12 shows the blood-flow velocity and Reynolds number as a function of vessel size. The reported Reynolds number are those describing pipe flow. In addition to average values, the maximum flow velocities and Reynolds numbers are also critical, reaching Re = 8500 in the aorta and 1000 in the arteries.

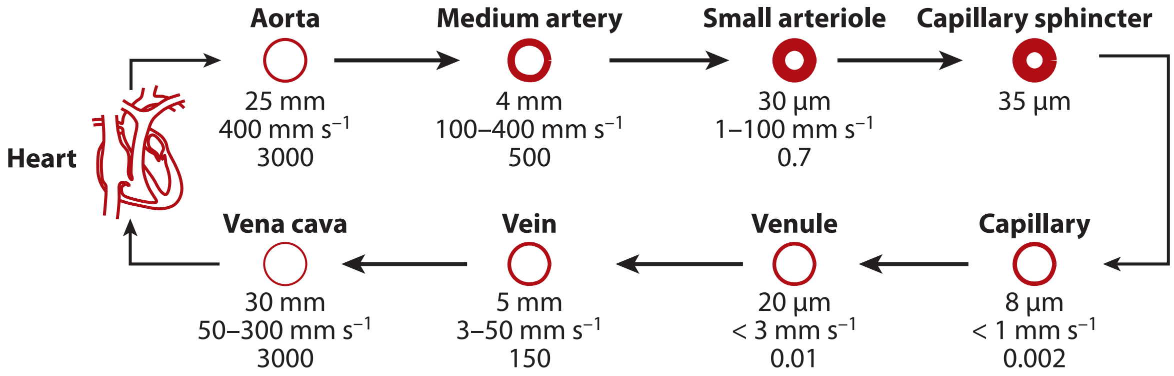


Figure 4.12: Schematic representation of the vessels of the cardiovascular system, with the inner diameter, average blood-flow speed, and Reynolds number. Taken from (Nelson et al., 2010)

Blood consists largely of a watery fluid named plasma, in which red and white blood cells measuring 5–10μm are suspended. Although blood has many properties (density 1060 kg/m^3, pH 7.38–7.42, surface tension 55–61 $\cdot 10^{-3}$ N/m) that are approximately the same as water, its apparent viscosity (3.5e-3 $\text{Pa} \cdot \text{s}$) is three times higher, primarily owing to suspended red blood cells. The assumption that the blood is an homogeneous fluid seems to be lose adequacy at the microscale. Depending on its size, a microrobot could very well experience a working environment closer to obstacle-filled plasma than blood as described by a homogeneous model.

Nearly every site of the body can be accessed by blood, so the circulatory system is arguably the most important application area for wireless microrobots. Some of the most promising applications for microrobots in the circulatory system include performing targeted drug delivery, removing plaque (rotational atherectomy), destroying blood clots (thrombolysis), acting as stents to maintain blood flow, acting as occlusions to intentionally starve a region of nutrition, and administering therapy for aneurysms. Microrobots could also carry electrodes for electrophysiology.

It is worth noticing that the small-diameter capillary network makes unlikely the possibility of round-trip navigation of the body for microrobot always moving with the flow. Consequently, a viable microrobot is likely to require the ability to move against the flow. On the contrary, a suitable designed nanorobot should not require this ability.

Part II

Original contribution

Chapter 5

Nanobot tasks and architecture

> "As a general rule, the greatest crimes never come to light because the greatest criminals, their perpetrators, are too clever to be caught."
> J. Futrelle, *The Scarlet Thread* (Futrelle, 1973)

The main results presented in this chapter have been published in (Cerofolini et al., 2010a) and (Cerofolini and Amato, 2012).

5.1 Introduction

A Nanorobot (since now on *nanobot*) is any artificial machine with overall size of the order of a few microns or less in all spatial directions, constituted by nanoscopic components with individual dimensions in the interval $1 - 10^2$ nm, and able to perform sophisticated functions like navigation, recognition, data collection and transmission (Requicha, 2003). In view of their limited size, nanobots are necessarily devices with very limited computational resources and their interaction can only happen locally. These facts suggest that the algorithms for their control could be based on swarm intelligence; in other words the algorithms could be inspired by the collective behavior of social-insect colonies and other animal societies (Bonabeau et al., 1999).

Ideas regarding swarm of nanobots have already been presented in literature. Methods that use self-assembling agents (nanobots) are discussed in (Requicha, 2003) and (Arbuckle and Requicha, 2004). Medical applications of nanobots are described in (Freitas, 2005; Cavalcanti and Freitas, 2005; Cavalcanti et al., 2006; Freitas, 2006; Hogg, 2007). In particular, chemical communication techniques for nanobots foraging in fluid environments relevant for medical applications are discussed in (Cavalcanti et al., 2006), while the problem of designing control systems for nanobots is analyzed in (Cavalcanti and Freitas, 2005).

The aim of this chapter is manifold. It intends to describe an architecture of nanobots

- addressed to monitoring *in vivo* the health state of complex systems like *homo sapiens sapiens*;

- able to embed sophisticated functions, like navigation, recognition, and data transmission; and

- suitable for being manufactured by processes compatible with today and likely tomorrow semiconductor industries.

5.2 Cancer—the crime

There are several theories of the origin of cancer. Most of them accept that genetic lesions have a major role in determining tumor phenotype, although evidence is accumulating that cancers of distinct subfamilies may derive from different cells of origin (Visvader, 2011). Without advocating any theory of tumor genesis (tissue or mesenchymal stem cell undergoing mutation or a hybridization with an oncogenic virus) and without pretending to formulate any theory of tumor growth, I list a a few general properties of cancerous cells that by themselves are able to account for some general behavior of the complex interaction between the tumor and the vascular system.

A cell becomes malignant essentially because a sequence of a few (about 5), usually extremely unlucky, errors in the replication of its genetic material.[1] In this way, cancer is essentially a disease of the elder.

Cancer is an extremely complex disease whose major features are listed in the following:

C_1 cancerous cells are mutations of tissue cells with unbound Hayflick limit[2]; this fact allows their unlimited reproduction provided that there is adequate room and sufficient amount of nutritive substances;

C_2 cancerous cells are aggressive and destroy their neighboring cells via the emission of proteolitic enzymes or, more likely, by acidification of their environment;

C_3 the metabolism (and thus the reproduction rate) of cancerous cells is much faster (by approximately a factor of 50) than that of non-malignant cells; and

C_4 the mutant cancerous cells retain much of the characters of the non malignant tissue cells.

The combination of all these features has dramatic effects:

- C_2 produces available room to growth thanks to the necrosis of nearby cells;
- C_3 implies that the faster reproduction rate of cancerous cells allows them to hold actually the room produced by C2 before non malignant cells;
- C_4 implies the immune response is not effective, and
- C_1 implies that in the absence of other factors the disease extends until the organism is able to feed it.

Capillaries are the smallest blood vessels in the body, and are responsible for actually delivering oxygen and other nutrients to all the cells of the human body. The capillary system is already optimized to its diffusion basin that cannot tolerate a large increase of O_2 consumption

---

[1] Hence the first rule for cancer prevention: avoid abnormal cell replication by reducing inflammation factors and excessive immune response.

[2] The fact that an organism's cells can only carry out a limited ($\simeq 50$) number of divisions, after which they die.

therein: in fact, moving in the frame of the minimum volume model (Kamiya and Togawa, 1972) and using physiological data of muscle blood flow and O_2 consumption rate in man during heavy exercise, Kamiya and coworkers calculated the optimum capillary number to be 3.7×10^{10} (Kamiya et al., 1987); since this number agrees well with the morphological data, the vascular-tissue system is constructed so as to attain the highest efficiency in O_2 transport to tissue at its maximum activity.

Since the metabolism of cancer cells is approximately 50 times faster than that of healthy cells (Brannon-Peppas and Blanchette, 2004), for solid tumors (to which the following analysis is restricted) a capillary able to sustain 2×10^3 healthy cells can feed approximately 40 cancerous cells. This fact has important effects on the dynamics of tumor growth in the early stages.

Consider now that each capillary is substantially unable to support any large additional request of O_2 to the diffusion basin—the growth of only 10 cancerous cells would impact significantly (approximately by 12%) on the oxygen uptake of healthy cells, so that the tissue structure must necessarily undergo heavy changes. The transformation of the entire diffusion basin into a cluster of cancer cells (a *tumorlet*) leaves a lot (approximately 98% of the entire mass supported by the capillary) of necrotic cells and is responsible for heavy tissue hypoxia (Bauer et al., 2007).

The growth of the maximum sustainable amount of cancerous cells is necessarily associated with the migration of the colony first toward the capillary and then along it toward the arteriole orifice where the capillary is originated. In so doing most of the tissue (original healthy cells and older cancerous cells) undergoes necrosis, so that one can imagine different scenarios:

1. the growth continues without destroying the arteriole-capillary-venule local circulation system;
2. the growth proceeds destroying locally the local circulation system in such a way as to produce the necrosis of the entire region (with healthy and malignant cells); and
3. the growth continues destroying the local circulation system, but in a way that allows the inner region of the tumor to be fed directly by the arteriole.

The first case might be taken as characteristic of benignant tumors, which are able to destroy the entire hosting tissue but do not involve the entire organism (except for the primary damage they produce). The second case may perhaps be viewed as a spontaneous remission (self-healing) of the disease. The third case is responsible for the transformation of a local disease in a systemic disease essentially because of the acceleration of the tumor growth rate produced by the exposure to blood of larger and larger fraction of the tumor.

The development of the tumor occurs in a relatively quiet manner up to the *angiogenic switch* (Folkman, 1971), when the tumor promotes a neo-vascularization able to sustain its subsequent explosive growth and its transformation from a local disorder into a systemic diseases (Welter and Rieger, 2010). Until the angiogenic shift has occurred (that does not necessarily happens) the tumor is referred to as in a *dormant phase*.

The tumor is generally believed to remain in the dormant phase up to a size of $1\ \mathrm{mm}$; the cancerous cells contained in a volume of diameter 1 mm are expectedly 2×10^4 (this amount follows from the maximum number of cells that can be contained in $1\ \mathrm{mm}^3$, about 10^6, assuming that the tumor in the dormant phase is a complex network of tumorlets, each fed by one capillary, where 98% of the biological material is necrotic).

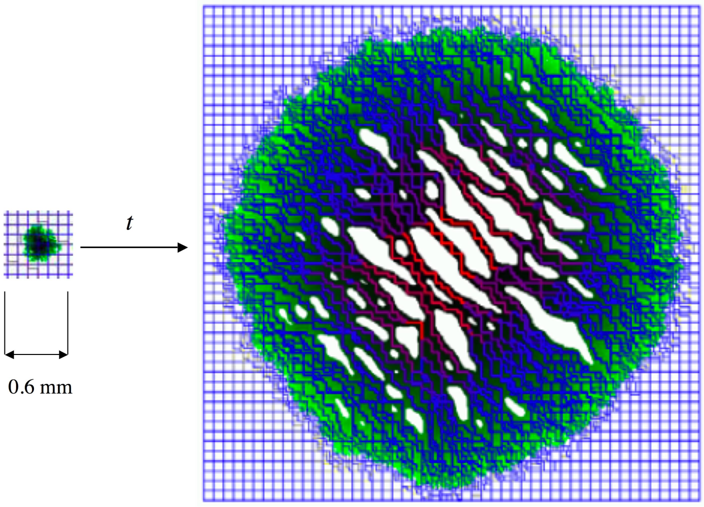


Figure 5.1: Dramatic morphological change of the solid tumor when its size switches from the submilllimetre (*left*) to the centimetre scale length (*right*). According to Bartha and Rieger (Bartha and Rieger, 2006), the capillary network is assumed to form a three-dimensional regular grid with step of $100\ \mu m$; all the remaining region is filled with cells. The tumor is characterized by proliferating cells, with the younger ones on its boundary and the oldest ones inside (increasing darkness indicating age). The tumor bulk is characterized by large regions (shown in white) of necrotic materials.

More difficult is to estimate the number of cancerous cells after the angiogenic switch; imaging is possible when the tumor has a larger size than $5\ mm$; a tissue with such a diameter could contain up to 10^8 cells; on another side, assuming that only 2% of the of the whole mass is formed by cancerous cells (as in the dormant phase) the tumor would contain 2×10^6 cells (see Fig. 5.1). I tentatively assume that a tumor at the imaging limit contains 10^7 cancerous cells.

What gives a hope in this picture is that in all stages the tumor is characterized by a large exposure to the blood stream so that it can be detected by the agents circulating in the blood. The above argument implies that a nanobot in a capillary could feel the metabolic pattern of the family of cells fed by the capillary itself, thus surveying the cells contained within a diffusion length.

5.3 General features of cancerous tissues—the clues

In view of the previous considerations, one fundamental issue of fight on cancer is its early diagnosis. How early it can be may be understood considering the phenomena occurring even in the early stages of cancer growth.

5.3.1 Hyperthermia

Needless to say the faster metabolism of cancerous cells and their accumulation at the arterial orifice generating the capillary is responsible for a local increase of temperature. This effect is so large that it is exploited not only in digital infrared thermal imaging to make diagnostic-quality images for breast cancer screening(Kennedy et al., 2009), but also hyperthermia is a type of cancer treatment in which the exposure to high temperatures (up to $45\,^\circ\mathrm{C}$) damages and kills cancer cells, usually with minimal injury to normal tissues. Hyperthermia may shrink tumors by killing cancer cells and damaging proteins and structures within cells (Luk et al., 1980; Steeves, 1992). Denoting with $\mathcal{C}_k$ a cluster of k cancerous cells, with a long right arrow a spontaneous and easy process and with a short left arrow a difficult and non-spontaneous process, the tumor growth can be sketched by means of the following quasi-equilibrium "reaction":

$$\mathcal{C}_k \underset{\leftarrow}{\longrightarrow} \mathcal{C}_{k+1} + \text{heat} \tag{5.1}$$

Irrespective of the detailed hyperthermia mechanism (increase of perfusion, permeability, pH, O_2 partial pressure, metabolic activity, or drug uptake), the inhibitory effect of heat on cancer can be understood in terms of Le Chatelier–Braun principle applied to Reaction (5.1)—as far as the disease produces a heat excess, imparting heat to the system depresses the evolution of the disease [3].

5.3.2 Hypoxia

Human solid tumors are considerably less well oxygenated than normal tissues.[4]. Hypoxia is essentially related to the fragility of the circulation system feeding the tumor. The collapse of weak portions of the tumor circulatory apparatus exposes the tissue to even hypoxic conditions and the tumor can tolerate such conditions only exploiting the glycolytic pathway[5]. In this pathway energy is produced by the transformation of glucose first to piruvic acid and then to lactic acid (rather than to CO_2 and H_2O). These acids by themselves are however unable to produce what is perhaps the most important effect on the nearby tissue—the pH reduction.

5.3.3 Acidity

The organism as a whole is a strongly buffered system: actually, it is able to regulate the pH of the blood to 7.40 ± 0.05. The tumor genesis produces a decrease of pH (as measured by microelectrodes and thus averaged over a distance non smaller that electrode diameter) by approximately 0.5 units (so that the H_3O^+ concentration increases by a factor of 3) (Tannock and Rotin, 1989).

---

[3] A naive justification could be grasped by considering that our organism defeats some diseases by elevating the body temperature

[4] Hypoxia leads to resistance to radiotherapy and anticancer chemotherapy, as well as predisposing to increased tumor metastases. However, it can be exploited in cancer treatment. One such strategy is to use drugs that are toxic only under hypoxic conditions, and the first drug of this class to enter clinical testing, tirapazamine, is showing considerable promise (Brown, 2000; Keith and Simon, 2007)

[5] Moreover, the character expressing the cell ability to survive in strongly hypoxic conditions can be selected during cancer evolution by the strong competitions among fast reproducing newly born malignant cells.

Assuming for the increase of tissue acidosis the same mechanism as in muscle during exercise[6] (Robergs et al., 2004), the increased acidity is not due to the anaerobic production of piruvic and lactic acid, but rather to the vivacious metabolic ATP hydrolysis.

The resulting high H_3O^+ production rate has dramatic effects: whereas the tumor cell succeeds in regulating the internal pH, the lower external pH on one side contributes (in combination with the proteolytic enzymes possibly produced by the cancerous cell) to destroy the nearby healthy cells and on the other side facilitates via Bohr's effect the loss of O_2 from the oxyhaemoglobin transported by red blood cells. In this way the fast metabolism of cancerous cells allows their aggression to the hosting tissue and facilitates oxygen uptake.

5.4 Diagnosis of cancer—smoking gun or frame evidence?

A lot of cytotoxic (i.e. toxic to cells) substances are known, either of natural origin or artificially synthesized. From the point of view of human pharmacology, what makes different one from another, distinguishing them in *poison* or *drug*, is their selectivity.

In this respect a drug is nothing but a poison able to destroy selectively guest or deviant organisms (from viruses to tissues or parasites) without excessive perturbation of the hosting organism. Of course, such a total selectivity is impossible (so that any drug has unavoidable side effects), but nonetheless the recognition of the pathogenic agent is a fundamental issue of any systemic (i.e., non surgical) treatment of the disease.

The selective recognition of specific antibodies occurring at the surface of cancer cells (the smoking gun), allowing their terminations using cytotoxic substances conjugated to ligands able to bind selectively to the antibodies, has ever been a *leit motif* of the fight on cancer. The following discussion leads however to conclude that this road is probably a *cul de sac*.

Cancer is certainly a disease of the elder, as confirmed by the power law of the dependence of its occurrence with age (Klawansky and Fox, 1984). However, it occurs even in children, and in those cases it is even more aggressive. In the light of evolution, this fact implies that *the advantage for the species, coming from the supposedly possible specialization of the immune system able to recognize and terminate cancer, is so small that the species has not payed the cost for such an evolution*. I see two reasons for that:

Genetic First of all, cancer is a disease characterized by a high heterogeneity, where each phenotype is able to develop autonomously (and often more fast) even when the others are destroyed.

Systemic The common features associated with cancer genesis and growth (hyperthermia, hypoxia and acidification) are the same as those characteristic of muscle under physical exercise. I propose this similitude as the reason why the immune systems does not recognize the tumor as a pathogenic state to terminate—because it is confused with a physiological state.

Getting success where evolution has failed is certainly a difficult task.

Leaving the mainstream of the fight on cancer, in the following I shall ignore the first case and concentrate on the way to overcome the limits posed by the possible confusion of cancer symptoms with the general features of muscle under exercise.

---

[6] I shall return later on this analogy

Before considering the details of our program, I however observe that the systemic confusion can also explain tumor angiogenesis. Although the detailed mechanisms of this process are not clear yet, taking as a model for angiogeneis the revascularization processes resulting from heavy physical exercise, one can indeed hypothesize that the angiogeneis is ultimately a *physiological* response of the organism to the increased demand of energy and O_2 (in this case by malignant cells).[7]

My program is thus that of *endowing the immune system with an artificial surveillance system devoted to detecting the simultaneous occurrence of hyperthermia, hypoxia and excessive acidity due to localized cancerous states without confusing it with the similar conditions produced under physiological conditions*.

In the light of this crime detection, the search of the perpetrator is not addressed to the identification of the smoking gun but rather to a frame evidence resulting from the simultaneous occurrence of three events and their persistence in time.

For that I imagine to use a swarm of agents, wandering through the organism along the blood stream, devoted to recognize hyperthermia, hypoxia and excessive acidity, resting in the region of detection to form a cluster so numerous to inform the hosting human of the presence of candidates to tumor.

Although the blood has been already pointed out as candidate medium for surveilling the organism, in the next section I'll compare it with two other possibilities: genes and cells.

5.5 Search of the appropriate level of surveillance — The scene of crime

The scientific theory characterizing the greatest progress in the second half of the 20th Century has certainly been the molecular biology of the gene. In extreme synthesis, this theory may be summarized in the Central Dogma of Biology: the deoxyribonucleic acids (DNAs) either make their copies (*replication*) or make ribonucleic acids (RNAs) (*transcription*), and RNAs make proteins (*translation*) (Watson et al., 2008). Taking into account that DNA duplication has a certain failure probability, the Central Dogma describes life as dominated by chance and necessity (Monod, 1970).

Molecular biology may collectively be viewed as a reductionistic theory of life where the healthy or ill state of any living organism is totally contained in its genetic code (DNAs) and the way it is expressed (proteins).

As far as the complete and continuous surveillance of the entire molecular pattern of an organism seems beyond any current possibility, the first problem is the identification of the appropriate surveillance level. In the following I shall consider three levels: genes, cells and blood.

---

[7] If this model were correct, cancer would be a source of paradoxes, with self healing due to necrosis, and cancerogenesis due to physiological response!

5.5.1 Gene level

The human genoma length in a single cell is about 2 m (Aerssens et al., 2001). Taking into account that the total number of cells in an adult organism is ca. 6×10^{13}[8], the total length of the human genetic material is around 1.2×10^{14} m. Even assuming the existence of a device

(A1) sufficiently non-invasive to probe the DNA without perturbing it,

(A2) able to 'read' each base in a negligible time,

(A3) able to manage the information, recognizing the correctness or not of the genetic code, and

(A4) moving along the sequence of bases at a speed of about 10 m s^{-1}[9],

reading the whole DNA with one detector would require about 10^{13} s ($\simeq 3 \times 10^5$ years). Controlling the entire genetic material in one year in the above hypothesis would require a system formed by 3×10^5 devices. Whether or not a system with this numerosity may be realistic, depends on the size of the constituting devices.

Although testing single molecular sites is currently possible (e.g., via scanning probe techniques), the apparatuses required for that have size on the metre length scale. Even assuming the ability to scale-down their size by factor of 10^2, the total system volume would have a lower limit of 3×10^5 cm^3 vs. a volume in the range $3 \times 10^3 - 1 \times 10^5$ cm^3 for *homo sapiens sapiens*.

Thus, although a system with so many devices is not manifestly absurd, the excessive optimism of assumptions (A1–A4) and the absence of any idea for scaling the technology of the probing techniques, discard the hypothesis that the appropriate scale for organism surveillance is the genetic one.

5.5.2 Cell level

Although there exist supramolecular complexes (like viruses, retroviruses, phages, prions, etc.) that share many of the properties collectively referring to living systems, there is a general consensus that the basic building block of life is the cell. Single cells are objects with diameter on a narrow length scale (typically in the interval $1 - 10$ μm for bacteria or $10 - 10^2$ μm for somatic eukaryotic cells) hosting a rich set of deeply interconnected physico-chemical, morphological, and functional properties.

Needless to say, the Central Dogma is too simplified and does not allow for the complicate feedback pathways regulating the duplication of DNA, its transcription into RNA, and the protein synthesis. In particular, the major physico-chemical property of living system—its *metabolism*—is not accounted for, even considering the control of the expressome (the set of

---

[8] A human body of 70 kg holds a volume of about 7×10^4 cm^3 and contains approximately 4×10^4 cm^3 of water (Guyton, 1976). The amount of extracellular water is about 1.5×10^4 cm^3 (Hansen and Koeppen, 2002). The complementary volume, 5.5×10^4 cm^3, is filled by the various kinds of cells. Assuming that each cell has an average volume of 10^3 μm^3 ($= 10^{-9}$ cm^3) the total number of cells would be 5.5×10^{13}.

[9] The speed of the hypothetical reading apparatus has been *heavily* overestimated from that of H$^+$ in water at room temperature (with mobility is about 4×10^{-7} m^{2}V^{-1}s^{-1}) assuming that it is moving under an electric field of 2.5×10^7 V m^{-1}, corresponding to a potential difference of 25 mV (the 'thermal voltage') over a distance of 10 Å.

molecules expressed by the Central Dogma) on the nucleome (the set of molecules carrying the information).

Proteins participate in all cellular functions. The function that allows the cell to be fed by the neg-entropy flow, necessary to preserve the complex and delicate cellular structure, is the enzymatic activity. The activity of enzymes, indeed, is at the origin of a diffusion field sustaining

- the flow of energy-rich substances from the environment into the cell for animal cells, or
- the flow of adequate substrates for the transformation of external energy (light) into energy-rich intermediate for vegetal cells, and
- the out-diffusion of waste materials for both.

The metabolic pathway of a cell does depend on its genome only through the way it is expressed, so that cells with different expressomes but the same genome usually have totally different metabolisms.

The net rates of anabolite destruction and catabolite production (localized by the membrane cell in the region it contains) generates a diffusion field. If the coupled reaction-diffusion equations describing the metabolism were completely known, it would be possible to predict the time-dependent concentration profiles of all species. Moreover, in spite of the fact that the problem is ill-posed (Engl et al., 2009), the knowledge of those equations would also allow the determination of the actual metabolite concentration profiles inside the cell provided that they are known outside the cell. In other words, if the standard ('healthy') metabolic pattern is otherwise known, an external sensor can assign a pathological state to a cell when its metabolic pattern deviates from the standard one by more than given confidence limits. The development of the model for assessing the health condition of the cell is beyond the scope of this work; for more details on this topic see (Engl et al., 2009; Cerofolini, 2010b).

Of course, to have a representative fingerprint of the cellular metabolic pattern, it must be mapped on a length scale smaller than the one characteristic of cell organelles and on a time scale consistent with life cycle. Reference (Cerofolini, 2010b) has shown that this goal is to some extent achievable by the current technology and fully consistent with the expected development of electronics in the post-Roadmap era (ITRS, 2007) and discussed how to exploit this property for guessing the standard metabolic pattern from *in vitro* experiments on tissue cultures.

To evaluate the health status of a cell from its metabolism, the duration of the surveillance cannot be a negligible fraction of the characteristic circadian time. Assuming optimistically a surveillance time of 10^3 s, in one year the nanobot would be able to test approximately 3×10^4 cells; in a tissue of 1 kg it will be able to explore approximately 3 parts per million of the whole tissue. The number of nanobots must be so chosen as to have an adequate statistical coverage of the tissue. An adult human body contains approximately 6×10^{13} cells. Assuming that each nanobot does really sense 3×10^4 cells per year, the exploration in one year of 0.1% of the whole cell population would require 2×10^6 nanobots. Assuming that each of them has a mass of 10^{-9} g (see later), the total mass of circulating nanobots should be of 2×10^{-3} g.

Although this strategy would provide an adequate statistical coverage without exceeding with invasiveness, it however suffers from a fundamental difficulty. Assume that an adequate chemical mapping is obtained detecting 10 metabolites in 10^2 regions (thus exploring the whole cell surface with a step of 10^3 nm). Such a map would imply the collection of 10^3 data. Assuming that each measurement requires 0.1 s (including positioning), mapping the entire cell would require approximately 10^2 s. The repetition of 10 mappings would thus require 10^3 s and their memorization should occupy $10 - 10^2$ kbit [10]. Mastering the numerous activities of the nanobot (sensing, recognizing of the chemical pattern, controlling the motion, managing the power, and so on) requires moreover a minimum of intelligence, so that the complexity of the circuitry cannot be smaller than 10^5 bit.

Any attempt to integrate such a logic in a miniaturized device currently makes sense only exploiting the complementary metal-oxide-semiconductor (CMOS) technology. Denoting with F the minimum producible feature size, the most dense architecture hitherto proposed for silicon integrated circuits (ICs) requires an area of $4F^2$ per cell of NAND flash memories. At the $F = 45$-nm node this implies a maximum density of 1.2×10^{10} cm^{-2} that should increase to 5×10^{10} cm^{-2} at the $F = 22$-nm node. Two other technological nodes have however been hypothesized for future CMOS ICs: the $F = 16$-nm node (allowing a maximum density of 1×10^{11} cm^{-2}) and eventually the $F = 11$-nm node (with maximum density of 2×10^{11} cm^{-2}) (Faggin, 2009).

At the ultimate node (likely achievable within 20 years) a complexity of 10^5 bit would occupy an area of approximately $1 - 2 \times 10^2$ $\mu\mathrm{m}^2$. This area, though comparable with that of somatic cells, is however too large to allow the nanobot to move inside the parenchima, where the interstitial water holds regions with size of about 1 $\mu\mathrm{m}$, thus making cell sensing useless.

5.5.3 Circulatory-system level

The organization of higher organisms is based on the specialization of the various organs to specific functions (locomotion, vision, hearing, digestion, etc.). This result is obtained specializing the expressome to the task, so that cells with the same genome are expressed within the same organism in very different forms.

All the functions of the organism are supported by organs requiring an adequate input of energy and neg-entropy; their needs are satisfied centralizing the collection of energy and neg-entropy and distributing them to the various organs through the circulatory system.

The circulatory system can be described as a complex tree forming a double canopy inside each organ. The apparatus is so built that energy-rich substances and O_2 can escape from it (thus feeding nearby cells) only in the canopy. This is generally possible only associating with each branch of the canopy a basin where the anabolites are distributed and the catabolites are collected and eventually delivered to to the secretion apparatus. This basin, where the transport of matter is dominated by diffusion, will be referred to as *diffusion basin.* The union of all diffusion basins must embed all the living parts of the organism [11]. In other words, the entire organism admits a kind of Voronoi tessellation where each basin forming the tessellation (V-basin) is constituted by all cells prevailingly fed by the same capillary.

The above argument implies that a nanobot in a capillary could feel the metabolic pattern

---

[10] The lower limit is taken assuming that the information is classified with a single bit (0, healthy; 1, ill); while for some metabolites (like the prostate specific antigen or transaminase) this coding has a meaning, metabolites related to circadian rhythm require a more accurate information (estimated in 10 bits).

[11] This structure could work as described, but in higher organisms it is complicated and (in view of the key role of evolution on structure-function relationship) made more efficient by another distributed system, devoted to increase the collection efficiency catabolites and other wastes—the lymphatic system

of the family of cells fed by the capillary itself, thus surveying the cells contained within a diffusion length.

The total number of capillaries can be estimated from the blood velocities in, and from the cross sections of, aorta and capillaries. Let A_1 and v_1 denote the aorta cross section and blood velocity therein, assume that the circulatory system may be modeled as a tree with branching ratio of 2[12], and denote with A_n and v_n the corresponding quantities at the n-th braching; if the blood circulatory system is a nearly closed system, one has

$$A_1 v_1 = 2^{n-1} A_n v_n.$$

Specializing this equality to the capillaries ($n = \overline{n}$), one gets

$$2^{\overline{n}-1} = A_1 v_1 / A_{\overline{n}} v_{\overline{n}}. \tag{5.2}$$

The determination of $\overline{n}$ requires thus a knowledge of A_1, v_1, $A_{\overline{n}}$, and $v_{\overline{n}}$. For the aorta $A_1 \simeq 10$ cm^2 and $v_1 \simeq 30$ cm s^{-1}. Less trivial is to specify $A_{\overline{n}}$ and $v_{\overline{n}}$. Both these quantities are subjected to large uncertainties. Assuming that the cross-section of capillaries is circular, one has $A_{\overline{n}} = \frac{\pi}{4} d_{\overline{n}}$, with $d_{\overline{n}}$ being the capillary diameter. The diameter of the capillary lumen is strictly related to the red blood cell (RBC) diameter, that on the mean values $7.8\ \mu\mathrm{m}$ (varying with its age). In fact, $d_{\overline{n}}$ must be sufficiently large to allow the passage of the smallest RBCs because otherwise the tissue would undergo necrosis; taking into account that RBCs may bent without destruction to a diameter of ca. $2.7\ \mu\mathrm{m}$ (Freitas, 1999), one gets

$$d_{\overline{n}} \gtrsim 3\ \mu\mathrm{m}. \tag{5.3}$$

On another side, the efficiency of oxygen exchange would greatly be reduced if the RBC traveled along the capillary without impacting the walls; evolution has thus selected capillaries with $d_{\overline{n}} < 12\ \mu\mathrm{m}$. The capillary diameter for mammalians is reported to be about $5\ \mu\mathrm{m}$ (Vink and Duling, 1996), whereas an interval of $5 - 8\ \mu\mathrm{m}$ is reported for humans (Chandran et al., 2007). To be concrete I shall assume $d_{\overline{n}} = 8\ \mu\mathrm{m}$, so that $A_{\overline{n}} = 5 \times 10^{-7}$ cm^2. The blood velocity in capillaries is related to their diameter; in human skin capillaries located perpendicularly to the skin surface the $v_{\overline{n}}$ was measured to be around 5×10^{-2} cm s^{-1} (Stücker et al., 1996), whereas the RBC velocity within the capillary network of the brain undergoes rapid fluctuations and manifests spatial heterogeneity ($5 - 18 \times 10^{-2}$ cm s^{-1}) (Hudetz, 1997). To be concrete, in Eq. (5.2) I shall assume $v_{\overline{n}} = 2 \times 10^{-2}$ cm s^{-1}. With the above values Eq. (5.2) gives $\overline{n} \simeq 352$.

If the entire organism contains indeed 6×10^{13} cells, each of the $2^{\overline{n}}$ ($\simeq 3 \times 10^{10}$) capillaries feeds in the mean about 2×10^3 cells. Of course, this value is a mean value resulting from the combination of the feeding ratios of cells with fast metabolism (like neurons, for which a feeding ratio of 1 may be assumed) with those of cells with slow metabolism (like bone cells).

5.6 Swarms of nanobots—detectives

The basic idea for the early detection of tumors is to inject into the circulatory system a swarm of 'detectives' able to identify the presence in the organism of malignant cells, even if they are just a few or, ideally, just one.

---

[12] Chapter 6 will be devoted to the fractal modeling of the circulatory system.

On another side, in certain cases it is possible to detect endogenous diseases at very early stages via chemical analysis or imaging. In fact, the detection of a single molecule is at the reach of mass spectrometry (Dunn, 2008), whereas the localization of a pathologic tissue on the length scale of 1 mm (associated with a cluster of only 10^5 cells) is already achieved. These facts cast serious doubts on the usefulness of an artificial system distributed inside the organism. To support our program I however note the following facts:

- A typical blood analysis addressed to the search of a pathological marker involves a volume ω of approximately 1 cm^3 vs. a total blood volume Ω of 5×10^3 cm^3, sampled in a region generally very distant from that where the marker was injected in the circulatory system. Denoting with N the total number of marker molecules released in the blood by the pathological tissue and assuming Poisson statistics, the probability p that the sampled volume ω contains at least 1 marker molecule is given by

$$\begin{aligned} p &= 1 - \exp(-N\omega/\Omega) \\ &\sim N\omega/\Omega \qquad\qquad \text{for } N\omega/\Omega \ll 1 \end{aligned} \tag{5.4}$$

  Thus, unless N is comparable with Ω/ω ($\simeq 5 \times 10^3$), Eq. (5.4) states that the probability of detecting a single molecule (even assuming its technical feasibility (Dunn, 2008)) is negligibly small.

- Both blood analysis and systematic imaging can realistically be carried out with a frequency (of $1 - 2$ checks per year) that may be too low to avoid the transformation of a small pathologic tissue in a serious systemic disease.

Both these arguments support the usefulness of a *distributed surveillance system* (DSS), i.e. a distributed system addressed to the continuous surveillance of the organism, with a chemical sensitivity on the single molecule scale, and able to explore the entire organism through the circulatory system.

A detailed definition of a DSS is still missing. Of course, this goal is extremely ambitious and is expected to require several decades. Nonetheless, sketching a scenario is not a mere speculation, but rather is useful to identify the nature of the problems posed by the definition of a DSS.

The surveillance system considered in this work is constituted by two parts:

- A *central unit*, externally accessible but permanently resident in the organism (e.g., as an earring)

- and a *swarm of nanobots*.

In (Cerofolini, 2010b) it was hypothesized that each nanobot is a self-propelled machine, taking energy from the environment, able to recognize and dock the target cell, to sense its membrane and neighborhood, to recognize its health state, to store the information, to transfer it to the central unit, and eventually (once allowed) to destroy the malignant cell. Today a swarm so done (actually an *auxiliary immune system*) seems beyond current possibilities; I now believe that something similar can be achieved specializing the agents of the swarm to diagnosis (*scouts*) or to therapy (*workers*). A roadmap for nanobot diagnostics can since now be defined; more difficult is instead to imagine a general framework for the use of nanobots in therapy.

A scout is nothing but a circulating nanolaboratory for blood analysis *in situ*. The swarm may in turn be formed by sub-swarms, each addressed to set of metabolites characteristic of the target tissue. The population of the various sub-swarms must be tuned to have an optimal surveillance of the organism.

The malignant cell identification task can be decomposed in two different aims.

1. The first aim is the effective location of the malignant cells,
2. the second aim is the reporting of this location outside the body (e.g., to an external unit).

Every living region of the body can become the origin of cancer. This implies that in principle the agent must be able to explore the entire body. Actually all the living part of the body are at a distance of a nearby capillary by at most an O_2 decay length; in other words, an agent moving in the body transported by blood is in principle able to feel the metabolic field of the body averaged over a distance of about $50\ \mu\mathrm{m}$. Taking into account that such an agent can explore about one capillary per minute, the time required for exploring the entire organism would be of approximately 2×10^{12} s, exceeding the human lifetime by approximately 3 orders of magnitude. Such an exploration strategy requires thus a swarm of agents, whose population is determined by the mean number of necessary visits per region. For example, assuming that each region requires one visit per month, the total number of agents would be approximately 7×10^5. In Chapter 7 this point will be further discussed.

To be transported by the stream along the entire system, the agent must be so small as not to obstruct the capillary. This condition gives a criterion about size, which cannot be much larger than that of a blood red cell: having already in mind a production technology, I assume that the area of the agent is on the scale of $10^2\ \mu\mathrm{m}^2$. Of course, the shape should in a sense be the opposite one to that of the red blood cell (whose lenticular shape is matched just to occlude the capillary and to bend under the stream pressure to undergo striction and to release as much O_2 as possible). The agent should thus have an ellipsoidal shape with major axis much longer than its minor axis.

The agent must perform a lot of sophisticated functions like navigation, recognition, data collection and transmission so that it must be viewed as a robot; on another side, the limited area available to host them implies that the necessary intelligence and sensing elements are implemented in nanoscale devices. The agent must thus be viewed as a nanobot.

5.6.1 Size and logic

In view of their limited size, nanobots have necessarily limited computational resources. The maximum density of devices that can be hosted in a *planar* circuit is given by $1/4F^2$ where F is the minimum producible feature size (the so called $4F^2$ architecture). At the forefront of the technology $F = 40$ nm, so that a planar surface of area $10^2\ \mu\mathrm{m}^2$ could host 1.6×10^4 devices (and thus be endowed with a logic of 16 kbits. At the end of the Roadmap (say, within the next 10 years)(ITRS, 2009) F should be reduced to 10 nm and the logic could be formed by 250 kbits. This value is likely achievable within the frame of the current technological paradigm as well from technological shifts like that based on the crossbar structure(Heath et al., 1998a; Snider et al., 2005; Cerofolini and Romano, 2008). The shift to molecular electronics would however facilitate the integration of sensing devices onto the nanobot (Cerofolini, 2010a; Cerofolini, 2010b). Further increase of density with planar arrangements seems to be in contrast with physical laws; however, if the crossbar structure is built with out-of-plane wire arrays, terascale integration seems possible(Cerofolini et al., 2010c; Cerofolini et al., 2011b) so that the logics hosted in a 10^2 $\mu\mathrm{m}^2$ nanobot could have a megabit complexity.

Just to have an idea of the computational possibility of such a device, I remind that the simplest natural living system that can survive on a well defined chemical medium is *Mycoplasma genitalium*. Mycoplasma genitalium has the smallest genome of any organism that can be grown in pure culture and its genomic complexity is just around 1 Mbit (Glass et al., 2006). This comparison suggests that the hypothesized nanobot can likely perform sophisticated jobs provided that it is endowed with suitable hardware for sensing and actuating and sophisticated algorithm to manage them.

At the present stage of knowledge, the hypothesized swarm is certainly far from being producible, but it is not an (irrational) dream because most of the critical steps required for its preparation have already been established.

5.6.2 Shape

On another side, to be able to explore all capillaries avoiding the obstruction of the capillary lumen and the formation of blood clots, the scout must be suitably shaped and have some degree of flexibility (for this point see 5.7.4). An ellipsoidal disk with major axis $2a$ of 50 $\mu\mathrm{m}$, minor axis $2b$ of 2.5 $\mu\mathrm{m}$ and height h of 1 $\mu\mathrm{m}$ is consistent with condition (5.3), could host 10^5 bits, and could be produced via planar technology building the device on silicon-on-insulator (SOI) substrate removing the substrate at the end of process[13].

The avoid that in the blood stream the nanobot undergoes rotation (that in arterioles, venules and capillaries might be harmful) it must also be endowed with suitably designed hydrodynamic appendages. When immersed in fluid in laminar flows, discs shaped with low aspect ratio are known to undergo lateral drifts toward the vessel wall (Lee et al., 2009) and this phenomenon can be exploited for their docking thereon.

Assuming for a while that such a chip (sketched in Fig. 5.2) can actually be built, they have shape and size allowing them to explore the whole organism through the circulatory, lymphatic, or secretion systems.

5.6.3 Sensors

Imagine for a while that specialized portions of the electronic part of the IC (the gates of selected transistors) are in contact with the external world and so functionalized that the adsorption of a specific molecule affects the transconductance of the underlying field-effect transistor (FET) by an amount so large to allow the detection of the phenomenon. In that case the considered transistors could be used as sensors.

The integration of computation and chemical sensing in the same circuit, however, is not easy. In fact, logics and memories are built in such a way as to decouple as far as possible the electronic device from the outer world. The basic constituents of the electronic circuit, the FETs, are usually buried beneath approximately $10 - 20$ dielectric and metallic layers of total thickness of a few micrometers; if the transistor is produced via the silicon-gate technology, the exposure to the outer world of the FET core—the gate oxide—is even conceptually impossible.

---

[13] Possible advantages of discoidal particles instead of spherical ones have been discussed in (Gentile et al., 2008)

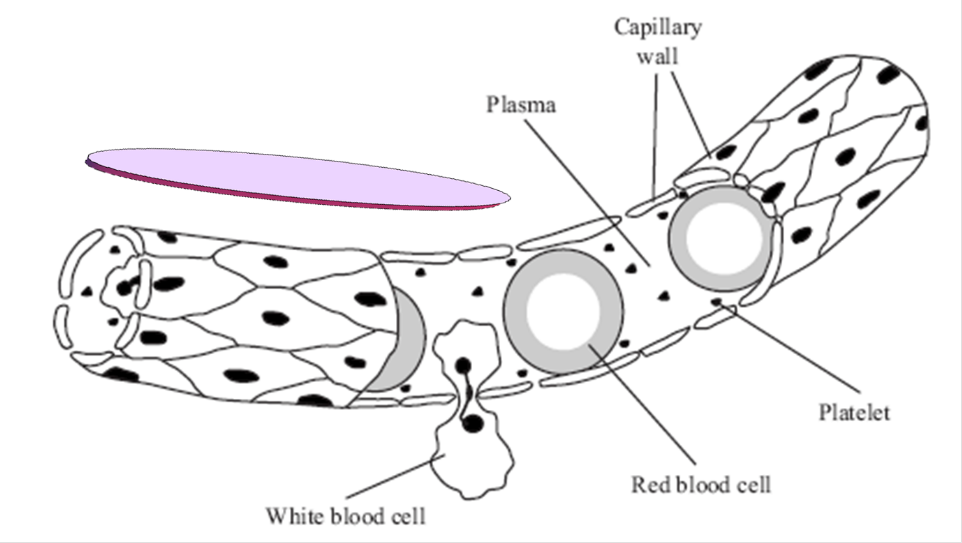


Figure 5.2: Pictorial view of a nanobot compared to a capillary (roughly on scale)

On another side, a new paradigm based on the crossbar structure has been proposed for the post-CMOS era. The crossbar allows a cross-point density even larger than the one achievable by CMOS FETs at the end of Roadmap, may be embedded in standard CMOS circuitry, can be built in the back-end stage of the fabrication process (Cerofolini et al., 2005a; Cerofolini et al., 2008a; Cerofolini and Romano, 2008; Cerofolini, 2010a), and can thus be easily put in contact with the environment. The crossbar architecture and technology are described in Chapter 8.

Scaling considerations suggest that one elementary charge is able to switch ON-OFF or *vice versa* an underlying transistor of area 10×10 nm^2. In fact, in current flash-memory technology, the information is recognized digitally, thus with almost 100% confidence, storing approximately 100 electrons in a floating gate of area 60×60 nm^2.

Reference (Cerofolini, 2010b) suggested a route for the crossbar functionalization via the growth of carbon nanotubes (CNTs) with a pattern consistent with that of the underlying crossbar. The hybridization of a nanoscopic crossbar with CNTs could thus allow the nanobot circuitry to sense the surrounding world provided that the CNTs are functionlized to sense the blood.

The use of CNTs as sensors requires their functionalization. Because of their poor reactivity, functionalizing CNTs is notoriously difficult. However, since CNTs are usually formed via catalytic growth on pre-deposited metal islands thatt are continuously segregated at the top of the growing CNT, the metal clusters can be exploited as Trojan horses for the functionalization (Cerofolini, 2010b).

The choice of the functionalizing agent is a delicate matter. Two major approaches seem possible, both made possible by the fact that, due to its tiny capacitance, the conductance state of a cross-point is controlled by really few (perhaps only 1) ions adsorbed on the nanotube [14].

The first approach is addressed to the detection of ionic species. For that the CNT (or, more easily, its metal cap) is derivatized with a molecule with a very large specificity toward the

---

[14] That a single elementary charge may be able to switch ON-OFF or *vice versa* an underlying transistor of area 10×10 nm^2 follows from scaling considerations: in current flash-memory technology, the information is recognized digitally, thus with almost 100% confidence, storing approximately 100 electrons in a floating gate of area 60×60 nm^2.

considered ion; see (Cerofolini, 2010b) for a few examples.

More difficult is the sensing of neutral species: for them, indeed, one must hypothesize that the CNT has been derivatized with suitable receptors reacting redox with the target species donating an electron to it (or *vice versa*). To allow the continuous operation of the electrode, however, the receptor must contain a sacrificial region that restores its pristine redox state with a longer time constant than the time required for the detection of the signal.

5.6.4 Power supply

Supplying power to the DSS is not trivial and most likely requires different solutions at different levels. The power required for the central unit is macroscopic and may be supplied either by external sources (batteries) or by implanted generators (stochastic microelectromechanical generators). More difficult is to supply power to the nanobots that, in view of their tiny size, can hardly tolerate such features.

In the absence of large source of external power, hybrid solutions must necessarily be found for fuelling and propelling nanobots. First of all, I note that while propulsion is most likely a fundamental function for workers, it is not important for scouts because they move with the blood stream. Rather, one can take advantage of the attempts to endow microdevices with biomotors, observing that *biomotors are reversible and can operate as engines too*. An enormous advantage of this solution is the fact that the motion does not require an alien input of energy—for that the chemical energy circulating with the blood in the organism (in the form of sugars and O_2) can be exploited.

Among the examples where the motion is imparted by the derivatization with biomotors I mention the following: Montemagno and Bachand have reported the construction of nanomechanical devices powered by biomolecular motors (Montemagno and Bachand, 1999); Kim and Breuer have described the successful use of live bacteria as mechanical actuators in microfabricated fluid systems (Kim and Breuer, 2008); and Behkam and Sitti have exploited bacterial flagella for propulsion and motion control of microscale objects (Behkam and Sitti, 2007).

5.6.5 Biomimetic coating

Biomimetic coatings designed to not activate immune response of the organism are at the reach of the current technology, as demonstrated by several demonstrators of lipid monolayers or bilayers (mimicking biological membranes) supported on solid or polymer surfaces on even large areas (Tanaka and Sackmann, 2005). Of course, the supported films are only weakly bound and are thus poorly stable; a larger stability, however, can likely be achieved by bonding covalently (e.g., via silanization) to the sensor surface molecules with carboxylic terminations, mimicking the outer surface of cells.

Alternatively, another route involving the hybridization of the nanobot with the membrane of red cells seems at the reach of near-future technology.

5.7 Nanobot tasks

As already mentioned, nanobots, due to their size, are necessarily devices with very limited computational and energetic resources, and able to have only strictly local interaction with the other nanobots. Consequently, it is not feasible to fancy that a single nanobot has on board the components and the energy to transmit a signal outside the body. However, although at first sight these facts seem to make the challenges really daunting, from nature we have experiences of tiny living beings (like ants or bees) with very limited resources but nonetheless able to perform complex tasks. In fact a task which cannot be tackled by a single individual, can be tackled by a swarm of individuals without the need for a central supervisor. This is the very idea at the base of *swarm intelligence* (Bonabeau et al., 1999), and its application for the control of swarm of micro- and nano-bots for medical applications have already been presented in literature (Martel and Mohammadi, 2010; Nagy et al., 2009; Requicha, 2003).

The first task of the scouts described in the previous section is to locate diseased cells, and to report their positions to a central unit outside the body. Due to the limited resources of the nanobots, direct wireless transmission of data to the central unit is, at present, a seemingly insurmountable barrier (Cerofolini et al., 2010a). Other (likely more efficient) organizations can however help in solving the problem. For instance, (Cerofolini, 2010b) hypothesized that in each squad the information is *transported* and reversed to a nearby stably implanted *local unit* where it is elaborated and *transmitted* via radio frequency to the central unit for the final elaboration in the light of the inputs coming from the other local units. After that, the relevant information is eventually transmitted to the external world.

These solutions pose, in any case, dramatic technological problems. While the passive transport through the circulatory, lymphatic or secretion apparatuses requires only passive appendages and docking may be driven by physical chemistry (through the conjugation of the nanobot to suitable specific receptors), localizing the targeted cell and returning to the local unit requires a sophisticated navigation system providing the nanobot with its coordinates.

However sophisticated might the algorithms be, a logic of 10^5 bits at most can hardly be considered sufficient for all the requested functions. Increasing further the bit density requires technological breakthroughs — like the adoption of crossbar technology (see Chapter 8).

The problem of reporting information outside the body could be solved by clustering enough nanobots around diseased cells. In fact, if this cluster is sufficiently large, it can trivially be imaged via x-ray computerized axial tomography (CAT). A more sophisticated application exploits the fact that the cluster may become large enough to behave as an antenna, able to send an electromagnetic pulse to the central units — the transmission of radiation in the millimeter band would require the clustering of ~ 10 nanobots only. A related problem (data collection via a swarm) has been discussed in (Winfield, 2000). To accomplish this task, the nanobots need to autonomously disperse in the capillary bed, take chemical sensor reading, mark the region where a positive signal is detected, and form a cluster in that region.

To accomplish the identification task the swarm of nanobots needs to accomplish the following sub-tasks:

1. autonomous dispersion in the capillary bed,
2. chemical sensor reading,
3. marking of the region where a positive signal is detected,
4. clustering in the marked region.

Once a cluster of a few nanobots formed around a malignant cell, it could be big enough to be imaged via computerised x-ray axial tomography or have enough energy to behave as antenna able to transmit a signal. The authors described a surveillance system together with a strategy to tackle the identification task in (Cerofolini et al., 2010a; Amato et al., 2011), here I focus on a general description of the most complex of these sub-tasks — the clustering around malignant cells.

5.7.1 Finding the diseased cells

According to the described strategy, when the nanobot enters a capillary it has to detect the temperature of the medium and the amount of lactic acid; in the case the nanobot feels values that can be considered symptoms of cancer actuates appendages for docking to nearby wall where remains anchored until the local pH is sufficiently low.

First of all, consider the time spent by the nanobot in the capillary, about 2 s, as follows from the length of capillary, $8 \times 10^{-2}\ \mathrm{cm}$ and the blood velocity therein, $5 \times 10^{-2}\ \mathrm{cm\,s^{-1}}$. This time is manifestly long enough to allow sufficiently accurate measurements.

The basic strategy is based on the docking of the nanobot on nearby capillary walls when it feels an abnormal increase of temperature and the simultaneous presence of piruvic acid and a high concentration of H_3O^+.

For the detection of the ambient temperature, I hypothesize two possible ways: (*i*) In the first approach, the nanobot is triggered by transient increase of temperature as measured by a built-in Seebeck thermocouple, whose sensing junction is exposed to the environment (and thus feeling the capillary temperature) while its reference junction is thermally isolated (and thus close to the temperature in the arteriole generating the capillary)[15]; (*ii*) in the second case the temperature is periodically detected under near equilibrium conditions by measuring the resistance of a semiconducting wire exposed to the environment.

Once allowed by the change of temperature, the nanobot looks for chemical signatures of the tumor. According to the general idea described in this work, the simultaneous presence of piruvate and hydronium ions can be assumed an adequate clue. Extracting the information, however, is not easy: the general scheme here proposed involves the gentle derivatization of crossbars (as described in (Cerofolini, 2010b)) with nicotinamide dinucleotide (NAD)[16] and the functionalization in its vicinity with lactate dehydrogenase (LDH). In this way, the following redox reactions are expected to occur very efficiently

$$\bullet\cdots\mathrm{NAD} + \mathrm{CH_3COCO_2H} + \mathrm{H^+} \overset{\mathrm{LDH}}{\rightleftharpoons} \bullet\cdots\mathrm{NAD^+} + \mathrm{CH_3CH(OH)CO_2H} \qquad (5.5)$$

where $\bullet\cdots\mathrm{NAD}$ denotes the derivatized crossbar. The higher is the concentration of piruvic acid and the acidity, the larger the reaction shift toward the lhs. The positive charge on the grafted NAD molecule will produce a positive voltage on the upper side of the underlying crossbar, and this signal can be taken as a marker for the simultaneous detection of $\mathrm{CH_3COCO_2H}$ and H^+.

---

[15] Remarkably enough, the silicon nanowires used for the definition of the vertical crossbar can be used as highly efficient Seebeck generators (Cerofolini et al., 2010b).

[16] 'Gentle' means that that the derivatization preserves the redox properties of NAD; in particular, the gentle derivatization should not interfere with Reaction (5.5).

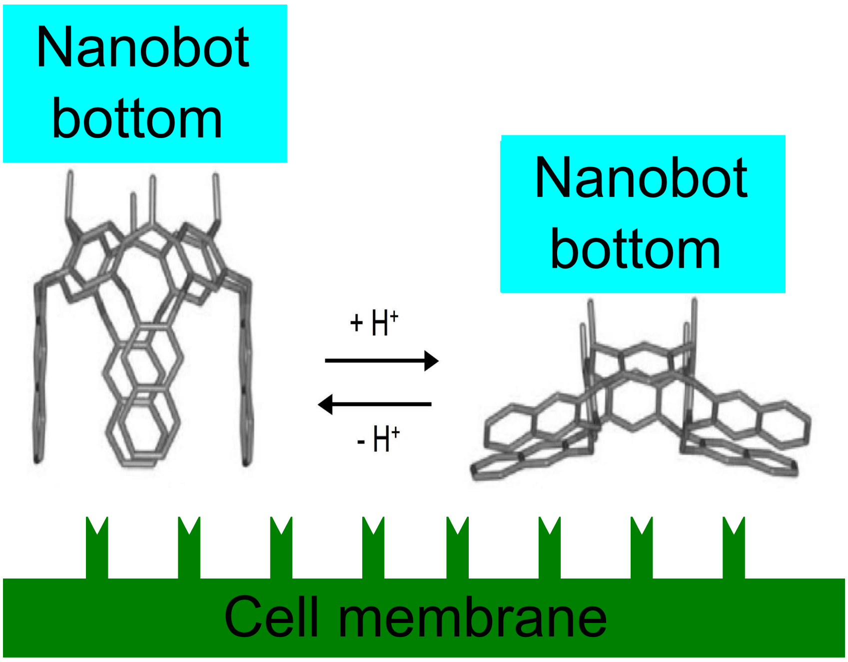


Figure 5.3: Conformational changes, resulting after proton capture on quinoxaline cavitands, allowing the nanobot docking on nearby capillary walls in the presence of piruvic acid and high acidity

5.7.2 Anchoring in the proximity of the diseased cells

Signalling the presence of the pathogenic chemical state requires that the nanobots behave collectively as a swarm. The collective behavior, in turn, is started from the docking of the nanobot to a nearby wall. That is achieved derivatizing the bottom of the nanobot with quinoxaline cavitands, as shown in Fig. 5.3. Due to the hydrophobic character in cave conformation, the termination will prevent the docking to epithelial cells. Imagine, however that once a marker molecule is detected the nanobot produces electrochemically (via electrolysis of water) H^+ and inject them at its bottom. Their capture by the nitrogen atoms will cause the opening of the cavity in the kite position, due to the Coulomb repulsion strength among positively charged nitrogen atoms. The protonated nitrogen atoms will then be attracted the negatively charged sites on the cell membrane forming the glue for the attachment of the nanobot to a nearby cell forming the capillary wall.

After docking, electrolysis is turned-off and the cave-kite conformation is eventually controlled by the ambient pH. If the initial low pH that produced the nanobot anchorage was produced by physiological reasons, the pH recovery to 7.4 will produce the detachment of the nanobot, while it will remain anchored to the wall until the low pH condition persists, Of course, such a behaviour is possible only via an extremely accurate tuning of the basic strength of proton-hosting site.

5.7.3 Clustering

In swarm robotics, spatial coordination between robots is often critical. When the robots are macroscopic, usually this coordination is achieved via local relative positioning sensors which are based, for example, on ultrasound or infrared technologies. Thus, by using this sensors, nearby robots can communicate and determine the bearing, orientation, and range of their neighbors.

In order to maintain a similar communication scheme even between nanobots, Ref. (Cavalcanti et al., 2006) proposed that each nanobot stores specific chemicals to be released for detection by other nanobots. However, another scheme not requiring these chemicals and their related mechanism can be proposed by exploiting a property of the microscopic world: the fact that *collisions are not an issue*. Of course, this is very different from what happens for macroscopic robots, where collision avoidance is a major issue to be taken into account. *Vice versa*, communication between nanobots can happen through direct physical clashes. In the following the focus is on this last approach which, to the best of our knowledge, has not been investigated yet.

communication-through-collision. A nanobot able to form a cluster has two main characteristics. First, it has terminations binding selectively to cells; these are the anchoring mechanism described in section 5.7.2. Second, it is endowed with two appendages (an *head* an a *tail*) able to bind to one another via weak non covalent forces, and disallowed to do that within the same nanobot by the robust constraints. The se two characteristics are depicted in Fig. 5.4.

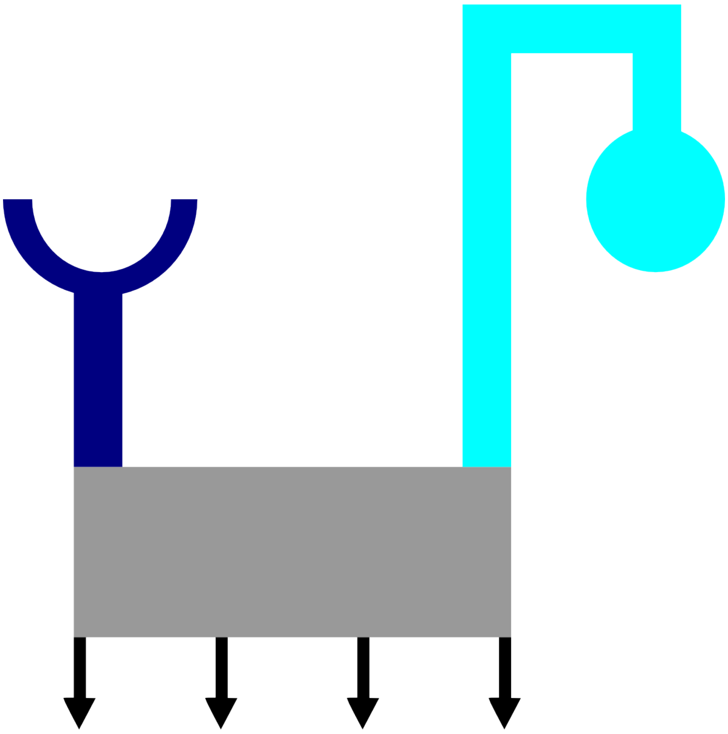


Figure 5.4: Conceptual scheme of a nanobot with appendages that allow the adhesion to cell membrane and the formation of clusters

It can be hypothesized that when two nanobots collide in liquid phase (say in blood), they remain paired for a short time. Although this can lead to the random formation of a cluster, however it may easily be destroyed by the thermal reservoir (Fig. 5.5). On the other side, clusters may be stabilized when they are formed on the membrane of a cell (Fig. 5.6), so that nanobots anchoring on unealthy cells become themselves sites for docking of other nanobots. I

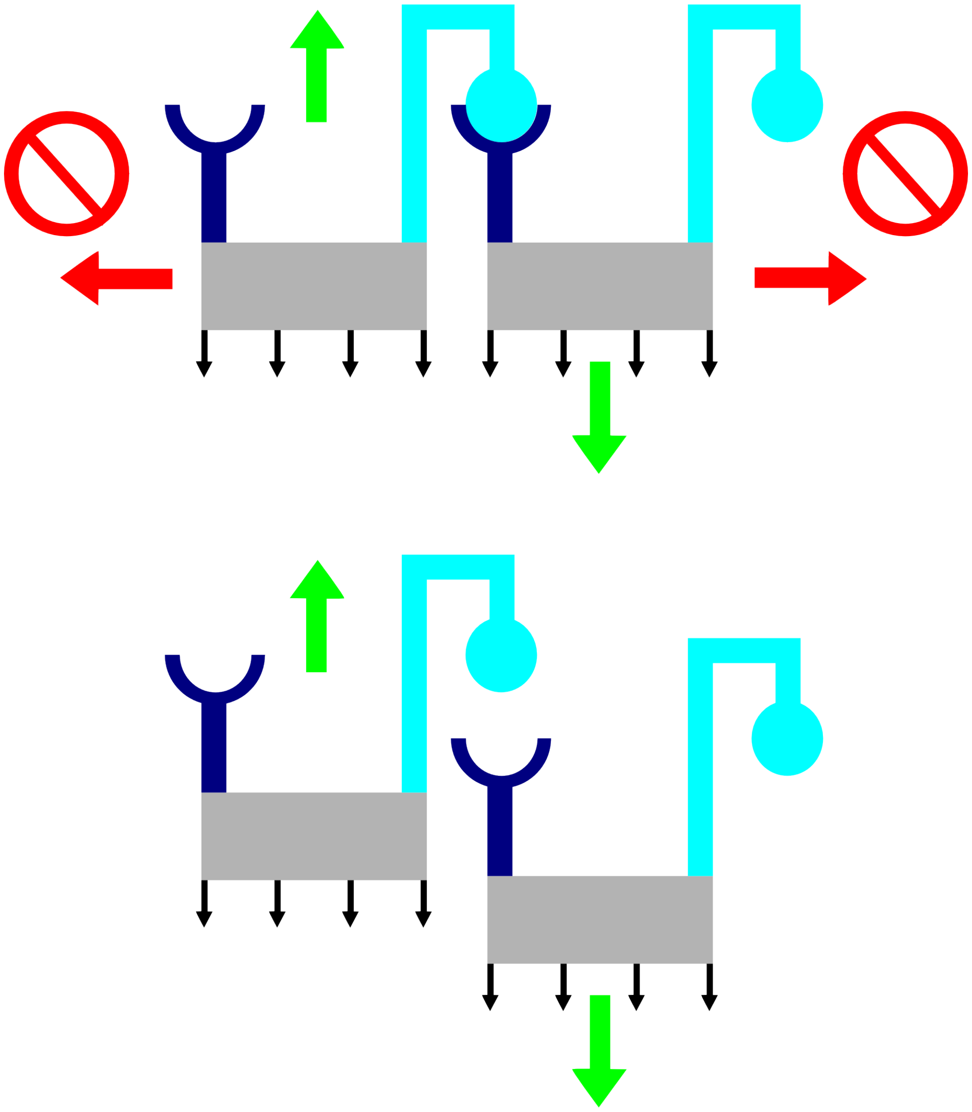


Figure 5.5: (*Top*) A cluster of two nanobots and the forces acting for its decomposition. (*Bottom*) The decomposition of the cluster is possible only along selected degrees of freedom.

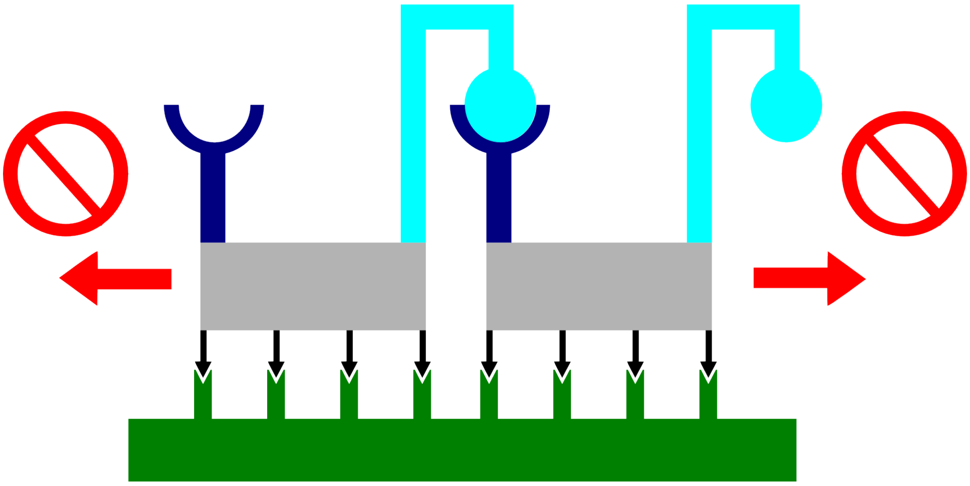


Figure 5.6: Freezing of the degrees of freedom allowing the decomposition of adsorbed clusters

call this mechanism *self-docking*. It has also the advantage of accelerating the formation of the cluster.

If each nanobot contains a metal wire connecting one docking side to the other, the formation of a cluster of N nanobots will result in a wire of length $2aN$. For the considered nanobot shape, a cluster with $N = 10$ will have a continuous wire of size of fraction of millimeter, thus able to transmit a signal to the central unit, for instance via an electromagnetic pulse obtained by discharge of built-in capacitors[17].

Actually, the pictorial sketches in Figures 5.4–5.6, though useful to illustrate the mechanism of cluster formation, are not representative of the actual technological solution. A realistic implementation of the nanobot head and tail could simply employ highly hydrophobic thiol-terminated noble metals. If they are suitably shaped (as in the example shown in Figure 5.7), they self-assemble spontaneously. Once the cluster is large enough (see Figure 5.8), it may be detectable via either external imaging or transmission of an electromagnetic pulse to the central unit. Of course, interface tension of water-hydrophobic region, adhesion energy to the tissue,

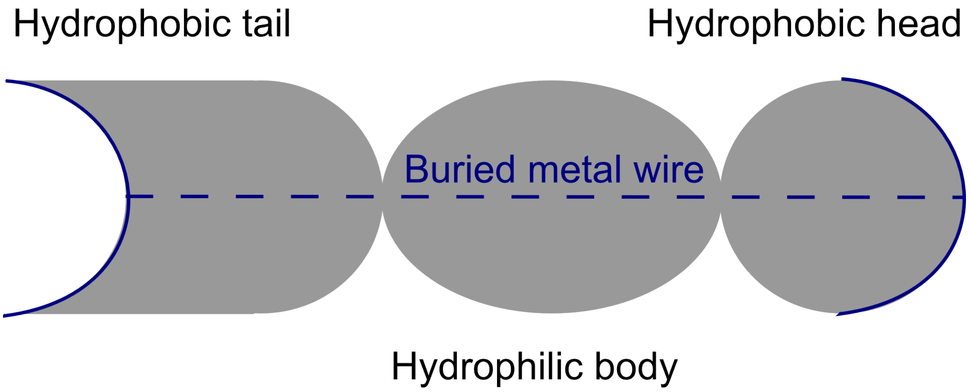


Figure 5.7: Plan view of a nanobot terminated, shaped to allow the spontaneous formation of a detectable cluster, and composed of smaller modules

---

[17] The central unit, in turn, will inform the person (or her or his physician) via cellular phone.

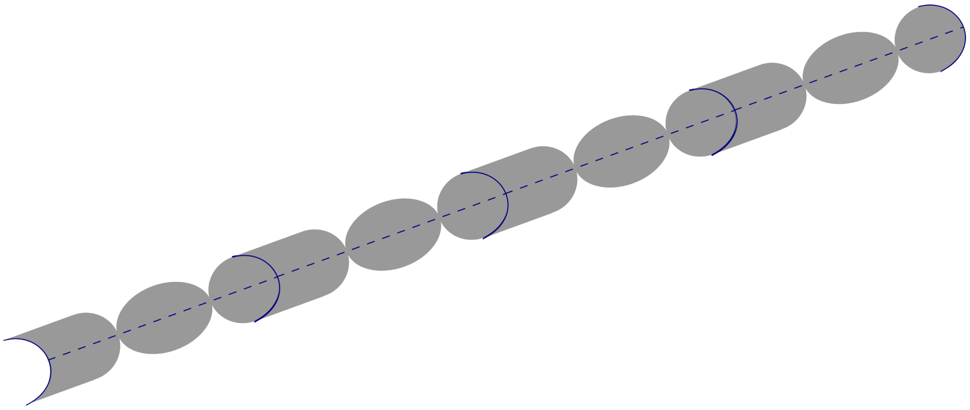


Figure 5.8: Nanobot cluster for labelling pathological tissues

and shape of the nanobot must be carefully designed to allow the formation of a detectable cluster.

For a swarm of $\simeq 5 \times 10^6$ agents, a nanobot anchored to a capillary wall will on the mean interact with another nanobot each 3 days, so that a cluster of 10 nanobots, sufficiently large to inform the central unit, will be formed in approximately one month. It is however noted that the above consideration applies to a situation where almost all nanobots are circulating. This occurrence is related to the fact that only few capillaries are decorated with a nanorobot. Quite paradoxically, the considered detection strategy works only in the early stage of the pathology and becomes progressively less effective while the disease progresses. The investigation of nanobot clustering techniques able to be effective even with the increasing of diseased cells is the object of Chapter 7.

5.7.4 Nanobots as disposables

The lifetime of erythrocytes is of about 10^2 days; leucocytes last from a few days to years. I can neither pretend that nanobots work indefinitely in the blood (think, for instance, of the redox detection of neutral substances exploiting a sacrificial substrate) nor tolerate that exhausted nanobots continue to reside and to be accumulated in the circulatory system. On another side, the secretion systems will limit the nanobot sojourn lifetime in the organism to a finite value $\tau_{\rm s}$. The optimal situation would be achieved if the nanobots were not released at all until they work but were immediately released when they are no longer active. An idea in this direction could be the following: the nanobot is formed by a set of smaller separated modules, processed all together on an SOI (silicon-on-insulator) substrate and glued before the separation from the substrate; see Figure 5.7. The glue has two functions: it imparts some degree of flexibility to the nanobot (thus making easier its motion in capillaries) and is progressively dissolved during nonobot functioning. Its complete dissolution will result in the formation of smaller fragments presumably eliminated with a time constant much shorter than $\tau_{\rm s}$.

5.8 Conclusions

As far as each part of the organism may develop potentially lethal endogenous diseases (I have especially in mind cancer), the nanobots must be able to explore all the organism. For that, mimicking the carriers of the immune system (the white cells), the basic idea for exploring the whole organism involves the exploitation of the blood circulatory system.

Any cell, indeed, to preserve its vital functions must continuously be fed with the carriers of neg-entropy (sugars and O_2) and the waste products (CO_2 and partly H_2O) must be taken way. In turn, that is possible only if all the living cells of the organism are within a diffusion length from the capillary system. Thus, as far as the nanobot has a size comparable with that of blood cells, it can explore the whole organism (within a diffusion length).

A great advantage of nanobots is the fact that they can check the markers of the pathologic tissue in the vicinity of the region where they are generated. After that, the scout will dock the capillary endothelial tissue eventually being the nucleus for a the formation of a scout cluster sufficiently large to give a detectable CAT contrast or sufficiently intelligent to transmit the information to the central unit, thus completing its goal.

In this chapter I described only the mechanism allowing the clustering of nanobots. Chapter 7 contains the description of the clustering strategy, and the results of their simulation in two different settings.

Chapter 6

The circulatory system

The main results presented in this chapter have been published in (Cerofolini and Amato, 2012).

6.1 Introduction

In section 5.5 the circulatory system as been identified as the most appropriate level of surveillance for the diagnosis of diseases. This means that the circulatory system is the environment in which I envision the nanobots will operate

This chapter is thus devoted to the description of the human circulatory system and of its role, and the development of a simplified model based on fractal methods. This model will then be used in the next chapter for the simulation of nanobots clustering algorithms.

The main result of this chapter is the fractal tree model of the systemic circulatory system described in section 6.5, that has been published in (Cerofolini and Amato, 2012). This model has been developed from scratch by introducing some equations able to link the physiological parameters of the circulatory system with the parameters of a symmetrical fractal trees.

Before introducing this model, two short sections review the physiological description of the circulatory system and the fractal trees respectively.

6.2 A brief physiological description of the circulatory system

It is common practice to view the cardiovascular system as composed of two distinguished parts (Raff and Levitzky, 2011):

1. the *pulmonary circulation*, composed of the right heart pump and the lungs
2. the *systemic circulation*, in which the left heart pump supplies blood to the systemic organ

The pulmonary and systemic circulations are arranged in series, that is, one after the other. Figure 6.1 shows a simplified scheme of the two circulations.

Blood flow through all organs is passive. It occurs only because the pumping action of the heart keeps arterial pressure higher than venous pressure. In particular the right heart pump

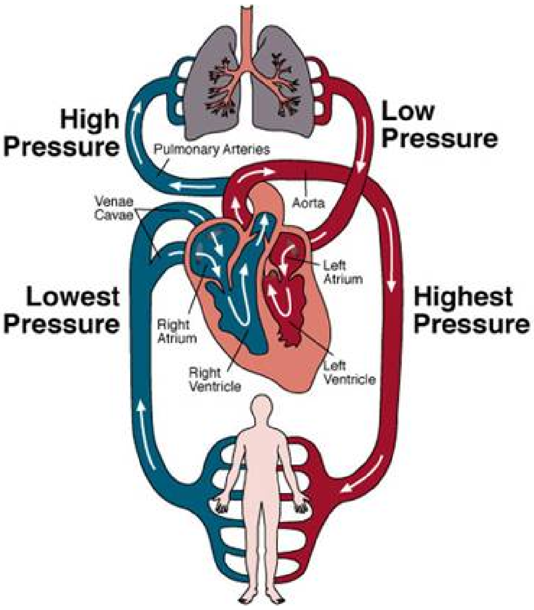


Figure 6.1: Simplified scheme of the circulatory system

provides the energy necessary to move blood through the pulmonary vessels, and the left heart pump provides the energy to move blood through the systemic organs.

Blood that is ejected into the aorta by the left heart passes sequentially through many different types of vessels before it returns to the right heart. The major vessel classifications are arteries, arterioles, capillaries, venules, and veins; see Figure 6.2. These consecutive vascular segments are distinguished from one another by differences in physical dimensions, morphological characteristics, and function.

Some representative physical characteristics are shown in Figure 6.2 for each of the major vessel types. It should be realized, however, that the vascular bed is a continuum and that the transition from one type of vascular segment to another does not occur abruptly.

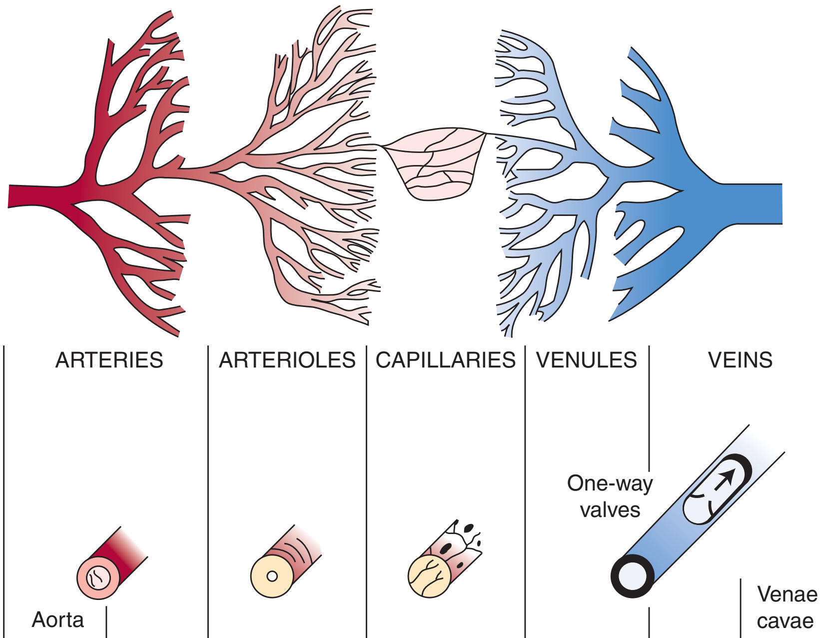
Arteries are thick-walled vessels that can expand to accept and temporarily store some of the blood ejected by the heart contraction (systole) and then, by passive recoil, supply this blood to the organs downstream during heart relaxation (diastole). Arterioles are not only smaller than arteries, but they also have a different structure (Raff and Levitzky, 2011).

Capillaries are the smallest vessels in the vasculature. In fact, red blood cells with diameters of about 7 $\mu\mathrm{m}$ must deform to pass through them. The capillary wall consists of a single layer of endothelial cells, which separate the blood from the interstitial fluid by only about 1 $\mu\mathrm{m}$.

After leaving capillaries, blood is collected in venules and veins and returned to the heart. Venous vessels have very thin walls in proportion to their diameters. Because of their thin walls, venous vessels are quite distensible. Venous vessels, especially the larger ones, also have one-way valves that prevent reverse flow.

6.3 A brief review of fractal trees

A fractal tree can be loosely defined as a trunk and a number of branches that each looks like the tree itself, thus creating a self-similar object. Often, these structures appear strikingly similar



	ARTERIES		ARTERIOLES	CAPILLARIES	VENULES	VEINS	
	Aorta						Venae cavae
Internal diameter	2.5 cm	0.4 cm	30 μm	5 μm	70 μm	0.5 cm	3 cm
Wall thickness	2 mm	1 mm	20 μm	1 μm	7 μm	0.5 mm	1.5 mm
Number	1	160	5×10^7	10^{10}	10^8	200	2
Total cross-sectional area	4.5 cm^2	20 cm^2	400 cm^2	4500 cm^2	4000 cm^2	40 cm^2	18 cm^2

Figure 6.2: Structural characteristics of the peripheral vascular system. Taken from (Raff and Levitzky, 2011)

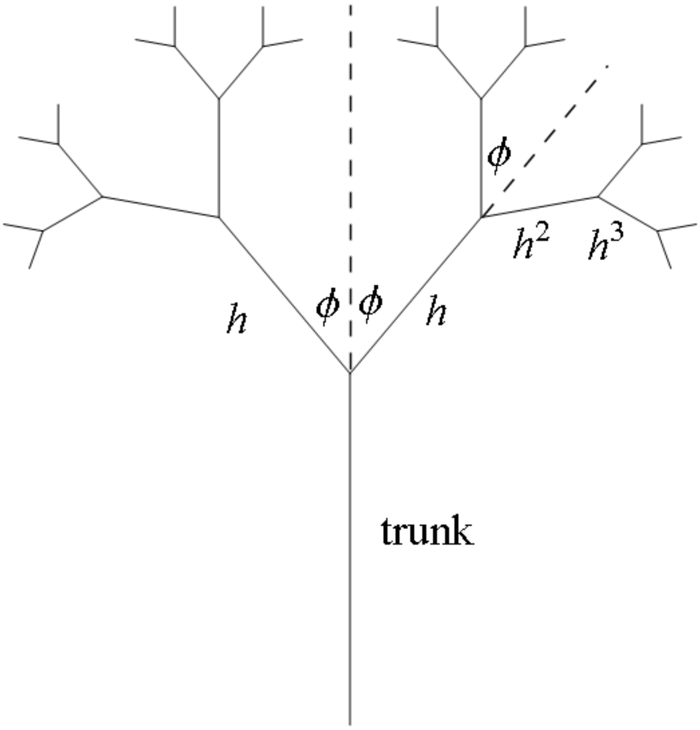


Figure 6.3: A symmetric tree, constructed through stage 4 with scaling ratio $h = 0.58$ and angle $\phi = 2/9\pi$, and $n_b = 2$. The trunk is assumed of length 1

to real trees. Fractals tree can be obtained by recursively applying a branching rule, starting form a given tree trunk (Mandelbrot, 1983).

Fractal trees are not object as simple as they might seem. First of all, a tree is the composition of two different parts: branch tips (the tree canopy), and branches. The canopy can be a fractal object by itself. In fact is the branch tips are disconnected they can form a *fractal dust*, i.e. a totally disconnected set with topological dimension equal to 0, but positive Hausdorff dimension (Falconer, 2003); the most famous example of fractal dust is the middle third Cantor set (Falconer, 2003). If each branch tip belongs also to some other branch, the canopy is a connected set and can results un fractal objects like the Koch curves (Mandelbrot and Frame, 1999).

6.3.1 Filiform fractal trees

Definition 2 (Symmetric fractal tree) *A symmetric fractal tree is described by the triple* $\mathcal{T} = (h, \phi, n_b)$, *where* h *is the scaling ratio at each level,* ϕ *is the angle of rotation, and* n_b *is the number of branches.*

The meaning of the parameter can be better grasped by looking at Fig. 6.3

With these parameters, the tree can be constructed by using the following affine transformation:

Definition 3 (2D tree transformation) *For a 2-dimensional symmetric fractal tree* $\mathcal{T} = (h, \phi, n_b)$,

h	ϕ	n_b	tree depth
0.66	$\pi/4$	2	4
0.58	$2\pi/9$	2	4
0.4	$\pi/4$	3	4
0.5	$4\pi/9$	3	4
0.7	$\pi/2$	2	12
0.5	$\pi/2$	3	10
0.65	$13\pi/18$	2	13
0.61	$29\pi/36$	2	10

Table 6.1: Parameters of 2-dimensional fractal trees shown in Figure 6.4

the corresponding affine transformations $t_j^{(2)}(x,y)$ *are given by the following equations:*

$$t_j^{(2)}(x,y) = y + h \begin{pmatrix} \cos(\theta_j) & \sin(\theta_j) \\ -\sin(\theta_j) & \cos(\theta_j) \end{pmatrix} \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} (y-x)$$

where $x = (x_1, x_2)$ *is the starting point of the parent branch,* $y = (y_1, y_2)$ *the ending point,* $\theta_j = 2\phi(j-1)/(n_b-1)$ *and* $j = 1, \ldots, n_b$.

Figure 6.4 shows some examples of 2-dimensional trees with different parameters listed in Table 6.1. Some of these parameters are taken from (Mandelbrot and Frame, 1999).

Definition 4 (3D tree transformation) *For a 3-dimensional symmetric fractal tree* $\mathcal{T} = (h, \phi, n_b)$, *the corresponding affine transformations* $t_j^{(3)}(x,y)$ *are given by the following equations:*

$$t_j^{(3)}(x,y) = y + hR_{(y-x)}(\eta_j) \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi) & -\sin(\phi) \\ 0 & \sin(\phi) & \cos(\phi) \end{pmatrix} (y-x)$$

where $x = (x_1, x_2, x_3)$ *is the starting point of the parent branch,* $y = (y_1, y_2, y_3)$ *the ending point,* $R_{(y-x)}(\eta_j)$ *is the rotation matrix around the 3-dimensional vector* $(y-x)$ *by* η_j *radians,* $\eta_j = 2\pi(j-1)/(n_b-1)$ *and* $j = 1, \ldots, n_b$.

Figure 6.5 shows some examples of 3-dimensional trees with different parameters listed in Table 6.2.

6.3.2 Thick-stemmed fractal trees

Thick-stemmed fractal "trees" are involved in lungs, vasculatures, botanical trees, river networks,… (Mandelbrot, 1983).

Definition 5 (Symmetric thick-stemmed fractal tree) *A symmetric thick-stemmed fractal tree is described by the quadruple* $\mathcal{T} = (h, b, \phi, n_b)$, *where* h *is the scaling ratio at each level, the ratio* b *between the diameters of parent and children branches,* ϕ *is the angle of rotation, and* n_b *is the number of branches.*

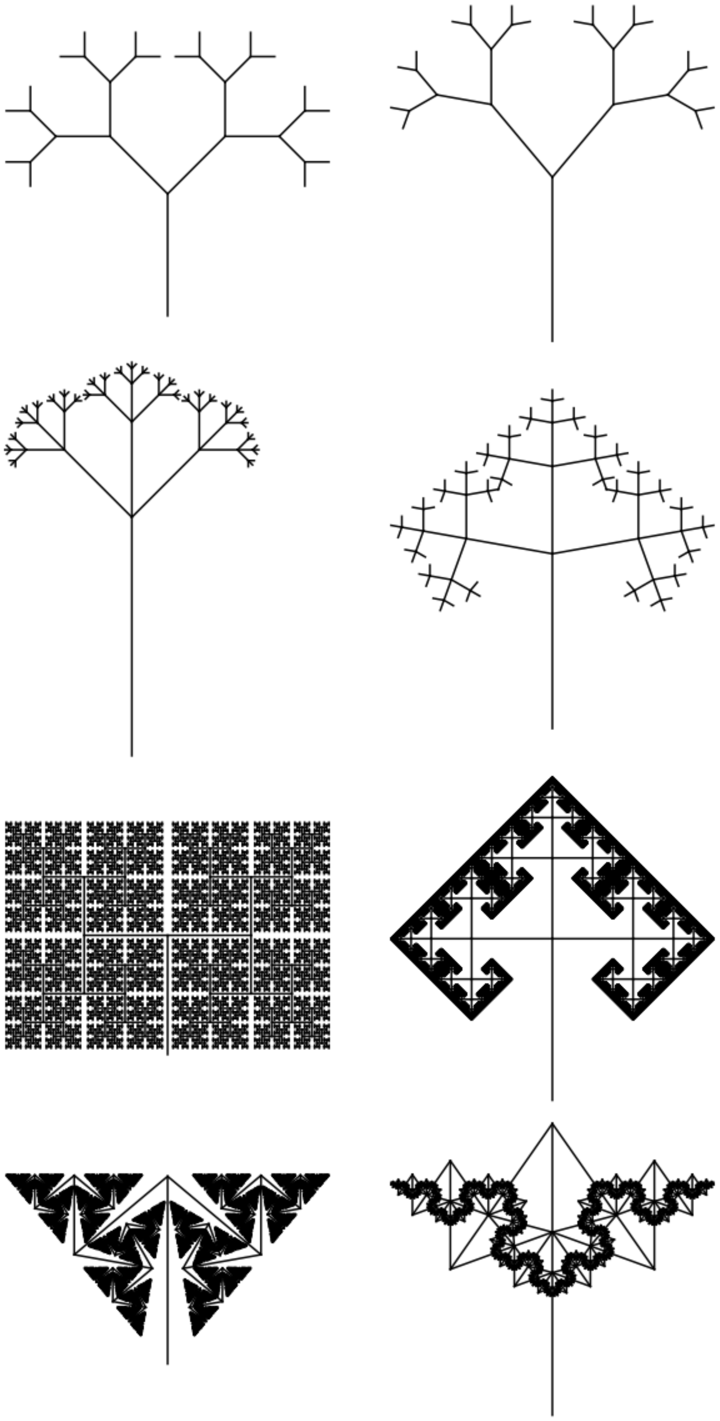


Figure 6.4: Examples of 2-dimensional fractal trees. Their parameteres are listed (row-by-row) in Table 6.1

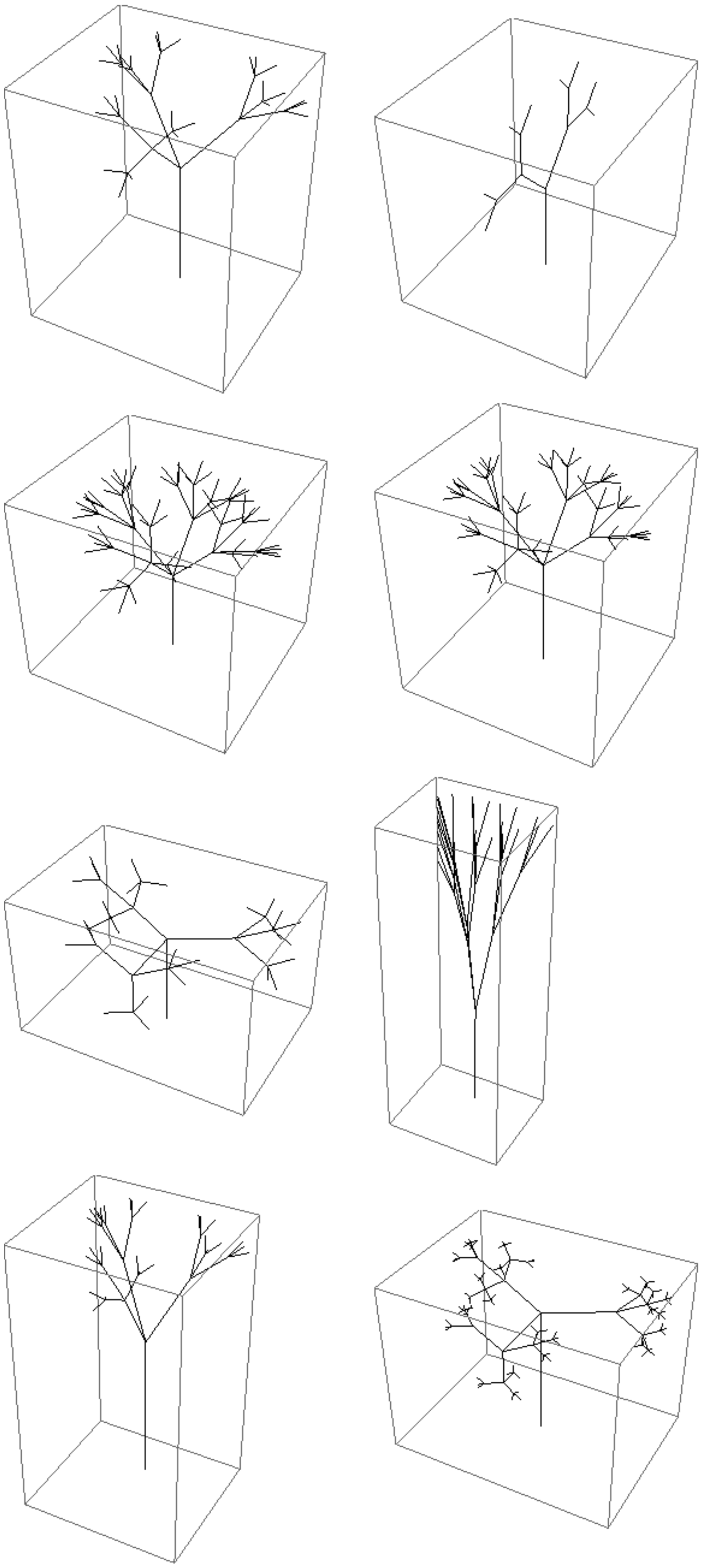


Figure 6.5: Examples of 3-dimensional fractal trees. Their parameters are listed (row-by-row) in Table 6.2

h	ϕ	n_b	tree depth
0.5507	$\pi/4$	3	3
0.66	$\pi/4$	2	3
0.66	$\pi/4$	4	3
0.5507	$\pi/4$	4	3
0.66	$\pi/2$	3	3
0.79	$\pi/18$	3	3
0.5	$\pi/6$	3	3
0.5	$\pi/2$	3	4

Table 6.2: Parameters of 3-dimensional fractal trees shown in Figure 6.5

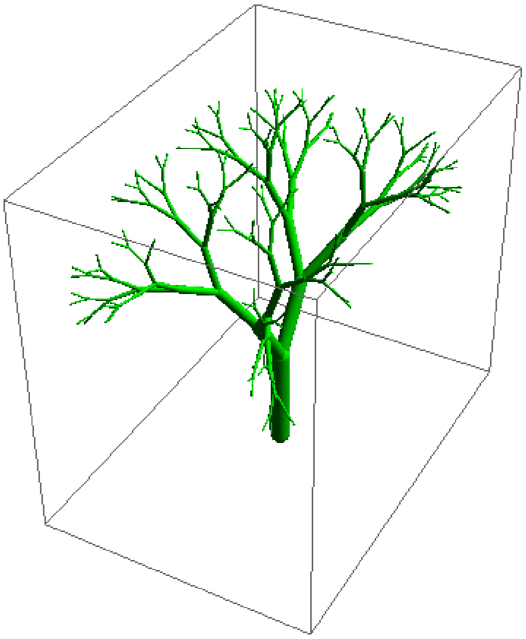


Figure 6.6: An example of a 3-dimensional thick-stemmed fractal tree with $h = 0.77$, $b = 1/\sqrt{2}$, $phi = 0.58$, and $n_b = 2$.

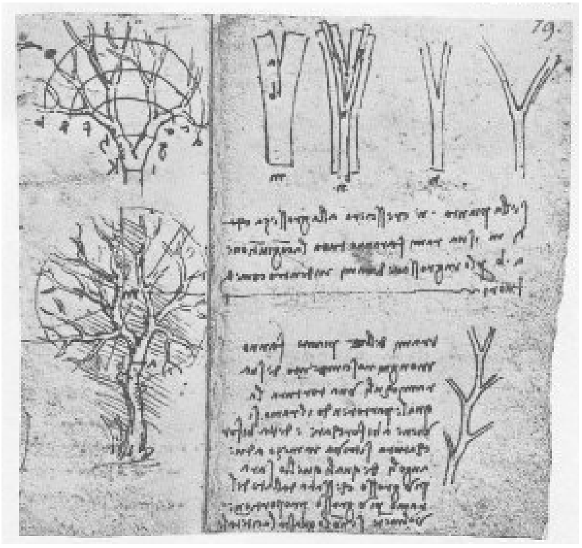


Figure 6.7: One of the plates contained in the chapter *Botany for painters* in Leonardo Da Vinci's notebooks.

Fig. 6.6 shows an example of Symmetric thick-stemmed fractal tree.

In (Mandelbrot, 1983) the relation between diameters at branching node i ($i = 1, 2, \dots, \overline{n}$) is given by the the following equation:

$$d_{i-1}^{\Delta} = d_i^{\Delta} + {d'_i}^{\Delta},$$

where d_{i-1} is the diameter of the parent segment, d_i and d'_i are the diameters of the children segments, and Δ is called *diameter exponent*.

Leonardo Da Vinci uses the same equation (with $\Delta = 2$) in describing botanical trees (Chapter VIII, Note 394 of his Notebooks (Richter, 1970)). Figure 6.7 shows the plate close to Leonardo's note.

In (Changizi and Cherniak, 2000) it is argued that in modeling the arterial networks via fractal trees, the values of the diameter exponent Δ (see Section 6.3.2) to be considered ought satisfy $2 \leq \Delta \leq 3$. In particular, in the same paper, $\Delta = 2.7$ is suggested as an adequate value for small arteries, while a value of close to 3 for capillaries.

Beside the diameter exponent, the geometry of a binary tree is characterized by two junction branch angles, ϕ_1 and ϕ_2, which are respectively the angle between the first child and the parent and the angle between the second child and the parent.

The rotation angles ϕ_1 and ϕ_2 at the bifurcation of blood vessels can be estimated from the ratio between the parent an children diameter (Murray, 1926):

$$\cos\phi_1 = \frac{d_{i-1}^4 + d_i^4 - {d'_i}^4}{2d_{i-1}^2 d_i^2}, \; \cos\phi_2 = \frac{d_{i-1}^4 - d_i^4 + {d'_i}^4}{2d_{i-1}^2 {d'_i}^2}. \tag{6.1}$$

These equations are based on the assumptions that the total work of the circulation is to be a minimum (Murray, 1926).

In particular, when symmetric trees are considered, equation (6.1) boils down to:

$$\cos\phi = \frac{1}{2}\left(\frac{d_{i-1}}{d_i}\right)^2, \tag{6.2}$$

because $d_i = d'_i$.

6.4 Distributing energy inside the body

The surface area A of any body of assigned regular form varies with its volume V as

$$V \propto A^{3/2}, \tag{6.3}$$

the constant of proportionality being characteristic of the shape. Any organism, whose single constituents require a continuous flow of energy[1] not to die, cannot increase its volume indefinitely preserving its shape. In fact, the required energy would increase as its volume V, whereas the energy input increases at most as its surface A. This means that if an organism grew continued indefinitely, according to (6.3), above a certain volume, a part of the organism

---

[1] In the case considered here the energy is chemical in nature, although there are system like the majority of the Vegetable Kingdom where such an energy is ultimately electromagnetic in character.

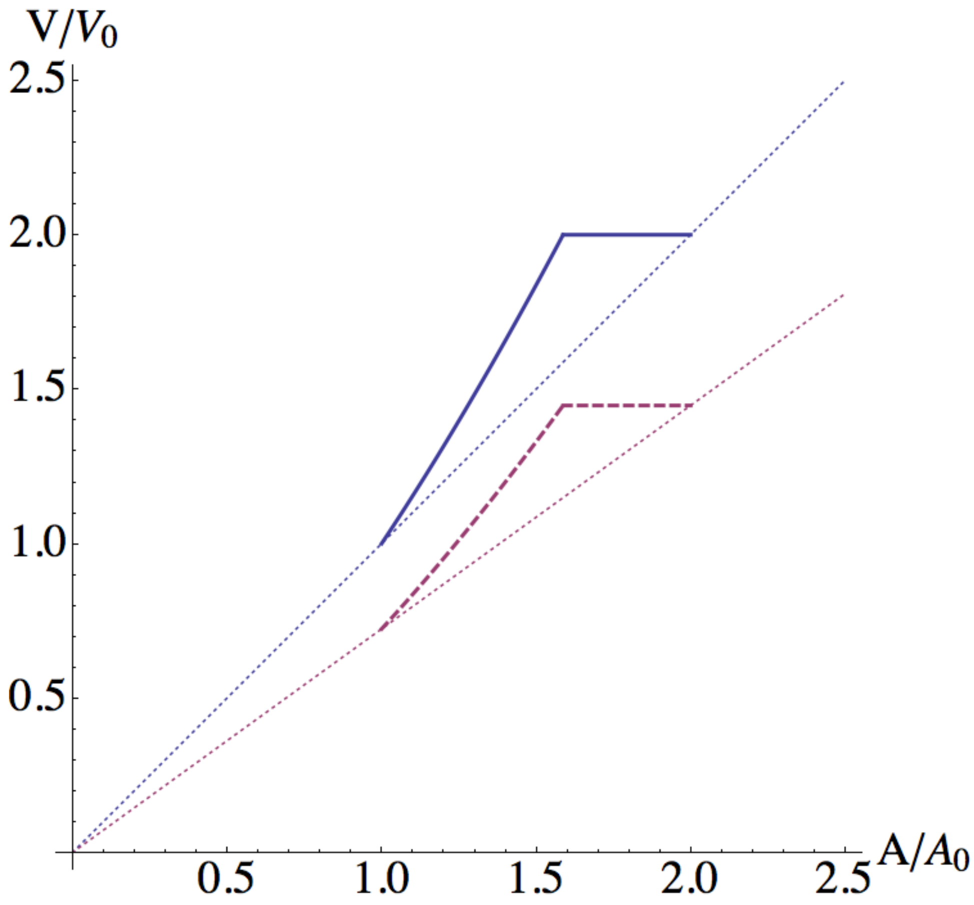


Figure 6.8: Plot of the volume V of a of single cell as a function of its outer area A, calculated assuming that the cell preserves its original shape during the entire G phase while changes continuously shape during the M phase until two new cells are formed. The solid line represents spherical organisms, while the dashed one represents (hypothetical) cubic organisms. V_0 is the initial volume of the sphere and A_0 is its area.

could not be any longer fed by an adequate energy. The strategy adopted by living organisms to continue their growth in spite of the limits imposed by (6.3) is reproduction.

The growth of single cells is an extremely complicated process with several phases an numerous check-points; to some extent it can however be described as formed by two phases:

- *growth* G, during which the cell increases its volume,
- and *mitosis* M, during which modifies its shape increasing its area without large change of volume (Aguda, 2006).

At the end of the G phase the cell attains a volume twice that it had at the beginning of its life; during M, together with important intracellular modification (formation of telomere and separation of the chromosomes) the cell changes its shape eventually assuming an area twice that it had at the beginning of its life. In this way, with the separation of the nucleus in two parts, the cell is in the condition of forming two copies of the parent cell. Figure 6.8 sketches this process.

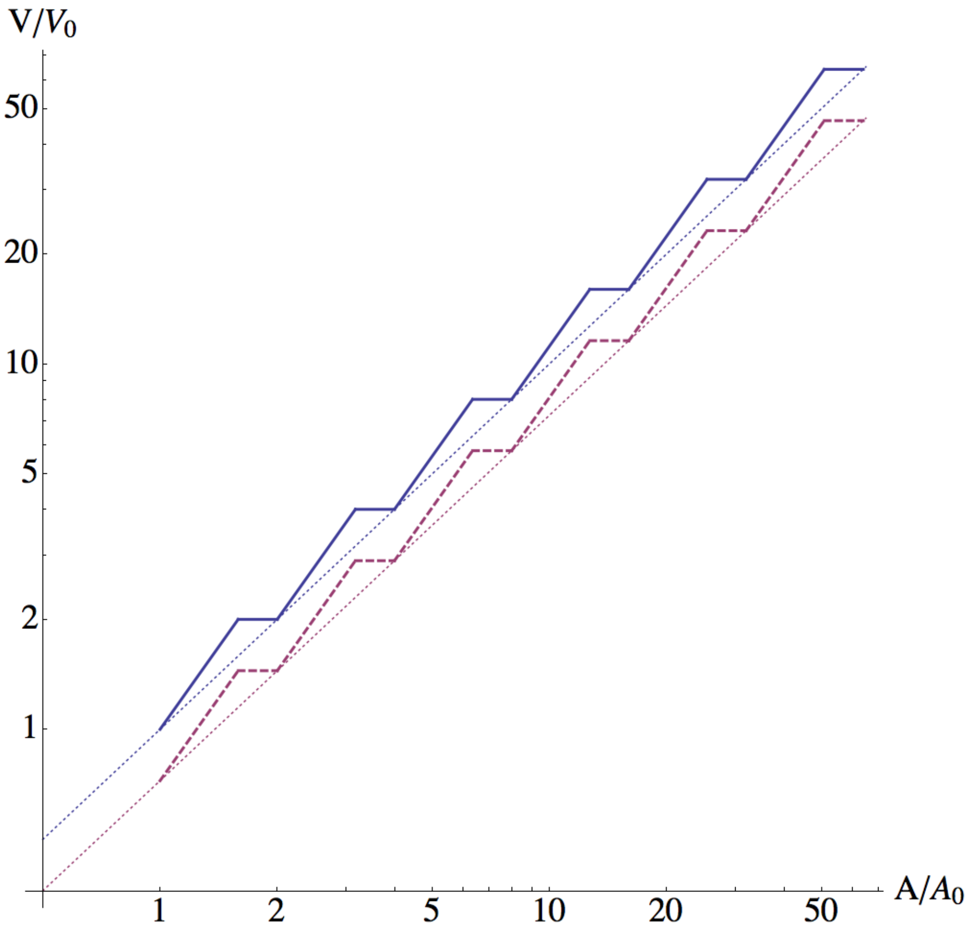


Figure 6.9: Log-log plot of the volume V of a system composed of single-cells organisms as a function of their total area A, for six generations. The solid line represents spherical organisms, while the dashed one represents (hypothetical) cubic organisms. V_0 is the initial volume of the sphere and A_0 is its area.

If each newly born cell may be viewed as isolated, the process can be reiterated indefinitely; Figure 6.9 shows that if the details of the GM cycle are ignored, the cell colony grows over many generations with total volume approximately proportional to the corresponding area,

$$V \propto A, \tag{6.4}$$

a behaviour that allows the functional characteristics of each cell to be preserved during growth (Cerofolini, 1981; Cerofolini, 1983a).

If the cells were homogeneously dispersed in an unbound medium, their number could increase indefinitely. In a limited environment or in the absence of motility, however, the cells cannot be assumed independent of one another, because the required energy increases in proportion to their number while the energy input can increase only with the size of the zone exposed to the medium.

The strategy adopted to overcome this limit by populations of single cells involves the growth of dendritic colonies with fractal-like structure. Much more interesting is the method adopted by higher organisms to solve this problem. The organization of higher organism is based on the specialization of the various organs to specific functions (locomotion, vision,

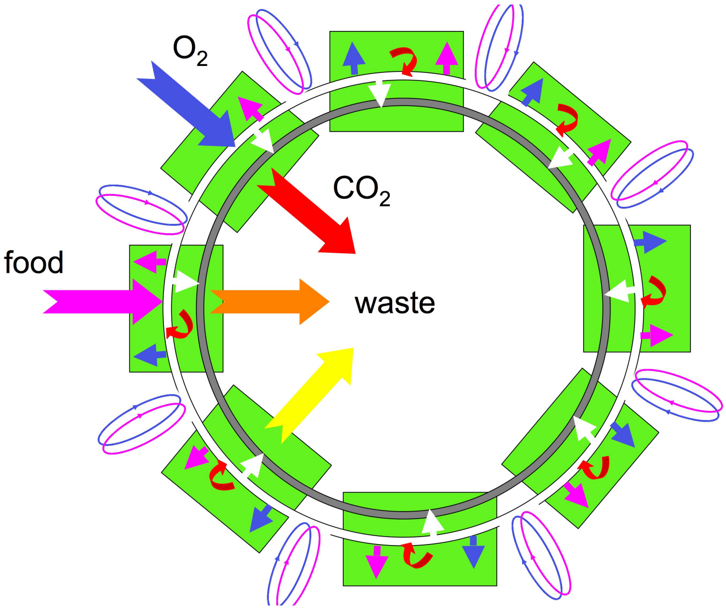


Figure 6.10: Lumped model of animal organisms emphasizing the distribution of O_2 and nutritive elements by the circulatory system and the collection of CO_2into the circulatory system and the passage of wastes into the lymphatic system

hearing, digestion, reproduction, etc.). This result is obtained specializing the expressome to the task, so that cells with the same genome are expressed in such a way as to generate, for instance, (*a*) nucleus-free cells (blood red cells) as well as poly-nucleated cells (muscle cells) in addition of course to mononuclear cells, (*b*) highly regular lenticular cells (again blood red cells) as well as dendritic cells (nerve cells), (*c*) cells undergoing frequent mitosis (epithelial cells) as well as cells not undergoing any reproduction for most of life time (again nerve cells), and certainly the list is not exhaustive.

All the functions of the organism are supported by organs requiring an adequate input of energy and neg-entropy; their needs are satisfied centralizing the collection of energy and neg-entropy and distributing them to the various organs through the circulatory system.

The circulatory system is so built that energy-rich substances and O_2 can out-diffuse from it (thus feeding nearby cells) only in the capillaries; conversely the waste materials produced by cell metabolism are eventually collected in nearby capillaries before being eventually delivered to the secretion apparatus. This basin, where the transport of matter is dominated by diffusion, will be referred to as diffusion basin. The union of all diffusion basins must embed all the living parts of the organism.

This structure can work as described, but in higher organisms it is complicated and, in view of the key role of evolution on structure-function relationship, made more efficient by another distributed system, devoted to to increase the collection efficiency catabolites and other wastes—the lymphatic system. Although this system could play a key role in therapy, it will be henceforth ignored and the focus will be on the blood circulatory system.

A lumped model of animal organism is sketched in Fig. 6.10, where the entire organism is depicted as formed by several (respiratory, digestive, urinary, cardiovascular, lymphatic, immune, etc.) systems linked by two complex space-filling systems (the circulatory and lymphatic systems). The chemicals necessary to the life are taken from the environment through the respiratory system (O_2) and digestive system (food) and distributed to all tissues of the organism through the circulatory system (that in turn must be fed by a subsystem with the same features—*vena venorum*); the circulatory and lymphatic systems are also involved in the elimination of the CO_2 and nitrogen-catabolite wastes by the respiratory and urinary systems, respectively.

6.5 Fractal-tree description of the circulatory system — into the scene of crime

Since the work of Mandelbrot, the fractal geometry of biological structures (Mandelbrot, 1983) has suggested that fractal methods could be used for modeling of human circulatory system. In particular Mandelbrot stresses the point that the seminal Harvey work (published in 1628 (Harvey, 1910)) led to a view of the circulation of the blood which asserts that both an artery and a vein are found within a (infinitely) small distance of nearly every point of the body. Stated differently, every point in nonvascular tissue should lie on the boundary between the two blood networks. Considering then that blood is expensive, the volume of all the arteries and veins must be a small percentage of the body volume, leaving the bulk to tissue. These criteria are apparently contradictory, since the tissue must be a topologically 2-dimensional shape (it is the common boundary of two 3-dimensional shapes), and it must have a nonnull volume. However the two above requirements are perfectly compatible in fractal analysis. In fact, tissues can be described as fractal surfaces whose topological dimension is 2 and whose fractal dimension is close to 3. Examples of this kind of fractals have been introduced by Osgood in 1903 (Osgood, 1903).

Both pulmonary and systemic circulation (see Section 6.2) can be described by a couple of trees forming a double canopy inside each organ (Harvey, 1910), where the the ultimate branches of the canopies are linked by filaments (the capillaries). I consider that each couple if formed by two identical trees (one arterial, the other venous) and that the trunks of all trees are rooted in the human heart. In recent years fractal analysis of blood vessel in any different parts of human body has been described. In particular in (Gabrys et al., 2005) the application of asymmetrical and symmetrical *fractal trees* has been analyzed.

The following features (Pocock and Richards, 2006) will be used as basis for the development of the fractal model of the circulatory system:

- the circulatory system is formed by a double (arterial and venous) tree;
- each tree has a branching factor of 2 and terminates after a number n of approximately 35 branchings;
- the trunks of the trees are connected (by heart) whereas the canopies are connected by 2^n ($\simeq 3 \times 10^{10}$) capillaries;
- each capillary feeds on the mean 2×10^3 cells, as follows from the total number of cells of an adult human (6×10^{13} cells);
- the average capillary diameter is in the range $4 - 8$ μm and its length is $0.5 - 1$ mm;
- the mean blood velocity in capillaries is $v_n = 2 \times 10^{-2}$ cm s^{-1} so that the mean time τ_n spent by the any particle flowing along the capillary is around $2 - 4$ s whereas the time required by blood for a complete cycle is about 60 s.

6.6 Symmetrical tree model: quantitative description

In the following I'll focus my attention on the the systemic circulation, but the same reasoning (with different parameters) could be applied to the pulmonary circulation as well.

For the sake of simplicity, let us suppose that the systemic circulation is composed by two binary thick-stemmed fractal trees described in Section 6.3.2. Moreover by using equation (6.2) the angle ϕ is directly related to the ratio b. In particular

$$\cos\phi = \frac{1}{2b^2}.$$

Given this, just 3 parameters are needed to characterize the fractal tree:

- the number n of branching (i.e, the depth of the tree)
- the ratio h between their length
- and the ratio b between the diameters of parent and children branches.

The description of the circulatory system as a double fractal tree whose canopies are linked by capillaries (common boundary of the trees) must satisfy the following five conditions:

$$2^n = N, \tag{6.5}$$

$$\frac{1-(2b^2h)^{n+1}}{1-2b^2h} = \frac{\frac{1}{2}v}{\pi r_0^2 l_0}, \tag{6.6}$$

$$\frac{1-(2b^2h)^{n+1}}{1-2b^2h} = \frac{\tau}{2}\frac{u_0}{l_0}, \tag{6.7}$$

$$(2b^2)^n = \frac{u_0}{u_n}, \tag{6.8}$$

$$h^n = \frac{1}{2}\frac{l_n}{l_0}, \tag{6.9}$$

where N is the number of capillaries, v is the blood volume, r_0 and l_0 the aorta radius and length, τ is the cycle time of the blood, u_0 and u_n are the speed of the blood in the aorta and capillaries, and l_n is the length of the capillaries. A full derivation of these equations is given later in this section.

In the above five equations the parameters n, b and h defining the fractal geometry of the tree are on the left, whereas the quantities on the right can be found in handbooks of anatomy and physiology. In view of the relative variability of these data from one source to another, the ones assumed in this work (and already used in previous chapters) are given in Table (Murray, 1926).

Of course the above equations are not consistent, so that I should limit to determine the fractal tree describing experimental data with minimum inaccuracy. With the aim of improving the accuracy of the search we add another geometric condition produced by the need that even the organs at the maximum distance from the aorta (hand and foot fingers) must be bathed by the capillary system. This condition reads:

$$\frac{\left(h\,(h^{n+1}-1)\,\frac{1}{2b^2}+h^{n+2}-1\right)^2+h^2\,(h^{n+1}-1)^2\,(1-\frac{1}{4b^4})}{(h^2-1)^2} = \left(\frac{D_{\mathrm{V}}}{2l_0}\right)^2, \tag{6.10}$$

parameter	value
N	3×10^{10}
v	5×10^3 cm^3
r_0	1 cm
l_0	6 cm
τ	60 s
u_0	26.5 cm s^{-1}
u_n	2×10^{-2} cm s^{-1}
l_n	0.08 cm

Table 6.3: Physiological parameters assumed in this work

where D_{V} is the Vitruvian diameter ($D_{\mathrm{V}} = 85$ cm) and ϕ is the average branching angle[2] ($\phi = 20$ °). Although the notion of Vitruvian diameter goes back to Da Vinci, to the best of our knowledge Eq. (6.10) has never been used yet. It is immediately understood merging the concept of circulatory system with the famous Da Vinci's drawing generally referred to as Vitruvian Man (hence the name; see Figure 6.11). The drawing (a blend of art and science during Renaissance) depicts a male figure in two superimposed positions with his arms and legs apart and simultaneously inscribed in a circle and square, and can be interpreted saying that the proportion of man is so done to admit a center (the heart) able to bathe the farthest organs (hand and foot fingers) irrespective of posture.

In the rest of this section a full derivation of the six conditions (6.5)–(6.10) is given.

6.6.1 Equation (6.5): Number of capilllaries

This equation simply states that the number N of the capillaries is equal to the number of the leaf of the tree (i.e., 2^n). As it will be shown later, this equation is critical for the stability of parameter estimation.

6.6.2 Equation (6.6): Blood volume

The blood volume v must be equal to the volume occupied by the systemic circulatory system. By assuming that each vessel of the circulatory system can be considered as a cylinder, its volume is given by $\pi r_i^2 l_i$. Thus:

$$\frac{v}{2} = \pi r_0^2 l_0 + 2\pi r_1^2 l_1 + 2^2 \pi r_2^2 l_2 + \ldots + 2^n \pi r_n^2 l_n.$$

---

[2] Since the tree is assumed symmetric, there is only one angle for both branching arteries.

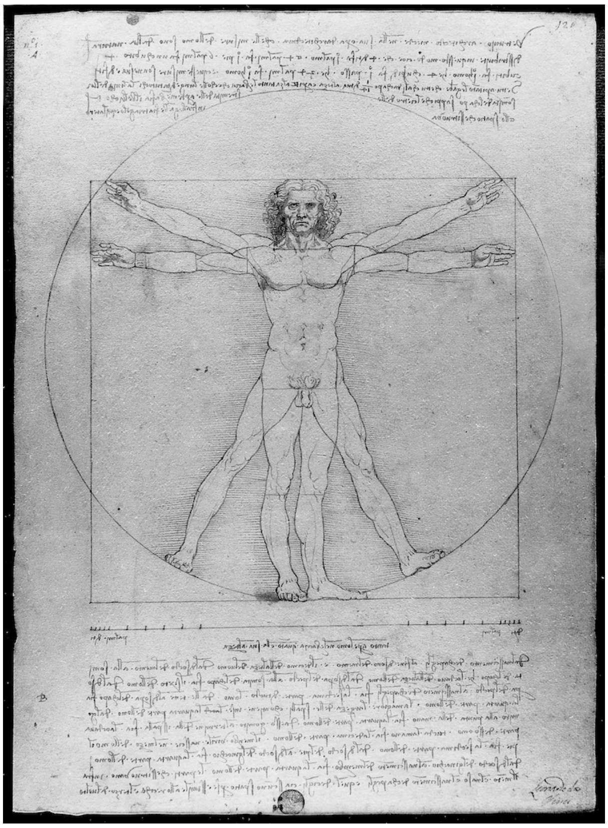


Figure 6.11: Da Vinci's Vitruvian Man

Now remembering that, by $r_i = br_{i-1}$ and $l_i = hl_{i-1}$, the previous equation becomes

$$\begin{aligned}
\frac{v}{2} &= \pi\left(r_0^2 l_0 + 2b^2 r_0^2 h l_0 + 2^2 b^4 r_0^2 h^2 l_0 + \ldots + 2^n b^{2n} r_0^2 h^n l_0\right) \\
\frac{v}{2} &= \pi r_0^2 l_0 \left(1 + 2b^2 h + 2^2 b^4 h^2 + \ldots + 2^n b^{2n} h^n\right) \\
\frac{v}{2} &= \pi r_0^2 l_0 \sum_{i=0}^{n} (2b^2 h)^i \\
\frac{v}{2} &= \pi r_0^2 l_0 \frac{1 - (2b^2 h)^{n+1}}{1 - 2b^2 h}
\end{aligned}$$

Equation (6.6) is simply obtained by rearranging the terms of the last equation.

6.6.3 Equation (6.7): Blood flow and total blood volume

The quantity of blood flowing trough the aorta during a complete cycle time τ of the blood,is given by $\tau\pi r_0^2 u_0$. The total volume of the blood in the circulatory system must be equal to this quantity. Then Equation (6.6) is obtained by following a reasoning similar to that used for Equation (6.6).

6.6.4 Equation (6.8): Blood velocity

The quantity of blood flowing trough the aorta is given by $\pi r_0^2 u_0$. Assuming that the circulatory system is a closed system, this quantity must be equal to the quantity of blood flowing trough all the capillaries:

$$\begin{aligned}
\pi r_0^2 u_0 &= \pi 2^n r_n^2 u_n \\
\pi r_0^2 u_0 &= \pi 2^n (b^n r_0)^2 u_n \\
\frac{u_0}{u_n} &= (2b^2)^n
\end{aligned}$$

6.6.5 Equation (6.8): Blood velocity

Half of the length l_n of the capillaries is obtained by applying the iteration rule $l_i = h l_{i-1}$.

6.6.6 Equation (6.10): Vitruvian diameter

The Vitruvian radius $D_{\text{V}}/2$ is the maximum distance that has to be covered by the circulatory system to carry the blood from the heart to the farthest capillaries. To simplify the analysis, the fractal tree is considered two-dimensional. In (Mandelbrot and Frame, 1999) it is shown that for two-dimensional trees whose branching angle ϕ is such that $0 \leq \phi \leq 3/4\pi$, the leaf of maximal distance from the root of the tree is achieved by the branches alternating left and right directions.

This leaf (or branch tip) is located at the point with x coordinate given by:

$$\begin{aligned}
x &= l_0 \left(h \sin\phi + h^3 \sin\phi + \ldots + h^{n-1} \sin\phi \right) \\
x &= l_0 h \sin\phi \sum_{i=0}^{n/2-1} h^{2i} \\
x &= l_0 \frac{h\left(h^{n+1}-1\right)\sin(\phi)}{h^2-1},
\end{aligned}$$

and y coordinate given by:

$$\begin{aligned}
y &= l_0 \left(1 + h\cos\phi + h^2 + h^3 \cos\phi + \ldots + h^n \right) \\
y &= \frac{h^{n+2}-1}{h^2-1} + \frac{h\left(h^{n+1}-1\right)\cos(\phi)}{h^2-1} \\
y &= l_0 \frac{h\left(h^{n+1}-1\right)\cos(\phi) + h^{n+2} - 1}{h^2-1},
\end{aligned}$$

where in both cases, n is assumed to be even.

Consequently the maximum distance $D_{\text{V}}/2$ is given by the following equation:

$$\begin{aligned}
\frac{D_{\text{V}}}{2} &= \sqrt{x^2+y^2} \\
\frac{D_{\text{V}}}{2} &= l_0 \sqrt{\frac{\left(h\left(h^{n+1}-1\right)\cos(\phi) + h^{n+2} - 1\right)^2 + h^2\left(h^{n+1}-1\right)^2 \sin^2(\phi)}{\left(h^2-1\right)^2}}
\end{aligned}$$

Figure 6.12 shows the canopy of the two-dimensional tree with $h = 0.86$ and $\phi = \pi/5$.

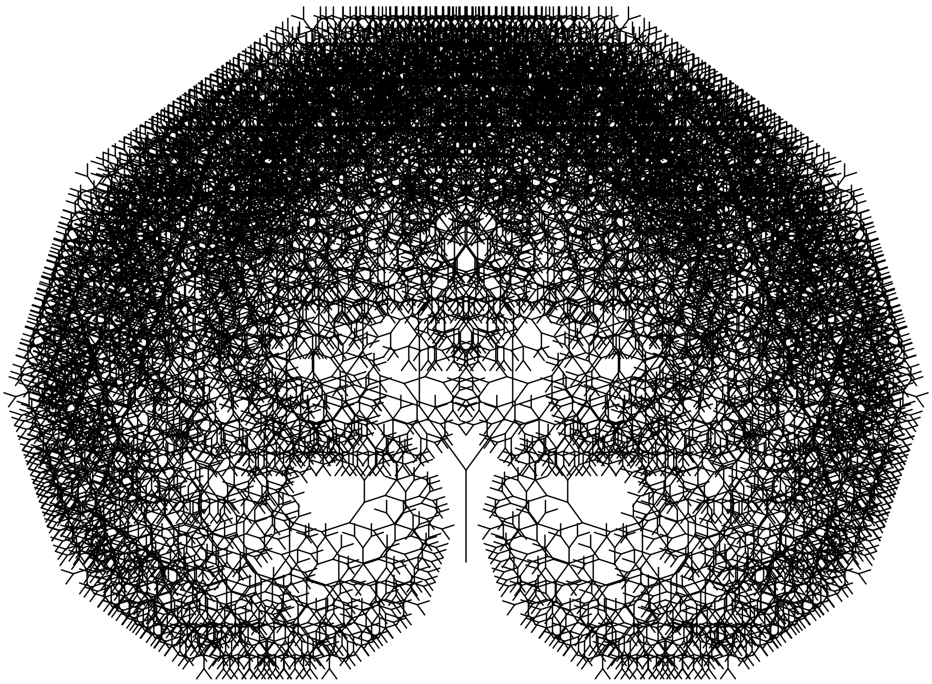


Figure 6.12: The canopy of the two-dimensional tree with $h = 0.86$ and $\phi = \pi/5$.

At last by using the fact that $\cos\phi = \frac{1}{2b^2}$ (obtained from (6.2)) the last equation can be rewritten as:

$$\frac{\left(h\left(h^{n+1}-1\right)\frac{1}{2b^2}+h^{n+2}-1\right)^2+h^2\left(h^{n+1}-1\right)^2\left(1-\frac{1}{4b^4}\right)}{\left(h^2-1\right)^2} = \left(\frac{D_{\mathrm{V}}}{2l_0}\right)^2 \tag{6.11}$$

6.7 Fractal tree parameters estimation

Of course the Equations (6.5–6.10) are not consistent so that we should limit to determine the fractal tree describing experimental data with minimum inaccuracy. Defining the relative

e_1	e_2	e_3	e_4	e_5	e_6
0.06	0.706	0.609	0.879	0.916	0.206

Table 6.4: Percentage error on the 6 conditions given in Section 6.7 associated to the optimal estimate of the fractal tree parameters

inaccuracies as follows

$$\begin{aligned}
\epsilon_1(n,b,h) &= \frac{2^n - N}{N} \\
&= \frac{1}{N}2^n - 1 \\
\epsilon_2(n,b,h) &= \frac{2}{v}\pi r_0^2 l_0 \frac{1-(2b^2h)^{n+1}}{1-2b^2h} - 1 \\
\epsilon_3(n,b,h) &= \frac{2}{\tau}\frac{l_0}{u_0}\frac{1-(2b^2h)^{n+1}}{1-2b^2h} - 1 \\
\epsilon_4(n,b,h) &= \frac{u_n}{u_0}(2b^2)^n - 1 \\
\epsilon_5(n,b,h) &= \frac{2}{l_n} l_0 h^n - 1 \\
\epsilon_6(n,b,h) &= \frac{2l_0}{D_{\mathrm{V}}}\sqrt{\frac{\left(h\left(h^{n+1}-1\right)\frac{1}{2b^2} + h^{n+2} - 1\right)^2 + h^2\left(h^{n+1}-1\right)^2\left(1-\frac{1}{4b^4}\right)}{\left(h^2-1\right)^2}} - 1
\end{aligned}$$

the condition determining n, b and h is thus given by

$$\sum_{i=1}^{6} \epsilon_i^2(n,b,h) = \min. \tag{6.12}$$

Given as constraints that $25 < n < 40$, $0 < b < 1$ and $0 < h < 1$, the optimal estimates of n, b and h are the following: $\hat{n} = 34.805, \hat{b} = 0.784$ and $\hat{h} = 0.866$ These estimates are coherent with other ones computed with different methods (Cerofolini et al., 2010a; Amato et al., 2011). Figure 6.13 shows a symmetrical fractal tree (whose depth is 9) with the above parameters. The percentage errors on the experimental data are given in Table 6.6: Moreover the estimate of b is close to the value 0.77 obtained by assuming as constant diameter exponent $\Delta = 2.7$ (the suggested value for human arterial blood vessel (Gabrys et al., 2005)).

The binary symmetric fractal tree with the parameters estimated by optimizing equation (6.12) is shown in Figure 6.13. Since $b \simeq 0.78$, it follows that $\phi \simeq \frac{\pi}{5}$.

To check the stability of these equations, the same optimization process has been run by excluding, in turn, one of the six equations. The obtained estimates are shown in table 6.5. On one side, from the table it is possible to note that the first equation is paramount in defining an accurate value of depth of the tree. In fact, without the condition $2^n = N$, the estimate $\hat{n}$ is set equal to the maximum allowed value (in this case 40). On the other side, the estimates of all the other parameters (and also n in the other conditions) are quite stable regardless of the lack of any single condition.

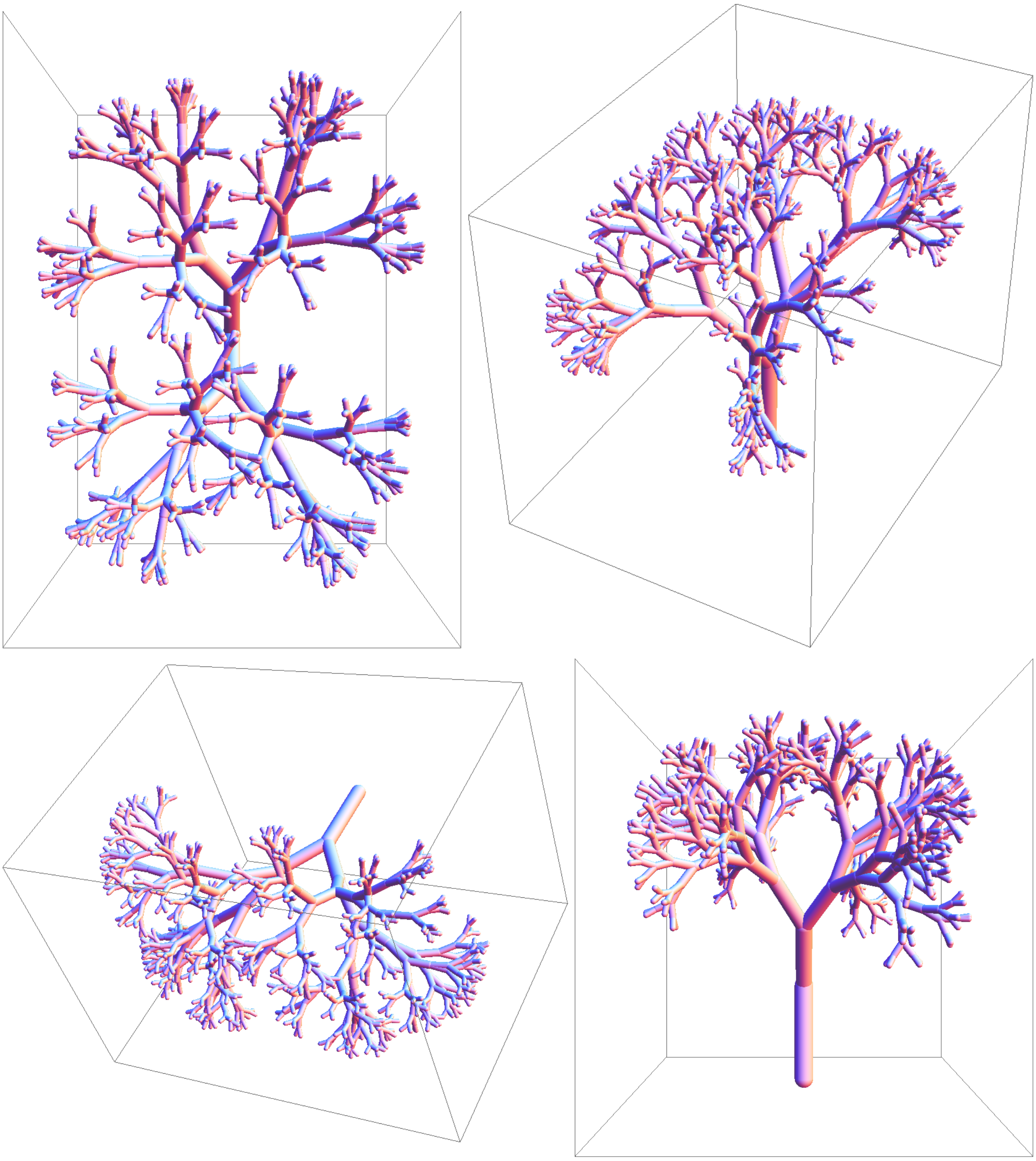


Figure 6.13: Four views of the fractal vascular symmetric tree with parameters obtained by optimizing (6.12). From up to down, left to right they are top view, 2 different isometric views, and front view. The tree is limited to first 7 branchings

excluded condition	$\hat{n}$	$\hat{b}$	$\hat{h}$
1	35.177	0.783	0.867
2	34.805	0.784	0.866
3	34.805	0.784	0.866
4	34.804	0.784	0.866
5	34.804	0.784	0.867
6	34.805	0.784	0.866

Table 6.5: Estimates of the parameters n, b and h considering in turn just five out of the six conditions given in Section 6.7

excluded condition	e_1	e_2	e_3	e_4	e_5	e_6
1	29.497	0.379	0.282	0.444	0.322	0.677
2	0.047	1.037	0.94	0.628	0.672	0.25
3	0.048	0.992	0.896	0.662	0.705	0.244
4	0.014	0.049	0.049	2.734	0.064	0.36
5	0.003	0.064	0.033	0.021	2.756	0.117
6	0.053	0.715	0.618	0.891	0.891	0.211

Table 6.6: Percentage error on the 6 conditions given in Section 6.7 associated to the estimates obtained by considering in turn just five out of the six conditions

6.7.1 Limits of the symmetrical tree model

The hypothesis that the fractal tree is symmetric has a large impact on the model. Consider the number of terminals. In a perfect symmetrical tree an exponentially increasing number of segments are found in each bifurcation level (2^n) and all leaves appear in the highest level only. In the asymmetrical tree the leaves emerge in the early stages of development as well[3].

Another point is the maximal length of a path from root to tree. In a symmetrical tree this length is always the same, while in an asymmetrical one there can be a large distribution of values. This is more realistic, considering the human circulatory system.

Another improvement could be obtained by considering that the degree of symmetry is a function of branching levels. In fact from the analysis in (Gabrys et al., 2005), asymmetrical vessel trees seem to dominate in circulatory system — while symmetrical are concerned in capillary bed only.

---

[3] The idea is that a leaf emerges when the raidus of the branch is below a given threshold (like 5 μm)

Chapter 7

Swarm-based algorithms for nanobots clustering

> *. . . it is in the perspective of miniaturization that swarm-based robotics becomes meaningful*, p. 21 (Bonabeau et al., 1999).

Some of the results presented here have been published in (Amato et al., 2011).

7.1 Introduction

Chapter 5 describes the nanobot architecture, and the task that the nanobots have to tackle to help the diagnosis of endogenous diseases. In brief, the task is to cluster around diseased cells in order to identify the positions of potential cancerous regions.

As already mentioned, a nanobot is any artificial machine with overall size of the order of a few micrometers or less in all spatial directions, constituted by nanoscopic components with individual dimensions in the interval $1 - 10^2$ nm (Requicha, 2003). A major challenge is the design of nanobots (i) addressed to monitoring *in vivo* the health state of complex systems like *homo sapiens sapiens*; (ii) able to embed sophisticated functions, like navigation, recognition, and data transmission; and (iii) suitable for being manufactured by processes compatible with today and likely tomorrow semiconductor industries.

In tackling these challenges it is important to take into account that, in view of their limited size, nanobots are necessarily devices with very limited computational resources and that their interaction can only happen locally. Although at first sight these facts seem to make the challenges really daunting, from nature we have experiences of tiny living beings (like ants) with very limited resources but nonetheless able to perform complex tasks. In fact it is enough to think about insects living in colonies (like ants, bees, wasps and termites) to find beautiful examples of activities apparently requiring intelligent planning and supervision: foraging, nest building, group transport, forming chain to pass obstacles,. . . In other words, swarm intelligence (Bonabeau et al., 1999) seems to be a key tool for designing nanobot control algorithms. In fact, ideas regarding swarm of micro- and nano-bots for medical applications have already been presented in literature (Martel and Mohammadi, 2010; Nagy et al., 2009; Requicha, 2003).

This chapter describes some nanobot clustering strategies that can be used for the identification of regions containing diseased cells. These strategies, that are based on the *communication-through-collision* mechanism described in Chapter 5, can be divided in three groups:

1. Only target (OT)
2. Self-Docking (SD)
3. Division of labor (DL) between explorer and recruits

To test these strategies I hypothesize two different settings. In the first setting, the search region is a square in which the nanobots freely move. This setting mimics the use of nanobots in test plate; for example for blood analysis.

In the second setting, the search region is a tree, and the nanobots flow trough the branches of the trees. After traversing the arterial tree, the nanobots enter in the venous tree and then come back in the aorta (which is the trunk of the arterial tree).

7.2 Nanobots clustering strategies

7.2.1 Only target (OT) strategy

The basic strategy for the use of nanobots is to not consider any interaction between them. In this approach each nanobots searches for the presence of diseased cells, and anchors to the wall of the capillaries once it senses it. The cluster is simply formed by the aggregation of several nanobots in the same area.

This strategy is mainly introduced to compare it with the other two strategies, which require some interaction between the member of the swarms.

Figure 7.1 shows the finite state machine of the nanobot control system for OT strategy.

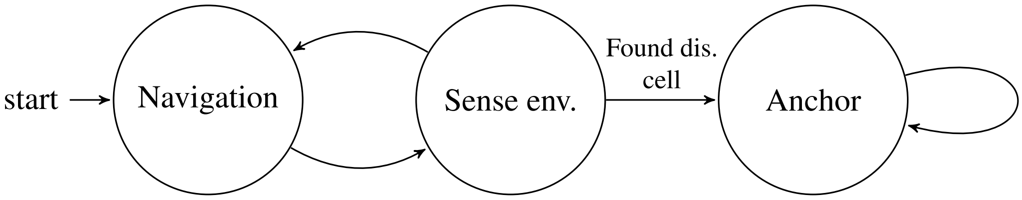


Figure 7.1: Nanobot finite state machine for OT strategy

7.2.2 Self-docking (SD) strategy

The second strategy introduces some interaction between the members of the swarm. In particular the communication mean is direct physical contact between nanobots, as described in Section5.7.

What happens in this strategy is that an anchored nanobot becomes itself a site for docking of other nanobots — the *self-docking* process giving the name to the strategy. Since it is more easy to recognize an anchored nanobot than sense the presence of diseased cells, this should accelerate the formation of the clusters.

Figure 7.2 shows the finite state machine of the nanobot control system for SD strategy.

7.2.3 Division of labor (DL) strategy

Before describing this strategy, a brief background on honey bees foraging is given.

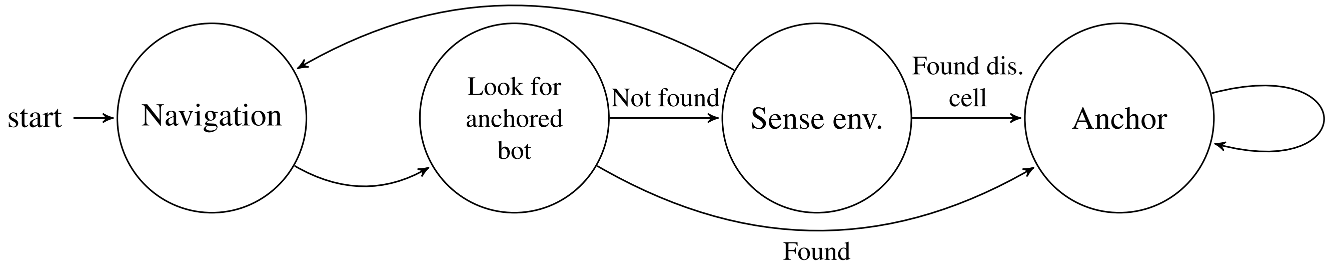


Figure 7.2: Nanobot finite state machine for SD strategy

Labor division in honey bee colonies

In acquiring its food, a colony of honey bees functions as a large, diffuse, amoeboid entity. This entity has the ability to extend itself over great distances and in multiple directions at the same time (covering an area of more than 100 km^2) to tap a vast array of food sources (Seeley, 1996). Colonies are highly skilled at choosing among different food sources, selectively exploiting those that are the most profitable.

Whenever flowers are in bloom, a honey bee colony must solve the difficult problem of wisely deploying its foragers in the surrounding countryside. In particular the colony needs to achieve a proper allocation of foragers among nectar sources. As shown in Figure 7.3 (left), the colony gathers nectar by deploying several thousand foragers across an array of flower patches in the environment. These patches differ in size and profitability, with the profitability of each patch a decreasing function of the total number of bees foraging there. To effectively exploit this array of resources, the colony must continuously gather information about the flower patches.

Every time a forager bee returns to her hive from a patch of flowers, she brings home not only food, but also information about her food source. She can share this knowledge by means of the *waggle dance* communication behavior (Seeley, 2010). Hence the colony acquires information from all its members that are actively gathering food (called *employed foragers*).

To obtain foraging success in a rapidly changing flower market, the colony must receive both updates on old food sources and reports on promising new ones. This is achieved by keeping two distinct type of employed foragers: *recruits* (or *exploiters*) and *explorers* (Seeley, 1996).

At any given moment nearly all foragers returning to a hive are employed foragers that have been exploiting food sources discovered hours or days before. Only a small fraction of the returning have been exploring for new food sources. These explorers come from the rank of unemployed foragers in a colony. Unemployed foragers are bees that need to locate a forage site. Most of the unemployed foragers follow the recruitment dances of their nest mates to find a forage site; hence are recruits. A few of the unemployed search new food sources on their own; hence are explorers. See Figure 7.3 (right).

Clustering strategies inspired by bee division labor

The basic idea behind the allocation of labor in honeybee colony can be applied to the clustering problem of nanobots. In fact it is possible to introduce two different type of nanobots: explorers and recruits. This means that DL strategy implies the use of heterogeneous swarms, like in the

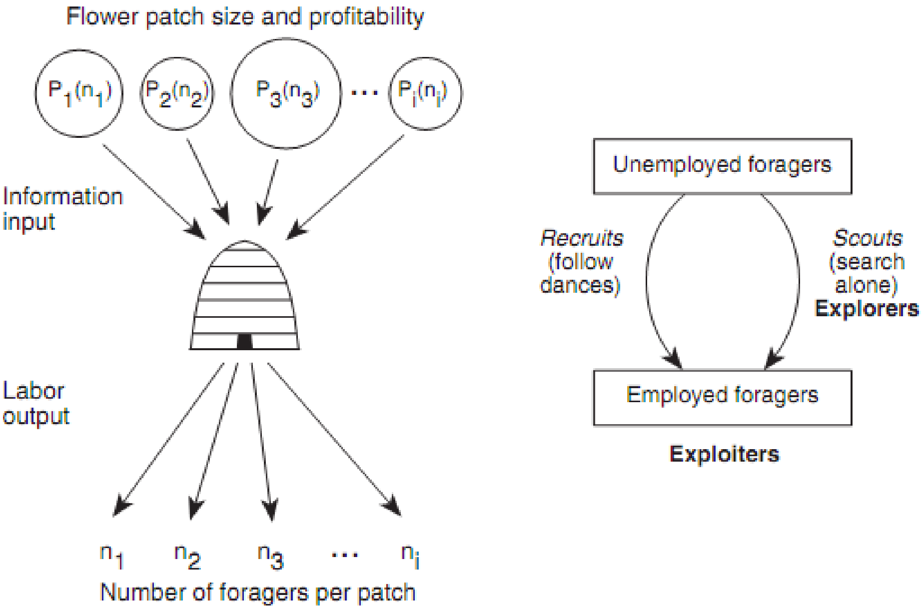


Figure 7.3: (*left*) Labor allocation problem to be faced by a honey bee colony in collecting nectar; (*right*) Explorer and Recruits among a colony's foragers. Taken from (Seeley, 1996)

SWARMANOID project (Dorigo et al., 2012).

The introduction of these two distinguished type of nanobots requires only small modifications to the nanobot architecture outlined in Chapter 5.

Definition 6 (Explorer Nanobot) *An* explorer *is a nanobot with sensors and tail, but without head*

Since an explorer nanobot has sensors, it is able to locate the region with presence of diseased cells; after finding one its anchors to capillary walls, and the presence of the tail allows other nanobots (with head) to form cluster with it by docking. However, not having the head, they are not able to dock to other nanobots, and in particular to anchored nanobots.

Definition 7 (Recruit nanobot) *A* recruit *is a nanobot with both head and tail, but without sensors*

In contrast with explorers, recruit nanobots do not have any sensor, and so cannot locate diseased cells. But they have tail and head, making them suitable for forming cluster once they find already anchored nanobots.

Figure 7.4 shows the finite state machine of the nanobot control system for DL strategy.

7.3 General simulation parameters

All the simulations described in the rest of the chapter share some common parameters.

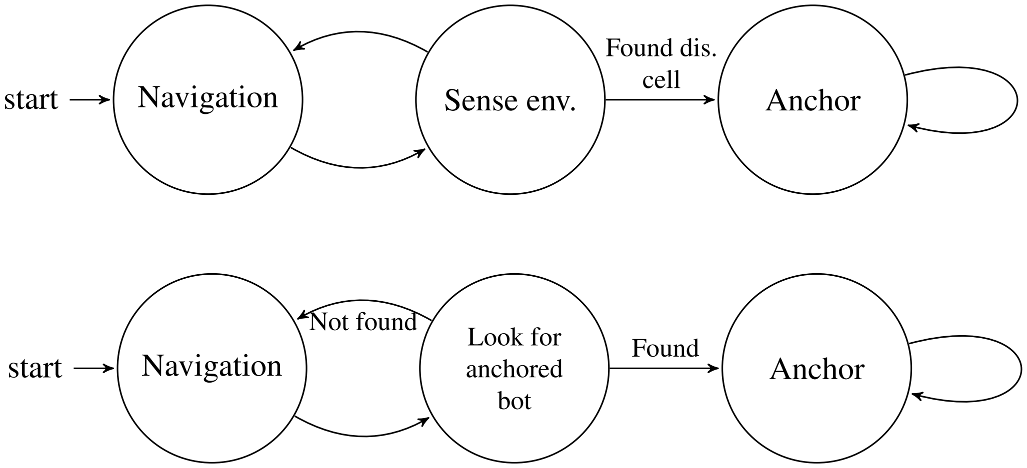


Figure 7.4: Nanobot finite state machine for DL strategy: Explorer (*top*) and recruit (*bottom*)

First of all, while OT and SD strategies have no parameters, the DL one has one: the proportion of explorer nanobots in the swarm. I considered six different values: $0.1, 0.2, 0.4, 0.6, 0.8, 0.9$. Consequently, in each test case 8 strategies are compared: OT, SD, DL0.1, DL0.2, DL0.4, DL0.6, DL0.8 and DL0.9[1].

The major issue to tackle considering the presence of multiple targets is that the nanobots could spread among many different targets. In such cases there will be many cluster (as worst case, each nanobot anchors close to a different target), but each of them will be too small to be identified by an outside unit (or too small to send a signal outside the body). Given this, it is important that at least one cluster is big enough to signal the presence and location of a target, in so doing giving and alarm about the presence of a disease. Consequently the plot representing the clustering performance as a function of the iterations shows the size the of biggest cluster obtained by any particular strategy.

7.4 First setting: Squared region

In the first setting, the search region is a square in which the nanobots freely move. The idea behind this setting is to mimic the use of nanobots in test plate; for example for blood analysis.

Consider a square area of 30×30 (arbitrary units), and that 50 nanobots are introduced in its lower left corner — at coordinate (0,0).

The sensing happens trough direct contact, and we assume that each nanobot can sense a molecule in a range of radius 1. At each time step every nanobot chooses a random direction, and moves 3 units along it. When a nanobot senses the target it stops moving and anchors to its current position.

In the SD and DL strategies these nanobots becomes itself a target and, since a nanobot is much larger than a molecule, it can be sensed in a range of radius 2.

A target is considered successfully identified if a cluster of at least 10 nanobots is formed around it.

---

[1] Here the ratio between the 2 types of bots is rigid, but it is also possible to change this ratio during the exploration; e.g., by introducing in later moments bots of a given type

7.4.1 Single target

As first test case, let's consider the situation in which there is just one target in the squared region. Moreover suppose that the target molecule to be identified is in position (20,20).

Figure 7.5 shows the movements of the bots in the region, starting from the bottom-left corner of the squared region and spreading through the region. For simplicity only the strategy OT, SD and DL0.2 are represented. In this particular simulation OT creates a cluster of 3 bots, SD of 9 bots, and DL0.2 does not create any cluster at all.

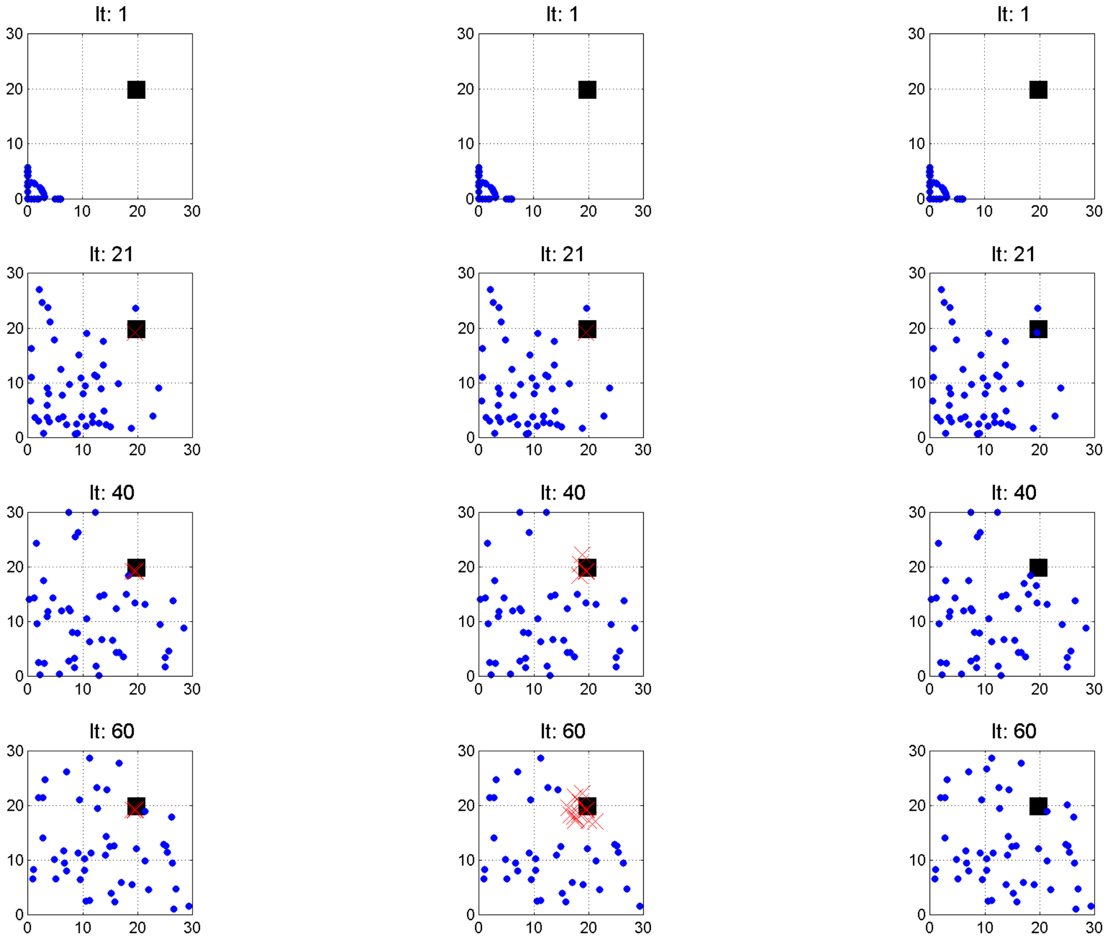


Figure 7.5: Squared region: Cluster (*) of nanobots (·) around the target (■) at different iterations by using the strategies OT (*left*), SD (*center*) and DL0.2 (*right*),

The results obtained by running 1000 simulations in the same condition are represented by the histogram of cluster sizes shown in Figure 7.6. The mean cluster size is 2.39 for OT strategy, 7.45 for SD, and 2.33 for DL0.2. In other words the SD strategy seems to perform consistently better than the other two. Moreover by looking at the histogram, it is possible to notice that not only the means, but also the distributions of cluster sizes are different. In fact OT and DL0.2 have an exponential-like distribution, and most of the time they do not form a cluster at all. While the SD strategy has a binomial-like distribution, with around 50% of the simulations resulting in a cluster size between 5 and 10.

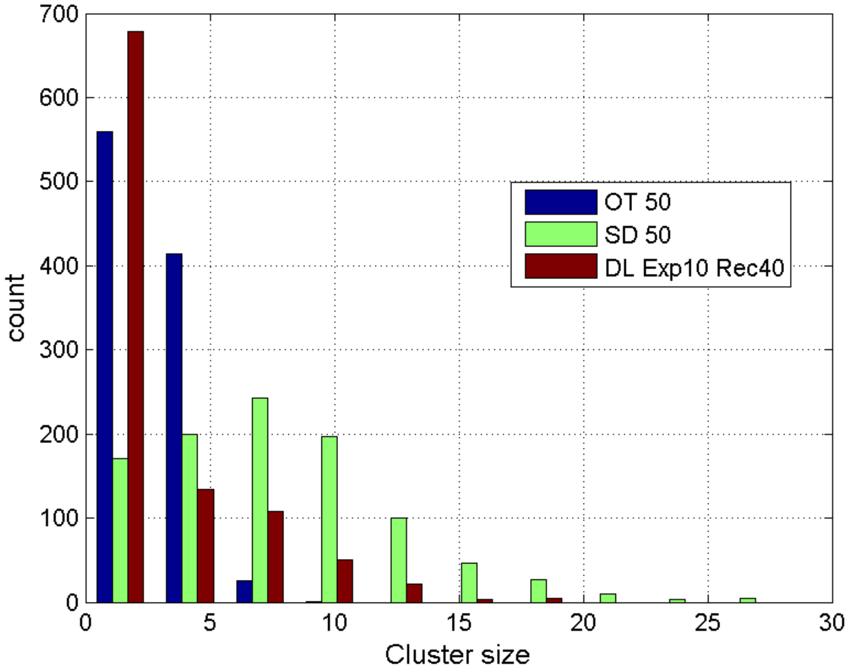


Figure 7.6: One target in squared region. Histogram of cluster size after 100 iterations considering OT, SD and DL0.2

Now let's extend the analysis to all the 8 considered strategies, and also to the effects related to the number of iterations. Again let consider the result of 1000 simulations (each one consisting of 250 iterations).

Figure 7.7 shows how the obtained cluster size changes as a function of the iterations. The SD strategy remains the most effective along the full range of iterations (from 0 to 250). Moreover, although up to 50 iterations all the DL strategies are less effective even of the OT one, they become more and more effective by increasing the number of iterations. This result can be explained by the fact that the DL strategies use less nanobots in searching for the target — a drawback when there is just 1 target.

In particular, after 100 iterations, the DL strategies with many explorers (like DL0.8 and DL0.9) tend to perform like the OT strategy (just slightly better). On the other side, DL strategies with many recruits (like DL0.1 and DL0.2) steeply improve the cluster size; although they are not able to fill the gap with the SD strategy.

To complete the picture, Figure 7.8 shows the histograms of the cluster sizes obtained by running these 1000 simulations. Here it is worth noticing that the cluster size distribution of DL0.1 and DL0.2 are almost uniform for sizes between 1 and 35, with the sole exception of the spike associated to size 0 (the absence of cluster), which accounts for $\simeq 25\%$ of the simulations for DL0.1 and $\simeq 12.5\%$ for DL0.2.

7.4.2 Many targets

In this section the performances of the clustering strategies will be analyzed considering the cases where the number t of targets ranges from 2 to 7. The targets will be distributed inside

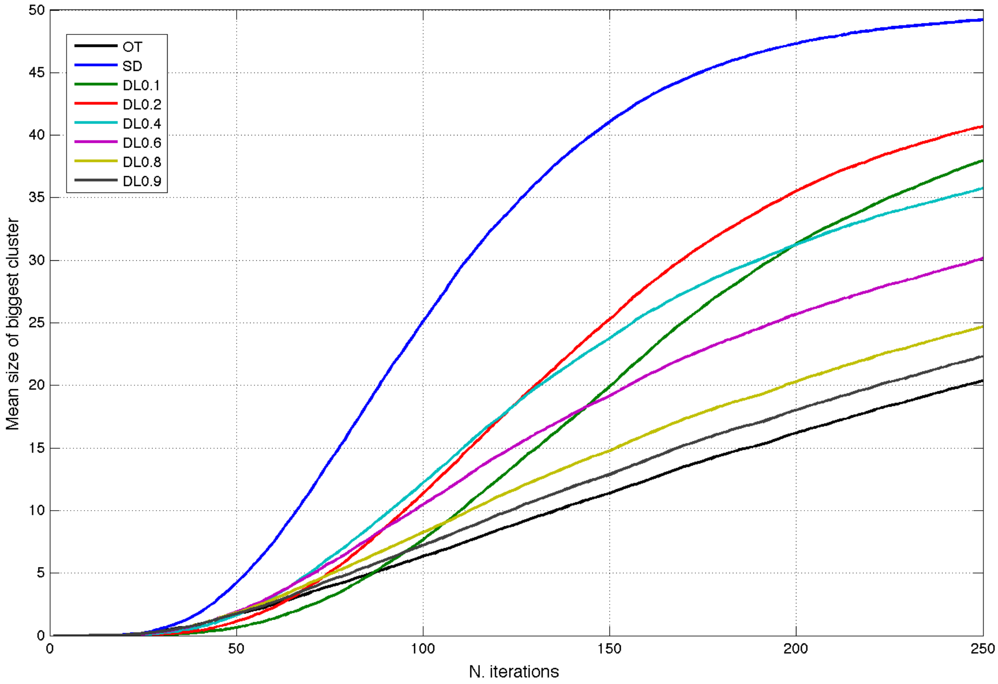


Figure 7.7: One target in squared region: Cluster size obtained as a function of the number of iterations

the region in the following way. For $1 < t < 7$, the targets will be at a fixed distance $d_1 = 20\sqrt{2}$ from the origin, and positioned at angles $\pi/(2t+2), 2\pi/(2t+2), \ldots, (t-1)\pi/(2t+2), t\pi/(2t+2)$. For $t = 7$ the first 4 targets will be positioned as for the case with $t = 4$, while the other 3 targets will be positioned at distance $d_2 = 20$ but following the same rule for the angles. The distribution of the targets for all the considered cases are shown in Figure 7.9.

Just to give an example of the evolution of the simulation with many targets, Figure 7.10 shows the first 60 iterations of the 3-target case. Also in this case for simplicity only the strategy OT, SD and DL0.2 are represented. In this particular example the sizes of the clusters around the three targets are $(5, 5, 1)$ for OT, $(13, 6, 7)$ for SD, and $(6, 1, 0)$ for DL0.2. By Looking at Figure 7.12, these values are confirmed by the average size of the biggest cluster obtained at iteration 60. However, the same figure shows that with the increasing of the number of iterations the size of the biggest cluster obtained by DL0.1 and DL0.2 strategies are greater than those obtained by SD.

The success of the DL strategies over the others (OT and SD) strategies is not peculiar of the 3-target case only. In fact by looking at the Figures 7.11–7.16, representing the cases from 2 to 7 targets, it is possible to find the same result.

All these figures share a common pattern. At the beginning of the iterations the SD strategies produces the greatest biggest-cluster-sizes. But then, there is always a crossover point in which the DL0.1 first and DL0.2 then outperform the SD strategy. This crossover point seems to be correlated with the number of targets; in particular the higher the number of targets, the sooner it happens. In fact for 2 targets the crossover point is at $\simeq 170$ iterations, for 3 at $\simeq 140$, for 4 at $\simeq 120$, for 5 at $\simeq 110$, for 6 at $\simeq 100$, and at last for 7 at $\simeq 80$.

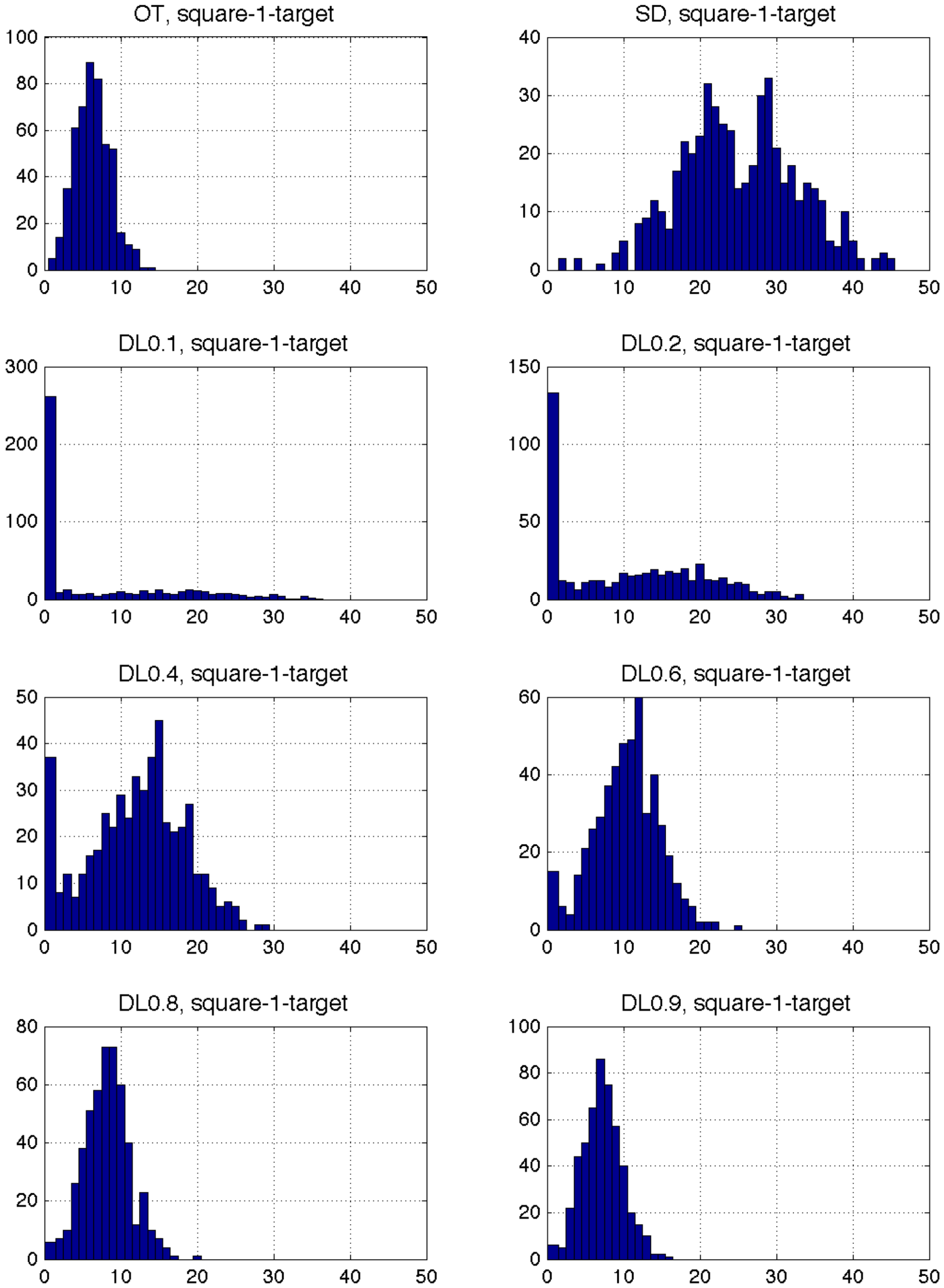


Figure 7.8: One target in squared region. Histogram of cluster size after 100 iterations.

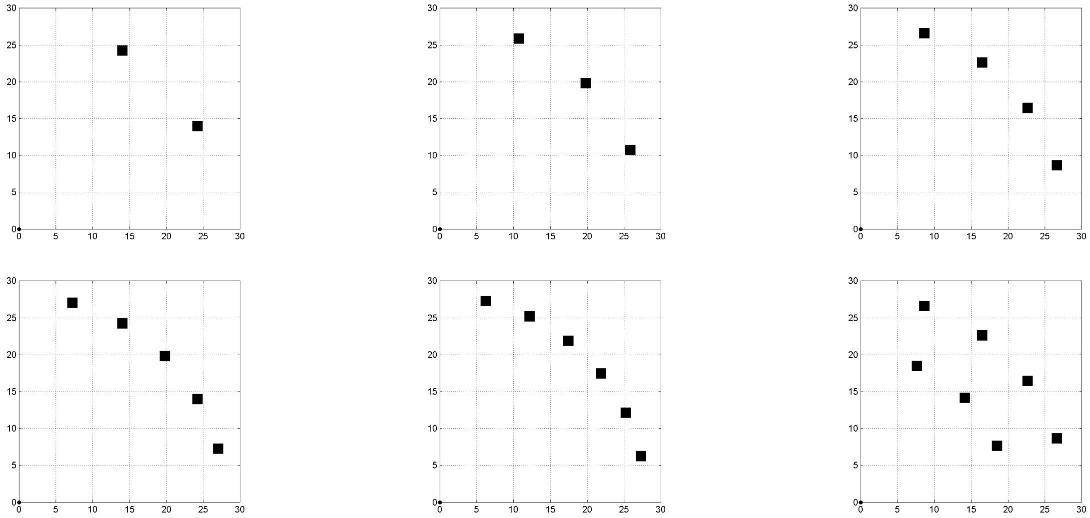


Figure 7.9: Squared region: target (■) positions for $t = 2, \ldots, 7$

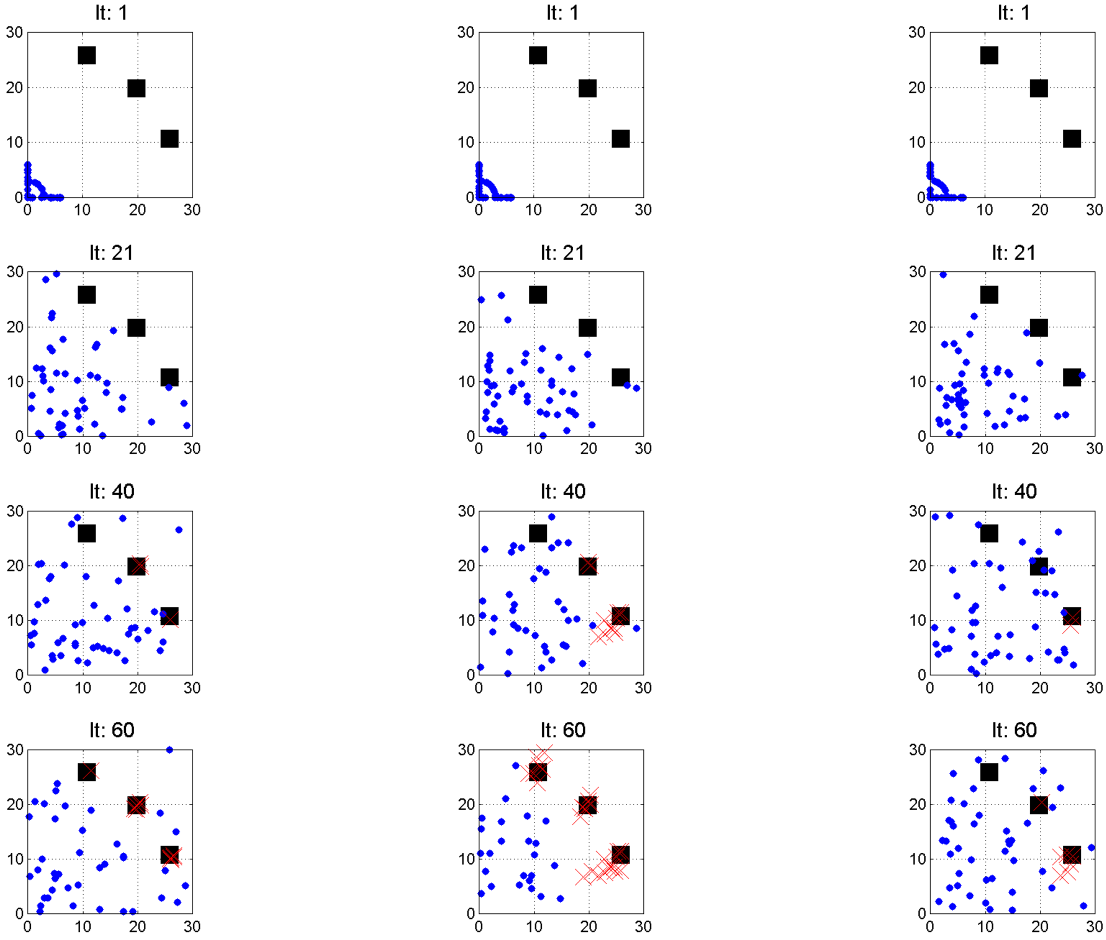


Figure 7.10: Three targets in squared region. Cluster (*) of nanobots (·) around the target (■) at different iterations by using the strategies OT (*left*), SD (*center*) and DL0.2 (*right*),

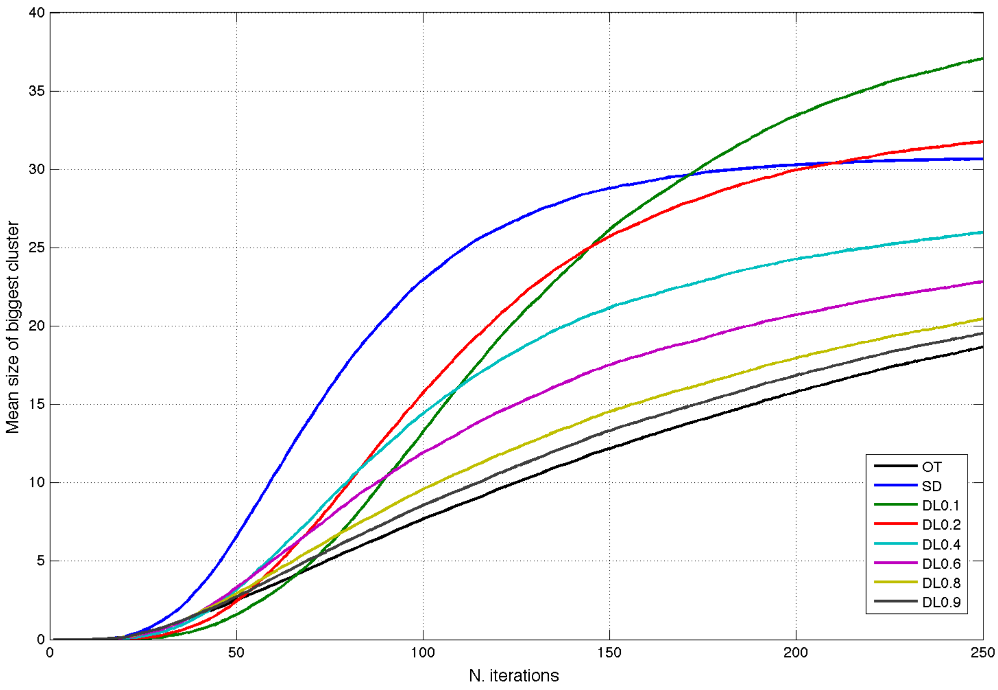


Figure 7.11: Two targets in squared region: Cluster size obtained as a function of the number of iterations

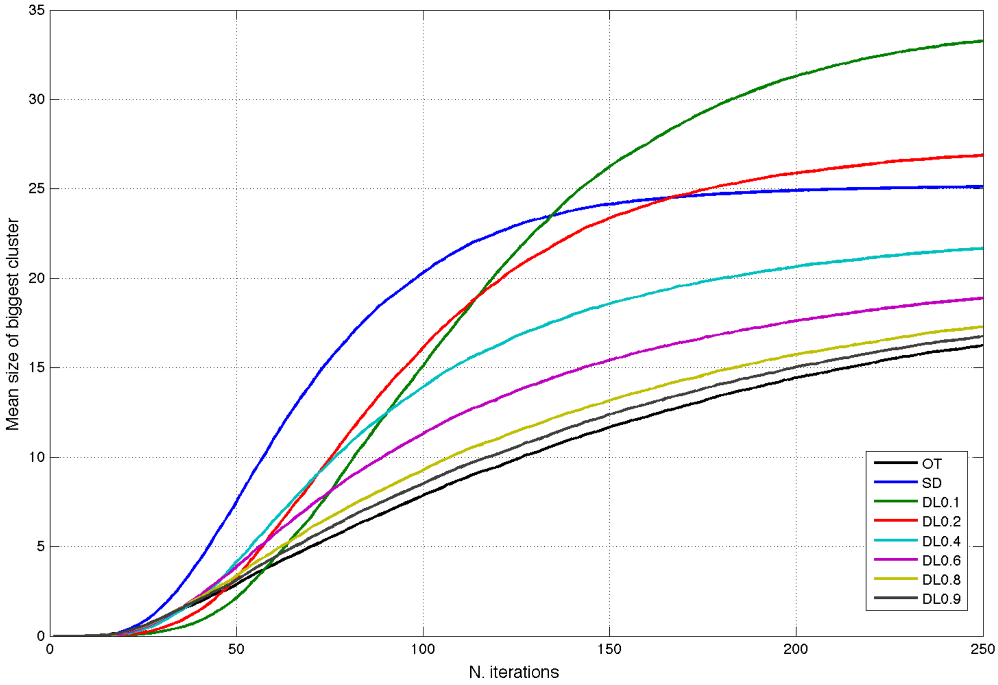


Figure 7.12: Three targets in squared region: Cluster size obtained as a function of the number of iterations

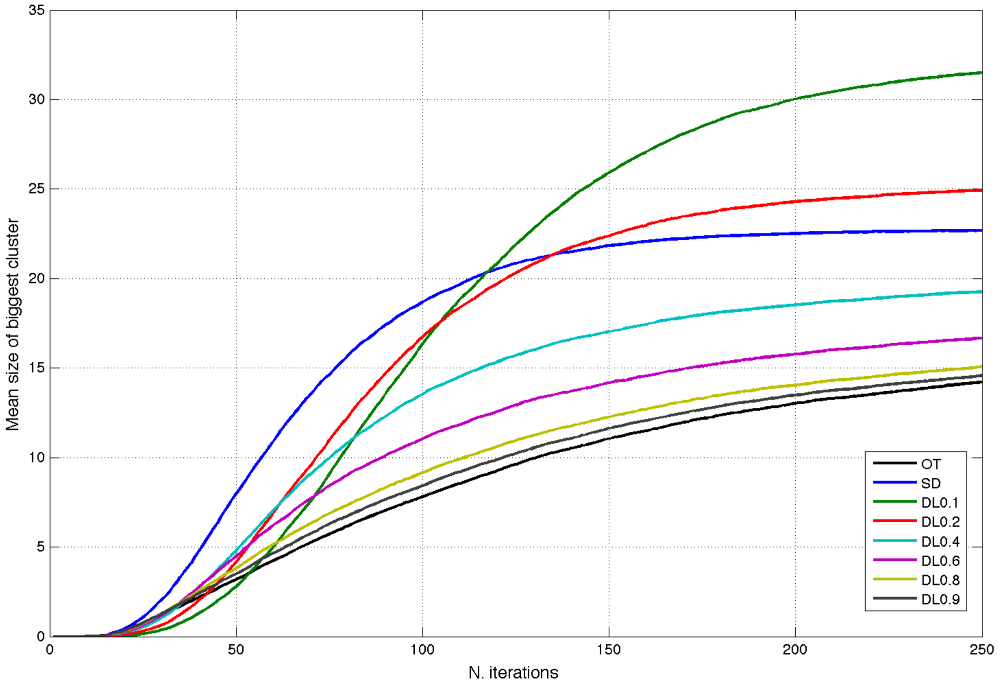


Figure 7.13: Four targets in squared region: Cluster size obtained as a function of the number of iterations

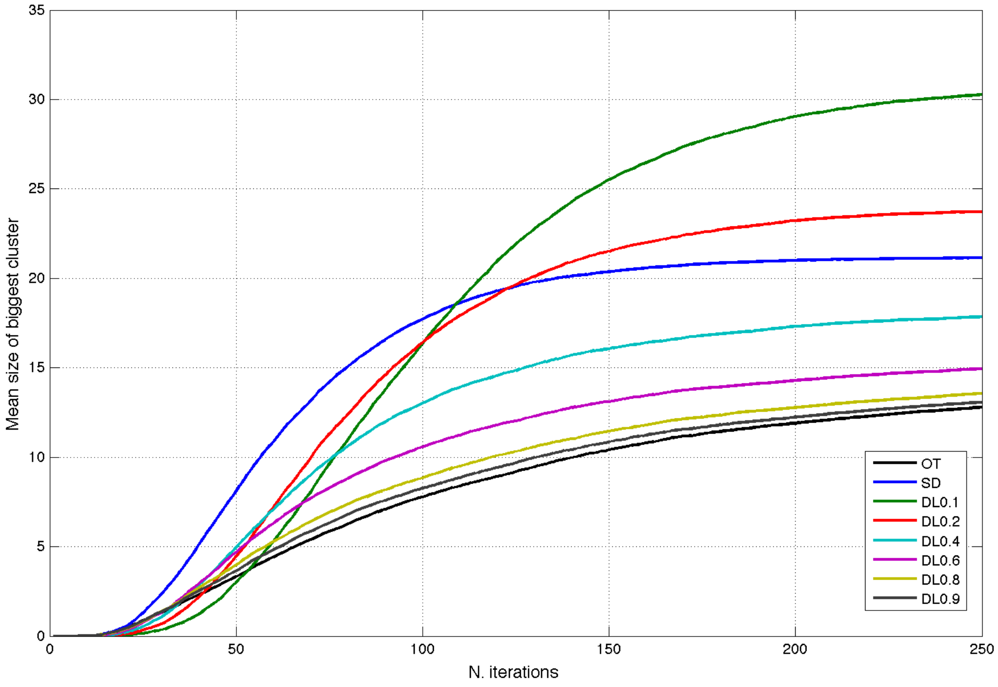


Figure 7.14: Five target in squared region: Cluster size obtained as a function of the number of iterations

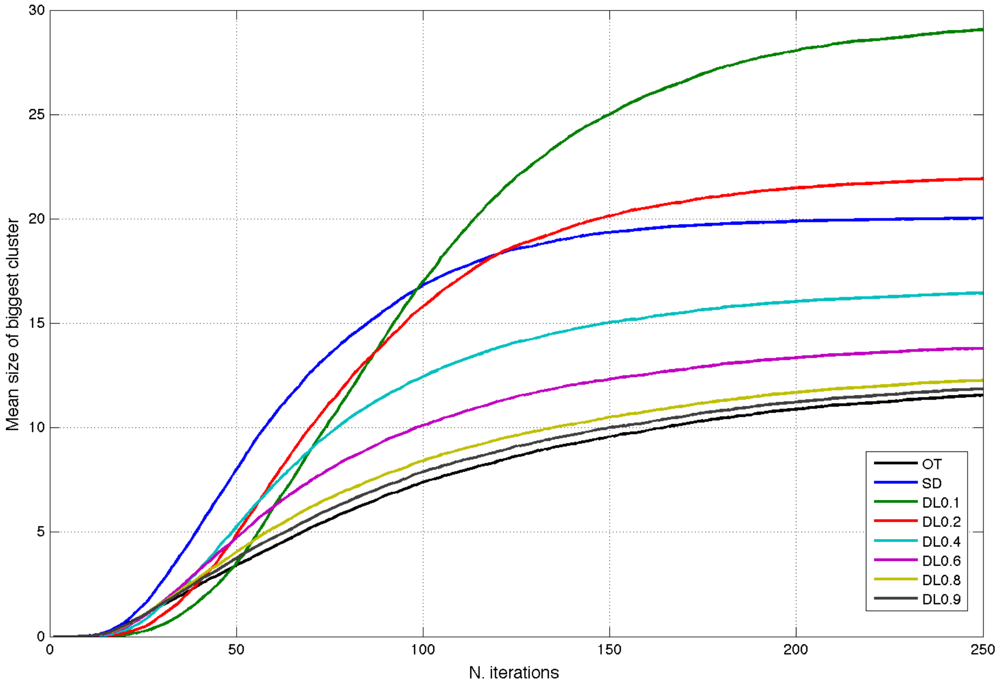


Figure 7.15: Six targets in squared region: Cluster size obtained as a function of the number of iterations

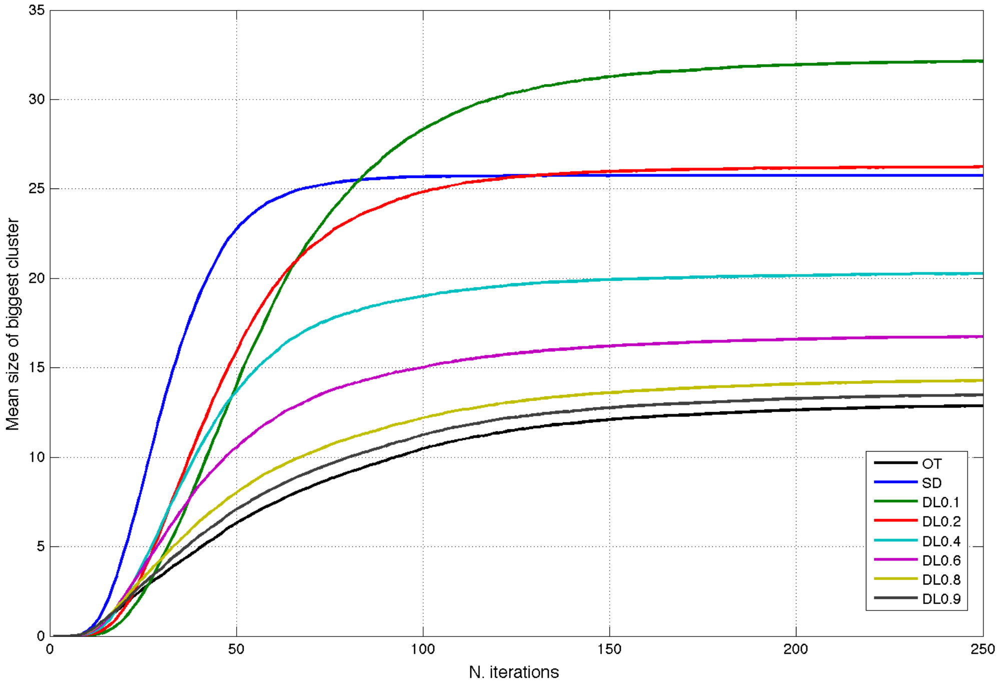


Figure 7.16: Seven targets in squared region: Cluster size obtained as a function of the number of iterations

To complete the picture Figures 7.17 and 7.18 show the histograms of cluster sizes after 100 iterations for the 6- and 7-target cases. In both cases, at such iteration DL0.1 and DL0.2 already took the lead on the size of the biggest cluster.

The fact that DL strategies (especially DL0.1 and DL0.2) tend to perform better with the increasing of the number of targets can be explained by considering the explorers-recruits interactions. Since just a small part of the swarm (the explorers) has the ability to sense new targets (and anchor close to them), it is more difficult for the swarm to spread equally among many different targets. Let's consider the extreme case in which the number t of targets is greater than the number of swarm members n. In such a case, the worst situation for OT and SD strategy is to have every nanobot anchored to a different cluster (thus having n clusters of 1 element each). But for the DLp strategies in the worst situation there are np clusters of $n(1-p)/p$ elements each.

A final note can be made for the 7-target case, where SD strategy quickly converge to a stable value. Since the first 3 targets are closer to the origin, they tend to attract many nanobots and SD rapidly creates a cluster of $\simeq 25$ elements. But at the same time all the other nanobots anchor to other targets. Thus increasing the number of iterations does not increase the size of the biggest cluster.

7.5 Second setting: double tree

In this settings the search space is a double symmetric binary tree, i.e. a simplified model of the systemic circulation as described in Chapter 6.

Figure 7.19 is a schematic representation of the double tree setting. The 2 trees are 2 identical symmetrical binary filiform tree — in these simulation only the tree topology is considered. And they are connected trough their leaves.

Consider a swarm of 50 nanobots, and that they start the search in the aorta (the tree trunk). At each iteration, The nanobots explore the tree by randomly choosing one of the two child branches at each bifurcation. Moreover, to simulate different velocities inside the blood flow, there is a probability p_s that the nanobot stays in the same branch.

The targets can be located only in the leaves of the top tree (the capillaries). If a nanobot lands in a leave containing a target, its has a probability p_d to detect it. After detecting a target a nanobot anchors in that leave. In the SD and DL strategies, every other nanobots passing in the same leave automatically anchors too.

If a nanobot has reached a tree leaf and it did not find any target, at the next iteration step it start to go trough the bottom tree (the venous tree). In this tree no search is performed. The nanobots simply flow through it until they come back to the trunk of the top tree; then it start the search again. This simulate the flow of the nanobots through the venous tree, and the re-entering in the arterial tree.

In the double tree simulations, the depth d of each tree is 6; i.e. they have 64 leaves. Moreover p_s = 1/10, and p_d=1/4. A complete round of the double tree requires the nanobots to traverse $2d$ branches. Since there is a probability p_s to stay in the same branches, on average the complete round needs $2d/(1-p_s)$ iterations. This number is equal to $\simeq 13.3$ with the given values of the simulation parameters.

Many of the reasonings made for the squared region can be applied also to this setting. Thus the rest of the section will proceed at a higher pace through the obtained results. Moreover only

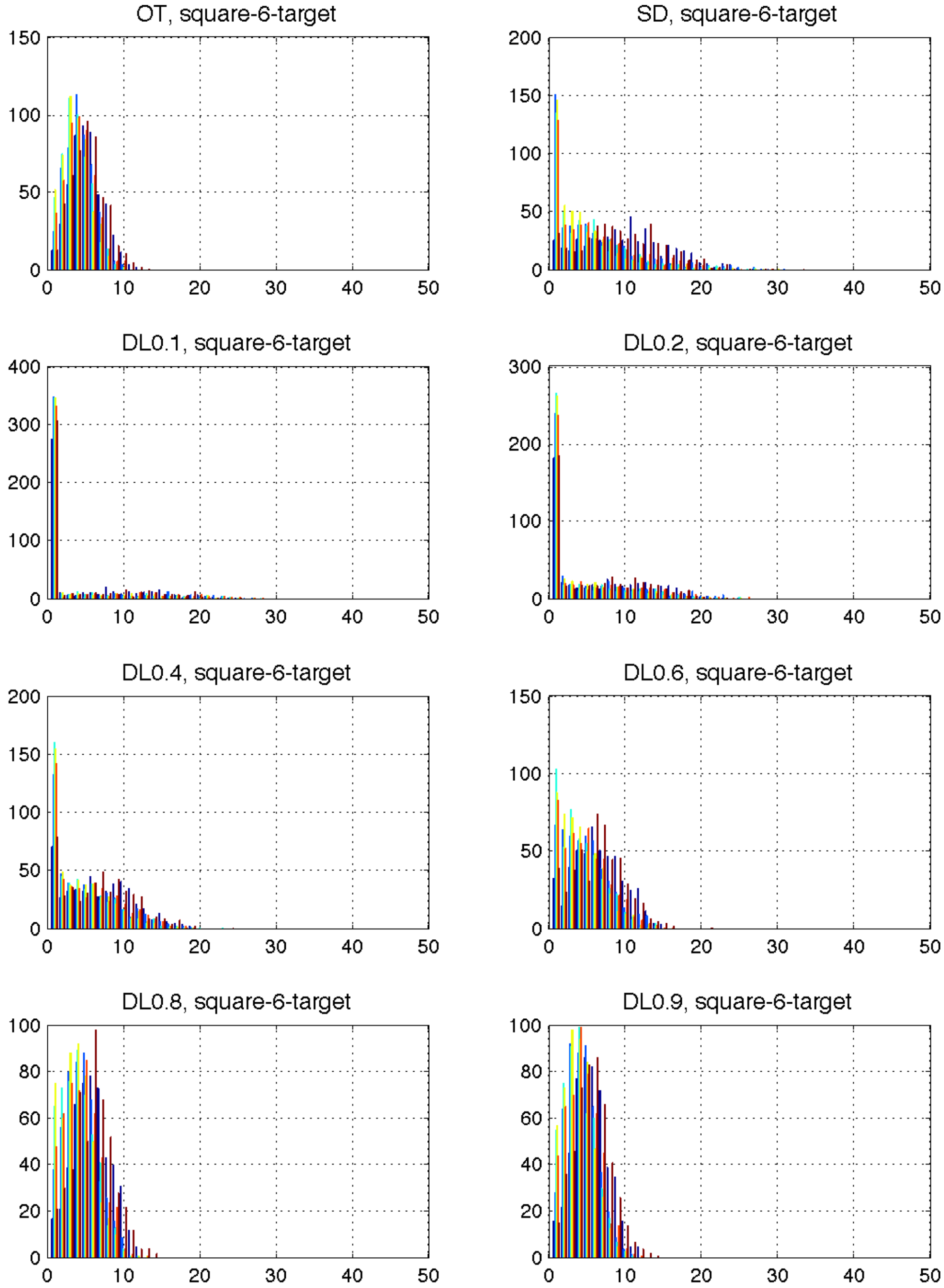


Figure 7.17: Six targets in squared region: Histogram of size of clusters after 100 iterations

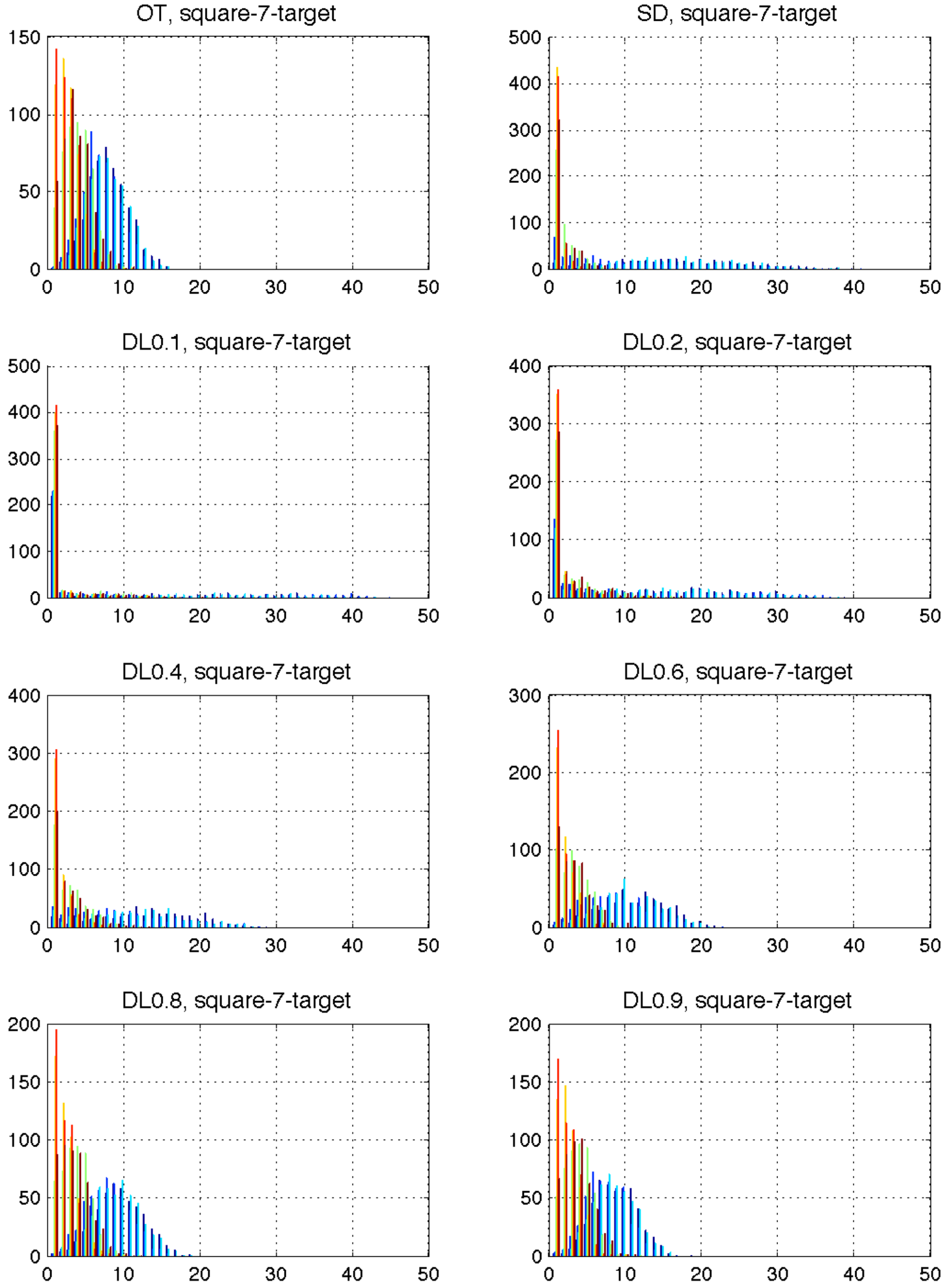


Figure 7.18: Seven targets in squared region: Histogram of size of the clusters after 100 iterations

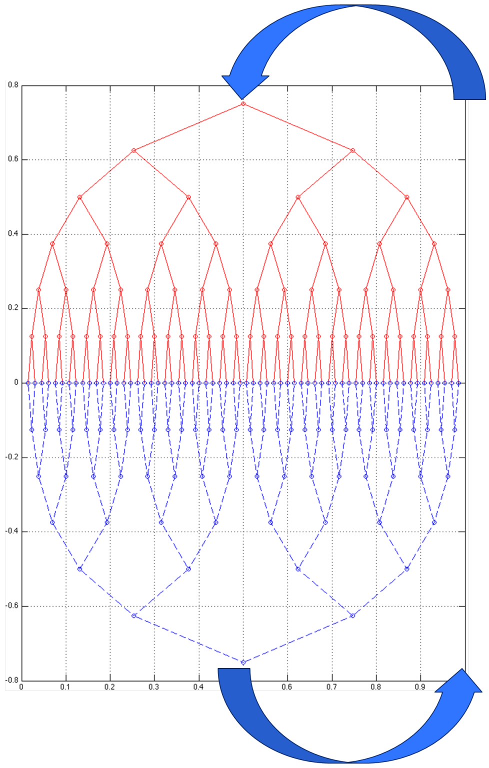


Figure 7.19: Schematic representation of the double tree setting

3 cases will be considered: 1 target, 2 targets, and 64 targets (i.e., a target in each leaf).

7.5.1 Single target

First of all, let consider the situation in which there is just one target in the double-tree region; precisely in the fourth leaf from the left.

Figure 7.20 shows the evolution of the simulation, up to 60 iterations, on a tree of depth 4. The numbers close to the branching points indicate how many nanobots are in that branch at the given iteration. For simplicity only the strategy OT, SD and DL0.2 are represented.

Now let consider the result of 500 simulations, each consisting of 3500 iterations. Figure 7.21 (*top*) shows the mean size of the cluster as a function of the number of iterations. As in the squared region, the SD strategy seems to the best strategy in the presence of just 1 target.

7.5.2 Many targets

Figure 7.22 shows the analysis of 500 simulations for the 2-target case, where each simulation consists of 2000 iterations. The 2 targets are located respectively in the fourth leaf from the left and in the fourth leaf from the right. From Figure 7.22 (*top*), it is possible to notice that at the beginning the SD strategy exhibits the best performance. But after a given number of iterations (here ~ 1200), the strategies DL0.1 and DL0.2 start to dominate. This behavior is consistent with what seen in the sqaured region setting.

In the last considered test case, there is a target in each of the 64 leaves of the tree. Figure 7.23 shows the results of 500 simulations, each consisting of 500 iterations. Here the strategies

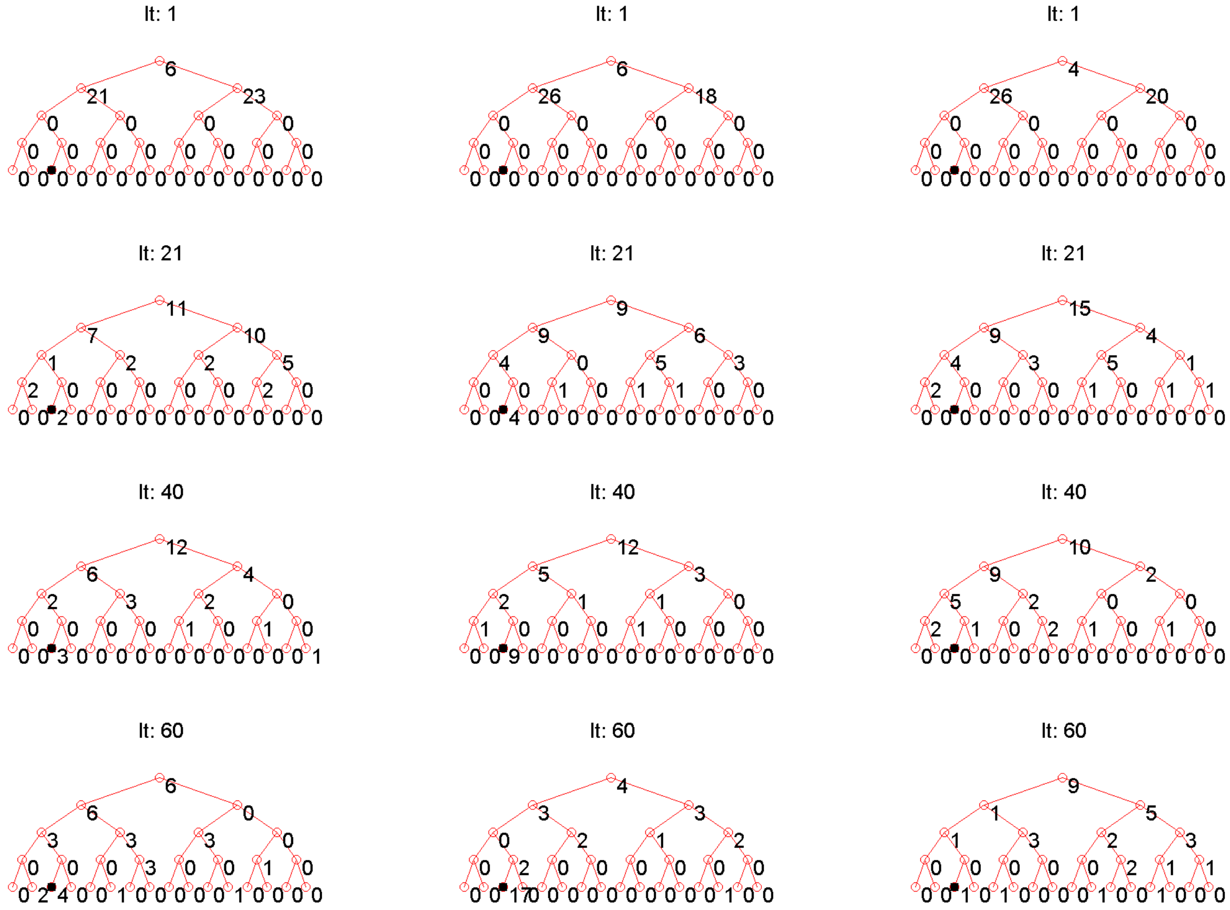


Figure 7.20: Evolution of the simulation in the double-tree region by using the strategies OT (*left*), SD (*center*) and DL0.2 (*right*). The target is indicated as a black square (■)

DL0.1 and DL0.2 become dominant very quickly; already after 60 iterations. Moreover also Dl0.4 first and DL0.6 then perform better that SD strategy. Actually SD and DL0.6 exhibit almost the same behavior.

It is worth noticing that if, for example, 10 members were the minimum size to make the cluster identifiable from the outside, the SD strategy would not issue any alarm (the average cluster size is less than 5)

At last also the crosspoint in which DL strategies start to exceed SD performances seems to follow a pattern similar to the one seen in the squared region.

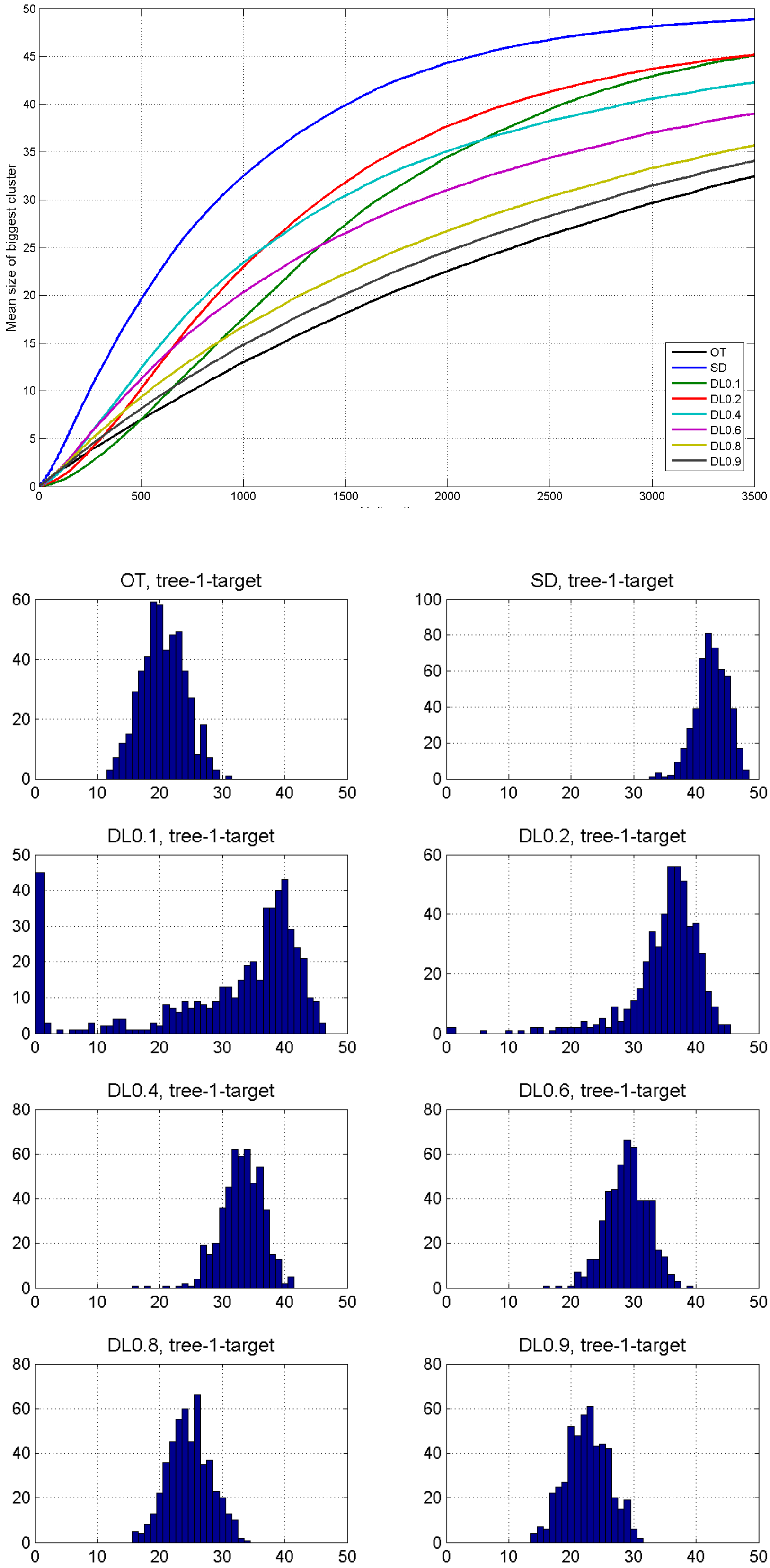


Figure 7.21: One target in double tree: Mean size of cluster as a function of the number of iterations (*top*), and histogram of size of the clusters after 1750 iterations (*bottom*)

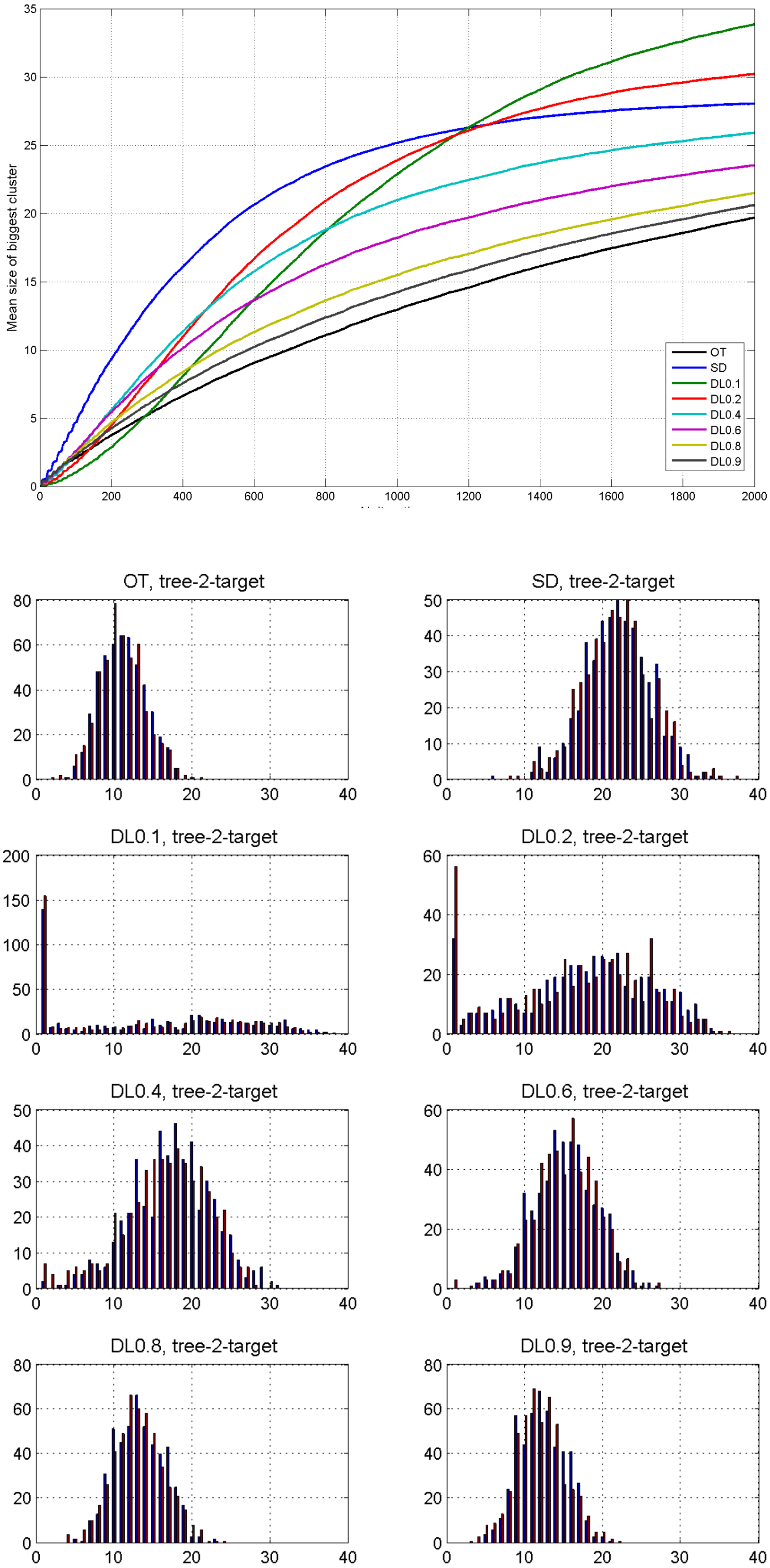


Figure 7.22: Two targets in double tree: Mean size of biggest cluster as a function of the number of iterations (*top*), and histogram of size of the clusters after 1000 iterations (*bottom*)

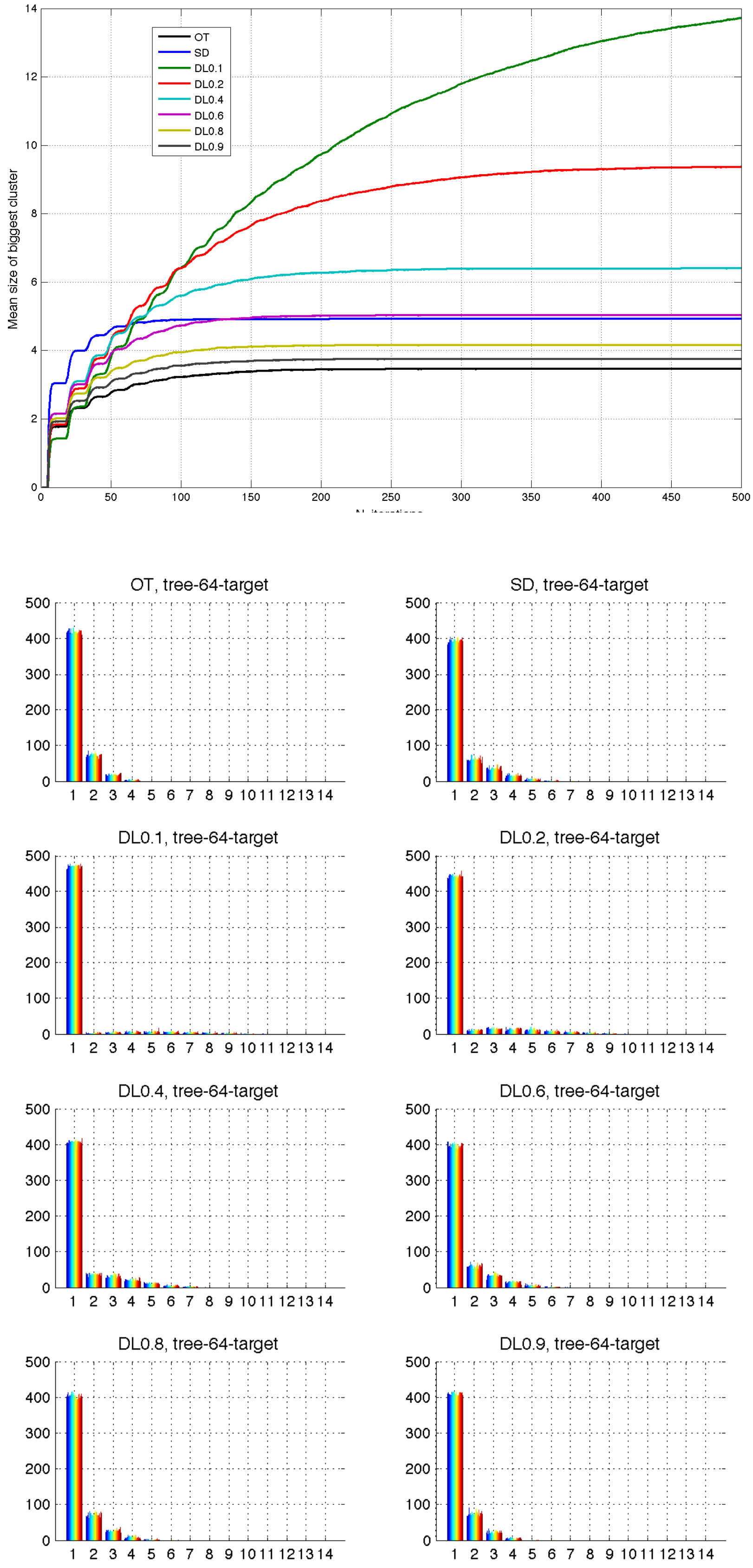


Figure 7.23: Sixty four targets in double tree: Mean size of biggest cluster as a function of the number of iterations (*top*), and histogram of size of the clusters after 150 iterations (*bottom*)

7.6 From double tree scenario to *in vivo* monitoring

The simulations shown so far involve a small swarm (50 members) and a tree with just 6 levels of branching. But the envisioned real applications will involve swarms with hundreds of thousands of members, and the environment will be a tree with around 35 levels of branching. The key questions to answer is how long the swarm will require to form a cluster in this condition, and how the size of the swarm impact this time. For the sake of simplicity, here the analysis will be restricted to the case of just 1 target.

7.6.1 How to estimate OT and SD cluster sizes

First of all, Figure 7.24 shows the results of the 1 target case in a double tree scenario where the depth of tree is 10 and the number of the bots is 30. Changing in this way the parameters of the simulation, the ratio leaves/nanobots goes from $64/50 \simeq 1.3$ to $1024/30 \simeq 34$. By noticing that the trends are almost the same to those of Figure 7.21 (in particular SD is still the best strategy, while OT tends to become the worst strategy.), I make the hypothesis that also with the real applications parameters the trends will be similar to these.

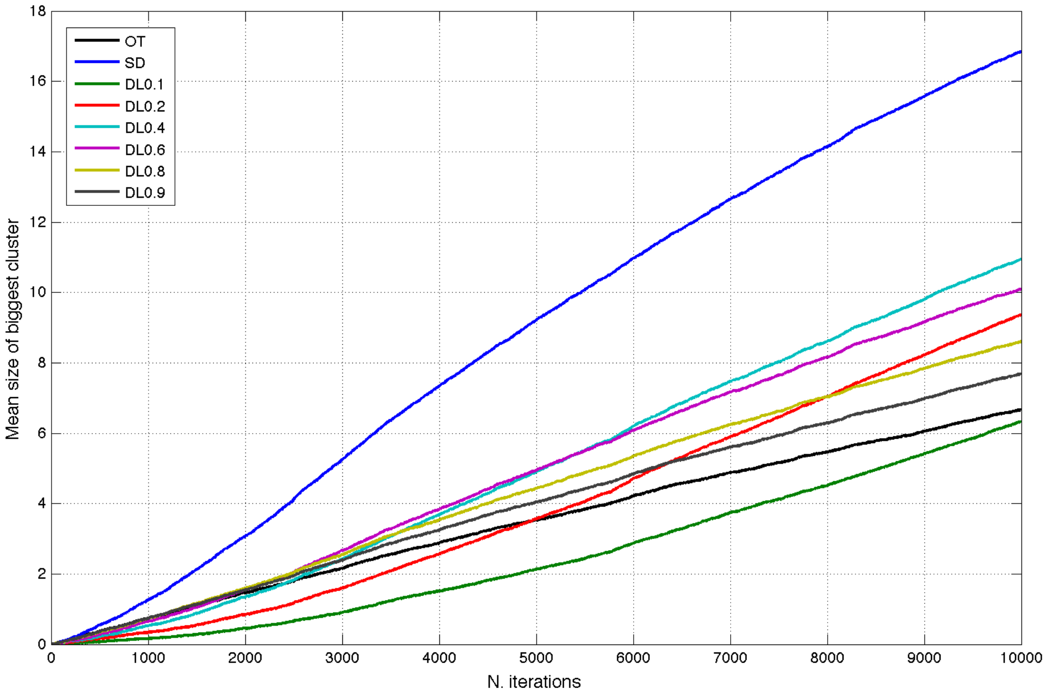


Figure 7.24: One target in double tree: Mean size of cluster as a function of the number of iterations (30 bots, tree of depth 9)

What already changed between the the 2 simulations (Figures 7.21 and 7.24) is the number of iterations needed to achieve a given cluster size. In fact the mean sizes achieved by the different strategies after 500 iterations in Figure 7.21, are obtained after 10^4 iterations in Figure 7.24 — i.e., a factor of 20 times between the number of required iterations.

The performance of the OT strategy in case of 1 target, can be estimated in the following way. The probability p that in a round inside the double tree a nanobot finds the target is:

$$p = \frac{p_d}{l},$$

where l is the number of leaves of the tree, and p_d is the probability to detect the target once a bot is in the leaf containing the target. In the simulations $p_d = 1/4$.

Assuming that the probability to find the target follows a binomial distribution, the average size $c_{\mathrm{OT}}(r)$ of the cluster formed by OT strategy after r complete rounds of the double tree can be approximated by the following equation:

$$c_{\mathrm{OT}}(r) = n\left(1 - \left(1 - \frac{p_d}{l}\right)^r\right), \tag{7.1}$$

where n is the number of bots in the swarm. An overestimate of the average size $c_{\mathrm{SD}}(r)$ of the cluster formed by SD strategy can be obtained in a similar way, by substituting in (7.1) $\frac{1}{l}$ instead of $\frac{p_d}{l}$; i.e., by assuming that each time a nanobot enter in the leaf containing the target, it finds it[2].

$$c_{\mathrm{SD}}(r) = n\left(1 - \left(1 - \frac{1}{l}\right)^r\right). \tag{7.2}$$

The relation between the number of iterations and the number of round is obtained by considering that, in every iteration, each nanobot moves to another branch with a probability $1 - p_s$. Since a complete round across a double tree requires to traverse $2d$ branches (where d is the depth of the tree), on average a round requires $2d/(1 - p_s)$ iterations. Figure 7.25 shows the estimated $c_{\mathrm{OT}}(r)$ and $c_{\mathrm{SD}}(r)$, and the values of OT and SD obtained by simulation. As expected for the OT strategy there is a good agreement, while for $c_{\mathrm{SD}}(r)$ is an upper bound for the SD strategy.

7.6.2 Estimating the time to form clusters in the circulatory system

Now, by using (7.1) and (7.2) and the fact that a complete round in the circulatory system requires 60 s, it is possible to estimate the time needed by a swarm of a given size to form a cluster in the capillary containing the diseased cells. Figure 7.26 shows that, if the swarm has 5×10^6 members, OT would need around 12 weeks to form a cluster of size 10, while SD would need around 4 weeks.

By inverting (7.1) and (7.2), it is possible to compute the number of rounds needed to achieve the cluster size c with a swarm of size n:

$$\begin{aligned} r_{\mathrm{OT}}(c,n) &= \frac{\log\left(1 - \frac{c}{n}\right)}{\log\left(1 - \frac{p_d}{l}\right)} \\ r_{\mathrm{SD}}(c,n) &= \frac{\log\left(1 - \frac{c}{n}\right)}{\log\left(1 - \frac{1}{l}\right)} \end{aligned}$$

Section 5.7.3 describes that a cluster of 10 nanobots should be enough to be spotted by x-ray or to be used as antenna. Figure 7.27 shows the weeks needed to form a cluster of such size

---

[2] This is an overestimation, because in SD strategy the probability to find the target is 1 only *after* at least a nanobot has already found it

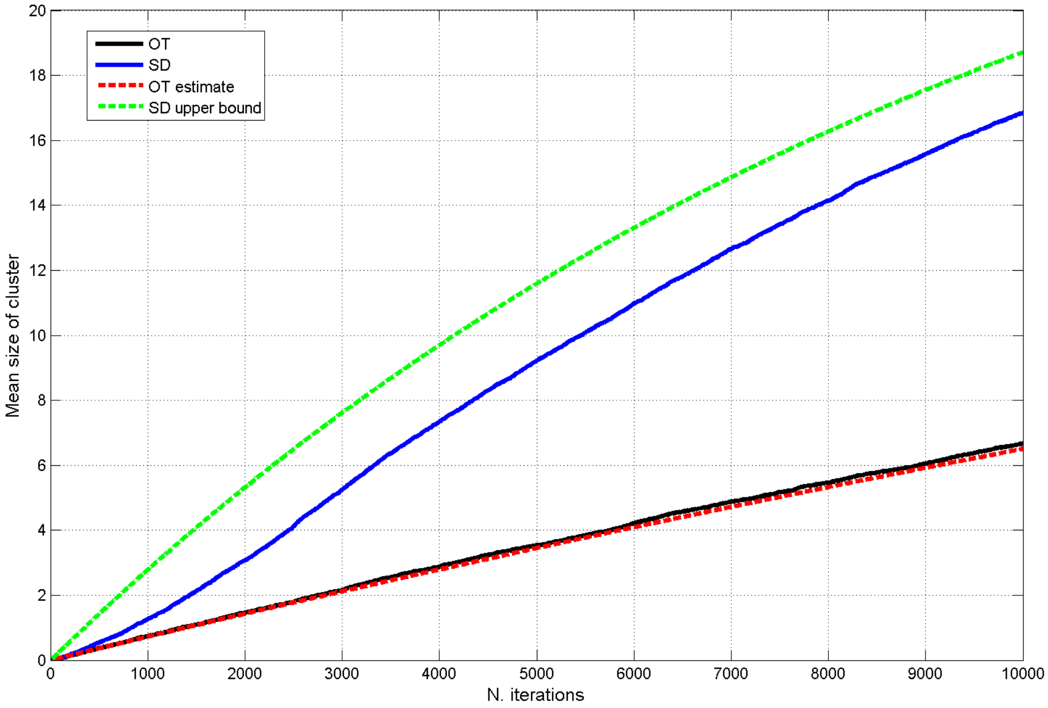


Figure 7.25: One target in double tree: Mean size of cluster as a function of the number of iterations(30 bots, tree of depth 9), and comparison with estimated values

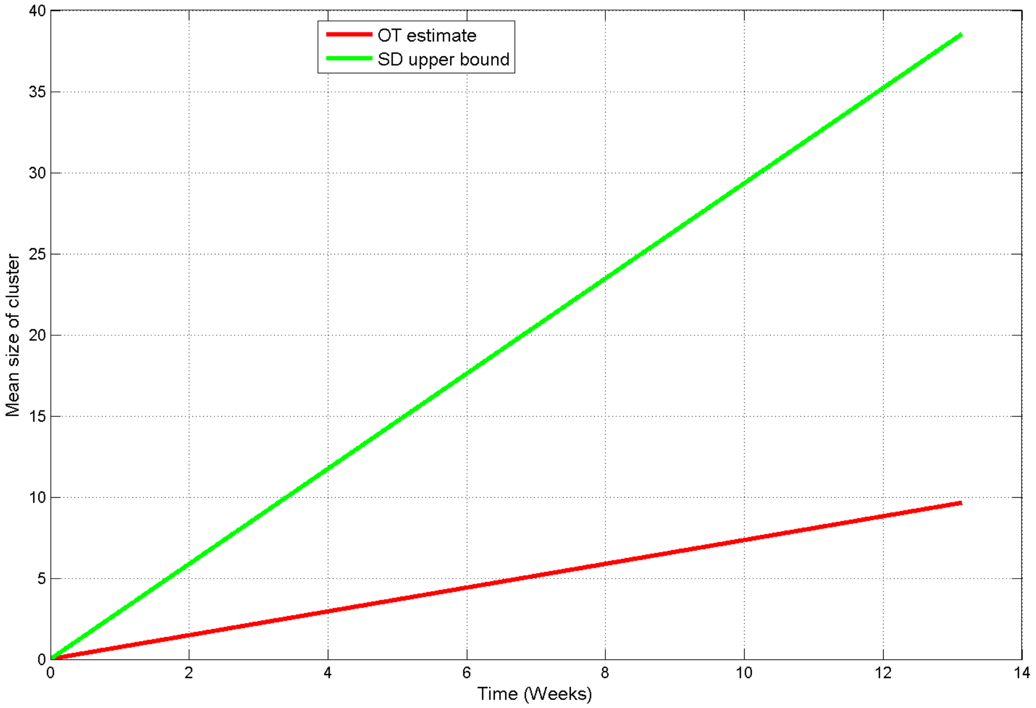


Figure 7.26: One target in double tree: Estimated size of cluster as a function of the time in week (5×10^6 bots, tree of depth 35)

as a function of swarm size. In log-log plot the curves for OT and SD are straight lines with the same slope but different intercepts. A swarm of 10^6 nanobots will need around 20 weeks to form the cluster, while a swarm of 10^7 nanobots will require just a couple of weeks.

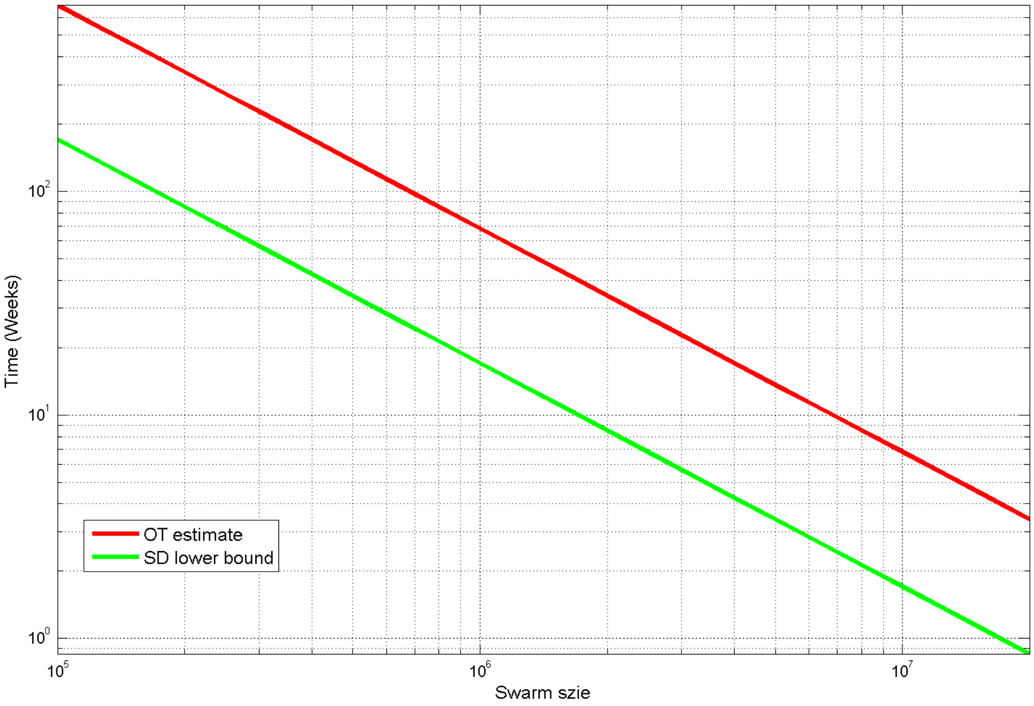


Figure 7.27: One target in a double tree of depth 35: Log-log plot of the time needed to form a cluster with 10 elements as a function of swarm size,

A swarm of 10^7 nanobots is not infeasible in terms of production, cost, and impact on the organism in which it will be injected. This claim is supported by 2 arguments:

- Remembering that the area of the envisioned ellipsoidal nanobot is about 100 μm^2 (see Section 5.6),a 200 mm silicon wafer could host almost 10^8 nanobots. Then, since in mass production phase the cost of a processed silicon wafer is about 1000$, a monitoring system employing a swarm of 10^7 could have a final price of a few thousands dollars; a price in line with that of advanced medical cares
- Assuming that the weight of each nanobot is $\sim 10^{-9}$ g, the total mass the swarm circulating inside the body would be just 10^{-2} g.

Chapter 8

Crossbar technology for nanobots

The main results presented in this chapter have been published in (Cerofolini et al., 2011a).

Nanobots are far from being manufacturable objects. Considering as a framework the technology developed by today (and likely tomorrow) semiconductor industry, a key point to enable nanobots feasibility is the availability of low-cost high-density (in the order of terascale bit per cm^2) integrated circuits. In fact, to implement the on-board control system for all the nanobot functions hundreds thousands (or millions) logic gates will be needed on an area of around 100 μm^2.

The most promising approach for reaching such densities is the *crossbar architecture*. In fact, as mentioned in section 5.6.3, it can enable at the same time higher densities (i.e., more gates for cm^2) and the growth of Carbon Nanotubes (CNT) to be used as sensors.

In this chapter, different techniques to exploit the crossbar architecture are presented.

8.1 Introduction

The evolution of integrated circuits (ICs) has been dominated by the idea of scaling down its basic constituent—the metal-oxide-semiconductor (MOS) field-effect transistor (FET). In turn, this has required the development of suitable lithographic techniques for its definition on smaller and smaller length scales. There are several generations of lithographic techniques, usually classified according to the technology required for the definition of the wanted features on photo- or electro-sensitive materials (resists): standard photolithography (436 nm, Hg *g*-line; refractive optics), deep ultraviolet (DUV) photolithography (193 nm, ArF excimer laser; refractive optics), immersion DUV photolithography (refractive optics), extreme ultraviolet (EUV) photolithography (13.5 nm, plasma-light source; reflective optics); electron beam (EB) lithography (electron wavelength controlled by the energy, typically in the interval $10^{-3} - 10^{-2}$ nm).

The industrial system has succeeded in that, but the cost of ownership has in the meanwhile dramatically increased, because of either the required investment per machine,

$$\text{DUV} \ll \text{immersion DUV} \ll \text{EUV},$$

or the throughput,

$$\text{EB} \ll \text{DUV}.$$

It is just the dramatic cost explosion necessary for the reduction of the feature size that throws doubts on the possibility of continuing the current increase of IC density beyond the next ten years.

Entirely new revolutionary technological device platforms, overcoming the CMOS (complementary MOS) paradigm, must likely be developed to enable the economical feasibility and scalability of electronic circuits to the tera scale integration (TSI).

On another side, the preparation of ICs with bit density as high as 10^{11} cm^{-2} seems now possible, with modest changes in the current production process and marginal investment for the fabrication facility, within a different paradigm. The new paradigm is based on a structure, where a crossbar embodies in each of its cross-points a functional material able to perform by itself the functions of a memory cell (Heath et al., 1998a).

The crossbar is indeed producible (via non-conventional lithography or even without any lithographic method) with geometry on the 10-nm length scale. Although the crossbar is not yet a circuit, it may nonetheless become a circuit if each cross-point contains a memory cell and each of them can be addressed, written, and read—that requires an external circuitry for addressing, power supply, and sensing. The best architecture for satisfying those functions is manifestly achieved embedding the crossbar in a CMOS circuitry (Cerofolini, 2007):

$$\begin{aligned}\text{TSI IC} &= \text{submicro CMOS IC} \cup \\ &\quad\ \text{nano crossbar} \cup \\ &\quad\ \text{nanoscopic cells.}\end{aligned}$$

This architecture reduces the problem of preparing TSI ICs to that of producing nanoscopic memory cells and inserting them into the cross-points of a crossbar structure. It would be a mere declaration of will were it not for the fact that molecules by themselves able to behave as memory cells have been not only designed (Aviram and Ratner, 1974; Joachim et al., 2000; Joachim and Ratner, 2005) and synthesized (Tour et al., 2001; Mendes et al., 2005), but also inserted in hybrid devices (Chen et al., 1999; Reed et al., 2001; Luo et al., 2002; Cerofolini and Ferla, 2002; Green et al., 2007). This fact opens immediately the possibility of a *hybrid route* to TSI ICs:

$$\begin{aligned}\text{hybrid TSI IC} &= \text{submicro CMOS IC} \cup \\ &\quad\ \text{nano crossbar} \cup \\ &\quad\ \text{grafted functional molecules.}\end{aligned} \tag{8.1}$$

In this approach, the transport properties of programmable molecules are exploited for the preparation of externally accessible circuits. Because of this, it is usually referred to as molecular electronics.

The hypothesized TSI IC has thus a hybrid structure, formed by a nanometre-sized kernel (the functionalized crossbar), linked to a conventional submicrometre-sized CMOS control circuitry (producible with currently achievable technologies) and hosting molecular devices (whose production is left to chemistry).

8.2 The crossbar process

A crossbar is nothing but the superposition of an array of n parallel conductive wires on an array of m parallel wires; the arrays are oriented perpendicularly (within a non-critical accuracy) to one another. The $m \times n$ overlapping regions are referred to as *cross-points* and are usually filled with material with desired electrical properties. If this material displays suitable electrical properties (like hysteresis in electrical conductance) the crossbar may open a new paradigm for the design and production of electronic devices.

That self-assembled monolayers on preformed gold contacts may behave as nanoscale memory elements was demonstrated for thiol-terminated π-conjugated molecules containing amino or nitro groups (Reed et al., 2001). The first demonstration of non-volatile molecular crossbar memories employed self-assembled monolayers of thiol-terminated rotaxanes as reprogrammable cells (Luo et al., 2002). The molecules were embedded between the metal layers forming the crossbar via a process that can be summarized as follows:

XB^1, deposition and definition of the first-level (‘bottom’) wire array;

XB^2, deposition of the active reconfigurable molecules, working also as vertical spacer separating lower and upper arrays; and

XB^3, deposition and definition of the second-level (‘top’) wire array.

Figure 8.1 sketches the XB process.

Although potentially revolutionary, the XB approach, with double metal strips, has been found to have serious limits:

- The organic active element is incompatible with high-temperature processing, so that the top layer must be deposited in XB^3 at room or slightly higher temperature. This need implies a preparation based on physical vapor deposition, where the metallic electrode results from the condensation of metal *atoms* on the outer surface of the deposited organic films. This process, however, poses severe problems of compatibility, because isolated metal atoms, quite irrespective of their chemical nature, are mobile and decorate the molecule, rather than being held as a film at its outer extremity (Service, 2003; Stewart et al., 2004a; Lau et al., 2004).

- A safe determination of the conductance state of bistable molecules requires the application of a voltage V appreciably larger than $k_\mathrm{B}T/e$ (with k_B being the Boltzmann constant, T the absolute temperature, and $-e$ the electron charge), say $V = 0.1 - 0.2$ V. Applied to molecules with typical length around 3 nm, this potential sustains an electric field, of the order of 5×10^5 V cm^{-1}, sufficiently high to produce metal electromigration along the molecules (Zhitenev et al., 2006).

- The energy barrier for metal-to-molecule electron transfer is controlled by the polarity of the contact, in turn increasing with the electronegativity difference along the bond linking metal and molecule (Stewart et al., 2004b). The use of thiol terminations for the molecule, as implicit in the XB approach, is expected to be responsible for high energy barriers because of the relatively high electronegativity of sulphur.

Even though the first difficulty can in principle be removed by slight sophistication of the process (for instance, as follows: spin-coating the organic monolayer with a dispersion of metal nanoparticles in a volatile solvent, evaporating the solvent, forming a relatively compact layer via coalescence of the metal particles, and compacting the resulting film by means of an

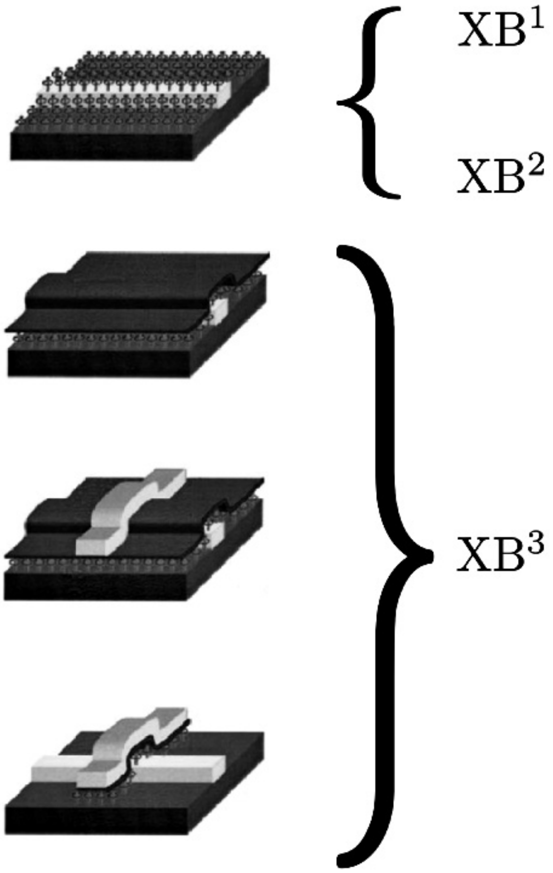


Figure 8.1: The first proposed crossbar-architecture fabrication steps

additional amount of PVD metal), the other difficulties are more fundamental in nature and require different materials.

A solution to the electromigration problem can be achieved preparing the bottom electrodes in the form of silicon wires (as done in (Green et al., 2007)), and the top electrodes in the form conducting π conjugated polymers (as suggested in (Akkerman et al., 2006)):

XB_{+}^{1}, deposition and definition of the bottom array of poly-silicon wires;

XB_{+}^{2}, deposition of the active (reconfigurable) element, working also as vertical spacer separating lower and upper arrays; and

XB_{+}^{3}, deposition and definition of the top array of conducting π-conjugated polymers.

The use of poly-silicon as material for the top array too seems impossible because it is prepared almost uniquely via chemical vapor deposition at incompatible temperatures with organic molecules.[1] The only way to overcome this difficulty consists thus in a process, XB_{*}, where the two poly-silicon arrays defining the crossbar matrix are prepared *before* the insertion of the organic element (Cerofolini and Ferla, 2002). Preserving a constant separation on the nanometer length scale is possible only via the growth of a sacrificial thin film on the first array before the deposition of the second one (Cerofolini and Romano, 2008):

XB_{*}^{1}, preparation of a bottom array of poly-silicon wires;

XB_{*}^{2}, deposition of a sacrificial layer as vertical spacer separating lower and upper arrays;

XB_{*}^{3}, preparation of a top array of poly-silicon wires crossing the first-floor array;

XB_{*}^{4}, selective chemical etching of the spacer; and

XB_{*}^{5}, insertion of the reprogrammable molecules in a way to link upper and lower wires in each cross-point.

The basic idea of process XB_{*}, of inserting the functional molecules after the preparation of the crossbar, is sketched in Figure 8.2.

Of the three considered processes (XB, XB_{+}, and XB_{*}), the one based on double-silicon strips is certainly the most conservative one and is thus expected to be of easiest integration in IC processing. For this reason (and for the possibility of using three-terminal molecules, see Sect. 8.4.3) the attention will be concentrated on the XB_{*} route.

8.3 Non-lithographic preparation of nanowires

The preparation of a crossbar requires the use of simple geometries—essentially arrays of dielectrically-insulated conductive wires. What is especially interesting is that wire arrays with pitch on the nanometer length scale are producible via non-lithographic techniques (NLTs). Not only is this preparation possible, but also the wire linear density already achieved with the the NLTs described in the following is smaller than the one achievable via the most advanced EUV or EB lithographies. Such NLTs exploit the following features:

---

[1] It is instead possible if the spacer is an inorganic film. This case, although not for molecular electronics, is interesting for memories based phase-change materials (Lankhorst et al., 2005) or mimicking the memristor (Strukov et al., 2008).

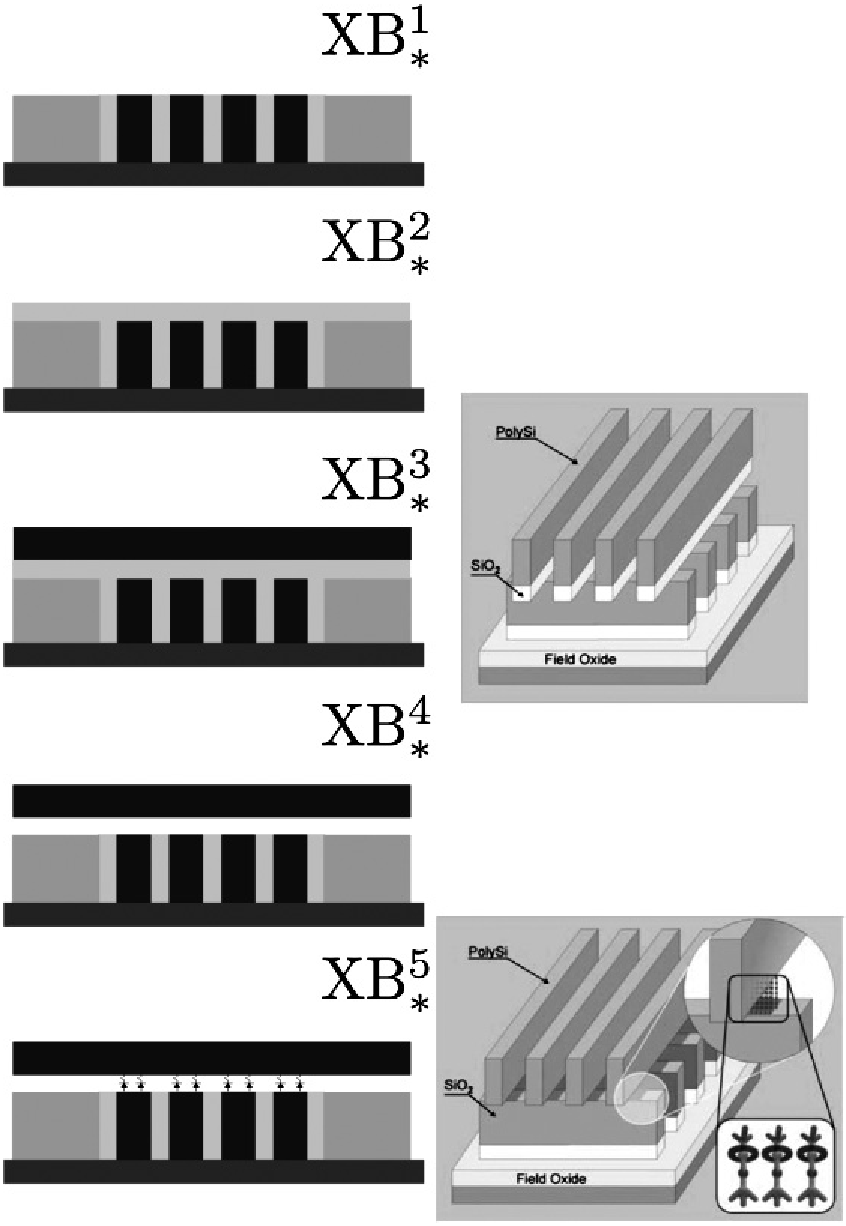


Figure 8.2: The basic idea of XB*: preparing the crossbar before its functionalization

(V), the ‘vertical’ control of film thickness, possible down to the sub-nanometer length scale provided that the film is sufficiently homogeneous; and

(V-to-H), the transformation of films with ‘vertical’ thickness t into patterns with ‘horizontal’ width w:

$$t \xrightarrow{\text{NLT}} w.$$

These techniques are *imprint lithography* and two variants of the *multi-spacer patterning technique*.

8.3.1 Imprint lithography

Imprint lithography (IL) is a contact lithography where properties (V) and (V-to-H) are exploited for the preparation of the mask. The process is essentially based on the sequential alternate deposition of two films, A and B, characterized by the existence of a preferential etching for one (say A) of them. After cutting at $90°$, polishing, and controlled etching of A, one eventually gets a mask formed by nanometer-sized trenches running parallel to one another at a distance fixed by the thickness of B (Natelson et al., 2000; Melosh et al., 2003). For instance, a contact mask for imprint lithography with pitch of 16 nm was prepared by growing on a substrate a quantum well via molecular beam epitaxy, cutting the sample perpendicularly to the surface, polishing the newly exposed surface, and etching selectively the different strata of the well (Melosh et al., 2003). The potentials of the preparation method based on superlattice nanowire pattern transfer (SNAP) are reviewed in Ref. (Wang et al., 2008b).

Actually a number of variants for transferring the pattern to the surface have been developed: molding, embossing, and stamping are the ones most frequently considered (Gates et al., 2005). In one of them (molding), after filling the trenches with a suitable polymer, the mask is used as a stamp, pressing it onto the surface; if the polymer has a higher affinity for the surface than for the mask, the pattern is transferred to the surface when the mask is eventually removed (Gates et al., 2005). The transfer of the polymer to the surface is possible without loss of geometry only if the trench is sufficiently shallow; this implies that the polymer must sustain a subsequent process where it is used as mask for the definition (via directional etching) of the underlying structure with a high aspect ratio. Another method (embossing), sketched in the right hand side of Figure 8.3, involves the pressure-induced transfer of the pattern from the mask to a plastic film, and its subsequent polymerization.

Imprint lithography is generally believed to have potential advantages over conventional lithography because its methods can be carried out in ordinary laboratory with no special equipments (Whitesides and Love, 2007). This situation is expected to make it easy to run along the learning curve to a mature technology. However, very little is known about the overall yield, eventually resulting in production cost, of this process (preparation of mask and stamp, imprint, etching) when the geometries are on the length scale of tens of nanometers.

Imprint lithography has been the matter of extended investigation (see for instance, Refs. (Gates et al., 2005; Whitesides and Love, 2007)) and will not be discussed here. Rather, this chapter is devoted to describe the multispacer patterning technique and to compare its two variants.

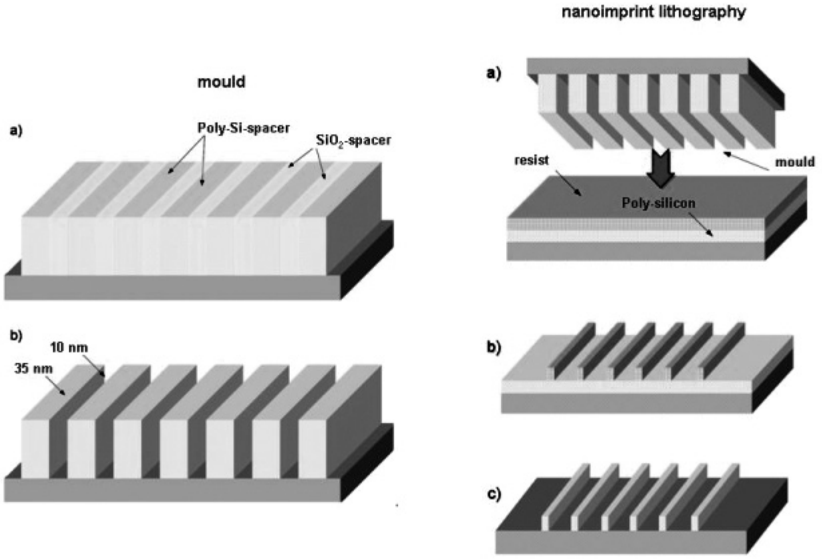


Figure 8.3: Preparation of mold for imprint lithography (*left*) and its use as contact mask (*right*); the multilayer has been supposed to be produced with cycles of sequential depositions of silicon and SiO_2

8.3.2 Spacer patterning technique

The multi-spacer patterning technique (S^nPT) is essentially based on the repetition of the spacer patterning technique (SPT). In turn, the SPT is an age-old technology originally developed for the dielectric insulation of metal electrodes contacting source and drain from the gate of metal-oxide-semiconductor (MOS) transistors.

The SPT involves the following steps:

SPT^0, the *lithographic definition* of a seed with sharp edge and high aspect ratio;

SPT^1, the *conformal deposition* on this feature of a film of uniform thickness; and

SPT^2, the *directional etching* of the film until the original seed surface is exposed.

If the process is stopped at this stage, it results in the formation of side walls of the original seed; otherwise, if

SPT^3, the original seed is removed via a *selective etching*,

what remains is constituted only by the walls of the seed edges. Figure 8.4 sketches the various stages of SPT.

This technique has been demonstrated to be suitable for the preparation of features with minimum size of 7 nm (Choi et al., 2003b; Choi et al., 2003a) and has already achieved a high level of maturity, succeeding in the definition, with yield very close to unity, of nanoscopic bars with high aspect ratio.

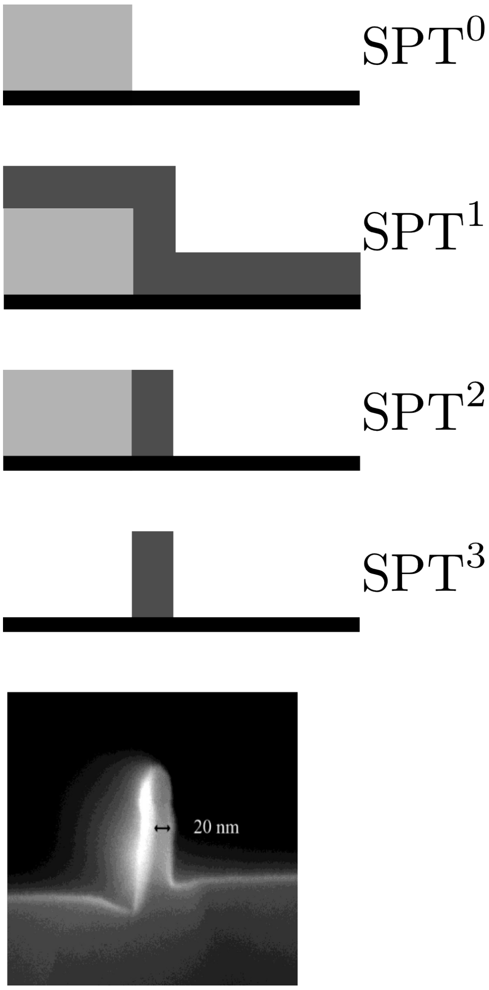


Figure 8.4: (*Up*) The spacer patterning technique: SPT^0, definition of a pattern with sharp edges; SPT^1, conformal deposition of a uniform film; SPT^2, directional etching of the deposited film up to the appearance of the original seed; and SPT^3, selective etching of the original feature. (*Down*) Cross section of a wire produced via SPT.

The SPT may be sophisticated via the deposition of a multilayered film; Figure 8.5 shows the sidewalls resulting from the deposition of a multilayer and compares it with what is really done in the original application of SPT—the insulation of the source and drain electrodes from the gate.

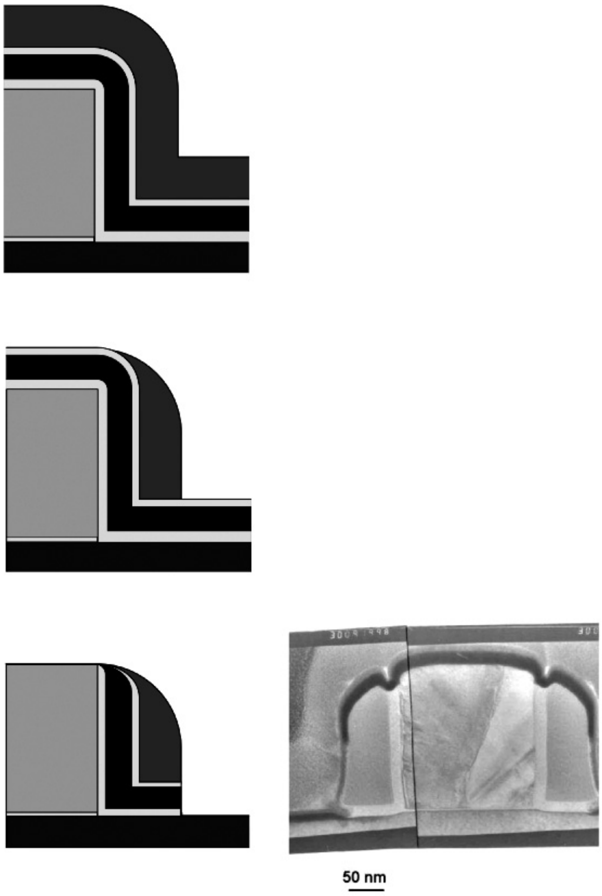


Figure 8.5: The original application of the SPT in microelectronics—dielectric insulation of the gate from source and drain electrodes

8.4 Multi-spacer patterning techniques

Two $\mathrm{S}^n\mathrm{PT}$ routes have been considered: the additive ($\mathrm{S}^n\mathrm{PT}_{+}$) and multiplicative ($\mathrm{S}^n\mathrm{PT}_{\times}$) routes.

The $\mathrm{S}^n\mathrm{PT}_{+}$ is recent and was proposed having in mind the preparation of crossbars for molecular electronics (Cerofolini et al., 2005b; Cerofolini et al., 2005a; Cerofolini, 2005). The $\mathrm{S}^n\mathrm{PT}_{\times}$ is instead much older: The first demonstrators were developed for the generation of gratings with sub-lithographic period (Flanders and Efremow, 1983); recently, however, this technique has been used for the preparation of wire arrays in biochips too (Choi et al., 2003a). Since no detailed comparison of the limits and relative advantages of these techniques is known, the following part will try to understand them on the basis of fundamental considerations.

8.4.1 Additive route—$\mathrm{S}^n\mathrm{PT}_+$

The $\mathrm{S}^n\mathrm{PT}_+$ is substantially based on n STP repetitions where *the original seed is not removed and each free wall of newly grown bars is used as a seed for the subsequent STP*. Each SPT_+ cycle starts from an assigned seed and proceeds with the following steps:

$\mathrm{S}^n\mathrm{PT}_+^1$, conformal deposition of a conductive material,

$\mathrm{S}^n\mathrm{PT}_+^2$, directional etching of this material up to the exposure of the original seed,

$\mathrm{S}^n\mathrm{PT}_+^3$, conformal deposition of an insulating material, and

$\mathrm{S}^n\mathrm{PT}_+^4$, directional etching of this material up to the exposure of the original seed.

The basic idea of the $\mathrm{S}^n\mathrm{PT}_+$ is shown in Figure 8.6: the upper part sketches the process; the lower part shows instead how poly-silicon arrays separated by SiO_2 dielectrics with sub-lithographic pitch (35 nm) can indeed be produced (Cerofolini et al., 2005b; Cerofolini et al., 2005a; Cerofolini, 2005).

The sketch in Figure 8.6 shows a process in which lines are additively generated onto a progressively growing seed, preserving the original lithographic feature along the repetitions of the unit process. The unit process is based on two conformal depositions of uniform layers (poly-silicon and SiO_2) each followed by a directional etching.

Figure 8.7 shows, however, that a similar structure could be obtained by a cycle formed by

$\mathrm{S}^n\mathrm{PT}'_+$, conformal deposition of a bilayer film (formed by an insulating layer deposited before the conductive one—the order of deposition is fundamental), and

$\mathrm{S}^n\mathrm{PT}''_+$, directional etching of this film up to the exposure of the original seed.

Consider an array with pitch P of lithographic seeds, each with width W (and thus separated from one another by a distance $P - W$). Denote with the same symbols in lower case, p and w, the corresponding sub-lithographic quantities. Starting from the said array of lithographic seeds, after n repetitions of SPT_+ any seed is surrounded by $2n$ lines (an example with $n = 4$ is shown in Figure 8.8), so that the corresponding effective linear density K_n of spacer bars is given by

$$K_n = 2n/P.$$

The example of Figure 8.8 shows also that the process can be tuned to preserve the constancy with n of wire width w_n, $w_n = w$, and pitch p_n, $p_n = p$. On the contrary, the spacer heights s_n decrease almost linearly with n,

$$s_n \simeq s_0 - \tau n \tag{8.2}$$

(with τ the spacer height loss per SPT cycle), at least for n lower than a characteristic value $n^{\max}$.

This unavoidable decrease is ultimately due to the fact that the conformal coverage of a feature with high aspect ratio results necessarily in a rounding off of the edge shape with curvature radius equal the film thickness and that the subsequent directional etching produces a non planar surface. Figure 8.9 explains the reasons for the shape of the resulting bars: Even

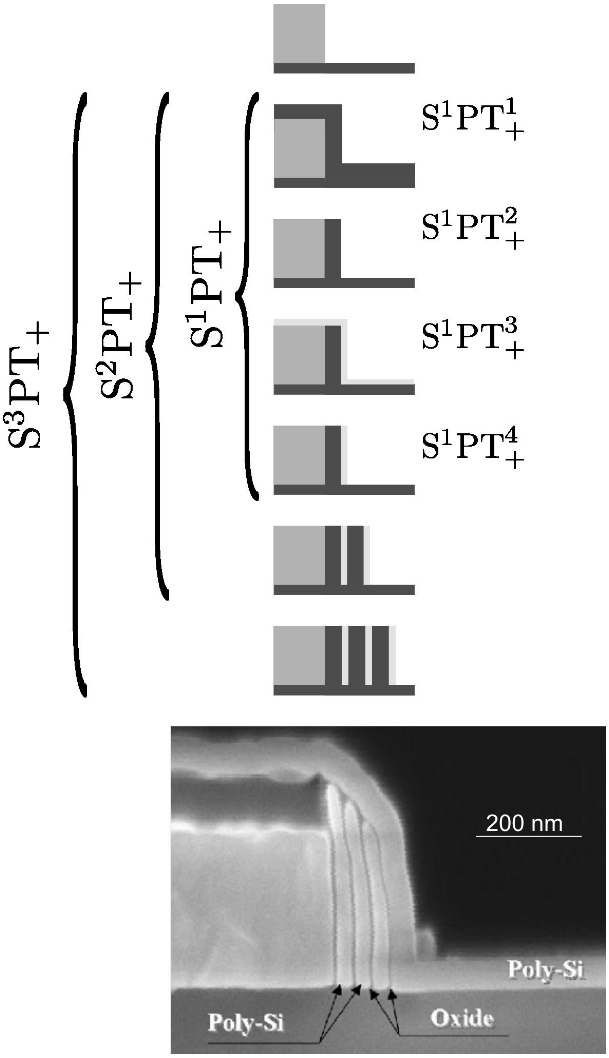


Figure 8.6: *Up*: the additive multi-spacer patterning technique. *Down*: an example of S^3PT_+ multi-spacer (with pitch of 35 nm and formed by a double layer poly-Si|SiO_2) resulting after three repetitions of the SPT_+

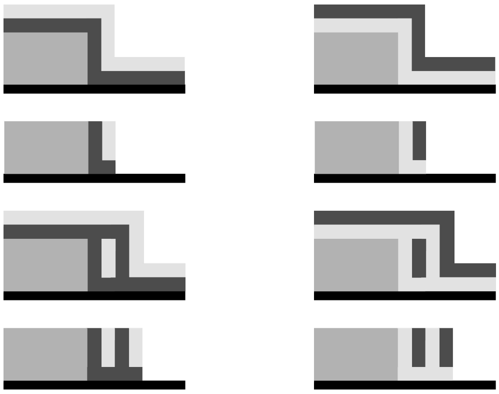


Figure 8.7: A variant of the additive multi-spacer patterning technique able to reduce the number of directional etching by a factor of 2 via sequential deposition of a bilayered film. Two possibilities are considered, consisting in the deposition first of either poly-silicon and then of SiO_2 (*left*), or the same layers in reverse order (*right*).

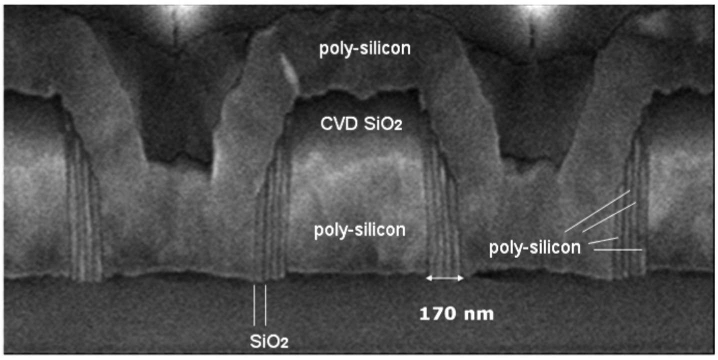


Figure 8.8: An example of S^4PT_+, showing the construction of four silicon bars per side of the seed

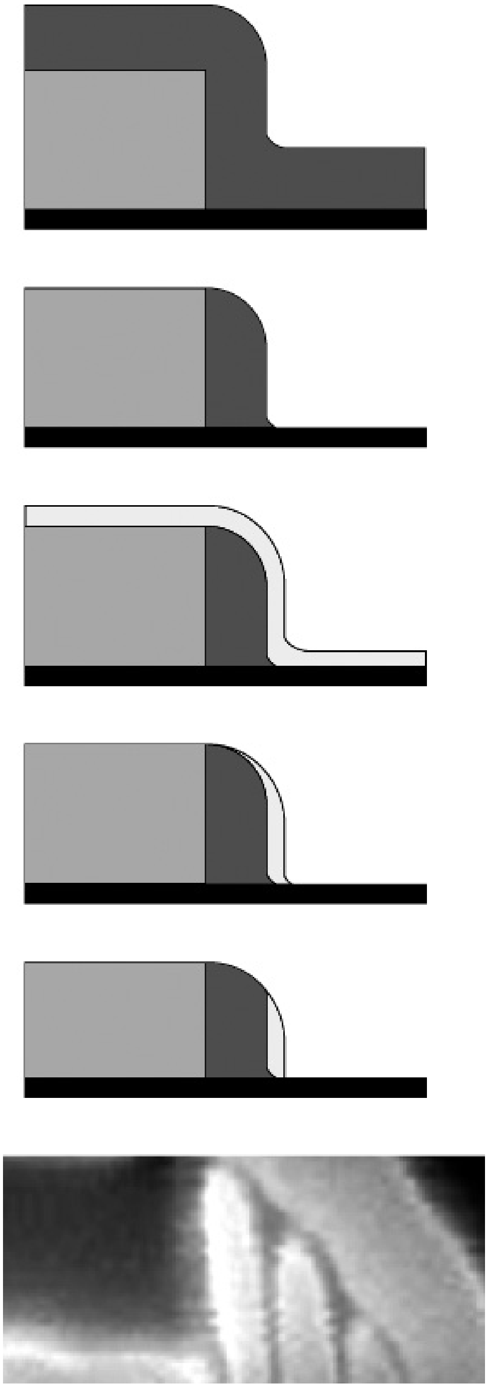


Figure 8.9: Shape of a sidewall resulting after ideal conformal deposition and directional etching (*top*; five sketches) and and an image of how it results in practice (*bottom*; magnification of the spacers shown in Figure 8.6)

assuming a perfect directional attack, the resulting spacer is not flat and there is a loss of height τ not smaller than t: $\tau \geq t$.

Figure 8.6 shows that the process can actually be controlled to have τ coinciding, within error, with its minimum theoretical value, $\tau = 1.0t$. Figure 8.8 shows however that τ depends on the process, and this can be tuned to have $\tau = 3.2t$.

Although the loss of height may seem a disadvantage, in Sect. 8.5.1 it will be shown how the controlled decrease of s_n with n may be usefully exploited.

Assuming the validity of Eq. (8.2) until s_n vanishes, the maximum number $n^{\max}$ of SPT_+ repetitions is given by $n^{\max} = s_0/\tau$; after $n^{\max}$ SPT_+ repetitions, the seed is lost and the process cannot continue further. In view of the availability of techniques for the production of deep trenches with very high aspect ratios, in this analysis s_0 (and hence $n^{\max}$) may be regarded as almost free parameters.

The optimum distance allowing the complete filling of the void regions separating the original lithographic seeds is therefore given by $P - W = 2n^{\max}p$. Hence, the maximum number of cross-points that can be arranged in any square of side P is given by $(2n^{\max})^2$, and the maximum effective cross-point density $\delta_+^{\max}$ achievable with the $\mathrm{S}^n\mathrm{PT}_+$ is given by

$$\begin{aligned} \delta_+^{\max} &= (2n^{\max}/P)^2 \\ &= \frac{1}{p^2}\left(\frac{1}{1 + W/2n^{\max}p}\right)^2 \end{aligned} \tag{8.3}$$

Equation (8.3) shows that $\delta_+^{\max}$ depends on the lithography (through W) and on the sub-lithographic technique $\mathrm{S}^n\mathrm{PT}_+$ (through $n^{\max}$ and p).

Just to give an idea of the maximum obtainable density, Figure 8.6 shows that $p = 35$ nm has already been achieved and $n^{\max} \simeq 10$ is at the reach of the $\mathrm{S}^n\mathrm{PT}_+$; for $W = 0.1$ μm (characteristic value for IC high volume production) and $p = 30$ nm, Eq. (8.3) gives $\delta_+^{\max} \simeq 8 \times 10^{10}$ cm^{-2}. The comparison of this prediction with the lithographically achievable cross-point density (currently of about 2×10^9 cm^{-2}), shows that $\mathrm{S}^n\mathrm{PT}_+$ allows the cross-point density to be magnified by a factor of about 40. This is however achieved only with the construction of 10 consecutive spacers per (bottom and top) layer. The spacer technology is a mature technology with yields close to unity also when employed in more complex geometries than single lines. The increase of processing cost implied by its repeated application in IC processing has bee discussed in Ref. (Cerofolini and Mascolo, 2006). Although this increase is moderate, the integration of so many SPT_+ cycles may however be not trivial and passes through the development of dedicated cluster tools.

It is however noted that the basic idea sketched in Figure 8.7 can be extended to remove at least partially this difficulty: the conformal deposition of a slab with n poly-silicon|insulator bilayers (the insulator being SiO_2, $\mathrm{Al}_2\mathrm{O}_3$,...) followed by its directional etching would indeed result in the formation of $2n$ dielectrically insulated poly-silicon wires. Figure 8.10 sketches the process.

The figure however shows that this dramatic simplification of the process can be done only for the deposition of the bottom array; its use for the top array would result in a distance of the top wire from the lower one varying with the order of poly-silicon layer.

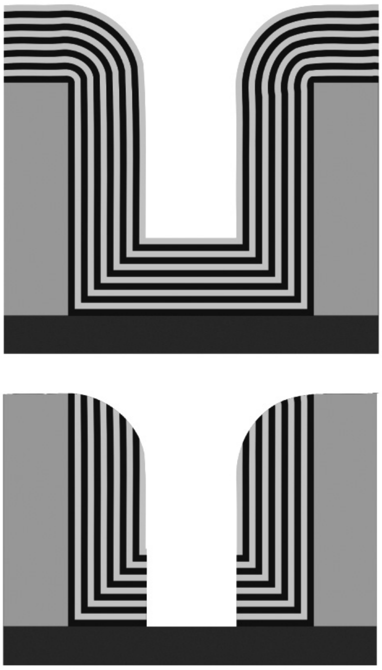


Figure 8.10: Structure resulting after the deposition of a slab of six poly-silicon$|$SiO$_2$ bilayers (*top*) in one shot followed by the conformal attack stopped with the exposure of the original lithographic seed, resulting in an array of 12 sub-lithographic wires (*bottom*)

8.4.2 Multiplicative route—$\mathrm{S}^n\mathrm{PT}_\times$

The SPT allows, starting from one seed, the preparation of *two* spacers (Choi et al., 2003b); in principle, this fact allows another, multiplicative, growth technique—$\mathrm{S}^n\mathrm{PT}_\times$. The multiplicative generation requires that both sides of each newly grown spacer are used as seeds for the subsequent growth—that is possible only if the original seed is etched away at the end of any cycle. In $\mathrm{S}^n\mathrm{PT}_\times$ each multiplicative $\mathrm{SPT}_\times$ cycle involves therefore the following steps:

$\mathrm{S}^n\mathrm{PT}^1_\times$, conformal deposition of a film on the seed,

$\mathrm{S}^n\mathrm{PT}^2_\times$, directional etching of the newly deposited film up to the exposure of the seed, and

$\mathrm{S}^n\mathrm{PT}^3_\times$, selective etching of the original seed.

Figure 8.11 sketches three $\mathrm{S}^n\mathrm{PT}_\times$ repetitions.

Assume that the process starts from a seed formed by an array with pitch P of lithographically defined seeds (lines) each of width W, the linear density K_0 of lines being thus given by $K_0 = P^{-1}$. If lower and upper arrays have the same linear density K_0, the lithographic cross-point density is K_0^2 and the repetition of n (bottom) plus n (top) $\mathrm{S}^n\mathrm{PT}_\times$ results in a sub-lithographic cross-point density $\delta_\times$ given by

$$\delta_\times = 2^{2n} K_0 \qquad (8.4)$$

For any assigned n this density is optimized maximizing K_0, i.e. minimizing P. The minimum value of P is determined by the considered lithography, while W is adjusted to the wanted value

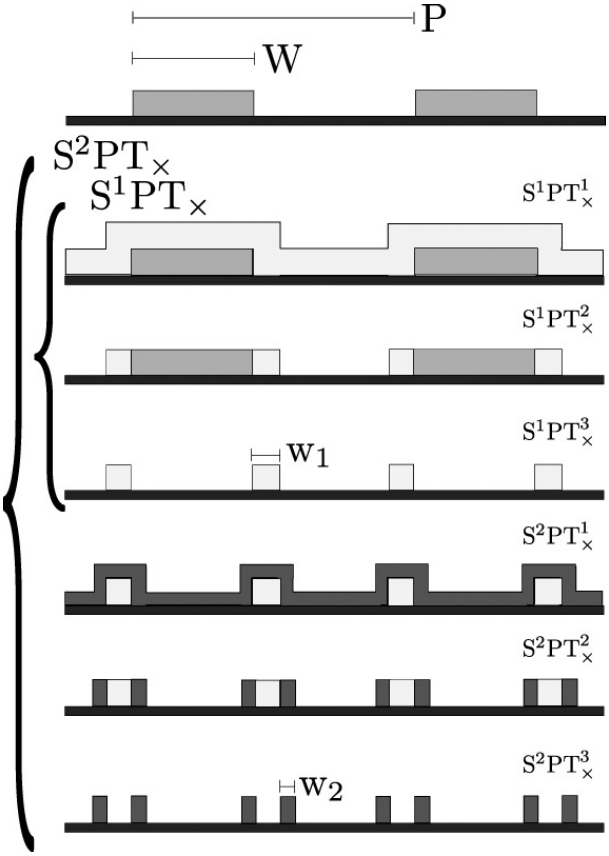


Figure 8.11: Two $STP_\times$ steps for the formation of a sub-lithographic wire array starting from a lithographic seed array

controlling exposure, etching, etc. The appropriate ratio W/P is obtained with the following considerations.

Let the seeds be formed by a given material A (to be concrete we shall think of it as SiO_2) and the $SPT_\times$ be carried out depositing another material B (again for concreteness, we shall think of B as poly-silicon) forming a conformal layer of thickness $t_1 = \varrho_1 W$, with $\varrho_1 < 1$. After completion of the multiplicative SPT cycle, the surface will thus be covered by an array of $2K_0$ wires per unit length each of width $w_1 = t_1 = \varrho_1 W$.

Let the process proceed with the deposition of a film of A with thickness $t_2 = \varrho_2 w_1 = \varrho_1\varrho_2 W$, with $\varrho_2 < 1$ (in the considered example, this process could be the oxidation of poly-silicon to an SiO_2 thickness t_2). After completion of the second $SPT_\times$ cycle, the surface will be covered by a spacer array of linear density 2^2K_0 each of width $w_2 = t_2 = \varrho_1\varrho_2 W$. It is noted that in $S^nPT_\times$ the seed material at the end of each $SPT_\times$ is inverted from A to B or *vice versa*, so that the material of the original seed must be chosen in the relation to the parity (even or odd) of n.

In the following, the attention is addressed to the search of the mask geometry that maximizes the spacer density. After n reiterations of the $SPT_\times$, the spacers will extend both beyond and beneath the original lithographic feature. The zone containing the spacers extends from the edge of the original lithographic feature both into the region separating them and into the region beneath the original feature by amounts l^n_{out} and l^n_{in} given by

$$\begin{aligned} l^n_{\text{out}} &= w_1 + w_2 + \cdots + w_n \\ &= W \sum_{k=1}^{n} \prod_{j=1}^{k} \varrho_j, \end{aligned} \tag{8.5}$$

$$\begin{aligned} l^n_{\text{in}} &= w_2 + \cdots + w_n \\ &= W \sum_{k=2}^{n} \prod_{j=1}^{k} \varrho_j. \end{aligned} \tag{8.6}$$

The estimate of l^n_{out} and l^n_{in} requires one knows the various ϱ_k; at this stage it is impossible to state anything about them. Without pretending to describe the actual technology, but simply to have quantitative (although presumably correct at the order of magnitude) estimates, we assume ϱ_k independent of k, $\forall k$ ($\varrho_k = \varrho$). With this assumption Eqs. (8.5) and (8.6) become

$$\begin{aligned} l^n_{\text{out}} &= W \sum_{k=1}^{n} \varrho^k \\ &= W \frac{\varrho}{1-\varrho} \left(1 - \varrho^n\right), \end{aligned} \tag{8.7}$$

$$\begin{aligned} l^n_{\text{in}} &= W \sum_{k=2}^{n} \varrho^k \\ &= W \frac{\varrho^2}{1-\varrho} \left(1 - \varrho^{n-1}\right). \end{aligned} \tag{8.8}$$

The least upper bounds of l_{out} and l_{in} are obtained taking the limit for $n \to +\infty$ in Eqs. (8.7) and (8.8): $l_{\text{out}} = \varrho/(1-\varrho)$ and $l_{\text{in}} = \varrho^2/(1-\varrho)$; moreover, already for relatively low values of n both ϱ^n and ϱ^{n-1} are negligible with respect to 1 so that we can reasonably assume

$$\begin{aligned} l^n_{\text{out}} &\simeq W\varrho/(1-\varrho), && (8.9)\\ l^n_{\text{in}} &\simeq W\varrho^2/(1-\varrho). && (8.10) \end{aligned}$$

For any W the optimum ϱ is obtained imposing the condition that all the region beneath the original lithographic feature is filled with non-overlapping spacers: $2l_{\text{in}} = W$. Inserting this condition into Eq. (8.10) gives

$$\varrho \simeq \tfrac{1}{2}. \qquad (8.11)$$

Similarly, the optimum size of the outer region is given by the following condition: $2l_{\text{out}} = P - W$. Inserting this condition into Eq. (8.9) gives

$$P \simeq 3W. \qquad (8.12)$$

Reference (Choi et al., 2003a) has demonstrated that three $\text{SPT}_\times$ repetitions on a lithographically defined seed result in nanowire arrays of device quality and suggests that very long wires can indeed be produced with a high yield. However, even accepting that the process have a yield so high as to allow the preparation of non-interrupted wires over a length (on the centimetre length scale) comparable with the chip size, if the lines are used as conductive wires of the crossbar its length is so high to have a series resistance larger than the resistance of the molecules forming the memory cell. This problem was considered in Ref. (Cerofolini, 2007), where it was shown that the crossbar memory can conveniently be organized in modules each hosting a sub-memory of size 1–4 kbits. This implies that each module must be framed in a region sufficiently large to allow the addressing of the memory cells. In this way the density calculated with Eq. (8.4) is an upper value to the exploitable density.

8.4.3 Three-terminal molecules

The use of molecules in molecular electronics is essentially due to the fact that they embody in themselves the electrical characteristics of existing devices. Characteristics of non-linear resistors, diodes and Schmitt triggers have been reported for two-terminal molecules; their use as non volatile memory cells is possible thanks to the stabilization of a metastable state excited by the application of a high voltage (thus behaving as a kind of virtual third terminal). Three-terminal molecules offer more application perspectives non only because they can mimic transistors but also because they could exploit genuine quantum phenomena like the Aharonov–Bohm effect (Cardamone et al., 2006).

The application potentials of three-terminal molecules can however be really exploited only if all terminals can be contacted singularly. This is manifestly impossible using the XB or XB_+ routes, but is possible in the XB_* framework. The major advantage of poly-silicon in the XB_* route is that it does not pose the problem of metal electromigration. However, the multispacer technology can also be adapted for the preparation of nanowire arrays of poly-silicon and metals in arrangements that

- allow the use of three-terminal molecules,
- facilitate the self-assembly of functional molecules, and

- avoid the problem of metal electromigration.

Assume that the top array defining the crossbar is formed by poly-silicon nanowires whereas the bottom array has a more complicate structure. Assume, as sketched Figure 8.12 (a), that each conformal deposition is formed by the following multilayer:

$\mathrm{SiO_2}$ | poly-silicon | $\mathrm{SiO_2}$ | metal | $\mathrm{SiO_2}$ | poly-silicon,

where the metal might be, for instance, platinum obtained via CVD from $\mathrm{PtF_6}$ precursor.

After an SPT, one gets the structure sketched in Figure 8.12 (b); a subsequent time-controlled selective etch of the metal electrode will result in the formation of a recessed region, as sketched in Figure 8.12 (c).

Observe now that thiol-terminated molecules self-assemble spontaneously on many metals (like platinum, gold, etc.) forming closely packed monolayers. Therefore, if the considered three-terminal molecules contain two alkyne and one thiol terminations, they can arrange in the cross-points in an ordered way, allowing their covalent grafting by simple heat treatment, as sketched in Figure 8.13. This figure also explains why the electric field at the metal-sulphur interface may be significantly lower than at the silicon-carbon interface.

8.5 The influence of technology on architecture

Device architecture and preparation procedure are strongly interlocked. This will become especially clear considering that crossbars obtained via $\mathrm{S}^n\mathrm{PT}_+$ may be linked to the external world via methods that cannot be extended to $\mathrm{S}^n\mathrm{PT}_\times$.

8.5.1 Addressing

If the availability of nanofabrication techniques is fundamental in establishing a nanotechnology, not less vital is the integration of the nanostructures with higher-level structures: Once the crossbar structure is formed, it is necessary to link it to the conventional silicon circuitry. This is especially difficult because the nanoworld is not directly accessible by means of standard lithographic methods—“the difficulties in communication between the nanoworld and the macroworld represent a central issue in the development of nanotechnology” (Roukes, 2007).

The importance of addressing nanoscale elements in arrays goes beyond the area of memories and will be critical to the realization of other integrated nanosystems such as chemical or biological sensors, electrically driven nanophotonics, or even quantum computers.

In the following the attention will however be limited to the problem of addressing cross-points in a nanoscopic crossbar structure by means of externally accessible lithographic contacts.

Several strategies have been adopted to attack this problem: Many of them involve materials and methods quite far from, if not orthogonal to, those of the planar technology (Huang et al., 2001; Zhong et al., 2003; DeHon et al., 2003; Likharev and Strukov, 2005; Strukov and Likharev, 2005; Beckman et al., 2005).

The consistent strategies with the planar technology are discussed in Ref. (Cerofolini, 2007). Of them, one can be applied to all crossbars irrespective of their preparation methods. According to this strategy each line defining the crossbar extends beyond the crossing

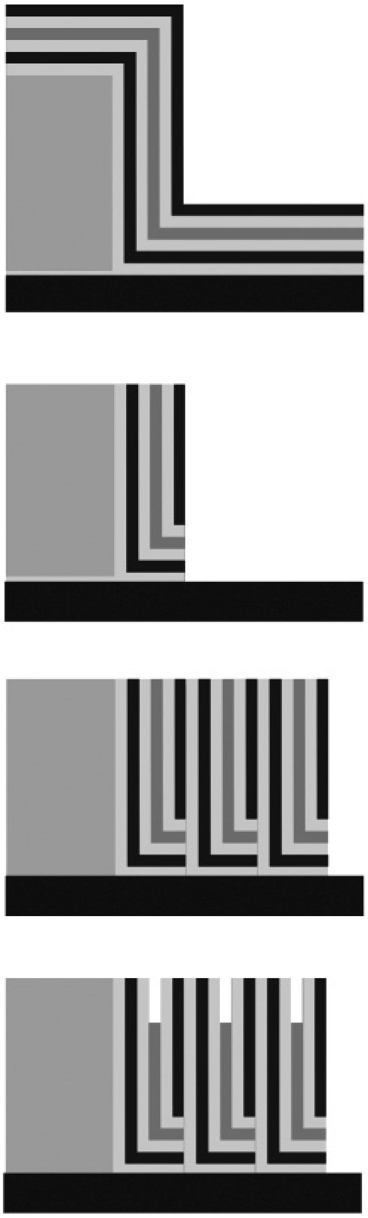


Figure 8.12: The structure resulting after (a) conformal deposition of a multilayer, (b) its directional etching, (c) three repetitions of the above processes, and (d) the time-limited preferential attack of the metal

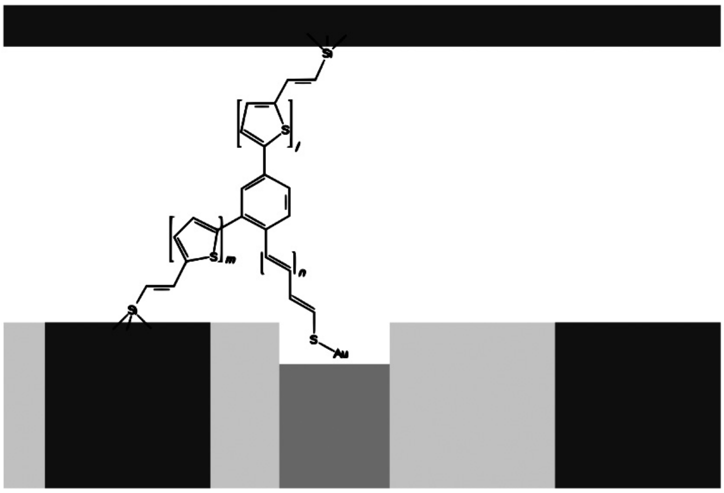


Figure 8.13: The three-terminal molecule after the formation of a self-assembled monolayer on the metal (a) and the grafting to the silicon (b). The equipotential surfaces, reported in the case that metal and bottom poly-silicon electrodes have the same potential, suggests that the metal surface is subjected to an electric stress much lower than the silicon surface.

region and in this zone it is used for addressing. This region is then covered with a protecting cap, that is etched away along a narrow (sublithographic) line misoriented with respect to the array by a small angle α. In this way the zones where the bars are not covered are separated by a distance that diverges for $\alpha \rightarrow 0$; thus, if α is sufficiently small, the separation between the zones no longer protected makes them accessible to conventional lithography and suitable for contacting the CMOS circuitry. In this method each line is linked separately from the others to the external circuitry—addressing n^2 cross-points requires therefore $2n$ contacts.

The multispacer technique (in particular, the $\mathrm{S}^n\mathrm{PT}_+$) permits however novel strategies for the nano-to-litho link in addition to the ones suitable for crossbar prepared with other techniques.

In the first of such strategies the original mask defining the seed is shaped with n indentations with size so scaled that

1, the first indentation is filled, with the fusion of the wires, after the first deposition,

2, the second indentation is filled, with the fusion of the wires, after the second deposition,

$\ldots,$

n, the n-th indentation is filled, with the fusion of the wires, after the n-th deposition.

Taking into account that the minimum distance between the centers of two adjacent fused layers is $W + 3p$ (say 150 nm) and that each contact requires the definition of a hole in a region with side $2p$ (say 70 nm), this technique allows the nano-to-litho link. Figure 8.14 shows the cross section demonstrating how filling indentation with the fusion of the central wires may render them accessible to lithography. Similarly to the strategy considered above, in this method each line is linked separately from the others to the external circuitry, so that addressing n^2 cross-points requires therefore $2n$ contacts.

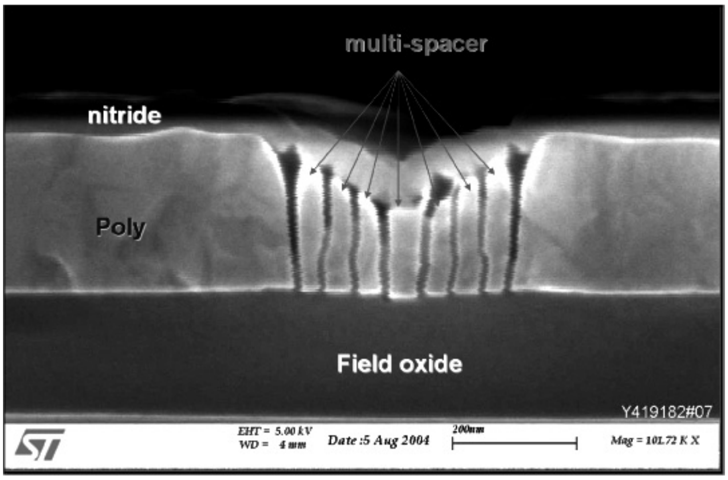


Figure 8.14: Cross section showing the fusion of the arms of the fourth wire grown on two sides of the indentation

The S^nPT_+ technique results wires with different heights. This fact can be exploited to inhibit or enable them by means of dielectrically insulated lithographically defined electrodes in the geometry described in Ref. (Cerofolini, 2007). A crossbar with $n \times n$ cross-points can therefore be addressed by controlling the conduction along the wires from one Ohmic contact to the other by means of $2 + 2$ electrodes. Therefore, for the addressing of n^2 cross-points this method needs $3 + 3$ electrodes only; however, as discussed in Ref. (Cerofolini, 2007), this architecture requires a complex elaboration of the information involving analog-to-digital conversion, with subsequent analysis and elaboration of data.

8.5.2 Comparing crossbars prepared with additive or multiplicative routes

While the repetition in additive way of n SPTs per (bottom and top) layers magnifies the lithographically achievable cross-point density K_0^2 by a factor of $(2n)^2$, the repetition in multiplicative way gives a magnification of 2^{2n}. The magnification factor increases quadratically for the additive way and exponentially with the multiplicative way.

To estimate numerically the process simplifications offered by $S^nPT_\times$ over SPT_+, consider for instance the case of the 3 $SPT_\times$ repetitions per layer. This would produce a magnification of the lithographic cross-point density by a factor of $2^3 \times 2^3$. Taking $W = 0.1$ μm, after 3 $SPT_\times$ repetitions the spacer width should be of 12.5 nm, with minimum separation of 25 nm. Taking into account Eq. (8.12), the cross-point density achievable with the repetition of 2×3 $SPT_\times$ would thus be almost the same as that obtainable with the repetition of 2×10 SPT_+ (7×10^{10} cm^{-2} vs. 8×10^{10} cm^{-2}).

Figure 8.15 shows in plan view a comparison between the following crossbars:

(a) a 2×2 crossbar obtained by crossing lithographically defined lines;

(b) a 16×16 crossbar obtained via S^8PT_+ starting from lithographically defined seeds separated by a distance allowing the optimal arrangement of the wire arrays; and

(c) an 16×16 crossbar obtained via $S^3PT_\times$ starting from lithographically defined seeds separated by a distance satisfying Eq. (8.12).

The figure has been drawn in the following hypotheses:

- the lithographic lines in (a) and (b) have width at the current limit for large-volume production, say $W = 65$ nm;
- the height loss τ is such that the maximum number of repetitions in the additive route is 8, and the sub-lithographic pitch is the same as shown in Figure 8.6; and
- the lithographic width of (c) is chosen to allow the minimum pitch to be consistent with the one obtained with the additive route ($W = 100$ nm), in this way producing sub-lithographic wires with width (12.5 nm) that has been proved to be producible (Choi et al., 2003a).

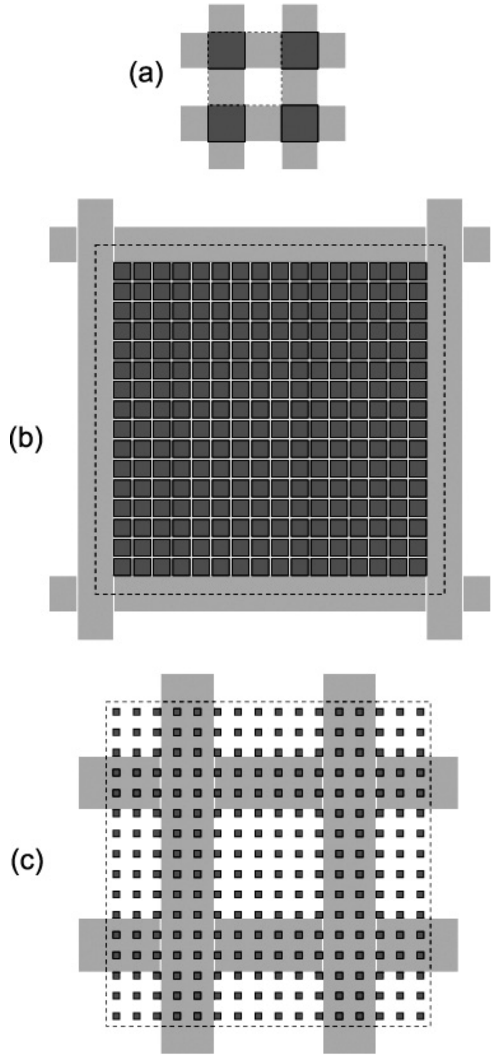


Figure 8.15: Plan-view comparison of the crossbars obtained (a) crossing lithographically defined lines, (b) using the lithographically defined lines above as seeds for S^8PT_+, and (c) using the lithographically defined lines above as seeds for $S^3PT_\times$. In each structure the square with dashed sides denotes a unit cell suitable for the complete surface tiling.

For $n \geq 3$ this comparison is so favorable to $S^nPT_\times$ to suggest its practical application. The following factors, however, would make S^nPT_+ preferable to $S^nPT_\times$:

1. If addressing the wires defining the crossbar are used also as addressing lines, they cannot run along the entire plane; their interruption for addressing reduces the available area.
2. In $\text{S}^n\text{PT}_\times$ all wires are produced collectively, and have the same height and material characteristics. On the contrary, in S^nPT_+ the wires are produced sequentially, each SPT_+ repetition produces wires of decreasing height, and the characteristics of the material (or even the materials themselves) may vary in a controlled way from one cycle to another.

Whereas at the first glance all the S^nPT_+ features described the second item may seem detrimental, the analysis of the Sect. 8.5.1 has clarified that they may be usefully exploited for cross-point addressing.

Deciding which route, between the additive and multiplicative ones, is actually more convenient for the preparation of hybrid devices depends on the particular application and circuit architecture. In fact, the multiplicative route (certainly less demanding for what concerns the preparation of the crossbar but more expensive for what concerns the nano-to-litho link) is presumably suitable for random access memories; on the contrary, the additive route (more demanding for the crossbar preparation, but much heavier for the elaboration of the signal) seems consistent with non-volatile memories.

8.5.3 Applications—not only nanoelectronics

The proposal of the S^nPT is quite recent (Cerofolini et al., 2005b; Cerofolini et al., 2005a), so that it has had only short time to be tested.

Electronics

For twhat concerns he preparation of crossbars, the attention has mainly been concentrated on one side on the verification of the possibility of scaling the S^nPT_+ to large values of n (the productions of arrays with $n = 3$ (Cerofolini et al., 2005b; Cerofolini et al., 2005a) and 4 (Cerofolini et al., 2007) have been reported) and on the other on the architectural impact of this technology (Ben Jamaa et al., 2007; Ben Jamaa et al., 2008; Ben Jamaa et al., 2009).

The electrical characterization of silicon nanowires has been the matter of really few works: Ref. (Wang et al., 2008b), addressed to the nanonowires produced via the SNAP technique,

The interest of a technology for the cheap production of nanowires is not limited to nanoelectronics but might extend to energetics.

Energetics

The large-scale availability of devices able to transform low-enthalpy heat (which would otherwise be dispersed into the environment) into electrical energy without complicate mechanical systems might impact the energy problem on a global scale. In principle the Seebeck effect provides a way for that. The ability of a material to operate as a Seebeck generator is contained in a parameter, ZT—a function of the Seebeck coefficient and of the electrical and thermal conductivities. The possibility of practical application is related to the occurrence of $ZT \gtrsim 1$.

From the technical point of view, Seebeck devices are already known, but materials with high Seebeck coefficient (multi-component nanostructured thermoelectrics, such as Bi_2Te_3/Sb_2Te_3 thin-film superlattices, or embedded PbSeTe quantum dot superlattices) are expensive. This fact has allowed the application of direct thermoelectric generators only to situations (e.g., space) where cost is not important or due to other factors (e.g., weight in space application).

This state of affairs would significantly change only in the presence of high efficient, low cost, materials. A new avenue to progress in such a direction has been the claim, by two independent collaborations (Hochbaum et al., 2008; Boukai et al., 2008), that silicon nanowires with width of $20 - 30$ nm and rough surfaces have $ZT \simeq 1$. Although silicon is certainly a cheap material and a lot of technologies are known for its controlled deposition, this result is of potential practical interest only if the production of nanowires does not involve electron beam or extreme ultra violet lithographies—hence the relevance of S^{n}PT.

It is also noted that the demonstrators of Refs. (Hochbaum et al., 2008) and (Boukai et al., 2008) employed wires with comparable height and width, thus characterized with a very high resistance. The S^{n}PT allows the preparation of wires with much higher aspect ratio (same width and much higher height—actually nanosheets). If it were possible to impart a sufficient roughness to the side walls of these sheets, that would result in much more efficient Seebeck generators.

8.6 Fractal nanotechnology

There is an application where the multiplicative route is manifestly superior: the preparation of fractal structures on the sub-lithographic length scale.

8.6.1 Fractals in Nature

The volume V of any body with regular shape (spheric, cubic,...) varies with its area A as

$$V = g_3 A^{3/2}, \tag{8.13}$$

with g_3 being a coefficient related to the shape (for instance, $g_3 = (1/6)^{3/2}$ for cubes, $g_3 = 1/6\sqrt{\pi}$ for spheres, etc.; a well known variational properties of the sphere guarantees that for all bodies $g_3 \leq 1/6\sqrt{\pi}$). For bodies with regular shape the ratio A/V diverges for small V and vanishes for large V.

Life can be preserved against the Second Law of thermodynamics only in the presence of a production of negative entropy in proportion to the total mass of the organism. This is achieved via the establishment of a diffusion field sustained by the metabolism inside the organism (Rashevsky, 1960). Since the consumption of energy and neg-entropy of living systems increases in proportion to V whereas the exchange of matter increases A, small unicellular organisms (like bacteria) have no metabolic problem due to their shape and can grow satisfying Eq. (8.13) preserving highly symmetric shapes (the smallest living bacterium, the pleuro-pneumonia like organism, with diameter $0.1 - 0.2$ μm, is spherical). On the contrary, larger organisms (even procaryotic, like the amoeba) may survive only adapting their shapes to have an energy uptake coinciding with what is required to preserve living functions (Cerofolini, 1981; Cerofolini, 1983b):

$$V = g_2 A, \tag{8.14}$$

with g_2 being a coefficient characteristic of two-dimensional growth. Moreover, the need to adapt itself to the variable environmental conditions is satisfied only thanks to the existence of an inner organellum, the endoplasmatic reticulum (Ploegh, 2007), which can fuse to the external membrane thus allowing an efficient change of area at constant volume (Cerofolini, 1981; Cerofolini, 1983b). This mechanism, however, can sustain the metabolism of unicellular organisms (like the amoeba) only for diameter of at most a few tens of micrometers, at most.

Larger organisms require the organization of the constituting cells in tissues specialized to single functions. This specialization, however, requires the formation of a vascular network (to transport catabolites to, and anabolites from, each constituting cell) whose space-filling nature implies a fractal character (Gazit et al., 1995). Nature prefers to manifest itself with fractal shapes in other situations too, like for the mammalian lung (to allow a better O_2-CO_2 exchange) (Shlesinger and West, 1991), the dendritic links of the neuron (to allow a high interconnection degree (Simons and Pellionisz, 2005)).

That smoothness is not a mandatory feature of the way how Nature expresses itself, but rather fractality is of ubiquitous occurrence in a large class of phenomena even in the Mineral Kingdom, has become clear with Mandelbrot's question about the length of Great Britain coast (Mandelbrot, 1983).

The fractality of Nature is manifestly approximate, the lowest length scale being ultimately limited by the atomic nature of matter. That surfaces may continue to have a scale invariance down to the atomic size became however clear only after Avnir, Farin and Pfeifer's study of the adsorption behavior of porous adsorbents (Pfeifer and Avnir, 1983; Avnir et al., 1983; Avnir et al., 1984).

Surprisingly enough, the artificial production of self-similar structures has remained largely unexplored. This is somewhat disappointing because the use of arrays or matrices with self-similar structures is potentially interesting even for applications. For instance, a matrix formed by the Cartesian product of two Cantor sets would provide a way for sensing with infinite probes a surface, leaving it almost completely uncovered. Although the minimum length scale is actually larger than the atomic one, that ideal case suggests the usefulness of the idea.

The lowest length scale of biological fractals is determined by cell size, say 10^4 nm. The preparation of self-similar structures on length scales between 10^2 and 10^4 nm is relatively easy via photolithography, but this scale seems inadequate for interesting applications like the highly parallel probing of single cells (as required, for instance, by the needs of systems biology (Hood et al., 2004)). Rather, for that application the appropriate length scale seems the one characteristic of nanotechnology, $1 - 10^2$ nm.

8.6.2 Producing nanoscale fractals via $\mathrm{S}^n\mathrm{PT}_\times$

Imagine for a moment that, in spite of the atomistic nature of matter and of the inherent technological difficulties, the multiplicative route can be repeated indefinitely. Remembering that the $(n+1)$-th step generates a set $\mathcal{S}_{n+1}$ that is nothing but the one at the n-th, $\mathcal{S}_n$, at a lower scale, the sequence $\{\mathcal{S}_0, \ldots, \mathcal{S}_n, \ldots\}$ defines a fractal; it will be referred to as multi-spacer fractal set. This fractal is self similar only if the height of each spacer varies with n as 2^{-n}, otherwise, the fractal is self-affine (Falconer, 2003). As mentioned above, the 'spontaneous' decrease of height with n, Eq. (8.2), renders the fractal self-affine. A self-similar fractal can be obtained at the end of process planarizing the whole structure with a resist and sputter-etching in a non-selective way the composite film until the thickness is reduced to $s_0/2^n$.

It is however noted that even assuming our ability to scale down the fabrication technology, the atomistic structure of matter limits anyway the above considerations to an interval of $1 - 2$

orders of magnitude, ranging from few atomic layers to the lower limits of standard lithography.

Having clarified in which limits the set $\mathcal{S}_n$ may be considered a fractal, it is interesting to compare it with other fractal sets. The prototype of such sets, and certainly the most interesting from the speculative point of view, is the Cantor middle-excluded set. Figure 8.16 compares sequences of three discrete processes eventually leading to the multi-spacer fractal set $\mathcal{S}$ and to the Cantor set $\mathcal{C}$. The comparison shows interesting analogies: Consider a multiplicative multi-spacer with $P = 2W$; if $w_n = \frac{1}{3}w_{n-1}$, the measure at each step of multi-fractal set coincides with that of the Cantor set. This implies that the multi-spacer fractal set has null measure. Similarly, it can be argued that the multi-spacer set, considered as a subset of the unit interval, has the same fractal dimension as the Cantor middle-excluded set—$\ln(2)/\ln(3)$ (Falconer, 2003). At each step the multi-spacer fractal set is characterized by a more uniform distribution of single intervals than the Cantor set; this makes the former more interesting for potential applications than the latter.

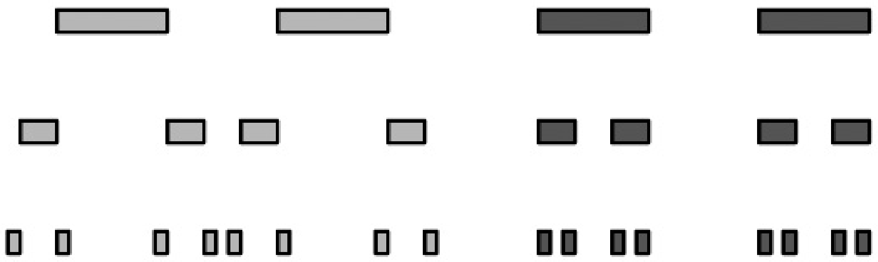


Figure 8.16: Generation of the multi-spacer set (*left*) and of the Cantor middle-excluded set (*right*)

Once one has one-dimensional fractal sets, two-dimensional fractal sets can be constructed taking their Cartesian products. Although the mixed product $\mathcal{C} \times \mathcal{S}$ is possible, in the following the attention will be concentrated on a comparison of the Cantor and multi-spacer fractal crossbars. Figure 8.17 compares their plan views showing

(A) a 16×16 crossbar obtained via $\mathrm{S^3PT_\times}$ starting from lithographically defined seeds separated by a distance satisfying Eq. (8.12), and

(B) a 16×16 crossbar obtained via $\mathrm{S^3PT_\times}$ starting from lithographically defined seeds and arranging the process to generate the Cantor middle excluded set.

Although the potential applications of the Cantor set are quite far, trying to reproduce it on the nanometer length scale seems of a certain interest. This is compatible with existing technologies; a possible process would involve

(C1) the lithographic definition of the seed (formed, for instance, by poly-silicon) generating the Cantor set,

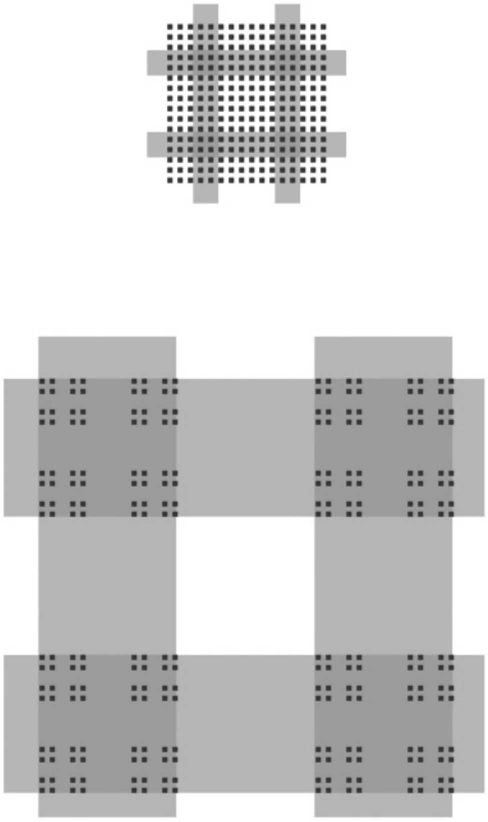


Figure 8.17: Plan-view comparison of the crossbars obtained via $\mathrm{S}^3\mathrm{PT}_\times$, and a corresponding Cantor middle-excluded set; the width of the lithographic seed in $\mathrm{S}^3\mathrm{PT}_\times$ has been assumed to to be at the technology forefront, whereas in the Cantor set it has been assumed sufficiently large to allow three reiterations of the process

(C2) its planarization (for instance, via the deposition of a low viscosity glass and its reflow upon heating),

(C3) the etching of this film to a thickness controlled by the exposure of the original seed,

(C4) the selective etching of the original film,

(C5) the conformal deposition of a film of the same material as the original seed (poly-silicon, in the considered example) and of thickness equal to 1/3 of its width,

(C6) its directional etching, and

(C7) the selective etching of the space seed (glass, in the considered example).

Figure 8.18 sketches the overall process (Cerofolini et al., 2008b).

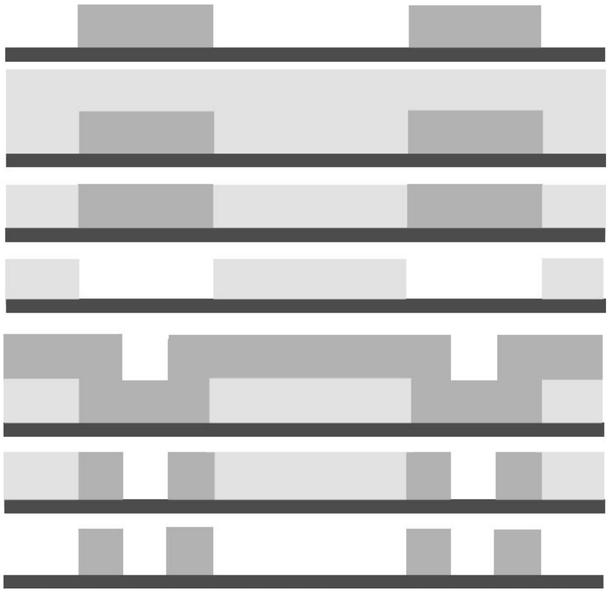


Figure 8.18: A process for the generation of Cantor's middle-excluded set

The preparation of fractal structures may appear at a first sight nothing but a mere exercise of technology stressing. The following examples suggest however the usefulness of a fractal technology:

1. The highly parallel and real-time sensing of single cells (as required, for instance, by the needs of systems biology (Hood et al., 2004) of by label-free immunodetection (Stern et al., 2007)) could be done with a minimum of perturbation (i.e., leaving almost completely uncovered the cell surface) by a matrix formed by the Cartesian product of two Cantor sets (Falconer, 2003).

2. Super-hydrophobic surfaces may be prepared controlling roughness and surface tension of non-wetting surfaces (Tuteja et al., 2007). Whereas surface tension is a material property, roughness can be controlled by the preparation. For instance, roughness may be achieved by imparting a suitable (fractal) relief on the surface.

3. If the S^{n}PT is used for the preparation of crossbar structures for molecular electronics, the functionalization with organic molecules of the cross-points can only be done after the preparation of the hosting structure. According to the analysis of Ref. (Cerofolini et al., 2007), this requires an accurate control of the rheological and diffusion properties in a medium embedded in a domain of complex geometry. Understanding how such properties change when the size is scaled and clarifying to which extent the domain can indeed be viewed as a fractal (so allowing the analysis on fractals (Kigami, 2001) to be used for their description) may be a key point for the actual exploitation of already producible nanometer-sized wire arrays in molecular electronics.

Chapter 9

Conclusions

Of course nanobots are a topic so broad and so far reaching that, currently, a PhD thesis can only sketch a hypothetical scenario identifying the nature of problems posed by their development and use for real applications. Nevertheless I believe this PhD thesis achieved 2 aims. First, I described an architecture of nanobots which is at the same time able to embed sophisticated functions, and suitable for being manufactured by processes compatible with today and likely tomorrow semiconductor industries. Second, I developed swarm intelligence algorithms needed for tackling the nanorobot tasks: clustering around diseased cells. These algorithms take into account all the restrictions and limitations associated with the proposed nanobot architecture.

The work I did for my PhD thesis led me to publish up to now 5 papers (Cerofolini et al., 2010a; Cerofolini et al., 2011a; Amato et al., 2011; Amato, 2011; Cerofolini and Amato, 2012)[1]. In the following I briefly summarize the 4 major achieved results.

Identified nanobot tasks for early diagnosis of diseases. The main idea for the early diagnose of diseases is to survey the state of the organism by injecting swarms of nanobots in the circulatory system. In fact, since the main role of the circulatory system is to feed the organism, the capillaries come close to any cells in the body. Thus Nanobots can be considered as circulating nano-laboratories for blood analysis in situ. Once injected in the circulatory system, the first task of the nanobots is to cluster around the diseased cells in order to help the identification of their positions; ideally even if there is just one diseased cell. To accomplish this task, the nanobots need to autonomously disperse in the capillary bed, take chemical sensor reading, mark the region where a positive signal is detected, and form a cluster in that region. The most complex subtask is the last one. Since in the microscopic world collisions are not an issue, communication between nanobots can happen through direct physical clashes. By equipping a nanobot with suitable appendages it can anchor on the surface of the capillaries and bind to other nanobots. All the here developed swarm-intelligence strategy are based on this communication-through-collision mechanism.

Defined high-level architecture of nanobots. If nanobots are shaped as ellipsoidal disk ($50\times 2.5\times 1\mu\mathrm{m}$) they could move quite freely through the entire circulatory system. Such a disk is in principle producible by employing the standard silicon technology used for integrated circuits,

---

[1] I also have been invited to popularize the nanobots theme in (Cerofolini and Amato, 2010)

but there are 2 major challenges to be tackled. On one side, even by using next generation silicon planar technology, such a disk could host 10^5 gates to be used for implementing all the needed circuitry. This number is quite low, but it could be increased by at least one order of magnitude by adopting the crossbar architecture. On the other side, endowing these nanobots with the navigation, sensing ad anchoring capabilities requires the exploitation of bio-molecular devices. In particular the sensors could be implemented by growing carbon nanotubes in the crosspoints of crossbar architecture. Regarding the clustering capabilities, the anchoring can be achieved by using quinoxaline cavitands, while the binding between nanobots could simply require the use of suitably shaped hydrophobic head and tail.

Developed a first fractal model of circulatory system. The systemic human circulatory system can be represented as the conjunction of 2 fractal trees. These 2 trees (the arterial one and the venous one) are connected trough their branch tips. Assuming that both tress are identical thick-stemmed binary symmetric trees, I developed a set of equations to connect the tree parameters with known physiological parameters of the circulatory system. By optimizing the tree parameters, the estimated physiological values stay in a reasonable error range ($< 3\%$); an indication that the fractal trees can be used at least as first order approximation of the circulatory system.

Developed and analyzed nanobots clustering strategies. Since the task of the nanobots is to cluster around diseased cells, I develop three different swarm-intelligence strategies to tackle this task: only target (OT), self-docking (SD) and division of labor (DL). In particular DL is inspired by how honeybee colonies organize the work during foraging. To obtain foraging success in a rapidly changing flower market, the colony must receive both updates on old food sources and reports on promising new ones. This is achieved by keeping two distinct types of employed foragers (Seeley, 1996), and a similar approach can be used for nanobots looking for diseased cells. Consequently in this strategy there are 2 types of nanobots: explorers and recruits. The explorers look for new target, while the recruits can only anchor to already anchored nanobots. To test the effectiveness of these strategies in dealing with different number of targets (i.e., the presence of one or more diseased cells in different parts of the body) I simulated two different scenarios: a square region and a double-tree. The square region mimics a static scenario (e.g., blood analysis inside a test tube), while the tree scenario mimics the circulatory system. In both scenarios SD turns out to be the best strategy when dealing with just one target, while DL is the best strategy dealing with many targets. At last, by generalizing the simulation results, it has been estimated that a swarm of 10^7 nanobots would be able to form a cluster of size 10 inside the human body in around 2 weeks. A swarm of such size can be produced by processing a single 200 mm silicon wafer, and the total nanobot mass circulating inside the body would be just 10^{-2} g.

Nanobots stay at the intersection of several research fields, including Computer Science, Medicine, Chemistry, Robotics and Technology. And indeed during my PhD I had the luck of collaborating with people from the Department of Computer Science, Department of Material Science and Department of Experimental Medicine of University of Milano Bicocca. I wish to thank them all.

Part III

Appendix

Appendix A

Abstract technology

The main results presented in this chapter have been published in (Amato, 2011)

A.1 Introduction

The fabrication of integrated circuits (ICs) involves a complicated sequence of processes, usually formed, for circuits of microprocessor or memory complexity, by more than 400 fabrication steps. They are built on starting from a substrate of bare single-crystalline silicon, via the deposition or etching of suitable materials (like poly-crystalline silicon, SiO_2, Si_3N_4, aluminum,...). At the end of these steps, the wafer is divided in many dice — each die being the IC object of the particular manufacturing. The fabrication steps change as a function of the desired physical and functional properties of the final circuit (Cerofolini and Meda, 1989; Waser, 2003).

The key process for IC production is certainly related to *patterning*. In this process a relatively thick film of photoactive monomer is uniformly deposited on the wafer and exposed to collimated light beam suitably masked (Rai-Choudhury, 1997). In this way the pattern defined on the mask is projected onto the photoactive material, referred to as *resist*, that consequently undergoes polymerization[1]. What makes interesting this process is the fact that the residual monomer can be selectively etched away leaving on the wafer the same pattern on the mask *suitably scaled to the wanted size through the optical system.* Of course the scaling factor has physical limits (see, for instance, Ref. (Ito and Okazaki, 2000)), however not of interest for this work.

Most of the other processes involved in the IC fabrication may be divided in four classes: conformal deposition, isotropic etching, directional deposition and directional etching. Moreover the etching processes can be selective; i.e., affecting a given material only. Once defined in rigorous mathematical terms, the processes above may be used (in non-conventional way) as building blocks of a nanotechnology; in this way they may be seen as the equivalent of straightedge and compass in Euclid's geometry, where their use in geometric constructions is replaced by the construction via conformal, directional or selective, deposition or etchings. The resulting theory is referred to as *abstract technology (AT).* An early version of this theory has been sketched in (Cerofolini et al., 2011a).

In the spirit of Erlangen program, Euclidean geometry is characterized by the properties

---

[1] Materials with this behavior are referred to as positive resists; any symmetric material, formed by a polymer undergoing depolymerization under irradiation, is referred to as negative resist.

that are invariant with respect to translations and rotation; in particular, the transformation of Euclidean geometry preserve the topological properties of any given body. That the operations defining the nanotechnology are not described by the Euclidean geometry is thus immediately understood observing that patterning usually destroys the topological properties of the resist, transforming in general the original, simply connected, layer into a set of non-overlapping, possibly multiply connected, regions. Hence the need of a formal description of technological processes.

Of course, the properties of a figure built with any real straightedge and compass are not exactly the same as those of the same figure constructed with the ideal straightedge and compass. Similarly, the use of real materials and processes produces structures that differ from those achievable via the ideal materials and processes. In fact, the proposed theory considers the geometric properties only, leaving out all the chemical and physical phenomena.

Nonetheless, it is convenient to develop the theory in abstract terms to enable the development of software (SW) tools able to forecast the result of the applications of different semiconductor fabrication techniques. In fact, at the best of our knowledge, the tools currently used are based on the analysis of the physical-chemical properties of the structured regions that define the wafer-scale reactor where the process is carried out, and are very computationally intensive; see e.g. (Camarda et al., 2007).

Together with the above operations one could certainly define other, more complicate (or perhaps even formally more interesting), operations. What is of uppermost importance here is that *there already exist combinations of materials and processes of the silicon technology mimicking the ideal behaviors described above*.

Before giving the notions of AT theory, the following section briefly review the main silicon fabrication processes.

A.2 Concrete technology

The materials considered in this article are well known in the silicon technology: A practical model for wafer is any body extending in two directions exceedingly more than the change of thickness resulting from the considered processes and such that only the processes are actually carried out on only one of its major surfaces. Before undergoing any operations such wafers have generally a homogeneous chemical composition and high flatness and are referred to as substrates. In the typical situations considered herein the substrates are slices of single crystalline silicon. The other materials are polycrystalline silicon (poly-Si), SiO_2, Si_3N_4 and various metals (Al, Ti, Pt, Au,. . .) (Plummer and Griffin, 2001).

The typical processes involved in the silicon technology are lithography (for patterning), wet or gas-phase etchings, chemical or physical vapor deposition, planarization, doping (typically via ion implantation), and diffusion. A special class of materials is formed by resists, i.e. photoactive materials undergoing polymerization (or depolymerization) under illumination. The processes of interest here are the following:

Lithography requires a preliminary planarization with a resist, its patterning (i.e., the definition of a geometry) via the exposure through a mask to light, the selective etching of the exposed (or unexposed) resist, the selective etching first of the region not protected by the patterned resist, and eventually of this material (Okazaki and Moers, 2003).

Table A.1: Shape resulting after etching or growth processes

	shape	
process	conformal	directional
attack	wet etching	sputter etching
	⟵ plasma etching	
	reactive ion etching ⟶	
growth	CVD	PVD

Wet etchings are usually isotropic and are used for their selectivity: HF_{aq} etches isotropically SiO_2 leaving unchanged silicon and Si_3N_4; H_3PO_4 etches isotropically Si_3N_4 leaving unchanged silicon and SiO_2; $HF_{aq} + HNO_{3\,aq}$ etches Si leaving unchanged Si_3N_4 (but has poor selectivity with respect to SiO_2) (Schneider and McClatchie, 2003).

Sputter etching is produced by momentum transfer from a beam to a target and results typically in poorly selective, highly directional etching. Selectivity can be imparted exploiting reactive ion etching.

Plasma etching can be tuned to the situation: via a suitable choice of the atmosphere it can be used for the isotropic selective etching of Si, or SiO_2 and Si_3N_4; it becomes progressively more directional and less selective applying a bias to to the body ('target') with respect to the plasma (Schneider and McClatchie, 2003).

Chemical vapor deposition is the typical way for the conformal deposition, that occurs when the growth is controlled by reactions occurring at the growing surface. Poly-Si grows well on SiO_2 but requires an SiO_2 buffer layer for the growth on Si_3N_4; conversely Si_3N_4 is easily deposited on SiO_2. Silicon can be conformally covered by extremely uniform layers of SiO_2 with controlled thickness on the subnanometer range via *thermal oxidation* (Erhardt, 2003).

Physical vapor deposition is the typical way for the deposition of metal films. The corresponding growth mode is the positive counterpart of directional etching. The conformal deposition of a metal layer can only be done via CVD from volatile precursors (like metal carbonyls or metallorganic monomers) (Erhardt, 2003).

Table A.1 summarizes the shape (conformal or directional) resulting from the above processes.

A.3 Abstract technology

In this section the formal counterparts of the above silicon fabrication processes are introduced.

A.3.1 Bodies and surfaces

The minimal characterization of a body is in terms of its geometry and constituting materials.

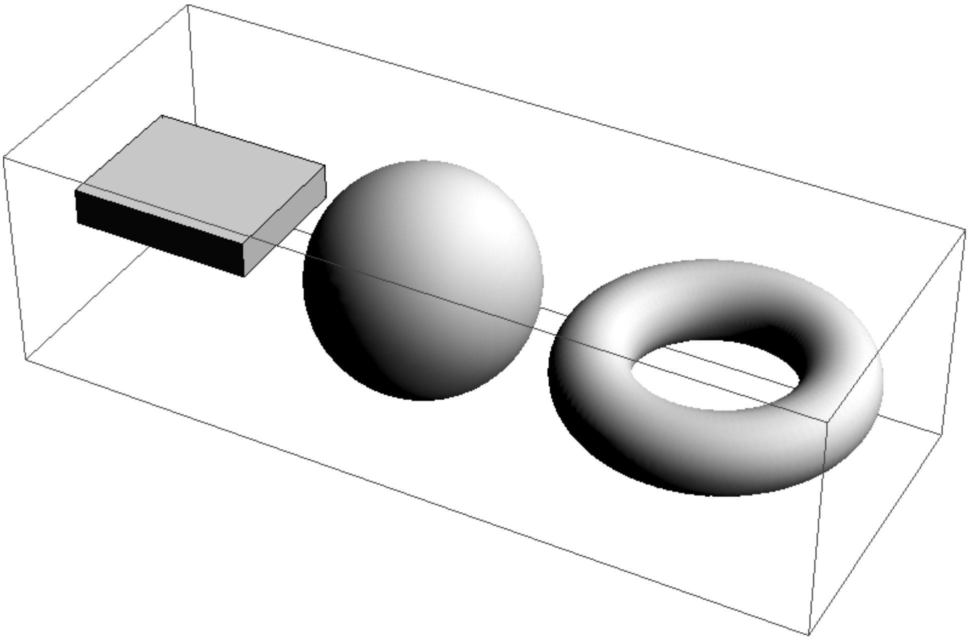


Figure A.1: Three examples of simple bodies

Definition 8 (Simple body) *A* simple *body $\mathcal{B}_\mathrm{X}$ is a 3-dimensional closed connected subset B of the Euclidean space $\mathbb{E}^3$ filled with a material X:*

$$\mathcal{B}_\mathrm{X} = (B, \mathrm{X}).$$

Figure A.1 shows three examples of simple bodies. The definition requires the set B to be connected, but does not forbid holes (like in the torus on the right of the figure).

What is a material is considered here a primitive concept. It may be a substance, a mixture, a solution or an alloy. The emphasis is on homogeneity—whichever region, however small, of B is considered, the constituting material is the same. When not necessary, the index denoting the constituting material will not be specified.

In view of the granular nature of matter, considering the limit for vanishing size is a nonsense. Here and in the following when dealing with the concept of vanishing length (as it implicitly happens when considering the frontier B^* of B) I mean that the property holds true on the *ultimate length scale*. Although the ultimate length scale is an operative concept related to the probe used for observing the body, I however have in mind the molecular one. Analogously, I consider a body as indefinitely extended when its size is much bigger than the size variations induced by the considered processes. In this sense, a *wafer* is a simply-connected simple body indefinitely extended over two dimensions.

Definition 9 (Composite body) *Consider the N-tuple of simple body $(\mathcal{B}_{\mathrm{X}^1}, \mathcal{B}_{\mathrm{X}^2}, \ldots, \mathcal{B}_{\mathrm{X}^N})$ formed by non-overlapping sets $(B_1, B_2, \ldots, B_N)$, with*

$$\forall I, J \left(B_i \cap B_j = B_i^* \cap B_j^* \right),$$

and let

$$B = \bigcup_{i=1}^{N} B_i.$$

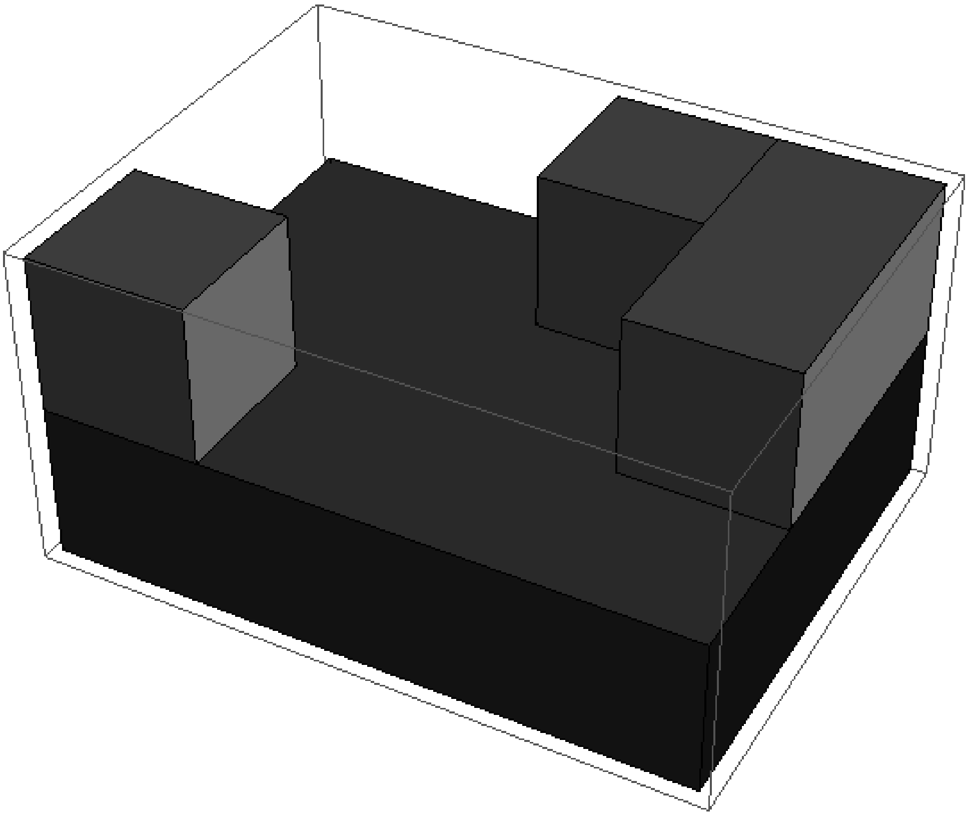


Figure A.2: Example of composite body containing two different materials (material1 in dark grey, and material 2 in light gray)

If B *is connected and the simple bodies* $\mathcal{B}_{\mathrm{X}^1}, \mathcal{B}_{\mathrm{X}^2}, \ldots, \mathcal{B}_{\mathrm{X}^N}$ *contain at least two different materials* X^I *and* X^J*, then the* N*-tuple is a* composite body.

Figure A.2 shows an example of composite body containing two different materials

Definition 10 (Interface) *Let a composite body* $\mathcal{B}$ *be formed by simple bodies* $\mathcal{B}_{\mathrm{X}^1}$ *and* $\mathcal{B}_{\mathrm{X}^2}$ *with* $\mathrm{X}^1 \neq \mathrm{X}^2$*; the region*

$$F_{1|2} = B_1^* \cap B_2^*$$

defines the interface between $\mathcal{B}_{\mathrm{X}^1}$ *and* $\mathcal{B}_{\mathrm{X}^2}$.

For any x in $F_{1|2}$ all neighborhoods, however small, contain materials X^1 and X^2.

Definition 11 (Total surface) *The set*

$$S_{\text{tot}} = B^*$$

is the total surface *of* $\mathcal{B}$.

Proposition 1 *Let a composite body* $\mathcal{B}$ *be formed by the* N*-tuple of simple body* $(\mathcal{B}_{\mathrm{X}^1}, \mathcal{B}_{\mathrm{X}^2}, \ldots, \mathcal{B}_{\mathrm{X}^N})$. *Then:*

$$S_{\text{tot}} = \bigcup_i B_i^* \setminus \bigcup_{j,k} F_{j|k}$$

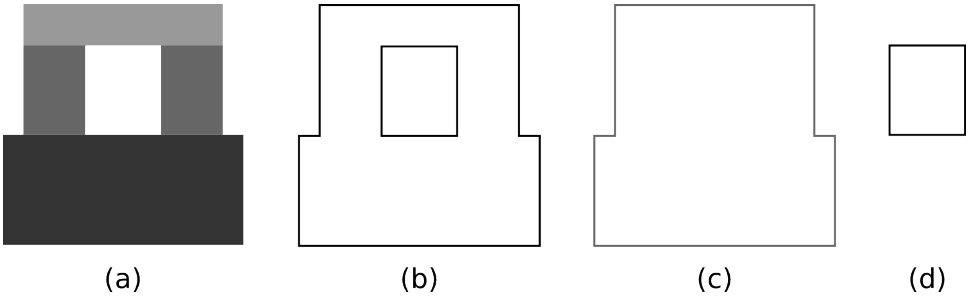


Figure A.3: A 2-dimensional composite body made of three materials (*a*), its total surface (*b*), its (outer) surface (*c*) and its inner surface (*d*)

Definition 12 (Surface) *Let $\mathcal{B}$ be a composite body, and $B_\bullet$ be the smallest simply-connected (i.e., without holes) set containing B. Then:*

$$S = B^*_\bullet$$

is the outer surface *(or simply* surface*) of $\mathcal{B}$.*

Definition 13 (Inner surface) *Let $\mathcal{B}$ be a composite body. Then:*

$$S_{\text{in}} = S_{\text{tot}} \setminus S$$

is the inner surface *of $\mathcal{B}$.*

Figure A.3 shows an example of composite body togheter with its total, outer and inner surface.

A.3.2 Conformal deposition and isotropic etching

Definition 14 (Delta coverage) *For any body $\mathcal{B}$ with surface S, the* additive delta coverage *$\overline{D}_{+\delta}$ of thickness δ is the set*

$$\overline{D}_{+\delta} = \{\mathbf{x} : \mid \mathbf{x} - \mathbf{y} \mid \leq \delta \wedge \mathbf{y} \in S \wedge \mathbf{x} \notin B_\bullet \setminus S\}$$

The delta coverage can be imagined to result from the application of an operator a_δ to B:

$$\overline{D}_{+\delta} = a_\delta(B).$$

The delta coverage $\overline{D}_{+\delta}$ is the “pod” of thickness δ covering the set B. The set obtained as union of B with its delta coverage is referred to a $B_{+\delta}$:

$$B_{+\delta} = B \cup a_\delta(B).$$

This operation can be reiterated, and the final set obtained after n iteration is referred to as $B^n_{+\delta}$:

$$B^n_{+\delta} = B^{n-1}_{+\delta} \cup a_\delta\big(B^{n-1}_{+\delta}\big),$$

where $B^1_{+\delta} = B_{+\delta}$ and $B^0_{+\delta} = B$.

The reiteration of a_δ is the base for the following definition:

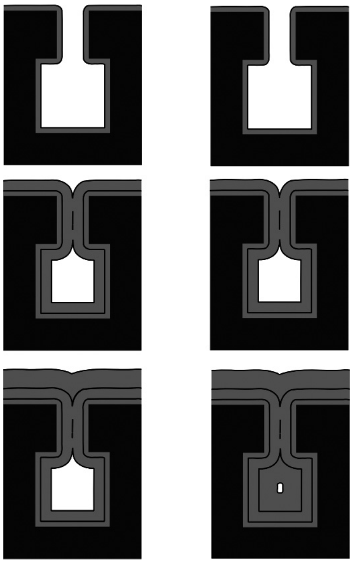


Figure A.4: The different behavior of conformal coverage (*left*) and delta coverage (*right*)

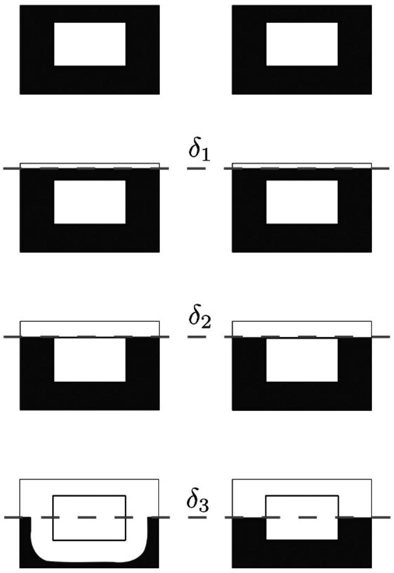


Figure A.5: The different behavior of conformal depletion (*left*) and delta depletion (*right*)

Definition 15 (Conformal coverage) *A* conformal coverage $C_{+\delta}$ *of thickness* δ *is the following limit*[2]

$$C_{+\delta} = \lim_{n\to+\infty} \bigcup_{i=0}^{n-1} a_\delta\big(B^i_{+\delta/n}\big).$$

In other words, the conformal coverage process can be thought of as obtained by applying infinitely many delta coverages; one after another, and each one of infinitesimal thickness. As shown in Figure A.4, in general conformal coverage and delta coverage of the same thickness can lead to different bodies. Moreover from the same figure it is clear that these transformations are not topological invariants.

Definition 16 (Conformal deposition) *A* conformal deposition *of a film of material Z of thickness* δ *is the pair* $(C_{+\delta}, \mathrm{Z})$

Note that the material Z may be different from the materials X^i forming the original body The following theorem is trivial:

Theorem 1 (Planarization) *Let* $d_{\mathcal{S}}$ *the diameter of* $\mathcal{B}$*. Irrespective of the shape of* $\mathcal{B}$ *the conformal deposition of a layer for which* $\delta \gg d_{\mathcal{S}}$ *produces a body whose shape becomes progressively closer and closer to the spherical one as* δ *increases. If* $\mathcal{B}$ *is indefinitely extended in two directions, it undergoes a progressive planarization.*

---

[2] Such a limit is naively clear, but its rigorous specification should require to elaborate many technical details that I prefer to skip in this appendix. The same considerations hold true anytime I introduce limits of this kind.

Isotropic etching can be defined in a way similar to that used for conformal coverage by introducing an operator s_δ, characterized by the following definitions:

$$\begin{aligned} s_\delta(B) &= a_\delta\big(\mathbb{E}^3 \setminus B_\bullet\big) \\ B^n_{-\delta} &= B^{n-1}_{-\delta} \setminus s_\delta\big(B^{n-1}_{-\delta}\big) \end{aligned}$$

In general, the operator s_δ is not the inverse of a_δ.

Definition 17 (Conformal depletion) *A* conformal depletion $E_{-\delta}$ *of thickness* δ *is the following limit:*

$$E_{-\delta} = \lim_{n\to+\infty} \bigcup_{i=0}^{n-1} s_\delta\big(B^i_{-\delta/n}\big)$$

Also in this case, it is true that in general the conformal depletion and delta depletion s_δ of the same thickness can lead to different bodies (see Figure A.5).

Definition 18 (Isotropic etching) *An* isotropic (non-selective) etching *is a conformal depletion of thickness* δ *applied to a body* $\mathcal{B}$ *regardless of its constituting materials*

The etching process just defined is non-selective. However the most part of etching are selective (see section A.3.4)

A.3.3 Directional processes

Definition 19 (Shadowed surface along $\mathbf{d}$) *Let* $\mathbf{d}$ *be a unit vector in* $\mathbb{E}^3$*. For any point* s *belonging to the surface* S *consider the straight line* l_s *starting from* s *and oriented as* $\mathbf{d}$*. If* l_s *intersects* S*, then* s *is said to be shadowed along* $\mathbf{d}$*. The set all shadowed points of* S *is referred to as the* shadowed surface $S_\mathbf{d}$ *along* $\mathbf{d}$ *of the body* $\mathcal{B}$*.*

Definition 20 (Directional delta coverage along $\mathbf{d}$) *The* directional delta coverage *along* $\mathbf{d}$ *is the set of points lying on the straight lines directed along* $\mathbf{d}$ *and going from the exposed surface to the same surface shifted along* $\mathbf{d}$ *by an amount* δ*:*

$$\overline{D}_{+\delta,+\mathbf{d}} = \{\mathbf{x} :\ \mathbf{x} = \mathbf{y} + a\mathbf{d} \wedge a \le \delta \wedge \mathbf{y} \in S \setminus S_\mathbf{d}\}$$

Figure A.6 shows the results of applying to a given set the delta coverage and the directional delta coverage along $\mathbf{d}$.

By using the directional delta coverage along δ instead of the delta coverage, it is possible to define the directional deposition:

Definition 21 (Directional deposition) *A* directional deposition $C_{+\delta,+\mathbf{d}}$ *of thickness* δ *along* $\mathbf{d}$ *is the following limit:*

$$C_{+\delta,+\mathbf{d}} = \lim_{n\to+\infty} \bigcup_{i=0}^{n-1} a_{\delta,+\mathbf{d}}\big(B^i_{+\delta/n,+\mathbf{d}}\big),$$

where $a_{\delta,+\mathbf{d}}$ *is the operator associated with the directional delta coverage and* $B^n_{+\delta,+\mathbf{d}} = B^{n-1}_{+\delta,+\mathbf{d}} \cup a_{\delta,+\mathbf{d}}(B^{n-1}_{+\delta,+\mathbf{d}})$.

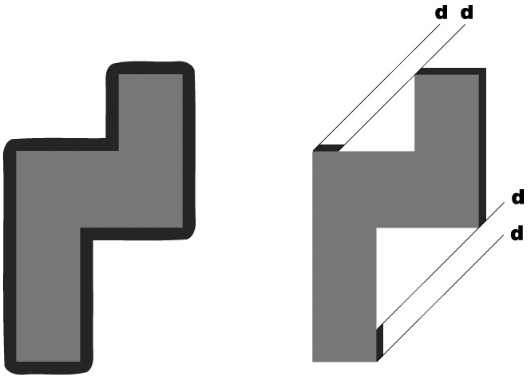


Figure A.6: Comparison between delta coverage (*left*) and directional delta coverage along $\mathbf{d}$ (*right*) of the same thickness δ

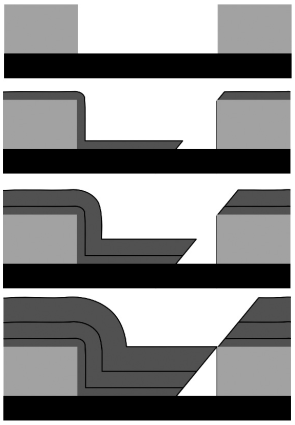


Figure A.7: An example of directional deposition

Figure A.7 shows an example of directional deposition. This process (as shown in the figure) can create holes inside the bodies.

Analogously, to define the directional depletion I introduce the directional delta depletion:

Definition 22 (Directional delta depletion along $\mathbf{d}$)

$$\overline{D}_{+\delta,-\mathbf{d}} = \{\mathbf{x} :\ \mathbf{x} = \mathbf{y} - a\mathbf{d} \wedge a \leq \delta \wedge \mathbf{y} \in S \setminus S_{\mathbf{d}}\}$$

Definition 23 (Directional etching along $\mathbf{d}$) *A* directional etching $E_{-\delta,-\mathbf{d}}$ *of thickness* δ *is the following limit:*

$$E_{-\delta,-\mathbf{d}} = \lim_{n \to +\infty} \bigcup_{i=0}^{n-1} s_{\delta,-\mathbf{d}}\big(B^{i}_{-\delta/n,-\mathbf{d}}\big),$$

where $s_{\delta,-\mathbf{d}}$ *is the operator associated to the directional delta depletion and* $B^{n}_{-\delta,+\mathbf{d}} = B^{n-1}_{-\delta,-\mathbf{d}} \setminus s_{\delta,-\mathbf{d}}(B^{n-1}_{-\delta,-\mathbf{d}})$.

A.3.4 Selective processes

No indication has hitherto been given about the dependence of the same unit process on the material to which it is applied. In general the effect depends on the material, and this difference is referred to in terms of selectivity.

Before introducing selectivity, I note that the surface S of a body $\mathcal{B}$ is, in general, composed by different materials. In particular, by using Proposition 1 it is possible to define $S_{\text{tot,X}}$, the subset of S_{tot} containing material X:

$$S_{\text{tot,X}} = \bigcup_{\text{X}^i=\text{X}} B_i^* \setminus \bigcup_{j,k} F_{j|k},$$

and then the (outer) surface S_{X} composed by material X:

$$S_{\text{X}} = S_{\text{tot,X}} \cap S.$$

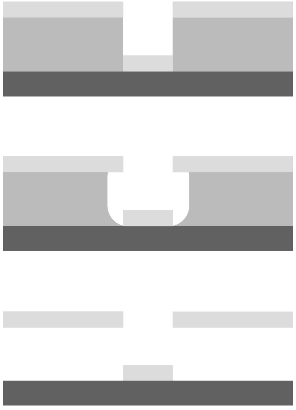


Figure A.8: An example of selective etching (“lift-off”)

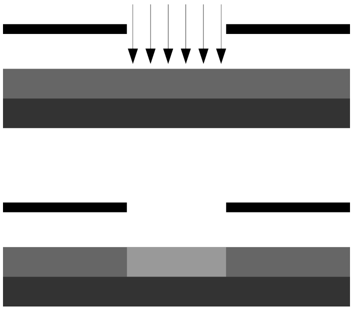


Figure A.9: The patterning process. The mask (the black bar) acts as a filter

Definition 24 (Selective delta depletion) *For any body $\mathcal{B}$ with surface S, the* selective delta depletion $s_{\delta,\mathrm{X}}(B)$ *of thickness δ is the set*

$$s_{\delta,\mathrm{X}}(B) = \{\mathbf{x} :\mid \mathbf{x} - \mathbf{y} \mid\leq \delta \wedge \mathbf{y} \in S_{\mathrm{X}} \wedge \mathbf{x} \notin B_{\bullet} \setminus S_{\mathrm{X}}\}$$

In other words, the selective delta depletion is a delta depletion that affects S_{X} only, not the whole surface S. I remark that, in general, $s_{\delta,\mathrm{X}}(B)$ is not a connected set.

Definition 25 (Selectivity) *Any process is said to be* selective *with respect to materials* X *when, applied to a composite body $\mathcal{B}$ with frontier S, affects S_{X} only.*

Definition 26 (Selective etching) *A* selective etching $E_{-\delta,\mathrm{X}}$ *of thickness δ is the following limit (in a sense to be defined):*

$$E_{-\delta,\mathrm{X}} = \lim_{n\to+\infty} \bigcup_{i=0}^{n-1} s_{\delta,\mathrm{X}}\big(B^{i}_{-\delta/n,\mathrm{X}}\big),$$

where

$$B_{-\delta/n,\mathrm{X}} = \big(B \setminus s_{\delta,\mathrm{X}}(B)\big) \cup \left(\bigcup_{\mathrm{X}^j\neq\mathrm{X}} B_j\right),$$

and the iterative process is defined as in the previous cases.

As example of selective etching is given in Figure A.8.

A.3.5 Patterning

The aim of patterning is to project the pattern defined on a given body (called *mask*) on a photoactive material (called *resist*); see Figure A.9. Consequently patterning is both a directive, along directions perpendicular to the mask, and a selective process with respect to resist material. Usually patterning involves also the scaling of the pattern on the mask. But here I suppose that the mask is already suitably scaled to the wanted size.

Figure A.10: Ideal directional etching (*left*); linear combination of directional and isotropic etching (with weight 5/6 and 1/6 respectively) to mimic imperfect anisotropy (*right*)

Definition 27 (Patterning) *Let $\mathcal{B}$ be a body, consider all the direction* $\mathbf{d}$ *perpendicular to the mask M, and let $D = \{\mathbf{d} : \mathbf{d}$ is not shadowed by M and $\mathbf{d}$ intersects $S_{\text{resist}}\}$. The* patterning process *directionally works along all $\mathbf{d} \in D$ transforming all the resist material connected to the surface S_{resist} in* polymerized resist.

Consequently the application of patterning to a simple body of resist material, implies the separation of it in two or more simple bodies composed by either resist or polymerized resist. Once part of the resist is polymerized, it can be selectively etched away.

At last, I note that in concrete the resist is deposed trough a dedicated technique called *spin-coating*. However, here I consider it as a standard conformal deposition with δ much greater than the roughness of the surface. In this way, following Theorem 1, a final surface with minimum roughness can be achieved.

A.3.6 Towards concrete processes

As pointed out in the introduction, the spirit behind AT theory is to define ideal processes to be used for the preparation of structures; in the same way that Euclidean geometry defines ideal straightedge and compass to build geometric figures. In so doing it is accepted that the resulting structures will differ from those produced by using real processes (and real materials too). However, there are many situations in which it is well known that the real process is far from being ideal.

Fortunately, in AT theory it is possible to cope with these situations by simply making suitable linear combinations of the ideal processes. For example an imperfect anisotropy etching can be modeled as a linear combination of of directional etching and isotropic etching. Figure A.10 shows the impact on the final shape of a trench of introducing such a linear combination. This example will have a particular relevance in the section discussing vertical arrangements of nanowires based on crossbar structures.

The etching processes described so far, together with their linear combinations, share the following property: They can grow in depth in an amount equal or greater than they can grow in width. Figure A.11 (*a* and *b*) shows this concept for the isotropic and directional etchings. However there are applications for which it would be interesting to have an etching process able to grow more in width than in depth (Figure A.11, *c*). An example is the production of power devices, where a high breakdown voltage can be sustained only by smooth shaping the border of active zone, currently obtained by a suitable sequence of masks. Moreover it is possible to

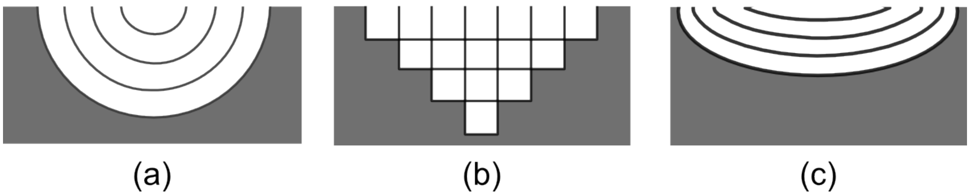


Figure A.11: Etching processes: (*a*) isotropic, (*b*) directional, and (*b*) expanding more in width than in depth

fancy that such an etching process, non-existing at best of our knowledge, could also play a role in the development of contact interfaces for vertical arrangement of (nano)wires.

A.4 Examples

In this section examples of application of AT theory are given both to build manufacturing structures and geometrical shapes.

A.4.1 Manufacturing examples

Here I'll consider 2 examples: the formalization of the spacer patterning technique (SPT), that has been considered in Section 8.3.2; and the vertical arrangement of poly Si nanowires based on the crossbar structure.

SPT

From the practical point of view, an SPT may be implemented with the following steps:

1. the formation (via CVD or thermal oxidation) of a relatively thick (say with thickness $t_{\mathrm{ox}} \simeq 100$ nm) film of SiO_2 on a silicon substrate,
2. the coverage of the film with resist film,
3. the photolithographic definition of a certain pattern,
4. the reactive ion etching (tuned to allow the edge to have a high aspect ratio) of the unprotected SiO_2 terminating when the original silicon is exposed,
5. the reoxidation of the silicon with the formation of a thin (say 10 nm) SiO_2 film,
6. the conformal deposition of a polycrystalline silicon film (of thickness $t_{\mathrm{SiO}_2} \ll \frac{1}{2} t_{\mathrm{Si}}$),
7. the directional sputter etching of the silicon terminating when the underlying SiO_2 substrate is exposed, and
8. the selective etching with aqueous solution of HF of the original seed.

In extreme synthesis, the SPT can be rationalized by the steps SPT^0–SPT^3 given in Section 8.3.2. Figure 8.4 (*up*) shows the formalization of the various stages of SPT. While Figure 8.4 (*down*) shows the actual result of the application of the SPT. The SPT may be sophisticated via the deposition of a multilayered film; Figure 8.5 shows the sidewalls resulting from the deposition of a multilayer and compares it with what is really done in the original application of SPT—the insulation of the source and drain electrodes from the gate. It can be stressed that the estimated shapes is highly similar to the actual one.

This example shows that it is possible to transform a thickness t in a width w. Since t can be controlled on the scale of 1 nm, while width can be produced lithographically on the scale of 100 nm, this procedure enables patterning geometries with dimensions much lesser of those allowed by photolithography.

As a preliminary step of SPT, it is needed to obtain a starting seed. Figure A.14 shows how such a seed can be obtained: (a) Bare wafer; (b) Conformal deposition of oxide; (c) Conformal deposition of resist; (d) Polymerization of resist through masking; (e) Selective etching of resist; (f) Selective etching of oxide; and (g) Selective etching of polymerized resist;

Nanowires

The crossbar structure is a serious candidate for next-generation electronics (Heath et al., 1998b; Snider and Williams, 2007; Cerofolini and Romano, 2008; Csaba and Lugli, 2009).

For a planar arrangement of the wire arrays defining the crossbar the maximum cross-point density is given by p^{-2} where p is the pitch of each array The maximum density is thus controlled by the nanowire diameter: assuming a wire diameter of 25 nm and that the pitch is twice the wire diameter, their arrays could have a maximum linear density of 2×10^5 cm^{-1} only, irrespective of the method used for their production. This limit can however be overcome, without reducing the wire diameter, *arranging the nanowires in vertical arrays* (Cerofolini et al., 2010c; Cerofolini, 2007; Cerofolini et al., 2005b).

The nanowires vertical arrangement can be rationalized as follows:

1. N repetitions of
   (a) Conformal deposition of insulator X
   (b) Conformal deposition of insulator Y
2. Imperfect anisotropic etching: directional etching + isotropic etching with weight 9/10 and 1/10 respectively (see Section 3.6)
3. Selective etching of Y
4. Conformal deposition of Z
5. Isotropic etching

Figure A.12 shows how these steps can be realized, while figure A.13 shows the actual SEM images of the vertical arrangement of nanowires produced in (Ferri et al., shed). It worths noticing that the subsequent application of conformal deposition and isotropic etching (step 4 and 5) does not lead back to the situation obtained by step 3; an example of the fact that the AT operations are neither invertible or commutative.

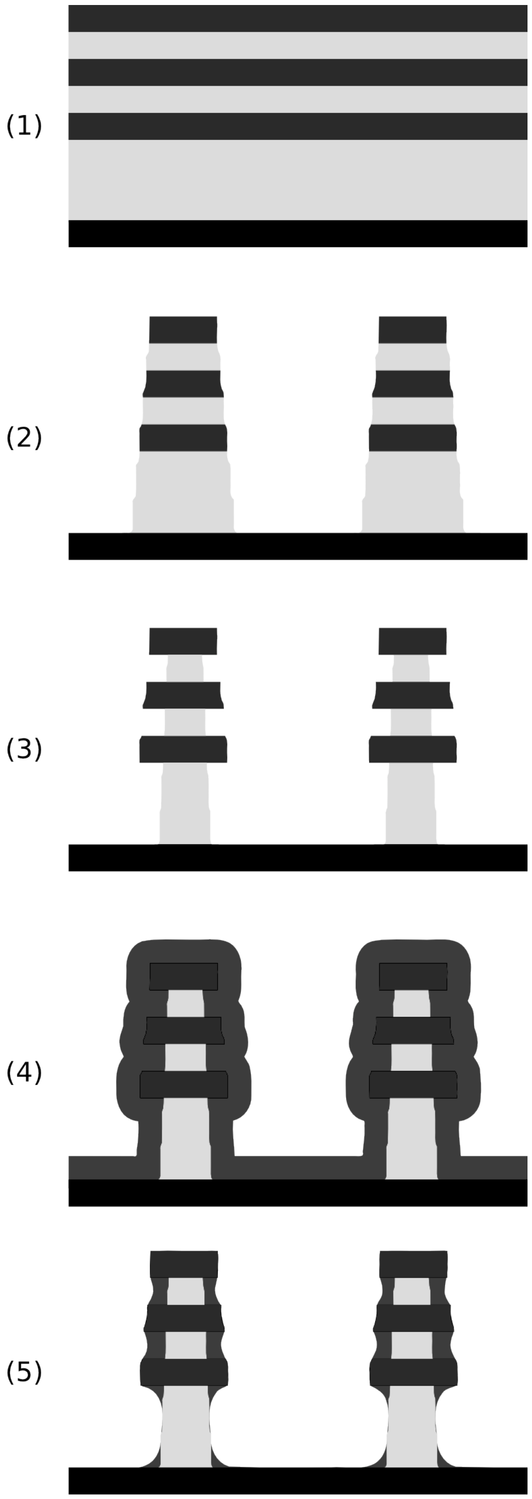


Figure A.12: Preparation of an uniform array of nanowires arranged perpendicularly to the surface

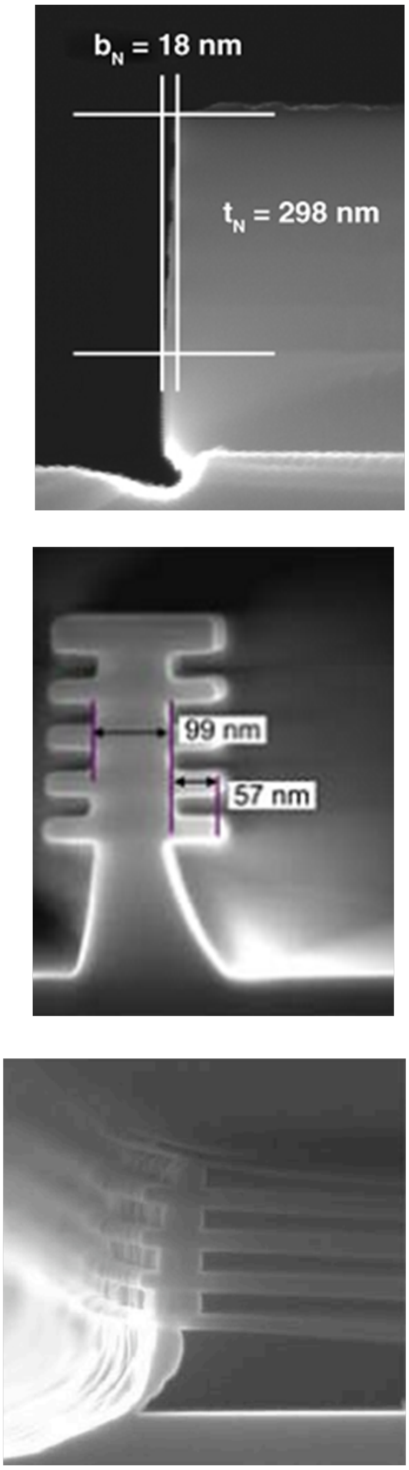


Figure A.13: SEM images of a vertical arrangement of nanowires (taken from (Ferri et al., shed)) corresponding to the AT steps (2), (3) and (5) respectively

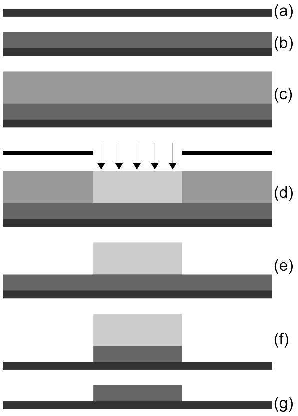


Figure A.14: The process to obtain a seed for the SPT starting from the bare wafer

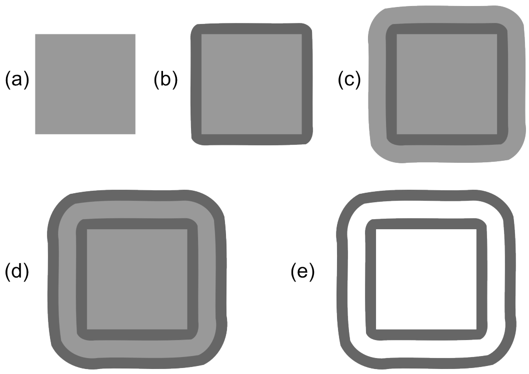


Figure A.15: Plan view of nested square borders. The underlying wafer is not shown

A.4.2 Geometrical examples

In this section I show an example of geometrical shape constructed with the AT operations instead of straightedge and compass. Figure A.15 shows two nested square borders obtained applying the following sequence: (a) Seed definition (see Figure A.14); (b) Conformal deposition of material Y; (c) Conformal deposition of material X; (d) Conformal deposition of material Y; and (e) Selective etching of material X.

A.5 Discussion

One of the considered example (SPT) shows that the processes of AT can be used to transform thickness into width. This fact is especially important because the current limit of photolithography (controlling width) is on the 100-nm length scale whereas thickness can be controlled on the 1-nm length scale. In this one has access to the nanoworld without employing extremely expensive processes like electron-beam lithography (Cerofolini and Ferla, 2002). This result may be achieved exploiting the same idea without however using the methods of AT (Wang et al., 2008a; Heath, 2008); it is however mentioned that while the SPT is consistent with the planar technology (Cerofolini et al., 2005a) (and also allows the preparation of complicated patterns (Cerofolini et al., 2008a)), those methods are not.

Of course, AT is a purely geometric theory so that it can provide only approximate descriptions of real physical processes. It works well only in the limit for which the described processes really exist. Actually none of the processes listed in Table 1 is ideal so that AT is only approximate. Since the considered processes modify the shape of the body, the theory is correct as far as their etching or deposition features do not depend on body shape. Since this property certainly fails on the nanometer length scale (actually the failure of this property may be considered as a definition of the nanoscopic regime), AT is not expected to be extensible to the nanometer length scale, where the shape can be adequately predicted only solving the reaction-diffusion equations describing the considered physical-chemical process.

The formalization of the processes of semiconductor manufacturing can be interesting *per se*, because it implies deeper understanding and generalizations of the current practice. However as interesting side-effect AT theory can be used also to give a sound framework for the exploitations of techniques from computational geometry and computer graphics in the modeling of semiconductor manufacturing.

In fact by representing a body with a vector object, the delta coverage of thickness δ can be obtained by offsetting the object's path (outside the object), i.e. by displacing it perpendicular to the path in each point. And the conformal coverage can be obtained by iterating this *offset path* operation with a small δ. Many vector drawing programs include it among their functionalities.

Moreover, in the directional processes, to identify the shadowed surface along a given direction it is possible to exploit the *shadow mapping* techniques commonly used in computer graphics.

The figures of this paper have been made by manually applying the AT theory in the vectorial graphics program Inkscape.

Bibliography

Abbott, J., Nagy, Z., Beyeler, F., and Nelson, B. (2007). Robotics in the small, part i: Microbotics. *Robotics Automation Magazine, IEEE*, 14(2):92 –103.

Adams, J. M. and Cory, S. (2007). The bcl-2 apoptotic switch in cancer development and therapy. *Oncogene*, 26(9):1324–1337.

Aerssens, J., Armstrong, M., Gilissen, R., and Cohen, N. (2001). The human genome: An introduction. *Oncologist*, 6:100–109.

Aguda, B. D. (2006). Modeling the cell division cycle. *Lect. Notes Math*, 1872:1–22.

Akkerman, H. B., Blom, P. W. M., de Leeuw, D. M., and de Boer, B. (2006). Towards molecular electronics with large-area molecular junctions. *Nature*, 441:69–71.

Amato, P. (2011). A formal theory of semiconductor manufacturing processes suitable for the exploitation of planar technology for 3d integration. *Science of Advanced Materials*, 3:444–454.

Amato, P., Masserini, M., Mauri, G., and Cerofolini, G. (2011). Early-stage diagnosis of endogenous diseases by swarms of nanobots: An applicative scenario. In Dorigo, M., editor, *Swarm intelligence: 7th International Conference, ANTS 2010, Brussels, Belgium, September 8-10, 2010. Proceedings*, volume 6234, pages 408–415. Springer Verlag, Berlin, Germany.

Arbuckle, D. and Requicha, A. (2004). Active self-assembly. In *ICRA*, pages 896–901. IEEE.

Aviram, A. and Ratner, M. (1974). Molecular rectifiers. *Chem. Phys. Lett*, 29:277–283.

Avnir, D., Farin, D., and Pfeifer, P. (1983). Chemistry in noninteger dimensions between two and three. ii. fractal surfaces of adsorbents. *J. Chem. Phys.*, 79:3566–3571.

Avnir, D., Farin, D., and Pfeifer, P. (1984). Molecular fractal surfaces. *Nature*, 308:261–263.

Bartha, K. and Rieger, H. (2006). Vascular network remodeling via vessel cooption, regression and growth in tumours. *J. Theor. Biol.*, 241:903–918.

Bauer, A., Jackson, T., and Jiangy, Y. (2007). A cell-based model exhibiting branching and anastomosis during tumour-induced angiogenesis. *Biophys. J.*, 92:3105–3121.

Beckman, R., Johnston-Halperin, E., Luo, Y., Green, J. E., and Heath, J. R. (2005). Bridging dimensions: Demultiplexing ultrahigh-density nanowire circuits. *Science*, 310:465–468.

Behkam, B. and Sitti, M. (2007). Bacterial flagella-based propulsion and on/off motion control of microscale objects. *Applied Physics Letters*, 90(2):023902.

Ben Jamaa, M. H., Atienza, D., Leblebici, Y., and De Micheli, G. (2009). Stochastic perturbative approach to design a defect-aware thresholder in the sense amplifier of crossbar memories. In *14th Asia and South Pacific Design Automation Conference*, Yokohama.

Ben Jamaa, M. H., Moselund, K. E., Atienza, D., Bouvet, D., Ionescu, A. M., Leblebici, Y., and De Micheli, G. (2007). Fault-tolerant multi-level logic decoder for nanoscale crossbar memory arrays. In *Proc. 2007 IEEE/ACM Int. Conf. on Computer-Aided Design*, pages 765–772, Piscataway, NJ. IEEE Press.

Ben Jamaa, M. H., Moselund, K. E., Atienza, D., Bouvet, D., Ionescu, A. M., Leblebici, Y., and De Micheli, G. (2008). Variability-aware design of multilevel logic decoders for nanoscale crossbar memories. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, pages 2053–2067.

Berg, H. (1993). *Random walks in biology*. Princeton Univ Pr.

Bonabeau, E., Dorigo, M., and Theraulaz, G. (1999). *Swarm intelligence: from natural to artificial systems*. Oxford University Press, Inc., New York, NY, USA.

Boukai, A. I., Bunimovich, Y., Tahir-Kheli, J., Yu, J.-K., Goddard III, W. A., and Heath, J. R. (2008). Silicon nanowires as efficient thermoelectric materials. *Nature*, 451:168–171.

Brannon-Peppas, L. and Blanchette, J. O. (2004). Nanoparticle and targeted systems for cancer therapy. *Advanced Drug Delivery Reviews*, 56:1649–1659.

Brown, J. M. (2000). Exploiting the hypoxic cancer cell: mechanisms and therapeutic strategies. *Mol. Med. Today*, 6:157–162.

Camarda, M., La Magna, A., and La Via, F. (2007). A kinetic Monte Carlo method on superlattices for the study of the defect formation in the growth of close packed structures. *J. Comput. Phys.*, 227(2):1075–1093.

Cancer Research UK (2011). Worlwide cancer incidence. `http://info.cancerresearchuk.org`.

Cardamone, D. M., Stafford, C. A., and Mazumdar, S. (2006). Controlling quantum transport through a single molecule. *Nano Lett.*, 6:2422–2426.

Cavalcanti, A. and Freitas, Jr., R. A. (2005). Nanorobotics control design: a collective behavior approach for medicine. *IEEE Trans Nanobioscience. 2005 Jun;4(2):133-40.*, 4:133–40.

Cavalcanti, A., Hogg, T., Shirinzadeh, B., and Liaw, H. (2006). Nanorobot communication techniques: A comprehensive tutorial. In *ICARCV*, pages 1–6. IEEE.

Cerofolini, G. (2007). Realistic limits to computation. ii. thetechnological side. *Appl.Phys.A*, 86:31–42.

Cerofolini, G. and Amato, P. (2012). Sensing strategies for early diagnosis of cancer by swarm of nanobots — an evidential paradigm. In Mavroidis, C. and Ferreira, A., editors, *Nanorobotics: Current Approaches and Techniques*, volume 6234, chapter 17. Springer Verlag, Berlin, Germany.

Cerofolini, G., Amato, P., and Romano, E. (2008a). The multi-spacer patterning technique: a non-lithographic technique for terascale integration. *Semiconductor Science and Technology*, 23(7):075020 (8pp).

Cerofolini, G. and Meda, L. (1989). *Physical Chemistry of, in and on Silicon*. Springer Verlag, Berlin, Germany, Berlin.

Cerofolini, G., Romano, E., and Amato, P. (2011a). Multi-spacer patterning technique. a general and versatile tool allowing the planar technology to face the nano era. In Sattler, K., editor, *Handbook of Nanophysics*. Chapter 17, CRC Press, Boca Raton, FL.

Cerofolini, G. F. (1981). Size, shape, growth and reproduction—towards a physical morphology. *Thin Solid Films*, 79:277–299.

Cerofolini, G. F. (1983a). The biomedium. adsorbed water as a model for the aqueous medium supporting life functions. *Adv. Colloid Interface Sci.*, 19:103–136.

Cerofolini, G. F. (1983b). The biomedium. adsorbed water as a model for the aqueous medium supporting life functions. *Adv. Colloid Interface Sci.*, 19:103–136.

Cerofolini, G. F. (2005). An extension of microelectronic technology to nanoelectronics. *Nanotechnol. E-Newslett.*, 7:5–6.

Cerofolini, G. F. (2010a). *Nanoscale Devices: Fabrication, Functionalization, and Accessibility from the Macroscopic World*. Springer Verlag, Berlin, Germany, Berlin.

Cerofolini, G. F. (2010b). Two routes to subcellular sensing. In Korkin, A., Krstić, P., and Wells, J., editors, *Nanotechnology for Electronics, Photonics, and Renewable Energy*, pages 153–182. Springer Verlag, Berlin, Germany, New York.

Cerofolini, G. F. and Amato, P. (2010). Swarms of nanobots for in vivo diagnosis of endogenous diseases. `http://asdn.net/asdn/life/nanorobots2.shtml`.

Cerofolini, G. F., Amato, P., Masserini, M., and Mauri, G. (2010a). A surveillance system for early-stage diagnosis of endogenous diseases by swarms of nanobots. *Advanced Science Letters*, 3:345–352.

Cerofolini, G. F., Arena, G., Camalleri, M., Galati, C., Reina, S., Renna, L., and Mascolo, D. (2005a). A hybrid approach to nanoelectronics. *Nanotechnology*, 16:1040–1047.

Cerofolini, G. F., Arena, G., Camalleri, M., Galati, C., Reina, S., Renna, L., Mascolo, D., and Nosik, V. (2005b). Strategies for nanoelectronics. *Microelectr. Eng.*, 81:405–419.

Cerofolini, G. F., Casuscelli, V., Cimmino, A., Di Matteo, A., Di Palma, V., Mascolo, D., Romanelli, E., Volpe, M. V., and Romano, E. (2007). Steps farther toward micro-nano-mole integration via the multispacer patterning technique. *Semicond. Sci. Technol.*, 22:1053–1060.

Cerofolini, G. F. and Ferla, G. (2002). Toward a hybrid micro-nanoelectronics. *J. Nanoparticle Res.*, 4:185–191.

Cerofolini, G. F., Ferri, M., Romano, E., Roncaglia, A., Selezneva, E., Arcari, A., Suriano, F., Veronese, G. P., Solmi, S., and Narducci, D. (2010b). Industrially scalable process for silicon nanowires for Seebeck generators. `arXiv:1011.4700v1`, downloaded on February 15, 2011.

Cerofolini, G. F., Ferri, M., Romano, E., Suriano, F., Veronese, G. P., Solmi, S., and Narducci, D. (2010c). Tera scale integration via a redesign of the crossbar based on a vertical arrangement of poly-Si nanowires. *Semicond. Sci. Technol.*, 25:095011/1–7.

Cerofolini, G. F., Ferri, M., Romano, E., Suriano, F., Veronese, G. P., Solmi, S., and Narducci, D. (2011b). Crossbar architecture for tera scale integration. *Semicond. Sci. Technol.*, 26:045005/1–8.

Cerofolini, G. F. and Mascolo, D. (2006). A hybrid micro-nano-molecular route for nonvolatile memories. *Semicond. Sci. Technol.*, 21:1315–1325.

Cerofolini, G. F., Narducci, D., Amato, P., and Romano, E. (2008b). Fractal nanotechnology. *Nanoscale Research Letters*, 3(10):381–385.

Cerofolini, G. F. and Romano, E. (2008). Molecular electronics *in silico*. *Appl. Phys. A*, 91:181–210.

Chandran, K., Rittgers, S., and Yoganathan, A. (2007). *Biofluid Mechanics: The Human Circulation.* CRC Press, Boca Raton, FL.

Changizi, M. A. and Cherniak, C. (2000). Modeling the large-scale geometry of human coronary arteries. *Can. J. Physiol. Pharmacol.*, 78:603–611.

Chen, J., Reed, M. A., Rawlett, A. M., and Tour, J. M. (1999). Large on-off ratios and negative differential resistance in a molecular electronic device. *Science*, 286:1550–1552.

Choi, Y.-K., Lee, J. S., Zhu, J., Somorjai, G. A., Lee, L. P., and Bokor, J. (2003a). Sublithographic nanofabrication technology for nanocatalysts and dna chips. *J. Vac. Sci. Technol. B*, 21:2951–2595.

Choi, Y.-K., Zhu, J., Grunes, J., Bokor, J., and Somorjai, G. A. (2003b). Fabrication of sub-10-nm silicon nanowire arrays by size reduction lithography. *J. Phys. Chem. B*, 107:3340–3343.

Corradi, P., Scholz, O., Knoll, T., Menciassi, A., and Dario, P. (2009). An optical system for communication and sensing in millimetre-sized swarming microrobots. *Journal of Micromechanics and Microengineering*, 19(1):015022.

Csaba, G. and Lugli, P. (2009). Read-out design rules for molecular crossbar architectures. *IEEE Trans. Nanotechnol*, 8:369–374.

Dasgupta, D. (2006). Advances in artificial immune systems. *Computational Intelligence Magazine, IEEE*, 1(4):40–49.

Decuzzi, P. and Ferrari, M. (2008). Design maps for nanoparticles targeting the diseased microvasculature. *Biomaterials*, 29(3):377 – 384.

DeHon, A., Lincoln, P., and Savage, J. E. (2003). Stochastic assembly of sublithographic nanoscale interfaces. *IEEE Trans. Nanotechnol.*, 2:165–174.

Donald, B., Levey, C., Mcgray, C., Paprotny, I., and Rus, D. (2006). An untethered, electrostatic, globally controllable MEMS micro-robot. *Journal of Microelectromechanical Systems*, 15(1):1–15.

Dorigo, M. (2005). SWARM-BOT: An experiment in swarm robotics. In Arabshahi, P. and Martinoli, A., editors, *Proceedings of the 2005 IEEE Swarm Intelligence Symposium*, pages 192–200.

Dorigo, M. and Şahin, E. (2004). Guest editorial. Special issue: Swarm robotics. *Autonomous Robots*, 17(2–3):111–113.

Dorigo, M., Floreano, D., Gambardella, L., Mondada, F., Nolfi, S., Baaboura, T., Birattari, M., Bonani, M., Brambilla, M., Brutschy, A., Burnier, D., Campo, A., Christensen, A., Decugnière, A., Caro, G. D., Ducatelle, F., Ferrante, E., Förster, A., Guzzi, J., Longchamp, V., Magnenat, S., Gonzalez, J. M., Mathews, N., de Oca, M. M., O'Grady, R., Pinciroli, C., Pini, G., Rétornaz, P., Roberts, J., Sperati, V., Stirling, T., Stranieri, A., Stuetzle, T., Trianni, V., Tuci, E., Turgut, A., and Vaussard, F. (2012). Swarmanoid: a novel concept for the study of heterogeneous robotic swarms. *IEEE Robotics & Automation Magazine*, page in press.

Dorigo, M., Maniezzo, V., and Colorni, A. (1996). Ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 26(1):29–41.

Dorigo, M., Trianni, V., Şahin, E., Groß R., Labella, T. H., Baldassarre, G., Nolfi, S., Deneubourg, J. L., Mondada, F., Floreano, D., and Gambardella, L. M. (2004). Evolving self-organizing behaviors for a swarm-bot. *Autonomous Robots*, 17(2–3):223–245.

Duncan, R. (2003). The dawning era of polymer therapeutics. *Nat Rev Drug Discov*, 2(5):347–360.

Dunn, W. (2008). Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Microcirculation*, 5:011001.

Edqvist, E., Snis, N., Mohr, R. C., Scholz, O., Corradi, P., Gao, J., Diéguez, A., Wyrsch, N., and Johansson, S. (2009). Evaluation of building technology for mass producible millimetre-sized robots using flexible printed circuit boards. *Journal of Micromechanics and Microengineering*, 19(7):075011.

Eigler, D. M. and Schweizer, E. K. (1990). Positioning single atoms with a scanning tunnelling microscope. *Nature*, 344(6266):524–526.

Encapsula Nano Sciences (2011). Technical summary - an introduction to lipid nanoparticles. `http://www.encapsula.com/company.html`.

Engl, H. W., Flamm, C., Kügler, P., Lu, J., Müller, S., and Schuster, P. (2009). Inverse problems in systems biology. *Inverse Problems*, 25:123014.

Erhardt, P. (2003). Film deposition methods. In (Waser, 2003), chapter 8, pages 201–221.

Faggin, F. (2009). Dalla microelettronica alla nanoelettronica (in italian). *Mondo Digitale*, 29(1):3–10.

Falconer, K. (2003). *Fractal Geometry: Mathematical Foundations and Applications*. Wiley.

Ferrari, M. (2005a). Cancer nanotechnology: Opportunities and challenges. *Nature Reviews*, 5:161–171.

Ferrari, M. (2005b). Nanovector therapeutics. *Current opinion in chemical biology*, 9(4):343–346. Cited By (since 1996): 56.

Ferrari, M. (2008). Nanogeometry: Beyond drug delivery. *Nat Nano*, 3(3):131–132.

Ferrari, M. (2010). Experimental therapies: Vectoring sirna therapeutics into the clinic. *Nat Rev Clin Oncol*, 7(9):485–486.

Ferri, M., Roncaglia, A., Solmi, S., Suriano, F., Cerofolini, G. F., Romano, E., and Narducci, D. (2010, To be published). Ultradense silicon nanowire arrays produced via top-down planar technology. *Microelectronic Engineering*.

Feynman, R. (1992). There's plenty of room at the bottom [data storage]. *Microelectromechanical Systems, Journal of*, 1(1):60–66.

Flanders, D. C. and Efremow, N. N. (1983). Generation of < 50 nm period gratings using edge defined techniques. *J. Vac. Sci. Technol. B*, 1:1105.

Folkman, J. (1971). Tumour angiogenesis: therapeutic implication. *N. Engl. J. Med.*, 285:1182–1186.

Freitas, R. (1999). *Nanomedicine, Volume I: Basic Capabilities*. Landes Bioscience, Georgetown.

Freitas, R. (2005). Current status of nanomedicine and medical nanorobotics. *J. Comput. Theor. Nanosci*, 2:1–25.

Freitas, R. (2006). Pharmacytes: An ideal vehicle for targeted drug delivery. *Journal of Nanoscience and Nanotechnology*, 6:2769–2775(7).

Futrelle, J. (1973). The scarlet thread. In Bleiler, E. F., editor, *Best "Thinking Machine" Detective Stories*, pages 48–76. Dover, New York, NY.

Gabrys, E., Rybaczuk, M., and Kedzia, A. (2005). Fractal models of circulatory system. symmetrical and asymmetrical approach comparison. *Chaos, Solitons & Fractals*, 24(3):707–715.

Gates, D. B., Xu, Q. B., Stewart, M., Ryan, D., Willson, C. G., and Whitesides, G. M. (2005). New approaches to nanofabrication. *Chem. Rev.*, 105:1171–1196.

Gazit, Y., Berk, D. A., Leunig, M., Baxter, L. T., and Jain, R. K. (1995). Scale-invariant behavior and vascular network formation in normal and tumor tissue. *Phys. Rev. Lett.*, 75:2428–2431.

Gentile, F., Chiappini, C., Fine, D., Bhavane, R., Peluccio, M., Cheng, M. M.-C., Liu, X., Ferrari, M., and Decuzzi, P. (2008). The effect of shape on the margination dynamics of non-neutrally buoyant particles in two-dimensional shear flows. *Journal of Biomechanics*, 41(10):2312 – 2318.

Glass, J., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M., Maruf, M., III, C. H., Smith, H., and Venter, J. (2006). Essential genes of a minimal bacterium. *Proc. Nat. Acad. Sci. USA*, 103:425–430.

Green, J. E., Cho, i. J. W., Boukai, A., Bunimovich, Y., Johnston-Halperin, E., Delonno, E., Y., L., Sheriff, B. A., Xu, K., Shin, Y. S., Tseng, H.-R., Stoddart, J. F., and Heath, J. R. (2007). A 160-kilobit molecular electronic memory patterned at 10^{11} bits per square centimetre. *Nature*, 405:414–417.

Groß R. and Dorigo, M. (2004). Cooperative transport of objects of different shapes and sizes. In Dorigo, M., Birattari, M., Blum, C., Gambardella, L. M., Mondada, F., and Stützle, T., editors, *Ant Colony Optimiza- tion and Swarm Intelligence – Proceedings of ANTS 2004 – Fourth International Workshop*, volume 3172, pages 107–118. Springer Verlag, Berlin, Germany.

Guyton, A. C. (1976). *Textbook of Medical Physiology*. W.B. Saunders, 5th edition.

Hanahan, D. and Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell*, 100(1):57–70.

Hanahan, D. and Weinberg, R. A. (2011). The hallmarks of cancer: The next generation. *Cell*, 144(5):646–674.

Hansen, J. T. and Koeppen, B. M. (2002). *Netter's Atlas of Human Physiology*. Icon Learning Systems, Teterboro, NJ.

Harvey, W. (1910). On the motion of the heart and blood in animals. In *Scientific papers; physiology, medicine, surgery, geology, with introductions, notes and illustrations*, volume 38 of *The Harvard classics*. P. F. Collier & Son, New York.

Heath, J. (2008). Superlattice nanowire pattern transfer (SNAP). *Acc. Chem. Res.*, 41:1609–1617.

Heath, J., Kuekes, P., Snider, G., and Williams, R. (1998a). A defect-tolerant computer architecture: Opportunities for nanotechnology. *Science*, 280:1716–1721.

Heath, J. R., Kuekes, P. J., Snidera, G. S., and Williams, R. S. (1998b). A defect-tolerant computer architecture: Opportunities for nanotechnology. *Science*, 280:1716–1721.

Hochbaum, A. I., Chen, R. K., Delgado, R. D., Liang, W., Garnett, E. C., Najarian, M., Majumdar, A., and Yang, P. (2008). Enhanced thermoelectric performance of rough silicon nanowires. *Nature*, 451:168–171.

Hogg, T. (2007). Distributed control of microscopic robots in biomedical applications. In Prokopenko, M., editor, *Advances in Applied Self-organizing Systems*, Advanced Information and Knowledge Processing, pages 147–174. Springer Verlag, Berlin, Germany, New York.

Hood, L., Heath, J. R., Phelps, M. E., and Lin, B. (2004). Systems biology and new technologies enable predictive and preventative medicine. *Science*, 306:640–643.

Huang, Y., Duan, X., Cui, Y., Lauhon, L. J., Kim, K.-H., and Lieber, C. M. (2001). Logic gates and computation from assembled nanowire building blocks. *Science*, 294:1313–1317.

Hudetz, A. (1997). Blood flow in the cerebral capillary network: a review emphasizing observations with intravital microscopy. *Microcirculation*, 4:233–252.

Ito, T. and Okazaki, S. (2000). Pushing the limits of lithography. *Nature*, 406:1027–1031.

ITRS (2007). International technology roadmap for semiconductors. `http://www.itrs.net/reports.html`.

ITRS (2009). International technology roadmap for semiconductors. `http://www.itrs.net/links/2009ITRS/Home2009.htm` on January 15, 2011.

James, N. (2011). *Cancer*, volume 267 of *Very Short Introductions*. Oxford University Press, Inc., Oxford.

Joachim, C., Gimzewski, J. K., and Aviram, A. (2000). Electronics using hybrid-molecular and mono-molecular devices. *Nature*, 408:541–548.

Joachim, C. and Ratner, M. A. (2005). Molecular electronics: Some views on transport junctions and beyond. *Proc. Natl. Acad. Sci.*, 102:8801–8808.

Kamiya, A., Takeda, S., and Shibata, M. (1987). Optimum capillary number for oxygen delivery to tissue in man. *Bull. Math. Biol.*, 49:351–361.

Kamiya, A. and Togawa, T. (1972). Optimal branching structure of the vascular tree. *Bull. Math. Biol.*, 34:431–438.

Keith, B. and Simon, M. C. (2007). Hypoxia-inducible factors, stem cells, and cancer. *Cell*, 129:465–472.

Kennedy, D. A., Lee, T., and Seely, D. (2009). A comparative review of thermography as a breast cancer screening technique. *Integr. Cancer Ther.*, 8:9–16.

Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In *Proc. IEEE Int'l Conference on Neural Networks, 4 (Perth, Australia)*, pages 1942–1948.

Kigami, J. (2001). *Analysis on Fractals*. Cambridge Univ. Press, Cambridge, MA.

Kim, M. and Breuer, K. (2008). Microfluidic pump powered by self-organizing bacteria. *Small*, 4(1):111–118.

Klawansky, S. and Fox, M. S. (1984). A growth rate distribution model for the age dependence of human cancer incidence: A proposed role for promotion in cancer of the lung and breast. *J. Theor. Biol.*, 111:531–587.

Lankhorst, M. H. R., Ketelaars, B. W. S. M. M., and Wolters, R. M. (2005). Low-cost and nanoscale non-volatile memory concept for future silicon chips. *Nature Mater.*, 4:347–352.

Lau, C. N., Stewart, D. R., Williams, R. S., and Bockrath, D. (2004). Direct observation of nanoscale switching centers in metal/molecule/metal structures. *Nano Lett.*, 4:569–572.

Lee, S., Ferrari, M., and Decuzzi, P. (2009). Shaping nano-/micro-particles for enhanced vascular interaction in laminar flows. *Nanotechnology*, 20:495101.

Likharev, K. K. and Strukov, D. B. (2005). CMOL: Devices, circuits, and architectures. In Cuniberti, G., Fagas, G., and Richter, K. B., editors, *Introducing Molecular Electronics*, pages 447–477. Springer Verlag, Berlin, Germany.

Luk, K. H., Hulse, R. M., and Phillips, T. L. (1980). Hyperthermia in cancer therapy. *West. J. Med.*, 132:179–185.

Luo, Y., Collier, C. P., Jeppesen, J. O., Nielsen, K. A., Delonno, E., Ho, G., Perkins, J., Tseng, H.-R., Yamamoto, T., Stoddart, J. F., and Heath, J. R. (2002). Two-dimensional molecular electronics circuits. *Chem. Phys. Chem.*, 3:519–525.

Mandelbrot, B. and Frame, M. (1999). The canopy and shortest path in a self-contacting fractal tree. *The Mathematical Intelligencer*, 21:18–27. 10.1007/BF03024842.

Mandelbrot, B. B. (1983). *The Fractal Geometry of Nature*. W. H. Freedman and Co., New York.

Martel, S. and Mohammadi, M. (2010). Using a swarm of self-propelled natural microrobots in the form of flagellated bacteria to perform complex micro-assembly tasks. In *Proceedings of the 2010 IEEE International Conference on Robotics and Automation (ICRA)*, Anchorage, Alaska.

Mavroidis, C. and Ferreira, A. (2009). Editorial: Special issue on current state of the art and future challenges in nanorobotics. *I. J. Robotic Res.*, 28(4):419–420.

Melosh, N. A., Boukai, A., Diana, F., Gerardot, B., Badolato, A., and Heath, J. R. (2003). Ultrahigh-density nanowire lattices and circuits. *Science*, 300:112–115.

Mendes, P. M., Flood, A. H., and Stoddart, J. F. (2005). Nanoelectronic devices from self-organized molecular switches. *Appl.Phys.A*, 80:1197–1209.

Mohanraj, V. and Chen, Y. (2006). Nanoparticles – a review. *Tropical Journal of Pharmaceutical Research*, 5:561–573.

Monod, J. (1970). *Le Hasard et la Necessité*. Editions du Seuil,, Paris.

Montemagno, C. and Bachand, G. (1999). Constructing nanomechanical devices powered by biomolecular motors. *Nanotechnology*, 10(3):225–231.

Murray, C. D. (1926). The physiological principle of minimum work applied to the angle of branching of arteries. *J. Gen. Physiol.*, 9(6):835–841.

Nagy, Z., Harada, K., Flickiger, M., Susilo, E., Kaliakatsos, I. K., Menciassi, A., Hawkes, E., Abbott, J. J., Dario, P., and Nelson, B. J. (2009). Assembling reconfigurable endoluminal surgical systems: Opportunities and challenges. *International Journal of Biomechatronics and Biomedical Robotics*, 1(1):3–16.

Natelson, D., Willett, R. L., West, K. W., and Pfeiffer, L. N. (2000). Fabrication of extremely narrow metal wires. *Appl. Phys. Lett.*, 77:1991–1993.

National Cancer Institute (2010). Cancer topics. `http://www.cancer.gov`.

Nelson, B., Kaliakatsos, I., and Abbott, J. (2010). Microrobots for minimally invasive medicine. *Annual Review of Biomedical Engineering*, 12(1):55–85.

Nelson, B. J. and Dong, L. (2010). Nanorobotics. In Bhushan, B., editor, *Springer Handbook of Nanotechnology*, pages 1633–1659. Springer Verlag, Berlin, Germany.

O'Grady, R., Pinciroli, C., Grofl, R., Christensen, A., Mondada, F., Bonani, M., and Dorigo, M. (2011). Swarm-bots to the rescue. In Kampis, G., Karsai, I., and Szathmary, E., editors, *European Conference on Artificial Life (ECAL 2009)*, volume 5777 of *Lecture Notes in Artificial Intelligence*, pages 165–172, Berlin and Germany. Springer Verlag, Berlin, Germany.

Okazaki, S. and Moers, J. (2003). Lithography. In (Waser, 2003), chapter 9, pages 225–247.

Osgood, W. F. (1903). A Jordan curve of positive area. *Transactions of the American Mathematical Society*, 4:107–112.

Pfeifer, P. and Avnir, D. (1983). Chemistry in noninteger dimensions between two and three. i. fractal theory of heterogeneous surfaces. *J. Chem. Phys.*, 79:3558–3565.

Pinciroli, C., Trianni, V., O'Grady, R., Pini, G., Brutschy, A., Brambilla, M., Mathews, N., Ferrante, E., Caro, G. D., Ducatelle, F., Stirling, T., Gutiérrez, A., Gambardella, L., and Dorigo, M. (2011). ARGoS: A modular, multi-engine simulator for heterogeneous swarm robotics. In *Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'11)*, pages 5027–5034, Los Alamitos, CA. IEEE Computer Society Press.

Ploegh, H. (2007). A lipid-based model for the creation of an escape hatch from the endoplasmic reticulum. *Nature*, 448:435–438.

Plummer, J. and Griffin, P. (2001). Material and process limits in silicon VLSI technology. *Proc. IEEE*, 89:240–258.

Pocock, G. and Richards, C. D. (2006). *Human Physiology. The Basis of Medicine*. Oxford Univ. Press, Oxford, 3rd edition.

Purcell, E. (1977). Life at low Reynolds number. *Am. J. Phys*, 45(1):3–11.

Raff, H. and Levitzky, M. G. (2011). *Medical Physiology : A Systems Approach*. McGraw-Hill Professional, 1 edition.

Rai-Choudhury, P. (1997). *Handbook of Microlithography, Micromachining, and Microfabrication*. SPIE Society of Photo-Optical Instrumentation.

Rashevsky, N. (1960). *Mathematical Biophysics*. Dover, New York, 3rd edition.

Reed, M. A., Chen, J., Rawlett, A. M., Price, D. W., and Tour, J. M. (2001). Molecular random access memory cell. *Applied Physics Letters*, 78:3735–3737.

Requicha, A. A. G. (2003). Nanorobots, NEMS, and nanoassembly. *Proceedings of the IEEE*, 91(11):1922–1933.

Richter, J. P. (1970). *The Literary Works of Leonardo Da Vinci*. Phaidon Publishers Inc., New York, 3rd edition.

Robergs, R. A., Ghiasvand, F., and Parker, D. (2004). Biochemistry of exercise-induced metabolic acidosis. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, 287:R502–R516.

Roukes, M. (2007). Plenty of room indeed. *Sci. Am. Repts.*, 17(3):4–11.

Roundy, S., Steingart, D.and Frechette, L., Wright, P., and Rabaey, J. (2004). Power sources for wireless sensor networks. In Karl, H., Wolisz, A., and Willig, A., editors, *Wireless Sensor Networks*, volume 2920 of *Lecture Notes in Computer Science*, pages 1–17. Springer Verlag, Berlin, Germany.

Sakamoto, J. H., van de Ven, A. L., Godin, B., Blanco, E., Serda, R. E., Grattoni, A., Ziemys, A., Bouamrani, A., Hu, T., Ranganathan, S. I., Rosa, E. D., Martinez, J. O., Smid, C. A., Buchanan, R. M., Lee, S.-Y., Srinivasan, S., Landry, M., Meyn, A., Tasciotti, E., Liu, X., Decuzzi, P., and Ferrari, M. (2010). Enabling individualized therapy through nanotechnology. *Pharmacological Research*, 62(2):57 – 89.

Schneider, S. and McClatchie, S. (2003). Material removing techniques—etching and chemical-mechanical polishing. In (Waser, 2003), chapter 10, pages 251–271.

Seeley, T. D. (1996). *The Wisdom of the Hive: The Social Physiology of Honey Bee Colonies*. Harvard University Press.

Seeley, T. D. (2010). *Honeybee Democracy*. Princeton University Press.

Service, R. F. (2003). Next-generation technology hits an early midlife crisis. *Science*, 302:556–559.

Seyfried, J., Szymanski, M., Bender, N., Estaña, R., Thiel, M., and Wörn, H. (2005). The I-SWARM project: Intelligent small world autonomous robots for micro-manipulation. In Azahin, E. and Spears, W., editors, *Swarm Robotics*, volume 3342 of *Lecture Notes in Computer Science*, pages 70–83. Springer Verlag, Berlin, Germany.

Shlesinger, M. F. and West, B. J. (1991). Complex fractal dimension of the bronchial tree. *Phys. Rev. Lett.*, 67:2106–2108.

Simons, M. J. and Pellionisz, A. J. (2005). Genomics, morphogenesis and biophysics: Triangulation of Purkinje cell development. *Cerebellum*, 4:1–9.

Sitti, M. (2001). Survey of nanomanipulation systems. In *Nanotechnology, 2001. IEEE-NANO 2001. Proceedings of the 2001 1st IEEE Conference on*, pages 75–80.

Sitti, M. (2007). Microscale and nanoscale robotics systems [grand challenges of robotics]. *Robotics Automation Magazine, IEEE*, 14(1):53–60.

Snider, G., Kuekes, P., Hogg, T., and Williams, R. (2005). Nanoelectronic architectures. *Appl. Phys. A*, 80:1183–1195.

Snider, G. S. and Williams, R. S. (2007). Nano/CMOS architectures using a field-programmable nanowire interconnect. *Nanotechnology*, 18:035204.

Song, Y. and Sitti, M. (2007). Surface-tension-driven biologically inspired water strider robots: Theory and experiments. *Robotics, IEEE Transactions on*, 23(3):578 –589.

Steeves, R. A. (1992). Hyperthermia in cancer therapy: Where are we today and where are we going? *Bull. N. Y. Acad. Med.*, 68:341–350.

Stern, E., Klemic, J. F., Routenberg, D. A., Wyrembak, P. N., Turner-Evans, D. B., Hamilton, A. D., LaVan, D. A., Fahmy, T. M., and Reed, M. A. (2007). Label-free immunodetection with CMOS-compatible semiconducting nanowires. *Nature*, 445:519–522.

Stewart, D. R., Ohlberg, D. A. A., Beck, P., Chen, Y., Williams, R. S., Jeppesen, J. O., Nielsen, K. A., and Stoddart, J. F. (2004a). Molecule-independent electrical switching in pt/organic monolayer/ti devices. *Nano Lett.*, 4:133–136.

Stewart, M. P., Maya, F., Kosynkin, D. V., Dirk, S. M., Stapleton, J. J., McGuiness, C. L., Allara, D. L., and Tour, J. M. (2004b). Direct covalent grafting of conjugated molecules onto Si, GaAs and Pd surfaces from aryldiazonium salts. *J. Am. Chem. Soc.*, 126:370–378.

Strukov, D. B. and Likharev, K. K. (2005). Prospects for terabit-scale nanoelectronic memories. *Nanotechnology*, 16:137–148.

Strukov, D. B., Snider, G. S., Stewart, D. R., and Williams, R. S. (2008). The missing memristor found. *Nature*, 453:80–83.

Stücker, M., Baier, V., Reuther, T., Hoffmann, K., Kellam, K., and Altmeyer, P. (1996). Capillary blood cell velocity in human skin capillaries located perpendicularly to the skin surface: measured by a new laser doppler anemometer. *Microvascular Res.*, 56:188–192.

Talmadge, J. E. and Fidler, I. J. (2010). AACR centennial series: The biology of cancer metastasis: Historical perspective. *Cancer Research*, 70(14):5649–5669.

Tanaka, M. and Sackmann, E. (2005). Polymer-supported membranes as models of the cell surface. *Nature*, 437(7059):656–663.

Tannock, I. F. and Rotin, D. (1989). Acid pH in tumors and its potential for therapeutic exploitation. *Cancer Res.*, 49:4373–4384.

Tour, J. M., Rawlett, A. M., Kozaki, M., Yao, Y., Jagessar, R. C., Dirk, S. M., Price, D. W., Reed, M. A., Zhou, C.-W., Chen, J., Wang, W., and Campbell, I. (2001). Synthesis and preliminary testing of molecular wires and devices. *Chem. Eur. J.*, 7:5118–5134.

Tuteja, A., Choi, W., Ma, M., Mabry, J. M., Mazzella, S. A., Rutledge, G. C., McKinley, G. H., and Cohen, R. E. (2007). Designing superoleophobic surfaces. *Science*, 318:1618–1622.

Vink, H. and Duling, B. (1996). Identification of distinct domains of macromolecules, erythrocytes, and leucocytes within mammalian capillaries. *Circulation Res*, 79:581–589.

Visvader, J. (2011). Cells of origin in cancer. *Nature*, 469:314–322.

Wang, D., Sheriff, B., McAlpine, M., and Heath, J. (2008a). Development of ultra-high density silicon nanowire arrays for electronics applications. *Nano Res.*, 1:9–21.

Wang, D., Sheriff, B. A., McAlpine, M., and Heath, J. R. (2008b). Development of ultra-high density silicon nanowire arrays for electronics applications. *Nano Res.*, 1:9–21.

Waser, R., editor (2003). *Nanoelectronics and Information Technology. Advanced Electronic Materials and Novel Devices*. Wiley–VCH, Weinheim.

Watson, J., Baker, T., Bell, S., Gann, A., Levine, M., and Losick, R. (2008). *Molecular biology of the gene*. Pearson Education, London, 6th edition.

Welter, M. and Rieger, H. (2010). Physical determinants of vascular network remodeling during tumour growth. *Eur. Phys. J. E*, 33:149–163.

Whitesides, G. M. and Love, J. C. (2007). The art of building small. *Sci. Am. Repts.*, 17:12–21.

Winfield, A. (2000). Distributed sensing and data collection via broken ad hoc wireless connected networks of mobile robots. In Parker, L. E., Bekey, G. A., and Barhen, J., editors, *DARS*, pages 273–282. Springer Verlag, Berlin, Germany.

Winfield, A. (2012). *Robotics*, volume 330 of *Very Short Introductions*. Oxford University Press, Inc., Oxford.

Wurman, P. R., D'Andrea, R., and Mountz, M. (2007). Coordinating hundreds of cooperative, autonomous vehicles in warehouses. In *Proceedings of the 19th national conference on Innovative applications of artificial intelligence - Volume 2*, IAAI'07, pages 1752–1759. AAAI Press.

Zhitenev, N. B., Jiang, W., Erbe, A., Bao, Z., Garfunkel, E., Tennant, D. M., and Cirelli, R. A. (2006). Control of topography, stress and diffusion at molecule–metal interfaces. *Nanotechnology*, 17:1272–1277.

Zhong, Z., Wang, D., Cui, Y., Bockrath, M. W., and Lieber, C. M. (2003). Nanowire crossbar arrays as address decoders for integrated nanosystems. *Science*, 302:1377–1379.