Operation and performance of the ATLAS semiconductor tracker in LHC Run 2

The ATLAS Collaboration

DOI
10.1088/1748-0221/17/01/P01013

Publication date
2022

Document Version
Final published version

Published in
Journal of Instrumentation

License
CC BY

Citation for published version (APA):
hhttps://doi.org/10.1088/1748-0221/17/01/P01013

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Operation and performance of the ATLAS semiconductor tracker in LHC Run 2

The ATLAS collaboration

ABSTRACT: The semiconductor tracker (SCT) is one of the tracking systems for charged particles in the ATLAS detector. It consists of 4088 silicon strip sensor modules. During Run 2 (2015–2018) the Large Hadron Collider delivered an integrated luminosity of 156 fb$^{-1}$ to the ATLAS experiment at a centre-of-mass proton-proton collision energy of 13 TeV. The instantaneous luminosity and pile-up conditions were far in excess of those assumed in the original design of the SCT detector. Due to improvements to the data acquisition system, the SCT operated stably throughout Run 2. It was available for 99.9% of the integrated luminosity and achieved a data-quality efficiency of 99.85%. Detailed studies have been made of the leakage current in SCT modules and the evolution of the full depletion voltage, which are used to study the impact of radiation damage to the modules.

KEYWORDS: Charge transport and multiplication in solid media; Particle tracking detectors (Solid-state detectors); Radiation damage to detector materials (solid state); Solid state detectors

ArXiv ePrint: 2109.02591
1 Introduction

The ATLAS experiment [1] has been collecting data to study the interactions of elementary particles since 2009, using proton-proton (\( p p \)) collisions provided by the Large Hadron Collider (LHC) [2] at CERN. The LHC produces the highest-energy proton beams in the world, providing a unique...
opportunity to search for physics beyond the Standard Model (SM), as well as to precisely measure SM processes at very high energy. The first operational period in 2009–2012, called Run 1, yielded physics data at centre-of-mass energies ($\sqrt{s}$) of 7 and 8 TeV. After the first long shutdown (LS1) in 2013–2014, dedicated to upgrading the accelerator and the detectors, Run-2 operation began in 2015 at $\sqrt{s} = 13$ TeV and continued until 2018. The total integrated luminosity of $pp$ collisions delivered to the ATLAS experiment by the LHC during Run 2 corresponds to 156 fb$^{-1}$. In addition to $pp$ collisions, special physics data-taking runs for heavy-ion physics were performed. The second long shutdown (LS2) in 2019–2021 is for machine and detector upgrades towards Run-3 operation, which is scheduled to start in 2022.

The ATLAS detector is a multipurpose detector composed of the inner detector (ID), the calorimeters and the muon spectrometer. The ID is located in an axial magnetic field of 2 T generated by a superconducting solenoid, and performs measurements of charged-particle trajectories and determines their momenta. The electromagnetic and hadronic calorimeters are outside the solenoid and measure the energies of electrons, photons and hadrons. Good hermeticity of the calorimetry is ensured by the forward calorimeters, enabling a precise measurement of the missing transverse momentum for each event. The muon spectrometer is the outermost part of the detector. It provides muon identification and momentum measurement using toroidal magnets, as well as a muon-based trigger.

The ID is divided into three subsystems: the pixel detector (Pixel), the semiconductor tracker (SCT) and the transition radiation tracker (TRT). The Pixel provides a four-layer measurement of charged-particle tracks, where the innermost layer is the insertable B-layer (IBL), which was installed during LS1 [3, 4]. It is located closest to the interaction point (IP) and hence it has a great impact on the reconstruction of primary and secondary decay vertices. The SCT is a precise silicon microstrip detector, which extends the tracking volume to radial distances of $299 < r < 560$ mm.\textsuperscript{1} It plays an important role in measuring the transverse momenta of charged particles. The outermost ID system is the TRT, consisting of 50 layers of drift tubes. It is used for charged-particle tracking and for electron identification, utilising transition radiation.

As shown in figure 1, the SCT consists of four layers of silicon-strip sensors in the barrel (numbered 3 to 6), and nine disks (1 to 9) in each of the endcaps. Each barrel layer consists of 12 rings of modules along the $z$-direction, with a total of 2112 barrel modules [5]. There are 988 modules per endcap [6], with each disk consisting of up to three rings, referred to as inner, middle and outer, at increasing radii from the beam-pipe. Each barrel module consists of four rectangular 285-$\mu$m-thick $p^+$-on-$n$ silicon sensors [7] with strips of pitch 80 $\mu$m. The sensors have 770 strips of length $\sim$6 cm, of which 768 are read out via wire bonds. The unbonded outermost strips are present for field uniformity at the sensor edges. Two sensors on each side of a module are wire-bonded together to form 12-cm-long strips. A second pair of identical sensors is glued back-to-back with the first pair at a stereo angle of 40 mrad. Each module also has two arrays of six ABCD chips [8], where each chip is wire-bonded to 128 strips, in order to amplify, shape and discriminate pulses from the charges collected on the strips. The ABCD chips are placed on hybrid boards, which are bridged over the sensors. A binary threshold of around 1 fC in the detected charge is used to define

\textsuperscript{1}ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point in the centre of the detector and the $z$-axis along the beam-pipe. The $x$-axis points from the interaction point to the centre of the LHC ring and the $y$-axis points upwards. Cylindrical coordinates $(r,\phi)$ are used in the transverse plane, $\phi$ being the azimuthal angle around the beam-pipe. The pseudorapidity $\eta$ is defined in terms of the polar angle $\theta$ as $\eta = -\ln \tan (\theta/2)$. 


a ‘hit’ in the SCT. A barrel module has dimensions of 6 cm × 12 cm. Most of the endcap modules have similar dimensions, but have a trapezoidal shape, except for the ‘short’ endcap modules on the inner rings of disks 2–6 and also on the middle ring of disk 8, which consist of just one sensor on each side and therefore have only half the length.

The sensors were provided by two manufacturers: Hamamatsu Photonics (HPK) and CiS. Most of the sensors have a crystal lattice orientation (Miller indices) of $<111>$, while 92 modules in the barrel, provided by HPK, have $<100>$. The difference in crystal orientation results in slightly different performance, such as different noise characteristics; this is discussed in detail in section 3. The silicon wafers produced by CiS to manufacture short endcap modules were oxygenated to increase the radiation hardness of these innermost sensors [7].

The SCT is cooled to about 0°C by an evaporative cooling system using C$_3$F$_8$ fluid [9]. A heater pad system is intended to provide a thermal barrier between the SCT and the TRT, because the TRT operates at room temperature. However, it was only partially operative in the barrel region; hence barrel layer 6 was operated 6°C warmer than the rest of the SCT to prevent mechanical stress on the TRT. The cooling system was originally driven by compressors; these were replaced by a thermostiphon system [10] in 2018. The SCT position was monitored using frequency-scanning interferometry in Run 1 [11]. In Run 2, the SCT alignment was performed fully offline using reconstructed tracks [12].

The SCT has been fully operational since the commissioning period using cosmic-ray data in 2008 [13]. It showed excellent performance during LHC Run 1, as reported in ref. [14]. In Run 2,

---

2 Hamamatsu Photonics K.K., 1126-1, Ichino-cho, Higashi-ku, Hamamatsu, Shizuoka, 435-8558, Japan.
3 CiS Forschungsinstitut für Mikrosensorik GmbH, Konrad-Zuse-Strasse 14, 99099 Erfurt, Germany.
Figure 2. Event display projected onto the $r$–$\phi$ plane, taken by an unbiased random trigger in a typical $pp$ collision run at $\sqrt{s} = 13$ TeV in May 2018. This event has approximately 60 $pp$ interactions, which is nearly the maximum experienced in Run 2. The outer four circles correspond to the four barrel layers of the SCT, while the inner four circles are the Pixel layers. The lines emerging almost radially from the centre (interaction point) represent reconstructed tracks of charged particles with transverse momentum greater than 0.5 GeV.

the experimental environment became more challenging for the SCT, because an instantaneous luminosity up to $2.1 \times 10^{34}$ cm$^{-2}$s$^{-1}$ was achieved, corresponding to about twice the design value. Consequently, the average number of $pp$ interactions per bunch crossing (pile-up), increased and the SCT was exposed to a high density of charged particles, as shown in figure 2. This affected the SCT data transmission rate and the SCT data quality, as discussed in section 2.

Radiation damage was also a key feature of SCT operation in Run 2. The radiation received by the SCT modules until the end of Run 2 is estimated to be equivalent to a fluence of up to $5.6 \times 10^{13}$ 1 MeV neutrons [15, 16]. The operational strategy for the SCT therefore needed to be modified to account for the changes in sensor properties caused by radiation damage. At the same time, this provided valuable data for studies of radiation damage to silicon sensors in a hadron collider experiment; in particular, the SCT uses $p^+$-on-$n$ type silicon sensors, for which the $n$-bulk of all sensors was expected to invert into quasi-$p$-type (type inversion) during the second half of Run 2.

Several improvements to the SCT system are discussed in this paper. The operation of the SCT in Run 2 is described in section 2, and its performance is assessed in section 3. The effects of radiation on the detector are discussed in section 4.

2 Operation of the SCT in LHC Run 2

The typical cycle of operation of the SCT in Run 2 followed the same procedure that had been established in Run 1. During periods of beam injection, energy ramping and optimisation, the SCT was put into the standby state, where the low voltage (LV) supplied to the ABCD chips was fully maintained, but the high voltage (HV) to the sensors was reduced to 50 V. This is sufficiently low so as to
avoid electrical breakdown if there was any significant beam loss, which could cause many particles to enter the SCT simultaneously. Once stable beam conditions were declared, the HV was automatically raised back to the operational voltage (typically 150 V), provided that the LHC collimators were at their predefined positions for physics data-taking, the background rates measured by the beam conditions monitor [17], the beam loss monitor and the forward detectors were low enough, and the hit occupancy of the SCT was as expected for the instantaneous luminosity. This ‘warm start’ procedure took about 30 seconds before the SCT reached its operational HV state. When the LHC beam was dumped, the HV was immediately ramped back down to its standby state. An operator in the ATLAS control room was responsible for monitoring the SCT data acquisition (DAQ) system, detector control system (DCS) and data quality (DQ). In case of problems, the operator attempted to recover normal operation by following predefined procedures or by calling an SCT expert if necessary.

The experimental conditions in Run 2 were challenging due to high pile-up, resulting in up to ~2% hit occupancy in SCT modules during periods of operation at the maximum LHC luminosity. This resulted in a very high load on the DAQ system. There was significant radiation damage to the SCT sensors, which required modification to the operational strategy. Nevertheless, the SCT was available for 99.9% of the integrated luminosity during Run 2. In terms of the DQ efficiency, which is the fraction of data usable for physics analyses relative to the recorded integrated luminosity, it achieved 99.85%. This was due to several improvements, in particular to the DAQ system in 2015–16, which are discussed in more detail in the following sections. As shown in table 1, 98.6% of the SCT elements remained active until the end of Run 2. During LS1, two of the modules that had been disabled during Run 1 were recovered. An additional 14 modules were disabled in Run 2. For ten of these modules, this was attributed to HV-related problems such as trips of the HV and unstable HV currents. The remaining four modules were disabled due to LV-related problems or other problematic on-detector components.

The operational HV for most sensors was kept at 150 V until 2017. In 2018, a small drop in hit efficiency of ~0.5% was observed. This was related to an increase in the full depletion voltage ($V_{FD}$), due to radiation damage. Based on HV scan data (discussed in detail in section 3.1), the operational HV of modules located in regions of high radiation was increased to 200–250 V. In Run 1, it was reported that 27% of the modules with CiS-manufactured sensors exhibited anomalous fluctuations of up to ~200 $\mu$A in their leakage current during $pp$ collisions [14], and consequently the operational HV of these modules was reduced by 30–40 V to suppress these fluctuations. This condition was retained at the start of Run 2. In 2018, however, their behaviour was found to be more stable, since the nominal leakage current of ~1000 $\mu$A exceeded these fluctuations. Hence, the operational HV of these modules was set back to the nominal value in order to recover their hit efficiency.
To keep the temperature of the SCT close to 0°C, the evaporative cooling system was operated stably at −20°C to −14°C, except for barrel layer 6, which was set to −10°C (see section 1). The cooling temperatures set depended on the location within the SCT, because of variations in the hydraulic impedance among the cooling pipes, and the effects of gravity, which resulted in higher fluid concentrations in the lower part of the detector. These temperatures were maintained throughout most of Run 2, except for two weeks during each winter shutdown period, when the evaporative cooling system was turned off for maintenance work and the modules warmed to room temperature, and also part of 2015. During that time, early in Run 2, the temperature of the evaporative cooling system was set to −10°C, because increased humidity within the ID volume resulted in a higher dew point around the SCT. The cause of this was identified as an over-pressure safety valve to the ID volume, which was unexpectedly open. This problem was fixed in October 2015, after which the temperature was restored to the nominal value. This small change in temperature had no impact on detector operation.

2.1 Data acquisition

The DAQ system of the SCT uses a binary readout architecture. Charge created in the sensor is immediately digitised and stored in a 132-cell-deep pipeline within the ABCD chip. On receipt of a trigger, the column of binary data associated with the bunch crossing is read out and transmitted to the off-detector DAQ system.

Figure 3 shows a schematic diagram of the DAQ system [14, 18]. It comprises 128 Readout Driver (ROD) and Back of Crate (BOC) cards. The BOC provides the optical interface between a ROD and up to 48 SCT modules; for each module the BOC transmits the clock and trigger via a single command stream and receives data back via two optical links (FE-links) from the six ABCD chips on each side of the module. For each level-1 (L1) trigger, the ROD processes the incoming data from up to 96 FE-links and combines those data into a single event fragment. The event fragment is then transmitted at the level-1 trigger rate on an optical link (S-link) [19] to the global DAQ system of ATLAS, where the data are buffered until a higher-level trigger decision is made. The commands and triggers, as well as a 40 MHz clock, are distributed by the trigger, timing and control (TTC) interface module.

Both the transmission (TX) and receiver (RX) optical paths incorporate redundancy mechanisms which can be used in the case of failure of individual links [19]. The TX redundancy allows the module to be configured to use the trigger command stream from the adjacent module; the RX redundancy allows the module to transmit data from all 12 ABCD chips using one of the two data links. The use of optical link redundancy at the end of Run 2 is shown in table 2. Although the total number of links using redundancy mode increased during Run 2, the fractions are still only 1.3% (1.9%) for the TX (RX).

In the following subsections, specific DAQ developments during Run 2 are discussed in more detail.

2.1.1 Optical data transmission

The dominant failure mechanism for the DAQ system during Run 1 was the failure of individual channels within the 12-channel array of the vertical-cavity surface-emitting laser (VCSEL) for the TX [19]. The loss of a channel meant that the module no longer received the clock and trigger
Tables and Figures

**Figure 3.** Schematic diagram of the SCT DAQ [14]. The front-end module is located on the detector, while the ROD, BOC and TTC Interface Module are located in the underground cavern outside the detector.

**Table 2.** Numbers of optical links configured using the redundancy mode at the end of Run 2. The numbers in parentheses are the status at the end of Run 1 [14].

<table>
<thead>
<tr>
<th>Link type</th>
<th>Barrel</th>
<th>Endcaps</th>
<th>Total</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>TX</td>
<td>36 (9)</td>
<td>19 (5)</td>
<td>55 (14)</td>
<td>1.3% (0.3%)</td>
</tr>
<tr>
<td>RX</td>
<td>53 (41)</td>
<td>99 (91)</td>
<td>152 (132)</td>
<td>1.9% (1.6%)</td>
</tr>
</tbody>
</table>

The loss of data fromWant to add a sentence here signals, resulting in the loss of data from that module. Although the TX redundancy could be applied to re-enable the module, failures of two adjacent channels resulted in the loss of data from one module until that TX plug-in was physically replaced. The peak rate of channel deaths in Run 1 was as high as 10–12 per week, and TX redundancy was used for up to 220 modules.

A number of TX replacement campaigns were carried out; failures in the original TX installation were attributed to inadequate precautions against electrostatic discharge by the manufacturer, and then in the second installation to the ingress of humidity due to a non-hermetic seal. A third installation, carried out early in 2012, used a new VCSEL product manufactured with a complete hermetic seal, leading to a dramatic reduction in the channel death rate. Nonetheless, a small rate of deaths continued (~5–10 per year) due to a coefficient of thermal expansion mismatch between the epoxy and the surface of the VCSEL array, leading to cracks in the VCSEL and the subsequent ingress of humidity.

During LS1, a final batch was produced incorporating the commercially available VCSEL array (10 Gbps LightABLE surface-mount parallel fibre-optic engine, Reflex Photonics), using a re-engineered TX plug-in to match the LightABLE to the BOC. Three of the eight ROD crates were fully populated with LightABLEs, while the BOCs in the remaining five crates continued to be equipped with the original TX, and they operated throughout Run 2. The LightABLEs have proved to be very reliable; only two channel deaths occurred, of which one was infant mortality. In the remaining ROD crates, failures were only about 2–3 per year, which is easily compensated for by...
2.1.2 Developments for operation with high pile-up

The SCT was designed to operate with 0.2%–0.5% occupancy in its 6.3 million sampled strips at an LHC pile-up rate of up to $\mu \sim 23$. The fundamental bottlenecks that restrict data-taking with increased occupancy, arising from higher pile-up, are the bandwidth limitations of the FE-links (which transmit bit streams at 40 Mbps) and the S-links (which transmit 32-bit words at 40 MHz, a throughput of 1.28 Gbps). For the FE-links, exceeding the bandwidth leads to a loss of data from the corresponding module. As shown in figure 4, the fraction of errors issued by the ABCD chips increased for high pile-up and trigger rates, in particular for modules relying on RX redundancy. If data throughput on the S-link exceeds ~90% of the bandwidth, the ROD starts to receive back-pressure from the S-link and asserts a busy flag (BUSY) to suspend the entire data-taking of the global DAQ system [18] via the TTC interface module. Extrapolations of the occupancy with increasing pile-up during Run 1 suggested that data flow in the FE-links and S-links would exceed bandwidth limitations at pile-up rates of $\mu \sim 87$ and $\mu \sim 33$ respectively, at a level-1 trigger rate of 100 kHz. These projections assumed optimum on-chip data compression: only hits from the in-time bunch crossing that exceed a threshold were read out, while hits from the preceding bunch crossing were simultaneously vetoed (the so-called ‘01X’ mode). In addition, the standard data-packing scheme for the ROD was used, which included readout of the hits from each strip over three consecutive bunch crossings. In any case, FE-links using the RX redundancy would reach the bandwidth limitation at a much lower pile-up level, because all data would then be read out through only one of the data links.

The TX redundancy mechanism. It is envisaged that this configuration will be maintained in Run 3, as the very low death rate does not justify the significant effort and risk involved in changing the TXs in the remaining five crates.
Table 3. Mitigation steps deployed to address SCT bandwidth limitations.

<table>
<thead>
<tr>
<th>Action</th>
<th>Type</th>
<th>Date (LS)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expanded DAQ</td>
<td>Hardware</td>
<td>2014</td>
<td>New ROD and BOC production</td>
</tr>
<tr>
<td>FE-link remapping</td>
<td>Hardware</td>
<td>2015</td>
<td>Reshuffle of data links to BOCs</td>
</tr>
<tr>
<td>Data compression</td>
<td>Firmware</td>
<td>2016</td>
<td>Improved data compression on the ROD</td>
</tr>
<tr>
<td>Chip masking</td>
<td>Software</td>
<td>2017</td>
<td>Implementation of on-the-fly chip masking</td>
</tr>
</tbody>
</table>

**Figure 5.** Number of FE-links exceeding the bandwidth limitation as a function of \( \mu \), at a level-1 trigger rate of 100 kHz. In the nominal configuration (black curve at right), signals are read out by two FE-links and therefore one FE-link transmits signals from up to six ABCD chips. For the FE-links using RX redundancy, where one FE-link transmits signals from up to 12 ABCD chips on a module, the dashed middle (solid left) curve corresponds to the case with chip masking enabled (disabled).

It was anticipated late in Run 1 that the LHC would provide higher luminosity, causing more pile-up. Furthermore, it was discovered early in Run 2 that the SCT hit occupancy had already increased significantly compared to Run 1; this was attributed to secondary interactions with material in the newly installed IBL services. In order to achieve a comfortable margin of safety for operation with pile-up at \( \mu \sim 60 \), the four mitigation steps listed in table 3, and described below, were developed.

**Expanded DAQ:** To address the pile-up limit imposed by the S-links, the numbers of RODs and BOCs, and consequently the number of S-links, were increased from 90 to 128 during LS1, thereby reducing the number of modules serviced by each ROD from 48 to no more than 36.

**FE-link remapping:** The mapping of the FE-links to the BOCs was re-optimised in order to ensure a more uniform data load on each ROD. This resulted in a flatter occupancy distribution across the S-links, removing spikes from the highest-occupancy S-links and thereby improving the stability of data-taking.
Data compression: A new highly efficient data-packing scheme (the ‘supercondensed mode’) was implemented on the RODs routinely from 2016 onwards. In this mode, clusters of up to 16 strips are packed into a single 16-bit word, resulting in a typical data size reduction of ~25% during \( pp \) collisions compared to the expanded mode. This data size reduction is achieved at the expense of some error flags which were superfluous to efficient data-taking. It significantly reduces the data size used to encode large clusters, which otherwise unduly bias the distribution.

Chip masking: This mechanism was developed in order to mask the noisiest chips during operation with the highest level of LHC pile-up. Lists of chips with the highest noise were prepared so that various chips could be disabled when different pile-up levels were exceeded. The value of \( \mu \) was continuously monitored by the LHC, and published within the ATLAS DAQ framework. This allowed reductions in the volume of hit data, so as to remain below the bandwidth limitations of the data links at the cost of a negligible loss of data. The mechanism worked ‘on the fly’, gradually reducing the number of masked chips to zero as the LHC luminosity, and hence the rate of pile-up, decreased during a data-taking run, effectively suppressing the ABCD chip errors, as shown in figure 4. The principle of chip masking has been demonstrated successfully, leading to all FE-links being operable within bandwidth constraints up to \( \mu = 60 \), as shown in figure 5. This concept was first developed to cope with high-pile-up conditions in 2017. At this time the number of LHC bunches was reduced in order to avoid unexpected beam dumps, which had been occurring frequently during 2017, and consequently the average \( \mu \) was increased to maintain the luminosity. However, this situation was mitigated in 2018, leading to a return to nominal pile-up conditions for the remainder of Run 2. Chip masking was therefore not deployed routinely; however, this capability is expected to be a valuable asset in efforts to maintain tracking efficiency if the level of pile-up increases significantly in Run 3.

The impact of the different mitigation steps related to S-links is further illustrated in figure 6, using extrapolations of measured data occupancy as a function of pile-up. When using the configuration from 2016, all S-links are seen to operate below a bandwidth occupancy of 90%, which shows that the bandwidth limitation problem was successfully solved.

2.1.3 Online module recovery

Due to intense radiation from the \( pp \) collisions, a charged particle occasionally passed through front-end electronics on the detector, causing an instantaneous error. Energy deposited in the \( p-i-n \) diodes in the modules, which receive the clock and trigger signals, resulted in desynchronisation of the module relative to the global ATLAS DAQ system. An automated module recovery mechanism monitored desynchronisation errors in data buffered downstream of the S-link and reconfigured desynchronised modules. This entire process, from detection of the desynchronisation error to the resumption of synchronised data arriving from the module, took 20–30 seconds. In high-pile-up conditions (\( \mu \sim 60 \)), module recovery was typically activated a few times per minute.

Single-event upsets (SEUs) within the internal threshold register of an ABCD chip caused the chip to become noisy or inefficient. If this happened on the configuration register, the readout
Figure 6. Number of S-links operating above the bandwidth occupancy threshold specified on the x-axis. During operation, once the S-link occupancy exceeds 90% of its bandwidth, the corresponding ROD starts to fall into the busy state, so that it becomes impossible to transfer data correctly. The bandwidth occupancy in 2012 (green line, right), 2015 (dashed black, middle) and 2016 (blue, left) is simulated using the extreme condition of a level-1 trigger rate of 100 kHz with \( \mu = 60 \). The line labelled 2012 shows the 90-ROD system used in Run 1, where many S-links were not able to handle the high data rate. Several S-links for endcap modules showed significantly higher occupancy than the other S-links, due to an imbalanced distribution of data rate per ROD. These features were significantly improved in the 2015 configuration with 128 RODs. In 2016 the supercondensed mode was introduced and the FE-link remapping was performed, in order to distribute the S-link occupancy more uniformly.

Figure 7. Evolution of the number of noisy chips in a run with and without global reconfiguration activated, for two typical runs in Run 2.

sequence of the ABCD chip was interrupted. The rate of SEUs was studied during Run-1 operation [14]. As expected, a higher SEU rate was observed as the hit occupancy increased. This could be understood approximately from earlier studies using test-beam data [20], and indicated that errors induced by the SEUs would become more important in higher-luminosity operation in
Run 2. Since it is difficult to identify these problematic chips online, a system was implemented to apply a global reconfiguration to the entire SCT DAQ system, instead of using a targeted module recovery mechanism. This process was activated every ~90 minutes and took ~1.2 seconds, so the loss of data is expected to be negligible. Figure 7 shows the number of noisy chips (defined as at least 120 channels with continuous high noise) as a function of time since the beginning of the run. Typically, the number of noisy chips was reduced to less than five at any time.

2.1.4 Handling ROD-busy assertions
Towards the end of Run 1, the SCT started to have an average data-taking inefficiency of ~0.9%, caused by an increase in the number of permanent busy assertions coming from the RODs. The impact on the ATLAS data-taking efficiency was limited by the automatic exclusion of the ROD from data-taking within about 0.2 seconds of the onset of BUSY. However, the re-integration of the ROD into data-taking mode imposed about 7 seconds of BUSY, as the ROD must be masked from incoming triggers during its reconfiguration. The overall time that the ROD was excluded from data-taking, from the initial ROD removal to its full re-integration, was ~20 seconds, including the ~7.2 seconds of dead time to ATLAS. The permanent ROD-busy signals were induced by unrecognised illegal incoming bit patterns on the FE-links. The impact on the DQ of taking data while the ROD was excluded was evaluated offline (discussed in section 2.3.1).

This problem was significantly mitigated early in Run 2 by ROD firmware improvements, as well as by imposing a limit on the bit-stream length for incoming events on the FE-links. Towards the end of Run 2, a ROD busy occurred once per day in typical pp run conditions, and the average SCT dead time was typically less than 0.1%.

2.2 Detector calibration
Routine calibration of the SCT was performed to keep the hit efficiency above 99% and the noise occupancy below $5 \times 10^{-4}$. The basic procedure is described here, while more details can be found in refs. [14, 18]. Optical calibrations were typically run first, to establish reliable communication for sending and receiving data via the data links. Digital tests were then performed to exercise and verify the functionality of the front-end chips, the mask registers and the pipeline buffers. Analogue optimisation was the main part of the calibration procedure, consisting of the trim range scan and the response curve test.

The trim range scan found the best configuration of the digital-to-analogue converters (DACs)\(^4\) to give a threshold equivalent to 1 fC for each ABCD chip. This was achieved by injecting a fixed 1 fC test charge into the chips and measuring the variation in hit occupancy while changing the threshold. The occupancy was expected to be 100% when the threshold was sufficiently low and 0% when the threshold was high. The optimal setting was found in the DAC configuration where the hit occupancy was 50% ($\upsilon_{50}$). This measurement also provided an estimate of the noise, from the Gaussian spread of the hit occupancy as a function of the threshold, which was modelled by an error function. In the response curve test, a similar threshold scan was performed using ten different values for the charge injected by the ABCD chips. The gain was estimated from the slope

\(^4\)There are two types of DACs implemented in each ABCD chip. An 8-bit DAC changes the threshold setting for all 128 strips belonging to the chip. Variations in individual strips are compensated by a 4-bit DAC. Details are fully described in refs. [14, 18].
of the curve relating the injected charge to vt50, and the input equivalent noise charge (ENC) was calculated from the output noise divided by the gain.

The calibration procedure also provided a list of problematic or unusable strips, which was uploaded to the conditions database to be used in the offline data reconstruction. Another important calibration was the noise occupancy test, in which the noise occupancy was measured for ten different threshold settings between 0.3 fC and 1.0 fC. This test was done occasionally to understand the noise level in terms of the noise occupancy. Results from the noise occupancy test as well as the response curve test are discussed in section 3.3.

The calibration procedure was typically performed every few weeks. Since the HV needed to be set to the nominal values during calibration, to ensure operational detector response, it could only be done without beam circulating in the LHC, for safety reasons in case of unexpected beam losses. The times available for calibration were dedicated calibration periods between LHC fills, and longer intervals in LHC operations to allow technical interventions. Due to improvements, the time between LHC fills was gradually reduced to about 1.5 hours. This allowed more physics data to be provided in the later years of Run 2, but the calibration time was then significantly limited. To speed up the calibration, the number of scan points for the trim range scan was reduced by 50%, for example, while still retaining reliability and robustness. The total time for the full SCT calibration was reduced to 1.5 hours. This allowed a calibration frequency of at least once per month, and the fraction of noisy strips was kept below 0.3%, as discussed in detail in section 2.3.2.

2.3 Data quality

2.3.1 Data quality monitoring and assessment

To assure high-quality data for physics analyses, it is essential to monitor the DQ during data-taking. Evaluation of the SCT DQ was based on the ATLAS experiment’s common DQ framework, which is summarised in ref. [21]. Several important parameters were monitored using histograms produced during the fast processing of sampled events [22]. These included information about the SCT performance, such as hit efficiency and occupancy, and also information from reconstructed tracks, such as the number of clusters associated with a track. These histograms were automatically produced and evaluated by the online data-quality monitoring framework [23], which would warn the DQ monitoring operators when a potential problem was identified, based on predefined criteria.

When the data had been fully reconstructed, the DQ was thoroughly assessed offline. When a problem was found, a ‘defect’ flag was set for the corresponding luminosity block. This is a period of time (typically 1 minute), during which the instantaneous luminosity, the detector and trigger configurations, and the other important run conditions are considered to be constant. The defect flags are stored in a database [24]. Several different defect flags were prepared, and classified as being either ‘intolerable’ or ‘tolerable’, depending on the severity of the problem and its impact on the data quality. The SCT-specific DQ flags were mostly the same as those used in Run 1 [14].

Data flagged with an intolerable defect are excluded from physics analyses. For the SCT, three major types of intolerable defect were defined:

- Significant acceptance loss for the tracking, because multiple RODs were excluded from the DAQ.
Table 4. SCT DQ defect summary in Run 2.

<table>
<thead>
<tr>
<th>Year</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luminosity recorded by ATLAS [fb⁻¹]</td>
<td>3.6</td>
<td>35.5</td>
<td>46.5</td>
<td>60.2</td>
<td>145.8</td>
</tr>
</tbody>
</table>

Defect description | Fraction of luminosity affected [%]

<table>
<thead>
<tr>
<th>Intolerable defects</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Major acceptance loss due to ROD exclusion</td>
<td>0.22 0.01 0.01 0.01</td>
</tr>
<tr>
<td>DAQ crate(s) excluded from readout</td>
<td>0.01 0.01 0.01 0.01</td>
</tr>
<tr>
<td>SCT HV not ready</td>
<td>0.01 0.10 0.04 0.01 0.04</td>
</tr>
<tr>
<td>SCT simulation data flag activated</td>
<td>0.01 0.01 0.01 0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tolerable defects</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 40 modules noisy</td>
<td>6.0 13 20 51 31</td>
</tr>
<tr>
<td>Minor acceptance loss due to ROD exclusion</td>
<td>14 5.1 0.68 0.55 2.0</td>
</tr>
<tr>
<td>Low efficiency</td>
<td>4.3 0 1.4 1.6 1.2</td>
</tr>
<tr>
<td>&gt; 40 modules with link errors</td>
<td>2.4 1.6 0.34 0 0.57</td>
</tr>
<tr>
<td>SCT not at standard HV</td>
<td>0.58 0.08 0.08 1.2 0.54</td>
</tr>
<tr>
<td>Trip of one power supply crate</td>
<td>0 0 0 0.06 0.03</td>
</tr>
</tbody>
</table>

- One or more DAQ crates (each holding about 12.5% of the modules) were excluded from the DAQ.

- The SCT HV was not in the ‘ready’ state: the HV was less than the operational setting, which was typically 150–250 V depending on module location.

The fraction and evolution of intolerable SCT defects during Run 2 is summarised in table 4. In 2015, the most significant DQ inefficiency was caused by multiple ROD failures. This was because the trigger rate was significantly higher than that in Run 1. This problem became less severe in later years, due to the ROD firmware improvements described in section 2.1.4. From 2016 onwards, high DQ efficiency was maintained. Manual adjustments to the SCT were sometimes necessary, for example when unexpected changes in beam conditions occurred. These led to inevitable small delays in turning the SCT back on, resulting in some data being recorded with intolerable defects because the SCT voltage was not ready. In 2018, a one-time incident resulted in significant DQ loss. An SCT simulation-mode data flag was unintentionally activated during a physics run. This inhibited further SCT data reconstruction, and consequently no track was provided to the high-level trigger. Software changes have been implemented to prevent this from happening again. Through the whole of Run 2, the DQ inefficiency due to SCT intolerable defects was 0.15%, corresponding to 217 pb⁻¹.

The tolerable SCT defect affecting the most integrated luminosity originated from having more than 40 noisy modules, where a noisy module is defined by the average noise occupancy being greater than 0.15%. The frequency of this defect increased in the later years of Run 2. Other tolerable defects were:
• Minor acceptance loss for tracking, because a single ROD was excluded from the DAQ.

• Hit efficiency less than 98% on average, and less than 99% when measured using only the first bunch crossing of a bunch train.

• More than 40 modules with errors in the FE-link.

• SCT being operated at non-standard HV values. This happened rarely, when the HV was scanned to study radiation damage to the sensors. The HV was kept sufficiently high that the impact on tracking is considered marginal.

• At least one power supply crate tripped (where one crate holds ~1.14% of the modules).

This final defect flag was introduced in 2018, because of a few incidents related to a tripped power supply crate observed during physics runs. This was attributed to aging of the components. The degraded components will be replaced to mitigate this problem in Run 3.

2.3.2 Noisy strip identification

Noisy strips were searched for using a special data stream triggered on empty LHC bunch crossings, so that no collision hits should be present. Events were triggered randomly at a rate of 2–10 Hz, giving sufficient data to determine the number of noisy strips run-by-run. When a strip had a noise occupancy greater than 1.5%, the strip was identified as being noisy and masked during offline data reconstruction.

Figure 8 shows the fraction of noisy strips as a function of date, for both Run 1 and Run 2. Typically, this number rose from about $10^{-4}$ to $3 \times 10^{-3}$, until a calibration was performed and the fraction of noisy strips dropped again. This was due to resetting the thresholds, which drift as charge is accumulated in the chips. On a shorter timescale, large run-by-run variations are observed. The main reason for these fluctuations was bit flips in the threshold register, as discussed in section 2.1.3. Although these strips were usually recovered by a global reconfiguration during the run, they are identified as noisy when the average noise occupancy in the run exceeds the threshold.

3 SCT performance

The performance of the SCT had to be maintained in the harsh experimental environment of LHC Run 2, with high levels of pile-up and radiation. To maximise the tracking efficiency for charged particles, the SCT is regularly calibrated to maintain a hit efficiency above 99%, while the noise occupancy remains below $5 \times 10^{-4}$. Hits found on contiguous SCT strips are grouped into a cluster, which forms the input to the tracking algorithm. It is therefore important that the distribution of cluster sizes remains stable.

These key parameters are studied and discussed in this section, using the abundant data sample collected during Run 2. Since the radiation levels depend on the location within the SCT, some performance results are shown separately for different regions. For convenience, the parameter $\eta_{\text{index}}$ is introduced to specify the position along the $z$-axis of the 12 modules forming a stave of the SCT barrel: $\eta_{\text{index}} = -6, \ldots, -1, 1, \ldots, 6$, with increasing $z$-coordinate. For the endcaps, the terms ‘inner ring’, ‘middle ring’ and ‘outer ring’ specify the radial positions of the modules.
3.1 Hit efficiency

3.1.1 Hit efficiency measurement

For a given SCT sensor, a hit efficiency for each track is defined by the equation

$$\varepsilon = \frac{N_{\text{cluster}}}{N_{\text{cluster}} + N_{\text{hole}}} \quad (3.1)$$

where $N_{\text{cluster}}$ and $N_{\text{hole}}$ denote, respectively, the number of clusters and the number of holes associated with the track. If no cluster is found along the track in an SCT layer where one is expected, its absence is counted as a hole, unless there is a known reason for this absence (such as disabled modules, links or chips, or transient errors), in which case it is excluded from the hit efficiency calculation. However, an exception is made for disabled or noisy strips, which can still be counted as a hole.

The method used to measure the hit efficiency is the same as that discussed in ref. [14]. The calculation uses only well-reconstructed tracks in the ID. Events with more than 500 tracks are rejected, because of the high probability of reconstructing fake tracks from incorrect combinations of clusters in such events. Furthermore, the tracks used are required to satisfy the following:

- transverse momentum $p_T > 1$ GeV;
- track fit quality identified as ‘good’;\(^5\)
- transverse impact parameter $|d_0| < 10$ mm;
- number of clusters on SCT sensors, excluding the sensor under consideration, $N_{\text{cluster}}^{\text{SCT}} \geq 6$;
- number of holes in Pixel detector, $N_{\text{hole}}^{\text{Pixel}} = 0$;
- number of clusters in Pixel IBL or B-layer (second innermost Pixel layer), $N_{\text{cluster}}^{\text{Pixel}} \geq 1$;
- incident angle with respect to SCT module surface, $|\phi_{\text{inc}}| < 40^\circ$.

\(^5\)All tracks must have a fit quality satisfying $\chi^2/N_{\text{dof}} < 3$ ($N_{\text{dof}}$ = number of degrees of freedom). Details are given in ref. [25].
In Run 2, the track-finding required the total number of clusters (holes) in the Pixel and SCT combined to be $N_{\text{Si\_cluster}} \geq 7$ ($N_{\text{Si\_hole}} \leq 2$). To avoid introducing any bias into the hit efficiency measurement from these requirements, only tracks that would be reconstructed regardless of the state of the sensor under consideration are used. Therefore, when excluding the sensor under consideration, $N_{\text{Si\_cluster}} \geq 7$ and $N_{\text{Si\_hole}} \leq 1$ are required. Additionally, any cluster unassociated with the track, but found within a distance $D_{\text{cut}}$ around the corresponding track position, is included in $N_{\text{cluster}}$ and excluded from $N_{\text{hole}}$ in eq. (3.1). The value of $D_{\text{cut}}$ is set to 200 $\mu$m unless stated otherwise.

Figure 9(a) shows the hit efficiency for each barrel layer and each endcap disk measured in a $pp$ collision run in July 2018, with a pile-up value $\mu$ between 25 and 50. The average hit efficiency measured using all bunches in a train ranges between 97% and 98%, which is slightly lower than the expected hit efficiency for the SCT sensors. This is due to the ‘01X’ readout mode, defined in section 2.1.2. A real hit is lost when, by chance, the strip also has another hit in the preceding bunch. If the hit efficiency is measured using only the first bunch crossing in each bunch train (called ‘first BC’), it is around 99%. The result using the first BC is therefore considered to be the intrinsic hit efficiency of the SCT sensor, while that determined using all bunches is also important because it impacts the tracking performance. Thanks to the redundant design of the SCT, consisting of multiple barrel layers and endcap disks typically providing eight SCT clusters per track, this 1% loss of hit efficiency has a negligible impact on the tracking efficiency. The fraction of bad strips in each layer, including both the problematic strips identified by the calibration (section 2.2) and the noisy strips (section 2.3.2), is superimposed on figure 9(a). There tends to be anti-correlation between the hit efficiency and the fraction of bad strips, indicating that the non-uniform distribution of hit efficiency across the SCT regions can be partially attributed to the bad strips.

Figure 9(b) shows the hit efficiency measured using barrel modules as a function of time since the start of a run, with the evolution of the level of pile-up superimposed. The intrinsic hit efficiency found using only the first BC is constant and remains above 99%. The difference between the intrinsic hit efficiency and the hit efficiency measured using all bunches is about 2% at the
beginning of the run, when $\mu = 50$, decreasing to about 1% at the end of the run, when $\mu = 25$. From this, the inefficiency induced by the 01X readout mode is seen to be proportional to the level of pile-up, as expected.

Systematic uncertainties in the hit efficiency are estimated as described in ref. [14]. All uncertainties are symmetrised by taking the larger of the upward and downward variations. Uncertainties arising from inclusion of unassigned clusters within $D_{cut}$ are estimated by varying the value of $D_{cut}$ between 0 and 500 $\mu$m. The change in the measured efficiency when a cluster is also required on the other side of the sensor in the same module is taken to be another systematic uncertainty. The intrinsic hit efficiencies using the first BC in a low-pile-up run are measured to be (99.00$\pm$0.53)% (endcap A), (99.26$\pm$0.21)% (barrel) and (98.83$\pm$0.51)% (endcap C). These values are consistent with 99% and the impact on the tracking efficiency is negligible.

### 3.1.2 Interstrip hit efficiency

A precise study of the hit efficiency as a function of position, for charged particles passing in between two strips of an SCT sensor (referred to as the interstrip hit efficiency), was performed using test beams [26] before the construction of the SCT. A slight drop in the hit efficiency was observed in the centre of the gap between two strips, after irradiation and in the presence of a magnetic field. A comparable feature can be investigated using data collected during Run 2 for the real SCT sensors, which have been irradiated during ten years of LHC operations. Figure 10 shows this efficiency as a function of relative $x$, the distance of a track from one strip towards the next. Position resolution results in an uncertainty of about 20 $\mu$m [14] in the estimate of relative $x$ for each track. For this measurement, only modules in barrel layer 3 with $|\eta_{index}| = 1$ are chosen, because they are located in the innermost layer, where high radiation fluence is expected, and they have tracks almost perpendicular to the sensors. In order to avoid inefficiencies from other sources, only modules that are fully active and have HV greater than 150 V are used. An incident angle of $\phi_{inc} = -5^\circ$ (with the direction defined in the same sense as for global $\phi$) was chosen because it is close to the Lorentz angle [14]. A small dip of 0.8% is observed in the interstrip hit efficiency in the middle of the gap between two strips, qualitatively reproducing the result obtained from test-beam data.

### 3.1.3 HV dependence of hit efficiency

During the second half of Run 2, the $n$-type bulk of the silicon sensors was expected to invert into quasi-$p$-type. For type-inverted sensors, the depletion region develops from the backplane towards the strips. The operational HV therefore needs to be sufficiently higher than $V_{FD}$ to allow the depletion region to reach the strips and maximise the charge collection efficiency. In order to confirm the optimal HV, as well as to study radiation damage effects, the HV dependence of the hit efficiency was occasionally measured during physics runs.

Figure 11 shows comparisons of hit efficiency as a function of HV for modules with $|\eta_{index}| = 1$ in barrel layer 3. The data were collected throughout Run 2, between November 2015 and September 2018. Until May 2017, the efficiency curves remained almost unchanged, rising sharply as the HV is increased. In November 2017, however, the increase in efficiency with HV was slower, as expected from type inversion. Between November 2017 and April 2018, this increase became slightly faster again, due to beneficial annealing which occurred during the year-end technical shutdown. Around the end of Run 2 (September 2018), the start of the efficiency plateau was at around 140 V, close
Figure 10. Interstrip hit efficiency as a function of relative $x$, the distance from one strip towards the next, with $\phi_{\text{inc}} = -5^\circ$. The measurement is performed using data from a low-pile-up run in July 2018. Only modules with $|\eta_{\text{index}}| = 1$ are used. All possible interstrip combinations are averaged. The blue shaded regions near 0 and 80 $\mu$m correspond to the strip implant positions.

Figure 11. Hit efficiency as a function of HV for modules with $|\eta_{\text{index}}| = 1$ in barrel layer 3, measured from November 2015 to September 2018.

to the original operational HV of 150 V. The hit efficiency was measured to be slightly lower in September 2018, and the operational HV was therefore increased to 250 V to recover the efficiency.

3.2 Cluster width and Lorentz angle

The cluster width is defined as the number of contiguous strips with hits that form a cluster. Tracks used to measure the cluster width must satisfy the following criteria, matching those used in the previous study [14]: $p_T > 0.5$ GeV, $|d_0| < 1$ mm, $N_{\text{SCT}}^{\text{cluster}} \geq 8$ and $N_{\text{Pixel}}^{\text{cluster}} \geq 2$. Other track selection requirements are the same as those used in the hit efficiency measurement (see section 3.1.1). Figure 12 shows cluster widths measured in each of the four barrel layers at a HV of 150 V. These
The mean cluster width distributions peak at 1 and have a long tail, with a mean width of about 1.2. A cluster width of 2 is induced by charge sharing between two neighbouring strips. The exponentially decreasing tail of the cluster width distribution up to ~10 is induced by δ-rays [14]; for the long tail component greater than ~10, contributions from multiple charged particles, where their hits are merged into one cluster, may be more prominent. Clusters with zero width are due to tracks with a hole in the corresponding barrel layer.

The HV dependence of the mean cluster width has been investigated using HV scan data collected during physics data-taking runs. The level of pile-up was constant to within 4%–7% during these HV scans. Figure 13(a) shows distributions of mean cluster width as a function of $\phi_{\text{inc}}$ for barrel layer 3 at five different HV values. The mean cluster width increases as the HV is increased up to 75 V, but then it decreases by a few percent as the HV is increased further to 150 V. This behaviour suggests there is a trade-off between better charge collection at higher HV, increasing...
Figure 14. Mean cluster width as a function of $\phi_{\text{inc}}$ for barrel layers 3 to 6 in a typical $pp$ collision run in August 2018. Modules with crystal orientation (a) <111> and (b) <100> are shown separately. For geometrical simplicity, only the sides with no stereo angle are included.

the cluster width, and faster charge collection, which decreases the cluster width by suppressing transverse charge diffusion. To evaluate the HV dependence of the cluster width, only tracks with $\phi_{\text{inc}} = [-5.0^\circ, -4.5^\circ]$, close to the minimum in cluster width, are used. Figure 13(b) shows the HV dependence of the mean cluster width measured in barrel layer 3, from data collected between 2015 and 2018 in Run 2. Until May 2017, the cluster width increased rapidly as the HV was increased to around 50 V. Since November 2017, however, the increase of the cluster width has become slower. This is consistent with the observed change in the HV dependence of the hit efficiency, discussed in section 3.1.3, and appears to be induced by type inversion. The peak in the mean cluster width around 70–100 V is more visible in the later curves; further investigation is needed to understand this feature.

The displacement from zero of the position of the minimum cluster width (MCW) in each of these distributions, $|\phi_{\text{MCW}}|$, gives a good estimate of the Lorentz angle, $\theta_L$, resulting from the deviation of the drift-charge direction from the electric field vector, due to the Lorentz force acting on the charge carrier in the axial magnetic field. The Lorentz angle in a magnetic field $\vec{B}$ is given by

$$\tan \theta_L = \gamma_H \mu_d |\vec{B}|,$$

where $\gamma_H$ and $\mu_d$ denote the Hall factor and the charge-carrier mobility, respectively. The charge-carrier mobility is affected by the HV, the thickness of the depleted region and the temperature. The Lorentz angle is important in the barrel modules, where the electric field from the HV is perpendicular to the magnetic field from the solenoid. When the incident angle of a particle equals the Lorentz angle, all induced charge drifts along the particle direction, giving the minimum cluster width. Only negatively charged tracks are selected, because the tilt of barrel SCT modules relative to the radial direction results in positively charged tracks populating only the lower end of the $\phi_{\text{inc}}$ distribution, below the minimum. The track selection requirement is lowered to $N_{\text{SCT}}^{\text{cluster}} \geq 7$, because the tighter requirement was found to bias the shape of the cluster width distributions.

The data are fitted using a convolution of the function [14]

$$f(\phi_{\text{inc}}) = a |\tan \phi_{\text{inc}} - \tan \phi_{\text{MCW}}| + b$$  \hspace{1cm} (3.2)
with a Gaussian distribution. The fitted parameters are $\phi_{MCW}$, $a$, $b$ and the width of the Gaussian distribution. Typical fit results are shown in figure 14; modules with crystal orientations of <111> and <100> are shown separately, because differences in the charge-carrier mobilities between these two types of sensor are expected to result in different values for the Lorentz angle. The displacements of $\phi_{MCW}$ from zero correspond to a shift of $\sim$10 µm in the position of the cluster on the sensor surface. The effect of the Lorentz angle must therefore be taken into account for precise tracking with the SCT. The layer-dependence of the minimum cluster width was marginal in Run 1 [14]. However, figure 14 clearly shows a large layer-dependence during Run 2, because the hit occupancy was significantly higher, resulting in a greater probability that hits induced by multiple charged particles are merged into a single large cluster. Although the four barrel layers are almost equally spaced, the minimum cluster width in barrel layer 6 is almost equal to that in barrel layer 5. This is due to the higher temperature of barrel layer 6 (see section 1), which induces faster diffusion of the charge carriers, and consequently wider dispersion over the sensor surface.

Figure 15 shows the variation of $\phi_{MCW}$ measured in barrel layer 3 between 2015 and 2018, corresponding to the evolution of the Lorentz angle. No corrections have been made to account for changes in parameters such as the HV setting or the cooling temperature; however, these were stable up to September 2018, when the operational HV was intentionally raised from 150 V to 250 V. The systematic uncertainty in $\phi_{MCW}$ is estimated by fitting an alternative asymmetric function, in which the parameter $a$ in eq. (3.2) is allowed to have different values above and below the minimum. The deviation from the nominal fit is taken to be the systematic uncertainty. During Run 2, $|\phi_{MCW}|$ increased continually, with a change of about 0.5° between 2015 and 2018. This is presumably induced by radiation damage to the sensors, which modifies the space-charge distribution and changes the mobility. For a sensor thickness of 285 µm, this change of 0.5° corresponds to a shift of $\sim$1 µm in the cluster position. This is negligibly small because the intrinsic position resolution of the SCT is about 20 µm [14]. However, in Run 3, where the total fluence of radiation will increase by a factor of two, monitoring of changes in the Lorentz angle may become more important.

**Figure 15.** Time dependence of $\phi_{MCW}$ between 2015 and 2018 in barrel layer 3. Only the sides with no stereo angle are shown. The error bars are statistical, while the shaded bands show the systematic uncertainties. The time when the nominal operational HV was raised from 150 V to 250 V is indicated by the vertical dashed line.
The final $\phi_{MCW}$ measurement shown in figure 15 corresponds to the period after the operational HV of all modules in barrel layer 3 was raised to 250 V, which reduced $|\phi_{MCW}|$ by about 0.7°. This behaviour is expected, because increased HV results in lower charge-carrier mobility [27], and thus a smaller deflection in the magnetic field. The change in the HV setting is taken into account when the SCT cluster position is used to determine a space point for tracking.

### 3.3 Noise measurement

One of the crucial conditions for maintaining high tracking efficiency is to keep the readout thresholds as low as possible. This is possible only when the input noise levels are sufficiently low. It is therefore important to understand the noise levels of radiation-damaged sensors during detector operation. Two methods are used to evaluate the sensor noise in dedicated SCT stand-alone calibration runs: (i) the response curve test, which measures the input equivalent noise charge (ENC, with the unit ‘number of electrons’), and (ii) the noise occupancy test. The former method reflects the width of the Gaussian-like noise distribution at around one standard deviation ($\sigma$), while the latter measures the tail of the noise distribution at around $4\sigma$ when a threshold is set at 1 fC, and it is therefore more sensitive to external noise sources.

Distributions of the chip-averaged response-curve noise and noise occupancy at a HV of 150 V are shown in figure 16(a) and figure 16(b), respectively. Results from February or March of 2015 and from September or November 2018 are compared, corresponding to the start and end of the Run-2 physics data-taking period. Results from the response-curve test are also summarised numerically in table 5. During Run 2, most of the chip-averaged noise levels increased marginally by 10%–20%.
The response curve test also provides measurements of the amplifier gain, as shown in figure 16(c). The gain decreased gradually by around 10% during Run 2. Since calibration runs were carried out several times each year (see section 2.2) to adjust the threshold to be 1 fC, the observed decrease in gain had only a marginal impact on the hit efficiency.

Figure 17(a) shows the evolution of ENC for barrel layer 3 and for endcap A disk 4, where measurements made through each of the two FE-links on either side of a module (referred to as ‘link-0’ and ‘link-1’) are shown separately. The endcap modules show significantly higher noise in link-0 than in link-1; the difference creates a double peak in the noise distributions shown in figure 16. Figure 17(b) shows the ENC for all modules in endcap A. Almost all of the HPK modules, except those in disks 1 and 9, have more noise in link-0, while the CiS modules show only a marginal difference. In the endcaps, all the modules are oriented so that the side with link-1 is attached to the support frame, while the other side with link-0 is generally exposed to the air. In the case of disks 1 and 9, however, the side with link-0 also faces the support structure or the plate of the thermal enclosure. It is therefore evident that the exposed HPK sensors show systematically higher levels of noise; however, its source is not yet clear. Nevertheless, the ENC level is at most 2000 electrons (0.32 fC), which is sufficiently low compared to the threshold of 1 fC.

In contrast to the endcap modules, barrel modules do not show a noise-level difference between link-0 and link-1. The noise from sensors with the <100> crystal orientation is consistently lower than that from sensors with the <111> orientation, reproducing the results from earlier studies presented in ref. [28].

### 3.3.1 HV dependence of noise

The HV dependence of the noise levels was measured periodically throughout Run 2 using response curve tests. Figure 18(a) shows results from tests performed between February 2013 and May 2019. In November 2016 a knee-like structure started to appear in these curves, and the HV value corresponding to this knee increased with time until November 2018. The same trend is observed for both crystal orientations <111> and <100>, although the absolute level of the noise is significantly different. This knee-like structure also appears in a noise occupancy scan performed in May 2019, which is shown in figure 18(b).

The noise level is largely determined by the interstrip capacitance, which is the dominant component of the overall sensor capacitance [7], and it is sensitive to the properties of the silicon at the surface of the sensor. The development of the depletion region near the surface shows a complex evolution during type inversion, because of the non-uniform interstrip space-charge distribution. The appearance of the knee-like structure is predicted to be a qualitative indicator of type inversion; such behaviour was reported, for example, in ref. [28]. Its evolution results from changes in the full depletion voltage. As shown in section 4.2, type inversion in late 2016 is consistent with a more quantitative analysis using the $I-V$ relation.

---

6 How the endcap modules are attached to the disks is shown in figure 1. Modules staggered to the left are attached to the disk and face the collision point, while modules staggered to the right are attached to the disk and face away from the collision point. The side of each module on which link-1 is located is the side attached to the disk.
Table 5. Numerical results for noise and amplifier gain from the response curve test. Mean values from March 2015 and from September 2018 are compared for two FE-links on different types of modules, corresponding to the groups in figure 16.

<table>
<thead>
<tr>
<th>Module type</th>
<th>ENC [$\times 10^3$ electrons]</th>
<th>Gain [mV/fC]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>March 2015</td>
<td>September 2018</td>
</tr>
<tr>
<td></td>
<td>Link-0</td>
<td>Link-1</td>
</tr>
<tr>
<td>Barrels 3–5</td>
<td>1.49</td>
<td>1.48</td>
</tr>
<tr>
<td>&lt;111&gt; HPK</td>
<td>1.39</td>
<td>1.39</td>
</tr>
<tr>
<td>&lt;100&gt; HPK</td>
<td>1.53</td>
<td>1.54</td>
</tr>
<tr>
<td>Barrel 6</td>
<td>1.44</td>
<td>1.43</td>
</tr>
<tr>
<td>&lt;111&gt; HPK</td>
<td>1.16</td>
<td>1.10</td>
</tr>
<tr>
<td>EC inner</td>
<td>1.09</td>
<td>1.15</td>
</tr>
<tr>
<td>&lt;111&gt; HPK</td>
<td>1.57</td>
<td>1.46</td>
</tr>
<tr>
<td>EC middle</td>
<td>1.59</td>
<td>1.63</td>
</tr>
<tr>
<td>&lt;111&gt; HPK</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>EC middle short</td>
<td>1.62</td>
<td>1.54</td>
</tr>
</tbody>
</table>
Figure 17. Chip-averaged noise. Only modules with HV between 145 V and 155 V are shown. (a) Evolution of the response-curve noise for barrel layer 3 (top), the outer ring of endcap A disk 4 (middle) and the middle ring of endcap A disk 4 (bottom). Measurements from link-0 and link-1, from modules with different crystal orientations and also from modules manufactured by HPK and CiS, are shown separately. (b) Chip-averaged response-curve noise for all modules in endcap A. From top to bottom, outer, middle and inner rings are shown. Measurements from link-0 and link-1, and also from HPK modules and CiS modules, are shown separately. The nine segments separated by vertical dashed lines correspond to disks 1 to 9.

Figure 18. HV dependence of the noise for modules in barrel layer 3. (a) Response-curve noise measured between February 2013 and May 2019. (b) Noise occupancy measured in May 2019.
4 Radiation damage

During LHC operations, the SCT silicon sensors are irradiated by various species of particles with a wide range of energies, from the TeV-scale to thermal neutrons, resulting from $pp$ collisions at the interaction point, or from back-scattered secondary particles from the calorimeters or the accelerator collimators. After receiving radiation, the SCT performance is changed due to the creation of additional states within the semiconductor band gap. The SCT was designed to endure up to about 700 fb$^{-1}$ at a $pp$ collision energy of 14 TeV [29]. Up to the end of 2018 the accumulated integrated luminosity was 156 fb$^{-1}$ at 13 TeV during Run 2 and 29 fb$^{-1}$ at 7–8 TeV during Run 1, so a safe margin for further operation remains.

Nevertheless, the intense radiation flux in the environment of the ID has modified various SCT parameters such as the leakage current, the full depletion voltage and the probability of charge trapping. The expected fluence around the SCT is shown in figure 19, with the numerical values calculated at the positions of each barrel layer and endcap disk listed in table 6. These are based on simulation using the FLUKA [30] software package and a detailed description of the ATLAS detector geometry [31]. Inelastic $pp$ collisions were simulated using the PYTHIA 8 event generator [32, 33] with the MSTW2008lo parton distribution functions [34] and the A3 set of tuned parameters [35]. The integrated luminosities delivered by the LHC at $\sqrt{s} =$ 7, 8 and 13 TeV correspond to 5.6, 23.2 and 159.4 fb$^{-1}$, respectively. These slightly larger luminosity values than in ref. [21] result from the inclusion of beam collision periods outside the stable beam time when data collection occurred.

Both the total ionising dose (TID) and the non-ionising energy loss (NIEL) affect the performance of the SCT sensors. The TID as well as neutron fluences around the SCT were monitored by a dedicated online radiation monitoring system [36], in which the TID was measured using radiation-sensitive $p$-MOSFET transistors, while the neutron fluences were measured using forward- and reverse-biased $p-i-n$ diodes. There is good agreement between these measurements and the simulated radiation background in Run 2, as presented in ref. [37].
Table 6. Simulated 1-MeV-neutron equivalent fluence in units of $10^{13} \text{ cm}^{-2}$ up to the end of Run 2. Radiation backgrounds originating from $pp$ collisions are assumed to be symmetric about the origin along the $z$-axis. The $r$ and $z$ positions listed correspond to the centres of the modules.

| $|\eta|_{\text{index}}$ | $|z|$ [mm] | Barrel 3 ($r = 299 \text{ mm}$) | Barrel 4 ($r = 371 \text{ mm}$) | Barrel 5 ($r = 443 \text{ mm}$) | Barrel 6 ($r = 514 \text{ mm}$) |
|------------------------|------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| 1                      | 61.9       | 4.30                          | 3.41                          | 2.85                          | 2.45                          |
| 2                      | 184.0      | 4.37                          | 3.46                          | 2.87                          | 2.47                          |
| 3                      | 308.7      | 4.47                          | 3.51                          | 2.92                          | 2.52                          |
| 4                      | 431.2      | 4.53                          | 3.58                          | 3.00                          | 2.57                          |
| 5                      | 555.3      | 4.65                          | 3.71                          | 3.09                          | 2.66                          |
| 6                      | 679.1      | 4.80                          | 3.84                          | 3.20                          | 2.76                          |

| Disk | $|z|$ [mm] | Inner ($r = 304.6 \text{ mm}$) | Middle ($r = 396.5 \text{ mm}$) | Outer ($r = 499.4 \text{ mm}$) |
|------|------------|-------------------------------|-------------------------------|-------------------------------|
| 1    | 853.8      | –                             | 3.86                          | 3.07                          |
| 2    | 934.0      | 5.12                          | 3.89                          | 3.08                          |
| 3    | 1091.5     | 5.20                          | 3.98                          | 3.19                          |
| 4    | 1299.9     | 5.27                          | 4.14                          | 3.31                          |
| 5    | 1399.7     | 5.37                          | 4.21                          | 3.40                          |
| 6    | 1771.4     | 5.69                          | 4.47                          | 3.67                          |
| 7    | 2115.2     | –                             | 4.94                          | 4.01                          |
| 8    | 2505.0     | –                             | 5.42                          | 4.66                          |
| 9    | 2720.2     | –                             | –                             | 5.33                          |

The instantaneous leakage current increases are approximately proportional to the NIEL. Too large a leakage current might damage the SCT sensors due to thermal runaway. Radiation damage also modifies the doping concentration, which is proportional to $V_{\text{FD}}$. As damage accumulates, the number of donors in the $n$-bulk material decreases, while the number of acceptors effectively increases. This leads to type inversion of the $n$-bulk into $p$-bulk material, with a further increase of $V_{\text{FD}}$. After type inversion, the HV applied needs to be kept above $V_{\text{FD}}$ to achieve full hit efficiency. If $V_{\text{FD}}$ becomes too high, this might therefore lead to degradation of the tracking performance. Although these changes in the bulk properties were taken into account in the design of the SCT, improved understanding of these characteristics is important for safe operation of the SCT.

In this section, the evolution of the leakage current and the full depletion voltage are discussed; leakage currents from both the pixel detector and the SCT are reported in ref. [38]. Other effects of radiation damage, such as SEUs and changes in hit efficiency, cluster width and noise, are discussed in sections 2 and 3, respectively. Variation of the charge-trapping probability is also important, but its impact on the SCT at the end of Run 2 is expected to be marginal [39, 40] and it is therefore not discussed in this paper.
4.1 Leakage current

The main source of leakage current is thermally generated electron-hole pairs. The leakage current is therefore highly dependent on the sensor temperature, $T_{\text{sensor}}$, as well as on the accumulated radiation damage. In order to evaluate the evolution of the leakage current under the same conditions throughout all periods, the measured current on the HV supply line (HV current), $I_{\text{meas}}$, is normalised to a reference temperature, $T_{\text{norm}} = 0^\circ\text{C}$, using the relationship

$$
\left(\frac{I_{\text{leak}}}{I_{\text{meas}}}\right) = \left(\frac{T_{\text{norm}}}{T_{\text{sensor}}}\right)^2 \exp \left[\frac{E_{\text{gen}}}{2k_B} \left(\frac{1}{T_{\text{norm}}} - \frac{1}{T_{\text{sensor}}}\right)\right],
$$

where $I_{\text{leak}}$ is the normalised leakage current, $E_{\text{gen}}$ is the effective electron-hole pair generation energy of $1.21 \pm 0.01 \text{ eV}$ [41], and $k_B$ is Boltzmann’s constant.

The value of $T_{\text{sensor}}$ for each module is determined from a hybrid-board temperature, $T_{\text{hybrid}}$, measured using negative temperature coefficient thermistors mounted on the hybrid boards. However, there is an offset between these temperatures, $\delta T = T_{\text{hybrid}} - T_{\text{sensor}}$, due to the thermal impedance between the sensor and the thermistor, which depends strongly on the mechanical and thermal structure of each board. It differs from module to module, due to small variations in the thermal impedances among the hybrid circuits, sensors and cooling pipes. In Run 1 [14], $\delta T$ was estimated using thermal simulations based on the finite element method (FEM). Compared with the barrel modules, the endcap modules had larger uncertainties on these estimates. Moreover, a systematic difference in the raw leakage current between the two endcaps existed, which could not be explained by the FEM estimation. It was therefore difficult in Run 1 to estimate the correct endcap sensor temperatures using the thermistors. In order to improve the estimate of $I_{\text{leak}}$ for the endcap modules, a method to determine $\delta T$ using cooling temperature scans was developed during Run 2.

4.1.1 Development of temperature correction method

A value of $\delta T$ for each module was estimated using dedicated measurements, in which the cooling temperature was varied. With the LV supplies to the hybrid circuits turned off, $T_{\text{sensor}}$ was considered to be equal to $T_{\text{hybrid}}$, due to the absence of any hybrid heating sources. After turning on only the HV, the cooling temperature was then increased from approximately $-20^\circ\text{C}$ to $0^\circ\text{C}$. In this scan, the relation between $T_{\text{hybrid}} \approx T_{\text{sensor}}$ and $I_{\text{meas}}$ at various temperatures was measured for each module. Two representative examples are shown in figure 20(a), together with the results of fits using eq. (4.1). Once the cooling temperature had returned to its nominal value, the LV was then turned on. Since heat was generated by the hybrid circuit, $T_{\text{hybrid}}$ increased to become significantly higher than $T_{\text{sensor}}$, while $T_{\text{sensor}}$ remained sensitive to $I_{\text{meas}}$ and so could be deduced for each module using the fitted curve. Hence values of $\delta T$ were obtained.

A systematic uncertainty of $1^\circ\text{C}$ in $\delta T$ is assigned for all modules. This was estimated by performing an alternative fit using only the lowest pair of $T_{\text{hybrid}}$ data points and comparing the result with that from the full fitted curve. Smaller contributions arising from the stability of the leakage current measurements, the digitisation of the temperature measurements and the uncertainty in $E_{\text{gen}}$ are also included within this $1^\circ\text{C}$ uncertainty.

Distributions of $\delta T$ for modules in barrel layers 3, 4 and 5, and in each of the endcap disks are shown in figure 20(b). Barrel layer 6 cannot be cooled to a temperature lower than $6^\circ\text{C}$, as explained
in section 1, so modules in layer 6 were excluded from the $\delta T$ measurement. The average $\delta T$ for barrel modules is about 4°C. As expected, $\delta T$ is larger for the endcap modules, with broader distributions and averages in the range ~15°C–17°C. There is a difference of about 2°C between the average values of $\delta T$ measured in endcap C and in endcap A. There has been a difference in the cooling temperatures set for endcap A and endcap C since the initial commissioning of the SCT, in order to compensate for a measured difference in $T_{\text{hybrid}}$ between the two endcaps. The cause of this difference in $T_{\text{hybrid}}$ is unknown, but it may also result in the systematic difference in $\delta T$ between the two endcaps.

### 4.1.2 Leakage currents at the end of Run 2

Figure 21 shows normalised leakage currents per unit volume for all barrel and endcap modules in November 2018, at the end of Run 2, with the HV set to 150 V. Four modules with abnormally high currents and 42 permanently disabled modules have been excluded. These plots display a lateral view of the SCT, with the horizontal axis corresponding to the $z$-coordinate, and modules with the same $r$- and $z$-coordinates, but different azimuthal locations, grouped together within the same vertical bin.

Almost all modules in the same group have consistent leakage currents to within about 3%. Despite the difference of about 2°C between the average values of $\delta T$ found in the two endcaps, as shown in figure 20, their leakage currents agree once they are normalised to 0°C, giving confidence in the validity of this method for correcting the sensor temperatures. No appreciable differences in leakage currents are observed among sensors from different manufacturers, with different crystal orientations ($<111>$ vs $<100>$), or between the standard or oxygenated silicon wafers.

There is a trend toward larger leakage currents in regions at higher $|z|$ for all radii. In the barrel layers, the normalised leakage currents of modules at $|\eta_{\text{index}}| = 1$ (closest to $z = 0$ mm) are about 3% smaller than those of modules at $|\eta_{\text{index}}| = 6$ (near $|z| = 680$ mm). This is in contrast to the end of Run 1, when a small excess in leakage current was measured in the central region of barrel layer 3 and uniform distributions were observed in the other barrel layers [14]. The trend seen in figure 21 is consistent with the radiation fluences listed in table 6. These results are compared with

---

**Figure 20.** (a) Measured currents as a function of $T_{\text{hybrid}}$ for typical barrel and endcap modules with the LV off. The blue curves are obtained from fits to the measured points using eq. (4.1). The red point displaced from each curve shows the hybrid temperature measured with the LV on; the temperature differences $\delta T$ are indicated by the arrows. (b) Distributions of $\delta T$ for different categories of SCT modules.
Figure 21. Measured leakage currents per unit volume, normalised to 0°C, with HV at 150 V, for (a) barrel and (b) endcap modules in November 2018. Modules with the same $r$- and $z$-coordinates, but different azimuthal locations, are grouped together within the same vertical bin, arranged from $\phi = 0^\circ$ on the left to $\phi = 360^\circ$ on the right. Modules in endcaps A and C, as well as from different sensor manufacturers, are plotted in different colours. Horizontal solid and dashed bars indicate model predictions from the Hamburg model, using conversion factors estimated from FLUKA and Geant4 transport simulations. The main difference arises from the simplified detector geometry used by FLUKA, in comparison with the Geant4 simulation. The uncertainties in the predictions are about 15%, and are not shown in the plot to maintain visibility.

4.1.3 Evolution of leakage currents

Hybrid temperatures, and the voltages and currents of all HV power supplies were monitored continuously since the start of data collection and their values were stored in a database. The normalised leakage currents can therefore be calculated precisely throughout Run 2, with the $\delta T$ correction applied, and compared with variations in operating conditions. In LS1 and other technical shutdowns, as well as during unexpected power outages, when $T_{\text{hybrid}}$ data were not available, the temperature of the closest cooling pipe is used instead. This is a reasonable approximation, because there was no heat generation in the SCT volume during these periods.

Typically, 150 $I_{\text{meas}}$ data points are recorded for each module every hour. During most physics runs, $I_{\text{meas}}$ drops by $0.2\%$–$2\%$ as the instantaneous luminosity decreases, consistent with expecta-

The current-related Hamburg model has been commonly used to describe the annealing effects of the radiation-induced leakage current. It has an exponential and logarithmic term with parameters obtained experimentally in a series of accelerated annealing tests. Details of the parameterisation are discussed in ref. [39].
tions from self-heating of the sensor and heat from the front-end chips. A simple time-weighted average is used to determine a single value of the normalised leakage current for each module in each physics run. During LS1, technical stops and periods with no physics runs, $I_{\text{meas}}$ values recorded during calibration runs are used instead. These two sources of data show consistency to within 1%.

The HV current measured for a module depends on the HV applied to the sensor. Above the full depletion voltage, the HV current increases by several percent for each additional 100 V. Also, due to the filter resistance of about 12 kΩ on the HV supply line, the true voltage on a sensor is lower than the voltage applied by the power supply; the difference is about 20 V for modules in barrel layer 3, for which the operational HV was increased to 250 V in late 2018. This introduces an additional uncertainty of up to about 1% in these $I_{\text{meas}}$ values.

Figure 22 shows the evolution of normalised leakage currents during Run 2 for four representative groups of modules: modules at $\eta_{\text{index}} = 1$ in barrel layer 3, modules at $\eta_{\text{index}} = 6$ in barrel layer 6, outer ring of endcap C disk 9 and inner ring of endcap A disk 5. Values of $T_{\text{sensor}}$ were around $-1^\circ \text{C}$, $+5^\circ \text{C}$ and $-7^\circ \text{C}$ for modules in barrel layer 3, barrel layer 6 and the two endcaps, respectively, during data-taking periods, while they were $>15^\circ \text{C}$ during LS1 and winter technical shutdowns. As described in section 2, in 2015 all modules except those in barrel layer 6 were set about 5°C warmer to avoid condensation in the SCT volume. Leakage currents increased by more than one order of magnitude over the course of Run 2. The leakage current decreased by 20%–30% during each winter shutdown due to annealing.

Leakage current measurements are compared with two model predictions in figure 22: the Hamburg model [42, 43] and the Sheffield model [45, 46]. The Sheffield model predicts about 15% more leakage current than the Hamburg model. Systematic uncertainties in these model predictions are estimated by varying each parameter of the respective model by ±1 standard deviation and summing the components in quadrature. Additional uncertainties are also included in the temperature measurements (1°C) and the delivered luminosities (1.7%). Uncertainties from the FLUKA simulation are not included; however, these comparisons provide an estimate of the quality of the simulation as well as the uncertainty in the current-related damage factors. The model predictions are consistent with the data at a level of around 30%. The partial annealing which occurred during the winter shutdown periods is also modelled with similar precision. Discontinuities in the data-to-model ratios between years may be due to errors in the estimation of $T_{\text{sensor}}$, for which the $\delta T$ correction was performed only once a year during shutdowns, as well as from other imperfections in the model predictions.

### 4.2 Full depletion voltage

Since all SCT sensors are expected to have been type-inverted, it is essential to understand $V_{\text{FD}}$ and to predict its future evolution. If space charge is uniformly distributed within the bulk silicon, then the leakage current should be proportional to the depleted volume, which develops as a function of the square-root of the HV. Once full depletion has been reached, the leakage current remains nearly constant. Therefore, a point of abrupt change in the variation of $I_{\text{meas}}$ as a function of the HV ($I$–$V$ curve) provides an estimate of the value of $V_{\text{FD}}$.

---

*This is an alternative model to the current-related Hamburg model. Its current-related damage constant was obtained using actual SCT sensors irradiated at $-10^\circ \text{C}$. It parameterises the annealing behaviour using a sum of five exponential functions with time constants given by the proton and neutron irradiation tests. Exact formulae for both the Hamburg and Sheffield models are briefly summarised in appendix A of ref. [14].
Figure 22. Evolution of normalised leakage currents for four groups of modules: from left to right, barrel layer 3 (modules at $\eta_{\text{index}} = 1$), barrel layer 6 (modules at $\eta_{\text{index}} = 6$), endcap C (outer ring of disk 9) and endcap A (inner ring of disk 5). The main plots show leakage current data (red points) compared with predictions from the Hamburg model (blue line with uncertainty bands). Data points were measured during physics runs or in SCT stand-alone calibration runs. The lower two sets of plots show ratios of data to model predictions from the Hamburg and Sheffield models. Both models use the same conversion factors estimated from FLUKA and Geant4 transport simulations. Coloured bands correspond to $\pm1\sigma$ uncertainties in the model prediction, but uncertainties from the FLUKA simulation are not included.

Figure 23. Measurement of $V_{FD}$. (a) Evolution of the relation between the HV current and the HV ($I-V$ curve) for one module in barrel layer 3, from June 2009 to November 2019. (b) Example of linear fits made to an $I-V$ curve. The scale of the horizontal axis is linear for the square-root of the HV. The red and blue dots are the data points used for the two linear fits.
Figure 24. Distributions of $V_{FD}$ for modules in the four barrel layers. From top to bottom, results from November 2011, March 2017 and November 2018 are shown.

Figure 25. Evolution of $V_{FD}$ estimated from the $I$–$V$ curves. From left to right, results from barrel layer 3, outer ring of endcap disk 9 and inner ring of endcap disk 6 are shown. Green points show initial values of $V_{FD}$ determined from $C$–$V$ measurements. Predictions from the Hamburg model are shown by lines with model uncertainty bands for standard (red) and oxygenated (blue) silicon sensors.

Figure 23(a) shows the evolution of the $I$–$V$ curve from one of the modules in barrel layer 3, with 15 different measurements made between 2009 and 2019 superimposed. As expected, the $I$–$V$ curve consists of a lower region, in which $I_{\text{meas}}$ rises sharply, and a higher region where the rate of
increase is less. As shown in figure 23(b), separate linear fits are made to the data in each region, and the crossing point of these two lines is considered to be $V_{FD}$. Values of $V_{FD}$ obtained from all modules in the barrel region on three different dates are shown in figure 24. In November 2011, all barrel layers had similar values of $V_{FD}$ around 50 V. In 2016 and early 2017, almost all modules had small values of $V_{FD}$, indicating a significant decrease in the donor concentration in the silicon bulk. By late 2018, values of $V_{FD}$ had increased again, clearly indicating type inversion. Barrel layer 3 (innermost layer) has the largest values of $V_{FD}$, around 70 V, while barrel layer 6 (outermost layer) has the lowest, around 30 V.

A similar analysis was performed for various periods during Run 2. The results are shown in figure 25, separately for three representative groups of modules: barrel layer 3, the outer ring of endcap disk 9 and the inner ring of endcap disk 6 (oxygenated sensors). The evolution of $V_{FD}$ is compared with predictions from the Hamburg model, where radiation-induced changes of the space-charge concentration are expressed by four main components: removal of initial donors, creation of stable acceptors, creation of short-term acceptors and a reverse annealing term, where a constant space-charge distribution over the bulk volume is assumed. Overall, the model reproduces qualitatively the $V_{FD}$ changes observed in the data, including the type inversion in 2016–17. Short-term annealing effects are more prominent in the data, however. The thermal history of the detector is included in the model. For example, the SCT was kept at room temperature for two weeks in March 2019, because of maintenance work on the evaporative cooling system. Annealing occurred at a faster rate than predicted during this period, as can be seen in figure 25. Although this behaviour is not fully understood, the larger than expected decrease of $V_{FD}$ helps to maintain safe detector operating conditions. The expected increase of $V_{FD}$ is at most 150 V, so this result allows continued SCT operation during LHC Run 3.

5 Conclusion

Operating conditions for the SCT detector during LHC Run 2 were more challenging than they had been in Run 1, because of the increased instantaneous luminosity and rates of pile-up, resulting in higher data transmission rates and more radiation damage. Nevertheless, the SCT was available for 99.9% of the integrated luminosity and achieved a data-quality efficiency of 99.85%. This was due to various improvements made to the SCT DAQ system: introduction of additional RODs and BOCs, more aggressive data compression, and automatic recovery mechanisms for SCT modules and RODs. The fraction of noisy strips often increased up to 0.3%, but these increases were temporary as recovery could be achieved by regular detector calibrations. The SCT calibration time was significantly reduced to fit into the shorter intervals available between LHC fills in the later years of Run 2.

Key parameters that determine SCT performance are high hit efficiency and low noise occupancy. During Run 2 the hit efficiency of the SCT remained around 99% or higher. Noise occupancy was kept below $5 \times 10^{-4}$, although noise levels increased by 10%–20% during Run 2. Noise levels and their evolution were observed to differ between the different types of sensors.

It is important to understand the effects of radiation damage to the SCT sensors. This was studied in terms of the leakage current and the full depletion voltage, which were monitored throughout Run 2. A method was developed to correct for the offset between the sensor temperatures and the temperatures measured using thermistors on the hybrid boards. This enabled leakage currents
in the endcap modules to be determined more accurately. Leakage currents in the SCT modules increased by more than a factor of ten during the four years of Run-2 operation. Measured values agree with predictions made by the Hamburg model to within ~30%. The full depletion voltage was measured from the HV dependence of the leakage current. Overall trends, such as type inversion, are consistent with predictions from the Hamburg model, while short-term annealing effects were more prominent in the data.

In summary, a number of new operational developments and improvements were made, which allowed stable operation of the SCT with high efficiency during Run 2. Performance parameters such as hit efficiency, cluster width and noise levels, as well as radiation damage, have been measured and understood. In LHC Run 3 an integrated luminosity of about 200 fb\(^{-1}\) is expected with a level of pile-up similar to that in Run 2. The experience and improved understanding of the detector acquired during Run 2 will be important to allow the safe and stable operation of the SCT until the end of Run 3.

Acknowledgments

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhi, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; SSTC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; ANID, Chile; CAS, MOST and NSFC, China; Minciencias, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF and MPG, Germany; GSRI, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MEiN, Poland; FCT, Portugal; MNE/IFA, Romania; JINR; MES of Russia and NRC KI, Russian Federation; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DSI/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; COST, ERC, ERDF, Horizon 2020 and Marie Skłodowska-Curie Actions, European Union; Investissements d’Avenir Labex, Investissements d’Avenir Idex and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and GIF, Israel; Norwegian Financial Mechanism 2014-2021, Norway; NCN and NAWA, Poland; La Caixa Banking Foundation, CERCA Programme Generalitat de Catalunya and PROMETEO and GenT Programmes Generalitat Valenciana, Spain; Göran Gustafssons Stiftelse, Sweden; The Royal Society and Leverhulme Trust, United Kingdom.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (U.K.) and BNL (U.S.A.), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in ref. [47].
References


[17] V. Cindro et al., The ATLAS beam conditions monitor, 2008 JINST 3 P02004.

[18] A. Abdesselam et al., The data acquisition and calibration system for the ATLAS semiconductor tracker, 2008 JINST 3 P01003.


