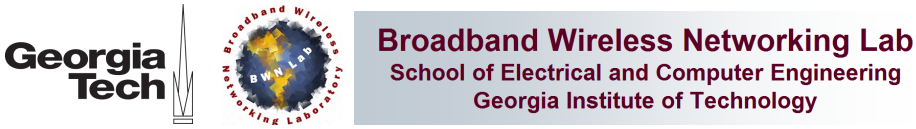


Capacity and Delay of Bacteria-Based Communication in Nanonetworks

Final Year Project

Escola Tècnica Superior de Telecomunicacions de Barcelona (ETSETB)

Author:
Albert Mestres Sugrañes

Advisor:
Dr. Ian F. Akyildiz

February 10, 2012

"Intelligence is the ability to avoid doing work, yet getting the work done"
– Linus Torvalds

Acknowledgements

I would like to thank everybody who has made possible this thesis and my stay at the Georgia Institute of Technology. First of all, I would like to thank Prof. Ian F. Akyildiz for giving me the opportunity to work on my thesis under his supervision and for the help and guidance he provided me during my stay at the BWN lab.

I would also like to acknowledge all members of the BWN Lab, where I enjoyed an amazing atmosphere. The members of the nanocommunication group deserve a special mention for the advices and the support I received from them: Massimiliano Pierobon, Josep Miquel Jornet, Raül Gómez, Francesco Rocca Cirasa, Sam Meng and Ignacio Llatser.

I would like to acknowledge the N3Cat faculty who offered me this opportunity, Eduard Alarcón, Josep Solé and Albert Cabellos, and all the staff of ETSETB that made possible this experience. I would also like to thank the economical support from ETSETB, AGAUR and Obra Social Bancaja, whixh I greatly appreciated.

I would like to dedicate some words to all the people I met during my stay in Atlanta, people from around the world that made me not only enjoy my research time, but also every moment I spent in Atlanta. Thank you. I do not want to forget all the friends from UPC who have accompanied me in all over those five years, and the rest of the friends I have in Barcelona.

Last but not least, I will be eternally thankful to my family for their love, support and understanding throughout my studies.

Abstract

Nanotechnology is enabling the miniaturization of electrical and mechanical components, which lead into the possibility of designing devices which can operate at the nanoscale. One promising application is in the nanosensing field, where different fields are potentially interested, such as the biomedical, environmental and industrial.

These nanodevices are expected to be strongly limited in capabilities and in actuation range, which implies a low performance when acting alone. In order to cover large areas and expand their capabilities it is necessary a collaborative effort among nanodevices by enabling the their interconnection into communication networks, also known as nanonetworks.

The classical communication paradigms are not valid in the nanoscale, so novel communications alternatives have been proposed in order to enable the communication between nanodevices. This work is based on flagelled bacteria communication, which consists of the use of bacteria to carry information. The nanodevice encodes the information in a DNA molecule, then the molecule is copied inside a bacterium and finally it moves using natural processes to the receiver, where the information is read.

The work we presented in this thesis is focused on the analysis of the capacity and the delay of this communication paradigm. In previous work, the objective was to define how a bacteria-based system works and to propose basic communication steps, giving less importance to analytical results. For example, no analytical expression was obtained for the propagation of bacteria, but the numerical values were obtained from a simulation tool.

The first contribution of this work is a complete analysis of the effect of the mutations in the reliability of the bacteria-based communication system by taking into account that multiple mutations can affect the information as it propagates through the system.

The second contribution of this work is an analysis of the propagation of the bacteria. Bacteria move randomly in the medium, but these movements are biased in order to follow a certain chemical, which is used to direct the bacteria towards receiver. As a result of the analysis we obtain an expression for the delay which avoids the use of simulation tools in the analysis of the propagation of information in flagellated bacteria-based communication systems.

Finally, we introduce two completely different communication systems, namely, the first synchronous and the second asynchronous. The synchronous system assumes that the nanodevices are synchronized, so a basic timing can be used among them. This system can present some advantages, but it is characterized by higher complexity in terms of logic elements, high power consumption... The asynchronous system does not assume any kind of synchronization, but it is characterized by some limitations, such as a lower efficiency in long range transmissions.

For both systems, the capacity and the delay are analyzed by using the tools developed in the first part of the work described in this thesis. From this analysis, we are able to compare both system, and discuss their advantages and disadvantages. We conclude that the adequated system depends on the characteristics of the network.

Resum

La nanotecnologia permet la miniaturització de components elèctrics i mecànics, i obre la possibilitat de dissenyar dispositius capaços d'operar a la nanoescala. Una aplicació interessant es troba en el camp dels nanosensors, on actualment s'estudien aplicacions per la biomedicina, la indústria i el medi ambient.

Es creu que els nanodispositius seran limitats tant en capacitats com en l'àrea d'actuació, cosa que implica un baix rendiment quan actuen aïllats. Per tal de cobrir grans àrees i ampliar el seu rendiment cal l'esforç col·laboratiu entre nanodispositius, interconnectant-los formant xarxes de comunicació, anomenades nanoxarxes.

Els paradigmes clàssics de comunicació no són vàlids a la nanoescala, per tant, s'han hagut de proposar noves tècniques de comunicació per tal de poder comunicar els diversos nanodispositius. Aquest projecte se centra en les comunicacions basades en bacteris flagel·lats; el nou paradigma es basa en l'ús de bacteris per tal de transportar la informació. El nanodispositiu codifica la informació en una molècula d'ADN, després copia la molècula dins del bacteri i finalment aquest es dirigeix fins al receptor usant un procés biològic anomenant quimiotaxi, on el nanodispositiu receptor descodifica la informació.

El treball presentat en aquest projecte se centra en l'anàlisi de la capacitat i del retard en aquest nou paradigma de comunicació. L'objectiu de la feina feta anteriorment sobre aquest tema va ser definir com funcionaria una xarxa basada en bacteris flagel·lats i en definir els passos bàsics de comunicació, donant menys importància als resultats analítics. Per exemple, no es va trobar cap expressió per definir la propagació dels bacteris, utilitzant simuladors per caracteritzar-ne la propagació.

La primera contribució del present projecte és una anàlisi completa de l'efecte de les mutacions en la fiabilitat d'aquest sistema, tenint en compte que es podem produir múltiples mutacions que afectaran la informació mentre aquesta es va desplaçant a través del sistema.

La segona contribució del projecte és una anàlisi de la propagació dels bacteris. Els bacteris es mouen de forma aleatòria en el medi, però aquest moviment està esbiaixat per tal de poder seguir unes determinades substàncies químiques, fenomen que és aprofitat per dirigir els bacteris cap als receptors. Com a resultat de l'anàlisi, trobem una expressió per al retard, fet que ens permet prescindir del les simulacions per estudiar la propagació de la informació en els sistemes basats en bacteris flagel·lats.

Finalment, s'introdueixen dos sistemes de comunicació totalment diferents, anomenant el primer sistema síncron i el segon sistema asíncron. En el sistema síncron, es considera que els nanodispositius estan sincronitzats entre ells, i que per tant, poden utilitzar una mateixa base de temps entre ells. Aquests sistema pot presentar alguns avantatges, però els sistemes sincronitzats es caracteritzen per algunes limitacions, com ara major complexitat i major

consum. En el sistema asíncron no se suposa cap sincronització entre els nanodispositius, però un sistema d'aquestes característiques presenta alguns desavantatges, com ara una menor eficiència en transmissions de llarg abast.

Per a cadascun dels sistemes s'estudia la capacitat i el retard usant les eines desenvolupades en la primera part del projecte. A partir d'aquesta anàlisi, podem comparar els dos sistemes, examinant-ne els avantatges i desavantatges i concloure que el sistema òptim depèn de les característiques de la xarxa.

Resumen

La nanotecnología permite la miniaturización de componentes eléctricos y mecánicos, y abre la posibilidad de diseñar dispositivos capaces de operar en la nanoescala. La nanotecnología encuentra una aplicación interesante en el campo de los nanosensores, donde actualmente se estudian aplicaciones para la biomedicina, la industria y el medio ambiente.

Se cree que los nanodispositivos serán limitados tanto en capacidad como en área de actuación, lo que implica un bajo rendimiento cuando actúan aislados. Para cubrir grandes áreas y ampliar el rendimiento de los nanodispositivos es necesario el esfuerzo colaborativo entre ellos, interconectándolos formando redes de comunicación, denominadas nanoredes.

Los paradigmas clásicos de comunicación no son válidos en la nanoescala, por lo que se han tenido que proponer nuevas alternativas para poder comunicar los nanodispositivos. Este proyecto se centra en las comunicaciones basadas en bacterias flageladas; el nuevo paradigma se basa en el uso de bacterias para transportar la información. El nanodispositivo codifica la información en una molécula de ADN, después copia la molécula dentro de la bacteria y, finalmente, ésta se dirige hasta el receptor usando un proceso biológico denominado quimiotaxis, donde el nanodispositivo receptor decodifica la información.

El trabajo presentado en este proyecto se centra en el análisis de la capacidad y el retardo en este nuevo paradigma de comunicación. La labor realizada anteriormente en este tema se dedicó a definir el funcionamiento de una red basada en bacterias flageladas y a definir los pasos básicos de comunicación, dando menos importancia a los resultados analíticos. Por ejemplo, no se encontró ninguna expresión para definir la propagación de las bacterias, utilizando simulaciones para caracterizar la propagación.

La primera contribución del presente trabajo es un completo análisis del efecto de las mutaciones en la fiabilidad del sistema estudiado, teniendo en cuenta que múltiples mutaciones pueden afectar la información mientras se desplaza la bacteria a través del sistema.

La segunda contribución consiste en un análisis de la propagación de las bacterias. Las bacterias se mueven de forma aleatoria en el medio, pero este movimiento está sesgado con el objetivo de seguir ciertas substancias químicas, fenómeno que se aprovecha para dirigir las bacterias a los receptores. Como resultado del análisis obtemenos una expresión analítica que describe el retardo de propagación de las bacterias, lo que evita el uso de simulaciones para caracterizar la propagación de las bacterias en sistemas de comunicacion basados en bacterias flageladas.

Finalmente, se introducen dos sistemas completamente diferentes, uno síncrono y el otro asíncrono. El primero considera que todos los nanodispositivos están sincronizados, lo que permite usar la misma base de tiempo en todos los nanodispositivos; este sistema presenta algunas ventajas, pero se caracteriza por una mayor complejidad y mayor consumo energético.

El sistema asíncrono no presupone ningún tipo de sincronización entre nanodispositivos, pero se caracteriza por algunas limitaciones, como un menor rendimiento en transmisiones de largo alcance.

Para los dos sistemas se estudia la capacidad y el retardo usando las herramientas desarrolladas en la primera parte del proyecto. Después de este análisis, se pueden comparar los dos sistemas y discutir sus ventajas e inconvenientes. Se concluye que el sistema óptimo depende de las características de la red.

Contents

List of Figures

List of Tables

Chapter 1

Introduction to Nanonetworks

In the last years, Wireless Sensor Networks (WSNs) have revolutionized the communication world. This new paradigm is based on the collaborative effort among tens or hundreds of small devices in order to monitor physical or environmental conditions, such as pressure, temperature or pollutants [1]. Current technologies do not allow to use WSN in specific environment, such as intra-body applications, where non-invasive sensors are needed. This leads to consider new techniques in the sensing and communication paradigms.

We expect nanotechnology to be the technology that will enable the development of novel devices, small enough to be considered non-invasive. Nanotechnology is providing new materials and new techniques that can be used to develop devices a few hundreds of nanometers in size. The tasks that one of these nanodevices can accomplish are limited in terms of complexity. However, the collaborative effort among these nanodevices will expand the capabilities of a single nanodevice. In this context, it is necessary to develop novel communication techniques to enable the communication among nanodevices [2].

1.1 Nanotechnology, Nanodevices and Applications

In 1959, the Nobel laureate physicist Richard Feynman, in his famous speech "There's Plenty of Room at the Bottom" described for first time, in a bottom-up approach, how the manipulation of individual atoms or molecules would provide more functional man-made devices. But the the term *Nanotechnology* was used for first time in [3], when Norio Taniguchi defined nanotechnology, in 1974, as: *Nanotechnology mainly consists of the processing of separation, consolidation and deformation of materials by one atom or one molecule.*

1.1.1 Nanotechnology

Nowadays, nanotechnology field is very diverse, ranging from the miniaturization of conventional devices to completely new approaches based upon molecular assembled devices; from

studying the properties of atom size materials to developing complete new material with desired properties. Different fields from engineering and science, such as chemistry, physics, electrical engineering, biology, are contributing to nanotechnology. Nanotechnology is also enabling new options in the nanoscale, such the possibility of directly interacting with biological component, that opens a broad range of new applications not available with previous technologies.

One of the most promising applications of nanotechnology is in the field of nanosensing [4, 5, 6, 7]. A nanosensor, usually referred as nanodevices, is a device that can detect events in the nanoscale that can not be detected using other technologies. For example, a nanosensor is able to detect the presence of virus, bacteria or cancerous cells [8, 9] in concentrations up to one part per billion [10, 11]. In order to achieve that, this nanosensors need a nanosensing unit to gather the information, a mechanism to send and receive data and a processing unit to coordinate these operations.

As in a WSN, the range of the sensors is limited, and it is necessary to deploy several sensors in order to cover large areas. In order to obtain information from the nanodevices and control them, it is necessary to establish communication among them.

1.1.2 Nanodevices

We define a nanodevice as an integrated device of around 10-100 μm^2 in size with limited capabilities. The main function of a nanodevice is to process the information collected in the nanosensing unit, and being able to forward this information to other nanodevices. No complete nanodevice has been realized yet but different disciplines involved in the nanotechnology converge to a common concept of nanodevice.

We envision different ways to achieve the design and fabrication of nanodevices, as summarized in Fig. 1.1. The top-down approach consists of reducing the size of current components to the nanoscale. The other approach is based on the design of the nanodevice from nanoscale component. This components can be human-made, such as graphene derivates (Nanomaterial-based approach) or organic, such as proteins, hormones, parts of cells, etc. (Biological approach).

In this work, we focus on the biological alternative. This alternative is characterized by the imitation of biological components in order to copy natural processes. We envision that two options will be possible in that scenario: hybrid nanodevices or completely biological nanodevices. The hybrid approach follows a classical design, where the device is divided into independent nanocomponents, biological or human-made, which are interconnected. The architecture of these nanodevices was introduced in [2] and it is shown in Fig. 1.2. An example of an hybrid device is studied in [12, 13]. The completely biological nanodevices approach consists in modifying existing cells to perform different functions. Cells can be

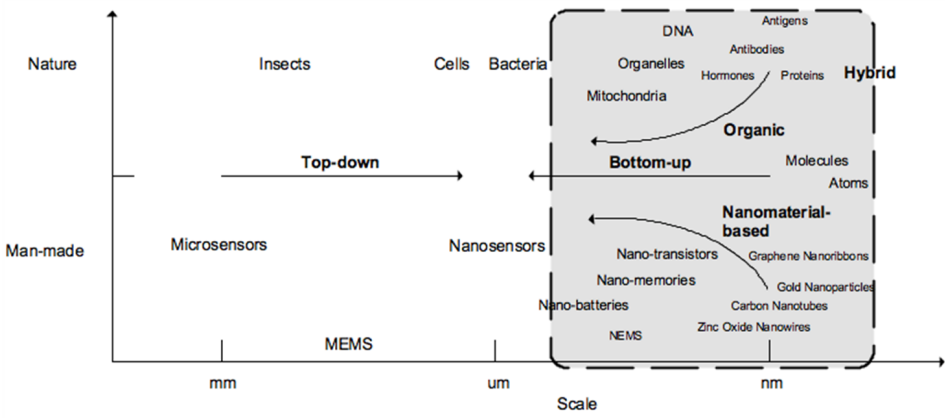


Figure 1.1: Different approaches to design nanodevices

considered nanodevices, since they perform the basic functions described for a nanodevice: energy harvesting, sensing, processing and communications.

A nanodevice has the perform different functions, which are done by different nanocomponents: nano-battery, nano-processor, nano-memory, nano-sensors and nano-transceivers. The nano-battery is the energy source of the nanodevice, the nano-processor controls it, the nano-memory stores all the necessary information, the nano-sensors gather the information and the nano-transceiver allows the communication among nanodevices. A cell has internal membranes and structures which perform these function, such as the nucleus, which is the processor and the memory of the cell.

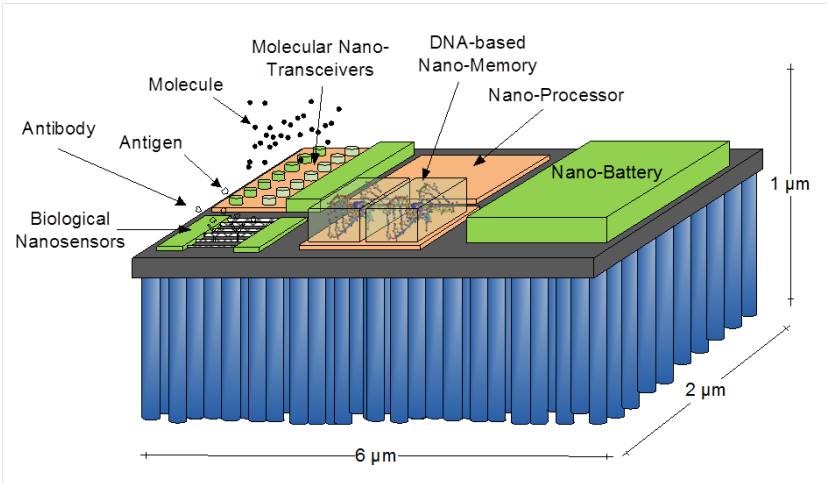


Figure 1.2: Components of a hybrid biological nanodevice

Nano-battery or Nano-power Unit

Cells obtain energy harvesting components from the medium such as glucose, amino acids, fatty acids and oxygen and process them to synthesize Adenosine TriPhosphate (ATP). This molecule is stored and then used to produce the energy needed by the cell. We envision that

biological nanodevices will also use molecule harvesting and chemical reactions to produce the energy they need to work.

Nano-processor

Different approaches can be studied for the nano-processor in the biological nanodevices. DNA computing [14] is the most interesting approach, which consists of taking advantage of the biological processing of the DNA to do computing tasks. In the hybrid approach, classic binary processors can be used.

Nano-memory

DNA can be used not only in a nano-processor, but also in the fabrication of nano-memories, as introduced in [15]. DNA is a molecule that stores already all the information needed by a living being; for example, the human chromosome stores 3 billions of base pairs [16].

Nano-sensors

Nano-sensors are the most important part of the nanodevice, being able to sense some characteristics of the medium. Authors in [17] studied biological and chemical sensors. Other nano-sensors based on nanomaterials, such as carbon nanotubes, are studied in [18, 19].

Nano-transceivers

Nano-transceivers are the part of the nanodevice able to establish communication with other nanodevices. The nano-transceivers depends on the communication paradigm used. In the biological approach it will be able to release and receive biological or chemical components, such as bacteria, pheromones or other molecules.

1.1.3 Applications

The reduced size of the nanosensor enables the use of Wireless NanoSensor Networks (WNSNs) in new environments, giving new applications in which classical sensor networks can not apply. We classify these applications into the following four groups:

Biomedical Applications

Biomedical applications are the most promising field in WNSNs. The nanoscale nature of the cells and organs has made impossible the use of current technologies in intra-body applications. These applications include glucose detection, sodium or cholesterol problems [20, 21] or the detection of infections or cancer tissue [22]. A wireless macro-scale interface will collecte the information from the nanosensors and report it to a health center.

Environmental Applications

Nanosensors can be used to release composites and control plantations of trees, bushes, plants or herbs [23, 24]. Other environment applications are pollution control or biodegradation [5].

Industrial and Consumer Goods Applications

Automatic control of the temperature, quality control procedures, self-cleaning nanomaterial, food and water quality control are some examples of the WNSNs. The use of WNSNs in consumer goods is referred with the term Internet of Nano-Things [25].

Military Applications

WNSNs can be used to detect different attacks, such as nuclear, biological and chemical threats. Other applications are advance camouflages, smart clothes able to self-regulate the temperature of the soldier or even detect if the soldier has been injured.

1.1.4 Types of nanonetworks

In the previous subsections, we defined nanodevices as units able to communicate among themselves. We define a nanonetwork as a network whose nodes are nanodevices and which uses novel communication techniques working at the nanoscale.

Two different paradigms have been proposed [2, 26] in order to establish communication among nanodevices. Molecular nanonetworks were proposed in order to mimic communications among biological beings and cells. After the study of new nanomaterials such as graphene, the electromagnetic nanonetworks were also proposed.

Molecular Nanonetworks

These are based on molecular communication, which is a paradigm based on the transmission and reception of encoded information into molecules [2, 27, 28]. This molecular nanonetworks can be divided on the basis of the distance range. Fig. 1.3 shows this classification into short, medium and long range.

In general, in molecular nanonetworks information propagates by molecule diffusion. Diffusion describes the spread of particles through random motion in order to avoid difference in the concentration. The study of the diffusion process from the communication point of view allows to obtain general formulas for the capacity and the delay that can be applied in most of the molecular cases. In [29], a channel model for this diffusion-based molecular nanonetworks is provided. In [30, 31], the diffusion noise and the reception noise are analyzed.

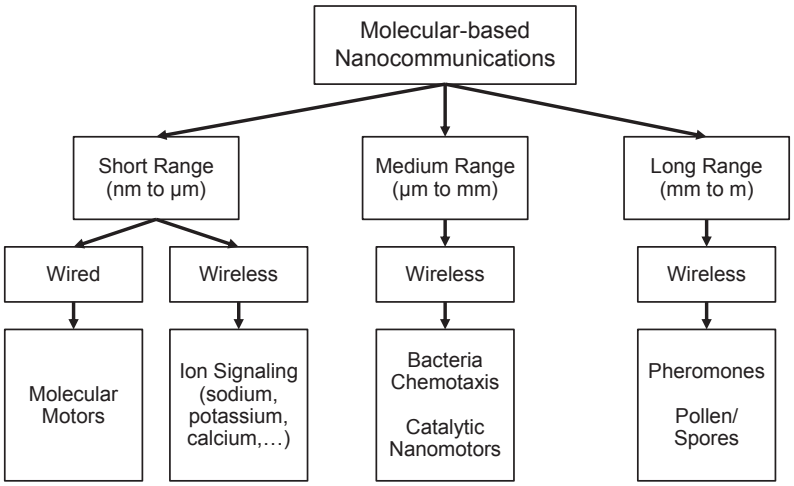


Figure 1.3: Classification of different molecular communication techniques according to the distance.

Different types of molecular nanonetworks have been proposed. The short range methods try to mimic the biological communication among cells. Molecular motors are proteins or proteins complexes that transform chemical energy into mechanical work at a molecular scale and are able to move and transport information molecules. Molecular motors move along molecular rails called microtubes, and they can be considered as wired connecting a network. A wireless alternative for the short range nanonetworks is ion signaling, this is a mechanism used by cells in order to communicate to the neighbors cells.

In the medium range, two options have been proposed: bacteria based communications and catalytic motors [32]. Bacteria based communications are studied in this thesis, and introduced in the next section. Catalytic nanomotors are synthetic particles that are able to propel themselves and can be used to carry DNA plasmids, which encode the information from transmitter to a receiver [33]. The main problem of the catalytic motors is that magnetic fields are needed in order to lead the particle to the receiver.

Finally, in the long range, two options have been proposed: pheromones and pollen or spores. Both options try to mimic natural long range communications in the biology. In [27] different long range nanonetworks are studied.

Electromagnetic Nanonetworks

These are based on the exchange of information through electromagnetic radiation, which is realized through novel nanomaterials and their properties. This paradigm is much closer to the classic communications, but taking into account all limitations and new characteristics of the nanoscale.

Graphene is a novel nanomaterial that will enable electromagnetic communications in the nanoscale. It has already been demonstrated [34] that a graphene antenna can be used in the nanoscale to radiate electromagnetic waves in the terahertz band. This band is located between the microwaves and the far infrared (300 GHz - 3 THz). For communication purposes, the terahertz band is one of the least explored spectral range, since it presents important withdraws in the macro-scale. For this, in [35], a novel propagation model was introduced, which describes the attenuation and noise that a signal suffers when it propagates through the medium in the presence of water molecules.

1.2 Bacteria-Based Communication in Nanonetworks

In this section, the state of the art of the flagellated bacteria communication is presented. It was first introduced in [32], and later developed in [36, 37]. In this previous work, the idea to use flagellated bacteria to communicate different nanodevices was presented, and the basic concepts of the communication were analyzed.

Bacteria are microorganisms composed only by one prokaryotic cell and they live in a liquid medium. They are able to randomly move using the flagellum. But in the presence of specific substances, called attractant, bacteria can direct their movements according to the gradient of the substance in the environment: this process is called chemotaxis. The propagation of the bacteria is deeply analyzed in Chapter 3: the idea is to take advantage of this phenomenon to be able to send the bacteria to a specific point.

Bacteria are used as carrier of information, which is encoded in a double-stranded DNA molecule, called plasmid, and transported by them. The use of DNA allows a higher throughput than other molecular approaches, due to the high information density of the DNA.

In the following section, we introduce the architecture of the network, each component and the basic communication steps.

1.2.1 Nanonetwork Components

A Bacteria-Based Nanonetwork (BBN) is composed of nodes and carriers, as shown in Fig. 1.4. Nodes are nanodevices, described in the last section, but with some specific characteristics needed in a system based on flagellated bacteria. Carriers are bacteria, which physically transport the information from a transmitter to a receiver. Nodes and carriers are floating within a liquid medium. All these components and their networking are described in this subsection.

Environment

All nodes and carriers are floating within a liquid medium provided with nutrients for the bacteria and energy sources for the nanodevices. We consider a 3D medium in general, but a

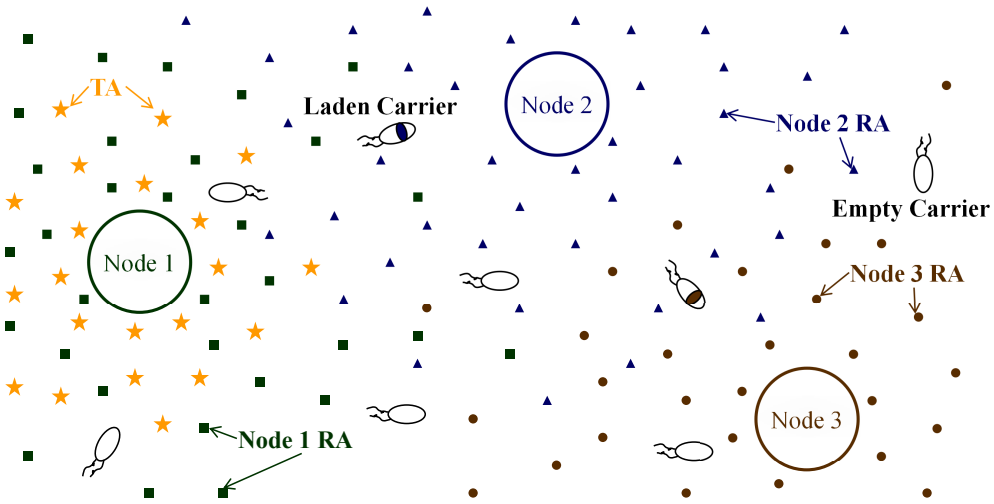


Figure 1.4: A Bacteria-Based Nanonetwork (BBN) with 3 nodes, 2 messages in transit and "empty" carriers. Laden carriers follow specific RA, represented by geometric figures and colours, that are emitted by each node.

planar (2D) or linear (1D) medium can also be used. Attractants released by nanodevices in order to attract bacteria are also present in the environment, which allow bacteria to direct their movement.

In a real scenario, this system can be deployed inside the body: gastrointestinal tract, blood vessel... or outside the body: in a glass, pool... The only condition is that the medium has to be compatible with bacteria.

Carriers: Bacteria

Bacteria are one of the oldest organisms on the Earth: the first bacteria appeared about 4 billions years ago. Since then, they evolved developing skills that allow them to move and convert chemicals into energy. Cilia and flagella are some of the mechanisms that allow the bacteria to move. Flagella are semi-rigid cylindrical structures that are rotated and function much like the propeller on a ship. Chemotaxis is the phenomenon in which bacteria direct their movements according to the concentration of specific substances (attractants) in the environment.

Among all possible bacteria that have the ability to move we focus on Escherichia Coli (E. Coli). This bacterium, shown in Fig. 1.5, is one of the most studied, and its complete genome sequence is known [38]. E. Coli is approximately 2 μm long and has a diameter of 1 μm and it has between 4 and 10 flagella located at the cell membrane that allow it to move. E. Coli has a huge number of chemical receptors surrounding its membrane that allow them to sense the environment for the presence of attractant particles and move towards them.

Bacteria will be distributed in the BBN forming a uniformly distributed population, spanning the whole medium. Bacteria are used by nodes when they need to send information.

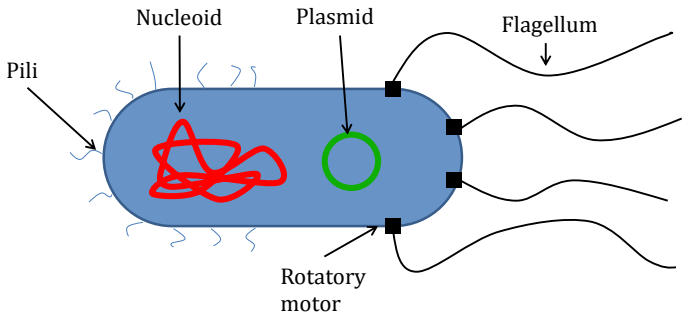


Figure 1.5: Escherichia Coli (E. Coli)

Bacteria are responsible for carrying the DNA molecule which encodes the information from the transmitter to the receiver by taking advantage of the chemotaxis phenomenon. The propagation of the bacteria is described and studied in Chapter 3. The population density can be maintained by regulating the replication cycle [39] using Quorum Sensing (QS). QS is a system of stimulus and response, used by many species of bacteria, such as E. Coli, than can be used to control their behavior.

E. Coli can sense at least 12 different attractants using different chemoreceptors that cause independent chemotaxis response [40]. Furthermore, there are E. Coli mutants that present no response towards some specific attractants [41]. This can be used to direct bacteria to a specific receptor which is releasing a specific attractant. So, it is needed to modify the behavior of the bacteria in order to be able to direct its movement to the right destination.

We need to assume that will be possible to modify bacteria to use already existing biological parts and that these parts can be combined as desired to obtain specific behaviors. All the biological parts are controlled by genes that implement certain functionalities in bacteria, such replication or chemotaxis. These modifications have been achieved by different scientific teams: [42, 43].

Nodes: Nanodevices

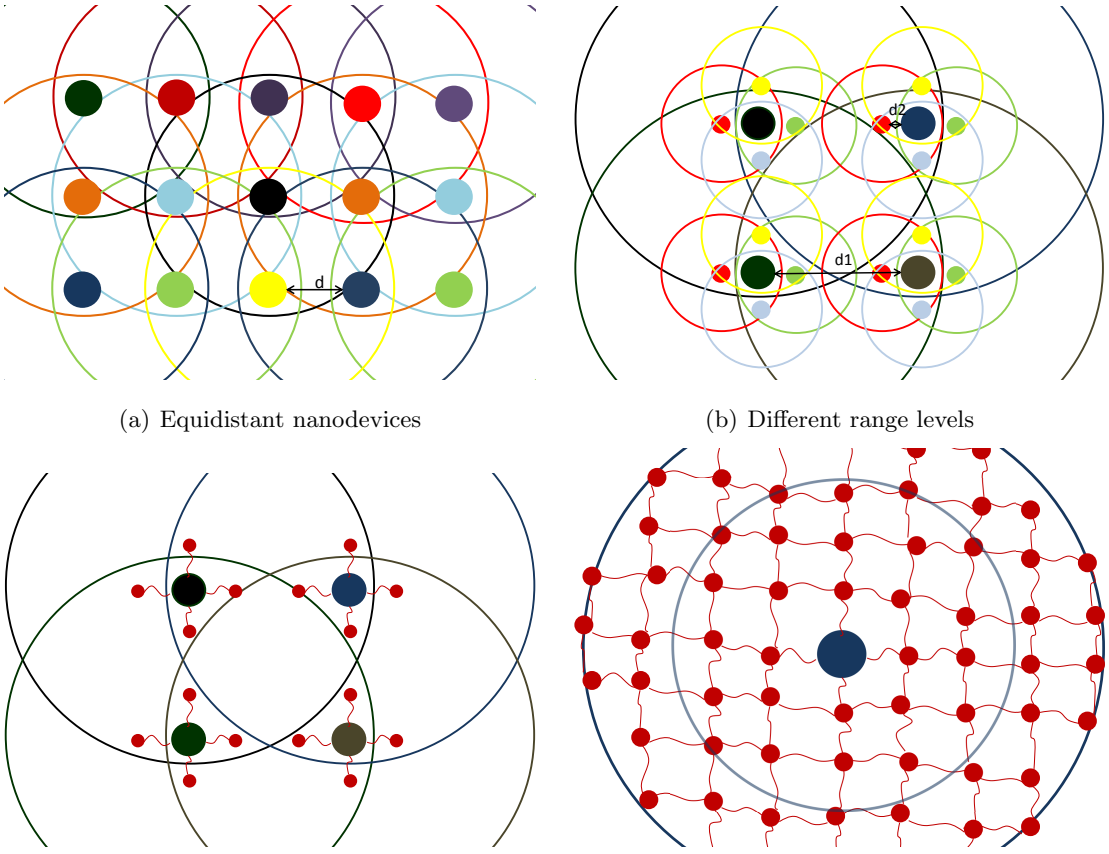
The nodes of a flagellated bacteria based nanonetwork are nanodevices, such as those described in section 1.1.2, but with some specific characteristics. First of all, a nanodevice will be able to work with DNA, which means that each nanodevice contains a DNA Processing Unit (DPU), that is able to encode and decode arbitrary DNA. The DNA computing is a new paradigm of computing, [14], and it has been introduced by different authors. Examples include the DNA-based Turing Machines [44, 45].

Nanodevices used in this type of nanonetwork are also required to be an active part in in the reception of information. Bacteria direct their movement according to the attractant released by the nanodevice, so each nanodevice has to be continuously releasing attractant in order to

be able to receive information at any time (Receiving Attractant or RA). Different attractants should be used in the nanonetwork in order to avoid interferences among nanodevices. When a node has to send information, "empty" bacteria are required in order to send information. To get the "empty" bacteria, an attractant which all "empty" bacteria are sensitive to is released by the node in order to attract some of them (Transmitting Attractant or TA).

1.2.2 Nanonetwork Architecture

As described in the previous subsection, the network is composed by nanodevices floating in a liquid medium and bacteria that are used to carry information. Nanodevices release different attractants in order to attract bacteria that are sensitive to these attractants, but the number of attracts available, which is around 15 [40], and the limited range of the attractant [37] make necessary to use multiple hops to reach all the nanodevices and to use a network address to identify the nanodevice in the network. This address is a unique identifier in the network used to identify the final destination. Attractant is used to identify the following nanodevice in the path.



(a) Equidistant nanodevices

(b) Different range levels

(c) Different paradigms per level

(d) Unidirectional communication

Figure 1.6: Different nanonetwork architectures

By using these addresses we propose a basic nanonetwork architecture where all the nanodevices are arranged in a perfect grid, equally separated, shown in Fig. 1.6(a). This can be

realized in a 2D scenario, such as any surface covered by a liquid, or in a 3D scenario, such as any volume filled with a liquid. The addressing in these architecture is an important and open research problem which needs to be further investigated and where concrete solutions are needed. Other architectures are also possible. One example is having different layers in the network, in which there are different kinds of nodes, some are only able to communicate with the immediate neighbors and others that can reach further nodes, Fig. 1.6(b). Other examples can be proposed combining this communication paradigm with other paradigms. For a high density distribution of nodes, a short range communication can be used to communicate immediate neighbors and the flagellated bacteria can be used to communicate nanodevices placed further apart, Fig. 1.6(c). Another option is to use flagellated bacteria communication as a unidirectional method, in order to obtain information from nodes, and use another paradigm to send information to nodes, Fig. 1.6(d).

1.2.3 Basic Communication Steps

In this section, the needed steps as well as the biological phenomena involved in the communication are described, which were first introduced in [37]. The communication can be divided into six steps: encoding, encapsulation, propagation, decapsulation, relaying and decoding. A scheme of all the process is shown in Fig. 1.7.

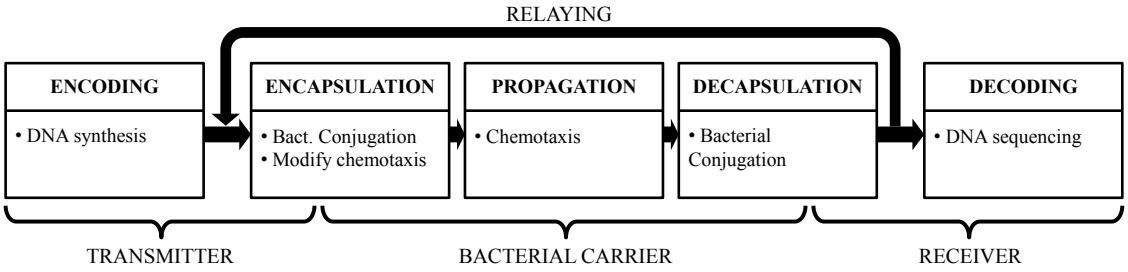
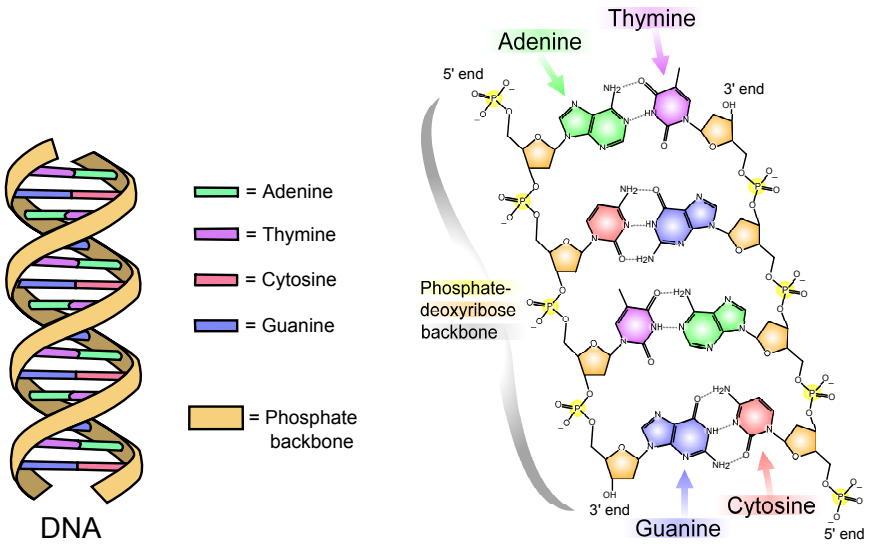


Figure 1.7: Overview of the basic communication steps in a Bacteria-Based Nanonetwork (BBN)

Encoding

The first step consists of encoding the information to be transmitted in a double-stranded DNA molecule. This is done by the DNA Processing Unit (DPU) of the nanodevice.

DeoxyriboNucleic Acid (DNA) is a long molecule composed of two polymers of nucleoides, with each nucleoide from one polymer bonded to one in the other polymer and forming a base pair (bp). Polymers are represented in yellow in Fig. 1.8, the nucleoides and bps are in each bound of the polymers. Each polymer extends in opposite direction, with the asymmetric ends of each polymer called 5' and 3'. Each nucleoide contains one of four possible bases: adenine (A), cytosine (C), guanine (G) or thymine (T). The base in a nucleoide determine the base in the other nucleoide of the pair: adenine is always with thymine and cytosine is with



(a) Scheme of a DNA molecule (b) Atomic composition of a DNA molecule

Figure 1.8: Structure of a DNA molecule

guanine. So the possible structures for a base are AT, TA, CG and GC. Only one nucleoide of the base it is needed to describe a DNA sequence, for example 5'ACTG3' is the sequence of the Fig. 1.8(b).

The DNA chain used to encode information in our BBN forms a plasmid. A plasmid is a circular DNA molecule separated from the chromosomal DNA of the bacteria, that is capable of self-replication and self-transfer. A plasmid can be up to 1.6 mbp (mega bp) [46] long.

The DNA encoded in the plasmid is divided into three parts: the transfer region, the routing region and the message region. The transfer and routing regions constitute the active section of the plasmid, which are expressed or interpreted as biological instructions. The message region contains the arbitrary information.

1. **The transfer region** is present in typical plasmids as the F factor of E. Coli and it is 33 kbp (kilo bp) [47]. This region contains the genes necessary for self-replication and transmission of the plasmid.

2. **The routing region** contains the genes that modify the behavior of the "empty" bacteria. These genes encodes new proteins or inhibit genes in the chromosomal DNA [48] for various purposes:

   - *Deactivation of the chomotaxis towards transmitters*, using a specific protein that disables chemotaxis (disable chemoreceptors) towards the specific attractant (TA) used to attract "empty" bacteria.
   - *Activation of the chemotaxis toward the receiver*, using a specific protein that enables chemotaxis (enable chemoreceptors) towards a specific attractant (RA). This protein identifies the physical address of the message.

- *Inhibition of bacterial replication [39]* in order to avoid an exponential growing of the number of bacteria with the same message, since this would overload the receiver.
- *Enabling of programmed death [49] on timeout.* This is used to prevent delivery of messages that have accumulated a very long delay by making bacteria suicide at a specific time after plasmid reception using a acTTLacro gene.

The size of the routing region is expected to be in order of the size of the transfer region, i.e., tens of kpb.

3. **The message region** contains the base pairs that are not interpreted as biological instructions, such as the destination network address and the message body, which occupies most of the length of the plasmid.

Encoding an arbitrary DNA sequence in the plasmid has a problem: once in the bacterium, the arbitrary sequence would be expressed and could lead to the synthesis of proteins that could interfere with the message delivery. To avoid this, it is necessary to disable the expression of the message by encoding it as inactive DNA, and this inevitably reduces the number of possible sequences that can be encoded.

One way to inhibit the DNA expression is to avoid promoter sequences. Promoters are specific DNA sequences where the Ribonucleic acid (RNA) polymerase can bind to start the transcription. Promoters are varied in the DNA of an organism, but they share subsequences called consensus promoters. Avoiding these sequences is one possibility to prevent DNA expression.

The time needed to synthesize the DNA depends on the speed of the DPU. To date, no practical implementation of DPU has been realized, but we envision that it will be in the order of the bacteria DNA replication speed, $r_{DNA} = 1400bp/s$, [50]. That means that the time needed to encode a plasmid of 1.6 Mbps will be approximately 19 min.

Encapsulation

In this step, the plasmid generated in the previous step, which contains the message, is transferred to the bacterium. First, the transmitter emits TA in order to attract "empty" bacteria carriers in the vicinity. This can happen simultaneously with the encoding step, so it will not affect the delay. When a bacterium reaches the transmitter, a copy of the plasmid is transferred to the bacterium using BC. BC is a biological phenomenon which involves the transfer of genetic material between two bacterial cells: a donor, who first has the plasmid, and the acceptor, who receives a copy of it.

This process can be mimicked to transfer the plasmid from a nanodevice. The transmitter first attaches to the bacterium using a pilis, which will retract to get the bacterium in physical

contact with the transmitter. The membranes of the transmitter and the carrier connect, and a single-stranded DNA (ssDNA) molecule is unwounded from the plasmid and transferred to the carrier. After the ssDNA has been transferred, transmitter and carrier separate and the ssDNA in the carrier replicates into a full plasmid. Note that in nature, the plasmid is also reconstructed in the donor bacterium (whose role is played by the transmitter in our case). A scheme of the BC is shown in Fig. 1.9

Once the plasmid is transferred to the bacterium, the active section of the plasmid is expressed and the bacterium modifies its behavior according to the routing region, as detailed in the last point, and the propagation starts.

The time needed to encapsulate the DNA is the sum of the time needed for the BC and the time needed to activate the chemotaxis towards the receiver. The time needed for the BC depends on the size of the plasmid, but it is copied at a speed around 833 bp/s [51]. That means that the time needed to encode a plasmid of 1.6 Mbps will we approximately 32 min. To activate the chemotaxis, the active part of the plasmid (around 50 kbps) must be activated,, with a bacterial transcription rate of 2500 nucleoides per minute [52], which corresponds to 20 min.

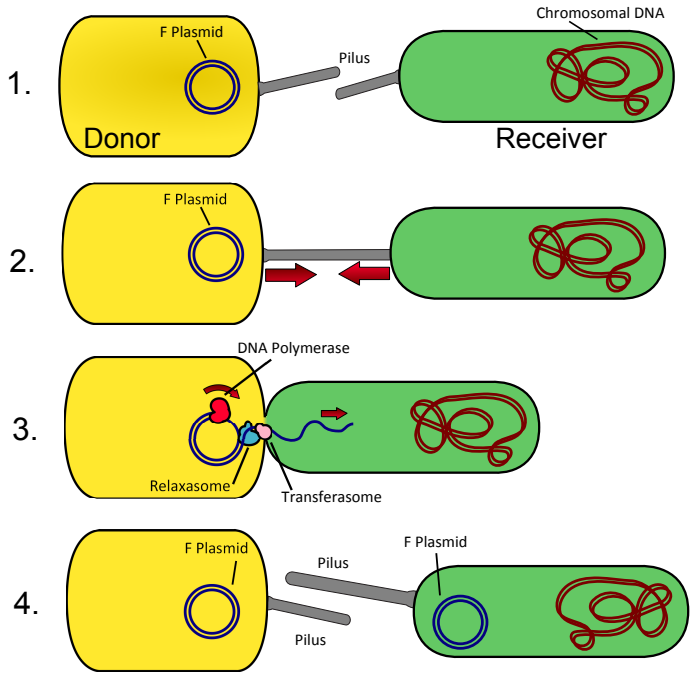


Figure 1.9: Schematic drawing of Bacterial Conjugation (BC)

Propagation

In the propagation step, the carrier swims from the transmitter to the receiver by using bacterial chemotaxis. The delay is limited by the programmed death in the active part of the plasmid, described in the encoding step by the Time To Live (TTL). If a message is

not received after that time, it is considered lost, and the bacteria which carry them will die (suicide).

For a reliable system, we must consider that after a laden bacterium dies, its membrane breaks and all its components are released to the medium. The plasmid floating in the medium could be incorporated by another bacterium through a natural process called transformation, and the transformed bacterium could complete its delivery. The ability to uptake naked DNA is called competence, and to avoid this problem, non-competent bacteria have to be used [53].

There is no reason for bacteria to die in a controlled medium. However, it is possible that some bacteria die during the propagation, and this has to be considered part of the message loss probability.

The time needed for a bacteria to swim to the receiver is random, in the order of minutes or hours depending on the distance. The propagation of the bacteria is widely analyzed in Chapter 3.

Decapsulation

This step uses BC, as explained in the encapsulation step, but in this case, the carrier is the donor and the nanodevice is the acceptor. After receiving the message, the receiver kills the bacterium, which would try to re-deliver the message otherwise.

A laden carrier might encounter another carrier during the propagation and pass a copy if the plasmid to it, resulting in undesired message duplication. We can use surface exclusion [54], which is used by naturally-occurring plasmids to ensure that they are not transferred to bacteria that already have a copy if the plasmid. This mechanism would be adapted so that BC can only happen between a node and a carrier.

In a network based on flagellated bacteria, each one is emitting a different attractant, and bacteria may reach, by change, the wrong node. The receiver determines if the message is for itself by checking the physical and network address in the message. If the physical address does not match, the message is discarded.

The time needed to decapsulate the information is given by the BC speed as well, described in the encapsulation step.

Relaying

Once the message is decapsulated, the receiver checks the destination network address in the message. If it matches with the local network address, the message is decoded. If it does not match, the routing region of the plasmid is updated by substituting the RA in the plasmid by the RA in the local routing table that matches the destination address of the message, and the plasmid is encapsulated again.

The time needed to to relay a bacteria is the time needed to modify the plasmid. One option to do that is by using restrictions endonucleases [55], a natural phenomenon able to modify foreign DNA, such as plasmids. This phenomenon is effective in order of 5 minutes [56], and we estimate a total of 10 minutes as a typical relaying time.

Decoding

In this step, the receiver uses its DPU to decode the message by the sequencing of the plasmid. Sequencing is the process of determining the primary structure, the sequence of nucleoides, of a DNA strand. Then, the message can be processed further at the receiver. This step concludes the communication process.

The time needed to sequence the DNA depends on the speed of the DPU, and we envision it will be similar to the time needed to synthesize the plasmid.

1.3 Motivation and Contributions

The communication alternatives for nanodevices are still very limited. Flagellated bacteria communication is a novel biological approach to enable the communication among nanodevices. Bacteria-based communications are first introduced in [32] among other communication paradigm useful in a medium range scenario. In [36], a deeper analysis is done by introducing the biological concepts that will make possible the use of bacteria to communicate information in the future. Finally in [37], the basic communication steps are analyzed, the delay of the system is obtained using simulations and some values for the capacity are found in a simple system by using the values previously obtained from the simulations.

Lots of work need to be done in this topic in order to make this communication paradigm a reality. In this context, we propose different study directions. We perform a deep study of the propagation of the bacteria in order to get an analytical formula for the delay and to avoid the use of simulations. We also perform a complete analysis of the mutations that can modify the information during the propagation of the information. Then we introduce and define two different systems, namely, the synchronous and the asynchronous, in order to study the capacity and the delay in different situations. Finally we compare both systems and concluded the work.

1.4 Structure

The rest of this work is organized as follows. In Chapter 2 we study the effect of the mutations in the reliability of the information. In Chapter 3, we present a mathematical model of the propagation of the bacteria. In Chapter 4 and in Chapter 5, two different communication systems are introduced, one synchronous and the other asynchronous, together with an analysis

of the capacity and the delay. In Chapter 6, we compare both systems and we point out the advantages and disadvantages of each system. Finally, in Chapter 7, we conclude our work.

Chapter 2

Effect of Mutations in a Bacteria-Based Communication System

2.1 Introduction

In the communication paradigm presented in the introduction, the information is carried and propagated by flagellated bacteria. Bacteria carry the information encoded in DeoxyriboNucleic Acid (DNA) plasmids, which are double-stranded DNA molecules, as explained in Sec. 1.2.3. These DNA molecules are formed by base pairs (bps), which at the same time are formed by two different nucleotides, one determined by the other. Each nucleotide can be one of four possible bases (amino acids) and they are the basic information units of the DNA.

In this chapter, we analyze how the errors or mutations affect the capacity of the system based on the flagellated bacteria communication paradigm . During the communication process, the information bearing DNA molecule is generated (also called synthesized or encoded), copied multiple times when the information needs to be relayed in order to reach the destination and finally read (also called sequenced or decoded). In each of these steps, errors can modify the information encoded, decreasing the capacity of the DNA.

Fig. 2.1 shows the model used to analyze the errors that can occur during the communication. The model is divided into three blocks: the DNA encoding or synthesis, the propagation process, which includes the multiple copies done in the relaying process, and the DNA decoding or sequencing. The errors in the encoding step are characterized by the probability of their occurrence $p_{e_{ENC}}$, the errors occurring during the multiple copies are characterized by the probability $p_{e_{PROP}}$ and the errors occurring during the decoding step are characterized by the probability $p_{e_{DEC}}$.

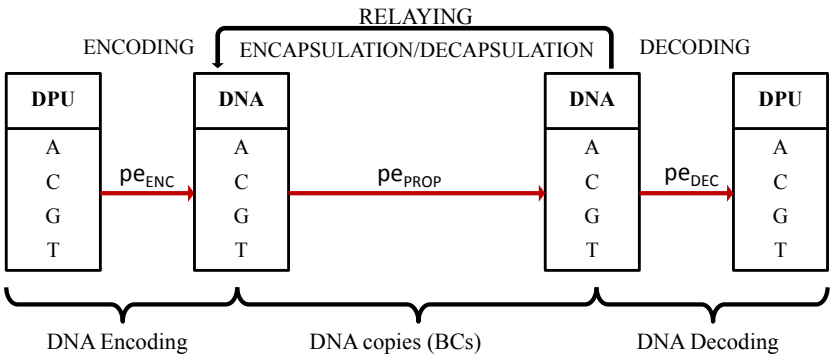


Figure 2.1: Model of the errors that can occur during the communication

The rest of the chapter is organized as follows. In Section 2.2, the errors that modify the information during the multiples copies of the plasmid are studied and an expression of the error probability is given as a function of the number of hops and it is validated using simulations. In Section 2.3, the same analysis is expanded to the complete system by incorporating the encoding and decoding processes. In Section 2.4, the information capacity the flagellated-bacteria based communication paradigm is studied. Finally, in Section 2.5, we conclude this chapter by commenting the obtained results..

2.2 Analysis of the Error Probability During the Multiple Encapsulation and Decapsulation Processes

In general, a communication between two nanodevice is multi-hop. During the communication, the information encoded in the bacterium visits multiple nanodevices that relays the information. In all of these relaying processes, the plasmid is encapsulated and decapsulated many times using Bacterial Conjugation (BC), as described in section 1.2.2. For example, in a one hop scenario, the plasmid is encapsulated in the transmitter, decapsulated in the nanodevice used to relay the information, encapsulated again after updating the information on the next destination, and finally, in the receiver, the plasmid is decapsulated again. This means a total of four times. In each process, mutations can occur and change the base in any bp of the plasmid. Successive mutations in the same base can change the error or even correct it.

The objective of this section is to calculate the probability of error when the information reaches the final destination, in oder words, after multiple processes of encoding and decoding. We start by calculating the probability of bit error and after that the probability of error of a base pair.

A mutation is an error in a bp during the BC process which changes a nucleotide for one of the other three possible nucleotides. We assume that the mutation probability (usually known as mutation rate) is known, independent and constant for all base pairs. In fact, for

the natural BC process among living cells, this mutation probability has been studied in [57], and it has been determined that the mutation rate in that case is 10^{-8}.

2.2.1 Bit Error Probability

In this analysis, it has to be taken into account that each base pair can carry 2 bits of information, since it can contain four possible bases; so a mutation in a base pair can modify one or two bits. This can be observed on the left part of Fig. 2.2. It also has to be taken into account that a successive error in the same bp can correct or change the number of bits modified, as shown on the right part of Fig. 2.2.

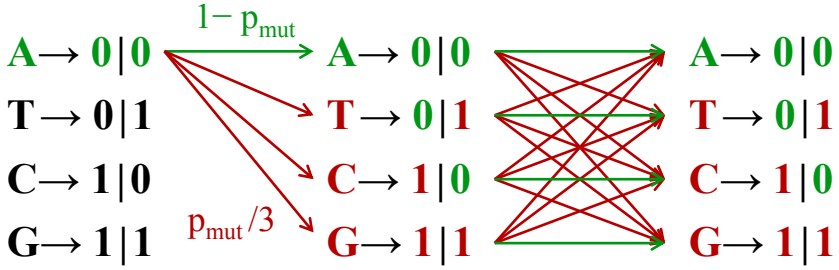


Figure 2.2: Schematic drawing of the errors that can occur during two BC processes

The probability of error if a bit of one BC process is obtained by taking into account how many bits are changed when a mutation occurs. A mutation changes the base for any of the other three possibles bases: two of them having one bit different and the third having both bits different. This means that, on average, two of three bits are modified when a mutation occurs, so the bit error probability is given by the following expression.

$$p_e(1BC) = \frac{2}{3}p_{mut} \tag{2.1}$$

where p_{mut} is the mutation rate.

The probability of bit error after two BC processes, which corresponds to one hop, has to take into account that a second error in the same base can change the number of wrong bits. This is given by (2.2), where the second term is negative since it takes into account that the second error can correct some of the bits modified by the first error. because it is more likely than

$$p_e(2BC) = \frac{4}{3}p_{mut} - \frac{8}{9}p_{mut}^2 \tag{2.2}$$

Theoretical Analysis

The objective is to obtain a general formula as a function of the number of hops by taking into account that each hop implies two BC processes or copies. In this analysis we take into

account that a mutation can occur at any time and different mutations can occur at the same base.

The analysis is done performed by following a recursive fashion. It consists on the definition of a base case and a set of relations which reduce all other cases toward the base case, in our case by increasing the number of hops. In the C^{th} process, if no mutation occurs, the error probability is $p_{ok} \cdot p_e(C-1)$, where p_{ok} is the probability of not having a mutation, equal to $1 - p_{mut}$. In case a mutation occurs, if no bit is modified, the probability of error is $\frac{1}{3}p_{mut} \cdot p_e(C-1)$, since the probability that a mutation does not modify a bit is $\frac{1}{3}$. In case a mutation occurs and the bit is modified, the probability of error is $\frac{2}{3}p_{mut} \cdot (1 - p_e(N))$. The base case is: the probability of error is zero if no copies are done. This can be mathematically expressed as follows:

$$p_e(C) = \begin{cases} 0 & \text{if } C = 0 \\ p_{ok} \cdot p_e(C-1) + \frac{1}{3}p_{mut} \cdot p_e(C-1) + \frac{2}{3}p_{mut} \cdot (1 - p_e(C-1)) & \text{if } C > 0 \end{cases} \tag{2.3}$$

By simplifying the last expression, the following recursive equation is obtained:

$$p_e(C) = \begin{cases} 0 & \text{if } C = 0 \\ \left(1 - \frac{4}{3}p_{mut}\right) \cdot p_e(C-1) + \frac{2}{3}p_{mut} & \text{if } C > 0 \end{cases} \tag{2.4}$$

which, if solved, allows us to obtain:

$$p_e(C) = \frac{2}{3}p_{mut} \sum_{i=0}^{C-1} \left(1 - \frac{4}{3}p_{mut}\right)^i \tag{2.5}$$

The probability of bit error is given by the following final expression:

$$p_e(C) = \frac{1}{2}\left(1 - \left(1 - \frac{4}{3}p_{mut}\right)^C\right) \tag{2.6}$$

where C is the number times the plasmid is copied (number of BC processes), or, in other words, $N = \frac{C}{2}$ is the number of hops done in total; and p_{mut} is the mutation rate.

Simulations

The objective is to validate the probability of bit error obtained in (2.6). The simulations done consist of the repetition of the mutation effect for copiy or step 10000 times. In each step, it is decided randomly if an error occurs and in that case, using the given mutation rate, and the DNA base pair is changed if it is determined that a mutation has occured. After all the steps, the number of wrong bits in the base is determined and finally averaged among all the repetition.

The results of the simulation are shown in Fig. 2.3, where it can be observed that the simulated data (colored lines) matches with the theoretical values obtained using the probability of error expression (2.6) (dotted lines).

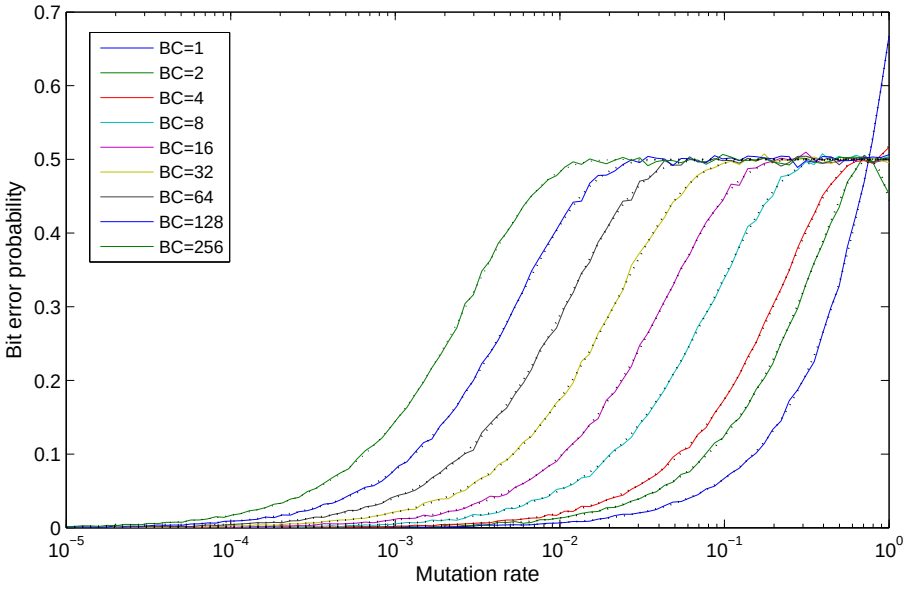


Figure 2.3: Comparison of the simulated bit error probability (colored lines) and the theoretical bit error probability (dotted lines) as a function of the mutation rate for a different number of BC processes

2.2.2 Base Error Probability

In this subsection, by starting from the probability of error in bits analyzed in the last subsection, we obtain the probability of error for the base pairs.

In this case, the system has to be analyzed as a quaternary system, where each error changes the base, considered the symbol, and future errors can correct the error or change the symbol to an other wrong symbol. In Fig. 2.2, the possible errors during two steps of propagation are shown. which can occur changes a symbol into one of the other four symbols. In case a symbol is affected by an error twice, the second error can change the wrong symbol either back to the first correct one or to any other symbol.

After one copy of the plasmid, the probability of error is directly the mutation rate. After two steps, it has to be taken into account that the second error can correct the first error, so the error probability is:

$$p_e(2BC) = p_{ok}p_{mut} + p_{mut}\left(p_{mut} + \frac{2}{3}p_{mut}\right) = 2p_{mut} - \frac{4}{3}p_{mut}^2 \tag{2.7}$$

where p_{mut} is the mutation rate and p_{ok} is the probability of not having a mutation, which is expressed as $p_{ok} = 1 - p_{mut}$.

Theoretical Analysis

The objective is to determine the quaternary error probability by following a similar approach as for the binary error probability. If no mutation occurs during the C^{th} copy, then the probability of error is $p_{ok} \cdot p_e(C-1)$. On the contrary, if there is an mutation in that step, the probability of error is the mutation rate multiplied by the probability for successive errors to not correct the error. Since $\frac{2}{3}$ of the mutations will not correct the error, the probability of error when a mutation occurs is $p_{mut}\left(1 - p_e(C-1) + \frac{2}{3}p_e(C-1)\right)$. The base case is the same than the binary case: the probability of error is zero if no copies are done. This can be mathematically expressed as follows:

$$p_e(C) = \begin{cases} 0 & \text{if } C = 0 \\ p_{ok} \cdot p_e(C-1) + p_{mut}\left(1 - p_e(C-1) + \frac{2}{3}p_e(C-1)\right) & \text{if } C > 0 \end{cases} \tag{2.8}$$

where p_{ok} is the probability of not having a mutation, which is equal to $1 - p_{mut}$. By simplification of (2.8), the following recursive equation is obtained:

$$p_e(C) = \begin{cases} 0 & \text{if } C = 0 \\ \left(1 - \frac{4}{3}p_{mut}\right) \cdot p_e(C-1) + p_{mut} & \text{if } C > 0 \end{cases} \tag{2.9}$$

By solving this recursive equation, we obtain a geometric expression, similar that in the binary case. After further simplification, the probability of error of a base pair is given by the following expression:

$$p_e(C) = \frac{3}{4}\left(1 - \left(1 - \frac{4}{3}p_{mut}\right)^C\right) \tag{2.10}$$

where C is the number of copies and p_{mut} is the mutation rate.

Expressions (2.6) and (2.10) are identical with the exception of a constant. This can be explained since both probabilities take into account the same mutations, but interpreting the errors in a different way: an error on a base changes the quaternary symbol but it only changes $\frac{2}{3}$ bits that are carried by the base pair. As a consequence, the relationship with both probabilities is:

$$p_{e_{BIN}}(C) = \frac{2}{3}p_{e_{BASE}}(C) \tag{2.11}$$

This relationship can be used in all cases to obtain one probability if the other one is know. From now on, only the base error probability will be calculated.

2.3 Analysis of the Error Probability in the Complete Communication Scheme

In this section, the probability of error of the complete system is studied, which means that the errors that can occur when DNA is generated (encoded) and sequenced (decoded) are introduced in the analysis. The complete model of the system was introduced in Fig. 2.1, and the probability of error during the propagation was studied in the previous section.

We envision that it will be possible to generate an arbitrary sequence of DNA in the nanoscale, which has already been done in macroscale [58, 59]. We foresee that if an error can occur in a biological process such as the BC, also an error can occur during the manipulation of the DNA, and we assume that this probability of error will be characterized.

In the current scenario, we assume that errors can occur when the DNA is encoded with a probability $p_{e_{ENC}}$ and more errors can occur when the DNA is decoded with a probability $p_{e_{DEC}}$. So three sources of errors are considered: encoding, multiple copies and decoding. By taking into account all the three possible sources of errors and all the combination among them, the following expression is obtained:

$$\begin{aligned} p_e(N) = p_{e_{ENC}} + p_{e_{PROP}}(N) + p_{e_{DEC}} - \frac{4}{3}p_{e_{ENC}}p_{e_{PROP}}(2N) - \frac{4}{3}p_{e_{ENC}}p_{e_{DEC}} \\ -\frac{4}{3}p_{e_{PROP}}(2N)p_{e_{DEC}} + \frac{16}{3}p_{e_{ENC}}p_{e_{PROP}}(2N)p_{e_{DEC}} \end{aligned} \tag{2.12}$$

which takes into account that only one error can occur during the synthesis, the propagation or the sequencing, two errors can occur in any pair of processes, and three errors affect all the processes.

In order to obtain a simple expression, we can assume that $p_{e_{ENC}} = p_{mut} = p_{e_{DEC}}$, which is reasonable because each process is a manipulation of the plasmid. In that case, the probability of error of a base pair in the complete system is given by the following expression:

$$p_e(N) = \frac{3}{4}\left(1 - \left(1 - \frac{4}{3}p_{mut}\right)^{2N+2}\right) \tag{2.13}$$

where N is the number of hops.

2.3.1 Numerical Results

As we stated above, the mutation rate during a BC in the nature is 10^{-8}. By assuming that this mutation rate will not change if the process is done in a nanodevice and the probability of error during the synthesis and sequencing will be equal to the mutation rate (2.13), we obtain the results shown in Fig. 2.4.

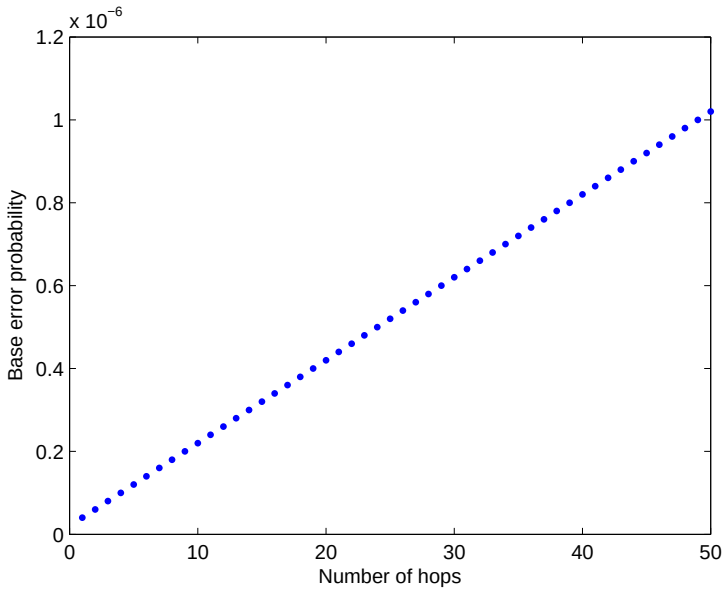


Figure 2.4: Base error probability assuming that the mutation probability is the same in all steps and equal to 10^{-8} as a function of the number of hops

It can be observed that the error probability increases linealy when the number of hops is small, since, when the mutation rate is very small, it is very unlikely to have two mutations in the same base. In that case, the expression of the probability of error can be approximated as:

$$p_e(N) \approx 10^{-8} \cdot (2\mathrm{N} + 2) \tag{2.14}$$

where N is the number of hops.

2.4 Effect on the Capacity

In this section, the objective is to calculate the capacity of a base pair in the system described in the introduction. In that system, the transmitter sends the message encoded using a quaternary alphabet, and the receiver has to decode it. The symbols of this message are modified during the propagation with a certain probability, as it is studied in the previous sections. When an error affects a base pair, the symbol is changed for any of the other three with the same probability. This system can be considered as a quaternary symmetric channel with the following transition matrix:

$$p\left(y|x\right) = \begin{pmatrix} 1-p & \frac{p}{3} & \frac{p}{3} & \frac{p}{3} \\ \frac{p}{3} & 1-p & \frac{p}{3} & \frac{p}{3} \\ \frac{p}{3} & \frac{p}{3} & 1-p & \frac{p}{3} \\ \frac{p}{3} & \frac{p}{3} & \frac{p}{3} & 1-p \end{pmatrix} \tag{2.15}$$

where p is the probability of error of a base calculated in the previous section (2.12). The capacity can be analyzed as follows, by starting from the Shannon expression for the capacity [60].

$$\begin{aligned} C &= \max_{p(x)} I(X,Y) \\ &= \max_{p(x)} \{H(Y) - H(Y|X)\} \\ &= \max_{p(x)} \left\{ H(Y) - \sum_x p(x) H(Y|X=x) \right\} \\ &= \max_{p(x)} H(X) - H\left(1-p_e, \frac{p_e}{3}, \frac{p_e}{3}, \frac{p_e}{3}\right) \\ &= 2 - H\left(1-p_e, \frac{p_e}{3}, \frac{p_e}{3}, \frac{p_e}{3}\right) \end{aligned} \tag{2.16}$$

where p_e is the probability of error of a base pair and $H(a,b,c,d) = -\left(a\log_2(a) + b\log_2(b) + c\log_2(c) + d\log_2(d)\right)$ is the quaternary entropy function [60] (where a, b, c and d are the probabilities of the symmetric channel and $a+b+c+d=1$). In order to simplify the notation, we define:

$$H_4(p) = H\left(1-p, \frac{p}{3}, \frac{p}{3}, \frac{p}{3}\right)) = -\left((1-p)\log_2(p) + 3\left(\frac{p}{3}\right)\log_2\left(\frac{p}{3}\right)\right) \tag{2.17}$$

The maximum of the capacity is when four symbols are equiprobable, which means that $H(X) = 2$.

The quaternary probability if error has to be used instead of the binary probability of error since the errors on the bits are not independent, since one error in a base pair causes one or two consecutive erroneous bits.

By assuming that the mutation rate is the same as the encoding/decoding error probability, the expression of the capacity of a base pair as a function of the number of hops (N) is:

$$C_{bp} = 2 - H_4\left(\frac{3}{4}\left(1 - \left(1 - \frac{4}{3}p_{mut}\right)^{2N+2}\right)\right) \tag{2.18}$$

The capacity of bacterium has to take into account the number of base pairs encoded, so it is $C_{bact} = B \cdot C_{bp}$. In Fig. 2.5, the normalized capacity of a bacterium is represented as a function of the mutation rate for a different number of BC processes is shown. We can observe that for small mutation rate, e.g. 10^{-8}, the normalized capacity does not decrease if the number of hops is reasonable. When the number of hops is higher than 10^4, the capacity decreases only 0.1 %. For a higher mutation rate, the capacity decreases sooner, but it has to be a really high mutation rate to affect any communication in a reasonable number of hops.

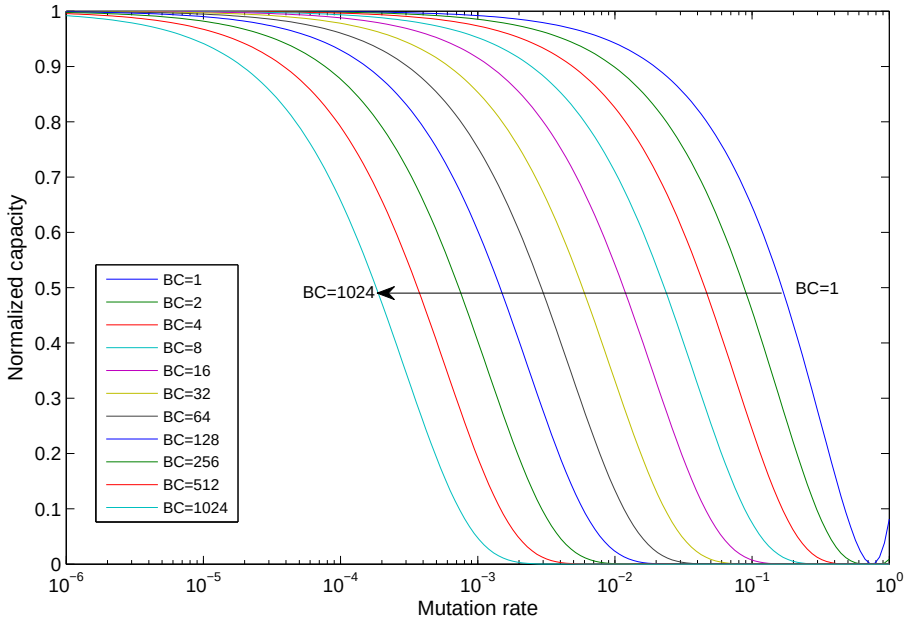


Figure 2.5: Capacity as a function of the mutation rate for a different number of BC processes

2.4.1 Numerical Results

In the last section, numerical results of the base error probability were obtained using the mutation rate found in the nature. In this subsection, the effect of the mutations in the capacity is studied assuming that mutation rate. By using the approximation (2.14), the capacity of a base pair can be written as follows:

$$C_{base}(N) \approx 2 - H_4\left(10^{-8}(2N+2)\right) \approx 2bits/base \tag{2.19}$$

That means that if the mutation rate does not increase doing the processes artificially in the nanoscale, the capacity of a base pair practically will not be modified, being approximately two bits per base. Ans the capacity of a bacterium in this communication system will be approximately proportional to the number of base pairs encoded in the plasmid:

$$C_{bact} \approx BC_{base}(N) \approx 2Bbits/bacterium \tag{2.20}$$

2.5 Conclusions

In this chapter, the effect of the mutations which can occur during the communication process was studied and a close formula for the probability of error of a base pair was obtained. No numerical approximation was applied in the study and the probability of error is expressed as a function of the number of hops, thus obtaining a general formula valid in all cases. Then,

these formulas were used to calculate the capacity of a base pair, as a function of the mutation rate and the number of hops. Finally, some numerical results were obtained by using mutation rate values similar to the values found in the nature.

If parameters related to the biological process found in the nature are used and no more sources of errors are introduced in the nanoscale, then, according to the expressions detailed in this section, the effect of the mutations can be considered negligible. This is true if the number of hops is reasonable. For example, after ten thousands hops, the decrease of the capacity is still very small, but the time needed only for relaying the information ten thousands times is around 60 days.

Chapter 3

Propagation Model of Bacteria in a Nanoscale Communication System

3.1 Introduction

The communication paradigm studied in this work, introduced in Section 1.2, uses bacteria as information carriers, which physically transport the information, which is encoded in DeoxyriboNucleic Acid (DNA) molecules, from one point to another point. Bacteria are able to swim in a liquid medium and can direct their movement by following to the concentration of some chemicals, according to a phenomenon which is called chemotaxis [40]. The receiver nodes or nanodevices are continuously releasing one of these chemical substances, called attractant, in order to attract the bacteria, which can be programmed to follow that specific attractant, so that they can finally reach the correct receiver. The reference bacteria used in this work is Escherichia Coli (E. Coli), because it is one of the flagellated bacteria most widely studied.

The objective of this chapter is study the propagation of a bacterium when it is released in a certain point of the system, the transmitter, and it is attracted to the receiver. In previous work, the propagation of a bacteria was only simulated, by using a 2D enviroment in [36] or a 3D enviroment in [37]. In this work, the propagation analyzed, and close-form expressions for the propagation and for the delay are obtained.

The rest of the chapter is organized as follows. In Section 3.2, the characteristics of the movement of bacteria and the chemotaxis are described. In Section 3.3 an analysis of the propagation is conducted and expressions for the propagation and for the delay are obtained. In Section 3.4, the obtained analytical results are validated by using a simulation tool. Finally, in Section 3.5 the work presented in this chapter is concluded.

3.2 Characteristics of the Bacteria Movement

E. Coli moves in series of "runs" and "tumbles". In each run, the flagella motors spin counterclockwise and the bacterium swims approximately in straight line. A tumble is a small period of time between two different runs which the bacterium moves erratically because some filaments are spinning clockwise. During a running period, the bacterium senses the amount of attractant in the environment several times by using its chemoreceptors. By comparison of the obtained measurements, the bacterium is able to decide whether the attractant concentration is increasing or decreasing. If the attractant concentration is increasing, the running time is longer, but if the concentration is decreasing, the running time remains constant, as if no attractant is detected. This difference of the running time enables bacteria to direct their movement where the attractant concentration is higher.

In the recent decades, an exhaustive research has been conducted in understanding how the flagellar motor of bacteria works. Its structure, parts and how these parts are assembled are well known [38]. Information like the fuel they use, the torque that can generate at different speeds and what controls the likelihood of the direction changes are well studied. However, it is still unknown what makes the bacterium run or tumble and what makes the bacterium change from one state to another. For this reason, bacterial mobility still has a random component that is being widely studied [40]. This random component makes impossible to do a deterministic analysis of the propagation and the different random components in the movement make the microscopic approach unfeasible.

Some studies have been conducted in order to describe the mobility of bacteria. Research on this topic is mostly experimental, and the results presented are based on the observation. As we stated, the movement of bacteria is based on the alternation of runs and tumbles, during the run, the bacterium moves approximately in straight line, and during the tumble the bacterium changes the movement direction. This can be modeled as a random walk. The effect of the chemotaxis makes some runs longer than others, depending on the direction of the movement. This effect biases the random walk, which is named biased random walk. In this section, this biased random walk of bacteria will be described by expressing run time as a function of the attractant concentration gradient, the tumble time, the correlation between the direction of two consecutive runs and the changes of direction during a run.

During the runs, bacteria swim at a speed of 20 μm/s. During the movement, bacteria process the attractant concentration in order to increase the running time when the attractant concentration is increasing. This process behaves as a linear time-invariant system [61], the shape of that system is shown in Fig. 3.1. The positive and negative lobes are equal, which means that the system is stable and inactive when the concentration is constant. When the concentration increases, the run time (λ) increases as the concentration gradient increases, but there is still controversy about the actual gain of the system [62, 63]. When the concentration

decreases, the run time time remains constant, the same as when the concentration is constant. We can define the tumble rate (α) as the inverse of the run time, and it is set as:

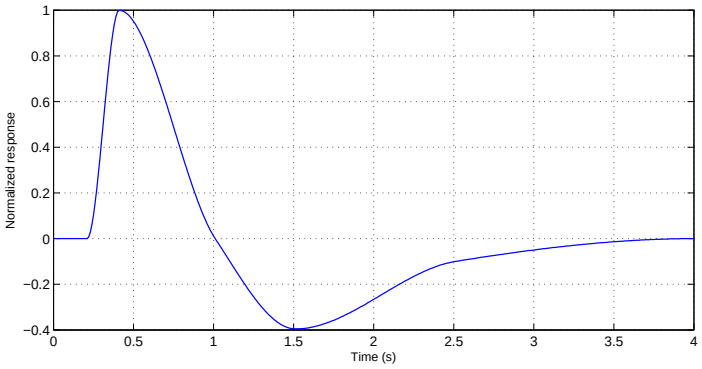


Figure 3.1: Normalized impulse response of E. Coli

$$\alpha(t) = \frac{1}{\lambda} = \begin{cases} \alpha_0 - \bar{g}\int_0^\infty I(\tau)C(t-\tau)d\tau & \text{if } \int_0^\infty I(\tau)C(t-\tau)d\tau > 0 \\ \alpha_0 & \text{if } \int_0^\infty I(\tau)C(t-\tau)d\tau < 0 \end{cases} \tag{3.1}$$

where α_0 is the base tumble rate, without attractant, and for E. Coli is 1 s; $I(t)$ is the impulse response of the bacterium; $C(t)$ is the attractant concentration sensed during the movement of the bacterium; and $\bar{g}$ sets the magnitude of the response to changes in attractant concentration.

It has been empirically observed that the running time is random and it follows an exponential distribution, so the event "the bacterium state changes from run to tumble" can be modeled as an exponential random variable, where the average value is given by the previous expression (3.1). The probability that this event occurs in a certain time t is:

$$P(t, \lambda) = \lambda e^{-\lambda t} \tag{3.2}$$

During runs, the trajectory is not perfectly straight due to the a phenomenon called rotational diffusion. This effect causes bacteria to change their direction by a mean square angular derivation on each axis of $\langle \theta^2 \rangle = 2D_r t$, where D_r is the rotational diffusion coefficient and t is the time. The rotational diffusion coefficient can be calculated as: $D_r = \frac{kT}{8\pi\eta a^3}$, with k the Boltzman constant, T the temperature in Kelvin, η the dynamic viscosity ($\eta = 0.27kg/(ms)$), and a the radius of the bacterium if it is modeled as a sphere. For E. Coli, the rotational diffusion is $0.062rad^2/s$.

During the tumble, bacteria stop, change the direction and continue moving. The tumble time follows also an exponential distribution, but in that case, the average time remains constant; for E. Coli that time is 0.1 s. The new direction is correlated with the old direction, and it can be modeled by the angular correlation, which is defined as the mean cosine of

the angle between successive directions of motion; for E. Coli, the angular correlation is approximately 0.5.

If there is no attractant, the movement of a bacterium can be studied as a random walk, where the direction of two consecutive movements is correlated and where the runs are not completely straight. In the presence of attractant, the movement is biased, which provokes a biased random walk that leads the bacterium to the destination. Fig 3.2 shows an example of a biased random walk in 2D.

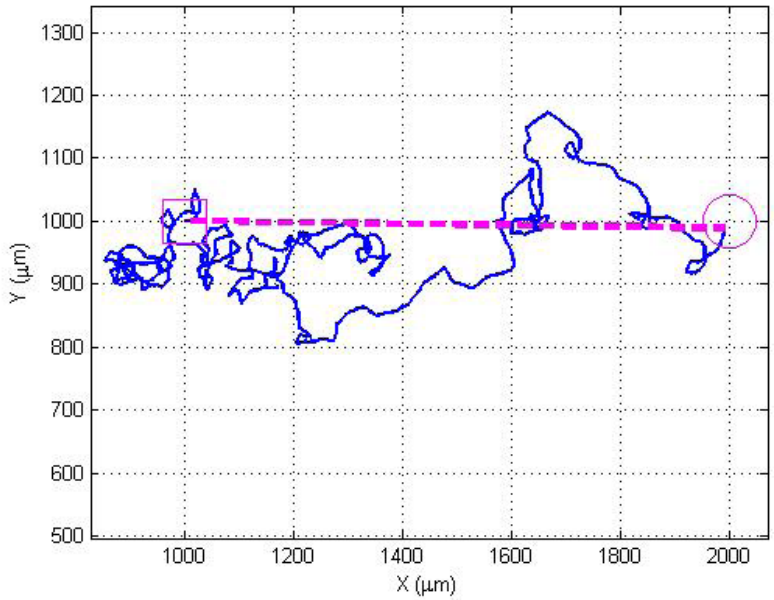


Figure 3.2: 2D biased random walk

3.3 Theoretical Analysis

As we stated just before, the movement of the bacteria can be considered as a random walk. One approach to study the movement of the bacteria is by using the diffusion theory. According to the diffusion theory, the concentration of diffusion particles at any point can be known as a function of the time. Alternatively, by normalizing the number of particle, the result can be also understood as the probability of finding one particle at an arbitrary location in the system. We use this approach to study the propagation of a bacterium, where the result is understood as the probability of finding a bacterium in a certain location at a specific time.

In [64], the movement of bacteria is analyzed from the random walk point of view. Mathematically, the propagation of bacteria is defined as a random walk whose collision rate depends continuously on position and on direction of motion. According to the flux equation, the flow of bacteria per unit area, is given by the following expression:

$$J = -D\nabla c\left(\overline{x},t\right) + v_{drift}(\overline{x})c\left(\overline{x},t\right) \tag{3.3}$$

where J is the bacterial flux, c is the concentration of bacteria as a function of position and time, D represents the diffusion coefficient of the bacteria and v_{drift} is the drift velocity of the system.

The diffusion coefficient of the bacteria, D, takes into account how the bacteria swim in the medium without attractant. It is defined as:

$$D = \frac{v^2}{n\left[\alpha_0\left(1-\Theta\right) + \left(n-1\right)D_{rot}\right]} \tag{3.4}$$

where v is the swimming velocity of bacteria, α_0 is the base tumble rate, the inverse of the base run time without attractant. Θ is the angular correlation between consecutive runs, n is the number of the dimensions where the problem is solved and D_{rot} is the rotational diffusion coefficient.

The drift velocity of the system, v_{drift}, takes into account the effect of the presence of attractant in the movement of the bacteria, the chemotaxis. This is proportional to the gradient of the concentration of attractant. The expression is given by:

$$v_{drift}(\overline{x}) = \frac{v^2 g\nabla C\left(1-\Theta\right)}{n\left[\alpha_0\left(1-\Theta\right) + \left(n-1\right)D_{rot}\right]} \tag{3.5}$$

where g is the gain, it has the dimensions of inverse attractant concentration and is proportional to the gain of (3.1). C is the concentration of attractant as a function of the position, and ∇C is the gradient of the attractant concentration.

The continuity equation states that particles can not be created or destroyed; this means that the number of particles in the system must be constant. In the bacterial analysis, this principle is expressed by the following relation:

$$\frac{\delta c\left(\overline{x},t\right)}{\delta t} = -\nabla J\left(\overline{x},t\right) \tag{3.6}$$

By substituting the flux equation (3.3) into the continuity principle (3.6), we obtain the equation that describes the movement of the bacteria, which is similar to the Smoluchowski equation [65]:

$$\frac{\delta c\left(\overline{x},t\right)}{\delta t} = D\nabla^2 c\left(\overline{x},t\right) - v_{drift}(\overline{x})\nabla c\left(\overline{x},t\right) \tag{3.7}$$

This is a Partial Differential Equation (PDE), a differential equation involving a unknown function, and more than one independent variables and partial derivatives with respect of those variables. The solutions to this PDE are all the functions that describe the movement

of bacteria. If we want a specific solution we need to impose an initial condition and the boundary condition. In our case we consider an infinite space, so we do not need to impose a boundary condition. The bacterium is released in the position of the transmitter, which means that the initial condition can be written as:

$$c\left(\overline{x},0\right)=\delta\left(\overline{x}-\overline{x}_{TX}\right) \tag{3.8}$$

where $\delta\left(\overline{x}\right)$ is the Dirac delta function in the n-dimensional space considered.

The particular solution of this PDE is a function that describes the probability of finding a bacterium in the certain location $\overline{x}$ at a specific time t when this bacterium is released in the position defined in the initial condition and it is attracted by the attractant concentration defined in the drift velocity term of the PDE. In general, this equation has no close solution and has to be solved by using numerical methods [66]. In a simple case, by considering one dimension and by assuming constant drift velocity, the PDE can be solved analytically and a closed-form expression can be obtained.

The final objective of this analysis is to know the distribution of the time needed by a bacterium to reach the receiver, that is, the propagation delay. This can be computed by integrating the solution of the PDE over the volume of the receiver, (3.9). The function obtained is a monotonic increasing function, since the bacteria that reach the volume of the receiver do not leave this volume. The function is zero at $t=0$, and starts to increase after a certain time. If the probability of reaching the receiver is one, which means that the loss probability is zero, the function also reaches one. Otherwise, if the loss probability is higher, the function does not reach one.

$$arr\left(t\right)=\int_{\langle Rx\rangle}c\left(\overline{x},0\right)dVol \tag{3.9}$$

where $\langle Rx\rangle$ is the volume of the receiver.

If the loss probability is zero, the function is directly the Cumulative Distribution Function (cdf) that describes the time needed by a bacterium to reach the receiver. Otherwise, we need to normalize the function in order to define it as a cdf.

3.3.1 Constant Drift Velocity

As stated above, it is possible to find a close-form solution to the PDE if the drift velocity is constant and by solving the equation using one dimension. The 1D case can also be used to solve radial systems, where the center is the position of the receiver and the position of the transmitter corresponds to the distance from the transmitter to the receiver.

A constant drift velocity implies a concentration gradient which is impossible to obtain by releasing attractant from the receiver due to the diffusion process of the particles, but it is a simple way to obtain a close-form solution to the PDE, and it is useful to analyze the result.

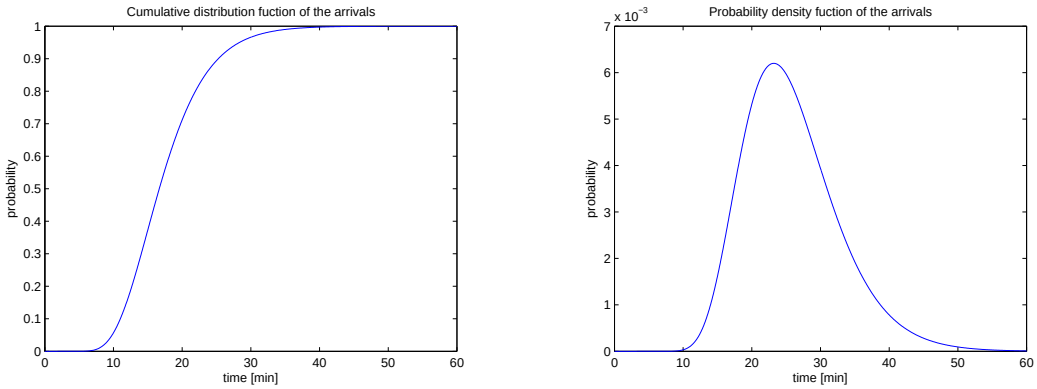


Figure 3.3: Probability Density Function (pdf) and Cumulative Distribution Function (cdf) of the arrival.

The initial condition and the drift velocity are defined as follows: the bacterium is released in $x = 0$ and the drift velocity, v_{drift}, is constant and positive, which implies that the bacterium moves to the positive numbers. By solving this system, the solution of the PDE is:

$$c\left(x,0\right) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{\left(x - v_{drift}\cdot t\right)^2}{4Dt}} \tag{3.10}$$

The form of the equation is the same than the equation that describes the particle diffusion when the particles are released in a certain point [67], but it is modified by the effect of the chemotaxis. Since the drift velocity is considered constant, all the particles are moving together, which does not affect the diffusion process.

By integrating this equation in the area of the receiver, we found the expression that describes the distribution of the time needed by a bacterium to reach the receiver (3.11). Then, the density distribution can be found by taking the derivative of the equation in (3.12). To calculate this expression, we integrate from the position of the receiver to infinity. This is because the drift velocity is considered constant and does not take into account the position of the receiver.

$$F_{arr}(t) = \frac{1}{2}\mathrm{erfc}\left(\frac{R - v_{drift}\cdot t}{\sqrt{4Dt}}\right) \tag{3.11}$$

$$f_{arr}(t) = \frac{v_{drift}t + R}{4\pi t^{3/2}}\mathrm{e}^{-\frac{\left(\frac{R - v_{drift}t}{\sqrt{4Dt}}\right)^2}{4\pi t}} \tag{3.12}$$

where R es the position of the receiver, $F_{arr}(t)$ is the cdf and $f_{arr}(t)$ is the Probability Density Function (pdf).

In Fig. 3.3, a particular solution of the system is shown. The following values are used: $v = 20\mu m/s, \alpha_0 = 1s^{-1}, \Theta = 0.5, n = 3, g\nabla C = 0.02\mu m^{-1}, D_{rot} = 0.062rad^2/s$

3.3.2 General Solution: Numerical Method

In a real scenario, the drift velocity is determined by the attractant concentration gradient that a nanodevice can create. To obtain the gradient, each nanodevice is continuously releasing particles of attractant from its position, and the gradient is created by the diffusion of these particles. The equation which describes the distribution of particles continuously released from a certain point by diffusion in the steady state is [67]:

$$C\left(r\right) = \frac{1}{10^6}\frac{Q}{4D\pi r} \tag{3.13}$$

where r is the distance from the point of release, C is the concentration, Q is the attractant emission rate and D is the diffusion coefficient of the attractant in the medium. The usual diffusion coefficient for the attractants that can be sensed by E. Coli is around $10^{-9}m^2/s$.

The attractant distribution decreases inversely proportional with the distance, and the peak is in the position where the attractant is released. In order to calculate the drift velocity we need to compute the gradient of the attractant. In a three dimensional space, the gradient is defined as: $\nabla C\left(\overline{x},t\right) = \left(\frac{\partial C(\overline{x},t)}{\partial x}, \frac{\partial C(\overline{x},t)}{\partial y}, \frac{\partial C(\overline{x},t)}{\partial z}\right)$

The concentration gradient of the attractant created by a nanodevice (3.13) can be expressed as a function of the distance to the nanodevice as follows:

$$\nabla C\left(r\right) = \frac{1}{10^6}\frac{Q}{4D\pi r^2} \tag{3.14}$$

In Fig. 3.4 we show the attractant concentration in the steady state and the gradient, calculated using $D = 10^{-9}m^2/s$ and $Q = 10^{-10}mols/s$. We observe that the attractant concentration decreases very fast $(1/r)$, and the gradient of the attractant decreases even faster $(1/r^2)$. This causes bacteria to advance really slow when they are far from the receiver or even to get lost.

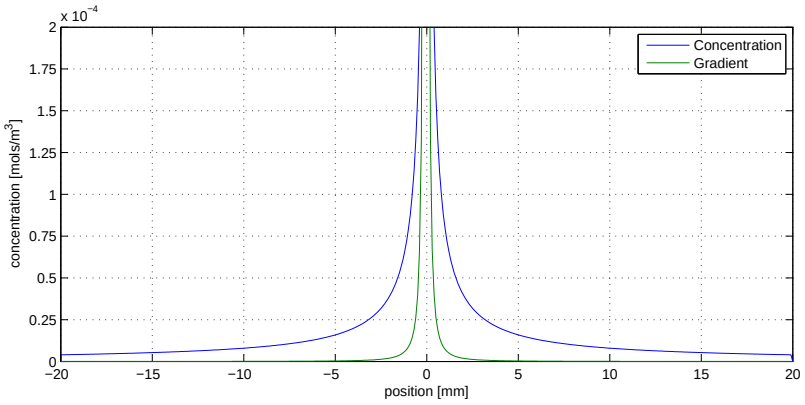


Figure 3.4: Concentration of attractant

As stated above, it is not possible to obtain a close-form solution of the PDE that defines the movement of bacteria when the term of the drift velocity is not constant, but a solution can

be obtained by using numerical methods. Numerical methods are often used in engineering to obtain the solution of differential equation.

The numerical method used to solve this PDE is the Finite Difference Method (FDM). This method allows to obtain a numerical solution easily when the problem does not have boundary conditions. This method is based on the use of finite difference equations to approximate the derivatives. The time and the space are discretized and the method is applied to find the values of the discretized function.

An example of the application of the method in a one dimension scenario, is shown in (3.15). To obtain the solution, the $c(\overline{x}, t)$ is calculated for all the discretized points in the space, Δx, and for each increment of time Δt. In order to obtain a numerical solution of the PDE for a higher dimension, the other partial derivatives have to be taken into account in the gradient. In Fig. 3.5 we show different instants of the numerical solution of a two dimension scenario.

$$\frac{c(\overline{x}, t) - c(\overline{x}, t - \Delta t)}{\Delta t} = D\frac{c(\overline{x} + \Delta x, t - \Delta t) - 2c(\overline{x}, t - \Delta t) + c(\overline{x} - \Delta x, t - \Delta t)}{\Delta x^2} - \\ -v_{drift}(\overline{x})\frac{c(\overline{x}, t - \Delta t) - c(\overline{x} - \Delta x, t - \Delta t)}{\Delta x} \tag{3.15}$$

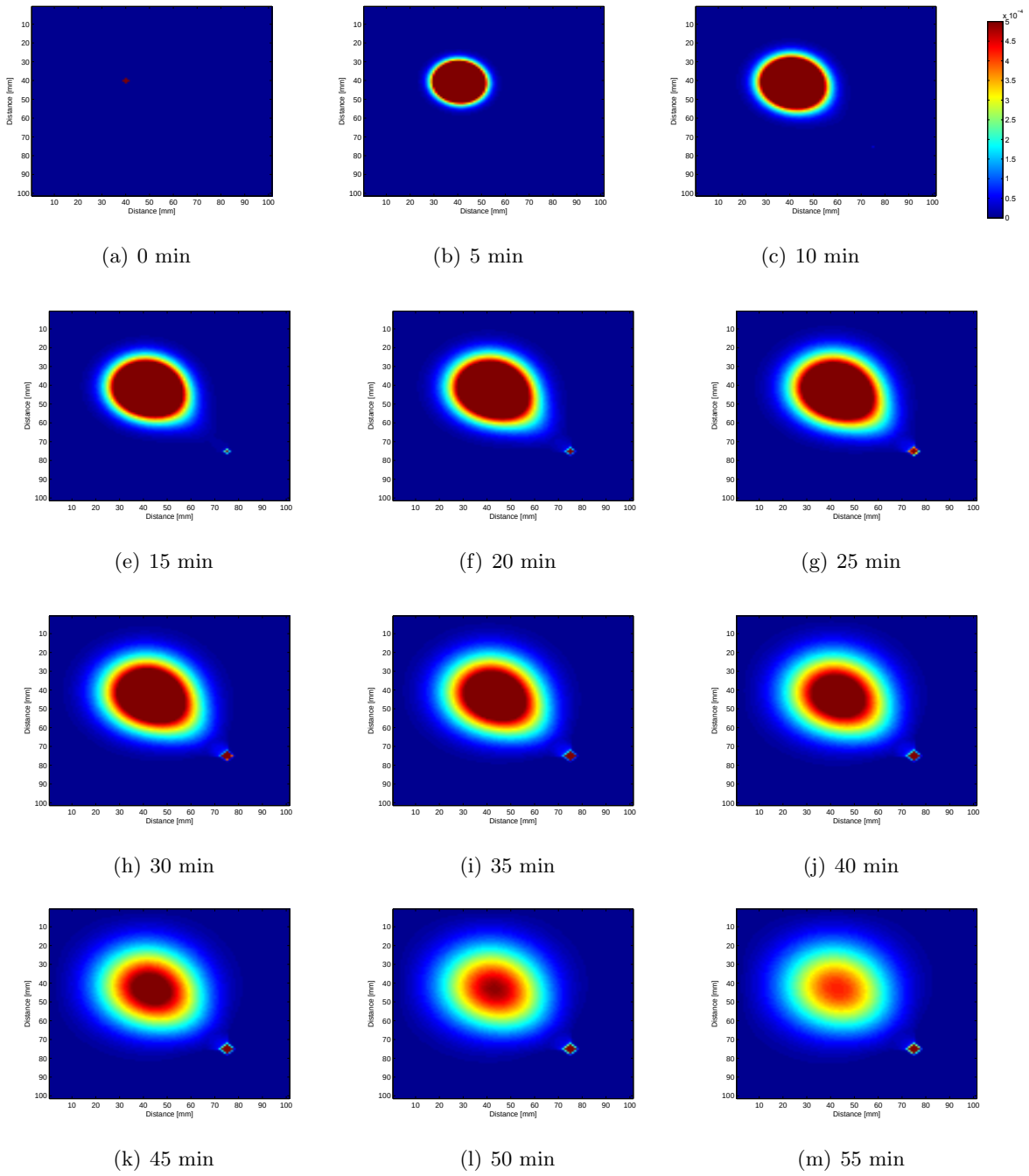
Once the numerical solution from the PDE is obtained, which corresponds to the distribution of the bacteria, we can calculate the distribution of the time needed by the bacterium to reach the receiver. This can be obtained by computing numerically the integral of the region in the receiver. The distribution of the arrivals is shown in Fig. 3.6 for three different distances between the transmitter and the receiver.

3.4 Comparison with the Simulator

The objective of this section is to validate the theoretical model with the simulation tool previously developed in [37].

The simulation tool works according to the bacterial chemotaxis model presented before in Sec. 3.2. It simulates the runs, with the rotational diffusion; and the tumbles, when the bacterium turns in a correlated angle. The attractant sensing process is also simulated by computing continuously the convolution of the attractant concentration in order to determine the run time. To perform the simulation, the time is divided in time steps of 0.01 s and in each time step, the position, the state and the changing probability are updated.

To validate the model, the propagation of a bacterium was simulated multiple times and averaged, and a simulated version of $c(\overline{x}, 0)$ was found. The comparison between the simulated data and the data obtained using the numerical method is shown in Fig. 3.7.



(a) 0 min (b) 5 min (c) 10 min

(e) 15 min (f) 20 min (g) 25 min

(h) 30 min (i) 35 min (j) 40 min

(k) 45 min (l) 50 min (m) 55 min

Figure 3.5: Probability Density Function (pdf) of the position of a bacterium in the 2D case every 5 min

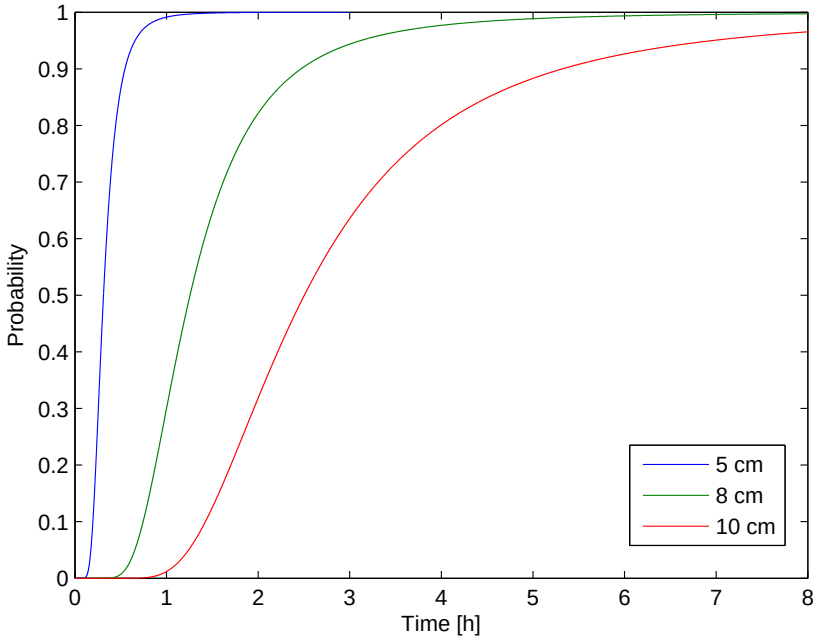
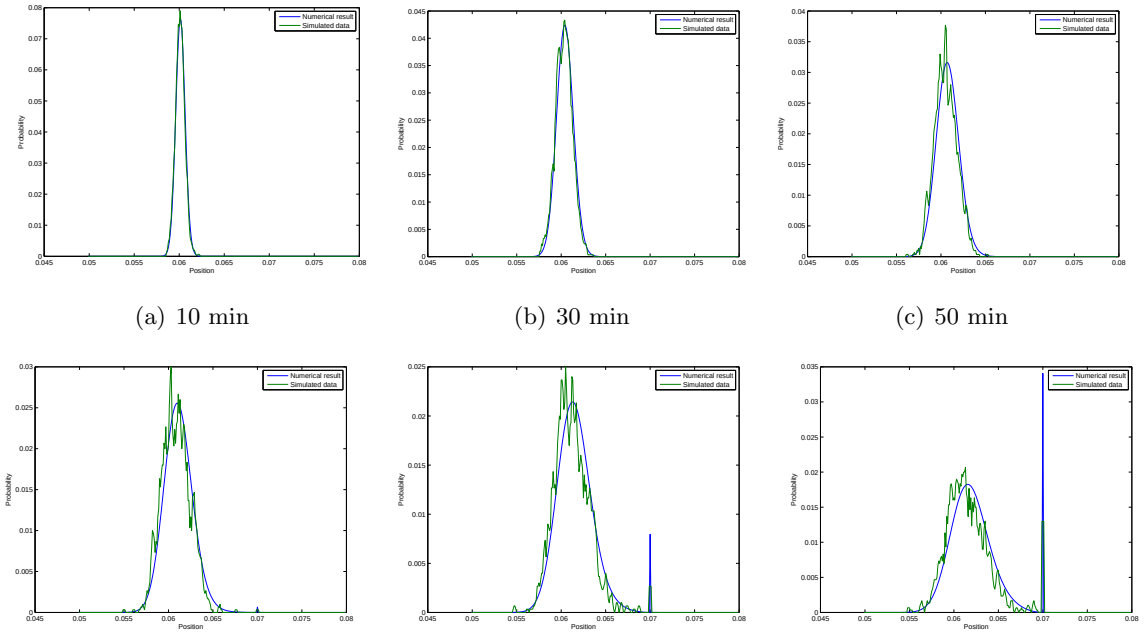


Figure 3.6: Distribution of the arrival when the distance is 5 mm, 8 mm or 10 mm



(a) 10 min (b) 30 min (c) 50 min

(d) 70 min (e) 90 min (f) 110 min

Figure 3.7: Comparison with the simulation tool

The distribution of the arrivals is validated by averaging out the result of simulating the time needed by a bacterium to reach the receiver. The comparison is shown in Fig. 3.8

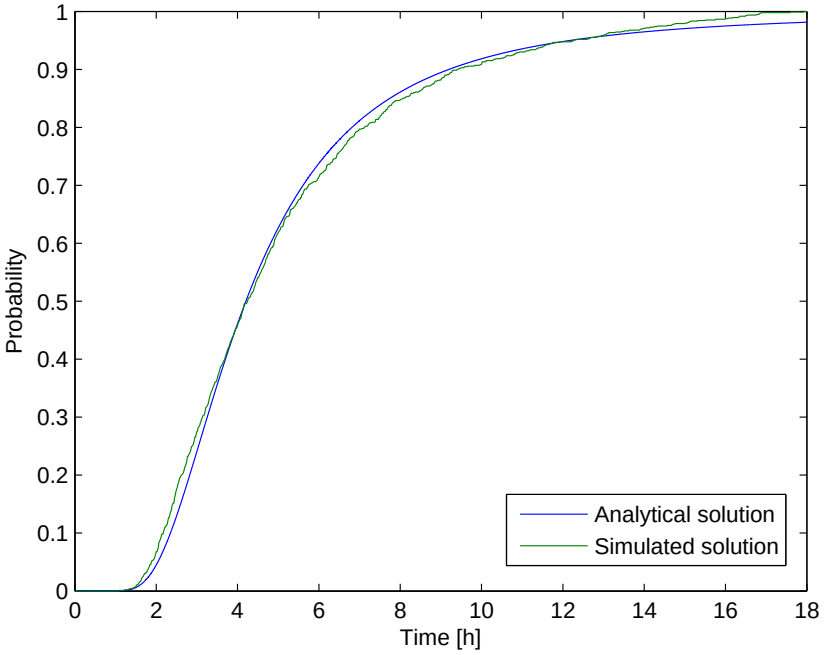


Figure 3.8: comparison of the simulated data and the analytical solution for 10 mm

Once we analyze the simulation results and we compare them with the analytical results we conclude that both models match. But there are some advantages and disadvantages in each case.

The advantages of the simulation tool against the analytical model are that while some effects can be taken into account in the simulation, they cannot be included in the analytical model. These effects are the saturation of the chemotaxis process [68] and the eye-of-a-needle effect [69] when the bacterium is approaching to the receiver. This effects can be avoided if the quantity of attractant released does not saturate the bacterium and if the radius of the receiver is enough to avoid this effect.

The advantages of the analytical model against the simulation tool are that the numerical method used to solve the PDE is much faster than averaging enough samples of the simulator and that the rounding errors of the simulation tool are avoided. The rounding errors become important when the simulation tool is processing small differences of attractant.

We can conclude that the theoretical model is valid to obtain the delay of the bacteria when the chemotaxis is not affected by saturation.

3.5 Conclusions

In this chapter, we proposed a theoretical model for the propagation of bacteria based on the diffusion theory. By using an analytical model, we avoided the use of simulation tools in order

to characterize the propagation and obtain the distribution of the time needed by a bacterium to reach the receiver.

We first described how the bacteria move and focused on the chemotaxis process that allow bacteria to direct their movement to a source of attractant. We realized that the propagation of bacteria attracted to a certain point can be analyzed as a biased random walk; and we used the diffusion theory in order to obtain an equation which describes the propagation of the bacteria in a medium in the presence of attractant. The result of that equation allowed us to obtain the distribution of the delay of the bacteria. Moreover, we compared both the theoretical results and the results of the simulation in order to validate the model.

The analysis of the results showed that the mathematical model is valid, with the exception of the case of the saturation of the bacteria. Even if a close-form solution can not be obtained, the numerical method used to obtain the solution is much faster than the simulations. Numerical methods are widely used in engineering, specially in PDE, since this kind of equations often do not have a close-form solution.

Chapter 4

A synchronous Communication Scheme for Bacteria-Based Communication in the Nanoscale

4.1 Introduction

The final objective of the study of flagellated bacteria is to be able to establish a communication among nanodevices. In the introduction, the basic communication steps and different possible architectures were introduced. The objective of this and the following chapter is to study different systems based on flagellated bacteria.

In Chapter 2, the effect of the mutations which modify the information during the propagation of the bacteria was studied, by assuming a multi-hop communication, but without taking into account other sources of errors during the propagation, such as the loss probability. In Chapter 3, the propagation of a bacterium attracted by a receiver was studied and the distribution of the delay needed by one bacterium to reach a receiver was obtained.

In this chapter, we introduce a synchronous system based on the synchronization among nanodevices. We focus on the study of the system capacity and the delay by using the results obtained in the previous chapters.

The rest of the chapter is organized as follows. In Section 4.2 the synchronous system presented is defined. In Section 4.3 the delay and the capacity are calculated in a point-to-point scenario and in Section 4.4 the analysis is extended to the multi-hop scenario. Finally, the chapter is concluded in Section 4.5.

4.2 System Definition

For this communication scheme, we assume that all the nanodevices are synchronized. This synchronization allows to establish common timing between nanodevices.

In this system, the time is divided into slots of duration T, the slot duration is the period of time during which the bacterium has to reach the receiver in order to achieve a correct reception. If it does not reach the receiver during the time slot, the receiver does not consider the bacterium and the information is lost. During each time slot, more than one bacterium can be sent, but only one needs to reach the receiver. In order to maximize the probability of one bacterium to reach the receiver, a delay d is used between the time the bacteria is released and the time the receiver can receive it: in this way we can maximize the probability of receiving a bacterium.

If the final destination is not in the range of the transmitter, bacteria are sent to an intermediate nanodevice and this relays the bacteria until the information reaches the final destination. The relaying process was introduced in Section 1.2.3 and the peculiarities of the relaying process in this system are analyzed in Section 4.4.

In the study of the system, we use the simplest nanonetwork architecture introduced in Section 1.2.2. We consider the following assumptions:

- A constant flux of data is sent from the transmitter to the final destination
- A 3D environment is considered
- All nanodevices are arranged in a perfect grid
- The steady state is assumed for the attractant concentration
- The interferences between attractants are not considered
- The possible saturation in intermediate nodes is not considered

A graphical representation of the timing of the system is shown in Fig. 4.1, where a burst of bacteria is sent, relayed two times and finally received.

4.3 Point-to-Point Scenario

In this section the delay and the capacity for the point-to-point scenario are analyzed. This is a specific case of a multi-hop communication but it allows to start a simple analysis and study its behavior.

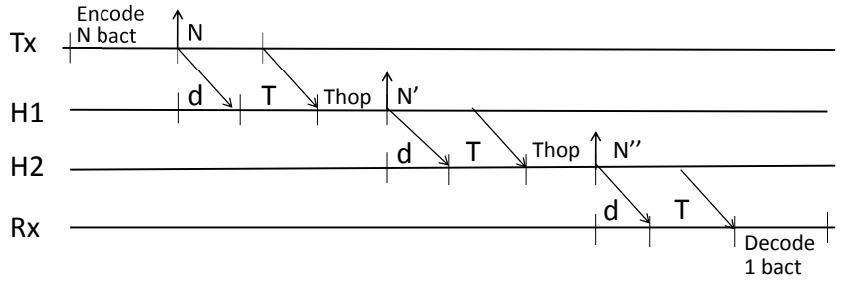


Figure 4.1: Graphical representation of the synchronous scheme

4.3.1 Delay

The complete delay of all the communication takes into account the time needed to encode all the plasmids in the bacteria, the delay waited before the time slot, the duration of the time slot and the time needed to decode the information:

$$delay = t_{tx}(B, N) + d + T + t_{rx}(B) \tag{4.1}$$

where d and T are the delay and the duration of the time slot respectively, $t_{tx}(N, B)$ is the time needed to encode and encapsulate N bacteria with B base pairs in the message region and $t_{rx}(B)$ is the time needed to decapsulate and decode one bacterium with B base pairs in the message region:

$$t_{tx}(B, N) = N \cdot t_{enc}(B) + t_{BC}(B) + t_{actChem} \tag{4.2}$$

$$t_{rx}(B) = t_{BC}(B) + t_{decode}(B) \tag{4.3}$$

The time for the encoding/decoding the DNA, for encapsulation/decapsulation the plasmid (Bacterial Conjugation) and the time needed to activate the chemotaxis were introduced in Section 1.2.3. It has to be taken into account that part of the plasmid contains other information besides of the message, and the time needed to encode that part has to be considered as well. The times needed by a plasmid of 1.6 Mbps, which contains 1.55 Mbps of information, are: $t_{enc} = t_{dec} \approx 19min$, $t_{BC} \approx 32min$ and $t_{actChem} \approx 20min$

4.3.2 Capacity of the System

The capacity of the system depends on the probability of the bacterium to reach the receiver, the information carried by the bacterium and the degradation of the information caused by the mutations during the communication. The information carried by a bacterium depends on the number of base pairs encoded and the degradation operated by the mutation, which was studied in Chapter 2. The probability of a bacterium to reach the receiver in this system

corresponds to the probability of receiving at least one bacterium in the time slot. The capacity expressed in bacteria/s is:

$$C_{bact}(d,T,N) = \frac{P_{Rx1bac}(d,T,N)}{T} \tag{4.4}$$

where d is the duration of the delay, T is the duration of the time slot and N is the number of bacteria used. To obtain the capacity of the system in bits/s, we have to take into account the number of base pairs, each of them contributing to the capacity.

$$C_{system}(B,d,T,N) = B \cdot C_{base}(N)\frac{P_{Rx1bac}(d,T,N)}{T} \tag{4.5}$$

The probability of receiving at least one bacterium can be calculated from the bacterium loss probability:

$$P_{Rx1bact}(d,T,N) = 1 - P_{1Loss}(d,T)^N \tag{4.6}$$

where $P_{1Loss}(d,T)$ is the bacterium loss probability. This probability can be calculated by using the propagation model introduced in Chapter 3, but other errors can effect the propagation of the bacterium. These errors are mutations in the active zone of the plasmid, that do not modify the message but can modify the behavior of the bacterium.

$$P_{1Loss}(d,T) = P_{PathLoss}(d,T) + P_{errBact} \tag{4.7}$$

where $P_{PathLoss}(d,T)$ is the probability of one bacterium not to reach the receiver in the time slot due to the random walk and $P_{errBact}$ is the probability that a mutation affects the active zone.

The first probability can be calculated as follows:

$$P_{PathLoss}(d,T) = 1 - (arr(T+d) - arr(d)) \tag{4.8}$$

where $arr(t)$ is the function obtained in chapter 3 that describes the distribution of the arrivals to the receiver.

The probability that a mutation in the active zone affects the propagation is at most the probability of having an error in the active zone. Taking into account that the size of the active zone is 50 kbps, the probability is at most:

$$P_{errBact} \leq 1 - (1 - P_{err}(1))^{5 \cdot 10^4} \tag{4.9}$$

where P_{err} is the probability of error in a base pair, obtained in (2.12).

Once we have the expression which defines the probability of receiving at least one bacterium during the time slot, we complete the theoretical expression by defining the capacity of the system (4.5). At this point, for a certain architecture of the nanonetwork, there are four free variables: the delay, the duration of the time slot, the number of sent bacteria and the number of base pairs encoded in each bacterium. The objective is to maximize the capacity with the contraint given by the following relationship:

$$\begin{aligned} & T \geq T_{enc}(B, N) \\ & B \in [0, \text{Bacteria Capacity}] \\ & N \in \mathbb{N} \end{aligned} \tag{4.10}$$

where $T_{enc}(B, N)$ is the time needed to encode the DNA of the plasmid, (4.2). Bacteria Capacity is the maximum number of base pairs that plasmid can carry.

4.3.3 Numerical Results

In the last subsection, theoretical formulas to calculate the capacity and the delay were introduced. In this subsection some numerical results are presented. The capacity (C) is calculated by optimizing the free variables introduced previously: delay (d), time slot duration (T), number of base pairs encoded (B) and number of bacteria sent (N). The system delay is also calculated by including also the time needed to prepare all bacteria (t_{tx}) and the time needed to decode one bacterium (t_{rx}). The arrival distribution used in the calculations is found as introduced in Chapter 2, with a releasing rate of $10^{11} mol/s$.

In the following table, these results are computed by using different distances among nanodevices.

Distance	5 mm	8 mm	10 mm	12 mm
$P_{Rx1bact}$	0.996	0.817	0.599	0.4345
d	5 min	33 min	74 min	134 min
T	71 min	90 min	109 min	128 min
B	1.55 mbp	1.55 mbp	1.55 mbp	1.55 mbp
N	1	2	3	4
C	725 bps	554 bps	443 bps	362 bps
delay	198 min	264 min	343 min	441 min
t_{tx}	71 min	90 min	109 min	128 min
t_{rx}	51 min	51 min	51 min	51 min

Table 4.1: Numerical values of the capacity calculation

It can be observed that the time needed to receive a bacterium is constant, since only one bacterium needs to be decoded. Also, the time needed to transmit the bacteria depends on

the the number of bacteria sent and it coincides with the duration of the time slot. This is because the duration of the time slot has to be at least as long as this time to encode all the released bacteria in the following time slot; at the same time, the optimal duration of the time slot corresponds to the minimum possible in order to receive the bacteria when it is more probable in their arrival distribution. It can be proved that the period of time where it is more probable to receive bacteria coincides with the position of the time slot. The number of bacteria released in each time slot is increased when the probability of receiving at least one bacterium over the duration of the time slot gets higher.

Finally, it can be observed that, for short distance between nanodevices, higher performance is achieved in the capacity and the delay since the propagation time is shorter, even if the time needed to decode one bacterium and the time needed to encapsulate and activate the chemotaxis can not be reduced.

4.4 Multi Hop Scenario

In this scenario, the final destination can not be reached directly from the transmitter, so other nanodevices have to be used in order to relay the information to the final receiver. The relaying system was explained in detail in Section 1.2.3, and it summarized as follows: each nanodevice has an unique address among the network and uses one of the possible attractants, the number of attractants is limited, they have to be reused. Bacteria carry information about the final network address and the attractant that has to be followed in each hop encoded in the plasmid. In each hop, the plasmid is modified in order to update the attractant that has to be followed.

As it was stated in the introduction of this chapter, only one bacterium needs to reach the receiver in order to receive the data. In a multi-hop scenario, this brings two possible solutions:

Resend only the bacteria received

In this option, each intermediate nanodevice only sends (relays) the bacteria that reach the receiver during the time slot. This option is simpler than the other, which is poposed next, but it is less reliable since the number of relayed bacteria per hop reduces during the propagation.

Send each time the same number of bacteria

In this option, each intermediate nanodevice sends the same number of bacteria. This option is more complex and increases the time spent in each hop but it also increases the probability of receiving bacteria.

The choice between the two solutions affects both the delay and the capacity, so both options are studied in the following analysis.

4.4.1 Delay

The delay of the full system takes into account the encoding time, the propagation time of all hops, the time needed to relay bacteria at the intermediate nodes and the time needed to decode the information at the destination:

$$delay = t_{tx}(B, N) + S\dot{(}d + T) + (S - 1) \cdot t_{hop}(B, N) + t_{rx}(B) \tag{4.11}$$

where S is the number of hops, and $t_{hop}(N)$ depends on the solution used.

For the first solution, when only bacteria that reach the receiver are sent again, the time needed in the intermediate nodes is:

$$t_{hop}(B, N) = t_{BC}(B) + N \cdot t_{encAddress} + t_{BC}(B) \tag{4.12}$$

where it has to be taken into account that in the worst case, all bacteria can reach the receiver within the time slot. Morover, the encapsulation process and the decapsulation process can be made in parallel and the time $t_{encAddress}$ needed to modify the attractant the bacterium is sensitive to is constant, as described in Section 1.2.3.

For the second option, when the same number of bacteria is sent at each intermediate nanodevice, the time needed at the intermediate nodes is:

$$t_{hop}(B, N) = t_{BC}(B) + N \cdot t_{enc}(B) + t_{encAddress} + t_{actChem} + t_{BC}(B) \tag{4.13}$$

where it has to be taken into account that in the worst case, only one bacterium can reach the receiver within the time slot, and after its reception, it is necessary to sequence the information ($t_{dec} = t_{enc}$), modify the address, regenerate $N - 1$ copies and encapsulate them.

4.4.2 Capacity

The expression of the capacity in the multi-hop scenario has the same form than the expression for the point-to-point scenario (4.4), but in this case, the numerator contains the probability of receiving at least one bacterium in the final receiver, after all hops.

In the first solution, when the lost bacteria are not encoded again by the relay node, it is necessary that at least one bacterium will reach the final receiver, while the others can be lost at any hop or can reach the receiver as well. This can be modeled by following a recursive fashion, where the base case is as follows: if the number of hops is zero, the probability of receiving a bacterium is one. The recursive case is modeled as follows: in each step, any number of bacteria but one can be lost and the successive probability has to take into account the number of lost bacteria. The mathematical model is given by the following expression:

$$P_{Rx1bact}(N,S) = \begin{cases} \sum_{i=0}^{N-1} \binom{N}{i} (1-P_{1Rx})^i \, {P_{1Rx}}^{N-i} P_{Rx1bact}(N-i,S-1) & \text{if } S>0 \\ 1 & \text{if } S=0 \end{cases} \tag{4.14}$$

where $P_{Rx1bact}$ and P_{1Rx} are also dependent on d and T, not shown in the expression in order to facilitate the reading.

In the second solution, all bacteria are encoded and sent again in each hop, so the probability of receiving at least one bacteria at the final destination corresponds to the probability of receiving at least one bacterium in each intermediate node. This can be expressed as follows:

$$P_{Rx1bact}(N,S,d,T) = \left(1-(1-P_{1Rx}(d,T))^N\right)^S \tag{4.15}$$

Once we have these expression, the expression of the capacity is as follows::

$$C_{ch}(B,d,T,N,S) = B \cdot C_{base}(S) \frac{P_{Rx1bac}(d,T,N,S)}{T} \tag{4.16}$$

and can be maximized taking into account the same conditions than the point-to-point scenario (4.10)

4.4.3 Numerical Results

As we did for the point-to-point scenario, some numerical results are calculated in this subsection, for both relaying solutions introduced. First, the optimal number of bacteria, the capacity and the delay are shown and analyzed when the distance between nanodevices is 10 mm. Then, the same results are shown using different distances and we analyze how the distance affects this variables.

In Fig. 4.2, the evolution of the number of bacteria needed is shown as a function of the number of hops. The number of sent bacteria is a whole number, and this causes the capacity and the delay to change the trend as the optimal number of bacteria changes. Another thing to be observed is that using the second solution (each intermediate nanodevice sends the same number of bacteria) the number of bacteria needed is smaller and almost constant, while when using the other solution, the number of bacteria keeps increasing with the number of hops. Less bacteria have to be sent in the second solution because the same number of bacteria is guaranteed at each hop.

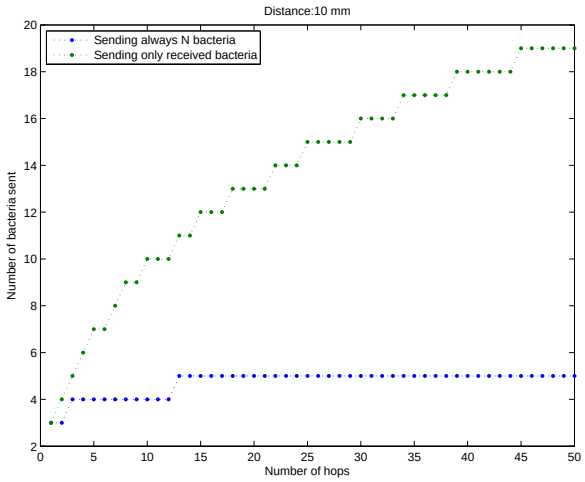


Figure 4.2: Evolution of the number of bacteria needed for each relaying option as a function of the number of hops

In Fig. 4.3, the evolution of the capacity is shown as a function of the number of hops. It can be observed that the capacity decreases as the number of hops increases, but in the first solution the capacity decreases much quickly than the second solution. This is because the bacteria that get lost are not substituted.

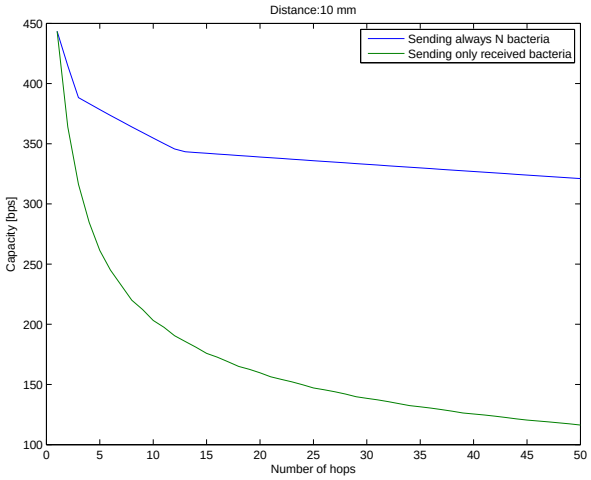


Figure 4.3: Evolution of the capacity for each relaying option as a function of the number of hops

In Fig. 4.4, the evolution of the delay is shown as a function of the number of hops. It can be observed that the delay in the first solution is higher, even if the relaying time in each hop is shorter. This is because the duration of the time slot has to be higher in order to assure that the maximum number of bacteria reach the intermediate nanodevices.

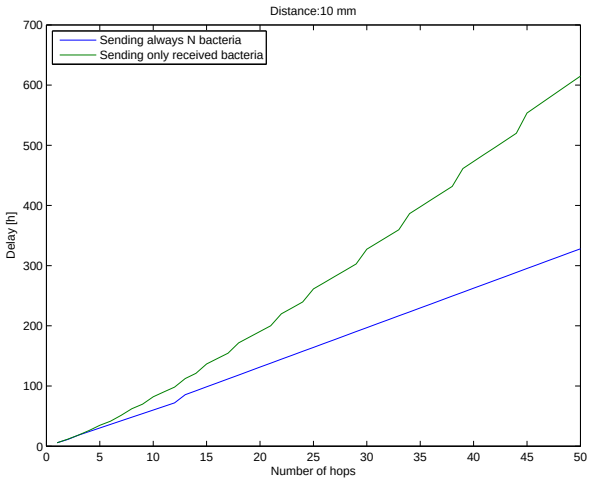
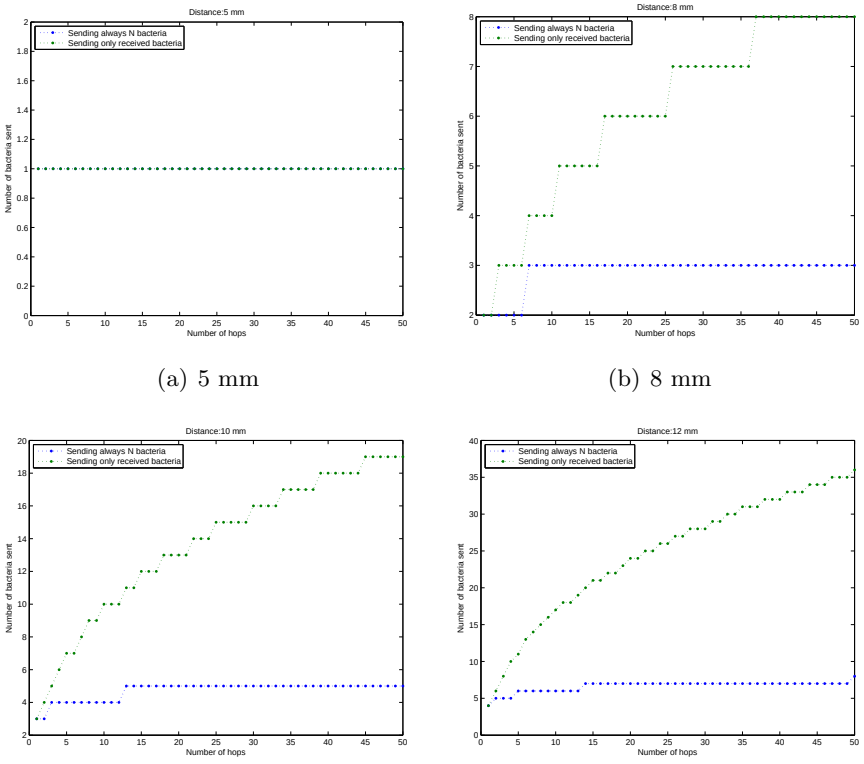


Figure 4.4: Evolution of the delay for each relaying option as a function of the number of hops

In Fig. 4.5, Fig. 4.6 and Fig. 4.7, the number of bacteria needed, the capacity and the delay are analyzed using different distances among nanodevices. We observe that when the distance among nanodevices is small, both solutions are equivalent, because the probability of receiving a bacterium within a time slot is almost one (the duration of the time slot is at least the time needed to encode one bacterium).

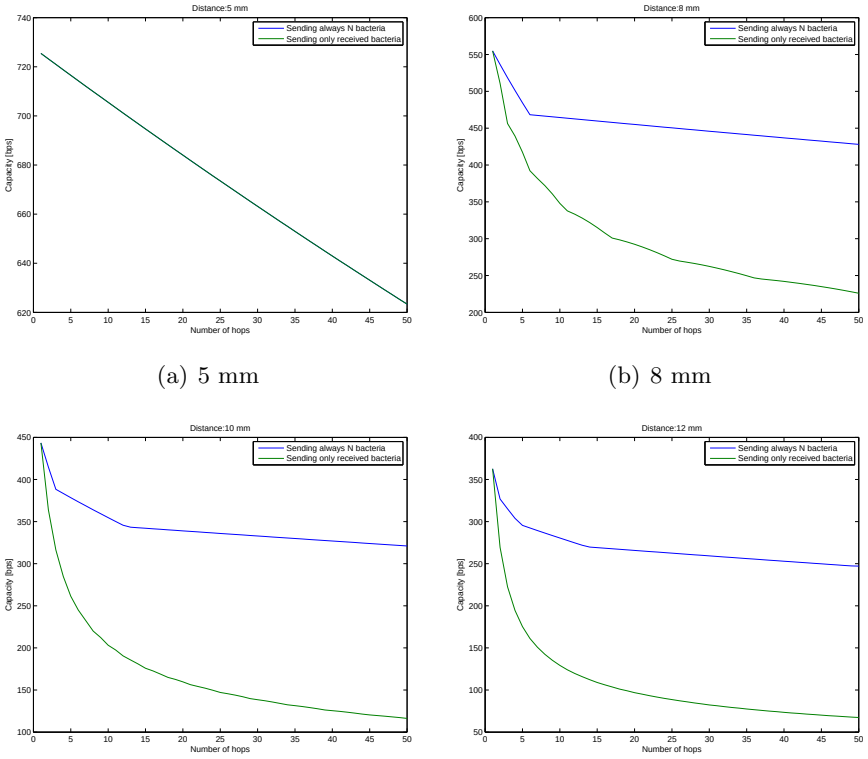


(a) 5 mm (b) 8 mm

(c) 10 mm (d) 12 mm

Figure 4.5: Evolution of the number of bacteria needed for both options as a function of the number of hops using different distances among nanodevices

It can be observed in Fig. 4.5 that the number of bacteria needed increases for both solutions as the distance between nanodevices increases, but the increment is higher in the first solution. This is because, as the distance increases, the probability of receiving a bacterium within the time slot decreases and more bacteria are needed to counteract the decrease in capacity.



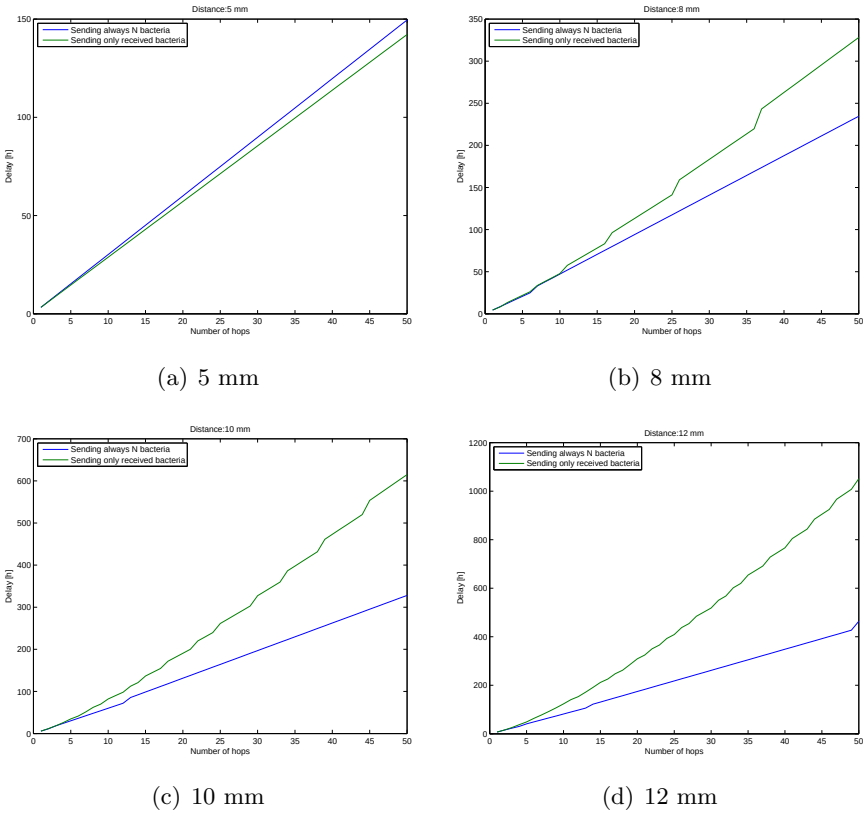
(a) 5 mm (b) 8 mm

(c) 10 mm (d) 12 mm

Figure 4.6: Evolution of the capacity for both option as a function of the number of hops using different distances among nanodevices

It can be observed in Fig. 4.6 that the maximum capacity decreases as the distance increases, and that the capacity of the second solution is always higher. The difference on the capacity values is due to the reliability when using the same number of bacteria in each hop.

It can be observed in Fig. 4.7 that the delay increases as the distance increases, and that the delay is not proportional to the distance. For example, if we double the distance, the increment of the delay is more than the double. This is because the effect of the chemotaxis is reduced by the distance. Another interesting thing to be observed is that the difference of the delay in both the solutions increases as the distance increases.



(a) 5 mm (b) 8 mm

(c) 10 mm (d) 12 mm

Figure 4.7: Evolution of the delay for both option as a function of the number of hops using different distances among nanodevices

4.5 Conclusions

In this chapter, we introduced a new communication scheme for a system based on flagellated bacteria. The scheme is characterized by the synchronization among nanodevices. By assuming this synchronization, we defined the communication scheme and calculated the capacity and the delay, first in a point-to-point scenario and finally in a multi-hop scenario.

After analyzing the results, we discussed some advantages and disadvantages. The main advantage of the synchronous system is that the delay is constant and known for all the information sent, which means that the message will be received after a certain time in the same order that it was sent. Another important advantages is that it is efficient in a long range transmission, when it needs a huge number of hops. But a synchronous system presents also an important drawback: this kind of systems is more complex to implement.

Chapter 5

An Asynchronous Communication Scheme for Bacteria-Based Communication in the Nanoscale

5.1 Introduction

In the last chapter, we introduced a synchronous scheme for a flagellated bacteria communication system, which allows a synchronous communication among nanodevices in the network. In this chapter we introduce an asynchronous communication system, and we calculate the delay and the capacity for the point-to-point scenario and for the multi-hop scenario. An asynchronous system is easier to implement since it does not need any kind of synchronization among nanodevices but its reliability can be smaller.

The rest of the chapter is organized as follows. The system is defined in Section 5.2. Then the simple case of the point-to-point scenario is analyzed in Section 5.3 and then extended to the multi hop scenario in Section 5.4. Finally, the chapter is concluded in Section 5.5.

5.2 System Definition

For this communication scheme, we assume that nanodevices are not synchronized. This means that it is not possible to establish any kind of timing for realizing the communication. Bacteria can reach a nanodevice at any time or get lost, due to the random nature of the propagation.

The transmitter is continuously releasing bacteria, at the maximum rate determined by the encoding speed. Bacteria propagates through the system, using intermediate nanodevices to relay the information in the case where the destination is not in the range of the transmitter, until they arrive to the destination.The relaying process was introduced in Section 1.2.3.

A time out (T_{out}) can be used to avoid having lost bacteria swimming in the medium. In an ideal medium, where there are no interferences of attractant, bacteria can not be lost, but even in that case, they may need a long time to reach the receiver. In a real situation, if a bacterium swims far from the correct receiver during its random walk, it can be attracted by another nanodevice, which will cause the bacterium to get lost. For that reason, it is interesting to consider a time out even if we do not consider interferences between attractants.

We use the same architecture of the synchronous scheme, introduced in Section 4.2. These are the assumptions:

- A constant flux of data is sent from the transmitter to the final destination
- A 3D environment is considered
- All nanodevices are arranged in a perfect grid
- The steady state is assumed for the attractant concentration
- The interferences between attractants are not considered
- The possible saturation in intermediate nodes is not considered

A graphical representation of the timing of the system is shown in Fig. 5.1, where two bacteria are sent, relayed two times and finally received.

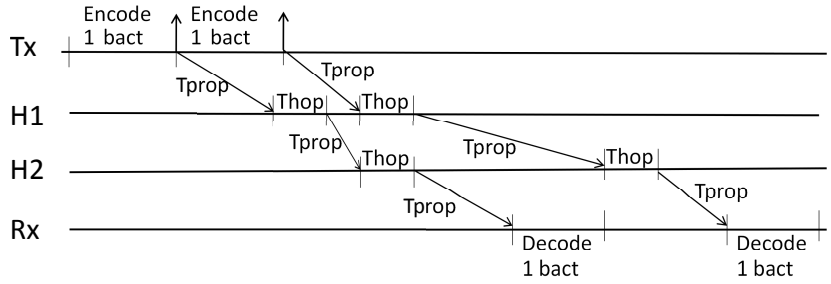


Figure 5.1: Graphical representation of the asynchronous scheme

5.3 Point to Point Scenario

In this section the delay and the capacity for the point-to-point scenario are analyzed. This is a specific case of a multi-hop communication but it allows to start a simple analysis and study its behavior.

5.3.1 Capacity

The capacity of this communication scheme is limited by the information that can be sent per unit of time by one nanodevice, and it is reduced by the bacteria loss probability and by the mutations in the plasmid. It can be expressed as follows:

$$C_{ch}(B) = C_{base} \cdot r_{efectiveEncode}(B) \cdot p_{Rx} \tag{5.1}$$

where p_{Rx} is the probability of receiving a bacterium and $r_{efectiveEncode}(B)$ is the effective encoding time taking into account that for each bacteria, the transfer and routing region have to be encoded independently of how many base pairs are encoded. The optimal amount of information to be encoded is reached when the all the base pairs contained in a plasmid are used, since in that case the ratio of useful information is maximum.

$$r_{efectiveEncode}(B) = \frac{\text{bits}}{\text{time}} = \frac{B}{\frac{B}{r_{enc}} + \frac{A}{r_{enc}}} = r_{enc}\frac{B}{B+A} \tag{5.2}$$

where A is the size of the active region. We apply the following values: $r_{enc} = 1400bp/s$, $B = 1550kbb$ and $A = 50kbps$, it gives an effective rate of $r_{effective} = 0.969 \cdot r_{encode} = 1356bp/s$.

The probability of receiving a bacterium is slightly different in this communication scheme, since there is no time slot, the loss probability is given by the mutations in the active zone that can modify the behavior of the bacteria and by the time out, if it is used.

$$p_{Rx} = 1 - p_{Tout} - p_{errBact} \tag{5.3}$$

where $p_{errBact}$ was calculated in the previous chapter, in (4.9). $p_{Tout} = arr(T_{out})$, since $arr(t)$ describes the distribution of the time needed by a bacterium to reach the receiver, as calculated in Chapter 3. If no time out is used, p_{Tout} corresponds directly to the loss probability, since $p_{Tout} = arr(\infty) = P_{Bactloss}$.

5.3.2 Delay

In that scheme, the delay is a random variable, since the nanodevices are not synchronized and the propagation time is also a random variable. In the point-to-point scenario, the delay has to take into account the time needed to encode the bacterium, which is constant, the time that the bacterium needs to reach the receiver, which is a random contribution, and the time needed to decode the information, which is constant as well. Mathematically it can be written as follows:

$$delay = t_{tx} + t_{prop} + t_{rx} \tag{5.4}$$

where t_{tx} is the time needed to encode one bacterium, as described in (4.2) and t_{rx} is the time needed to decode one bacterium, as described in (4.3), (both of them constants). t_{prop} is the random variable that describes the propagation of the bacterium.

The Probability Density Function (pdf) that describes the delay of the system can be writted as follows:

$$f_{delay}(t) = f_{arr}\left(t + t_{tx} + t_{rx}\right) \tag{5.5}$$

where f_{arr} is the pdf that describes the time needed by a bacterium to reach the receiver, introduced in Chapter 3. An example of this function is shown in Fig. 5.2

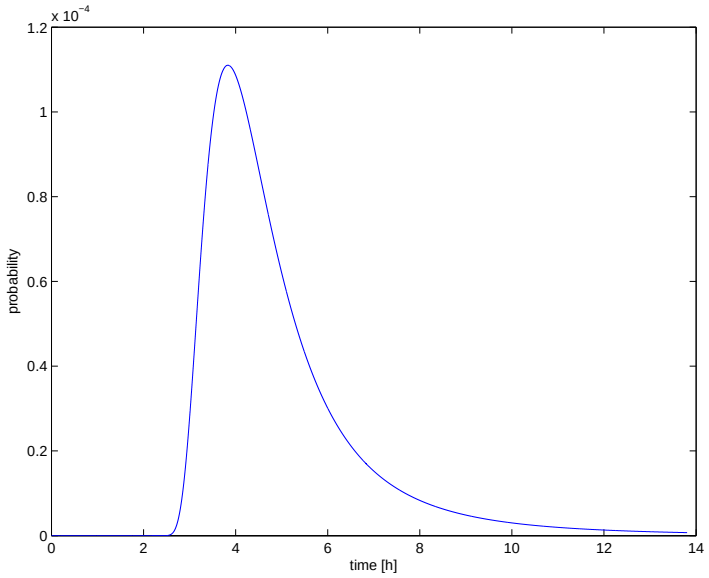


Figure 5.2: Probability Density Function (pdf) of the delay when the distance is 10 mm

5.3.3 Numerical Results

In Table 5.1, the capacity of the system, the average delay, the variance of the delay and the percentile 99 of the delay are shown per different T_{out} and different distances between nanodevices. The capacity is higher when no time out is used, but the reason for considering it was explained in the introduction of the chapter.

It can be observed that for a small distances, the time out does not affect the system. This is because the propagation time is short at small distances. For higher distances, it can be observed that when a time out is used, the capacity decreases but the delay is smaller and has less variance.

		Tout = 5h	Tout = 8h	No Tout
5 mm	Capacity	2712 kbps	2712 kbps	2712 kbps
	P_{Rx}	1	1	1
	Average delay	144 min	144 min	144 min
	Variance	1.76 min	1.78 min	1.79 min
	Percentile 99	181 min	181 min	181 min
8 mm	Capacity	2681 kbps	2709 kbps	2712 kbps
	P_{Rx}	0.989	0.998	1
	Average delay	208 min	211 min	213 min
	Variance	32 min	42 min	67 min
	Percentile 99	369 min	414 min	435 min
10 mm	Capacity	2395 kbps	2618 kbps	2712 kbps
	P_{Rx}	0.883	0.966	1
	Average delay	4.54 h	4.85 h	5.19 h
	Variance	0.91 h	1.91 h	7.26 h
	Percentile 99	6.9 h	9.33 h	14.75 h
12 mm	Capacity	1659 kbps	2293 kbps	2712 kbps
	P_{Rx}	0.611	0.846	1
	Average delay	5.35 h	6.16 h	7.69 h
	Variance	0.816 h	2.47 h	33.1 h
	Percentile 99	7 h	9.85 h	28.1 h

Table 5.1: Numerical results for the asynchronous system

5.4 Multi Hop Scenario

In the asynchronous communication system, the multi-hop scenario is the replication of the point-to-point scenario in each hop of the system. The difference is that the information is already encoded and only the attractant the bacterium is sensitive to (physical address) has to be changed. When a nanodevice receives a bacterium, first, it decapsulates the plasmid, it checks the network address (the address of the destination), and if it does not match, it modifies the physical address and it relays again the bacterium.

5.4.1 Capacity

The expression of the capacity is the same presented in the point-to-point scenario (5.1) but, the probability of receiving the bacterium takes into account all the hops. That probability can be written as follows:

$$p_{Rx} = (1 - p_{Tout} - p_{errBact})^S \tag{5.6}$$

where S is the number of hops.

5.4.2 Delay

The delay is the sum of the time needed to encode a bacterium, the multiple times needed to relay it, the time needed to decode it, which are all constants, and the time needed to propagate during all hops, which is a random variable.

The delay can be written as:

$$delay = t_{tx} + t_{prop}(S) + (S-1) \cdot t_{hop} + t_{rx} \tag{5.7}$$

t_{hop} is the time needed to read the network address and encode the following physical address, which was described in (4.12). $t_{prop}(S)$ is the random variable that describes the summation of the S hops. Giving the fact that the propagations in the hops are independent, the propagation time can be calculated as the S times convolution of the the pdf that describes the time needed by a bacterium to reach the receiver, introduced in Chapter 3.

$$f_{prop}(t, S) = f_{arr}(t) * f_{arr}(t) * \cdots * f_{arr}(t) \tag{5.8}$$

The pdf that describes the delay of the system can be written as follows:

$$f_{delay}(t) = f_{prop}\left(t + t_{tx} + (S-1) \cdot t_{hop} + t_{rx}\right) \tag{5.9}$$

5.4.3 Numerical Results

The capacity and the average delay are calculated in this subsection. We use a distance of 10 mm for the analysis and different values for the time out: 5 h, 8 h and no time out.

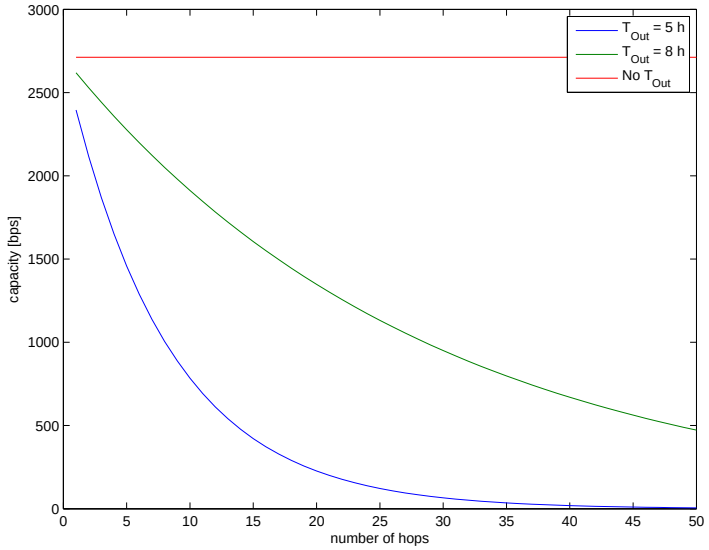


Figure 5.3: Capacity of the asynchronous system as a function of the number of hops using different T_{Out}

In Fig. 5.3, the capacity of the system is obtained. As expected, in an ideal system and without time out, the capacity is constant and determined by the encoding speed of the nanodevice, since the effect of the mutations in the active zone is negligible. When a time out is used, the capacity decreases exponentially as the number of hops increases. This is because the bacterium has to reach all nanodevices and the information is lost if the bacterium is lost in any of the hops.

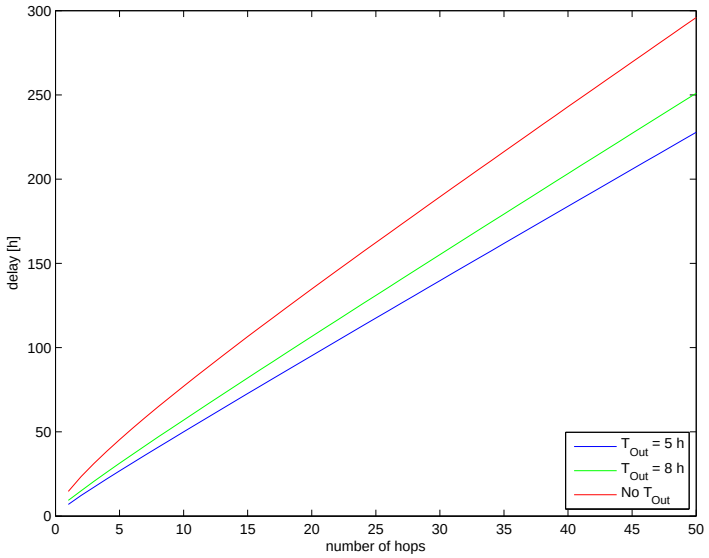


Figure 5.4: Average delay of the asynchronous system as a function of the number of hops using different T_{Out}

In Fig. 5.4, the average delay is obtained. We observe that the delay is higher as the time out increases, but the difference is small due to all the constant contributions that affect the

delay. On the contrary, the variance shows important differences in function of the value for the time out used. For example, for 50 hops, when $T_{Out} = 5h$, the variance is 46 h; when $T_{Out} = 8h$, the variance is 96 h; and when no T_{Out} is used, the variance is 311 h.

5.5 Conclusions

In this chapter, we introduced a new communication scheme for a system based on flagellated bacteria. In contrast with the scheme presented in the previous chapter, the system presented here is asynchronous. For this scheme, we calculate the capacity and the delay, first in a point-to-point scenario and finally in a multi-hop scenario.

In the analysis of the system, a time out can be used or not, but it is interesting to consider a time out since a real network will have interference between attractants that will make some bacteria swim away from the receiver and get lost. In that case, the capacity decreases exponentially as the number of hops increases, which gives a poor performance in long range transmissions, where the number of hops is high. The delay of the system is reasonable (for a system based on flagellated bacteria), but the variance of the delay is very high, which makes impossible to predict the order in which the packets are received.

Chapter 6

Comparison of the Syncronous and Asynchronous Schemes

6.1 Introduction

In the previous two chapters, two different communication schemes based on flagellated bacteria, the first synchronous and the second asynchronous, were introduced. For each communication scheme, we studied the capacity and the delay, and then we obtained some numerical results. The objective of this chapter is to compare both systems and to discuss the advantages and disadvantages of each scheme.

This chapter is organized as follows. In Section 6.2, the delay of both communication scheme is compared. The comparison of the capacity is done in Section 6.3. Finally, we conclude the chapter in Section 6.4.

6.2 System Delay Comparison

In Fig. 6.1, the delay for the synchronous scheme and the average delay for the asynchronous scheme are shown. We used a distance of 10 mm between nanodevices and a releasing rate of attractant of 10^{-11} mols/s. In the synchronous scheme, the delay is calculated for both relaying options, and in the asynchronous scheme, the delay is calculated using different values for the time out.

It can be observed that there is not an important difference in the delay for both communication schemes. On the one hand, the delay in the synchronous system is constant and known, which means that the receiver knows when it has to receive the information and that the information reaches the receiver in the correct order. On the other hand, the average delay of the asynchronous system is smaller, but the delay is a random variable and it has

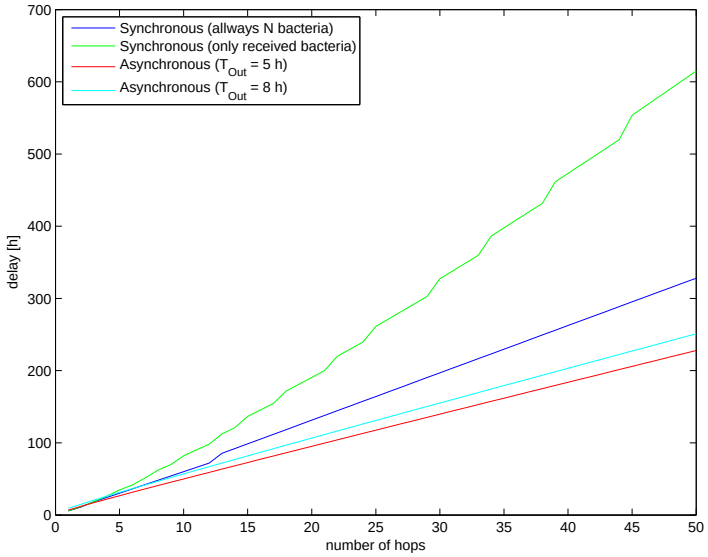


Figure 6.1: Comparison of the delay in the synchronous and the average delay in the asynchronous system as a function of the number using different options

high variance, which means that the information can reach the receiver at any moment and with no fixed order.

The delay in both cases is very high. This is something inherent to molecular nanocommunications [29], and the systems based on flagellated bacteria are not an exception. There are two reasons for the high delay of this communication paradigm. First of all, the time needed to process the DNA and to copy the message into the bacterium is very high; second, the propagation time of bacteria is very high. The extremely high delay forces to develop new protocols highly delay-tolerant.

6.3 System Capacity Comparison

In Fig. 6.2, the capacity of the system is shown for the synchronous and the asynchronous scheme, where we used a distance of 10 mm between nanodevices and a releasing rate of attractant of 10^{-11} mols/s. We used both relaying options for the synchronous scheme and different values for the time out for the asynchronous scheme. It has to be remarked that the case where no time out is used is not shown, because, as discussed in the previous chapter, in a real system there will be always a loss probability due to the attractant interferences due to other nanodevices.

We observe that the capacity of the asynchronous scheme is much higher than the capacity of the synchronous scheme for a small number of hops. This is because the bacterium does not need to complete a long path in order to reach the final destination, so it does not benefit from all the advantages given by the synchronous system.

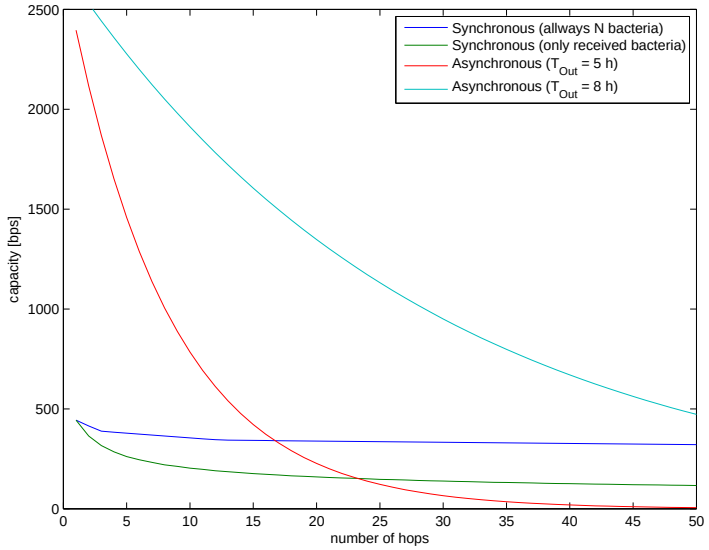


Figure 6.2: Comparison of the capacity in the synchronous and asynchronous system as a function of the number using different options

When the number of hops increases, the capacity of the asynchronous scheme decreases much faster than in the synchronous scheme, this is because the reliability given by the synchronous scheme increases the value of the capacity. On the contrary, the capacity of the synchronous scheme remains almost constant as the number of hops increases. This means that when the number of hops is high, a higher capacity can be achieved using the synchronous scheme.

The capacity in both schemes is around hundreds of bits per second. This is an extremely high capacity for molecular nanocommunications, where usual values are around tens of bits per second [29]. This is because in the system based on flagellated bacteria, the information is sent encoded in DNA and carried by bacteria, which correspond to sending huge information packets. In other molecular systems, the information is carried by modulating the concentration of some particles.

6.4 Conclusions

In this chapter we compared the delay and the capacity of the synchronous and asynchronous schemes. We observed that the delay is quite high in both schemes, but in the asynchronous scheme it is a random variable which has a high variance. We also observed that the capacity is higher in the asynchronous system when the number of hops is small, while decreases slower in the synchronous system. In both cases we obtain higher capacity values compared to other molecular nanocommunications.

Taking into account the results of this chapter, we conclude that the choice between the synchronous or the asynchronous schemes as the most suitable strongly depends on the

particular scenarios in which we deploy our bacteria-based communication scheme.

The synchronous system is appropriate when:

- The order of the information is important
- The delay has to be known and constant
- The number of hops is high

The asynchronous system is appropriate when:

- Neither the order nor the exact moment when a packet is received is important
- The number of hops is small

Chapter 7

Open Issues and Conclusions

7.1 Open Issues

The field of nanotechnology in general, and the research on Wireless NanoSensor Network (WNSN) in particular, are attracting the attention of scientists and engineers from different backgrounds. Interesting results have been obtained in the nanonetworking field, but at the same time, new challenges have arisen. There is still a long way to go before the first nanonetwork can be successfully realized.

Regarding this work, there are some issues that have to be solved. The addressing and the routing are the next challenges to be addressed, which correspond to the study of how the network addresses will be assigned, how the different attractants will distributed and how the routing tables will be created. Another interesting topic that should be addressed is to propose new communication schemes which combine the advantages of the synchronous and asynchronous schemes introduced in this work.

Additionally, some other work must be done in complementary fields, such as biology, in order to make possible the use of flagellated bacteria in communication at the nanoscale. Regarding the DNA Processing Unit (DPU), the scientists and the engineering community have to perform a deeper investigation in computing using DeoxyriboNucleic Acid (DNA) and also in how to synthesize and sequence DNA at the nanoscale. Another important direction of research is how the use of plasmids with arbitrary DNA affects the behavior of bacteria. Finally, another interesting direction that needs further research efforts regards how the attractant gradient can be created, taking into account the small dimensions of the nanodevice and the interferences between attractants released by other nanodevices. Once all these problems are studied, we envision that the use of flagellated bacteria will be one of the most useful and exciting approaches for communication at the nanoscale.

7.2 Conclusions

This thesis continues the work done in the Broadband Wireless Networking Lab (Georgia Institute of Technology) (BWN Lab) concerning the use of flagellated bacteria in communications at the nanoscale. The main purpose of this work has been to provide analytical models of the different concepts of the bacteria-based communication paradigm, and, finally, to obtain analytical results for the capacity and the delay, by stemming from the concepts introduced in previous work.

Firstly, we studied how the mutations and errors in the DNA affect the reliability of the flagellated bacteria system. During a multi-hop communication, the plasmid is copied multiple times and each time the plasmid can be affected by some mutations. We described mathematically the probability of error of a base pair after multiple possible mutations. Moreover, we verified this calculation by means of simulations. Then, we analyzed how these mutations affects the capacity, by providing an expression valid for any mutation rate and any number of hops. Finally, we concluded that by taking into account the mutation rate found in nature and if no other source of mutations are introduced, the effect of the mutations in the capacity is negligible.

Secondly, we studied analytically the propagation of bacteria. In the previous work, the propagation was only simulated, and the data obtained in the simulations was used to evaluate the capacity of the system. On the contrary, we obtained an expression for the propagation delay of bacteria through the study of the movement of the bacteria as a biased random walk and by applying the diffusion theory. The expression we provided does not have close-form solution, but numerical methods can be used in order to solve it.

After that, we introduced two different communication schemes, the first synchronous and the second asynchronous. For both systems, we studied the capacity and the delay, by using the propagation model and the mutation study introduced in the first part of the thesis. Some numerical results have been given in order to be able to compare both systems. We determined that both schemes present some advantages that make each scheme more suitable depending on the situation, for example, a situation where the number of hops is really low, the asynchronous scheme is more suitable. We envision that a good communication scheme for a flagellated bacteria communication will have some characteristics of both schemes.

We foresee this work to be an important step foward in the flagellated bacteria communication research, but there is still a lot of work that has to be done before a system based on flagellated bacteria can be used.

Appendix A

Acronyms

ATP Adenosine TriPhosphate

BBN Bacteria-Based Nanonetwork

BC Bacterial Conjugation

bp base pair

BWN Lab Broadband Wireless Networking Lab (Georgia Institute of Technology)

cdf Cumulative Distribution Function

DNA DeoxyriboNucleic Acid

DPU DNA Processing Unit

E. Coli Escherichia Coli

FDM Finite Difference Method

PDE Partial Differential Equation

pdf Probability Density Function

QS Quorum Sensing

RNA Ribonucleic acid

ssDNA single-stranded DNA

TTL Time To Live

WSN Wireless Sensor Network

WNSN Wireless NanoSensor Network

Bibliography

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer Networks*, 38(4):393 – 422, 2002.

[2] Ian F. Akyildiz, Fernando Brunetti, and Cristina Blázquez. Nanonetworks: A new communication paradigm. *Computer Networks*, 52(12):2260 – 2279, 2008.

[3] Norio Taniguchi. On the basic concept of nanotechnology. *Proc. Intl. Conf. Prod. Eng. Tokyo*, 2:18–23, 1974.

[4] Christofer Hierold, Alain Jungen, Christoph Stampfer, and Thomas Helbling. Nano electromechanical sensors based on carbon nanotubes. *Sensors and Actuators A: Physical*, 136(1):51 – 61, 2007.

[5] Jordi Riu, Alicia Maroto, and F. Xavier Rius. Nanosensors in environmental analysis. *Talanta*, 69(2):288 – 301, 2006. ¡ce:title¿1st Swift-WFD workshop on validation of Robustness of sensors and bioassays for Screening Pollutants¡/ce:title¿ ¡ce:subtitle¿1st SWIFT-WFD 2004¡/ce:subtitle¿ ¡xocs:full-name¿1st Swift-WFD Workshop on Validation of Robustness of Sensors and Bioassays for Screening Pollutants¡/xocs:full-name¿.

[6] Yaping Dan, Ye Lu, Nicholas J. Kybert, Zhengtang Luo, and A. T. Charlie Johnson. Intrinsic response of graphene vapor sensors. *Nano Letters*, 9(4):1472–1475, 2009.

[7] Bo Zhang and Tianhong Cui. An ultrasensitive and low-cost graphene sensor based on layer-by-layer nano self-assembly. *Applied Physics Letters*, 98, 2001.

[8] Gang Peng, Ulrike Tisch, and Hossam Haick. Detection of nonpolar molecules by means of carrier scattering in random networks of carbon nanotubes: Toward diagnosis of diseases via breath samples. *Nano Letters*, 9(4):1362–1368, 2009.

[9] Padmavathy Tallury, Astha Malhotra, Logan M Byrne, and Swadeshmukul Santra. Nanobioimaging and sensing of infectious diseases. *Advanced Drug Delivery Reviews*, 62(4-5):424 – 437, 2010.

[10] F. Schedin, A. K. Geim, S. V. Morozov, E. W. Hill, P. Blake, M. I. Katsnelson, and K. S. Novoselov. Detection of individual gas molecules adsorbed on graphene. *Nat Mater*, 6:652–655, 2007.

[11] C. Roman, F. Ciontu, and B. Courtois. Single molecule detection and macromolecular weighting using an all-carbon-nanotube nanoelectromechanical sensor. *4th IEEE Conference on Nanotechnology*, pages 263–266, 2004.

[12] A.J. Cadby, R. Dean, C. Elliott, R.A.L. Jones, A.M. Fox, and D.G. Lidzey. Imaging the fluorescence decay lifetime of a conjugated-polymer blend by using a scanning near-field optical microscope. *Advanced Materials*, 19(1):107–111, 2007.

[13] Nicholas P. Reynolds, Stefan Janusz, Maryana Escalante-Marun, John Timney, Robert E. Ducker, John D. Olsen, Cees Otto, Vinod Subramaniam, Graham J. Leggett, and C. Neil Hunter. Directed formation of micro- and nanoscale patterns of functional light-harvesting lh2 complexes. *Journal of the American Chemical Society*, 129(47):14625–14631, 2007.

[14] Johann Elbaz, Oleg Lioubashevski, Fuan Wang, Francoise Remacle, Raphael D. Levine, and Itamar Willner. Dna computing circuits using libraries of dnazyme subunits. *Nat Nano*, 5:417–422, 2010.

[15] Junghuei Chen, Russell Deaton, and Yu-Zhen Wang. A dna-based memory with 'in vitro' learning and associative recall. *Natural Computing*, 4:83–101, 2005. 10.1007/s11047-004-4002-3.

[16] Vega genome browser 45: Homo sapiens.

[17] Richard F. Taylor. *Handbook of chemical and biological sensors*. IOP, 1996.

[18] Chanda Ranjit Yonzon, Douglas A. Stuart, Xiaoyu Zhang, Adam D. McFarland, Christy L. Haynes, and Richard P. Van Duyne. Towards advanced chemical and biological nanosensorsan overview. *Talanta*, 67(3):438 – 448, 2005. Nanoscience and Nanotechnology.

[19] Chunyu Li, Erik T. Thostenson, and Tsu-Wei Chou. Sensors and actuators based on carbon nanotubes and their composites: A review. *Composites Science and Technology*, 68(6):1227 – 1249, 2008.

[20] J. Matthew Dubach, Daniel I. Harjes, and Heather A. Clark. Fluorescent ion-selective nanosensors for intracellular analysis with improved lifetime and size. *Nano Letters*, 7(6):1827–1831, 2007.

[21] Jianping Li, Tuzhi Peng, and Yuqiang Peng. A cholesterol biosensor based on entrapment of cholesterol oxidase in a silicic sol-gel matrix at a prussian blue modified electrode. *Electroanalysis*, 15(12):1031–1037, 2003.

[22] Ibtisam E. and Tothill. Biosensors for cancer markers diagnosis. *Seminars in Cell & Developmental Biology*, 20(1):55 – 62, 2009. A Special Edition on Biosensors and Development of Pigment Cells and Pigment Patterns.

[23] Martin Heil and Juan Carlos Silva Bueno. Within-plant signaling by volatiles leads to induction and priming of an indirect plant defense in nature. *Proceedings of the National Academy of Sciences*, 104(13):5467–5472, 2007.

[24] CM Pieterse and M Dicke. Plant interactions with microbes and insects: from molecular mechanisms to ecology. *Trends in plant science*, 12:564–569, 2007.

[25] Ian F Akyildiz and Josep Miquel Jornet. The internet of nano-things. *Ieee Wireless Communications*, 17(December):58–63, 2010.

[26] Ian F. Akyildiz and Josep Miquel Jornet. Electromagnetic wireless nanosensor networks. *Nano Communication Networks*, 1(1):3 – 19, 2010.

[27] Llus Parcerisa Gin and Ian F. Akyildiz. Molecular communication options for long range nanonetworks. *Computer Networks*, 53(16):2753 – 2766, 2009.

[28] T. Suda, M. Moore, T. Nakano, R. Egashira, and A. Enomoto. Exploratory research on molecular communication between nanomachines. *Genetic and Evolutionary Computation Conference (GECCO)*, Late Breaking Papers, 2005.

[29] M. Pierobon and I.F. Akyildiz. A physical end-to-end model for molecular communication in nanonetworks. *Selected Areas in Communications, IEEE Journal on*, 28(4):602 –611, may 2010.

[30] M. Pierobon and I.F. Akyildiz. Diffusion-based noise analysis for molecular communication in nanonetworks. *Signal Processing, IEEE Transactions on*, 59(6):2532 –2547, june 2011.

[31] M. Pierobon and I.F. Akyildiz. Noise analysis in ligand-binding reception for molecular communication in nanonetworks. *Signal Processing, IEEE Transactions on*, 59(9):4168 –4182, sept. 2011.

[32] M. Gregori and I.F. Akyildiz. A new nanonetwork architecture using flagellated bacteria and catalytic nanomotors. *Selected Areas in Communications, IEEE Journal on*, 28(4):612 –619, may 2010.

[33] Maria Gregori, Ignacio Llatser, Albert Cabellos-Aparicio, and Eduard Alarcón. Physical channel characterization for medium-range nanonetworks using catalytic nanomotors. *Nano Communication Networks*, 1(2):102 – 107, 2010.

[34] Josep Miquel Jornet and Ian F. Akyildiz. Graphene-based nano-antennas for electromagnetic nanocommunications in the terahertz band. In *Antennas and Propagation (EuCAP), 2010 Proceedings of the Fourth European Conference on*, pages 1 –5, april 2010.

[35] J.M. Jornet and I.F. Akyildiz. Channel capacity of electromagnetic nanonetworks in the terahertz band. In *Communications (ICC), 2010 IEEE International Conference on*, pages 1 –6, may 2010.

[36] Maria Gregori, Ignacio Llatser, Albert Cabellos-Aparicio, and Eduard Alarcón. Physical channel characterization for medium-range nanonetworks using flagellated bacteria. *Computer Networks*, 55(3):779 – 791, 2011.

[37] Luis C. Cobo and Ian F. Akyildiz. Bacteria-based communication in nanonetworks. *Nano Communication Networks*, 1(4):244 – 256, 2010.

[38] Frederick R. Blattner, Guy Plunkett, Craig A. Bloch, Nicole T. Perna, Valerie Burland, Monica Riley, Julio Collado-Vides, Jeremy D. Glasner, Christopher K. Rode, George F. Mayhew, Jason Gregor, Nelson Wayne Davis, Heather A. Kirkpatrick, Michael A. Goeden, Debra J. Rose, Bob Mau, and Ying Shao. The complete genome sequence of escherichia coli k-12. *Science*, 277(5331):1453–1462, 1997.

[39] T Ogura and S Hiraga. Mini-f plasmid genes that couple host cell division to plasmid proliferation. *Proceedings of the National Academy of Sciences*, 80(15):4784–4788, 1983.

[40] J Adler. Chemotaxis in bacteria. *Harvey Lect.*, 72:195–230, 1978.

[41] Gerald L. Hazelbauer, Robert E. Mesibov, and Julius Adler. Escherichia coli mutants defective in chemotaxis towards specific chemicals. *Proceedings of the National Academy of Sciences*, 64(4):1300–1307, 1969.

[42] Jesse Stricker, Scott Cookson, Matthew R. Bennett, William H. Mather, Lev S. Tsimring, and Jeff Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456:516–519, 11 2008.

[43] Octavio AU Tsimring Lev AU Hasty Jeff AU Danino, Tal AU Mondragon-Palomino. A synchronized quorum of genetic clocks. *Nature*, 463:326–220, 01 2010.

[44] Yaakov Benenson, Tamar Paz-Elizur, Rivka Adar, Ehud Keinan, Zvi Livneh, and Ehud Shapiro. Programmable and autonomous computing machine made of biomolecules. *Nature*, 414:430–434, 2001.

[45] Paul Wilhelm Karl Rothemund. A dna and restriction enzyme implementation of turing machines. In *Dimacs series in discrete mathematics and theoretical computer science*, pages 75–119.

[46] Turlough M. Finan, Stefan Weidner, Kim Wong, Jens Buhrmester, Patrick Chain, Frank J. Vorhlter, Ismael Hernandez-Lucas, Anke Becker, Alison Cowie, Jrme Gouzy, Brian Golding, and Alfred Phler. The complete sequence of the 1,683-kb psymb megaplasmid from the n2-fixing endosymbiont sinorhizobium meliloti. *Proceedings of the National Academy of Sciences*, 98(17):9889–9894, 2001.

[47] L S Frost, K Ippen-Ihler, and R A Skurray. Analysis of the sequence and gene products of the transfer region of the f sex factor. *Microbiol. Mol. Biol. Rev.*, 58(2):162–210, 1994.

[48] Andrew Fire, SiQun Xu, Mary K. Montgomery, Steven A. Kostas, Samuel E. Driver, and Craig C. Mello. Potent and specific genetic interference by double-stranded rna in caenorhabditis elegans. *Nature*, 391:806–811, 1998.

[49] Hanna Engelberg-Kulka, Boaz Sat, Myriam Reches, Shahar Amitai, and Ronen Hazan. Bacterial programmed cell death systems as targets for antibiotics. *Trends in Microbiology*, 12(2):66 – 71, 2004.

[50] Stephen Cooper and Charles E. Helmstetter. Chromosome replication and the division cycle of escherichia coli br. *Journal of Molecular Biology*, 31(3):519 – 540, 1968.

[51] Chandler Fulton. Continuous chromosome transfer in escherichia coli. *Genetics*, 52(1):55–74, 1965.

[52] Robert C. King, William D. Stansfield, and Pamela Khipple Mulligan. *A dictionary of genetics*. Oxford University Press, 2006.

[53] A D Grossman. Genetic networks controlling the initiation of sporulation and the development of genetic competence in bacillus subtilis. *Annual Review of Genetics*, 29(1):477–508, 1995.

[54] M Achtman, N Kennedy, and R Skurray. Cell–cell interactions in conjugating escherichia coli: role of trat protein in surface exclusion. *Proceedings of the National Academy of Sciences*, 74(11):5104–5108, 1977.

[55] Christoph Kessler and Vicentiu Manta. Specificity of restriction endonucleases and dna modification methyltransferases a review (edition 3). *Gene*, 92(12):1 – 240, 1990.

[56] Changbei Ma, Zhiwen Tang, Kemin Wang, Weihong Tan, Xiaohai Yang andWei Li, Zhaohui Li, and Xiaoyuan Lv. Real-time monitoring of restriction endonuclease activity using molecular beacon. *Analytical Biochemistry*, 363(2):294–296, 2007.

[57] John W. Drake, Brian Charlesworth, Deborah Charlesworth, and James F. Crow. Rates of spontaneous mutation. *Genetics*, 148(4):1667–1686, 1998.

[58] Michael J. Czar, J. Christopher Anderson, Joel S. Bader, and Jean Peccoud. Gene synthesis demystified. *Trends in Biotechnology*, 27(2):63 – 72, 2009.

[59] Jay Shendure and Hanlee Ji. Next-generation dna sequencing. *Nature*, 26:1135–1145, 2008.

[60] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 1948.

[61] Steven M. Block, Jeffrey E. Segall, and Howard C. Berg. Impulse responses in bacterial chemotaxis. *Cell*, 31(1):215 – 226, 1982.

[62] Ravi Jasuja, Yu-Lin, David R. Trentham, and Shahid Khan. Response tuning in bacterial chemotaxis. *Proceedings of the National Academy of Sciences*, 96(20):11346–11351, 1999.

[63] Victor Sourjik and Howard C. Berg. Receptor sensitivity in bacterial chemotaxis. *Proceedings of the National Academy of Sciences*, 99(1):123–127, 2002.

[64] Mark J. Schnitzer. Theory of continuum random walks and application to chemotaxis. *Phys. Rev. E*, 48(4):2553–2568, Oct 1993.

[65] M. V. Smoluchowski. Über Brownsche Molekularbewegung unter Einwirkung äußerer Kräfte und deren Zusammenhang mit der verallgemeinerten Diffusionsgleichung. *Ann. Phys.*, (24):1103–1112, 1916.

[66] Stanley J. Farlow. *Partial Differential Equations for Scientists and Engineers*. J. Wiley, 1993.

[67] William H. Bossert and Edward O. Wilson. The analysis of olfactory communication among animals. *Journal of Theoretical Biology*, 5(3):443 – 469, 1963.

[68] Robert Mesibov, George W. Ordal, and Julius Adler. The range of attractant concentrations for bacterial chemotaxis and the threshold and size of response over this range. *The Journal of General Physiology*, 62(2):203–223, 1973.

[69] Frank Spitzer. *Principles of Random Walk*. Springer, 2001.