



**HAL**  
open science

# Markov model of quantum fluctuations at the transition to lasing of semiconductor nanolasers

Arthur Vallet, Laurent Chusseau, Fabrice Philippe, Alain Jean-Marie

## ► To cite this version:

Arthur Vallet, Laurent Chusseau, Fabrice Philippe, Alain Jean-Marie. Markov model of quantum fluctuations at the transition to lasing of semiconductor nanolasers. *Physica E: Low-dimensional Systems and Nanostructures*, 2019, 105, pp.97-104. 10.1016/j.physe.2018.08.028 . hal-01888162

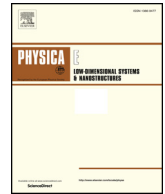
**HAL Id: hal-01888162**

**<https://hal.science/hal-01888162>**

Submitted on 25 Oct 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Markov model of quantum fluctuations at the transition to lasing of semiconductor nanolasers



Arthur Vallet<sup>a,\*</sup>, Laurent Chusseau<sup>a</sup>, Fabrice Philippe<sup>b</sup>, Alain Jean-Marie<sup>c</sup>

<sup>a</sup> IES, Univ Montpellier, CNRS, Montpellier, France

<sup>b</sup> LIRMM, Univ Montpellier 3, CNRS, Montpellier, France

<sup>c</sup> INRIA, Univ Montpellier, CNRS, LIRMM, Montpellier, France

## ARTICLE INFO

### Keywords:

Nanolaser  
Semiconductor laser  
Markov chain  
Laser threshold

## ABSTRACT

A Markov model of semiconductor nanolaser is constructed in order to describe finely the effects of quantum fluctuations in the dynamics of the laser, in particular by considering the transition to lasing. Nanolasers are expected to contain only a small number of emitters, whose semiconductor bands are simulated using true carrier energy states. The model takes into account carrier-carrier interactions in the conduction and valence bands, but the result is a huge Markov chain that is often too demanding for direct Monte-Carlo simulation. We introduce here a technique to split the whole chain into two subchains, one referring to thermalization events within the bands and the other to laser photonic events of interest. The model is applied to the analysis of laser transition and enlightens the coexistence of a pulse regime triggered by the quantum nature of the photon with the birth of the known coherent cw regime. This conclusion is highlighted by calculated time traces. We show that on the ultrasmall scale of nanolasers, we are unable to define perfectly the threshold.

## 1. Introduction

Nanolasers are becoming increasingly important in the scientific and technical community. Their potential applications for high-bit-rate optical interconnections can solve the need for an exponential increase in information flow, even in microelectronics at inter or intra-chip levels, for example. Regardless of the application and due to their ultrasmall size, nanolasers have unique advantages for this type of requirement as low consumption becomes an essential need. Their realization relies historically on semiconductor technologies based on heterostructures [1,2], but was extended recently to spasers [3] whose plasmonic effects allow resonant cavities of a volume much lower than  $\lambda^3$ , with  $\lambda$  the wavelength of operation in vacuum [4–6]. Because of these small ultimate sizes, only a few individual emitters can be involved in a single device, which results in a very small number of photons emitted and an increasing importance of the fluctuations due to the quantum nature of processes, either optical or electrical. In semiconductor laser technology, the tiny number of emitters is either reached with quantum dots or with very small VCSELs, although even some commercial VCSELs behave like nanolasers [7]. Laser models must then account for this quantization to move from continuous variable descriptions like rate equations [8] toward a quantum microscopic description of the device [9]. The expected physical answers are

then a better description of the transition from incoherent to coherent emission [10,11] and the associated ultimate noise performances of a device [12,13].

One class of very accurate models able to describe nanolasers relies purely on quantum mechanics [14]. It numerically includes only very few two-levels emitters ( $N_e \leq 5$ ) in the optical cavity [15,16] because the numerical complexity of the density matrix increases exponentially with  $N_e$ . However, in spite of this limitation these models have shown that coherence effects between excited atoms are smoothed as soon as their count exceeds a few units. The consequence is that the extensive consideration of these coherences is not necessary to give a detailed account of a real nanolaser.

If an accurate description of the quantum-well (QW) active medium is required, a quantum-mechanical nonequilibrium theory for the coupled carrier-photon system in semiconductor micro cavity lasers was built to analyze the laser dynamics and the stationary emission properties [17]. It shows dynamical response of the microlaser strongly influenced by nonequilibrium carrier effects.

Alternatively a quantum semiclassical description of the excited states and photon bath is much simpler and allows a complete stochastic description of both the photon and atom populations. Numerical models may rely either on a Fokker-Planck equation [18,19] or on a detailed account of the underlying birth-death Markov processes [20].

\* Corresponding author.

E-mail address: [vallet@ies.univ-montp2.fr](mailto:vallet@ies.univ-montp2.fr) (A. Vallet).

<https://doi.org/10.1016/j.physe.2018.08.028>

Received 7 May 2018; Received in revised form 20 August 2018; Accepted 27 August 2018

Available online 04 September 2018

1386-9477/ © 2018 Elsevier B.V. All rights reserved.

On the one hand, this representation was used to elucidate the population oscillation in nanolasers or the transition to lasing in high- $\beta$  nanolasers [11,21]. Models however include a large number of emitters that are only simulated within the approximation of small Gaussian fluctuations. If fewer emitters are considered, a more rigorous modeling is possible using the exact algorithm of Gillespie [22] applied to the birth-death Markov process. This has allowed the description of sub-Poissonian light generation from regularly pumped semiconductor lasers [12] or the increase of the apparent threshold of a nanolaser because of photon-carrier correlation [13].

The model proposed in this document falls within the latter framework and considers microlasers with QW active medium. It takes into account the semiconductor electronic band description [12,23] and finely reproduces the carrier-carrier interactions within conduction and valence bands. This is obtained owing to an evenly-spaced representation of energy levels within bands that was shown separately in perfect agreement with the grand-canonical density of states expected in semiconductors [24]. Although satisfactory from the point of view of accuracy, this model suffers from its lack of numerical efficiency if too many electrons have to be simulated in each band due to their thermalization. Extending results of [12,23] to more realistic nanolasers operating at room temperature would thus have required excessive CPU time. In the present paper we use a known principle on the Markov chains [25,26] to split our birth-death process into a thermalization part and a photonic events part to speed up the simulation process. Extreme CPU time improvements are demonstrated owing to a subsequent analytic account of the thermalization that we consider within the thermodynamics canonical framework between each pair of photonics events. As a result, our new model reproduces all known nanolaser peculiarities, especially in the transition to lasing [10,11,13,27] but with a realistic semiconductor occupancy.

## 2. Material and methods

Our picture of a semiconductor (SC) laser consists of a quantized electromagnetic field resonant with a cavity containing  $m$  photons and a finite number of electrons occupying the two upper bands of the SC: the conduction band (CB) and the valence band (VB), cf. Fig. 1. A finite number of energy values is available for each electron, their levels being evenly-spaced into the CB and the VB, separated by the SC bandgap. We fix the number of energy levels at  $B$  per band and the energy separation between levels to  $\delta$ . Recalling the fermion nature of electrons, each level can be occupied by one electron at maximum using a Pauli exclusion principle without spin consideration. Moreover we chose  $B$  electrons only for the whole system so as half of the energy levels are filled.

Interactions between electrons and photons are prescribed to occur only at a specified pair of energy levels whose energy difference becomes that of a photon stored in the cavity or emitted outside. Correspondingly, only one frequency is supported by this laser model, other frequency modes being considered filtered out by the cavity. For convenience, lasing levels are located in the middle of the VB and the CB. This will bring useful symmetry in further analytical calculations and is not restrictive for the physical description of the system.

The laser is represented by a Markov chain in continuous time. Individual states of the chain are completely described by the photon number and the electron repartition over the set of  $2B$  energy levels. Since photonic events may occur only at lasing levels, they can be listed as follows: If a hole fills the CB lasing level (CBL) and an electron fills the VB lasing level (VBL), a photon can be absorbed with a rate  $gm$  upgrading the electron at CBL. In this expression,  $g$  figures the laser gain expressed as the inverse of a time. Conversely, with an electron in CBL and a hole in the VBL, a photon can be emitted with a rate  $g(m + \beta)$  downgrading the electron to VBL, with  $\beta$  the fraction of the spontaneous emission coupled to the lasing mode. A photon emitted in another mode is lost with a rate  $g(1 - \beta)$ . For convenience,  $g = 1$  is

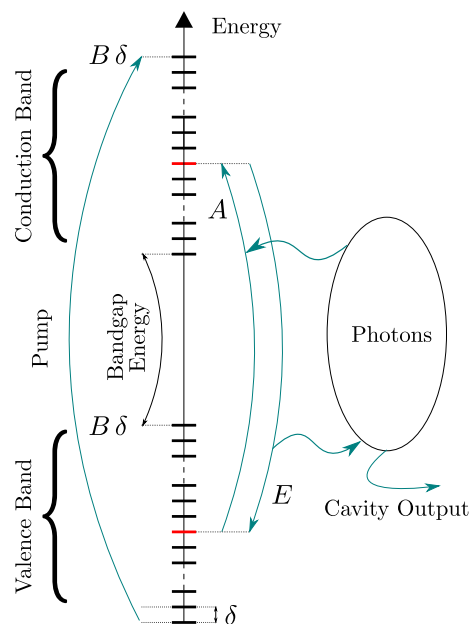


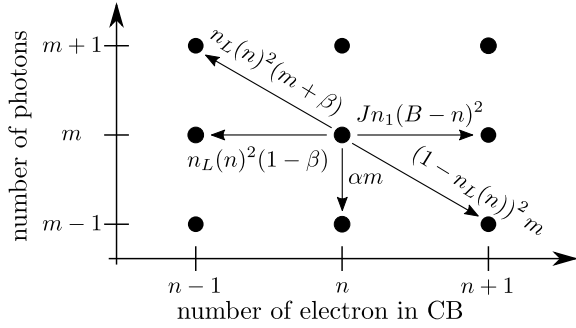
Fig. 1. Schematic of the semiconductor laser model. Bands (VB and CB) are figured by  $B$  evenly-spaced energy levels separated by  $\delta$ . The semiconductor gain medium is coupled to the optical cavity where photons are stored and can escape with a rate  $\alpha$ . Emission  $E$  and absorption  $A$  apply only to laser levels prescribed arbitrarily in the middle of the bands. The external pump promote the lowest energy electron in VB to the upper energy level in CB.

assumed everywhere as it defines the time unit. It is omitted in the sequel and therefore actual photon generation and absorption rates are simply  $m + \beta$ ,  $1 - \beta$  and  $m$ . According to band occupations, the energy balance is ensured by pump events which promote the electron in the VB lowest energy to the CB highest level with a constant rate  $J$  (Poissonian pump). On the other side, cavity losses (photon exit) are the useful external signal of the laser produced at a rate  $\alpha m$ , with  $\alpha^{-1}$  the photon lifetime in the cavity.

Markov chain trajectories are numerically obtained using the Gillespie algorithm [22], which gives a stochastically rigorous time evolution of the process and does not require time discretization. Further averaging produces first- and second-order statistics for electron and photon distribution. Taking into account the ergodicity of the process makes it possible in practice to take averages either in time or with multiple trajectories.

Crucial information for calculating the temporal evolution of photonic variables within trajectories are the probabilities of occupation of CBL and VBL by an electron. The latter, called occupancies, may be known according to the different ensembles of statistical physics. Although extensively used in SC physics, a rough application of the grand-canonical ensemble without the account of the feedback induced by photon emission and absorption leads to a nonsense in that occupancies will be constant regardless of the photon flux, discarding the population clamping that gives the coherent property of a laser.

On the other side, the microcanonical ensemble [28] applies only on isolated systems. Although it was shown to give excellent band description [24], it is unpractical here because either photonic events or thermalization events within CB or VB modify energy. That is why we are moving from such a microcanonical model to a microscopic Markov model (MMM) [12,23], which adds the possibility for an electron to downgrade its energy with a rate  $p$  and upgrade it with the same rate multiplied by the Boltzmann factor. This produces a correct electron repartition within bands and consequently on CBL and VBL. As is, the model is very accurate and correctly represents a laser based on the picture of a quantized gain medium interacting with a quantized field intensity. However, it is very inefficient in CPU time because of the



**Fig. 2.** Representation of the birth-death Markov process out of the state  $(n, m)$ . Arrows are the transition rates with their respective values:  $n_L(n)$  is the occupancy at lasing level in CB with  $n$  electrons in the band,  $n_1(B-n)$  is the occupancy at lower level in VB in the same conditions,  $\beta$  is the fraction of the spontaneous emission coupled to the lasing mode,  $J$  is the pump intensity and  $\alpha$  the optical cavity escape rate. Occupancies are calculated in §3.

requirement to calculate the trajectory evolution for a very large number of thermal events and a very small number of photonic events, which are the only ones to produce outputs relevant for laser physics.

As a result, the Markov chain describes a dynamic system with two very different time scales: a slow-changing chain with photonic events and pumps, and a fast-moving chain with thermal events. A known result in Markov theory shows the possible splitting of the initial chain into two subchains [25,26]. In between two photonic events, the physical system evolves towards the canonical ensemble since no particle exchange occurs outside whichever band. Assuming each band becomes stationary during this time, it is possible to make an efficient shortcut in the initial chain by calculating the thermal evolution of the system separately. The major advantage is a possible analytical resolution of this thermal chain producing the electron presence in each level, at fixed temperature, with  $n$  as sole parameter. A state is therefore perfectly defined only with the couple  $(n, m)$  and considerable improvement in CPU time consumption is expected. The whole set of possible transitions into and out of this state is plotted on Fig. 2. The probability of having an electron distribution that allows absorption is then the product of the probability of having an electron in the VBLL and a hole in the CBLL. Conversely, the probability of being in a distribution that favors emission is the product of the probability of having an electron in the CBLL and a hole in the VBLL.

We call this particular model the canonical Markov model (CMM) of a SC laser, and it relies at each step on the knowledge of the photon number  $m$  and CBLL and the VBLL occupancies. The calculation of all values for all rates in Fig. 2 is now complete, thus defining all transitions of the CMM. It should be compared to the other two models describing laser dynamics with a birth-death Markov process [20,29]. The major difference lies in the calculation of the rates, which in our case takes into account the occupation of the laser level and not only the total number of excited states. It is therefore expected to be more in line with semiconductor lasers. Moreover, it should be noted that Rice et al. used  $\beta$  as a consistent emission rate multiplier, unlike us and Roy-Choudhury et al. For nanolasers with a limited number of photonic events, this way is more adapted because it counts all possible events, which certainly allows a better account of the optical noise via quantum jumps [12].

In addition to Monte Carlo simulation, working with the CMM further offers the possibility to solve numerically for the stationary distribution of the process and other measures of interest. The stationary distribution of the state  $(n, m)$ , say  $\pi(n, m)$  is known to satisfy a set of linear equations that can be summarized in the matrix equation

$$\pi R = 0. \quad (1)$$

Here,  $R$  is the rate matrix formed of the transition rates, completed on the diagonal with the negative of the total rate out of each state. Solving

numerically for this equation first requires truncating the infinite state space by ignoring transitions leading to states with  $m > M$ , for an appropriate  $M$ . This  $M$  should be chosen so that the probability mass thus ignored is less than some precision requirement. The resulting linear system is still large but manageable using iterative methods, see e.g. Ref. [30].

### 3. Theory

The purpose of this model section is to bridge the gap between the two models, the MMM and the CMM. Finally the occupancies at lasing levels and a convergence criterion will be exhibited.

#### 3.1. Microscopic model and occupancies

In the microscopic description of a SC laser, a state is represented as a couple  $(o, m)$ , where  $m$  is the number of photons in the cavity, and  $o = (o_1, \dots, o_{2B})$  is a sequence of 0 or 1 which describes the repartition of the  $B$  electrons among the  $2B$  energy levels:  $o_k = 1$  means that the  $k$ th level is occupied. The indices of the CBLL and VBLL are denoted by  $L$  and  $\ell$  respectively. These levels are assumed to be symmetric with respect to the energy gap between the two bands, that is,  $\ell + L = 2B + 1$ .

The Markov chain that models the laser operation is defined by the transitions and rates given in Table 1. For a transition from state  $s = (o, m)$  to state  $s' = (o', m')$  to exist, states  $s$  and  $s'$  must satisfy  $o_k = o'_k$  for all levels  $k$  except those two indicated in the table.

It will be shown elsewhere by using classical criteria (e.g. Chap II in Ref. [31]) that the latter chain is ergodic, and that the stationary occupancies  $n_k$  satisfy the following symmetry. For each level index  $k \in [1, 2B]$ ,

$$n_k = 1 - n_{2B+1-k}. \quad (2)$$

#### 3.2. Markov chains without thermal transitions

Almost all occurrences of transition in the MMM are thermal ones. This is because their rates are by far the greatest ones: In Table 1 they are labeled  $p$  and  $q$ , with  $q \equiv \exp(-\delta/k_B T)$  the Boltzmann factor that includes the lattice temperature  $T$ . Nonetheless laser operation is mainly described by the *other* transitions. Thermal transitions are only useful in determining whether the lasing and pumping levels are occupied or not, and they neither change the number  $m$  of photons nor the number  $n$  of electrons in the CB. Therefore, we may get rid of them by clustering the states with same  $n$  and  $m$  values, and weighting transition rates by coefficients  $\gamma(n, m)$  in  $(0, 1]$  accounting for the likelihood of fitting microscopic states. A state of the corresponding Markov chain is then described by a couple  $(n, m)$ , and the transitions are given in Table 2.

For instance,  $\gamma_E(n, m)$  is interpreted as the probability of being in a

**Table 1**  
Transitions in the microscopic model.

	Conditions on $s, s'$	Rate $R_{ss'}$
$m' = m$	$o_1 = o'_{2B} = 1$ $o'_1 = o_{2B} = 0$	$J$
$m' = m - 1$	$o_\ell = o'_L = 1$ $o'_\ell = o_L = 0$	$m$
$m' = m + 1$	$o_L = o'_\ell = 1$ $o'_L = o_\ell = 0$	$m + \beta$
$m' = m$	$o_L = o'_\ell = 1$ $o'_L = o_\ell = 0$	$1 - \beta$
$m' = m - 1$		$\alpha m$
$m' = m$	$o_k = o'_{k+1} = 1$ ( $k \neq 2B$ ) $o'_k = o_{k+1} = 0$ ( $k \neq B$ )	$p q$
$m' = m$	$o_k = o'_{k-1} = 1$ ( $k \neq 1$ ) $o'_k = o_{k-1} = 0$ ( $k \neq B + 1$ )	$p$

**Table 2**  
Transitions in the generic fast model.

Conditions on $s, s'$		Rate $R_{ss'}$
$m' = m$	$n' = n + 1$ ( $n < B$ )	$J\gamma_J(s)$
$m' = m - 1$	$n' = n + 1$ ( $n < B$ )	$m\gamma_A(s)$
$m' = m + 1$	$n' = n - 1$ ( $n > 0$ )	$(m + \beta)\gamma_E(s)$
$m' = m$	$n' = n - 1$ ( $n > 0$ )	$(1 - \beta)\gamma_E(s)$
$m' = m - 1$	$n' = n$	$am$

microscopic state allowing a coherent emission after the occurrence of a large number of successive thermal events in the previous chain. The clustered chain is called generic because coefficients  $\gamma$  may be chosen in various ways. It is also ergodic whatever the choice.

Using the Fermi-Dirac distribution for a given pseudo-Fermi level is the first trial we may think of. Since the  $\gamma$  coefficients do not depend on  $(n, m)$  anymore, except for the forbidden values of  $n$  at which they vanish, the coupling between  $n$  and  $m$  almost disappears. Separate numerical experiments have shown that the distribution of  $m$  is merely geometric in that case, which is not expected in correct laser operation. We next show why using the canonical distribution fits our goal better, and is easily implemented in the case of equidistant energy levels.

### 3.3. Canonical model and occupancies

In this subsection, we study the subchain of the MMM that only consists in thermal events. If the events in Table 1 are restricted to thermal ones, the induced chain only explores the finite subset  $S(n, m)$  of all states with  $m$  photons and  $n$  electrons in the CB, where  $m, n$  are given by the initial state. Once limited to  $S(n, m)$ , the chain is irreducible, reversible, and satisfies  $q^{u_s}R_{ss'} = q^{u_{s'}}R_{s's}$ . Here,  $u_s$  is the energy of state  $s = (o, m)$ , that is,  $u_s = \sum_k o_k$ , and  $R_{ss'}$  is either 0 or  $p$  or  $pq$ . Accordingly, the stationary distribution is the canonical one with respect to  $n$ . Assuming that the latter distribution is reached between two rare events of the microscopic chain, each  $\gamma$  coefficient may then be defined as the canonical probability of being in a state allowing the corresponding event.

Observe that the two bands evolve independently of each other, and that symmetry relation (2) still holds. Therefore, the  $\gamma$  coefficients may be written

$$\gamma_J = n_1^2, \quad \gamma_A = n_e^2, \quad \gamma_E = (1 - n_e)^2. \quad (3)$$

The validity of the CMM is supported by the fact that, in each band, the subchain converges very quickly. It is well-known that the speed of convergence to  $\pi$  of the distribution  $\pi_t$  at time  $t$  is controlled by the greatest negative eigenvalue  $\lambda$  of rate matrix  $R$ :

$$\|\pi_t - \pi\| \leq C \exp(\lambda t) \quad (4)$$

for a constant  $C$  depending on the chosen norm. Formal computations for small  $B$  values give

$$\lambda = -p[(1 - \sqrt{q})^2 + 4\sqrt{q} \sin^2(\pi/2B)] \quad (5)$$

independently of  $n$ . We conjecture this also holds for large  $B$ . Then, at room temperature,  $\lambda \approx -4 \times 10^{-4}p$  while reasonable  $p$  values are of magnitude  $10^5$ .

Let us finally turn to the effective computation of the  $\gamma$  coefficients in the CMM defined in (3). Occupancies  $n_1$  and  $n_e$  only depend on the number  $N = B - n$  of electrons in VB. We claim they may be computed recursively thanks to the following relation. Let  $k$  be any index in  $[1, B]$ ,

$$n_k(0) = 0, \quad n_k(N) = \frac{q^k(1 - q^N)}{(q^{N-1} - q^B)} [1 - n_k(N-1)]. \quad (6)$$

Since the value of  $n$  only turns to  $n \pm 1$  at each rare event of the MMM, the array  $n_k(N)$ ,  $N = 0 \dots B$ , may be precomputed so that updating with (6) is done immediately.

The proof of (6) goes as follows. Since canonical occupancies do not

depend on the ground energy value, we may consider the VB as a system of  $N$  fermions distributed on  $B$  levels of energy  $1 \dots B$ . It is well-known that the partition function of such a system reads as follows (see, e.g. Ref. [32])

$$Z(N) = q^{\frac{N(N+1)}{2}} \prod_{i=1}^N \frac{1 - q^{B+1-i}}{1 - q^i}. \quad (7)$$

Define  $W(N, u)$  as the number of configurations of the system whose energy is  $u$ , and  $M_k(N, u)$  as the number of them that have a particle at level  $k$ . Removing that particle leads to the following identity:

$$M_k(N, u) = W(N-1, u-k) - M_k(N-1, u-k). \quad (8)$$

Accordingly,

$$\begin{aligned} n_k(N) &= \sum_{u \geq 0} M_k(N, u) \frac{q^u}{Z(N)}, \\ &= q^k \frac{Z(N-1)}{Z(N)} [1 - n_k(N-1)], \end{aligned} \quad (9)$$

which simplifies to (6) by using (7).

## 4. Results

We will first concentrate on the cross-validation of the CMM versus the MMM, which was already used to predict ultimate noise performances of SC lasers [12] or the stability of a two-mode regime [23]. Second we will use the CMM to question the threshold definition in nanolasers already addressed by many authors [10,11,20,29].

### 4.1. Model differences

The comparison of the two models was done using the same physical parameters for both. We chose singlemode laser parameters similar to those of a previous mode competition analysis [23]. They are all given in Table 3 where a gain constant of  $1 \text{ ns}^{-1}$  is implicitly considered. This gain constant usually appears as a multiplier of the  $R_{ss'}$  rates for coherent transitions and was omitted in Tables 1 and 2 since its value was implicitly assumed to be the inverse of the time unit.

The very first comparison between the MMM and the CMM concerns the average photon count,  $\langle m \rangle$ , and variance,  $\text{var}(m)$ , within the cavity taken over 12 trajectories of the Monte Carlo simulation. Data are given in Table 4 and first two lines evidence similar although not identical output values. Since error margin estimates due to the reduced number of trajectories involved are far less than these differences, the physical origin of this discrepancy has to be found elsewhere. Average and variance for the CMM have been confirmed in the third line of the Table with the direct resolution of the steady-state solutions using (1) that are in perfect agreement with Monte Carlo results.

The MMM exhibits average and variance  $\approx 4\%$  lower than that of the CMM. This can only be achieved with a reduced effective pumping of about the same magnitude arising because of the intrinsic pumping mechanism in the MMM: a pump event requires the simultaneous presence of an electron in the highest level in CB and of a hole in the lowest level in VB. This configuration does not allow a new pump event until these levels have been released and occupied again. It is thus mandatory to wait to recover stationarity and allow a new pump event.

**Table 3**  
Common laser parameters for model comparisons.

Parameter	Value	Unit/Remark
$B$	801	
$\delta$	1	meV
$\beta$	1	
$\alpha$	0.6	$\text{ns}^{-1}$
$p$	50,000	$\text{ns}^{-1}$
$q$	0.962	@ $T = 300 \text{ K}$
$J$	35	$\text{ns}^{-1}$

**Table 4**  
Model comparisons.

Model	$\langle m \rangle$		var( $m$ )	
MMM	56.23	$\pm 0.14$	110.3	$\pm 2.5$
CMM	58.32	$\pm 0.09$	115.6	$\pm 2.2$
CMM <sup>a</sup>	58.333	$\pm 0.001$	115.04	$\pm 0.01$
MMM <sup>b</sup>	57.69	$\pm 0.16$	115.7	$\pm 2.4$

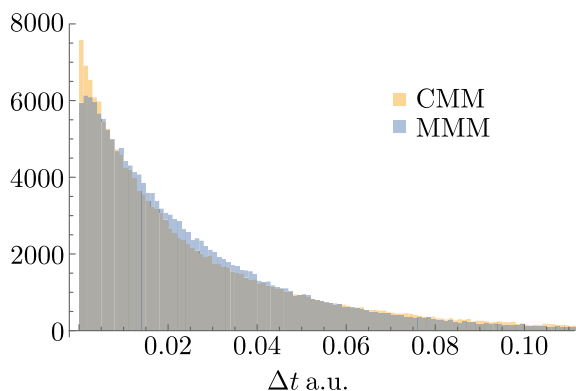
<sup>a</sup> As solved using (1).

<sup>b</sup> MMM with  $p = 200\,000\text{ ns}^{-1}$ .

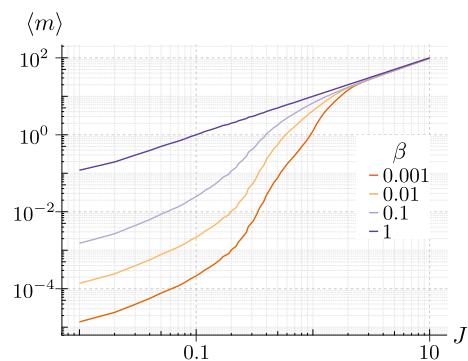
On the contrary in the CMM this time is considered infinitely short and the residual occupancy of the highest level in the CB (respectively, the lowest level in the VB) is almost negligible (respectively, extremely close to unity). The consequence is an effective pump rate lower than expected with the MMM that is strongly dependent on the thermalization parameter  $p$ .

This pump bias effect was confirmed in the framework of the MMM by reproducing the same simulations with a 4 times greater thermalization rate, *i.e.*  $p = 200\,000\text{ ns}^{-1}$ . The new average and variance, given in the fourth line of Table 4 are now very close to that of the CMM, but exact convergence would have required an even greater  $p$ -value leading to excessive CPU time consumption. This last characteristic is the biggest difference between the two models: where the MMM requires more than 1 h of calculation with  $p = 50\,000\text{ ns}^{-1}$  only 1 s is required with the CMM. This difference is further increased by a factor of 4 if  $p = 200\,000\text{ ns}^{-1}$ , making extensive simulations almost impracticable.

If the bottleneck encountered by the levels involved in laser pumping is responsible for the change in the mean value of  $m$ , a similar phenomenon may occur for the levels involved in coherent emission. Fig. 3 shows the histogram of waiting times between two emission events under Table 3 conditions for both the MMM and the CMM. As a result, the waiting time distribution is nearly exponential for the CMM. This is no longer the case if we consider the MMM whose distribution shows that it is unlikely that two very close emission events will occur. Again, some laser level bottlenecks impose a minimum waiting time between two emissions to fill the upper level and empty the lower level, a condition requiring at least two thermal events. Even if for very small lasers it seems impossible to have two very close photonic events in time, this is not the case for lasers of mesoscopic size where multiple emission sites exist. Consequently, if the MMM is well adapted to ideally small lasers the CMM is more in line with the physical reality of the achievable components. Moreover, it is still possible to encapsulate the emission within multiple exponential processes in series (Erlang distribution) in order to reproduce factually the impossibility of an almost simultaneous double emission without becoming non-Markovian. Such a technique will reproduce with the CMM the histogram of Fig. 3



**Fig. 3.** Histogram of the time interval between two emission events as deduced from trajectories calculated with the MMM and the CMM. Parameters are taken from Table 3.



**Fig. 4.** Light vs pump characteristics given by the CMM as a function of  $\beta$ . The external emitted intensity is  $\alpha \langle m \rangle$ .  $B = 10\,000$ ,  $\alpha = 0.1\text{ ns}^{-1}$ ,  $\delta$  and  $q$  are taken from Table 3.

obtained for the MMM, and will also preserve the extreme efficiency of the CMM simulation with all the subtlety of particle interactions and quantum jumps within the laser.

#### 4.2. Threshold

The question of the laser threshold has been revived in the last years because of the advent of very small lasers involving only a few emitters [11]. Far below (respectively far above) this threshold, the emission is predominantly incoherent (respectively coherent) and well understood in the corresponding framework whatever the overall size of the laser (number of emitters) or its  $\beta$ . For the first order output  $\langle m \rangle$ , our model produces light vs pump characteristics that are quite classic (see Fig. 4) with a typical ‘S’-shape whose bending is usually considered as the threshold. These curves are comparable to those obtained with similar models, the only notable difference being found in the positioning of the curvature which depends on pumping in Ref. [20] and not here, as was also the case in Ref. [29]. This discrepancy is due to whether or not  $\beta$  is taken into account in the stimulated emission rate, which is different between models as we noticed in §2.

Anyway for  $\beta = 1$ ,  $\langle m \rangle$  grows linearly with  $J$  according to a slope  $\alpha^{-1}$  which is imposed by the conservation of particles in our model. For lower  $\beta$ -values, the linear evolution is only present at high pump and drops off at  $J \approx 1\text{ ns}^{-1}$  with a depth all the more important as  $\beta$  is weak. On the contrary when  $\beta = 1$ , the laser seems ‘thresholdless’ but it is well-known that a distinction must always be made between incoherent spontaneous emission at very low pump and coherent stimulated emission at high pump [33]. Whatsoever, the bend in these light-pump characteristics is a poor estimator of the threshold because of the difficulty of defining it at high  $\beta$ -values and of its inaccuracy.

Another threshold definition involves the noise via the Fano factor, which is defined as  $\mathcal{F} = \text{var}(m)/\langle m \rangle$ . The larger  $\mathcal{F}$ , the more important the photon noise produced by the laser, a situation that occurs preferably at the threshold where the competition between spontaneous and stimulated emission is most important. The maximum of  $\mathcal{F}$  is therefore a better threshold estimator [20] also able to distinguish between the transition from incoherent to coherent emission in ideal  $\beta = 1$  devices. Fig. 5 plots  $\mathcal{F}$  as a function of pumping in log-log scale for various  $\beta$ -values. We clearly see an increase of the maximum of  $\mathcal{F}$ , together with a shift of the position of this maximum towards higher  $J$ , when  $\beta$  decreases. As a result the corresponding threshold is expected to increase, a subtlety difficult to catch in a linear scale with a similar model (see Ref. [29]). At the same time, the logarithmic scale also reveals that  $\mathcal{F}(J)$  looks like a double spike system with a shoulder around  $J \approx 0.5\text{ ns}^{-1}$ .

The depinning of the number of carriers in CB, the fact that  $\langle n \rangle$  exceeds its clamped value at high pumping, is shown in Fig. 6 for various  $\beta$  conditions. In a grand-canonical view and provided that the

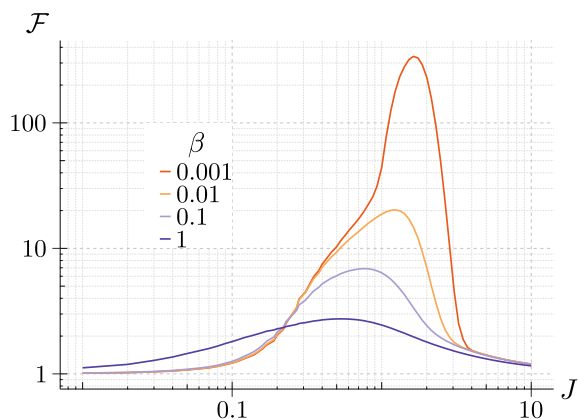


Fig. 5. Fano factor  $\mathcal{F}$  vs pump  $J$  as a function of  $\beta$ .  $B = 10\,000$ ,  $\alpha = 0.1 \text{ ns}^{-1}$ ,  $\delta$  and  $q$  are taken from Table 3.

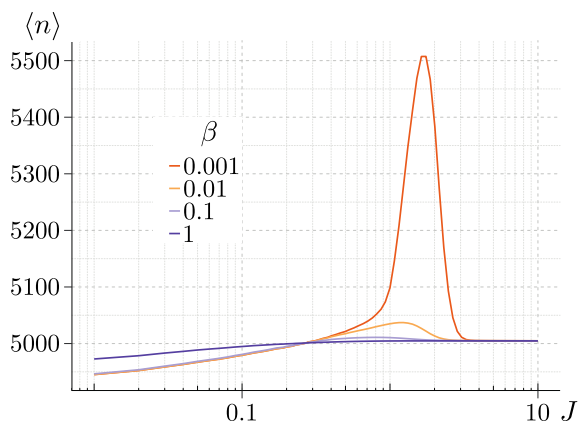


Fig. 6. Average number of electrons in CB  $\langle n \rangle$  vs pump  $J$  as a function of  $\beta$ .  $B = 10\,000$ ,  $\alpha = 0.1 \text{ ns}^{-1}$ ,  $\delta$  and  $q$  are taken from Table 3.

CB size is large enough,  $\langle n \rangle$  will be the Fermi level. It increases regularly starting from low pump values and is fixed for higher pumps, as can be expected with a classical laser theory. Between these two extremes an excess of carriers that can be very large (up to 10% at  $\beta = 10^{-3}$  in Fig. 6) appears as a peak for the expected threshold pumping values. Such a depinning of the Fermi level was already observed by Roy-Choudhury et al. [13]. It was shown responsible of a delayed threshold with respect to the predictions done with rate equations because of the quantum correlations between photon and electron populations.

Fig. 7 represents a small fraction of a trajectory calculated near the threshold as determined by the maximum of  $\mathcal{F}$  for  $\beta = 10^{-3}$ . The photon trace shows that the laser alternates between off- and on-states. During on-states the  $m$  and  $n$  instantaneous photon and electron in CB counts remain fairly stable until the last photon in the cavity has escaped. The average of  $n$  during this on-state then corresponds to the

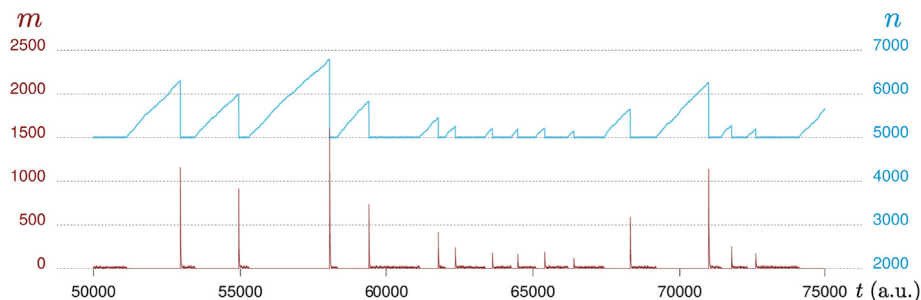


Fig. 7. Small portion of a time trajectory for electrons in CB,  $n$ , and photons stored in the cavity,  $m$ , calculated at  $J = 1.7 \text{ ns}^{-1}$ , the maximum of  $\mathcal{F}$  shown in Fig. 5.  $B = 10,000$ ,  $\beta = 10^{-3}$ ,  $\alpha = 0.1 \text{ ns}^{-1}$ ,  $\delta$  and  $q$  are taken from Table 3.

clamped value reached at high pumping (see Fig. 6), suggesting a coherent laser during this period. The laser then goes off for an average duration of  $\beta^{-1}$ . The lack of photons causes an increase of  $n$  at a speed of  $J - 1 + \beta$ . This extinction period ends when a photon is emitted spontaneously, acting as a trigger of an exponential growth of  $m$  because of stimulated emission, stopping when the  $n$  excess has been consumed. At that point the cavity reaches its maximum photon count that subsequently decreases as governed by the cavity lifetime  $\alpha^{-1}$ . In the trace of Fig. 7 the nanolaser then switches in continuous mode until the last photon exits the cavity again, thereby shifting once more in the pulse regime. Obviously this pulse regime is very similar to gain-switching but it is here triggered by the quantum nature of photon instead of the speed of a gain change driven by carrier injection. It is called quantum gain-switching (QGS) in the sequel. Optical pulses are very asymmetrical and their shape is close to a single-side decreasing exponential, of height proportional to the accumulated  $n$ , and of width at half maximum identical for all:  $\approx \ln 2/\alpha$ .

If  $J$  is a little bit lower, for instance close to the shoulder of  $\mathcal{F}$  identified on Fig. 5, the trajectory is now composed almost completely of QGS peaks. The companion regime of coherent emission at pulse end progressively vanishes because of the too weak pump that cannot maintain population inversion. When  $J$  decreases the average QGS peak amplitude is approximately maintained while their number decreases. If  $J$  diminishes further, the number of QGS peaks still decreases but now their average amplitude is also reduced. We assume that the origin of the shoulder observed on Fig. 5 is that the amplitude of the peaks is first maintained.

On the contrary, if  $J$  exceeds the value reached at Fano maximum, the number of QGS peaks per time unit decreases. As a major result, the Fano maximum exactly corresponds to the most important mixing between the two regimes, in perfect agreement with the noise maximum. The close inspection of the trajectory straightforwardly explains the overshoot on  $n$  shown in Fig. 6 because the occurrence of QGS peaks requires a non-stationary carrier concentration in excess as compared to cw. The greater the QGS peak amplitude and their number, the higher the corresponding overshoot.

An other illustration of the change of regime at threshold is given in Fig. 8.  $P(n)$ , the probability of having  $n$  electrons in the CB, is plotted as a function of  $J$  for  $\beta = 10^{-3}$ . These data were obtained by the matrix resolution of (1). At low  $J$ -values,  $P(n)$  is a single peak whose maximum position shifts towards higher values:  $n = 4994$  at  $J = 0.2 \text{ ns}^{-1}$ ,  $n = 5022$  at  $J = 0.5 \text{ ns}^{-1}$  and  $n = 5094$  at  $J = 1 \text{ ns}^{-1}$ . Starting from  $J = 1 \text{ ns}^{-1}$ , a second maximum appears at the fixed position  $n = 5006$  corresponding of the birth of the cw coherent emission. Conversely the other peak vanishes. In between, *i.e.* for example at  $J = 1 \text{ ns}^{-1}$ , the  $n$  distribution is bimodal, thus reflecting the coexistence of the two different laser regimes at the cw threshold. Notice that at  $J = 1.7 \text{ ns}^{-1}$ , where the maximum of  $\mathcal{F}$  happens,  $P(n)$  exhibits a unique peak, but remains over  $5 \times 10^{-4}$  for the whole scale of  $n$  in Fig. 8. When  $J$  increases further, this peak continues to grow while the secondary high  $n$  peak disappears.

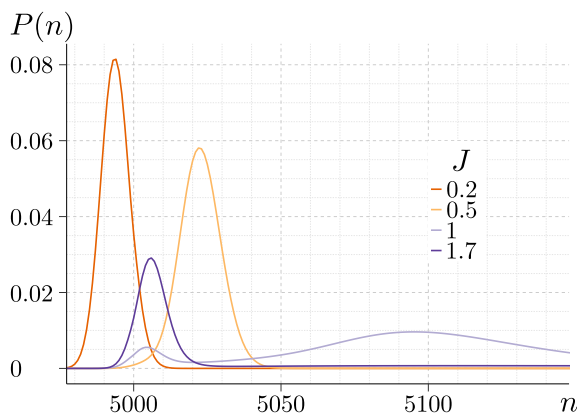


Fig. 8. Probability of presence of  $n$  electrons in the CB for various pump values  $J$ .  $B = 10,000$ ,  $\beta = 10^{-3}$ ,  $\alpha = 0.1 \text{ ns}^{-1}$ ,  $\delta$  and  $q$  are taken from Table 3.

## 5. Conclusion

We built a Markov model of the semiconductor nanolaser to take into account the realistic distributions of the carriers in the conduction and valence bands. This is achieved using evenly-spaced energy levels that allow both carrier-carrier interactions and semiconductor physics-compliant occupations. Although precise, the complete Markov chain of such a microscopic model (the MMM) is however very inefficient for calculating the temporal evolution of successive photonic events that occur in a nanolaser since most of the time is spent moving electrons to reach the thermalization of the bands.

Assuming that carrier and photonic events have very different timescales, the Markov chain can be divided and the two parts are treated separately. We proposed an analytical model for carriers based on the hypothesis of thermalization in the canonical ensemble between any two photonic events. This results in a direct assessment of laser level occupancy that depends only on the number of carriers in the band and allows another model (the CMM) to be built with a considerable improvement in CPU time of at least a few thousands. The two models compare favorably, with the possible exception of a bottleneck that appears at laser levels with the MMM and is not reproduced with the CMM. If physically relevant, this effect can be effortlessly added to the CMM without leaving Markov's framework through Erlang distributions. This will be discussed elsewhere. The extreme speed of the new model illustrated the physical characteristics during the transition to lasing of nanolasers as a function of  $\beta$  that would not have been obtained with the MMM because of its extreme CPU time requirement.

The numerical results underline the difficulty of precisely defining the threshold of nanolasers, either with the first-order output (intensity) or second-order output (Fano factor), as has already been pointed out in the literature [10,13,27]. This is particularly well illustrated by the temporal traces obtained from Monte Carlo evolutions as in Fig. 7 and which present a double behavior in time: first a fairly stable optical cw operation associated with a clamped carrier number until the last photon in the cavity disappears, and then a pulse regime where the number of carriers increases linearly because of the pump until an intense optical pulse is emitted. We have called this regime “quantum gain switching” because it is triggered by the first spontaneous photon emitted in the laser mode. Such a description is closely linked to the very small size of the laser as it is only possible with a very small number of excited photons and carriers. We also believe that this behavior is responsible for the asymmetric increase in the Fano factor observed during the laser transition in Fig. 5 which was also shown in Ref. [13], and the bimodal distribution of the number of carriers in the CB where we identified each of the two operating regimes by the two observed bumps.

The simultaneous existence of these two laser operating modes

makes it impossible to precisely define the beginning of the coherent cw emission since it is mixed with numerous pulses which consume part of the excited states. No very precise pump value corresponding to the threshold can be defined for a nanolaser, and moreover the transition is all the less defined as  $\beta$  is higher, a mandatory condition for a real nanolaser with a very low gain. An advantage will emerge if we are able in the future to control the beginning of the emission of these pulses, a nanolaser thus becoming a very compact and low power source of coherent optical pulses, useful for example in quantum information. Nevertheless, the control trigger has yet to be invented.

## Acknowledgments

Authors acknowledge G. L. Lippi for fruitful discussions. This work was supported by the french Agence Nationale de la Recherche under the Swiss-French IDYLIC contract ANR-15-CE24-0034.

## References

- [1] K.J. Vahala, Optical microcavities, *Nature* 424 (6950) (2003) 839–846.
- [2] R. Chen, T.-T.D. Tran, K.W. Ng, W.S. Ko, L.C. Chuang, F.G. Sedgwick, C. Chang-Hasnain, Nanolasers grown on silicon, *Nat. Photon.* 5 (3) (2011) 170–175.
- [3] D.J. Bergman, M.I. Stockman, Surface plasmon amplification by stimulated emission of radiation: quantum generation of coherent surface plasmons in nanosystems, *Phys. Rev. Lett.* 90 (2003) 027402.
- [4] M. Noginov, G. Zhu, A. Belgrave, R. Bakker, V. Shalaev, E. Narimanov, S. Stout, E. Herz, T. Suteewong, U. Wiesner, Demonstration of a spaser-based nanolaser, *Nature* 460 (7259) (2009) 1110–1112.
- [5] K. Ding, C. Ning, Metallic subwavelength-cavity semiconductor nanolasers, *Light Sci. Appl.* 1 (7) (2012) e20.
- [6] S. Wang, X.-Y. Wang, B. Li, H.-Z. Chen, Y.-L. Wang, L. Dai, R.F. Oulton, R.-M. Ma, Unusual scaling laws for plasmonic nanolasers beyond the diffraction limit, *Nat. Commun.* 8 (1) (2017) 1889.
- [7] T. Wang, D. Aktas, O. Alibart, É. Picholle, G. P. Puccioni, S. Tanzilli, G. L. Lippi, Nontrivial Photon Statistics in Small Scale Lasers, *ArXiv e-prints* arXiv:1710.02052.
- [8] L.A. Coldren, S.C. Corzine, Diode Lasers and Photonic Integrated Circuits, John Wiley & Sons, New York, 1995.
- [9] C. Gies, J. Wiersig, M. Lorke, F. Jahnke, Semiconductor model for quantum-dot-based microcavity lasers, *Phys. Rev.* 75 (1) (2007) 013803.
- [10] W.W. Chow, F. Jahnke, C. Gies, Emission properties of nanolasers during the transition to lasing, *Light Sci. Appl.* 3 (8) (2014) e201.
- [11] G.P. Puccioni, G.L. Lippi, Stochastic simulator for modeling the transition to lasing, *Optic Express* 23 (3) (2015) 2369–2374.
- [12] L. Chusseau, J. Arnaud, Monte-Carlo simulation of laser diodes sub-poissonian light generation, *Opt. Quant. Electron.* 34 (2002) 1007–1023.
- [13] K. Roy-Choudhury, S. Haas, A. Levi, Quantum fluctuations in small lasers, *Phys. Rev. Lett.* 102 (5) (2009) 053902.
- [14] H. Haken, Laser Theory, Springer Berlin Heidelberg, Berlin, Heidelberg, 1984, pp. 99–173 Ch. Fully quantum mechanical solutions of the laser equations.
- [15] M. Elk, Numerical studies of the mesoscale, *Phys. Rev.* 54 (1996) 4351–4358.
- [16] K. Roy-Choudhury, A. Levi, Quantum fluctuations and saturable absorption in mesoscale lasers, *Phys. Rev.* 83 (4) (2011) 043827.
- [17] F. Jahnke, S.W. Koch, Many-body theory for semiconductor microcavity lasers, *Phys. Rev.* 52 (2) (1995) 1712.
- [18] M.O. Scully, W.E. Lamb Jr., Quantum theory of an optical maser I. general theory, *Phys. Rev.* 159 (1967) 208–226.
- [19] R. Loudon, The Quantum Theory of Light, Clarendon Press, Oxford, 1983.
- [20] P.R. Rice, H. Carmichael, Photon statistics of a cavity-QED laser: a comment on the laser–phase-transition analogy, *Phys. Rev.* 50 (5) (1994) 4318.
- [21] A. Lebreton, I. Abram, N. Takemura, M. Kuwata-Gonokami, I. Robert-Philip, A. Beveratos, Stochastically sustained population oscillations in high- $\beta$  nanolasers, *N. J. Phys.* 15 (3) (2013) 033039.
- [22] D.T. Gillespie, Markov Processes: an Introduction for Physical Scientists, Academic Press, San Diego, 1992.
- [23] L. Chusseau, F. Philippe, F. Disanto, Monte Carlo modeling of the dual-mode regime in a quantum well and quantum dot semiconductor lasers, *Optic Express* 22 (5) (2014) 5312–5324.
- [24] J. Arnaud, J.M. Boé, L. Chusseau, F. Philippe, Illustration of the Fermi-Dirac statistics, *Am. J. Phys.* 67 (1999) 215–221.
- [25] H.A. Simon, A. Ando, Aggregation of variables in dynamic systems, *Econometrica* 29 (2) (1961) 111–138.
- [26] P.J. Courtois, Error analysis in nearly-completely decomposable stochastic systems, *Econometrica* 43 (4) (1975) 691–709.
- [27] T. Wang, G.P. Puccioni, G.L. Lippi, Dynamical buildup of lasing in mesoscale devices, *Sci. Rep.* 5 (2005) 15858.
- [28] F. Philippe, J. Arnaud, L. Chusseau, Statistics of Non-interacting Bosons and Fermions in Micro-canonical, Canonical and Grand-canonical Ensembles: a Survey, *arXiv math-ph/0211029* (2003).
- [29] K. Roy-Choudhury, A.F.J. Levi, Quantum fluctuations in very small laser diodes, *Phys. Rev. A* 81 (2010) 013827.



- [30] W.J. Stewart, *Introduction to the Numerical Solution of Markov Chains*, Princeton University Press, 1994.
- [31] S. Asmussen, *Applied Probability and Queues* vol. 51, Springer Science & Business Media, 2008.
- [32] G.E. Andrews, *The Theory of Partitions* vol. 2, Cambridge University Press, 1998.
- [33] H. Yokoyama, Physics and device applications of optical microcavities, *Science* 256 (5053) (1992) 66–70.