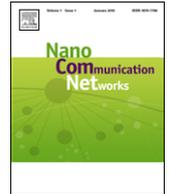




Contents lists available at ScienceDirect

Nano Communication Networks

journal homepage: www.elsevier.com/locate/nanocomnet

Bacteria-based communication in nanonetworks

Luis C. Cobo*, Ian F. Akyildiz

Broadband Wireless Networking Lab, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332–0250, United States

ARTICLE INFO

Article history:

Received 24 November 2010

Accepted 11 December 2010

Available online 21 December 2010

Keywords:

Bacteria-based nanonetwork

ABSTRACT

This paper describes a Bacteria-based Nanonetwork for communication between eukaryotic cell sized nano devices. The communication is achieved by the exchange of DNA molecules which are transported by bacteria guided by chemotaxis. First, the modules of the network are described and all the biological phenomena that support the basic communication steps are explained in detail. Then an analytical model is developed to assess the communication range and the network performance in terms of capacity and end-to-end delay by considering the available information about the biological mechanisms used. As there are no appropriate estimates of the propagation delay introduced by bacterial chemotaxis, our newly developed simulator is introduced which helps us to obtain the statistics on bacteria propagation. Finally, by combining the analytical model with the simulation results, a network performance in terms of end-to-end delay, capacity and end-to-end throughput is obtained which is 4 orders of magnitude higher than the other molecular communication approaches.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Nanotechnology is making it possible to develop new materials and devices with atomic precision. Nano-scale devices, such as nanosensors, offer revolutionary health-care, industrial and military applications. However, the small size of such devices limits the capabilities of any single one of them. Nanonetworking studies how to take full advantage of their power by enabling them to coordinate and communicate. Recent research has underlined the need for new communication paradigms to interconnect these devices, since traditional techniques are not directly applicable at this scale [4]. The approaches for nano-communications can be divided into two groups: electromagnetic (EM) and molecular. On the one hand, EM approaches [5] aim at harnessing the unique features of novel materials such as graphene or carbon nanotubes to implement antennas and transmitters adequate for the nano-scale. On the other hand, molecular communication

techniques [32,22] draw inspiration from nature to encode information in molecules that are exchanged between transmitter and receiver.

Molecular communication is specifically suited for application in biological nano-sensors, since it is already in use by naturally-occurring nano-devices, namely cells. The latest advances in synthetic biology [20] suggest that the first fully-functional man-made nano-devices will be based on biological building blocks. These bio-hybrid devices will be obtained by a combination of selective breeding and genetic engineering over existing organisms, combining functionalities found in different species to obtain the desired behavior. The applications of these devices span from cancer diagnosis and treatment to the detection of chemical attacks.

Unfortunately, the molecular communication approaches proposed so far have important drawbacks, such as very low capacity or the need to deploy a complex infrastructure. One of the most interesting ideas to overcome these problems is the use of bacteria to carry messages encoded in DNA from transmitter to receiver [21]. This must not be confused with Quorum Sensing (QS), which has been thoroughly investigated by Bonnie Bassler [40]. QS is used by bacteria to recognize the presence of other

* Corresponding author.

E-mail addresses: luisca@gatech.edu (L.C. Cobo), ian@ece.gatech.edu (I.F. Akyildiz).

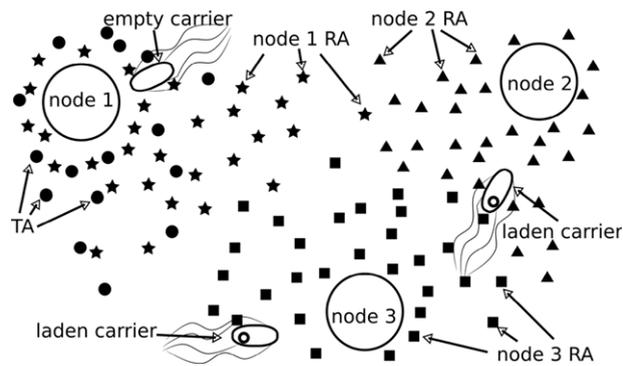


Fig. 1. A Bacteria-based Nanonetwork with 3 nodes, 2 messages in transit and an empty carrier. Laden carriers follow the specific RA that is emitted by their destination.

individuals from the same or similar species and to coordinate the behavior among different bacteria, but cannot be used to transmit arbitrary messages.

The use of DNA allows a higher throughput than other molecular approaches, given the high information density of DNA. At one bit per nm^3 [2], we could store up to 13 billion terabytes in the size of a typical laptop hard drive. However, the use of bacteria and DNA presents a series of challenges that have not been addressed yet.

In this paper, we introduce a Bacteria-based Nanonetwork (BN) that addresses the problems inherent to the use of bacteria and DNA for point to point communication. The paper is organized as follows: In Section 2, we describe the major components of the network, while in Section 3 we explain each step in the communication process. We then analyze the network in terms of end-to-end delay, capacity and communication range in Section 4. In Section 5, we describe our new bacteria simulator which is developed to estimate propagation delay. Using the previous mathematical analysis and the new simulator, we demonstrate in Section 6 the performance that can be achieved in a BN and summarize our conclusions in Section 7.

2. Nanonetwork architecture

The *Bacteria-based Nanonetwork* (BN) is composed of nodes and carriers, as shown in Fig. 1. Nodes communicate by exchanging DNA molecules. Carriers are the bacteria, which take a DNA molecule encoding the message from the transmitter, move towards the receiver, and then deliver the DNA molecule. The nodes and carriers are floating in a liquid medium with ample nutrients for them. In this section, we provide a detailed description of carriers and nodes.

2.1. Carriers: bacteria

The bacteria that carry information among the nodes are rod-shaped and around $2 \mu\text{m}$ long and $1 \mu\text{m}$ wide. This is the typical size of *Escherichia coli* [7], which is the reference organism for the design of our carriers.

We assume that it 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. A *biological part* is a set of genes that implement a certain

functionality in bacteria, such as chemotaxis or reproduction. There are many examples of teams who have succeeded in the task [37,38], and there already exists a freely accessible *Registry of Standard Biological Parts* with more than 12 thousand parts available.

In the BN, bacteria form a uniform population spanning the whole medium. They maintain a uniform population density by regulating their replication cycle [30] using Quorum Sensing. In this population, *mutations* can make the carriers deviate from their expected behavior. *E. coli*, e.g., has a mutation rate of 10^{-8} [12] per base pair (*bp*) per generation. With a total genome composed of $4.6 \cdot 10^6 \text{bp}$ [27], this means that mutations are rare, but will occur in a large population. In the future, we will address how to minimize them and mitigate their effects in the long-term performance of the system.

Bacteria are capable of self-motion in a liquid medium using flagella, long tail-like appendages that they rotate to propel themselves. *Chemotaxis* is the process by which bacteria sense, via chemoreceptors, the concentration gradient of specific particles and swim towards higher concentrations of beneficial particles called *attractants*.

E. coli can sense at least 12 different attractants using different chemoreceptors that cause independent chemotactic responses [3]. Additionally, there are *E. coli* mutants that present no response towards specific attractants [25]. This means that bacteria can be directed towards specific nodes that emit a given attractant. Different attractants may cause a chemotactic response of different strength in different bacteria, but we consider the responses homogeneous as, the mechanism being the same, homogeneity can be achieved simply by selective breeding of bacteria strains.

2.2. Nodes: Nano-machines

Nano-machines are devices with features in the nano-scale that are able to perform computing, sensing and/or actuation tasks [4]. We envision that the nodes will be bio-hybrid nano-machines ranging from 5 to $100 \mu\text{m}$ in diameter, which is the size range of most eukaryotic cells. Nodes contain a *DNA Processing Unit* (DPU), which will be able to encode an arbitrary strand of DNA. This DPU will result from research in autonomous DNA computing devices, such as DNA-based Turing Machines [6,14,34]. The

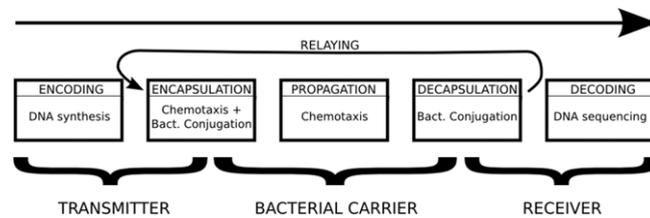


Fig. 2. Overview of the basic communications steps in a Bacteria-based Network.

nodes are considered static, but slight movements of the nodes do not affect the channel. Nodes can communicate directly with their neighboring nodes or can use other nodes as relays to reach nodes that are farther away using a multi-hop path. In the rest of this paper, the terms *node* and *nano-machine* are used interchangeably.

Nodes emit attractants, creating a concentration gradient of these particles, as described in Section 4.3, with the peak concentration at their own position. Every node transmits at a constant rate a specific attractant, which we call the *reception attractant (RA)*, that guides carriers delivering a message to the node. Additionally, when a node wants to transmit a message, it emits *transmission attractant (TA)*, which is common for all nodes and attracts nearby empty carriers.

Nodes are identified by a two-tier address system, where each node has a *physical address* and a unique *network address*. The addressing in these networks is an important and open research problem which needs to be investigated further and concrete solutions are needed. Each message in transit contains the network address of the destination node and the physical address of the next node on the path to destination. The physical address is related with the RA a node emits. Carriers perform chemotaxis towards the RA associated to the physical address in the message. The number of RAs is finite but, as shown in Section 4.3, the area of influence of a node is limited, so RAs can be reused in space. The network address is necessary to remove the ambiguity among distant nodes that use the same RA.

BNs relaying scheme is supported by an *RA-based routing table* in the DPU of each node, with network addresses as key and RAs as yield. The address assignment and routing table contents are interesting and open research problems. In particular, the lack of broadcast messages in BNs makes the problem much more challenging. When a transmitter wants to send a message to a specific destination, it will search for the corresponding reception attractant (RA) in the routing table. If the destination is in direct range of the transmitter, this RA will be the RA of the destination. If not, the RA will be the RA of the next hop in the route to the destination. This next hop will send the message to the following one and so on until the message reaches its destination.

3. Basic communication steps

In this section we describe the biological phenomena involved in the Bacteria-based Nanonetwork, which are summarized in Fig. 2.

3.1. Encoding

The DNA Processing Unit in the transmitter encodes the message to transfer in a double-stranded DNA molecule. DNA is a long molecule composed of two polymers of nucleotides, with each nucleotide from one polymer bonded to one in the other forming a *base pair* (bp). Each polymer runs in opposite directions, with the asymmetric ends of each polymer being called 5' and 3'. Each nucleotide contains one of four possible bases: adenine (A), cytosine (C), guanine (G) or thymine (T). The base in a nucleotide determines the base in the other nucleotide in the pair so that they are either AT or CG. Each bp thus can encode 2 bits, since there are 4 possible options for the base of one of the nucleotides of the pair, and the other base is determined by the first.

This DNA forms a *plasmid*, i.e., a circular DNA strand capable of self-replication and self-transfer, via *bacterial conjugation*, to new bacteria. The plasmid can be up to 1.6 mbp (mega base pairs) [16] long, and it is divided into 3 parts: the *transfer region*, the *routing region* and the *message region*. Transfer and routing regions constitute the *active section* of the plasmid, i.e., the section that is to be expressed or interpreted as biological instructions.

- (1) *The transfer region*: is present in typical plasmids as the F factor of *E. coli* and is 33 kbp (kilo base pairs) long [18]. This region contains the genes and structures necessary for self-replication and transmission of the plasmid.
- (2) *The routing region*: contains a set of genes that implement the behavioral differences between **empty** bacteria, with no data to deliver, and **laden** bacteria carrying a DNA message. These genes encode new proteins or inhibit genes in the bacterial DNA [17] for various purposes:

- **Deactivate chemotaxis towards transmitters.**
- **Activate chemotaxis towards the receiver**, using a specific protein that enables chemotaxis towards a specific RA. This protein is the *physical address* of the message.
- **Inhibit bacterial replication** [30]. To avoid the number of bacteria with the same message growing exponentially, since this would overload the receiver.
- **Enable programmed death** [15] **on timeout**. This is used to prevent delivery of messages that have accumulated a very long delay by making bacteria *suicide* a specific time after plasmid reception.

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

- (3) *The message region*: contains 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 poses 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, we disable the expression of the message by encoding it as *inactive DNA*, at the cost of reducing the available bits per base pair.

One way to inhibit DNA expression is to avoid promoter sequences. Promoters are specific DNA sequences where the RNA polymerase can bind to start the transcription. Promoters are varied in the DNA of a single organism, but they all share subsequences called *consensus promoters*. Avoiding these sequences prevents DNA expression.

3.2. Encapsulation

In this step, the plasmid generated in the encoding step is transferred to a carrier. First, the transmitter emits TA to attract empty carriers in the vicinity. The TA emission can happen simultaneously with the encoding step. When a bacterium is close enough to the transmitter, a copy of the plasmid is passed to the bacterium using *bacterial conjugation*. The transmitter first attaches to the carrier using a *pilus*, which will retract to get the carrier in physical contact with the transmitter. The membranes of the transmitter and bacterium connect and a single-stranded DNA molecule (ssDNA) is unwound from the plasmid and transferred to the carrier. After the ssDNA has been transferred, the 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, i.e., the transmitter. However, the remains of the plasmid in the donor can be discarded in our case.

Once the plasmid is transferred to the carrier, the *active section* of the plasmid is expressed, the bacterium regulates its behavior as detailed in Section 3.1, and the propagation step starts.

3.3. Propagation

In this step, the carrier swims from transmitter to receiver using bacterial chemotaxis. The delay is limited by the programmed death in the active part of the plasmid described in Section 3.1 by the TTL (Time To Live) gene. Messages not received after a certain time can be considered lost, as bacteria carrying them will suicide.

For a reliable TTL mechanism, we must consider that after a bacterium dies, its membrane breaks and its components are released to the medium. The naked plasmid floating in the medium could be incorporated by another carrier through a process called transformation, and the transformed bacterium could then finish the delivery. The ability to uptake naked DNA is called competence, and to avoid this problem we use non-competent bacteria [24].

There is no reason for bacteria to die while propagating in a controlled medium. However, if there is a probability of bacterial death in propagation, it can be considered part of the message *loss probability*.

3.4. Decapsulation

This step uses bacterial conjugation, as explained for the encapsulation step, with the carrier as donor and

the receiver as acceptor. After receiving the message, the receiver kills the bacterium, which would try to re-deliver the message otherwise.

A laden carrier might encounter an empty one and pass a copy of the carried plasmid to it, resulting in undesired message duplication. However, we can use surface exclusion [1], which is used by naturally-occurring plasmids to ensure they are not transferred to bacteria that already have a copy of the plasmid. For our purposes, this mechanism would be adapted so that bacterial conjugation can only happen between a node and a carrier.

Still, in case there are different nodes in the medium, each one emitting a different attractant, a bacterium might reach, by chance, the wrong node. The receiver determines if the message was for itself by checking the *physical address* in the message. If it does not match, the message is discarded, so the possibility of the message being relayed to the wrong node must be accounted for in the *loss probability*.

3.5. Relaying

Once the message is decapsulated, the receiver checks the destination *network address* in the message. If it matches the local network address, the message is decoded. If it does not, the RA in the plasmid is substituted by the RA in the local routing table that matches the destination address of the message and the plasmid is resent.

In Fig. 3 we can see an example of a network architecture using the Bacteria-based Nanonetwork. Only 6 RAs, represented by different colors and symbols, are used, but the network is formed by an arbitrary number of nodes. Here we have two types of nodes: client nodes and gateway nodes. Clients have a small area of influence due to a low attractant emission rate. They are associated with a nearby gateway and they use the gateway RA for all their outgoing messages, so they do not need to maintain a routing table. Gateways have complete routing tables and can use different RAs in their outgoing messages. Client RAs can be reused if the clients are associated to different gateways, and gateway RAs can be reused too if the gateways are distant enough. It is possible to send a message among any two nodes in the network and the architecture could be extended into a bi- or tri-dimensional lattice of gateways.

3.6. Decoding

In this step, the receiver uses its DPU to sequence the message region of the plasmid. Sequencing is the process of determining the primary structure, i.e., the sequence of nucleotides, of a DNA strand. Then the message can be decoded and processed at the receiver. This step concludes the communication process.

4. Analytical model

In this section we develop an analytical model for the performance evaluation of the Bacteria-based Nanonetwork in terms of end-to-end delay, capacity and communication range. We use the already known values for the times of the processes involved in the communication,

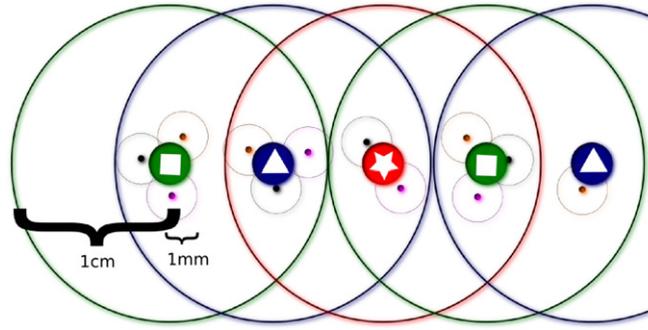


Fig. 3. BN example. The color and symbol of each node indicates the receiver attractant they emit and the circle around them their area of influence. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

except for the propagation time due to bacterial chemotaxis, which is obtained with our newly developed simulator in Section 5. Then we will obtain the capacity of the network so it can be compared with other approaches, such as the one proposed in [13].

4.1. Delay

In the Bacteria-based Nanonetwork, the delay corresponds to the sum of delays of each step in the communication process. For a single hop, the delay t_{hop} of a message of b bits to a destination at distance d in a medium with carrier density p is given by

$$t_{hop} = t_{encod}(b) + t_{encap}(p, b) + t_{prop}(d) + t_{decap}(b) + t_{decod}(b), \quad (1)$$

where

- $t_{encod}(b)$, the encoding delay, is the time needed to synthesize the plasmid at the transmitter. This parameter depends on the speed of the DPU, but we envision that it will be in the order of the bacterial DNA replication speed, $r_{DNA} = 1400$ bp/s [11].
- $t_{encap}(p, b)$ is the encapsulation delay, given by the expression $t_{encap}(p, b) = t_{carr}(p) + t_{bc}(b) + t_{exp}$, where
 - $t_{carr}(p)$ is the time needed for the transmitter to attract a carrier by emitting TA. For an adequate density of carriers, and if the transmitter starts emitting attractants while the encoding is taking place, this delay will not be significant.
 - $t_{bc}(b)$ is the time needed for bacterial conjugation. It depends on the size of the plasmid, which is physically pushed from donor to acceptor at a speed of around 833 bp/s [19].
 - t_{exp} is the time necessary for the expression of the genes that activate chemotaxis towards the receiver. With a bacterial transcription rate of 2500 nucleotides per minute [33] and an active part of the plasmid of 50 kbp, it is in the order of 20 min.
- $t_{prop}(d)$ is the propagation delay, i.e., the time required by bacteria to move from transmitter to receiver. To study this phenomenon, we use our simulator described in Section 5.
- t_{decap} is the delay for decapsulation, which is given by $t_{bc}(b)$, described above, the time for the bacterial conjugation process that will take place at the receiver.

- t_{decod} is the decoding delay, i.e., the time needed by the receiver to sequence the DNA in the received plasmid. We envision that it will be in the order of the encoding delay.

For N hops, assuming non-congested intermediate nodes, the total delay becomes

$$t_{tot} = t_{encod}(b) + N \cdot t_{encap}(p, b) + \sum_{i=1}^N t_{prop}(d_i) + N \cdot t_{decap}(b) + (N - 1) \cdot t_{rel} + t_{decod}(b), \quad (2)$$

where the term t_{rel} is the relaying delay, in which an intermediate node modifies the RA in the plasmid so that it will go to the correct next hop. With the restriction endonucleases being effective in time orders of 5 min [28], we estimate 10 min as a typical relaying delay.

4.2. Capacity

The Bacteria-based Nanonetwork presents unique features that demand a radically original approach for the study of its capacity. In principle, it may seem that the capacity of the network will be reached when there are so many carriers in transit that it is not possible for each carrier to move freely and perform chemotaxis from transmitters to receivers. This, however, will hardly be the limiting factor in this network. We can draw an analogy with the cargo capacity of the whole international merchant navy. In this case, the capacity would be reached when there are so many ships on the sea that they cannot sail to their destination ports. However, the actual limit will be reached much sooner, when all the ports in the world are at full capacity loading or unloading cargo from the ships as fast as they can. In this network, thus, the capacity is not determined by the channel itself but by the endpoints.

In our network, too, the maximum capacity is determined by the endpoints, and specifically by the speed at which plasmids can be built. This speed is in the order of the DNA replication rate of bacteria, r_{DNA} . For a multi-hop path, the routing done at the intermediate nodes needs to modify just a small part of the plasmid, and is thus faster than the encoding at the endpoints in a non-congested network.

As discussed in Section 3.1, to make the message part of the plasmid inactive, it will be necessary to avoid

consensus promoter sequences. For *E. coli*, such sequences are 5'TATAAT3' at position −35 (i.e. 35 bp before the bp where transcription starts) and 5'TGACA3' at position −10. We can thus disable transcription inserting a C after each 5'TATAA3' sequence. With this change we can obtain a rate of bits per base, bbp , of

$$bbp = \frac{5}{3} \cdot \frac{5}{1024} + 2 \cdot \frac{1019}{1024} = 1.998, \quad (3)$$

since bases that are part of a 5'TATAA3' sequence are encoding 10 bits in 6 bases and those bases that are not, still encode 2 bits per base. The probability of a base being part of such a sequence is $\frac{5}{1024}$ since there are 1024 possible 5-base sequences, and the base could occupy any of the 5 positions in it.

To obtain the capacity of the system, we must account also for the *error probability* and the *loss probability*. The error probability per base, p_{err} , is the probability of selecting the wrong base when replicating a DNA strand. In bacteria, $p_{err} = 10^{-8}$ per bp. Since two copies of the plasmid, one per bacterial conjugation, are made from receiver to transmitter and assuming that the errors are equiprobable among any two bases, the end-to-end total error probability for one hop is then

$$p_{err-1hop} = \left(1 - \frac{1}{3}p_{err}\right)p_{err} + (1 - p_{err})p_{err}, \quad (4)$$

where the first term in the addition accounts for the possibility that an error occurs in the first DNA copy and is not fixed by chance by an error in the second copy. The second term in the addition accounts for an error in the second copy.

The n –hop error probability is modeled after the single hop error probability as a geometric series. Considering the case where a second error on the same base at a later hop restores a corrupted base, by chance, to the correct one, we obtain

$$p_{err-n-hop} = 1 - (1 - p_{err-1hop})^n - \frac{1}{3}p_{err-1hop}^2 \cdot (1 - p_{err-1hop})^{n-2}. \quad (5)$$

We consider that the probability of three errors on the same base during a transmission is negligible.

The loss probability in one hop is given by

$$p_{loss} = 2p_{bc-int}(b) + p_{tout} + p_{act-err} + p_{prom} + p_{wrong-dest} + p_{death} - p_{mult}, \quad (6)$$

where p_{bc-int} is the probability that the bacterial conjugation process is interrupted for a plasmid of b bits, and p_{death} the probability of the carrier dying in transit. These two factors need to be estimated experimentally but can be 0 in a completely controlled medium. p_{tout} is the timeout probability, the probability that the propagation time will exceed the chosen TTL, which can be estimated using the simulator described in Section 5.

$p_{act-err}$ is the probability that an error is made when replicating the active section of the plasmid, transfer and routing regions. An error in this part may make the bacterium unable to either swim or transfer the plasmid

to the receiver. For an active area of l base pairs, $p_{act-err} = 1 - (1 - p_{err})^l$.

p_{prom} is the probability that a promoter appears in the message because of a mutation, causing the bacterium to express arbitrary DNA, which may incapacitate the bacterium to propagate to the receiver. We compute the probability that the *consensus promoter* appears in the message because of a mutation, but p_{prom} is lower since this is a necessary but not sufficient condition for a valid promoter and, furthermore, even if there is a valid promoter it might not affect the bacterium propagation. The probability that a mutation in a base will form the consensus promoter is $\frac{12}{222}$, since the consensus promoter is composed of 12 bases (12 possible positions the base might occupy in the sequence, and a chance of 1 in 4^{11} that all the other bases with respect to this one are correct). Then, for a message of k bases,

$$p_{prom} < \frac{12}{222} \cdot k \cdot \frac{1}{3}p_{err}. \quad (7)$$

$p_{wrong-dest}$ is the probability that a carrier delivers the message to the wrong node. It depends on the density and geometry of the node distribution, but we consider it is non-significant for topologies where nodes in range do not have other nodes in their line of sight and the distances between nodes are orders of magnitude higher than the node size.

p_{mult} is the probability of more than one loss condition happening to the same message. We consider this condition too rare to have a significant impact in the total loss probability.

The loss probability for multiple hops obtained again as a geometric series $p_{lossn-hop} = 1 - (1 - p_{loss})^n$.

The capacity per base is then

$$C_{base} = bbp - H(p_{err-n-hop}), \quad (8)$$

where $H(p) = -(p \log_2(p) + (1 - p) \log_2(1 - p))$ is the binary entropy function [35]. Finally, the total capacity of the network is

$$C = C_{base} \cdot r_{DNA} \cdot (1 - p_{lossn-hop}). \quad (9)$$

Since different bacteria do not cause interference to each other, this capacity can be simultaneously attained by every transmitter in the network as long as the intermediate nodes are not saturated.

4.3. Range

The range for a hop is the maximum distance at which a bacterium sensitive to a specific attractant will be able to find an emitter of that attractant. This range is determined by the concentration gradient of attractant created by the receiver, since bacteria will only find their way to the receiver if the steepness of the gradient is adequate to be sensed by them. This gradient depends on the attractant emission rate Q . The range is higher for higher Q , but, if Q is too high, the gradient near the receiver will be too steep, and bacteria will saturate and not find the receiver.

The concentration distribution of a molecule emitted at a constant rate of Q mol/s from a fixed point in an unbounded 3-dimensional space is [9]:

$$U(r, t) = \frac{1}{1000} \frac{Q}{4D\pi r} \operatorname{erfc} \left(\frac{r}{\sqrt{4Dt}} \right) \quad (10)$$

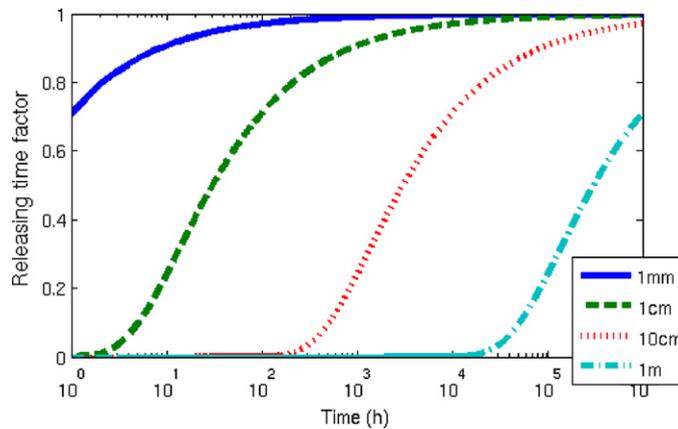


Fig. 4. Evolution of the concentration gradient time factor over time for different distances.

with d the distance from the point of release in m , t the time during which the particles have been released in s and D the diffusion coefficient in m^2/s . The resulting concentration is in molar. For infinite time, the concentration reaches its stationary level $U(r) = \frac{1}{1000} \frac{Q}{4D\pi r}$.

The limiting factor of the range will be given by the time-dependent factor of Eq. (10), $erfc\left(\frac{r}{\sqrt{4Dt}}\right)$, which is plotted in Fig. 4 over a logarithmic time scale for different distances. This factor determines how long it takes for the attractant concentration at a given distance from the point of emission to go from 0, at the moment the attractant starts to be emitted, to its stationary level, when the time factor approaches 1. We consider that a concentration gradient starts to be effective at a certain distance when the concentration reaches 20% of the stationary level. The time until such level is reached is the *preparation time*. In Fig. 4, we show that for a distance of 10 cm the concentration is effective only after 1000 h. We believe that larger preparation times are impractical and thus, consider the maximum range for a hop in the BN is in the order of tens of cms.

5. Simulation of bacteria propagation through chemotaxis

5.1. Motivation

The purpose of the bacteria simulator is to provide a measure of the time that bacteria need to move, using chemotaxis, from transmitter to receiver, since adequate estimates of this have not yet been published. Additionally, the simulator provides the timeout probability, i.e., the probability that a bacterium has not reached yet the receiver after a specific time. The stochastic nature of chemotaxis along with the high number of parameters involved in the process make simulation the best tool for these estimations. Some works have characterized chemotaxis using differential equations [31] that consider whole populations of bacteria and relate the population density and attractant density over time. This approach, however, cannot be used in our case of modeling the behavior of a single bacterium moving from a specific origin to a specific destination.

We can find a number of chemotaxis simulators in biology literature [10,39] that focus on the internal chemotac-

tic signal pathway. However, a different tool is needed to examine the complete behavior of bacteria when moving to specific locations. A simulator similar to ours is implemented in [23]. However, its architecture model makes it impossible for us to measure the delay in distances greater than 1 mm. Moreover, the simulation in [23] is based on a 2 dimensional medium. With our simulator, it is possible to effectively simulate any distance useful for bacterial chemotaxis in a 3D medium. The difference between the 2D and 3D medium is important. Having one more degree of freedom, it is more difficult for bacteria to reach the exact location of the transmitter, causing an *eye of a needle* effect that increases the delay. In a non-biased random walk, for example, the probability of reaching any point of a regular lattice as the number of steps approaches infinity is 1 in 2D, but only 0.34 in the 3D case [41].

5.2. Model

Our simulator follows the most widely accepted model for bacterial chemotaxis in Biology literature. A bacterium is, at any moment, either in a *running* or *tumbling* [7] state and alternates continuously between both. The duration of each state is exponentially distributed, with a mean duration of 1.25 s for runs and 0.17 s for tumbles. During runs, bacteria swim in an approximately straight direction and in tumbles they spin on their position, changing the direction of the next run.

Bacteria have a short-term memory [29] spanning a few seconds, so that bacteria notice when the attractant concentration is increasing. When this happens, they perform longer runs, which provokes a *biased random walk* that leads bacteria to their destination.

The simulator divides the time into steps of 0.01 s. At each time step, the probability of state transition is computed and the bacterium state is updated. Then the position and direction of the bacterium are determined. The rest of this section explains the model in more detail.

(1) *Runs*: During runs, bacteria swim at a speed of 20 μm per second, but their trajectory is not perfectly straight. Rotational diffusion causes bacteria to change their direction by a mean square angular deviation on each axis of

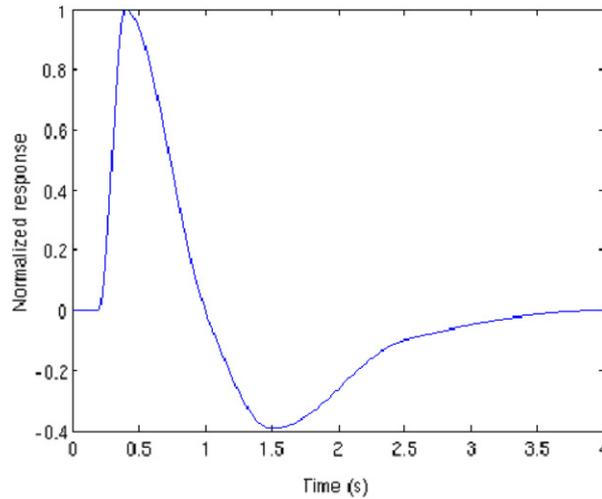


Fig. 5. Normalized impulse response vs. time.

$\langle \theta^2 \rangle = 2D_r t$, where D_r is the rotational diffusion coefficient and t the time. Approximating the bacterium as a sphere of radius a , $D_r = \frac{kT}{8\pi\eta a^3}$, with k the Boltzman constant, T the temperature in Kelvin and $\eta = 0.027$ g/cm s the dynamic viscosity. Rotational diffusion is implemented by randomly adding or removing at each time step an angular deviation equal to the mean square angular deviation to the current pitch and yaw.

(2) *Tumbles*: During tumbles, bacteria spin and pick a new direction for the next run. The new direction is correlated with the direction of the last run, the difference between both directions on a planar projection being on average 62° with a standard deviation of 26° . We compute the change of direction after a tumbling phase as a normal distribution with the mentioned parameters, truncated between 0 and π and normalized. Two values from this distribution are added to the pitch and yaw with the signs being determined by coin flips.

(3) *Impulse response*: Bacteria process attractant concentration with a system that behaves as a linear time-invariant system [8] when it is not saturated. While the shape of the impulse response, shown in Fig. 5, is known, there is still controversy about the actual gain of the system [36,26]. The positive and negative lobes of the impulse are equal, which means that the system is stable and inactive at a constant concentration. The impulse response varies for different organisms and attractants, but we focus now on a specific response to characterize the channel. We will study the advantages and feasibility of different impulse responses in the future.

The mean run time is $t_R = \min(10, 1.25 + k_s \cdot \max(0, h * s))$, where h is the impulse response, s is the concentration signal and $*$ the convolution operator. If the result of the convolution is zero or negative, the mean run time is the base mean run time, 1.25 s. k_s is a constant chosen to make a lineal concentration increase of 50 mM/ μ m, which is the saturation concentration, result on a mean run time of 10 s. Bacteria will be insensitive to concentration gradients steeper than the saturation concentration, keeping the maximum 10 s mean run time for them.

6. Performance evaluation

Our performance evaluation is based on known characteristics of biological phenomena used in the channel and the simulator described in Section 5 for the propagation delays. In the simulations, unless otherwise stated, the receiver has a radius $r_{rcv} = 50$ μ m, and emits attractant at a rate $Q = 10^{-11}$ mol/s. The attractant has a diffusion coefficient $D = 10^{-9}$ m²/s, like the most studied attractant, aspartate. Fig. 6 shows, from left to right and top to bottom, the xz , xy , and yz projections and the perspective of a bacterium run from a transmitter to a receiver 1 cm apart, plus the evolution of the distance to receiver with time.

6.1. Range

As explained in Section 4.3, the time constraints limit the range of a single hop to tens of cms. It is necessary to check by simulation if, below this limit, a concentration distribution created by a single and fixed attractant emitter can attract the bacteria. In Figs. 7–9, we show the evolution of bacteria distance to receiver with time for different Q , considering infinite emission time, averaged over 1000 runs, for 3 different initial distances: 1 mm, 1 cm and 10 cm, respectively. These images show that for 1 mm distances, the best rate is $Q = 10^{-12}$, but at 1 cm that rate attracts the bacteria very slowly and $Q = 10^{-11}$ is more effective. $Q = 10^{-9}$ saturates the bacterium at distances smaller than 4 cm from the receiver, but it is the only one sensed by bacteria 10 cm away. Given this, to attract a bacterium 10 cm away, a receiver would need to emit several independent attractants at different rates, each attractant driving bacteria in the right direction in a different distance range.

6.2. Delay

The delay for one hop, detailed in Section 4.1 can be split in fixed delay and propagation delay. The first comprises every step in a one-hop communication process except

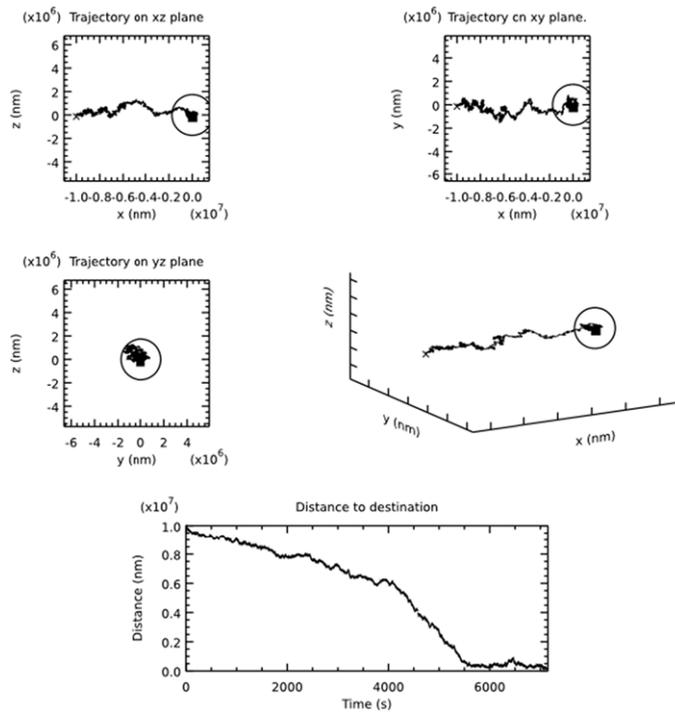


Fig. 6. Sample simulation for $r_{rcv} = 50 \mu\text{m}$, $Q = 10^{-11} \text{ mol/s}$, $D = 10^{-9} \text{ m}^2/\text{s}$, $d = 1 \text{ cm}$.

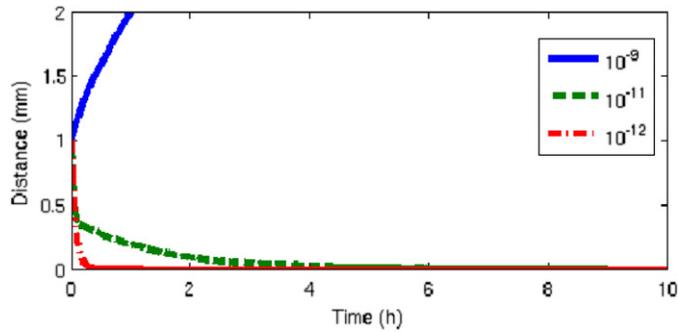


Fig. 7. Distance from receiver vs. time for initial $d = 1 \text{ mm}$ and different RA emission rates.

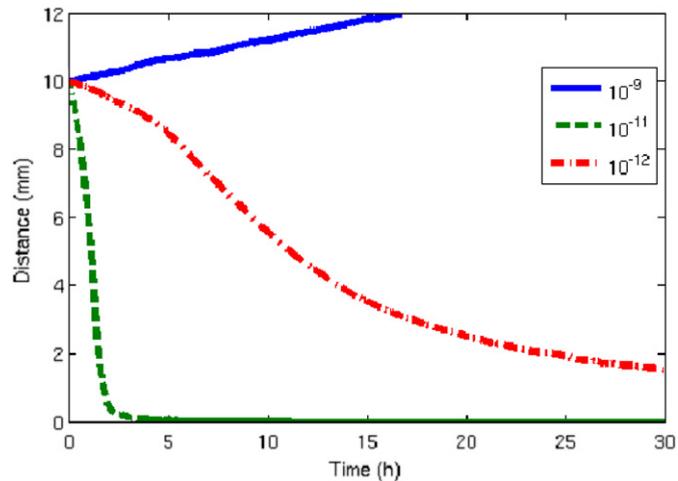


Fig. 8. Distance from receiver vs. time for initial $d = 1 \text{ cm}$ and different RA emission rates.

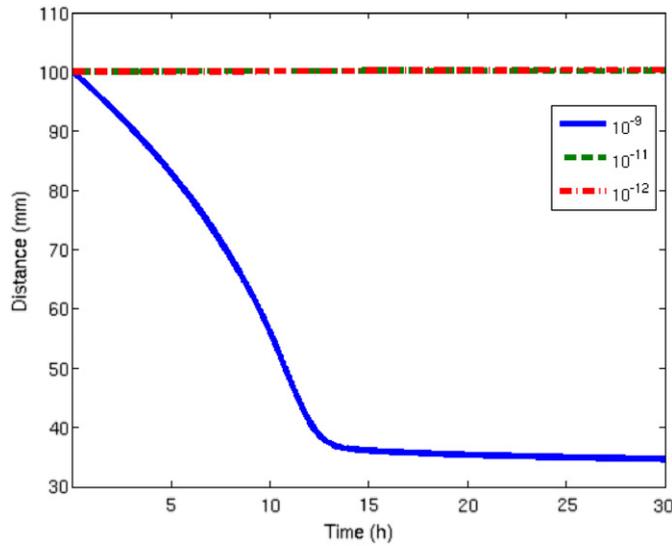


Fig. 9. Distance from receiver vs. time for initial $d = 10$ cm and different RA emission rates.

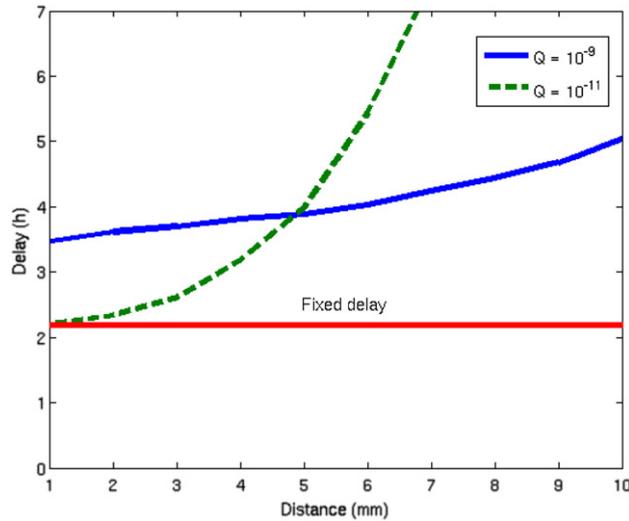


Fig. 10. Average total delay vs distance for different emission rates.

for the propagation and can be considered constant and equal to 122 min. The propagation delay is highly variable and depends on the distance to the transmitter, the RA emission rate of the transmitter, the receiver radius and the specific bacteria behavior. Fig. 10 shows the mean total delay, averaged over 1000 simulations for each case, for different emission rates and initial distances, from 1 mm to 1 cm. We see again that different emission rates have different areas of effectiveness.

Regarding the propagation delay variability, Figs. 11 and 12 show the delay distribution for $d = 1$ mm, $Q = 10^{-9}$ and $d = 1$ cm, $Q = 10^{-11}$. As has been proposed before [23], the delay distribution resembles a Gamma distribution.

For multi-hop, each intermediate node has to de-capsulate the plasmid, modify it for the next hop, and encapsulate it again, which takes a total time of 90 min.

This means that it is not practical to subdivide a long hop in faster, smaller hops. Even though the total propagation time in the many-hops route can be lower than in the 1-hop, the routing delays will cause a higher total delay. Using 1 cm hops then, BN offers a delay, for a distance of n cm, of $(4.5n - 0.5)$ h.

6.3. Capacity

We follow the model in Section 4.2 to compute the capacity of the BN using a 1.6 mbp plasmid divided in a 50 kbp active section and a 1.55 mbp message. $p_{err-1hop} = 2 \cdot 10^{-8}$. In a controlled medium, $p_{bc-int}, p_{death} = 0$. $p_{wrong-dest} = 0$ too for a typical topology. Additionally, the simulation shows that for a timeout large enough, like 20 h, p_{tout} can also be considered as 0. p_{prom} for a message size of 1.55 mbp is less than $1.43 \cdot 10^{-8}$. The probability of an error

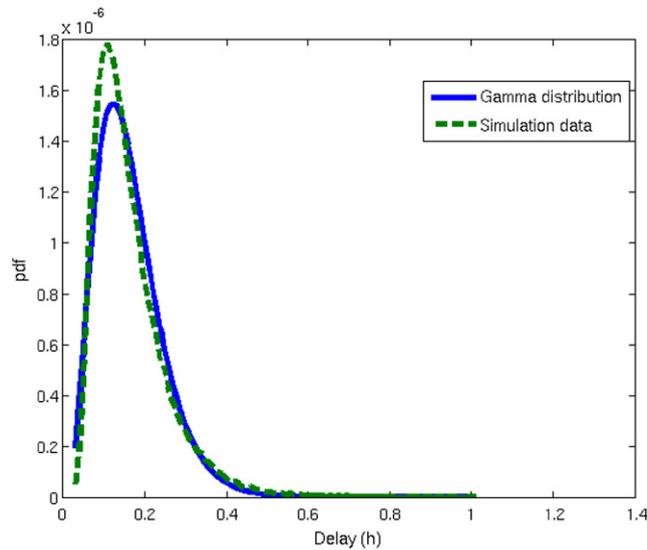


Fig. 11. Propagation delay distribution for $d = 1$ mm, $Q = 10^{-9}$.

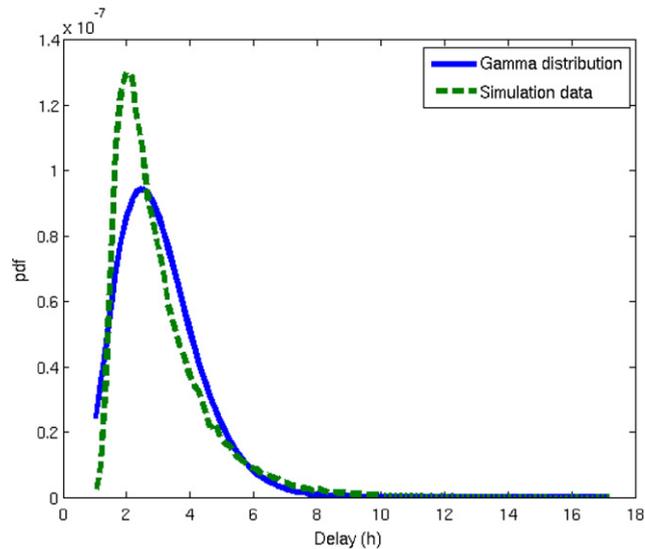


Fig. 12. Propagation delay distribution for $d = 1$ cm, $Q = 10^{-11}$.

in the active section of the plasmid is $p_{act-err} = 5 \cdot 10^{-4}$, a value that can be considered as the total loss probability.

With $\text{bpp} = 1.998$, the capacity per base for one hop is then $C_{base} = 1.998 - H(2 \cdot 10^{-8}) \approx 1.998$. The total capacity for one hop is $C = 2796$ bps (bits per second). This compares very favorably against other molecular communication approaches, such as the one proposed in [13], where a one dimensional space is assumed and a capacity below 0.1 bps is achieved and a delay in the range of 10 h is obtained.

For multiple hops, the error probability grows, but for 100 hops it is 10^{-6} , which still has a negligible influence in C_{base} . The capacity can be rewritten, depending on the number of hops, n , as

$$C = C_{base} \cdot r_{DNA} \cdot (1 - p_{loss})^n, \quad (11)$$

from which the capacity with n hops is 99.5% of the capacity with $n - 1$ hops.

7. Conclusions

Bacteria-based Nanonetworks use bacteria as carriers of messages encoded in DNA molecules. This novel approach to molecular communications offers significant improvements, such as a capacity 4 orders of magnitude higher than previous techniques. We envision that this type of network will be extremely useful for applications such as data collection from nanosensors, or command submission for nanoactuators.

In this paper we describe the different components of the network, nodes and carriers, and the various steps in the communication process along with the biological

phenomena that support them. This demonstrates the feasibility of the BN and, additionally, allows us to develop a mathematical analysis of the network based on the available information on the biological phenomena involved. In this analysis, we provide expressions for the end-to-end delay, capacity and communication range of the system. Given the lack of reliable estimates of the bacterial propagation time, we have developed a new simulator to obtain appropriate statistics on this time. Using our mathematical analysis and our simulator, we conduct a performance evaluation of the network that shows its superiority over other approaches.

In the future, we plan to address some of the many interesting research issues that BN presents, such as address assignment, multi-hop path creation, multi-attractant receivers, mathematical analysis of propagation delay, the effect of saturated intermediate nodes on capacity, or a more detailed description of each of the steps in the communication process. This paper provides a roadmap for what we believe will be one of the most useful and exciting approaches for molecular communication.

Acknowledgement

The authors would like to thank Ozgur Akan, Eylem Ekici, Tommaso Melodia, Dario Pompili and Mehmet Can Vuran for their constructive criticism which helped to improve the quality of the paper. This material is based upon work supported by the US National Science Foundation under Grant no. CNS-0910664 and Obra Social “la Caixa”.

References

- [1] M. Achtman, N. Kennedy, R. Skurray, Cell-cell interactions in conjugating *Escherichia coli*: role of traT protein in surface exclusion, Proceedings of the National Academy of Sciences of the United States of America 74 (11) (1977) 5104.
- [2] L. Adleman, Molecular computation of solutions to combinatorial problems, Science 266 (5187) (1994) 1021–1023.
- [3] J. Adler, Chemotaxis in bacteria, Harvey Lectures 72 (1978) 195–230.
- [4] I.F. Akyildiz, F. Brunetti, C. Blázquez, Nanonetworks: a new communication paradigm, Computer Networks (Elsevier Journal) 52 (2008) 2260–2279.
- [5] I.F. Akyildiz, J.M. Jornet, Electromagnetic wireless nanosensor networks, Nano Communication Networks (Elsevier Journal) 1 (1) (2010) 3–19.
- [6] Y. Benenson, T. Paz-Elizur, R. Adar, E. Keinan, Z. Livneh, E. Shapiro, Programmable and autonomous computing machine made of biomolecules, Nature 414 (6862) (2001) 430–434.
- [7] H. Berg, *E. coli* in Motion, Springer Verlag, 2003.
- [8] S. Block, J. Segall, H. Berg, Impulse responses in bacterial chemotaxis, Cell 31 (1) (1982) 215–226.
- [9] W. Bossert, E. Wilson, The analysis of olfactory communication among animals, Journal of Theoretical Biology 5 (3) (1963) 443–469.
- [10] D. Bray, M.D. Levin, K. Lipkow, The chemotactic behavior of computer-based surrogate bacteria, Current Biology 17 (1) (2007) 12–19.
- [11] S. Cooper, C.E. Helmstetter, Chromosome replication and the division cycle of *E. coli* B/r, Journal of Molecular Biology 31 (3) (1968) 519–540.
- [12] J.W. Drake, B. Charlesworth, D. Charlesworth, J.F. Crow, Rates of spontaneous mutation, Genetics 148 (4) (1998) 1667–1686.
- [13] A.W. Eckford, Nanoscale Communication with Brownian Motion, in: *Conference on Information Sciences and Systems, 2007, CISS'07*, Mar. 2007, pp. 160–165.
- [14] J. Elbaz, O. Lioubashevski, F. Wang, F. Remacle, R.D. Levine, I. Willner, DNA computing circuits using libraries of DNAzyme subunits, Nature Nanotechnology 5 (6) (2010) 417–422.
- [15] H. Engelberg-Kulka, B. Sat, M. Reches, S. Amitai, R. Hazan, Bacterial programmed cell death systems as targets for antibiotics, Trends in Microbiology 12 (2) (2004) 66–71.
- [16] T.M. Finan, S. Weidner, K. Wong, J. Buhrmester, P. Chain, F.J. Vorhölter, I. Hernandez-Lucas, A. Becker, A. Cowie, J. Gouzy, B. Golding, A. Pühler, The complete sequence of the 1,683-kb pSymB megaplasmid from the N₂-fixing endosymbiont *Sinorhizobium meliloti*, Proceedings of the National Academy of Sciences of the United States of America 98 (17) (2001) 9889–9894.
- [17] A. Fire, S. Xu, M. Montgomery, S. Kostas, S. Driver, C. Mello, Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*, Nature 391 (6669) (1998) 806–811.
- [18] L.S. Frost, K. Ippen-Ihler, R.A. Skurray, Analysis of the sequence and gene products of the transfer region of the F sex factor, Microbiological Reviews 58 (2) (1994) 162–210.
- [19] C. Fulton, Continuous chromosome transfer in *E. coli*, Genetics 52 (1) (1965) 55.
- [20] D. Gibson, J. Glass, C. Lartigue, V. Noskov, R.Y. Creation of a bacterial cell controlled by a chemically synthesized genome, Science 329 (5987) (2010) 52–56.
- [21] M. Gregori, I.F. Akyildiz, A new nanonetwork architecture using flagellated bacteria and catalytic nanomotors, IEEE Journal on Selected Areas in Communications 28 (4) (2010) 612–619.
- [22] M. Gregori, I. Llatser, A. Cabellos-Aparicio, E. Alarcón, Physical channel characterization for medium-range nanonetworks using catalytic nanomotors, Nano Communication Networks (Elsevier Journal) 1 (2) (2010).
- [23] M. Gregori, I. Llatser, A. Cabellos-Aparicio, E. Alarcón, Physical channel characterization for medium-range nano-networks using flagellated bacteria, Computer Networks (ISSN: 1389-1286) (2010).
- [24] A. Grossman, Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*, Annual Review of Genetics 29 (1) (1995) 477–508.
- [25] G. Hazelbauer, R. Mesibov, J. Adler, *E. coli* mutants defective in chemotaxis toward specific chemicals, Proceedings of the National Academy of Sciences 64 (4) (1969) 1300.
- [26] R. Jasuja, Y. Lin, D.R. Trentham, S. Khan, Response tuning in bacterial chemotaxis, Proceedings of the National Academy of Sciences of the United States of America 96 (20) (1999) 11. 346–51.
- [27] U. Lehninger, D. Nelson, M. Cox, Lehninger Principles of Biochemistry, Wh Freeman, 2005.
- [28] C. Ma, Z. Tang, K. Wang, W. Tan, X. Yang, W. Li, Z. Li, X. Lv, Real-time monitoring of restriction endonuclease activity using molecular beacon, Analytical Biochemistry 363 (2) (2007) 294–296.
- [29] R.M. Macnab, D. Koshland, The gradient-sensing mechanism in bacterial chemotaxis, Proceedings of the National Academy of Sciences 69 (9) (1972) 2509–2512.
- [30] T. Ogura, S. Hiraga, Mini-F plasmid genes that couple host cell division to plasmid proliferation, Proceedings of the National Academy of Sciences of the United States of America 80 (15) (1983) 4784–4788.
- [31] K.J. Painter, T. Hillen, Volume-filling and quorum-sensing in models for chemosensitive movement, Canadian Applied Mathematics Quarterly 10 (4) (2002) 501–543.
- [32] M. Pierobon, I. Akyildiz, A physical end-to-end model for molecular communication in nanonetworks, IEEE Journal on Selected Areas in Communications 28 (4) (2010) 602–611.
- [33] P.K.M. Robert, C. King, William D. Stansfield, A Dictionary of Genetics, Oxford University Press, USA, 2006.
- [34] P. Rothmund, A DNA and restriction enzyme implementation of Turing machines, DNA Based Computers 6 (1996) 75–119.
- [35] C.E. Shannon, A mathematical theory of communication, The Bell System Technical Journal 27 (1948) 379–423. 623–656.
- [36] V. Sourjik, H.C. Berg, Receptor sensitivity in bacterial chemotaxis, Proceedings of the National Academy of Sciences of the United States of America 99 (1) (2002) 123–127.
- [37] J. Stricker, S. Cookson, M. Bennett, W. Mather, L.S. A fast, robust and tunable synthetic gene oscillator, Nature 456 (7221) (2008) 516–519.
- [38] L. Tsimring, J. Hasty, T. Danino, O. Mondrago, A synchronized quorum of genetic clocks, Nature 463 (7279) (2010) 326–330.
- [39] N. Vladimirov, L. Løvdok, D. Lebedez, V. Sourjik, Dependence of bacterial chemotaxis on gradient shape and adaptation rate, PLoS Computational Biology 4 (12) (2008) e1000242.

- [40] C.M. Waters, B.L. Bassler, Quorum sensing: cell-to-cell communication in bacteria, *Annual Review of Cell and Developmental Biology* 21 (2005) 319–346.
- [41] E.W. Weisstein, Pólya's Random Walk Constants. From MathWorld—A Wolfram Web Resource.



Luis Carlos Cobo received a Degree in Telecommunication Engineering from the School of Electrical Engineering, Universidad Politécnica de Madrid (UPM), and a Master of Science in Electrical Engineering from the Illinois Institute of Technology in 2006. He then worked for two years in the San Francisco-based consulting company Cozybit Inc., where he developed the wireless mesh networking stack used in the One Laptop per Child (OLPC) project and the one currently included in the Linux kernel. Currently, he is pursuing his Ph.D. at the Georgia Institute of Technology, with a fellowship from Obra Social “la Caixa”. He is a student member of the IEEE and the ACM.



Ian F. Akyildiz is the Ken Byers Chair Professor in Telecommunications with the School of Electrical and Computer Engineering, Georgia Institute of Technology (Georgia Tech), Atlanta, and the Director of the Broadband Wireless Networking Laboratory and the Chair of the Telecommunication Group at Georgia Tech. In June 2008, Dr. Akyildiz became an honorary professor with the School of Electrical Engineering at Universitat Politècnica de Catalunya (UPC) in Barcelona, Spain. He is also the Director of the newly founded N3Cat (NaNoNetworking Center in Catalunya). He is also an Honorary Professor with University of Pretoria, South Africa, since March 2009. He is the Editor-in-Chief of *Computer Networks* (Elsevier) Journal and the founding Editor-in-Chief of *Ad Hoc Networks* (Elsevier) Journal, *Physical Communication* (Elsevier) Journal and *Nano Communication Networks* (Elsevier) Journal. Dr. Akyildiz serves on the advisory boards of several research centers, journals, conferences and publication companies. He is an IEEE FELLOW (1996) and an ACM FELLOW (1997). He received numerous awards from IEEE and ACM. His research interests are in nanonetworks, cognitive radio networks and wireless sensor networks.