A Fokker-Planck-based Monte Carlo method for electronic transport and avalanche simulation in single-photon avalanche diodes

Remi Helleboid, Denis Rideau, Isobel Nicholson, Jeremy Grebot, Bastien Mamdy, Gabriel Mugny, Marie Basset, Megan Agnew, Dominique Golanski, Sara Pellegrini, et al.

A Fokker-Planck-based Monte Carlo method for electronic transport and avalanche simulation in single-photon avalanche diodes.

Rémi Helleboid, Denis Rideau, Isobel Nicholson, Jeremy Grebot, Bastien Mamdy, Gabriel Mugny, Marie Basset, Megan Agnew, Dominique Golanski, Sara Pellegrini, Jérôme Saint-Martin, Marco Pala, Philippe Dollfus

***Abstract*—We present an efficient simulation method for electronic transport and avalanche in single-photon avalanche diodes (SPAD). Carrier transport is simulated in the real space using a particle Monte Carlo approach based on the Fokker-Planck point of view on an advection-diffusion equation, that enables us to reproduce mobility models, including electric fields and doping dependencies. The avalanche process is computed thanks to impact ionization rates implemented using a modified Random Path Length algorithm. Both transport and impact ionization mechanisms are computed concurrently from a statistical point of view, which allows us to achieve a full multi-particle simulation. This method provides accurate simulation of transport and avalanche process suitable for realistic three-dimensional SPADs, including all relevant stochastic aspects of these devices, together with a huge reduction of the computational time required, compared to standard Monte Carlo methods for charge carrier transport. The efficiency of our method empowers the possibility to precisely evaluate SPADs figures of merit and to explore new features that were untrackable by conventional methods.**
An extensive series of comparisons with experimental data on state-of-the art SPADs shows a very good accuracy of the proposed approach.

***Index Terms*—Advection-Diffusion Monte Carlo (ADMC), avalanche breakdown probability, electronic avalanche, electronic transport, drift-diffusion, jitter, Monte Carlo simulation, single-photon avalanche diode (SPAD), technology computed aided design (TCAD), Fokker-Planck**

I. Introduction

Single-photon avalanche diodes are avalanche photodiodes, operated above breakdown voltage, in which a single photon absorption can lead to a self-sustained electron-hole avalanche in the junction. These devices are used in many applications such as Fluorescence Lifetime Imaging Microscopy [1], Time-of-Flight (TOF) and Light Imaging Detection and Ranging (LiDAR) [2], [3]. Predictive numerical simulation of SPAD behavior is a key step to improve the performance of these devices in terms of output parameters such as Photon Detection Efficiency (PDE), timing jitter, afterpulsing and dark-count-rate [4].

A great deal of effort has already been dedicated to finding accurate methods to simulate the SPAD operation and to extract reliable electrical characteristics of a given device architecture without the need of expensive and time-consuming manufacturing and characterization. The state-of-the-art methodologies rely on solving the Boltzmann transport equation (BTE) by means of particle Monte Carlo simulation in the phase-space, either under frozen field to investigate the details of avalanche process [5], [6] or self-consistently coupled to Poisson's equation to study all stages of SPAD operation [7]. This stochastic approach of transport can reproduce the statistical behavior of the SPAD, but it is known to be computationally very intensive, which makes it unsuitable for SPADs with sizes of several micrometers. To address this issue, many empirical models have been developed to extract SPAD's characteristics [8]–[12]. While these models have proven themselves to be accurate for simple architectures, they are not able to reproduce the intrinsic statistical behavior of SPAD devices. Recently, some attempts to address a statistical calculation with a reduced simulation time have been made [13], [14], but at the cost of strong simplifications such as the need to artificially split the device into an absorption and a multiplication region. These models assume that the electric field is constant in the multiplication region and treat the avalanche as a one-dimensional process. Hence, such methods cannot succeed in simulating state-of-the-art SPAD devices in which the electrostatics is highly non-uniform and bent in order to increase the active absorption volume [15].

This work aims at gathering efficient empirical models within a stochastic Monte Carlo method, able to simulate accurately and efficiently the statistical figures of SPADs of any architecture, without any simplification of the device geometry.

Our approach relies on interpreting a transport advection-diffusion equation as a Fokker-Planck equation and to solve it for each particle with a Monte Carlo method operating only in the real space. Concurrently, the impact ionization process is considered through a probabilistic Random Path Length algorithm, which randomly creates new particles, according to an ionization coefficient model.

The coupling of these two stochastic numerical processes leads to a multi-particle simulation that allows us to extract figures of interest for the SPAD operation such as avalanche breakdown probability and timing jitter.

This work was supported by the French "Association Nationale Recherche Technologie" ANRT and the French GeSPAD Project of the Agence Nationale de la Recherche (ANR-20-CE24-0004).

Corresponding author: Rémi Helleboid (email: remi.helleboid@gmail.com).

Rémi Helleboid, Jérôme Saint-Martin, Marco Pala and Philippe Dollfus are with the Université Paris–Saclay, CNRS, Centre de Nanosciences et de Nanotechnologies, Palaiseau, France.

Rémi Helleboid, Denis Rideau, Jeremy Grebot, Marie Basset, Dominique Golanski and Gabriel Mugny are with STMicroelectronics, Grenoble and Crolles, France.

Isobel Nicholson, Megan Agnew and Sara Pellegrini are with STMicroelectronics, Edinburgh, United Kingdom.

Rémi Helleboid, is with the CEA LETI, Grenoble, France.

The strength and accuracy of our method is underlined and confirmed by comparing numerical results with characterization results for state-of-the-art SPADs. In what follows, for clarity purposes, our Monte-Carlo approach will be referred to as the *Advection-Diffusion Monte Carlo* (ADMC).

II. Model

A. Transport model

Carrier transport in semiconductor devices has been widely studied through the Boltzmann transport equation, which describes the evolution of the density of carriers in real and phase spaces:

$$\frac{\partial f}{\partial t}(\mathbf{r},\mathbf{k},t) + \mathbf{v}\cdot\nabla_{\mathbf{r}} f + \frac{1}{\hbar}\mathbf{F}\cdot\nabla_{\mathbf{k}} f = \mathcal{Q}_{coll} f \tag{1}$$

where $f(\mathbf{r},\mathbf{k},t)\mathbf{dr}\,\mathbf{dk}\,dt$ is the probability of finding a particle, at time t, position $\mathbf{r}$, with a wave vector $\mathbf{k}$ within a small volume $\mathbf{dr}\times\mathbf{dk}\times dt$. $\mathbf{v}$ is the group velocity at position $\mathbf{r}$ and time t, $\mathbf{F}$ is the total external force applied to the carriers and $\mathcal{Q}_{coll}$ is the collision integral operator. Although this equation was very successful to model all regimes of semi-classical transport, its integration by the means of BTE Monte Carlo methods is known to be computationally heavy, which may prevent its use in many-particle simulations of large devices.
The complexity of this phase and real space equation can be lowered by integrating over all the possible $\mathbf{k}$ vectors, in the reciprocal space. This leads to the balance equation, which can, at the cost of some simplifications be turned into the well-known Jacoboni drift-diffusion equation [16]:

$$\frac{\partial f}{\partial t}(x,t) = -v_0\frac{\partial f}{\partial x}(x,t) + D_0\frac{\partial^2 f}{\partial x^2}(x,t) \tag{2}$$

where v_0 is the carrier drift velocity, D_0 the carrier diffusion coefficient and $f(x,t)$ the carrier density probability, said differently, if the studied system is made of N_e electrons, then the local charge density can be extracted as:

$$n(x,t) = N_e f(x,t)$$

This equation (2) is valid for a bulk semiconductor with constant doping level and uniform electric field, in a one-dimensional simulation. Yet, it can be generalized to a three-dimensional device simulation with varying-in-space doping level and electric field, and then becomes:

$$\frac{\partial f}{\partial t}(\mathbf{r},t) = -\mathrm{div}(\mathbf{v}(\mathbf{r},t)f(\mathbf{r},t)) + \mathrm{div}(\mathbb{D}(\mathbf{r},t)\nabla f(\mathbf{r},t)) \tag{3}$$

with $\mathbf{v}(\mathbf{r},t)$ and $\mathbb{D}(\mathbf{r},t)$, the space and time dependent drift-velocity and diffusion coefficient at, respectively, position $\mathbf{r}$ and time t.
The function f in (2) and (3) is basically the same as in (1) but the information on the wave vector $\mathbf{k}$ is lost, because f was integrated over all the reciprocal space.
To be complete, one must add initial and boundary conditions to this equation. The initial condition can be any density probability function, which represents the distribution of charges at the beginning of the simulation. The boundary conditions describe the behavior of charges at specific boundaries of the simulation domain. They can be either reflecting boundary conditions, for example at interfaces with insulators, or absorbing/emitting boundary conditions at metallic contact interfaces.
Usually, the drift-diffusion equation (3) is solved as a standard partial differential equation, by the means of deterministic numerical methods such as finite difference, finite element or box methods [17]. Those methods were explored in-depth to achieve good accuracy while keeping computational time within a reasonable range. Yet, in SPAD simulations, analysis of the statistical behavior is a crucial aspect of device operation, which cannot be achieved using deterministic methods.
Our model relies on the reformulation of the generalized equation (3) as a stochastic differential equation (SDE). Indeed, (3) can be interpreted as a particular form of a Fokker-Planck equation, which describes the spatio-temporal evolution of the probability density function of a continuous random process where both diffusive and dragging forces are present. A time-continuous random process is a stochastic process, i.e. a sequence of random numbers, indexed by a continuous set of values, it will be written, for example, $(\mathbf{Z}_t)_{t\geq 0}$. One can define the probability density of such a process at a given time t by a function $g(\cdot,t)$, for which $g(\mathbf{r},t)\mathbf{dr}$ is the probability for the process to occur in a small volume $\mathbf{dr}$ at position $\mathbf{r}$ at time t.
Said differently, if we define the probability law $\mathbb{P}_{\mathbf{Z}_t}$ of the position of the stochastic process of $(\mathbf{Z}_t)$ at time t, as:

$$\mathbb{P}_{\mathbf{Z}_t} : A \mapsto \mathbb{P}(\mathbf{Z}_t \in A) \tag{4}$$

with A a subset of $\mathbb{R}^3$, the probability density function of the random variable $\mathbf{Z}_t$, named $g(\cdot,t)$, associated to $\mathbb{P}_{\mathbf{Z}_t}$ is a function such as:

$$\mathbb{P}_{\mathbf{Z}_t}(A) = \int_A g(\nu,t)d\nu \tag{5}$$

We now expose the link between (3) and a time-continuous stochastic process. The so-called generalized drift-diffusion can be interpreted as a Fokker-Planck equation, for which solutions are temporal probability density function of the stochastic process $(\mathbf{X}_t)$, defined as follows [18]:

$$d\mathbf{X}_t = \mathbf{v}(\mathbf{X}_t,t)dt + \boldsymbol{\sigma}(\mathbf{X}_t,t)d\mathbf{B}_t \tag{6}$$

The process $\mathbf{X}_t$ can be directly interpreted as the position of a particle in real space. We thus have $\mathbf{X}_t$ the random variable representing the position of the particle at time t in the real space, $\mathbf{v}(\mathbf{X}_t,t)$ the drift velocity of the particle at its position and at time t, $\boldsymbol{\sigma}(\mathbf{X}_t,t)$ the diffusion matrix of the dynamics at position $\mathbf{X}_t$ and time t and $\mathbf{B}_t$ is the three-dimensional standard Brownian motion. The diffusion factor of (6) is linked to the electronic diffusion coefficient through the following relation:

$$\boldsymbol{\sigma}(\mathbf{X}_t,t) = \sqrt{2\mathbb{D}(\mathbf{X}_t,t)}.$$

The differential dynamics (6) can be reformulated under its integral form:

$$\mathbf{X}_t = \int_0^t \mathbf{v}(\mathbf{X}_s,s)ds + \int_0^t \boldsymbol{\sigma}(\mathbf{X}_s,s)d\mathbf{B}_s \tag{7}$$

The Brownian motion (B_t) is a time-continuous random process describing a succession of infinitesimal random steps that follow a centered Gaussian law with variance dt. Each step is

independent of the previous ones, which results in a continuous random walk with no bias and a variance: $\mathbf{Var}(B_t) = t$. It is possible to construct a Brownian motion of any dimension by stacking independent one-dimensional Brownian motions in a vector of the desired dimension.

Our model relies on the following proposition: a process defined by the SDE (6) or (7) has a probability density function, as defined in (5), that is a solution of the PDE defining the generalized drift-diffusion (3). Proof of this statement and of existence of such density function can be found in [19]. The particular form of the Fokker-Planck equation we are using can also be found under the name "Kolmogorov forward equation" which is often treated as a so-called "Kolmogorov backward equation with reverse time". For this proposition to be true, the boundary and initial conditions of the PDE and of the SDE must match. The most common initial condition for the stochastic process is to set deterministically its value at time $t = 0$, for example $Z_{t=0} = \mathbf{r}_0$, which amounts setting the initial condition of (3) to the Dirac delta function :

$$f : (\mathbf{r}, t = 0) \longmapsto \delta_{\mathbf{r}_0}(\mathbf{r}) \tag{8}$$

The process (6) can hence be seen as a random walk, with standard deviation $\boldsymbol{\sigma}(\mathbf{X}_t, t)$, which is linked to the diffusion of carriers, and biased by the drift velocity $\mathbf{v}(\mathbf{X}_t, t)$ due to external forces. This view enables us to see the random process $(\mathbf{X}_t)_{t\geq 0}$ as the position of a carrier over time.

We have thus built a path between the deterministic macroscopic description of the density of carriers (with the drift-diffusion equation) and a stochastic description of individual carriers submitted to drift and diffusion. This is of first importance in the context of SPAD simulation, where the modeling of the statistical behavior of the device, resulting from the stochastic behavior of carriers, is mandatory to estimate all their relevant figures of merit.

In the present study, we have assumed that the transport is isotropic, and thus that the velocity and the diffusion coefficient can be reduced to single scalar functions that depend on the electric field strength and the doping concentration, through carrier mobility. We show in section III that this choice is reasonable for device simulation of Silicon-based SPAD devices at room temperature. Yet, for other applications or simulation parameters, such as cryogenic temperatures, where the transport might be highly anisotropic, one can keep the diffusion coefficient under a full 3x3 matrix-form and adjust the velocity with respect to the crystallographic orientation of the electric field [20].

The drift-velocity for electrons and holes is computed by using the field, temperature and doping dependent model of Arora [21], together with a high-field saturation regime model from Canali's work [22].

Furthermore, the Einstein relation is assumed to hold:

$$D_{e,h}(\|\mathbf{F}\|, N_{total}) = \frac{\mu_{e,h}(\|\mathbf{F}\|, N_{total}) k_B T}{q} \tag{9}$$

with $D_{e,h}$ the diffusion coefficient, $\mu_{e,h}$ the charge mobility, for electron (e) and hole (h), k_B the Boltzmann constant, T the temperature and q the elementary charge. The mobility depends on the electric field strength $\|\mathbf{F}\|$ and the total doping concentration N_{total}. The diffusion coefficient thus depends on the local electric field and doping concentration, through the mobility. If necessary, one can straightforwardly replace the Einstein relation by a more sophisticated model for diffusion in order to take into account, for example, high-field effects [23], [24].

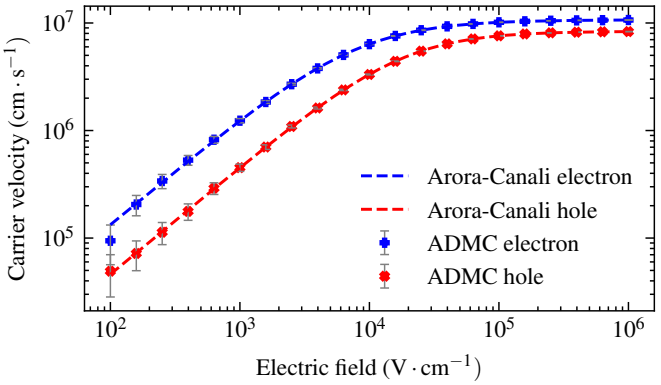

Fig. 1. Comparison of the carriers' drift velocity between input model [21], [22] and the Monte-Carlo results. Error bars represent the 99% confidence interval. The velocity is extracted by computing the distance over time ratio for 1000 electrons at different electric fields. The results show that our algorithm reproduces perfectly the input mobility model at any electric field.

The stochastic formulation of the drift-diffusion equation enables us to use stochastic numerical methods to simulate the transport of charges within the semiconductor material and thus gathering results over many simulations to extract the figures of interest. In this case the Monte-Carlo Euler-Maruyama method [18] is particularly well-suited, the algorithm is straightforward and goes as follows. We choose a constant time step dt, and we start from a given position $\mathbf{x}_0$. We call $\mathbf{X}^n$ the numerical approximation of $\mathbf{X}_t$ at time $n \cdot dt$. The Euler-Maruyama Monte-Carlo scheme is then given by:

$$\begin{cases} \mathbf{X}_0 = \mathbf{x}_0 \\ \mathbf{X}^{n+1} = \mathbf{X}^n + \mathbf{v}(\mathbf{X}^n, n \cdot dt)dt + \boldsymbol{\sigma}(\mathbf{X}^n, n \cdot dt)\mathbf{W}^n \end{cases} \tag{10}$$

where $(\mathbf{W}^n)_{n\geq 1}$ are random variables that follow a centered Gaussian distribution with variance dt. The transport properties of the carrier does not depend on the chosen time step dt, as the Brownian motion construction guarantee to have a standard deviation of $\sqrt{t}$ where t is the simulated time. Yet, because the Euler-Maruyama deterministic part is a numerical scheme of order one in time, the time step must stay relatively small, we found that a value of 10 fs was a good compromise between accuracy and computational time.

Our transport algorithm has been validated by checking that the random process results in the desired electron and hole velocities as displayed in figure 1, which demonstrates the accuracy of our method.

B. Impact ionization model

The peculiarity of our model is that the impact ionization modeling is fully coupled with the transport model. We have implemented a modified version of the "Random Path Length Algorithm" (RPLA) [25] [14].

In our model, the path integral (11) is computed on the fly, concurrently with the transport simulation, at each time step of the Monte-Carlo process. Thus, our model does not require the assumption that the device is divided into an absorption region and an avalanche region as in other previous works [13]. It allows us to simulate the full electron-hole avalanche process three-dimension space without the need for severe approximations on the device electrostatics in the so-called avalanche region.

Carrier multiplication is modeled through field-dependent rates of impact ionization [26], which is the number of impact ionization events a carrier experienced per distance unit.

The probability that an electron will not create an impact ionization event between two positions z and z' is given by :

$$P_{ii}(z'|z) = 1 - exp\left(-\int_z^{z'} \alpha(x)\,\mathrm{d}x\right) \tag{11}$$

where $\alpha(x)$ is the impact ionization coefficient at position x [27].

With the RPLA the occurrence of an impact ionization event is modeled by a random variable that follows the probability P_{ii}. A way of simulating such a probability law is to generate a random number variable r uniformly distributed between 0 and 1, then the integral expression (11) is computed along the carrier's trajectory, to get the position z for which $P_{ii}(z'|z) = r$. If such a z exists, then we consider that an impact ionization occurred at this location. The integral P_{ii} is then reset to zero for the particle and a new random number r is drawn. The created electron and hole will follow the same process, and might, eventually also experienced impact ionization.

In [13], the authors back-traced the timing of the impact-ionization event by assuming that the carrier's path in a so-called "avalanche-region" is a one-dimensional straight line at constant velocity. In our algorithm, we compute the integral $P_{ii}(z'|z)$ on the fly, consistently with the trajectory, and whenever $P_{ii}(z'|z)$ reaches or exceeds the random value r, an electron-hole pair is created at the current position of the parent carrier. The capacity of the multiplication model to be used consistently with the transport simulation is mandatory to accurately model the operation of state-of-the-art devices with highly bent electric fields, for which considering straight lines for carrier trajectories is very inaccurate [28]. The extraction of impact ionization coefficients in Silicon was performed by several authors either from measurements or from BTE Monte Carlo simulations. In this work we have used the experimental coefficients given by Van Overstraeten and De Man [29], but one can straightforwardly replace them by any other data.

Our multiplication modeling was tested by extracting the impact ionization coefficients back from simulation, at different electric fields. This is done by counting the number of impact ionization events on a given traveled distance and comparing the obtained coefficient with the values of the Van Overstraeten data used as an input, both for electrons and holes. The results are reported in figure 2 and show that our stochastic algorithm perfectly reproduces the input model of multiplication.

It was noted by several authors that modeling particle multiplication with impact ionization rates leads to the possibility

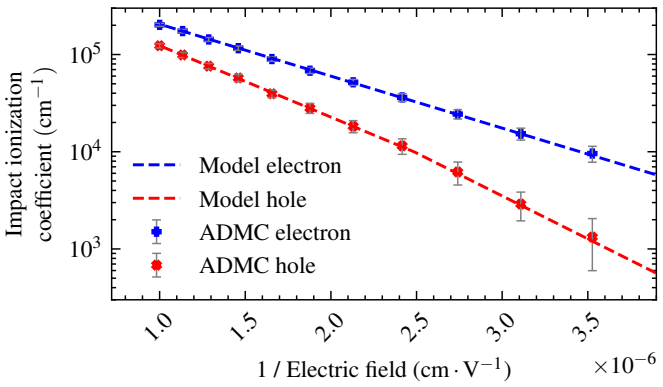

Fig. 2. Comparison of the carriers' impact ionization rates between input model [29] and the ADMC results. Error bars represent the 99% confidence interval. ADMC coupled with a RPLA accurately reproduces the input impact ionization model.

for a cold electron injected in a high electric field to create an impact ionization immediately after being injected. This is in contradiction with the fact that an impact ionization can occur only if the electron has gained enough energy from the electric field to be able to knock another electron out of the valence band [30], [31]. To accumulate this minimal threshold energy, the electron is required to travel a minimal mandatory distance in a high electric field. This length is called "dead-space" as no ionization can occur within this distance. Those considerations led to the creation of different models that take into account this energy threshold, among which the dead space model [32] and the effective field model [27] are the most used. The minimal energy required for a carrier to trigger an impact ionization was computed using theoretical considerations on the band structure and conservation of mass and momentum [33], [34]. It is also common to use effective values for this threshold energy, computed through conventional Monte-Carlo simulations and fitting procedures [35]. Those threshold energy values are often used as a fitting parameter, within an acceptable range of values. In this study we have used the following values for Silicon, the threshold energy for an electron to impact ionize is $E^e_{th} = 1.8$eV and $E^h_{th} = 2.4$eV for a hole, following [36].

For a particle injected with no energy in the system, we compute its energy gain from the electric field by extracting the electrostatic potential it traveled through, which is equivalent to integrating the electric field $\mathbf{F}$ over the particle path:

$$E_{gained} = \int_{Path} \left\|\overrightarrow{\mathbf{F}(z)}\right\| dz = \sum_{k=0}^{n} \left\|\overrightarrow{\mathbf{F}(\mathbf{X}^k)}\right\| \mathbf{v}(\mathbf{X}^k) \cdot dt \tag{12}$$

While E_{gained} is lower than the threshold energy E^e_{th} the impact ionization coefficient α in (11) is zero, and no impact ionization can occur. This non-local impact ionization model is called the "hard dead space model" because of the sharp transition between zero and non-zero impact ionization probability.

The need to avoid such a sharp transition, that is not observed in BTE Monte Carlo simulation results, has been discussed by several authors and is still under debate [37]. One might also

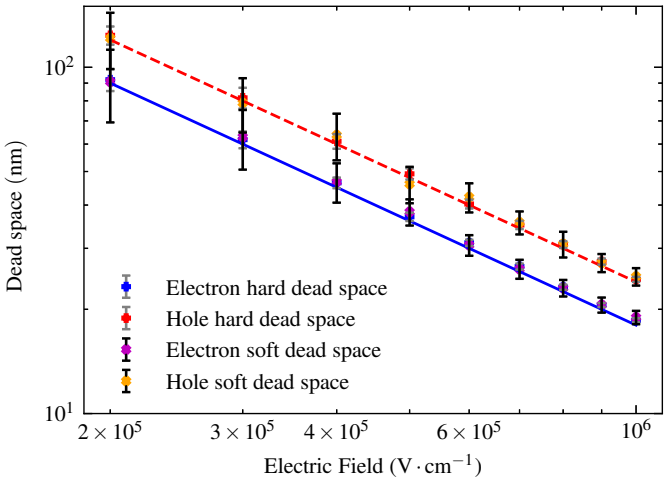

Fig. 3. Extracted dead space at various electric fields for hard and soft threshold models. Error bars represent the 99% confidence intervals. Plain blue and red lines represent theoretical values from formula (14) for electron and hole, respectively. Both models give values close to the theoretical calculation with an expected higher variability for the soft dead space model.

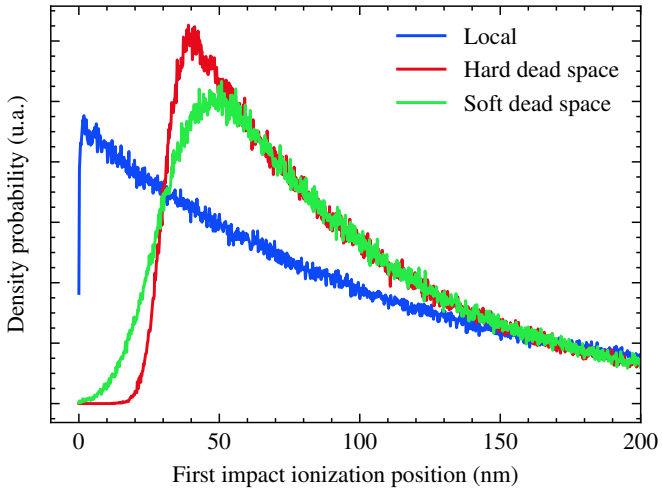

Fig. 4. Probability density function of the position of the first impact ionization for a cold electron injected in a constant electric field of 600kV/cm with three different impact ionization models. Each curve is made out of 1 000 000 single-particle simulation results.

want to take into account the residual energy of the carrier that engender the impact ionization, as well as the initial energy of the created electron and hole [38]. We did not focus on those topics as they would not significantly affect BrP or jitter result in SPADs. Yet we show that within our model, it is straightforward to implement models that would cope with those small-scale physical effects. Indeed, by randomly selecting the initial energy of the carriers on a given statistical distribution, one can retrieve a so-called "soft dead-space" [39] model. As an example, we have used a Gaussian distribution whose mean and standard deviation were set to roughly match results from BTE Monte Carlo.

When using a dead space impact ionization model, one must tweak the commonly used function for impact ionization rates for electrons α_e and holes α_h. Indeed, those values, as they come from measurements, intrinsically include the presence of the dead space. Following [40], the impact ionization coefficients for dead space models are transformed as such:

$$\alpha_{e,h}^{dead-space} = \frac{1}{\frac{1}{\alpha_{e,h}} + d_{e,h}} \tag{13}$$

where d_e and d_h are the dead space for electrons and holes, respectively.

The dead space can be extracted by simulating multiplication at different fields and looking at the distance traveled by an electron (d_e) or a hole (d_h) to reach its minimal threshold energy. We have checked that the obtained values matches the theoretical values for a constant electric field $\mathcal{E}$:

$$d_{e,h} = \frac{E_{th}^{e,h}}{\mathcal{E}q}. \tag{14}$$

The extracted values are shown in figure 3, one can observe the good agreement in both cases with the theoretical value for a constant electric field.

The difference in behavior of the impact ionization models is underlined by the probability density function of the distance traveled by a carrier before it suffered its first impact ionization, represented in figure 4. The local model enables impact ionization to occur immediately after the carrier was injected, while dead-space models give results where a certain distance is needed for the carrier to gain a threshold energy, necessary to trigger an impact ionization. The hard dead-space model results in a zone where the probability of triggering an impact ionization is strictly zero, while the soft model slightly relax this condition, with a low, but positive ionization probability.

C. Device simulation

Device simulation procedure begins with the design of the desired device followed by simulation of the manufacturing process with a commercial process simulator to obtain the final doping profile of the device. The electrostatics of the device is then calculated with standard commercial TCAD simulation software [41]. This preliminary workflow results in doping and electric field profiles in two or three dimensions. Our methodology then uses these profiles whenever it is necessary for the simulation, by interpolating the doping and the electric field within the mesh of the simulation. In what follows, the frozen field approximation is used [42], which means that the electrostatics of the device is not updated at each iteration of the Monte Carlo simulation, with respect to the new charge distribution. Our simulation could be extended to a so-called self-consistent device simulation, at the cost of computing, with a solver of the Poisson equation, the charge densities and its corresponding electric field after every given time step [43]. In Algorithm 1, the overall design of our methodology is represented, non-local approximations are not considered for clarity purposes. The algorithms with hard and soft dead-space models are shown in the Algorithms section of the supplementary material file. Figure 5 represents the trajectory of a single electron injected in the volume of the SPAD, which drifts and diffuses following the electric field, and finally

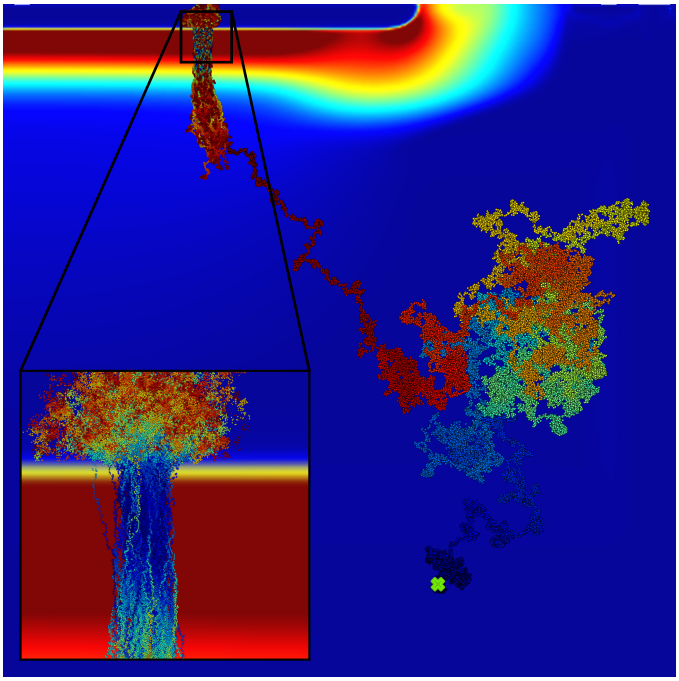

Fig. 5. Illustration of the device multi-particle simulation. The background map represents the strength of the electric field, the red region corresponds to the SPAD main junction. An electron is injected in the volume of the SPAD (yellow cross), it drifts and diffuses towards the junction and triggers several impact ionization events, leading to a self-sustained avalanche. The color of the trajectories' dot represents the intrinsic time of the particles. A zoomed view of the avalanche is shown in the inset. The third dimension was dismissed for visualization purposes.

creates an avalanche in the junction of the SPAD, visible in the inset.

Algorithm 1: ADMC algorithm

input : Starting position $\mathbf{x}_0$
input : Time step dt
input : Final time T_{max}
input : EndCondition
input : $\mathbf{v}(\mathbf{r},t)$ and $\sigma(\mathbf{r},t)$

$time \leftarrow 0$
$\mathbf{X}^0 \leftarrow \mathbf{x}_0$
$p_{ii} \leftarrow 0$
Generate $r \hookrightarrow \mathcal{U}(0,1)$

while $time \leq T_{max}$ ***and*** *no EndCondition* **do**
  for *each particle* **do**
    Generate $W^n \hookrightarrow \mathcal{N}(0,dt)$
    $\mathbf{X}^{n+1} = \mathbf{X}^n + \mathbf{v}(\mathbf{X}^n, n\cdot dt)dt + \boldsymbol{\sigma}(\mathbf{X}^n, n\cdot dt)\mathbf{W}^n$
    $p_{ii} = p_{ii} + \mathbf{v}(\mathbf{X}^n, n\cdot dt)\cdot dt \cdot \alpha_e(X^n)$
    if $p_{ii} \geq r$ **then**
      Impact Ionization Occurs : Inject a new electron and hole at position $\mathbf{X}^{n+1}$
      reset $p_{ii} = 0$
      reset $r \hookrightarrow \mathcal{U}(0,1)$
    if *particle hits an Ohmic contact* **then**
      Remove particle from the simulation

III. Results and discussion

A. Device architecture and simulation parameters

1) Device architecture: Our method was tested against characterization results from several SPAD devices. The considered SPADs are made of a main N on top of P regions that form the main junction, also hereafter called the avalanche region. A guard ring formed of lower p-type and n-type annular areas is also used to prevent lateral breakdown [44]. These SPADs are backside illuminated. The optical stack ends with an anti-reflective coating layer, the interface between the optical stack and the silicon is structured to enhance the light absorption [45]. This interface is passivated by the means of an additional P+ layer, to reduce the dark count rate due to the interfacial defects [15]. A schematic representation of a two-dimensional sectional view of the SPAD architecture is given in figure 6. The dimensions of the different regions as well as the doping level in each of them can be modulated in order to optimize the device performances.

2) Simulation parameters: The simulations were performed with a time step of 10 fs and a maximum simulation time of 5 ns. Simulations are typically made by injecting electrons in the volume of the SPAD, with starting points distributed on a grid of 50 $\times$ 50 points $\times$ 150 points, the values are taken to approximately match the SPAD's dimension ratios. On each of these points, 250 simulations are performed, leading to a total of around 10^8 simulations. During the simulation, if more than 200 particles are active (i.e. not absorbed by a contact), we consider that the simulation ends in an avalanche breakdown event. All the simulations are independent from each other, which allows us to parallelize the simulation process. One may note that, unlike in standard "free-flight"-based Monte Carlo, the time step in our method is constant and equal for every particle. This implies that all the particles are always synchronized in time, which is a strong advantage for parallelization over the particles, without performing costly synchronization. Simulations on three-dimensional SPADs, using the above set of parameters, performed with parallelization over the simulations, using 40 cores, take around 15 hours. Although comparison with standard BTE Monte Carlo highly depends on implementation, we found that our method is at least 200 times faster than typical Full-Band Monte Carlo simulation. The temperature is set to 300 K, to match the experimental conditions, for comparison purposes.

The only considered wavelength is 940 nm, which is the wavelength of the light emitted by the laser diode used in the experiments. At such large wavelengths, the absorption coefficient of silicon is very low, and the optical absorption can thus be considered uniform in the silicon volume thanks to the optical structuration that scatters the incoming light in all directions and the metal and oxide reflectors that reflect the light back into the silicon volume. For shorter wavelengths, the optical absorption might become non-uniform and one would have to couple the optical simulation with the ADMC simulation. This can easily be done by setting the number of simulations starting from a given position to a number proportional to the optical absorption at this position.

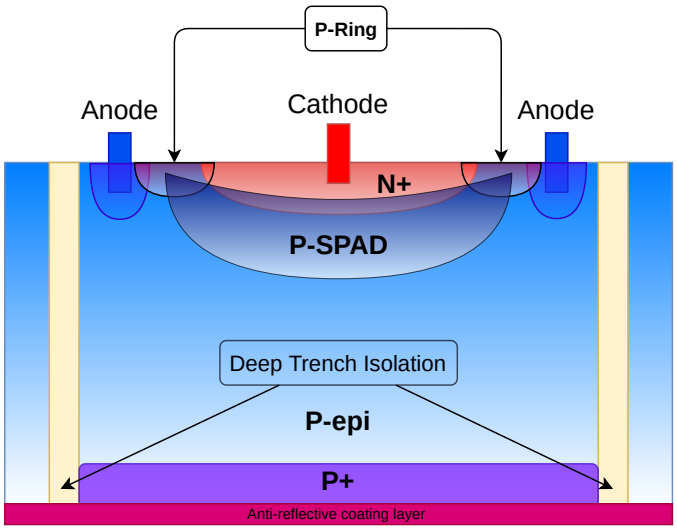

Fig. 6. Schematic sectional view of the considered SPAD architecture.

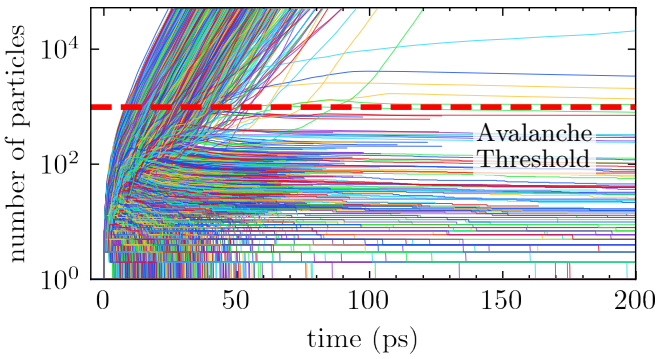

Fig. 7. Number of particles versus time for 20 000 simulations of a SPAD device. The simulations are synchronized via the time of the first impact ionization event. Particles are created through the impact ionization process and removed when collected at Ohmic contacts. The log scale allows us to appreciate the exponential growth of the number of particles when the avalanche mode is reached.

B. Avalanche breakdown probability

The detection efficiency of a SPAD is determined by, on one hand, its optical absorption (OA), and on the other hand its avalanche breakdown probability (BrP), that is, the probability that a carrier injected within the device will eventually trigger an avalanche that will be detected by the readout circuit. The probability for an incoming photon to trigger a count is called the Photon Detection Efficiency (PDE). One has also to take into account the fact that only a fraction of the sensor surface is photosensitive, this fraction is called the optical fill factor (FF). The optical absorption as much as the optical fill factor are ruled by the optical stack and the back-end circuit. Hence, only the avalanche breakdown probability can be optimized at a Silicon level.

Within our model, the breakdown probability is defined as the ratio of the number of simulations that turned into an avalanche over the total number of simulations performed. For a given simulation, if a certain threshold number of active particles is reached, then we consider that the avalanche mode is attained. This threshold number of particles can either be defined as the number for which a high enough current will be measured, and a count triggered, either an arbitrary threshold can be used, set to a number after which it is highly unlikely that the process will stop by itself. Figure 7 represents the number of active particles versus time for 20 000 simulations in a single SPAD. With this figure, it appears that if a simulation reaches 1000 active particles, it is then very unlikely that it will not lead to an avalanche (one event over 20 000).

In modern SPAD architecture the electrostatics is neither homogeneous nor isotropic [46], thus the avalanche breakdown probability has the same properties. In order to determine the local BrP, we performed several simulations with the same starting point $\mathbf{r_0}$, so we can determine the BrP of this location. By repeating this operation on many points within the device simulation domain, we create a map of avalanche breakdown probability. Such a map is shown in figure 8 (b), for each point the color represents the probability for an electron injected at this location to eventually trigger an avalanche. These kinds of maps were already produced by means of deterministic methods, such as solving the McIntyre equations over electric field lines [28]. Those methods could not take into account the statistical intrinsic behavior of the particles. With our method, a particle can, for example, overcome a potential barrier thanks to its thermal energy.

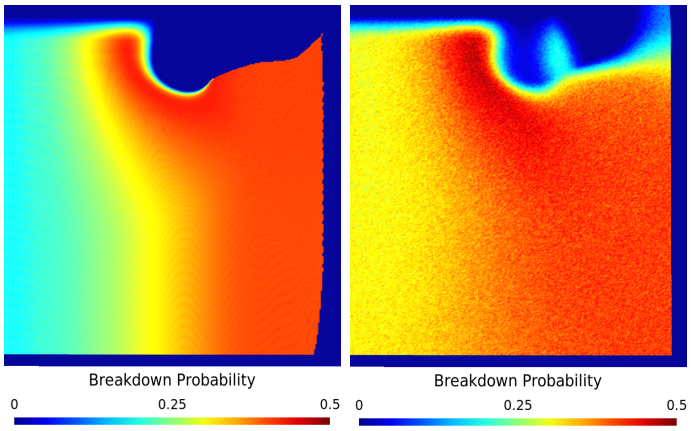

Fig. 8. Two-dimensional map of breakdown probability obtained by the method of Helleboid and al. using the McIntyre model over electric field lines [28] (left). Two-dimensional map of breakdown probability obtained from the ADMC method (right).

One can appreciate that, even though the breakdown probability maps are overall similar (see figure 8) with both methodologies, the one made with the ADMC method shows that some zones that seems to have a zero-BrP with the deterministic methods, have actually a low, but positive probability to be active.

On figure 9, many simulations from two single starting points are shown. For an electron injected near the top right anode (white star), either it enters directly in the peripheral junction, where the electric field is low, and cannot trigger any avalanche, or it diffuses downward enough to get around the ring region, to eventually be collected in the main junction

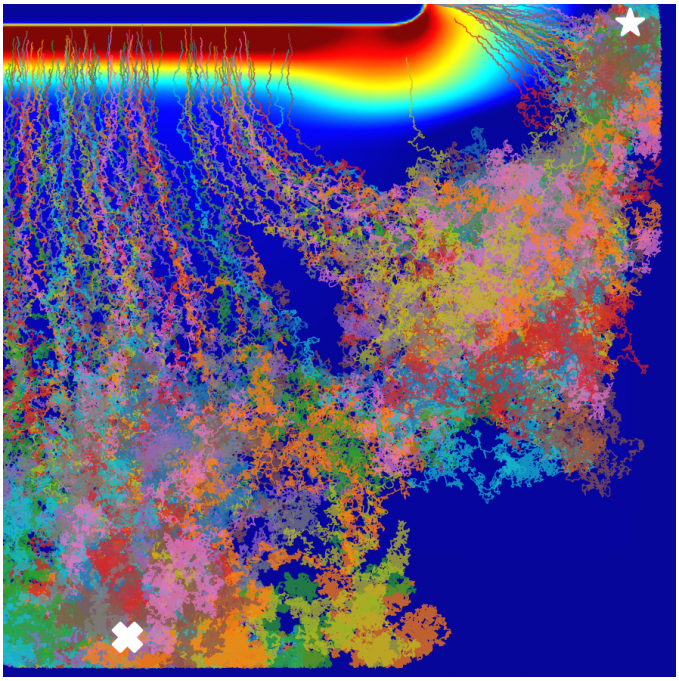

Fig. 9. Many electron trajectories before their first impact ionization. Only two starting points are considered, represented by the star and the cross. The important diffusion leads to very different paths, thus the particles enter the avalanche region at very diverse locations.

and trigger an avalanche. This second pathway cannot be appreciated with straight line or field lines method. Considering the random nature of transport is mandatory to fully capture the detailed breakdown mechanisms and understand the device operation.

Being able to accurately predict low breakdown probability areas is as well mandatory to simulate dark count rate precisely. Indeed, small areas can have a massive contribution to dark count due to localized clusters of defects with high Shockley-Read-Hall (SRH) generation rate [47]. This can be particularly the case in the near contact region (top left of figure 6) that are processed by means of heavily doped implant, that can lead to clusters of extending defects likely to enhance SHR generation rate. Dark count rate can be computed by integrating the result of the multiplication of the local breakdown probability by the local SRH generation rate [48]. Photon detection efficiency can also be computed by integrating the result of the multiplication of the local breakdown probability by the local photon absorption rate, that can be obtained from optical simulation.

C. Jitter

The main application of SPAD devices is ToF for three-dimensional imaging. In order to accurately determine the distance to an object, the time for a photon between its emission and its reception after being reflected by the object must be measured precisely. This measure is perturbed by different mechanisms: the limited time resolution of the light source, the transport time from the absorption location within the SPAD volume and the avalanche region, the time for the avalanche to grow up to measurable level of induced

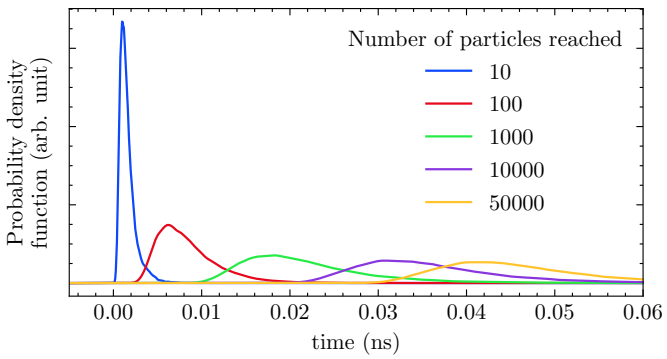

Fig. 10. Distribution of the time between the first impact ionization event and different number of particles thresholds. This figure illustrates the timing jitter due to the avalanche process itself. There is a not negligible spread of time for the avalanche to be established at measurable levels of currents.

current and finally, the finite precision of the readout circuit. The statistical distribution of the detection times for a given distance is called jitter. Among the different sources of jitter, we focus on those directly induced by the silicon device : the transport time and the avalanche build-up time.

The statistical aspect of the jitter is key, as a high discrepancy in the timing will be interpreted as a wrong distance, for example, an additional 1 ns leads to an error of 0.15 m in the distance hand.

Figure 10 illustrates the jitter due to the avalanche build-up. The distribution of the time interval between the first impact ionization and several numbers of active particles thresholds is shown. The spread of the distribution indicates that the time required to have a detectable current is a significant source of jitter. Building those statistical distributions requires an important number of results, which are hardly attainable with standard Monte Carlo device simulation. On figure 10, each curve gathers 10^6 simulations. While recent works explored analytically the statistics of the avalanche build-up under constant one-dimensional electric fields [49] [50], our method enables further investigations on three-dimensional arbitrary fields. Figure 11 exhibits some properties of the build-up jitter. The average plot highlights the exponential growth of the avalanche process, with a very stable rate after 1000 particles. The plot of the standard deviation emphasizes the fact that an important proportion of the build-up time fluctuation occurs during the beginning of the avalanche. Indeed after around 1000 particles, the standard deviation reaches a plateau. As a matter of fact, when the multiplication process is made of many particles, its timing properties are averaged, and the random history of a single particle has a lesser importance in the overall avalanche process.

In order to verify the accuracy of our model, we compared the jitter distribution obtained from our simulation with jitter characterization performed on several SPADs, manufactured within a CMOS technology process. Figure 12 gathers three comparisons between simulation and measurement of the jitter statistical distribution. We picked three diodes with significantly different electrostatics, to underline the robustness of our methodology against different architectures. Experimental and simulation parameters were kept identical for the three

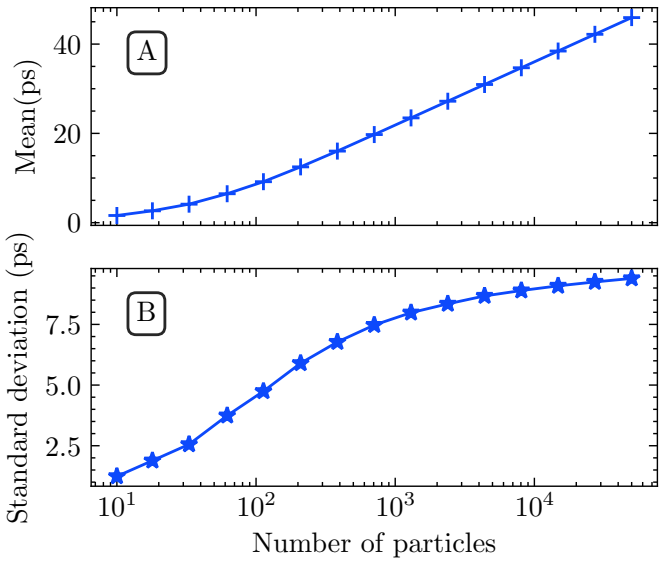

Fig. 11. Mean (A) and standard deviation (B) time to reach different numbers of particles in the avalanche process. The exponential growth of the avalanche process is clearly visible. After around 1000 particles, the standard deviation reaches a plateau, this indicates that the build-up time standard deviation is mostly determined by the timing of the first impact ionization events.

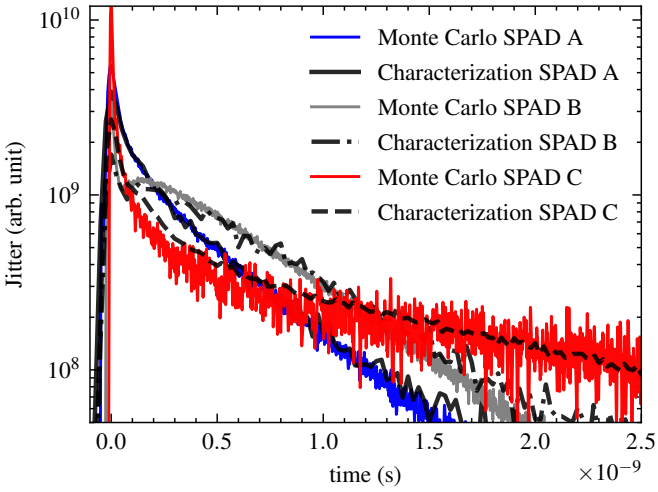

Fig. 12. Probability density function of the jitter for three different SPAD layouts. The densities extracted from simulations are compared to characterization results. The comparison shows an excellent agreement between simulation and measurement. All simulations were done in three-dimensional meshes, at the same excess bias as the characterization and each curve gathers around one million simulations, from starting points evenly distributed within the device volume.

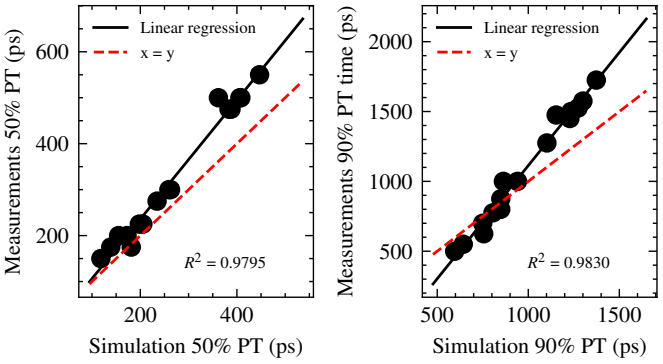

Fig. 13. Correlation plot of the time to 50% population time (PT) (left) and 90% population time (right) between simulation and characterization results. The linear regression for correlation is shown, together with the coefficient of determination which is around 98% for both correlations.

devices. All three SPADs have the same epi-layer thickness, SPAD A and B have a lighter P-SPAD implant, which helps increase the depletion region. SPAD C has a heavier P implant, leading to a narrower depletion region, which causes more diffusion of the electrons in the SPAD volume. This is reflected in the jitter distribution, as the SPAD C has a higher jitter tail than the other two SPADs.

The results show that for all devices, the simulation well reproduces the experimental results, despite very different distribution shapes. The beginning of the peak cannot be fully recovered because of the two extrinsic causes of jitter: the finite timing resolution of the laser pulse and of the readout circuit.

To assess the performance of a given device, it is often useful to reduce the full jitter distribution into a single number figure. We use a very common figure that is the time for which a given proportion of the population has a smaller jitter. For example, the 50% population time is the median of the timing jitter data.

We compared the 50% population time and the 90% population time for 16 different layouts. The correlation plots between values extracted from measurements and simulations are represented in figure 13. For both metrics, we found a very significant correlation between the simulation and the characterization. This corroborates the fact that the part of the jitter our model cannot fully assess is the same for all diodes, which is consistent with its origin (laser pulse and readout circuit), which is also the same among all silicon layouts.

IV. Conclusion

Since SPAD operation from carrier transport to multiplication process is intrinsically stochastic, an accurate modeling of such a device must include this randomness behavior. We introduced a new kind of Monte Carlo simulation that couples transport and multiplication mechanisms to achieve full multi-particle simulation. Our methodology, by its flexibility, makes it possible to carefully tune the transport as well as the impact ionization models. The simulations are utterly faster than any standard BTE Monte Carlo solver, which allows us to simulate the statistical behavior of the device with large amounts of data within reasonable computational time. The new methodology was tested against previous methods and characterization results to ensure its accuracy.

REFERENCES

[1] M. Perenzoni, N. Massari, D. Perenzoni, L. Gasparini, and D. Stoppa, "A 160 x 120 Pixel Analog-Counting Single-Photon Imager With Time-Gating and Self-Referenced Column-Parallel A/D Conversion for Fluorescence Lifetime Imaging," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 1, pp. 155–167, Jan. 2016, doi: 10.1109/JSSC.2015.2482497.

[2] D. Bronzi, F. Villa, S. Tisa, A. Tosi, F. Zappa, D. Durini, S. Weyers, and W. Brockherde, "100 000 Frames/s 64 × 32 Single-Photon Detector Array for 2-D Imaging and 3-D Ranging," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 20, no. 6, pp. 355–364, Nov. 2014, doi: 10.1109/JSTQE.2014.2341562.

[3] S. Lindner, S. Pellegrini, Y. Henrion, B. Rae, M. Wolf, and E. Charbon, "A High-PDE, Backside-Illuminated SPAD in 65/40-nm 3D IC CMOS Pixel With Cascoded Passive Quenching and Active Recharge," *IEEE Electron Device Letters*, vol. 38, no. 11, pp. 1547–1550, Nov. 2017, doi: 10.1109/LED.2017.2755989.

[4] D. Bronzi, F. Villa, S. Tisa, A. Tosi, and F. Zappa, "SPAD Figures of Merit for Photon-Counting, Photon-Timing, and Imaging Applications: A Review," *IEEE Sensors Journal*, vol. 16, no. 1, pp. 3–12, Jan. 2016, doi: 10.1109/JSEN.2015.2483565.

[5] J. D. Petticrew, S. Dimler, and J. S. Ng, "Simple Monte Carlo Simulator for modelling Linear Mode and Geiger Mode Avalanche Photodiodes in C++," *Journal of Open Research Software*, vol. 6, no. 1, May 2018, number: 1 Publisher: Ubiquity Press. doi: doi.org/10.5334/jors.212.

[6] D. Dolgos, H. Meier, A. Schenk, and B. Witzigmann, "Full-band Monte Carlo simulation of high-energy carrier transport in single photon avalanche diodes: Computation of breakdown probability, time to avalanche breakdown, and jitter," *Journal of Applied Physics*, vol. 110, pp. 084 507–084 507, Oct. 2011, doi: 10.1063/1.3652844.

[7] T. Cazimajou, M. Pala, J. Saint-Martin, R. Helleboid, J. Grebot, D. Rideau, and P. Dollfus, "Quenching Statistics of Silicon Single Photon Avalanche Diodes," *IEEE Journal of the Electron Devices Society*, vol. 9, pp. 1098–1102, 2021, doi: 10.1109/JEDS.2021.3127013.

[8] W. Oldham, R. Samuelson, and P. Antognetti, "Triggering phenomena in avalanche diodes," *IEEE Transactions on Electron Devices*, vol. 19, no. 9, pp. 1056–1060, Sep. 1972, doi: 10.1109/T-ED.1972.17544.

[9] P. Yuan, K. Anselm, C. Hu, H. Nie, C. Lenox, A. Holmes, B. Streetman, J. Campbell, and R. McIntyre, "A new look at impact ionization-Part II: Gain and noise in short avalanche photodiodes," *IEEE Transactions on Electron Devices*, vol. 46, no. 8, pp. 1632–1639, Aug. 1999, doi: 10.1109/16.777151.

[10] L. Pancheri, D. Stoppa, and G.-F. Dalla Betta, "Characterization and Modeling of Breakdown Probability in Sub-Micrometer CMOS SPADs," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 20, no. 6, pp. 328–335, Nov. 2014, doi: 10.1109/JSTQE.2014.2327791.

[11] F. Sun, Y. Xu, Z. Wu, and J. Zhang, "A Simple Analytic Modeling Method for SPAD Timing Jitter Prediction," *IEEE Journal of the Electron Devices Society*, vol. 7, pp. 261–267, 2019, doi: 10.1109/JEDS.2019.2895151.

[12] S. Gao, D. Issartel, R. Orobtchouk, F. Mandorlo, D. Golanski, A. Cathelin, and F. Calmon, "3D Electro-optical Simulations for Improving the Photon Detection Probability of SPAD Implemented in FD-SOI CMOS Technology," in *2021 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Sep. 2021, pp. 301–304, iSSN: 1946-1577.

[13] F. Rosset, A. Pilotto, L. Selmi, M. Antonelli, F. Arfelli, G. Biasiol, G. Cautero, D. De Angelis, F. Driussi, R. Menk, C. Nichetti, T. Steinhartova, and P. Palestri, "A model for the jitter of avalanche photodiodes with separate absorption and multiplication regions," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 977, p. 164346, Oct. 2020, doi: 10.1016/j.nima.2020.164346.

[14] A. Pilotto, P. Palestri, L. Selmi, M. Antonelli, F. Arfelli, G. Biasiol, G. Cautero, F. Driussi, R. H. Menk, C. Nichetti, and T. Steinhartova, "An improved random path length algorithm for p-i-n and staircase avalanche photodiodes," in *2018 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2018, pp. 26–30.

[15] E. Van Sieleghem, A. Süss, P. Boulenc, J. Lee, G. Karve, K. De Munck, C. Cavaco, and C. Van Hoof, "A Near-Infrared Enhanced Silicon Single-Photon Avalanche Diode With a Spherically Uniform Electric Field Peak," *IEEE Electron Device Letters*, vol. 42, no. 6, pp. 879–882, Jun. 2021, doi: 10.1109/LED.2021.3070691.

[16] C. Jacoboni, "Theory of Electron Transport in Semiconductors," in *Theory of Electron Transport in Semiconductors*, pp. 168–176.

[17] P. A. M. , *The Stationary Semiconductor Device Equations*, 1996. [Online]. Available: https://link.springer.com/book/10.1007/978-3-7091-3678-2

[18] G. A. Pavliotis, *Stochastic Processes and Applications : Diffusion Processes, the Fokker-Planck and Langevin Equations*, 2014th ed. New York: Springer-Verlag New York Inc., Nov. 2014.

[19] R. G. Gallager, *Stochastic Processes: Theory for Applications*, 1st ed. Cambridge University Press, 2013.

[20] C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials," *Reviews of Modern Physics*, vol. 55, no. 3, pp. 645–705, Jul. 1983, publisher: American Physical Society. doi: 10.1103/RevModPhys.55.645.

[21] N. Arora, J. Hauser, and D. Roulston, "Electron and hole mobilities in silicon as a function of concentration and temperature," *IEEE Transactions on Electron Devices*, vol. 29, no. 2, pp. 292–295, Feb. 1982, doi: 10.1109/T-ED.1982.20698.

[22] C. Canali, G. Majni, R. Minder, and G. Ottaviani, "Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature," *IEEE Transactions on Electron Devices*, vol. 22, no. 11, pp. 1045–1047, Nov. 1975, doi: 10.1109/T-ED.1975.18267.

[23] P. S. Cheung and C. J. Hearn, "The diffusion of electrons in semiconductors in high electric fields," *Journal of Physics C: Solid State Physics*, vol. 5, no. 13, pp. 1563–1572, Jul. 1972, doi: 10.1088/0022-3719/5/13/006.

[24] V. K. Arora, "Drift diffusion and Einstein relation for electrons in silicon subjected to a high electric field," *Applied Physics Letters*, vol. 80, no. 20, pp. 3763–3765, May 2002, doi: 10.1063/1.1480119.

[25] D. S. Ong, K. F. Li, G. J. Rees, J. P. R. David, and P. N. Robson, "A simple model to determine multiplication and noise in avalanche photodiodes," *Journal of Applied Physics*, vol. 83, no. 6, pp. 3426–3428, Mar. 1998, publisher: American Institute of Physics. [Online]. Available: https://aip.scitation.org/doi/10.1063/1.367111 doi: 10.1063/1.367111.

[26] S. M. Sze and K. K. Ng, "Imapct_ionization_physics of Semiconductor Devices," in *Physics of Semiconductor Devices*, 3rd ed. Hoboken, N.J: Wiley–Blackwell, Nov. 2006, pp. 45–47.

[27] R. McIntyre, "A new look at impact ionization-Part I: A theory of gain, noise, breakdown probability, and frequency response," *IEEE Transactions on Electron Devices*, vol. 46, no. 8, pp. 1623–1631, Aug. 1999, doi: 10.1109/16.777150.

[28] R. Helleboid, D. Rideau, J. Grebot, I. Nicholson, N. Moussy, O. Saxod, M. Basset, A. Zimmer, B. Mamdy, D. Golanski, M. Agnew, S. Pellegrini, M. Sicre, C. Buj, G. Marchand, J. Saint-Martin, M. Pala, and P. Dollfus, "Comprehensive modeling and characterization of photon detection efficiency and jitter tail in advanced SPAD devices," *IEEE Journal of the Electron Devices Society*, pp. 1–1, 2022, doi: 10.1109/JEDS.2022.3168365.

[29] R. Van Overstraeten and H. De Man, "Measurement of the ionization rates in diffused silicon p-n junctions," *Solid-State Electronics*, vol. 13, no. 5, pp. 583–608, May 1970, doi: 10.1016/0038-1101(70)90139-5.

[30] Y. Okuto and C. R. Crowell, "Ionization coefficients in semiconductors: A nonlocalized property," *Physical Review B*, vol. 10, no. 10, pp. 4284–4296, Nov. 1974, publisher: American Physical Society. doi: 10.1103/PhysRevB.10.4284.

[31] ——, "Threshold energy effect on avalanche breakdown voltage in semiconductor junctions," *Solid-State Electronics*, vol. 18, no. 2, pp. 161–168, Feb. 1975, doi: 10.1016/0038-1101(75)90099-4.

[32] M. Hayat, B. Saleh, and M. Teich, "Effect of dead space on gain and noise of double-carrier-multiplication avalanche photodiodes," *IEEE Transactions on Electron Devices*, vol. 39, no. 3, pp. 546–552, Mar. 1992, doi: 10.1109/16.123476.

[33] J. R. Hauser, "Threshold Energy for Avalanche Multiplication in Semiconductors," *Journal of Applied Physics*, vol. 37, no. 2, pp. 507–509, Feb. 1966, publisher: American Institute of Physics. doi: 10.1063/1.1708204.

[34] C. L. Anderson and C. R. Crowell, "Threshold Energies for Electron-Hole Pair Production by Impact Ionization in Semiconductors," *Physical Review B*, vol. 5, no. 6, pp. 2267–2272, Mar. 1972, publisher: American Physical Society. doi: 10.1103/PhysRevB.5.2267.

[35] A. Spinelli, A. Pacelli, and A. L. Lacaita, "Dead space approximation for impact ionization in silicon," *Applied Physics Letters*, vol. 69, no. 24, pp. 3707–3709, Dec. 1996, publisher: American Institute of Physics. doi: 10.1063/1.117196.

[36] C. H. Tan, J. C. Clark, J. P. R. David, G. J. Rees, S. A. Plimmer, R. C. Tozer, D. C. Herbert, D. J. Robbins, W. Y. Leong, and J. Newey,

"Avalanche noise measurement in thin Si p+-i-n+ diodes," *Applied Physics Letters*, vol. 76, no. 26, pp. 3926–3928, Jun. 2000, publisher: American Institute of Physics. doi: 10.1063/1.126823.

[37] N. Sano, T. Aoki, and A. Yoshii, "Soft and hard ionization thresholds in Si and GaAs," *Applied Physics Letters*, vol. 55, no. 14, pp. 1418–1420, Oct. 1989, publisher: American Institute of Physics. doi: 10.1063/1.101612.

[38] D. Dolgos, H. Meier, A. Schenk, and B. Witzigmann, "Full-band Monte Carlo simulation of single photon avalanche diodes," in *2013 IEEE Photonics Conference*, Sep. 2013, pp. 360–361, iSSN: 1092-8081.

[39] G. Rees and J. David, "Nonlocal impact ionization and avalanche multiplication," *Journal of Physics D: Applied Physics*, vol. 43, p. 243001, Jun. 2010, doi: 10.1088/0022-3727/43/24/243001.

[40] A. Spinelli and A. Lacaita, "Mean gain of avalanche photodiodes in a dead space model," *IEEE Transactions on Electron Devices*, vol. 43, no. 1, pp. 23–30, Jan. 1996, doi: 10.1109/16.477589.

[41] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. Vienna: Springer, 1984.

[42] J. Higman, K. Hess, C. Hwang, and R. Dutton, "Coupled Monte Carlo-drift diffusion analysis of hot-electron effects in MOSFETs," *IEEE Transactions on Electron Devices*, vol. 36, no. 5, pp. 930–937, May 1989, doi: 10.1109/16.299675.

[43] K. Hess, Ed., *Monte Carlo Device Simulation*. Boston, MA: Springer US, 1991.

[44] W. Wang, Y. Zhang, and Z. Wei, "High-Performance Structure of Guard Ring in Avalanche Diode for Single Photon Detection," *International Journal of Communications, Network and System Sciences*, vol. 10, no. 8, pp. 1–6, Aug. 2017, number: 8 Publisher: Scientific Research Publishing. doi: 10.4236/ijcns.2017.108B001.

[45] C. Liu, H.-F. Ye, and Y.-L. Shi, "Advances in near-infrared avalanche diode single-photon detectors," *Chip*, vol. 1, no. 1, p. 100005, Mar. 2022, doi: 10.1016/j.chip.2022.100005.

[46] K. Morimoto, J. Iwata, M. Shinohara, H. Sekine, A. Abdelghafar, H. Tsuchiya, Y. Kuroda, K. Tojima, W. Endo, Y. Maehashi, Y. Ota, T. Sasago, S. Maekawa, S. Hikosaka, T. Kanou, A. Kato, T. Tezuka, S. Yoshizaki, T. Ogawa, K. Uehira, A. Ehara, F. Inui, Y. Matsuno, K. Sakurai, and T. Ichikawa, "3.2 Megapixel 3D-Stacked Charge Focusing SPAD for Low-Light Imaging and Depth Sensing," in *2021 IEEE International Electron Devices Meeting (IEDM)*, Dec. 2021, pp. 20.2.1–20.2.4, iSSN: 2156-017X.

[47] A. Jay, A. Hémeryck, F. Cristiano, D. Rideau, P. Julliard, V. Goiffon, A. LeRoch, N. Richard, L. M. Samos, and S. D. Gironcoli, "Clusters of Defects as a Possible Origin of Random Telegraph Signal in Imager Devices: a DFT based Study," in *2021 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Sep. 2021, pp. 128–132, iSSN: 1946-1577.

[48] M. Sicre, M. Agnew, C. Buj, J. Coignus, D. Golanski, R. Helleboid, B. Mamdy, I. Nicholson, S. Pellegrini, D. Rideau, D. Roy, and F. Calmon, "Dark Count Rate in Single-Photon Avalanche Diodes: Characterization and Modeling study," in *ESSCIRC 2021 - IEEE 47th European Solid State Circuits Conference (ESSCIRC)*, Sep. 2021, pp. 143–146.

[49] P. Windischhofer and W. Riegler, "The statistics of electron–hole avalanches," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1003, p. 165327, Jul. 2021, doi: 10.1016/j.nima.2021.165327.

[50] W. Riegler and P. Windischhofer, "Time resolution and efficiency of SPADs and SiPMs for photons and charged particles," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1003, p. 165265, Jul. 2021, doi: 10.1016/j.nima.2021.165265.