The TT-PET project: a thin TOF-PET scanner based on fast novel silicon pixel detectors

Y. Bandi,* 1 M. Benoit, b F. R. Cadoux, b D.C. Forshaw, a R. Hänni, a D. Hayakawa, b G. Iacobucci, b S. Michal, b A. Miucci, a L. Paolozzi, b O. Ratib, b E. Ripiccini, b C. Tognina, a P. Valerio b and M. Weber a

a University of Bern,
Sidlerstrasse 5, Bern, Switzerland
b University of Geneva,
Quai Ernest-Ansermet 24, 1211 Geneve, Switzerland

E-mail: ybandi@lhep.unibe.ch

ABSTRACT: The TT-PET project aims at developing a compact Time-of-flight PET scanner with 30ps time resolution, capable of withstanding high magnetic fields and allowing for integration in a traditional MRI scanner, providing complimentary real-time PET images. The very high timing resolution of the TT-PET scanner is achieved thanks to a new generation of Silicon-Germanium (Si-Ge) amplifiers, which are embedded in monolithic pixel sensors. The scanner is composed of 16 detection towers as well as cooling blocks, arranged in a ring structure. The towers are composed of multiple ultra-thin pixel modules stacked on top of each other. Making it possible to perform depth of interaction measurements and maximize the spatial resolution along the line of flight of the two photons emitted within a patient. This will result in improved image quality, contrast, and uniformity while drastically reducing backgrounds within the scanner. Allowing for a reduction in the amount of radioactivity delivered to the patient. Due to an expected data rate of about 250 MB/s a custom readout system for high data throughput has been developed, which includes noise filtering and reduced data pressure. The realisation of a first scanner prototype for small animals is foreseen by 2019. A general overview of the scanner will be given including, technical details concerning the detection elements, mechanics, DAQ readout, simulation and results.

KEYWORDS: Gamma camera, SPECT, PET PET/CT, coronary CT angiography (CTA); Particle tracking detectors (Solid-state detectors); Instrumentation and methods for time-of-flight (TOF) spectroscopy; Timing detectors

*Corresponding author.
1 Introduction

The TT-PET project [1] aims to develop a Thin Time-of-flight PET scanner with 30 ps time resolution, capable of withstanding high magnetic fields, which will allow for integration in a traditional MRI scanner to provide complimentary real-time PET images. The very high timing resolution of the TT-PET scanner is achieved thanks to a new generation of Silicon-Germanium (Si-Ge) amplifiers, embedded in monolithic pixel sensors, which are being produced at IHP, using SG13S Si-Ge Bi-CMOS technology [2]. The scanner is composed of 16 detection towers and 16 cooling blocks arranged in a ring structure (figure 1(a)). The towers are composed of multiple ultra-thin pixel modules stacked on top of each other. This design enables us to perform depth of interaction (DOI) measurements and maximises the spatial resolution along the line of flight of the two photons emitted within a patient. This will result in improved image quality, contrast, and uniformity while drastically reducing backgrounds within the scanner; allowing for a reduction in the amount of radioactivity delivered to the patient. Due to the large expected data rate of about 250 MB/s, a custom readout system for high data throughput has been developed, which includes noise filtering and multiple stages of data suppression and buffering. The realisation of a first scanner prototype for small animals is foreseen by 2019, and has been funded by the SNSF (Swiss National Science Foundation).

2 Scanner layout

The scanner has been specifically designed for usage with commercially available small animal MRI machines such as the nanoScan MRI 3T [3]. The scanner and its supporting mechanics will be housed inside the MRI’s removable RF-coil, while all readout electronics will be housed outside the machine, where long kapton flex cables are used for interconnection. The scanner is ring shaped,
with an inner and external radius of 20 mm and 34 mm respectively, a total active length of 48.3 mm and total physical length of 100 mm. The scanner ring is divided into 16 identical but independent units called tower modules (figure 2(a)). A tower module is composed of a cooling block and a series of sensor modules which are stacked on top of each other. Each sensor module consists of a 50 µm thick, high Z, converter material foil glued to the reverse side of a 55 µm thick impedance controlled flex circuit. A 100 µm thick silicon pixel sensor is connected to the front side of the flex circuit using gold ball bumps, with diameters 50 µm before bumping (figure 3). Currently 60 layers of sensor modules are used per tower to maximise the scanners detection efficiency, while minimising the thickness, creating a superimposed sampling (sandwich) structure.

In comparison to the thick high density scintillating crystals (e.g. LYSO, LSO, or BGO) commonly used in commercial scanners, our tower modules are complex structures using silicon pixel sensors. The size and spacing of the towers in the scanner ring has been simulated to provide the best detection efficiency, while enabling space to provide the required cooling to the silicon sensors. For this reason, the tower design has a step like structure, as seen in figure 2(b). The detection layers are grouped together in stacks of 20 sensor modules. Each stack requires a different size of sensor module. The three sizes of sensor module have been named Type-A, -B, and -C where the only difference is their width along the radial axis, with widths of 7 mm, 9 mm and 11 mm respectively (figure 2(b)). As a result three versions of the same pixel sensor are required, whom share the same naming convention. The only notable different between the pixel sensor versions is their widths and therefore number of pixels.

Due to the process limitations of fitting all three sensor versions in a single reticule at IHP, the maximum achievable length of a single sensor is 24 mm. As a result, two sensors are placed end to end on their flex circuit per sensor layer. The sensors cover a total length of 48.3 mm including a 300 µm gap between adjacent sensors. For simplicity, it is assumed the edge periphery of each sensor will be 500 µ in all directions, leaving a total sensitive length of 46 mm. Each sensor version is segmented in the same way using 500 × 500 µm² pixels (figure 4). The total number of readout

Figure 1. (a) 3D CAD model of the insertable scanner ring. Tower Modules are coloured gold, while the cooling blocks are blue. (b) GEANT4 model implemented for basic simulations of scanner performance, only the towers modules are shown.
channels for the whole scanner is expected to be 1,474,560 pixels. Since the sensor is monolithic the readout electronics will be implemented into each pixel pad with a small amount of electronics in the sensor periphery. A conservative power consumption has been estimated to be 130 $\mu$W/mm$^2$ using electronic and thermal simulations. The chip peripheries along the 2 longest edges of the sensor is where the majority of heat will be dissipated from. For this reason, we have chosen to implement cooling blocks in between the Towers. Advances in 3D printing technology have enabled us to use 3d printable ceramics, such as Al-Nitride (heat conduction, 20 Wm$^{-1}$k$^{-1}$), to produce small cooling blocks with complex internal structures that do not require any complex assembly or leak testing. The final choice of ceramic is still under investigation detailed full scanner thermal simulations have been and continue to be performed. A final choice in material will be made after the full size prototype sensors have been characterised. The submission of these sensors for production at IHP is expected to be December 2017. The geometrical acceptance of the scanner has been calculated to be 78% assuming a point-like source in the center of the scanner based of the layout described above, with a single photon detection efficiency of 25% for perpendicular tracks with respect to the z-axis (through the ring structure) and radial axis.

![Figure 2](image1.png)

**Figure 2.** (a) A Tower Module with cooling block. The Tower consists of a sandwich structure of up to 60 Layers of 100 $\mu$m monolithic silicon pixels, 55 $\mu$m of kapton flex, and 50 $\mu$m of a high Z converter. (b) Tower layer stack drawing for scale.

![Figure 3](image2.png)

**Figure 3.** Cross-sectional view of a single detector layer. From bottom up, Grey: 50$\mu$m Lead Foil, Green: glue (3–5 $\mu$m), Orange: 55 $\mu$m flex circuit, Gold: gold ball bonds (10–25 $\mu$m after compression), and Blue: 100 $\mu$m silicon pixel sensor.
Figure 4. Top view illustration of Type-A pixel sensor showing the $500 \times 500 \, \mu m^2$ pixel pad structure. The blue lines represent the connection between each pixel to the front-end electronics located in the periphery. Three versions of this pixel sensor are required to increase the geometrical coverage of our scanner. Where each version is identical with the exception of a different width and therefore number of pixel rows, Type-A, -B, and -C, with widths of 7 mm, 9 mm and 11 mm respectively.

3 Sensor layout

Traditional silicon pixel detectors focus on accurate position measurement often disregarding hit time resolution. Whereas for this project, a time resolution of the order of 30 ps is required to accurately measure the TOF (time-of-flight) of a photon in a small animal PET scanner. To achieve this we have designed our own silicon pixel sensor. At the time of publication, various small test sensors with specific functionality have been produced and tested. Our sensor is designed to be a Monolithic CMOS sensor, using IHP’s SG13S Si-Ge Bi-CMOS technology, which allows germanium to be implanted into the in pixel/periphery circuitry. Both of these technology choices allow for drastic reductions in various negative parasitic effects compared to externally connected amplifier/discriminator circuits and take advantage of higher charge carrier mobility speeds in Si-Ge compared to Si to create faster front-end circuitry. The well structure of the CMOS sensor has been designed to allow for the application of higher reverse bias’s compared with more traditional monolithic devices. Allowing the bulk of the sensor to be fully depleted with a uniform internal electrical field, maximising charge carrier mobility collects the signal charge as soon as possible. The baseline sensor option is a 100 $\mu m$ thick CMOS sensor with relatively large pixels of $500 \times 500 \, \mu m^2$. This choice of sensor design introduces a series of challenges at the level of the front-end electronics. Such as working with detector capacitances as large as 1 pF/mm$^2$, while keeping noise levels low enough to minimise jitter on the ToT (Time over Threshold) and ToA (Time of Arrival) at the preamplifier. All on chip discriminated signals are multiplexed using a Time-to-Digital Converter (TDC) featuring 30 ps time binnning.

The main contributions to the time resolution of the sensor are:

$$\sigma^2_t = \sigma_{th}^2 + \sigma_n^2 + \sigma_{TDC}^2,$$

(3.1)

where $\sigma_{th} \propto \frac{n}{dV/dt}$ is the time-walk contribution proportional to the noise level N, $\sigma_n = \frac{n}{dV/dt}$ is the contribution of the electronic noise and $\sigma_{TDC} = \frac{\text{binning}}{\sqrt{12}}$ which is the TDC contribution and is expected to be negligible compared to other error sources. As a result, a fast pulse rise-time and low equivalent noise charge (ENC) are crucial parameters for the design of front-end electronics with a time resolutions of the order of a few tens of picoseconds.
The amplifier developed for the sensor is based on hetero-junction (Si-Ge) BJT from IHP microelectronics with a current gain $\beta = 900$ and a transition frequency $f_t = 250$ GHz. An ENC of about 500 electrons and a rise-time less than 1 ns have been measured on $1 \times 1 \times 0.1 \text{ mm}^3$ pixels with a 1 pF capacitance. As a proof of principal, a first measurement of the time resolution achievable using a 100 $\mu$m thick p-in-n pixel sensor coupled to external Si-Ge HBT transistors developed for use with another project was performed at CERN’s SPS pion beam facility. The time of flight was measured between two sensors of a known fixed position, operated with a fixed pulse arrival threshold of 2.3 mV. The time resolution was measured to be $106 \pm 1$ ps for 180 GeV pions [4]. Detailed electronic simulations were then performed, reflecting our sensors monolithic structure and higher $\beta / f_t$ values which result in improved time resolution. This combined with the larger charge and signal-to-noise ratio, caused by 511 keV photons, leads to time resolutions compatible with 30 ps.

4 Readout system

The Readout system is designed to be easily scalable and therefore has a modular structure, a simplified block diagram of this is shown in figure 5. From right to left, sensor layers in each detection tower are grouped into so called super modules. A typical super module is composed of 5 consecutive sensor module/layers attached electrically to a long flat dataflex circuit/cable. All the sensors in a super module are daisy-chained together inside the dataflex cable to simplify readout. In total, a standard detection tower will be composed of 12 super modules (4 of each type). All the dataflex cables for a given tower are connected to their own small custom FPGA board called the Tower Control (TC) board. Each TC board provides all electrical and communication services to its super modules, 8b10b encoding, and provides a first stage of temporary data storage/buffering. Given the current scanner design 16 TC boards will be used. All 16 TC boards are connected (electrically) to a single versatile link demo board (VLDB) board [5], which handles data multiplexing, and a second stage of data buffering. The VLDB provides an 4.8 Gb/s optical GBTx connection to send all the multiplexed data via one or more optical links to a single commercial Xilinx ZC706 [6] FPGA board, named, Central Trigger Processor (CTP). The CTP controls the data flow between itself and the TC boards and distributes a L0 160 MHz base clock to the rest of DAQ system. The CTP is interfaced to a computer for data storage and control via PCIe connection. The use a multi-stage data suppression and time-walk corrections make it possible to limit the fraction of random coincidences to be less than 1% of the total events which pass both stages of data suppression.

The current readout system is designed to handle a 50 MBq source placed in the center of the scanner providing the best geometrical coverage and resulting in a single photon detection rate of 19.2 MHz. Simulations have shown 10 MHz of these events pass the trigger check. Of these events, 2.4 MHz or 1.2 MHz of coincidences are expected to pass the second and final stage of data suppression check, which are then sent to the PC for data storage. This is equivalent to a total data flow of 1.7 Gbit/s from the readout system to the computer.

5 Scanner simulation

Dedicated Monte Carlo studies have been performed using FLUKA [7] and GEANT4 [8] to optimise the design of the detection towers. The scanner was subsequently simulated in GEANT4 using
Figure 5. Block diagram showing a simplified full readout chain from control PC (far left) to super modules (far right).

Em option4 [9] with a multiple scattering range factor of 0.001 to accurately reproduce single event scattering.

The GEANT4 simulation provides the hit information in the scanner in terms of position, time, and energy deposition in each sensor pixel for every event. A custom analysis tool was created to process this information in the following ways: the real hit position is transformed to the center-of-gravity of its corresponding pixel. The true hit time is smeared according to a gaussian distribution depending on the amount of energy deposited to account for the energy dependent response of our electronics.

The expected data rate for the scanner and detection towers was simulated and used to guide the development of the readout DAQ system. A point-like source of \( F^{18} \), with an activity of 50 MBq was simulated in the center of a cylindrical water volume with \( R=18 \) mm, to simulate the presence of a small animal with a radioactive tracer roughly 10x stronger than what would normally be used. The requirement for a single hit in the scanner is an energy deposition of at least 20 keV inside the sensor. While for a coincidence, two single hits have to be detected in a time window of less than 1 ns (\(|t_1 - t_2| < 1 \) ns). The hits have to occur in separate detection towers and the resulting Line of Response (LOR) must intersect with the water cylinder. This results in a coincidence rate of 3 MHz for a 50 MBq source, which corresponds to a sensitivity of 6%.

Figure 6(a) shows the hit distribution along the x-y axis. Most of the interactions are in the inner-most sensor layers closest to the source, which exponentially decreases along the radial axis. The 1D profile for a single tower module is shown in figure 6(b). The step like structure is an artefact of the scanner geometry. The TOF distribution of two hits in a coincidence is reported in figure 7(a). The width of the core distribution is 35 ps. Contributions to the TOF resolution come from the depth of interaction (DOI), detector time resolution, Compton scattering, \( \gamma \)-ray acollinearity, and the electron path in the detector after conversion. The extended tails are due to the energy deposition spectrum shown in figure 7(b).

The TOF core distribution (35 ps) stated above, can be improved using DOI information and the high granularity of the pixel sensors using, \( \Delta t_{\text{corr}} = \Delta t - \Delta \text{depth} \). Where \( \Delta \text{depth} \) is the difference in depth between the two hit distances, and \( \Delta t \) is the measured time difference between the coincidence hits. After correction, the core sigma width of the TOF distribution reduces to 29 ps.
Figure 6. (a) Distribution of hits in the scanner with an energy deposition of more than 20 keV. (b) Expected data rate per layer, the step-like structure is a result of the scanner geometry.

Figure 7. (a) Distribution of the time difference between two hits in a coincidence. (b) Distribution of the energy deposited in the silicon sensor.

6 Reconstruction

To study the expected performance of the scanner, such as the resolution and the possible dependence of the resolution along the radial axis. A first attempt at reconstructing point-like sources using five $^{18}F$ sources placed along the x-axis of the scanner in positions: 0, −4 mm, −8 mm, −12 mm, and -16 mm in a water cylinder of radius 18 mm placed at the center of the scanner. A common reconstruction algorithm named the Filtered Back Projection (FBP) algorithm was used. The reconstruction was performed with and without TOF information. Figure 8 shows the resulting reconstruction images, when using (right) and not using (left) TOF information. The inclusion of the TOF information clearly improves peak separation at each point-source location. The average resolution value, expressed as the width of the Gaussian used to fit the peaks, is 290 µm and 370 µm with and without TOF information respectively. Moreover, the resolution does not degrade as the hits are moved towards the scanner (along the radial direction). The observed variation in peak height is due to the change in geometrical efficiency. This is caused by the gaps between towers i.e. where the cooling blocks are located.
Figure 8. Reconstructed images of five point like $^{18}$F sources along the $x$ axis. (a) 2D $x$-$y$ profile without TOF information and (b) with TOF information. The corresponding 1D profiles with and without TOF information are shown in (c) and (d) respectively.

7 Conclusion

The development of a small animal TOF-PET scanner designed to be insertable in commercial MRI scanners is well underway, new novel technologies such as replacing the conventional high density, thick scintillating crystals, with multi-layered structures composed of high Z converter foils, ultra-thin flex circuits, and thin monolithic pixel sensors with embedded Si-Ge amplifiers are being explored. We are targeting 1,474,560 small $500 \times 500 \mu m^2$ pixels with a timing resolution of 30 ps. These improvements will translate to higher quality PET images with the aid of Depth of Interaction (DOI) corrections. Due to the large number of readout channels, a custom scalable readout DAQ system is under development. This aims at being easily scalable and provides excellent background rejection, data suppression, and real-time image processing. We have simulated the TOF measurement performance for the scanner to be 35 ps using Monte Carlo simulations, and the
first attempts of reconstructing point-like sources in a phantom structure show cleaner images with reduced gaussian point spread functions. Moreover, the quality of the reconstructed image does not degrade along the radius of the scanner.

References

[1] SNSF grant, *TT-PET project*, CRSII2-160808.
[2] https://www.ihp-microelectronics.com/en/start.html
[3] http://www.mediso.com/products.php?fid=2,11-pid=84
[4] M. Benoit et al., *100 ps time resolution with thin silicon pixel detectors and a SiGe HBT amplifier*, 2016 *JINST* **11** P03011.
[5] https://ep-news.web.cern.ch/content/versatile-link-demonstrator-board-vldb-evaluation-kit-gbt-ecosystem
[6] https://www.xilinx.com/products/boards-and-kits/ek-z7-zc706-g.html
[7] http://www.fluka.org/fluka.php
[8] http://geant4.web.cern.ch/geant4/
[9] https://twiki.cern.ch/twiki/bin/view/Geant4/LowePhysicsLists
[10] S.R. Cherry, J.A. Sorenson and M.E. Phelps, *Physics in nuclear medicine*, Elsevier, Amsterdam, The Netherlands (2012).