Direct Time-of-Flight TCSPC Analytical Modeling Including Dead-Time Effects

Foad Arvani, Tony Chan Carusone, The Edward S. Rogers Sr. Department of Electrical & Computer Engineering (ECE), University of Toronto, Toronto, Canada, Email: foad.arvani@isl.utoronto.ca

Abstract—**The optimization of a TCSPC system requires modeling which considers the design specifications and parameters of the target application under different operating scenarios. Since single-photon detection is fundamentally a stochastic process, extensive behavioral Monte Carlo simulations are normally used. Their accuracy depends upon computation time. However, the trend towards larger SPAD arrays and emerging complex TDC sharing architectures requires much faster simulation methods. In this paper, a simple, fast and accurate analytical model is presented to address this need. It accounts for dead time effects which result in missed photon counts through the analysis of inhomogeneous continuous time Markov chain. The effective received power and photon detection rate are determined and the corresponding analytical histogram is created. This histogram is the basis for calculating time of flight and can be used to explore architectural alternatives and accelerate design verification. Outputs of the presented analytical model match those of Monte Carlo simulations, and are produced considerably faster. The computation time improvement grows with array sizes and this enables parametric analysis of TCSPC system.**

Keywords—time of flight (ToF); time correlated single photon counting (TCSPC); SPAD; dead time; inhomogeneous continuous time Markov chain (ICTMC); Monte Carlo

I. Introduction

The interest in high performance three-dimensional (3D) imaging has grown in recent years due to immense demand in engineering, science and entertainment [1]. Much work has been carried out in the field of solid-state time-of-flight (ToF) 3D imagers. Evaluating the performance of state-of-the-art 3D imagers requires comprehensive simulations for the target application under various operating scenarios. System level simulations for single photon avalanche diode (SPAD) based ToF designs can be carried out by analytical modeling [2] or numerical Monte Carlo simulations. While Monte Carlo simulations generate realistic samples of the system output, its accuracy depends on the computation time which increases with the sensor array size. On the other hand, even though analytical physical models provide almost instantaneous results and parametrization, their accuracy has so far been limited. Furthermore, analytical modeling helps to develop design insight and facilitates parametric sweeps of design variables. Examples of design parameters are illumination power, illumination pulse duration, and time to digital converter (TDC) resolution and precision to name a few.

The focus of this work is to develop an analytical model of the time correlated single photon counting (TCSPC) process considering the dead time effect and the resulting missed counts.

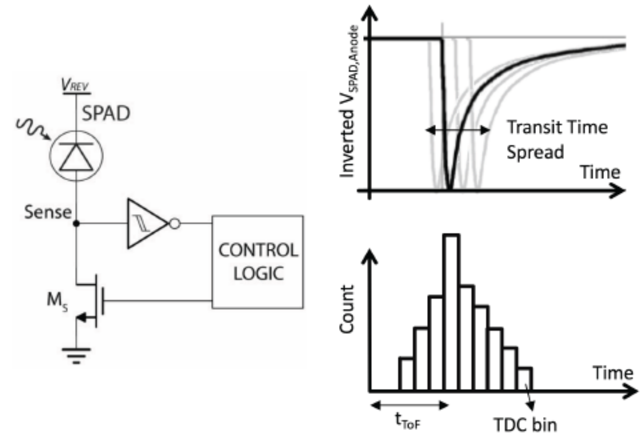

Fig. 1. Left: SPAD's variable-load quenching circuit (VLQC) [3]. The quenching element is a transistor with variable channel resistance. Right: TCSPC operation, Top: Single-photon detected pulses of SPAD, Bottom: Reconstructed signal histogram after the TDC

It facilitates system level simulation of SPAD-based direct ToF designs. The proposed model takes into account the time inhomogeneous nature of the received pulse along with the SPAD quenching and TDC dead time and the resulting instantaneous output matches that of Monte Carlo simulations. In this paper, initially, SPAD behavior and the time inhomogeneous nature of photon detection in TCSPC systems are briefly reviewed in part II. Later in part III, analytical modeling including the dead time effect is presented followed by Monte Carlo verification in part IV.

II. Background

A. SPADs in TCSPC and the Dead Time of the System

SPADs are semiconductor devices based on a p-n junction reverse-biased at a voltage that exceeds the breakdown voltage of the junction. At this bias point, the electric field is so high that a single charge carrier injected into the depletion layer can trigger a self-sustaining avalanche. The current rises swiftly with a sub-nanosecond rise-time to a macroscopic steady-state level in the mA range [3]. If the primary carrier is photo-generated, the leading edge of the avalanche pulse, marks the arrival time of the detected photon with picosecond time jitter [3]. The current continues to flow until the avalanche is quenched by lowering the bias voltage below the breakdown voltage, at which time the current ceases (see Fig. 1). To detect another photon, the bias voltage must be raised again above breakdown. The voltage will then be recharged by the flow of electric current into the parasitic capacitances, with the diode being ready to detect another carrier after this SPAD dead time. In TCSPC measurement, this process is repeated several times and the resulting time of flights which also include false triggers because of background illumination or dark counts are recorded by TDCs and a histogram is created [1] as shown is Fig. 1.

The dead time of the total system is calculated by adding up the dead time of TDCs. There have been some efforts in modeling of the SPAD based ToF systems including the dead time effects [2]. As they are investigating indirect time of flight measurement, they don't address TCSPC system behavior, they lack considering time inhomogeneity of detection, and they are limited to the special case of one TDC per pixel architecture. In this paper, we are interested in investigating the dead time effects of TDCs which are dominant in systems where TDCs are shared per sub-arrays especially in complex TDC sharing methods. Dead time is itself likely a random variable. Fortunately, the blocking probability resulting from the dead time effects depends only upon the mean dead time, μ and therefore does not require detailed treatment of the statistics of the dead time [4]. So, during periods when the system is busy, pulses leave the system according to a Poisson process of rate μ.

B. *Time-Inhomogeneous Photon Detection*

In ToF distance measurements, the received pulse incident on the SPAD is a time dependent reflected pulse (generally Gaussian) superimposed on background illumination resulting in a time varying signal. This inhomogeneous Poisson process does not have the stationary increment property, which complicates the analysis. However, since it is a shrinking Bernoulli process, it can be viewed as a homogeneous Poisson process over a non-linear time scale [4].

III. Analytical Modeling and Histogram

A. *Inhomogeneous Continuous Time Markov Chain (ICTMC) and Blocking Probability*

To consider the dead time in TCSPC systems, we are interested in finding the probability with which an arriving pulse is missed. This is important as missed counts degrade the effective signal counts and hence increase the variance of the measurement, reducing accuracy. We consider the general case where an array of SPADs are connected to in-column TDCs so that the arrival time of a pulse generated by any SPAD is measured by any available TDC in that column. Once measuring a pulse, the TDC is unavailable for a mean dead time. When all in-column TDCs are unavailable, the system is said to be in a blocking state. Time blocking refers to the proportion of time the system spends in the blocking state where all N TDCs are busy processing the photon detection times; and arrival blocking is the proportion of the arriving pulses which are missed. These two quantities are equal if the arrival process is Poisson and the dead times of the arrivals are independent and identically distributed [4].

The Poisson process that we have seen earlier is a continuous time Markov chain. However, equipped with the small-slot discrete-time approximation technique, we may analyze the Poisson process as a discrete-time Markov chain as well. The TCSPC system with dead time can be considered a loss system with no waiting time and can be modeled by an immigration-

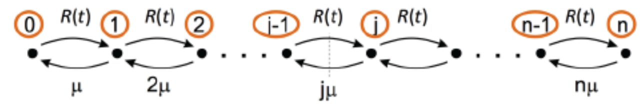

Fig. 2. The Markov chain transition diagram of a TCSPC system with dead time can be considered as a time inhomogeneous immigration-death process ($R_i(t) = R(t)$ and $\mu_i = i\mu$).

death process [4] which means that the new arrivals enter the system at random at a rate $R_i(t)$ which is independent of i, meaning that it is not dependent on the state of the system at time t, and the dead time is $\mu_i = i\mu$ which means the busy TDCs in the system become available again at random, at a simple linear rate μ per individual. These assumptions are valid when there is no appreciable crowding [4] as in the case of TCSPC systems [1]. The transition diagram for such a process is shown in Fig. 2. The labels on the arrows indicate transition probabilities which are conditional probabilities of going from some value of K to another value. Consider a random variable $N(t)$ which represents the number of occupied TDCs at time t. The Markov process with discrete states in continuous time is in principle defined when we have a set of functions [4],

$$p_{ij}(t,u) \quad prob\{N(u) \quad j \mid N(t) \quad i\}$$

The infinitesimal generator matrix $\boldsymbol{Q}$ is an array of numbers describing the rate a continuous time Markov chain moves between states. The rows of $\boldsymbol{Q}$ form a basis for a linear code with $q_{ii} \quad \lim_{h\to0} \frac{1 \;\, p_{ii}}{h}, q_{ij} \quad \lim_{h\to0} \frac{p_{ij}}{h}$ where h denotes the time increment [4].

The Markov Chain in Fig. 2 can be described with Kolmogorov forward differential equations. The matrix form of the Kolmogorov forward differential equations for the time-inhomogeneous Markov jump process is [4]

$$\frac{\partial \boldsymbol{P}(t,u)}{\partial u} \quad \boldsymbol{P}(t,u)\boldsymbol{Q}(u)$$

where for the case of Fig. 2 we have

$$\boldsymbol{Q}(u) \quad \begin{bmatrix} R(u) & R(u) & 0 & 0 \\ \mu & R(u) \quad \mu & R(u) & 0 \\ 0 & 2\mu & . & \vdots \\ 0 & \vdots & \vdots & \ddots \end{bmatrix}$$

Thus, the probability of each state is

$$\frac{\partial p_{ij}(t,u)}{\partial u} \quad R(u)\cdot p_{ij \;\, 1}(t,u) \quad (R(u)+j\mu)\cdot p_{ij}(t,u)$$
$$+(j+1)\cdot \mu p_{ij+1}(t,u)$$

The row vector i denotes the probabilities at time u corresponding to the initial state i and the initial condition is $\boldsymbol{P}(0,0) = \boldsymbol{I}$. Taking the nonzero terms, alternatively, we could think about of this equation as follows:

- Take the probability that the process goes from state i at time t to state j-1 at time u, and multiply this by the force of transition from state j-1 to state j at time u,
- Then add the probability that the process goes from state i at time t to state j at time u, multiplied by the force of transition that keeps the process in state j at time u,

- Then add the probability that the process goes from state i at time t to state j+1 at time u, and multiply this by the force of transition from state j+1 to state j at time u.

Practically, for TCSPC processes, there is far less than one detection per pulse on average. We therefore assume there is no appreciable crowding and that the histogram resolution (TDC LSB width) is much smaller than σ of the reflected pulse where $\sigma \quad \sqrt{\sigma_{Laser}^2 + \sigma_{SPAD}^2 + \sigma_{TDC}^2}$.

This results in the first photon detection rate which is a weakly time varying process in time slots whose width is the TDC resolution. Let $N(t)$ denote the number of TDCs in service and $p_j(t) \quad prob\{N(t) \quad j\}$ the probability of being in state j at time t. Then the state variable $N(t)$ constitutes a Markov process and we have

$$\overline{R(t)} \cdot p_{j \quad 1}(t) \quad j\mu \cdot p_j(t) \to p_j(t) \quad p_0(t) \frac{(\overline{R(t)} \cdot \tau_{dead})^j}{j!}$$

where $\overline{R(t)}$ denotes the average arrival rate in $[t \; t + \Delta]$. Given that $p_0 + p_1 + \cdots + p_n \quad 1$, the probability p_n that a new arrival is rejected because all TDCs are busy can be calculated as

$$p_n(t) \quad \frac{\frac{(\overline{R(t)} \cdot \tau_{dead})^n}{n!}}{\sum_{j=0}^{n} \frac{(\overline{R(t)} \cdot \tau_{dead})^j}{j!}}$$

This can be considered as a generalization of Erlang's formula [4] to a time inhomogeneous scenario. This equation shows a non-linear behavior with respect to the incoming arrival rate so superposition is not valid and $R(t)$ should include both background illumination and the reflected light pulse. The blocking probability as a function of arrival rate $R(t)$ and dead time τ_{dead} is shown in Fig. 3 for different TDC counts in the system.

B. The First Detection and Inter-Arrival-Time (IAT) for Time Inhomogeneous case

In TCSPC systems only the first photon that triggers the SPAD is detected [1]. In this section, the probability distribution function of the first arrival is determined. Let the last arrival occurred at time 0. The probability of an arrival occurrence in [t, t+Δ) is then given as

$$P(t \leq X < t + \Delta t) \quad \prod_{1}^{N \quad 1} [1 \quad \lambda(k\Delta)\Delta] \cdot [\lambda(N\Delta)\Delta] + o(\Delta)$$

Using the following approximations, $\prod_1^N [1 \quad \lambda(k\Delta)\Delta] \; exp[\sum_1^N \{ \quad \lambda(k\Delta)\Delta + o(\Delta)\}]$ and $\frac{\lambda(N\Delta)\Delta}{1 \quad \lambda(N\Delta)\Delta} \quad \lambda(N\Delta)\Delta + o(\Delta)$, the *IAT* density is given as

$$f(t) \quad \lim_{\Delta \to 0} \frac{P(t \leq X < t + \Delta t)}{\Delta} \quad \lambda(t) \exp\left[\quad \int_0^t \lambda(u)du \right]$$

The *IAT* density $f(t)$ represents the probability distribution function of the first arrival and it can be used to model $R(t)$ to create the detected arrival pulse sequence.

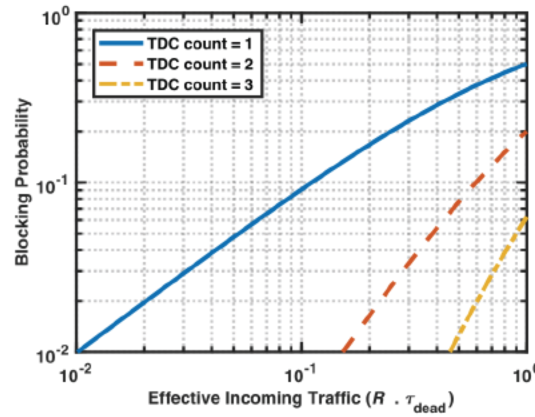

Fig. 3. The blocking probability versus photon detection rate for different TDC counts assigned to process the incoming traffic

C. Effective Received Pulse and Analytical Histogram

The power distribution of the reflected pulse at the detector can be described in time as

$$P \quad P_{Sig}(t) + P_{BG}$$

Then the effective incoming photon arrival rate as a consequence of reflected pulse and background illumination is as follows

$$R(t) \quad (\lambda_{BG} + \lambda_{Sig}(t)) \cdot exp\left\{ \quad \int_0^t (\lambda_{BG} + \lambda_{Sig}(u))du \right\}$$

where $\lambda_{BG} \quad \frac{P_{BG}}{E_{Photon}} \cdot \eta_{PDE}$, $\lambda_{Sig}(t) \quad \frac{P_{Sig}(t)}{E_{Photon}} \cdot \eta_{PDE}$ and η_{PDE} denotes the photon detection efficiency.

The effective rate after considering the dead time effect is

$$R_{Eff}(t) \quad R(t) \cdot [1 \quad p_n(t)]$$

Now the system can be treated as a normal light detector with zero dead time and sensitive to all incoming photons not just the first arrival. The reconstructed histogram after the TDC, can be calculated as follows

$$h(\tau) \quad \int_{\tau \quad \frac{\Delta t}{2}}^{\tau + \frac{\Delta t}{2}} R_{Eff}(t)dt$$

where $\tau \in \left[0, \Delta t, 2\Delta t, \ldots, (\frac{T_{Pulse}}{\Delta t} \quad 1)\Delta t\right]$, T_{Pulse} is the pulse repetition period, and Δt is equal to TDC's LSB width.

D. Position of the Peak; A Shift Towards Lower Values

Assume, for simplicity, the case in which λ_{BG} and τ_{dead} are zero. The typical diffusion tail present in SPAD timing responses which represents the after-pulsing is very often submerged under the background level of the TCSPC histogram [5], so for Gaussian illumination, the power distribution of the reflected pulse at the detector can be described in time as

$$P \quad \frac{E_{Rx}}{\sqrt{2\pi}\sigma} exp\left(\quad \frac{(t \quad t_{ToF})^2}{2\sigma^2} \right)$$

where E_{Rx} is the total energy of one pulse. Also assume that $\Delta t \ll \sigma$, then we have weakly time varying case and the corresponding analytical histogram can be calculated as

$$h(\tau) \approx R_{Eff}(\tau) \cdot \Delta t \cdot N$$

where N denotes the number of repetition which is equal to the number of Monte Carlo simulations. The position of the histogram peak, t_{peak}, is an estimation of t_{ToF} and it may be determined by setting $\partial h(\tau)/\partial \tau \quad 0$ as

$$\frac{t_{peak} \quad t_{ToF}}{\sigma^2} + \frac{E_{Rx} \cdot \eta_{PDE}}{E_{Ph} \cdot \sqrt{2\pi}\sigma} \exp\left[\quad \left(\frac{t_{peak} \quad t_{ToF}}{2\sigma^2}\right)^2 \right] \quad 0$$

Then $t_{peak} \quad j\Delta t \leq t_{ToF}$ and equity holds for the case of $E_{Rx} \cdot \eta_{PDE} \approx 0$. As nonzero λ_{BG} means higher detection rate $R(t)$ at lower values of time, it shifts t_{peak} further towards lower values so the statement holds true for real life scenarios as well.

IV. Verification with Monte Carlo Simulation

To validate the model, a series of the SPAD's first arrival detected pulses with rate $R(t)$ are statistically generated for an array of 3 by 4 through Monte Carlo simulation. These pulses are then processed for the most common TDC sharing architecture in 3D image sensors with SPADs wherein one TDC is shared within each column of the SPAD sensors. The pulse modulation frequency is set to f_{mod} = 100MHz, σ_{Laser} = 150ps, σ_{SPAD} = 30ps, and σ_{TDC} = 10ps. Under normal indoor operation conditions, λ_{BG} 33.4MPhoton/sec and an average transmitted pulse power of 10mW results in a received power level of approximately 1pW at a target distance of 30cm [6]. The histogram resolution is set to TDC LSB width of 50psec and simulations are run 500000 times for one measurement at 200 frames per second.

Eventually, the Monte Carlo simulation is run and dead time processing is performed on each pixel in the array. The pixel with maximum missed counts is selected to create the histogram. Fig. 4 shows a good match between the analytical histogram and the Monte Carlo result for two cases with different dead times of 5nsec and 10nsec. It's worth noting that the simulation time of Monte Carlo analysis strongly depends on the system simulation parameters such as incoming photon rate and dead time. Simulations are run with MacBook Pro with 2.7 GHz Intel Core i5 and 8GB of memory. For this case, the calculation time is approximately 5000 times faster for the analytical simulation and this improvement grows with the increase in SPAD array size.

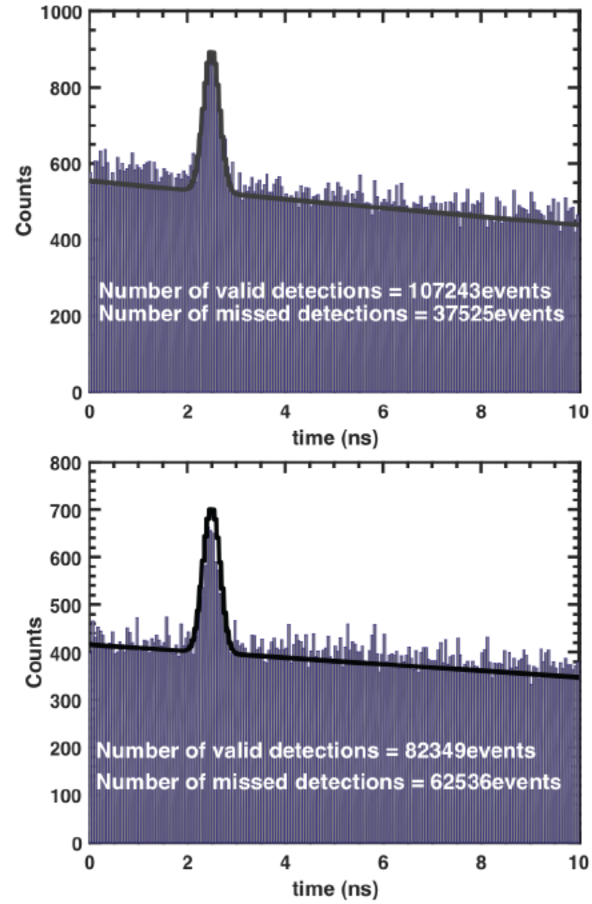

Fig. 4. Histogram of the 1st arrival with histogram resolution of 50psec resulting from Monte Carlo simulation along with the presented analytical modeling (solid line). The frame rate is 200 and repetition frequency is f_{mod} = 100MHz. Top: τ_{dead} = 5nsec, Bottom: τ_{dead} = 10nsec

Conclusions

In this paper, the need for faster simulation method for TCSPC systems is highlighted to support the emergent trend towards SPAD arrays with more elements and more complex detection architectures. An analytical model has been presented for the most common architecture in 3D image sensors with SPADs in which TDCs are shared per columns of SPAD sensors. The validity of this model has been verified through comparison with Monte Carlo simulations. It is therefore possible to use this model to accurately and quickly relate the detected ToF histogram for TCSPC 3D imagers to design parameters. The presented analytical model facilitates system level simulation of TCSPC systems, TDC architectural and design parameters sweep such as the number of TDCs, their resolution, dead time and precision, and it allows investigating more complex TDC sharing architectures.

Acknowledgment

Authors would like to thank Kapik Integration Inc. for their support.

References

[1] F. Remondino and D. Stoppa, *TOF range-imaging cameras* (TOF Range-Imaging Cameras). Springer, 2013, pp. 1-240.

[2] M. Beer, O. Schrey, B. J. Hosticka, and R. Kokozonski, "Dead Time Effects in the Indirect Time of Flight Measurement with SPAD," in *IEEE International Symposium on Circuits & Systems*, 2017.

[3] A. Gallivanoni, I. Rech, and M. Ghioni, "Progress in Quenching Circuits for Single Photon Avalanche Diodes," *IEEE Transactions on Nuclear Science*, vol. 57, no. 6, pp. 3815-3826, 2010.

[4] D. R. Cox and H. D. Miller, *The Theory of Stochastic Processes*. Taylor & Francis, 1977.

[5] M. J. D. Darek P. Palubiak, "CMOS SPADs: Design Issues and Research Challenges for Detectors, Circuits, and Arrays," *IEEE Journal of Solid-State Circuits*, vol. 20, no. 6, 2014.

[6] C. Niclass, "Single-Photon Image Sensors in CMOS: Picosecond Resolution for Three-Dimensional Imaging," PhD, EPFL, 2008.