UCG Report on the Phase-II Upgrade of the ATLAS ITk Pixel Tracker[1]

R. Calabrese, F. Forti, J. Goldstein, A. Honma, K. Krüger, M. Moll, M. Morandin, L. Musa, S. Nahn, P. Petagna, R. Roser, F. Simon, A.J.S. Smith, J. Spalding, W. Wisniewski

Process

At its February 2018 meeting the LHCC recommended approval of the ITk Pixel TDR, with the recommendations for several changes to reflect the interactions between ATLAS and the LHCC. The UCG review began simultaneously, with a kickoff meeting on February 28 at which ATLAS led us through the Cost Appendix for the Pixels, which we then reviewed over the next two weeks, and sent a large number of questions to ATLAS, which were discussed in a 2-hour interim Vidyo meeting on March 20, 2018. The UCG review itself took place April 10 at CERN, consisting of four hours of plenary presentations from the ATLAS Pixel group, followed by three hours of breakout sessions and an executive session. During the meeting panel members compiled a set of observations, comments and recommendations, which form the basis for this report. The revisions of the TDR were considered and accepted by the LHCC panel, finalizing the Scientific/Technical approval process.

We thank ATLAS for addressing our concerns and questions, and for their careful preparation for the review. The presentations were responsive and helped us greatly in our evaluation. The cost, schedule, resources and risks largely appear reasonable. Also, the "core" UCG members reviewed the confidential preliminary "money matrix" and observed that the project was well covered by the participating institutes and funding agencies. We are therefore pleased to report that the Pixel Upgrade is ready for approval.

Project Overview – the TDR

The primary motivation of the ITk pixel upgrade is to preserve the physics performance of the current Run II detector in an environment where much higher data rates, interactions per crossing, and radiation dose are expected. The primary metrics for design include tracking robustness against detector failures, minimizing the total cost and minimizing the CPU required for track reconstruction. The experiment presented a design to meet these requirements, referred to as an "inclined duals" layout. It uses a combination of "planar" and "3D" silicon sensor technologies. The 3D is used in layer 0 where one needs maximal radiation hardness. With this design, there is a possibility of 13 hits on a track in the barrel and 9 hits in the forward region. This design significantly reduces the amount of silicon in the detector, allows for high tracking efficiency even if the hit efficiency is less than ideal, and appears to perform well with today's software and 200 pile-up events.

We believe that the design described in the TDR meets the performance goals with appropriate robustness against operational failure of individual components, and feels that this is a solid baseline approach. However, at a high level, several design decisions remain, which will be described in later sections of this report. While most of these decisions are "cost-neutral", they can have a significant impact on the schedule and thus should be made as soon as feasible, and

[1] CERN-LHCC-2017-021 ; ATLAS-TDR-030

reflected in the schedule. To this end the collaboration has launched a “Layout Task Force” to further optimize the layout with conclusions expected this summer. ATLAS should finalize the layout as soon as possible and modify the schedule accordingly. These modifications, as well as the other decisions and changes that will be made over the course of the project, should be properly documented in "TDR addenda" or similar documents so that a proper record remains for what is actually built.

Management

The ITk Pixels upgrade project functions under an appropriate management structure, in which it and the ITk Strips project comprise the Phase II Inner Tracker upgrade. As in all long-term projects, there is likely to be turnover in senior management as the project transitions from the “TDR” stage to the “production” stage, and that success of this demanding and complex project will depend upon strong proactive leadership at all levels, and proactive succession planning. We have concerns at this point that a more vigorous and dynamic approach is needed to bring the project into production, when there is already time pressure to make several critical decisions to finalise the design. Looking further ahead, ATLAS needs to put in place forceful and energetic management at all levels to make the complex assembly model succeed, and to establish definitive processes for QC and QA, with well-staffed trained teams to implement them.

Cost Situation

The total cost of the ITk Pixel upgrade is 46.9M CHF, ~1.2M higher than at the time of the TDR after inclusion of 777K CHF for phase I FELIX boards needed for testing before the Phase II boards become available, and 400K CHF for additional infrastructure for CO2 storage on the surface. (The total cost of the ITk, including the strip detector and common items, is 122MCHF). Approximately 67% of the cost items still have quality factors 3 or 4, but the risk of a significant increase seems low because most of the estimates are based on similar previous systems. The funding profile fits well with the overall Phase II program.

Schedule

The schedule presented is already extremely tight, and is likely to expand as the final design is developed. To avoid slippages it is likely that the number of overlapping tasks will increase (the latest schedule has already changed to include overlap of module production and module loading). If impasses are to be avoided it is therefore essential for ATLAS to produce a resource-loaded schedule, including a detailed analysis of the critical path, and to strengthen project management in these areas. As an example described later in the report we recommend that the schedule include provision for an extra submission of the final front-end ASIC. We also see opportunities to contain or speed up the schedule without compromising quality, for example by reducing the module-loading time for the outer barrel.

Resources

The money matrix shows that within acceptable uncertainties all level 2 WBS items have sufficient EOI’s to cover the core costs, and the supply of manpower (FTE’s) is sufficient. Multiple engineers on subprojects gives stability. However, uncertainties in the schedule give concern because it is likely that more tasks will need to be performed in parallel than now planned, increasing the peak labor required. The highly distributed nature of the assembly makes it critical to have plenty of pro-active managers and supervisors.

Risks

Risk analysis is performed by estimating the probabilities, and the impacts on cost and schedule, classifying the risk levels as low, medium and high. Fortunately the cost risk is low, and the proposed mitigation plans for cost and technical risk look reasonable. However, as described above we have concerns that the schedule risks are serious, and not easy to mitigate. ATLAS should revisit the schedule as soon as the plan for the final front-end ASIC has been decided, and come up with a conservative plan.

Sensors

The Baseline sensor types for the hybrid pixel modules are 3D sensors (layer 0), 100um thick planar n in p sensors (layer 1) and 150um thick planar n-in-p sensors (layers 2, 3 and 4). Other options considered are passive CMOS sensors replacing planar sensors and an alternative module concept with fully monolithic CMOS pixel detectors. The latter are anticipated to form "drop-in modules," replacing hybrid pixel modules in layer 4 without any further change to the overall baseline system.

Decisions still open for the baseline sensors are: pixel cell size (50x50um2 or 25x100um2) and the optimization of the biasing structure. A plan was presented on how to make these decisions which are partly delayed by the non-availability of the readout ASIC. A yield figure for the production of 25x100um2 3D sensors should be determined as soon as possible to understand the feasibility and cost impact of the pixel size option for this sensor type. Highly specialized industry is supplying 3D sensors and bump bonding. While sufficient supply capacity has been presented, ATLAS is encouraged to coordinate the procurement with CMS in contracting with the same suppliers.

The CMOS drop-in solution for layer 4 is predicted to result in a cost reduction of 2.5 MCHF, and would reduce the load on hybridization vendors. The R&D program has obtained positive results on prototypes, but to be relevant for ATLAS the technology has to be developed into a fully functional front-end and drop-in module design, in time not to cause delays in the Pixel project.

ATLAS presented a schedule in which a review is planned for June 2018, to be followed by development of the monolithic CMOS chip building on RD53 know-how and technology in parallel to the baseline ATLAS FE. A decision on the CMOS option is indicated for Q3 2019. We judge the CMOS timeline to be extremely tight and have concerns whether full FE functionality could be ready in time. A further worry is on draining resources from the baseline ASIC development.

Recommendations :

1. The number of sensor options should be reduced as soon as possible to reduce design, testing and qualification efforts and flavours of modules.

2. The decision on the CMOS drop-in solution should be taken in June 2018 and not in Q3 2019.

Front End ASIC

The ATLAS Phase 2 pixel chip is in the 11th year of a 12 year development plan, jointly with the RD53 collaboration for the last 5 years. The current stage of development is reflected in the RD53A chip, which has essentially all the functional building blocks needed for the final chip, thus retiring most of the technical risk in the Front-end ASIC area. However, there remain custom features to be added, pertaining to special edge-pixel handling, acute radiation tolerance, trigger support, and serial power regulation, and some remediation arising from the RD53A testing campaign, in order to reach the production design.

The RD53A testing campaign has been underway since the beginning of 2018, and so far the chip is functioning as designed, with some minor design issues detected. Testing by the RD53 collaboration continues, and RD53A chips will be assembled into modules in the first half of 2018, to gain experience with multi-chip system functionality.

Two potential paths were presented for the development from RD53A to the final ATLAS pixel chip. Option 1 has two chip submissions, one immediately, based on the RD53B design, which just includes fixes for issues detected in the RD53A chip, followed by a second submission in roughly a year. Only the second submission would include the final design (e.g. the Low Voltage Shunt and the choice of the Front End and Read Out options). In Option 2, there would be only one submission, postponed for an additional 6 months into 2019 to allow continued design to produce a chip with all final chip functionality, with the expectation this will be the final chip.

There is general consensus that Option 2 provides the better path forward, as it gets a "final" chip into the hands of the proponents earlier than Option 1. There still remains considerable risk, however, that a second chip submission in 2020 will be needed, but in this case it will happen earlier, and after experience with all the final functionality. The committee believes it would be prudent to adjust the schedule to include this second submission in the baseline, with an "Opportunity" in the Risk Register that it may not be needed, rather than omit it from the baseline with a high probability "Threat" that it will be needed.

The possibility that ATLAS and CMS could use the same chip is very attractive and should be pursued with priority. However, this approach requires a thorough evaluation, for which there is a dedicated joint meeting with the LHCC on May 18.

Recommendations:

3. Proceed with Option 2, which guarantees optimization of manpower and early arrival (wrt option 1) of the 'final' chip. Include an additional submission in the baseline schedule, with an Opportunity in the Risk Register that it may not be needed.

4. ATLAS management should ensure that sufficient resources (manpower) are available for this item which is critical for all the project.

Other ASICS

The ATLAS pixel detector calls for several other ASICs: an aggregator and equalizer ASIC for the active data transfer cable, the so-called PSPP chip to handle power distribution within a module, and a DCS ASIC to control the PSPP chips. The PSPP chip is in 130 nm technology, while the other chips are 65 nm technology. We are pleased to note that the plans for these chips, including cost and schedule, seem well-founded.

Hybridization and Module Assembly

Allowing for some loss of modules in assembly and integration, the project is preparing to produce approximately 14,000 modules (for 10276 to be installed in the detector). The majority of the modules (for the outer barrel layers, and the rings) are "quad," meaning with four readout ASICs, with smaller numbers of single and dual modules for the inner barrel layers. The hope is that all quad modules can be identical, differing only in the cable pigtail that will depend on the module location. This will be determined through the work of the Layout Task Force that is expected to conclude June 2018.

Module assembly has two major stages: (1) the hybridization process that consists of the preparation of the bumps on the sensor and ASICs and flip-chip assembly of the two to form "bare modules", and (2) the assembly of the full working module with flex circuits for powering, control and readout. The plan is to do the hybridization stage in industry, sharing the work over approximately three small flip-chip vendors, with the module assembly carried out by a consortium of 10 institution clusters in the collaboration, each cluster consisting of one assembly site with additional local site(s) for the module testing. Following the qualification of vendors and assembly sites, the schedule includes 1 year for module preproduction for personnel training and development of quality and throughput, and 2 years for full production. Labor estimates were described.

The plan to qualify several flip-chip vendors is appropriate since the capabilities and interest of vendors is not clear. This will be assessed with a market survey by the end of 2018. The plan to pace this stage to the assembly throughput with small batches is appropriate to ensure high quality and limited wastage of parts. Two types of bumps, indium and solder, have been shown to be entirely equivalent in the final product, so this can be left to the vendors chosen. The plan includes 10 assembly clusters, geographically distributed. At present 9 of the 10 clusters are identified. These 9 include institutions with relevant experience from previous silicon detectors. ATLAS indicated that a similar distributed assembly process was used for previous projects and that this plan represents about a factor 2 increase in the number of assembly sites compared to the past.

The throughput assumption of 20 modules/week/cluster is set by the tooling foreseen, with the costs for fixturing, coldbox test systems, etc. for 30 modules/week included in the cost estimate. The 9 assembly sites and estimated labor appear sufficient to meet the baseline schedule. With modest increases in equipment cost and personnel (only on simple tasks not needing extensive training), a factor 2 increase in throughput to 40 modules/week/cluster seems feasible.

The project plans to build a limited number of single modules this summer, and many quad modules by the fall using the RD53A chip. The single modules will be used to radiation qualify sensors. The quad modules will be used to qualify bump and flip-chip processes as well as module assembly itself. The intent is to dice the ASICs to the physical size of the final chip, although the active RD53A circuit is smaller. This will allow final tooling to be tested and initial qualification of assembly techniques and sites in time for the module FDR in May 2019. A workshop planned for May 2018 will develop the detailed planning. The plan to build hundreds of modules with the RD53A chip is very important, particularly with the dicing for the full final chip size, to qualify the flip-chip process with thinned chips and to fully demonstrate module assembly. The project should ensure that sufficient parts are available to conclude as needed for the module FDR. In parallel, irradiation tests for all other materials (glues, potting compounds…) are ongoing. The DAQ system for module testing exists in an initial form using commercial components. This will be further developed to form production test systems.

Recommendations

5. The project should plan for tooling to support a throughput of 40 modules/week/cluster and to consider this rate in the hybridization vendor qualification. This can likely reduce production time from 2 years to ~1.5 years for very low cost, mitigating delays or vendor/quality problems during assembly.

6. We have concerns about the HV protection process where two insulators, BCB and Parylene are considered. Sufficient modules with the RD53A chip should be built with each approach to allow sufficient statistics to demonstrate that these materials can be used with high confidence.

7. Strong management is needed for such a distributed assembly process, with significant attention on the QA/QC. We recommend that there be a single person responsible for each of the key functions, namely: module assembly, hybridization, and QA/QC.

Local Supports, Global & Common Mechanics

The local, global and common mechanics rely on well-understood technology with which the collaboration has considerable expertise. The main risks are therefore understood to lie in the schedule, particularly with respect to "module loading," i.e. the assembly of individual detector modules into larger mechanical assemblies such as staves and rings. The risk register mitigates this by adding module loading lines, but the current estimates of effort available are insufficient.

A consortium in Prague (Unicorn Software, with sound experience in databases for the banking sector, plus three institutes) is developing a production database, and has taken full responsibility for the software development (possibly including a "soft EVMS" for cost invoicing to the institutes). This will include support and maintenance for the whole ~20 year duration through construction and operation. Experts from ITk are now in the process of defining the required technical content. The database will bear all the information related to each module/local support throughout the full production process, including QC and QA. A production logistics team will be put in place, with representatives of each production site and key experts covering all relevant technological aspects. It will be led by the Resource Coordinator (already appointed). The team will have direct on-line access to the production database allowing them to obtain in real time any necessary information and intervene promptly whenever needed.

Reworkability is a requirement, although it has not yet been demonstrated for the baseline local support design. A fall-back technique based on screwed connections is envisaged but has not been fully evaluated. The results shown from the "Baby Demo" exercise provide a new limit for the assumed minimum saturation temperature in the local supports. This creates a good safety margin releasing some pressure from the need of guaranteed TFM of the local supports during production, but this is not yet reflected in the specifications. No market survey is required for most mechanical components over the normal threshold, since the choice of resin and fibre system is made on technical grounds but is specific to a supplier.

A production process scattered over a large number of loading sites can be inefficient and inflexible with regard to management, and human and other resources. It is difficult to imagine that adding more loading sites really mitigates the risk, especially as available effort is already so critical. In discussion, however, it became clear that the current module loading schedule is very conservative and should be quite compressible, both through optimisation and a reduction of margins. There is also a significant safety factor in the fact that parallel lines in the same site can be activated if/when needed. For example there will be a line at CERN only devoted to reworking, which can be opened as a parallel module loading line whenever needed (using the same experienced team, but allowing them to work in parallel). This will limit the effects of starting at a green field site.

The database work by the commercial supplier has so far been entirely provided by goodwill. For serious work to progress a new hire needs to be made but funding from the project can only happen after formal project approval and the preparation of MOUs.

The ATLAS Institute Board will be informed of the need for more manpower dedicated to module loading at its next IB meeting. An official request will be made at the following one and will be codified in the MOUs. There is little point therefore in updating the effort survey. It is claimed that flexibility exists for splitting institutes' commitments on the use of funds from those of their human resources.

The mandate and constitution of the production logistics team looks very sensible, but its existence came as a surprise to the review team. This is part of the transition from a "TDR organisation" to a "production organisation", but the timescale for this transition is not clear.

The FDR (Q2 2019) must prove that reworkability during module loading is indeed achieved following the baseline of radiation-hard reworkable glue stated in the TDR. If not, the alternative technique based on screwed connections must be evaluated in all consequences (material, assembly time, cost). However, if FDR fails to prove the feasibility of the requirement, schedule and/or cost to completion and/or or physics performance will be affected, and a risk should be introduced to the register accordingly.

No significant technical, schedule or cost risks were observed in the global or common mechanics, provided that the technical coordination commitments for underground work are maintained and ITk only have to provide coordination.

Recommendations:

8. Management should explain the details of the preparation of the production database, and how this translates into people and responsibilities. The company must guarantee long-term maintenance of the software.

9. Management should clarify the timescale of appointment, composition and remit for the production logistics team, and a request should be made as soon as possible to find additional effort for module loading.

10. An amendment to the present local support specification should be issued, lowering the presently assumed value for the minimum saturation temperature.

11. A clear timeline for the PDR, FDR and PRR for the CO2 cooling system should be defined and agreed upon with EP-DT, in connection with CMS (as the system will be developed in common for the two experiments).

Pixel Services and off-detector electronics

ATLAS does not foresee having spares for possible losses during installation and operation for off-detector electronics, because the replacement of faulty modules will be covered by a warranty contract. However, we strongly feel that the possibility of a warranty that covers the project lifetime which might be longer than the obsolescence time for some of the components is not realistic. Moreover, the absence of an adequate number of spares would stop the operation of the full detector whenever one item would fail and for a period of time that is the company turnaround time for the repair or replacement of the faulty part. We therefore urge that the quantity of parts to be manufactured, both for services and for off-detector electronics, should be revised in order to include more realistic yield figures and spare parts to cover possible losses during installation and operation.

The overall strategy, cost, manpower and planning seem reasonable. Cost estimates for most of the items still have quality factor >3, which is rather natural at this stage, and the uncertainties are not at a level to compromise the overall pixel detector cost estimate.

Recommendations:

12. The quantity of parts to be manufactured, both for services and for off-detector electronics, should be revised in order to include more realistic yield (or working efficiency) figures and spare parts to cover possible losses during installation and operation.

13. The cost estimates for most of the items should be further refined in the course of 2018 in order to better consolidate the "budget estimates" to be included in the construction MoU.

Conclusions

This is a complex project involving new technology that still has important R&D and much design development ahead before launching production. It is therefore urgent to select among the various open options for technology and front-end ASIC development. The assembly model involves transfers of delicate components around the world, requiring tight management controls and QC/QA. The schedule is a real challenge, but there are opportunities to contain it if ATLAS gets on top of them early enough. The costs appear to be reasonably understood, so the main risk lies in the schedule.

We recommend Step 2 approval by the RB and RRB to allow resources to become available and MOU's to be signed. Vigorous oversight of the project, including external reviews, is essential.