INTERRUPT
INSIDE

Distributed monitoring and control using DDS

Ulstein simplifies both system architecture, development and testability with DDS.

Drones & wireless video

In a series of articles, Interrupt Inside will present the different systems of a drone, including the ground support equipment.

SoC FPGA evaluation guidelines

What are the differences between the designs from the three largest vendors on the market?

data : respons
EMBEDDED SOLUTIONS

WHAT'S INSIDE

ARTICLE SERIES: **The impact of consumer electronics on development of high-reliability products**
Article 3: END-OF-LIFE : In market segments with long development cycles companies may receive the first EOL notices before the product even enters series production.
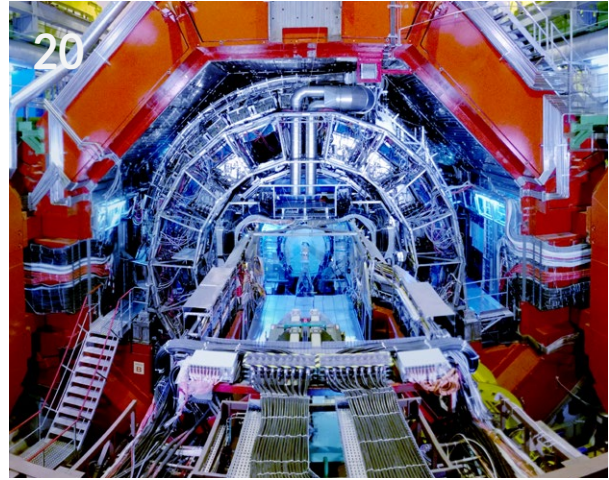Next article: Counterfeit electronic components

WELCOME TO
INTERRUPT INSIDE

Data Respons' vision, *a smarter solution starts from inside*, is our company's DNA described in one sentence. We really think that we can make the world smarter and we really think that this starts from the inside - whether it be inside the heads of our specialist engineers or new technology embedded into the world's products and solutions.

Our customers are typically excellent technology companies, often champions of their fields and driven by in-depth knowledge, product innovations and technology changes. Our daily work is in close collaboration with these R&D environments, where we often take the specialist role in our core areas. In this magazine, we want to highlight relevant technology areas and in-depth research.

This issue of Interrupt Inside has a wide range of articles with an in-depth read for everyone. On the software side you can read about Ulsteins journey in finding the right software solution for their distributed monitoring and control system, and on the hardware side we will talk about the challenges of End-Of-Life and Last-Time-Buy for long life products.

All of the articles are written by our own specialists and select guest writers. We welcome any feedback and suggestions from our readers.

Enjoy the reading!

KENNETH RAGNVALDSEN

ULS

DIS
MC
&

Ulstein has developed control systems for the maritime sector for decades, and are continuously seeking to improve their solutions and products to solve the demanding challenges their customers face. Recently, their search for an improved control system platform led them towards the Data Distribution Service (DDS). Find how their use of DDS simplifies both system architecture, development and testability.

BY: Rune Volden
R&D Manager
Ulstein Power & Control AS

CO WRITER: Preben Myrvoll
Principal Development Engineer
Data Respons

Technical DDS

The OMG Data-Distribution Service for Real-Time Systems (DDS) is the first open international middleware standard directly addressing publish-subscribe communications for distributed real-time and embedded systems. It offers abstracted communication schemes, where the different systems and applications can cooperate without a typical client/server architecture. Currently more than 10 companies or groups provide DDS middleware / products.

DDS goal is to provide the **right data,** at the **right place** and at the **right time**, providing a **global data space** for systems, ranging from machine domain (Edge) to Cloud. This is done via middleware, providing a portable application API and an underlying reliable and real time interoperability protocol (RTPS). It uses quality (Quality of Service - QoS) schemas to ensure that data transfer between participants is done according to mutually agreed standards.

Control system data should often be limited or filtered to be the **right data**, based on rate, content, etc. Being a data centric solution DDS understands the schema of the shared data, allowing for such advanced filtering (For instance: Filtering on a publisher to send ONLY temperature data when it's above 300 C is possible)

DDS dynamically discovers publishers and subscribers, the data they want to share and how they want to do so. Its self-forming nature then ensures that data is delivered to the **right place** even if consumers arrive late. It also detects loss of data or data producers (It implements QoS – enforced logical channel between each publisher – subscriber pair).

The balance of scarce system resources is needed to deliver the data at the **right time**. DDS middleware utilises QoS policies, for instance set by applications at runtime, to balance efficiency and determinism (For example, if a subscriber requires an update every 10ms and its matched publisher does not deliver, the system declares an error, enabling remedial action). QoS covers many characteristics such as urgency, importance, reliability, persistence and liveliness.

DDS provides a self-forming, scalable and distributed middleware, which gives the applications a **global shared data space**, and when you add characteristics such as deterministic performance, low latency / high throughput and high fault tolerance, it seems ideal for mission critical IoT and distributed control systems.

Also due to the dynamic and loosely coupled nature of these systems, DDS significantly reduces maintenance cost, since individual systems may be modified, added or upgraded without impact on the existing system.

SOME UNDERLYING TECHNICAL CONCEPTS

Relational data modelling: DDS addresses data in a manner similar to relational databases. It can manage data by both structuring related topics (by key-fields) and allows for ad-hoc queries and filters on content and time, so applications can extract specific data as needed.

Pub-sub messaging: DDS uses the publish/subscribe paradigm for dynamic discovery and primary management of data-flows between relevant DDS entities, including publishers, subscribers, durability services, recording and replay-services, and connected databases. Request-reply and other patterns are built on this powerful substrate.

Reliable multicast: The DDS standard wire protocol implements reliable multicast over plain UDP sockets, allowing systems to efficiently benefit from modern networking infrastructures.

Life cycle awareness: Unlike message-centric products, DDS offers explicit application support for information life cycle awareness. For instance, it detects, communicates, and informs applications about first and last appearances of data (topic instance) updates. This facilitates timely responses to new and outdated information.

IN USE

For large control systems with +10000's of I/O (sensors and the like), data exchange needs to be smart, reliable and efficient. DDS has been tested for this purpose in several mission critical systems within industries and domains such as power, medical, aviation and space, and the US Navy have used this standard for more than 10 years.

DDS can easily merge today's trends with yesterday's standards in a perfect manner. Interfaces, tools and libraries can easily convert data to and from DDS, to other fieldbus types, for instance Modbus, OPC (DA, UA), etc.

Using it from your application code is easy, and done via a standardised API. Thinking in a data centric way, one starts off by defining a set of Topics (holding data types, structures, etc) that you want to have on your data bus, then you create a Participant (to listen to data within a domain / your separate data space) which again hold DataReaders and/or DataWriters to write or read that Topic.

EXAMPLE:

I want to send and have someone receive the price of my delicious homemade strawberry ice cream. A «double» can hold the price such an item, but it would make no sense if I just sent that double on the wire, without any contextual information. So a Topic is created (I call it «IceCreamPrice»), using «double» as the data type, and thus enabling me to send, and someone to receive, this price. Then I create Data-Writers and -Readers (with some QoS settings) to send and receive that Topic. A simple setup, and now I'm ready to open my ice cream store.
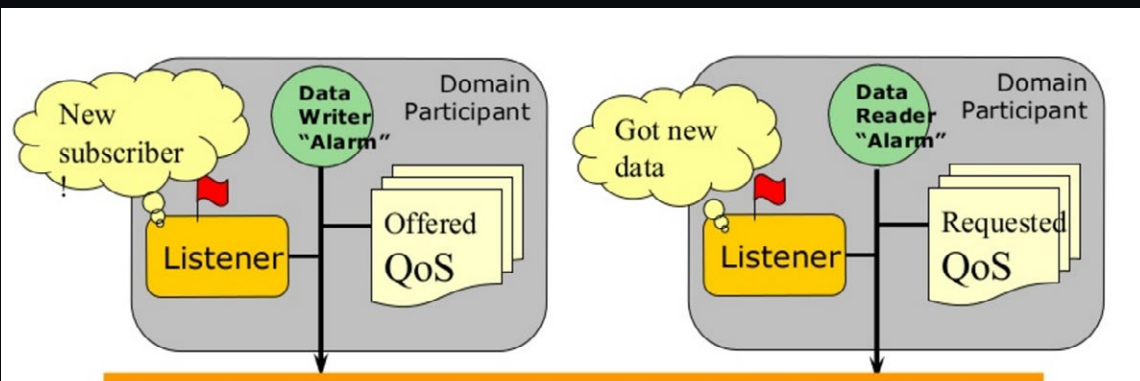
- **Participants** scope the global data space (domain)
- **Topics** define the data-objects (collections of subjects)
- **DataWriters** publish data on Topics
- **DataReaders** subscribe to data on Topics
- **QoS Policies** are used configure the system
- **Listeners** are used to notify the application of events

Ulstein's experience with DDS

BY: Rune Volden, R&D Manager
Ulstein Power & Control AS

Ulstein's experience with DDS started in 2013. May 14 that year, I got an email from a colleague regarding an alternative middleware. June 18th we had requested pricing n Open Splice/RTI Connext. We then started using Prism Tech's Open Splice DDS the first months for graphical user interface (GUI). November 5th 2013 we purchased RTI Connext licenses. My colleague then worked with RTI Connext DDS to implement the communication between GUI and the control system throughout 2014.

During 2014 we developed our IAS based on 3rd party control system middleware (CDP), and only used DDS as communication towards the GUI. Earlier we used a Modbus communication based on JSON, but this approach required much development work, not to mention testing, to get a good result.

In 2014 our IAS project met great challenges regarding system scaling, and handling the numbers of signals required by our customers. After repeated attempts with our former middleware, this approach was eventually cancelled. We made a thorough technical investigation from November 2014 to January 2015, as to how to build our future control system.

THE DDS GLUE
The conclusion was to develop our new control system in-house, in cooperation with Data Respons AS, using DDS as the fundamental block in the communication layer. This work started in February 2015. The IAS project was divided into teams working with documentation, Graphical User Interface design, Graphical User Interface implementation, Graphical editor, Control system kernel, Control system application, IO controller application and the Configurator tool.

Our automated systems experts started the documentation work in 2014, and all is done according to the guidelines and structure of the DNVGL's ISDS standard.

GRAPHICAL USER INTERFACE
In cooperation with Eggs Design we used the work from an earlier Ulstein Bridge Vision project as a starting point for the realisation of a graphical user interface, starting in January 2015. In March the implementation of the Graphical User Interface started in cooperation with The Qt Company. They also started developing a Graphical editor for us which makes it possible to get the complete control system including all graphics into one readable configuration.

CONTROL SYSTEM
The control system kernel was made from scratch, with a possibility to create and configure all internal components from XML, as a requirement. This was done in close cooperation with Data Respons, where they greatly contributed to get a strong and well tested kernel. The kernel then offers a communication layer towards the fieldbus layer (IO Controller) and graphical user interface.

The communication mostly utilises DDS (Data Distribution Services). We have
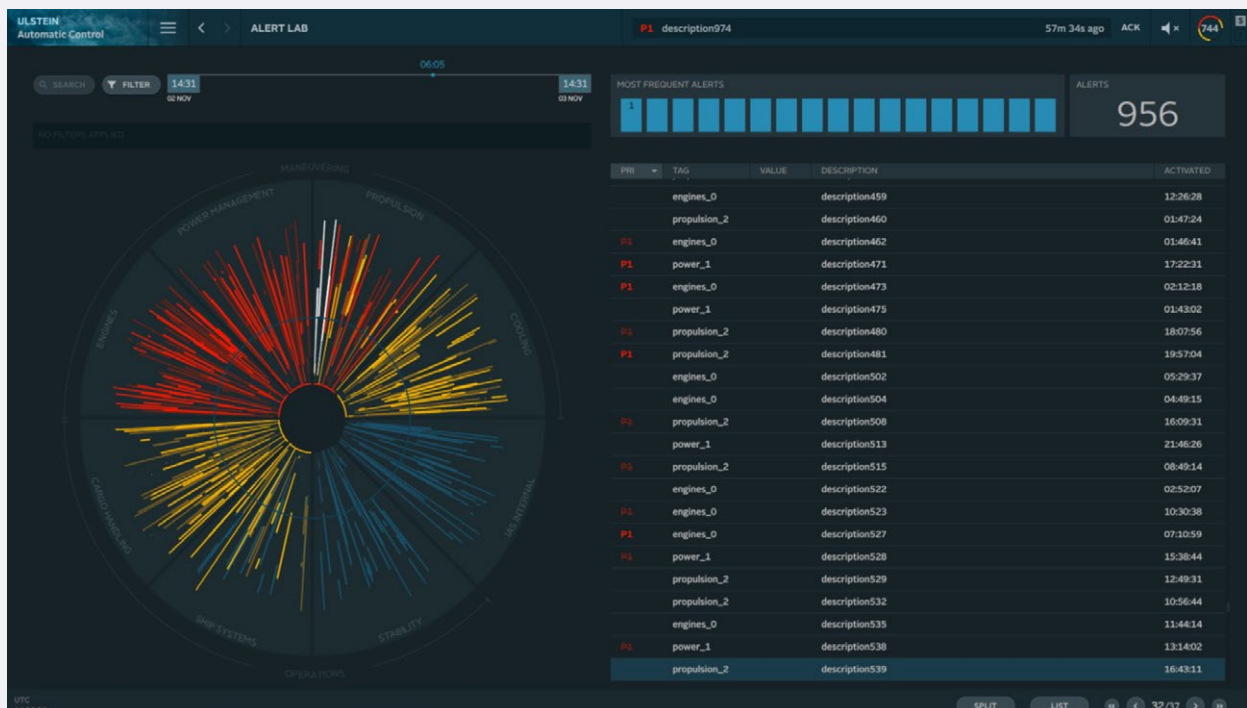
Alert Lab image, (Ulstein Power & control AS)
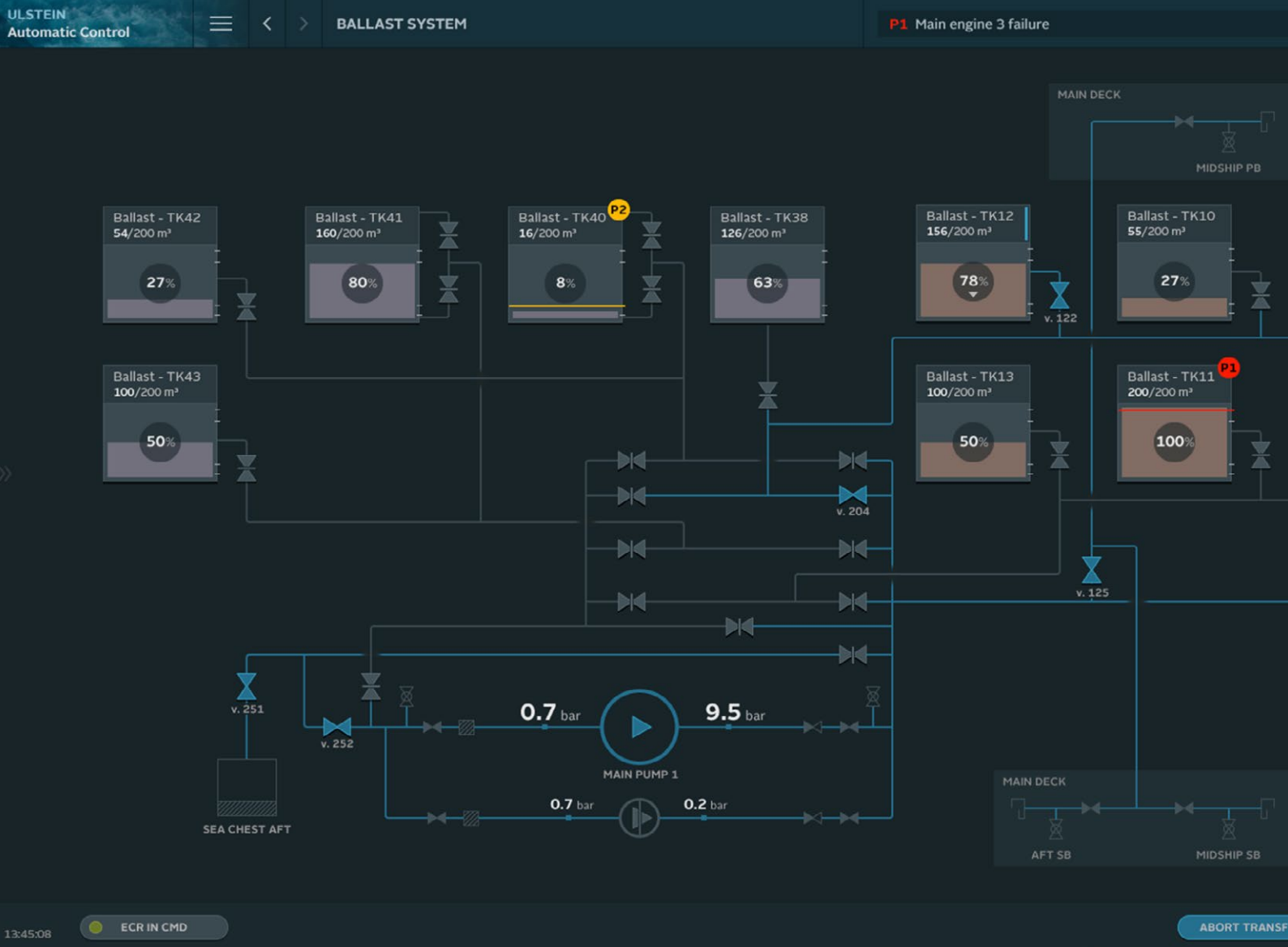
Ballast control system. (Ulstein Power & control AS)

evaluated several versions of DDS, but currently we use RTI Connext 5.2 in our systems. This is applicable for the control system, IO controller (Fieldbus ++) and graphical user interface also. DDS act's as the glue between all the different applications on various controllers, PC and workstations.

/// To get the delivery of control systems to the end customers as efficient as possible, with the highest possible flexibility, we have developed a configurator tool.

The control system application is also made in cooperation with Data Respons. In this case, legacy code is ported to the new system kernel, in addition to adding new code and functionality needed.

**CONFIGURATOR TOOL
/ CONFIGURE TO ORDER**

To get the delivery of control systems to the end customers as efficient as possible, with the highest possible flexibility, we have developed a configurator tool. This enables the application engineer to set a configuration of the control system in an easy, safe and understandable way, according to the customer's requirements. For an automation system on a ship this typically means adding
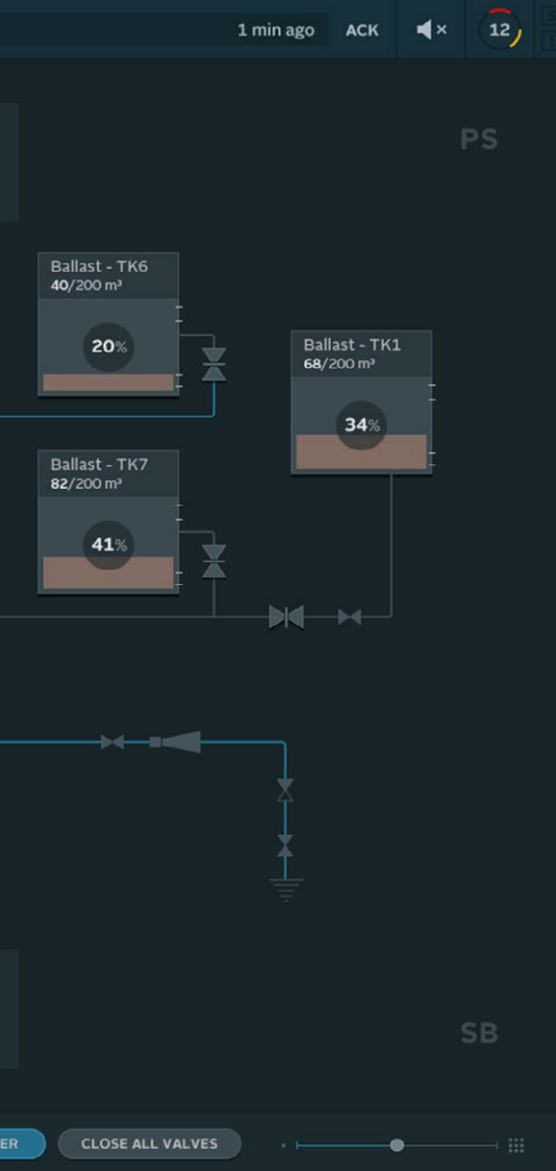
pumps, tanks, valves, pipes, switches, generator sets, propellers, motors etc., where each component has a control/ remote control, control logics, mimic and user interface design. In the existing SCADA platforms this is a very time consuming and comprehensive work. One of our great challenges is that the changes shall be executed fast and ef-

ficient in the final phase of large projects. Typical "last minute changes" can introduce human errors that everybody wants to avoid. The configurator tool we are developing will minimise this risk in an efficient way.

I/O CONTROLLER
Data Respons has also been involved in the development of the I/O controller application. This application process all I/O on that I/O controller, which for instance can have serial lines (RS422), CAN, Modbus RTU, analog and digital IO, and sends/receives data via DDS. Via the Configuration tool we can download the configuration of serial buses and CAN to the I/O controller. The configuration can include all required configuration, down to the node level, on a local fieldbus. With this automation, only limited changes are required via vendor specific tools. The I/O Controller application reads from and writes to the IO via the controller's I/O API. All maritime approved controllers with C++ API are basically of interest.

Our control system is not vendor specific. Currently we have at least three dif-

PS

Ballast - TK6
40/200 m³

20%

Ballast - TK1
68/200 m³

34%

Ballast - TK7
82/200 m³

41%

SB

CLOSE ALL VALVES

HIGHLY ADAPTABLE FOR VARIOUS MARKETS

Ulstein traditionally makes control systems for large ships, but by using DDS and having the flexibility of our new kernel, the ideas and the actual system can easily be adapted to most mission critical and normal control systems.

The underlying data model of control systems often consist of transporting digital and analog data, in addition to some business specific types. The framework to do so and the knowledge of how DDS works, already exists at Ulstein. We think we are in a position to offer control system expertise and software solutions to customers outside of our traditional domain also.

ferent suppliers of I/O and I/O controllers. Initially we start delivering systems with I/O controllers from Bachman, then Phoenix and Wago. Eventually all I/O Controllers with an environment and an API to create the I/O to DDS data transitions can be used. This gives our product flexibility, since often one supplier can't offer a complete solution but as a total they can.

TESTING
Data Respons recommended us to use test driven development, continuous integration, build servers and analysis tools at an early stage. We can now see that this saves us much time in rework and testing during both development and integration phases.

With DDS replacing components with "bots" or "mocks" for systems test is much easier, since they all use DDS to communicate. We have also started using Docker to run and simulate a multi controller environment network, and thus quite easily running large scale system and integration tests.

Due to the number of applications and controllers making up the system, continuous test and integration could have been a lot of work if not automated and thorough unit and system tests were applied continuously.
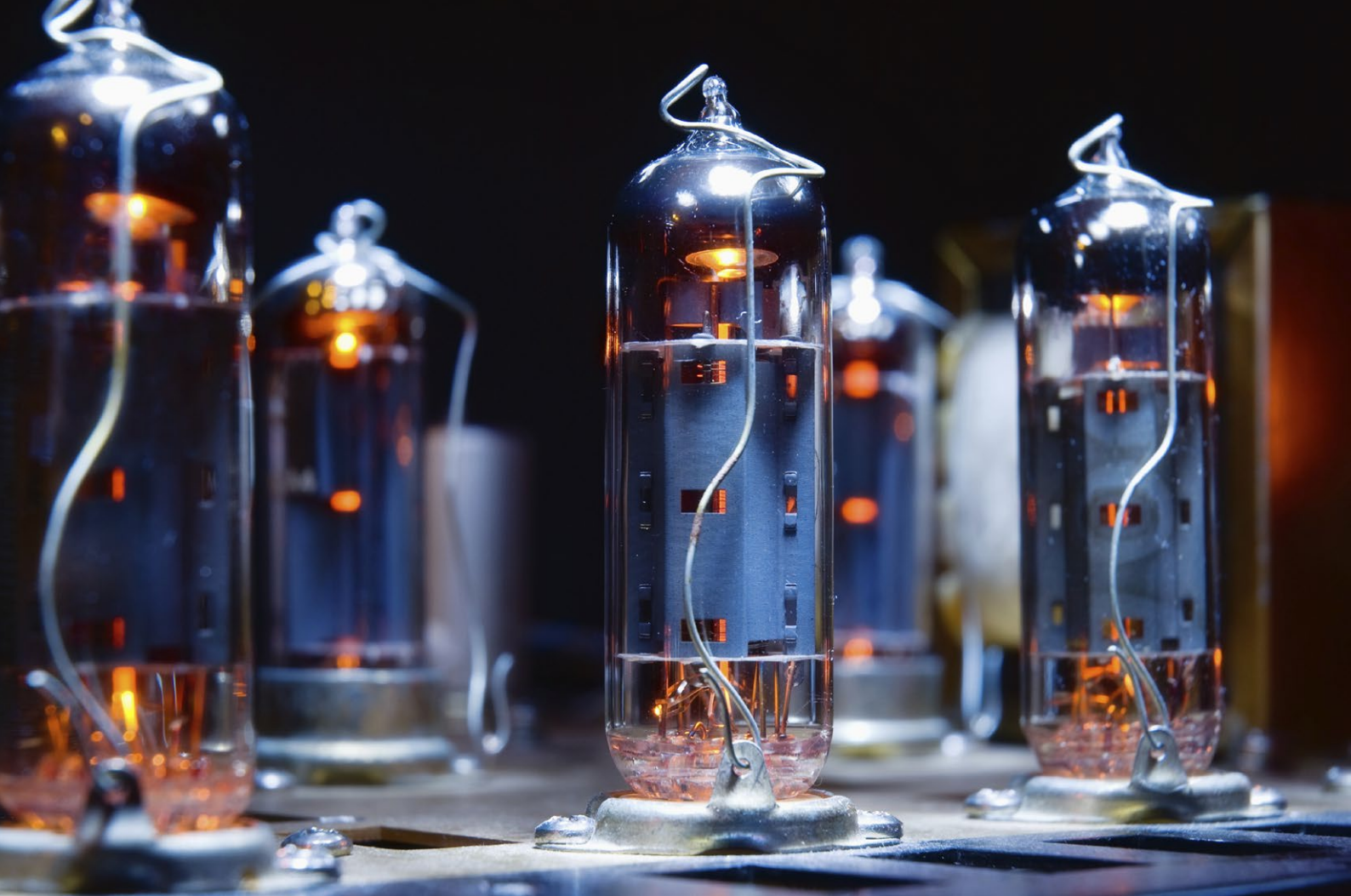
END-OF-LIFE

It happens with increasing frequency: a maker of electronic components issues a notification to customers announcing "End-of-Life (EOL) and Last-Time-Buy (LTB)" for certain chip components. The LTB date is the last date on which the vendor will accept orders for the parts concerned. Before this date, manufacturers using any of the affected parts in their products must determine once and for all how many parts they need. Upsides and market erosion are two in a range of uncertain variables to be considered when the same manufacturers attempt to balance the capital cost of buying too many parts with the opportunity loss of buying too few. In market segments with long development cycles, companies may receive the first EOL notices before the product even enters series production.

BY: Haldor Husby
Principal Development Engineer
Data Respons

MOORE'S LAW, YET AGAIN

At the heart of the problem of component obsolescence is Moore's law and the ever-shrinking dimensions of the transistor, the essential atomic building block of an integrated circuit. Moore's observation that the number of transistors in an integrated circuit doubles every second year has held true for 50 years, and the rate at which the number doubles is only now slowing down. Miniaturisation has brought steady improvements in aspects of performance such as switching speed and power consumption, but make no mistake, the primary driver for - and over-riding motivation behind this rapid technological development has been cost savings. A smaller transistor occupies a smaller area, and in the silicon world, cost is proportional to consumed area. Take Intel's recent transition from a 22nm to a 14nm process, which coincide with their change from the Haswell to the Broadwell processor families. Based on the same microarchitecture as the Haswell processors, the Broadwell nevertheless boast a 35% increase in the number of transistors while their die size has shrunk by 37%. Intel did not transition from a 22nm to a 14 nm process in order to offer us all better performance and added functionality (important processor performance parameters like clock speed levelled off a decade ago). Their sole motive was to convert the area reduction into a cost reduction. The added functionality was simply thrown in as an incentive to computer makers and consumers to switch to Broadwell as quickly as possible.

As a process with smaller feature sizes comes on line, a chip manufacturer starts to port its high-volume products to the new process in order to realise the cost reduction it offers and get a return on the cost of developing the process. Low runners are left behind and as their volumes drop, the manufacturer seeks to limit its obligations to customers and to shut the process down. This is when the manufacturer issues EOL notices.

On the one hand, market segments with lower volumes benefit tremendously from this: the incredible volumes of the consumer market supports a development model that delivers vast computing power at a very low cost. On the other hand, those segments with product lifetimes of a decade or more must learn to operate with a supply chain geared for product life times of only one or two years.

REACTIVE APPROACH

Obsolescence is nothing new, but it is becoming more and more serious as a product maintenance problem. What began as an irritation has gradually become a serious burden to many organisations. The fact that this has happened gradually may explain why many organisations still deal with it in a reactive manner.

Typically nothing happens until the company receives an EOL/LTB notice. That then triggers a frantic attempt to size up the last order, an activity that must normally be concluded within 180 days. Very few companies have forecasts allowing them to see a decade into the future with any precision, and the many uncertain variables that must be taken into consideration make it an almost impossible task to strike the right balance between opportunity and cost.

Buying and stockpiling components create other problems too. Even when the future need is correctly estimated, components in storage may represent a significant, if not unacceptable, amount of tied-up capital. Reserve charges and an increased reluctance among distributors to hold inventory for more than two years will affect the business case for continued manufacture of the product. Besides, electronic components go stale. After a couple of years in storage, the solderability suffers, leading to a higher yield loss. This is bad enough under normal circumstances, but it is much worse when the lost parts cannot be replaced.

Moreover, maverick lots in storage may go undetected for years causing yield and reliability problems long after the expiry of the guarantee and end of support from the supplier. A need for parts a long time after LTB may tempt buyers into the grey market, where all the fake components are. Although it can extend product life for a few years, the reactive approach is seldom viable.

PROACTIVE APPROACHES

To get a handle on obsolescence management it is helpful to categorise parts along these lines:

1. Unique, single sourced parts critical to function (examples are processors and FPGAs)
2. Integrated parts which may have complex functionality but are to some extent standardised and with multiple sources (memory and power components)
3. So called Chiclets or popcorn parts. Standardised parts available from many suppliers (passive components, logic gates etc.)

Each category may then be associated with an obsolescence risk calculated on the basis of the probability of obsolescence and the consequence if it happens. Once obsolescence is viewed in terms of risk, it ceases to be an unpredictable and devastating "force majeure" and becomes instead something that can be managed with well-established techniques of risk tracking and mitigation. Category A components represent the highest obsolescence risk, which means they must be given priority in a good obsolescence management strategy and tracked more closely than the other categories. And while an embedded circuit board may comprise a few hundred individual part numbers, there will typically be no more than a handful of category A components among them. The majority will fall under category C which may safely be tracked only casually or not at all. The task is suddenly far less daunting.

Mitigation, when EOL strikes, will also be different for each category. For category

Haswell (top) and Bradwell (bottom). Die shrink in the transition from a 22nm to a 14nm process (source: Intel)

A components – which are by nature unique – there may be little option other than to perform a Last Time Buy to forecast. But for category B parts the primary mitigation technique is to search for and evaluate replacement parts. When this is performed up-front and resulting in an Approved Vendors List (AVL), the obsolescence of parts have no particular consequence and may be handled as part of the normal procurement and manufacturing activity.

Another element of a proactive obsolescence management is the use of planned technology refresh cycles. Keeping a product alive for decades will often require one or more design revisions to replace obsolete components. Based on the anticipated lifespan of the category A components, the refresh interval may – to great advantage - be decided already
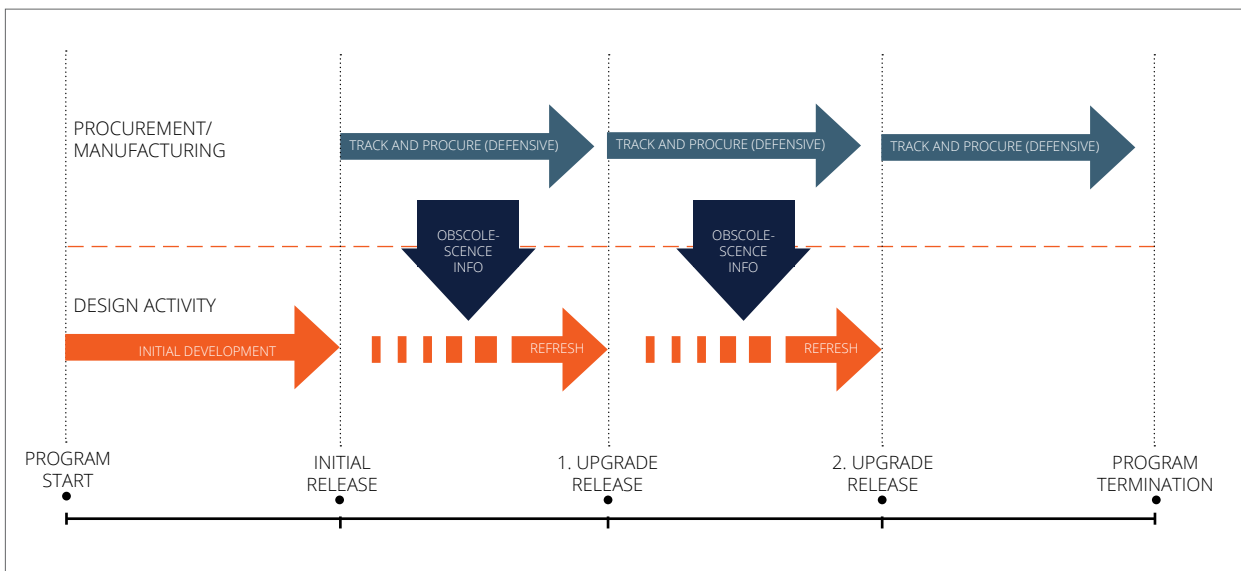
is often tight and the necessary design resource already allocated to other tasks. Planned revisions, on the other hand, contributes quality to the design and future proofing, as the purpose of the revision is to consider the life expectancy of the entire design not just the obsolete parts.

Thirdly, making plans for technology refresh cycles during the initial design phase forces designers explicitly to consider the life expectancy of the parts they choose, and this will in turn influence key design decisions such as part selection and architecture. A design is a series of trade-offs made under pressure of time and cost, and product life expectancy does not always get the attention it is due in the process.

the intended market for a given component. Parts intended for the automotive market will be around for much longer than those meant for tablet computers.

Individual parts may perish, but the overall architecture must stay the course. It might limit the scope of mid-life technology updates, and it helps reduce the extent of the re-qualification that is required after a design change. An enduring architecture favors a modular design with thin and standard interfaces between sub-systems. The general drive towards higher integration obstructs this goal, but designers should be conscious that a very integrated or convoluted design is hard to maintain and may require an extensive re-design effort.

Also crucial in its bearing on choice of architecture is the realisation of the fact

| PROCUREMENT/ MANUFACTURING | TRACK AND PROCURE (DEFENSIVE) | TRACK AND PROCURE (DEFENSIVE) | TRACK AND PROCURE (DEFENSIVE) |

OBSCOLESCENCE INFO

OBSCOLESCENCE INFO

DESIGN ACTIVITY

INITIAL DEVELOPMENT   REFRESH   REFRESH

PROGRAM START   INITIAL RELEASE   1. UPGRADE RELEASE   2. UPGRADE RELEASE   PROGRAM TERMINATION

Planned technology refresh cycles makes the logistic managable

during the initial design phase. Doing this divides the active life span of the product into shorter periods of defensive logistic obsolescence management, each separated by a technology refresh release. The first advantage of this is that, whereas forecasting and stockpiling for 10 years might not ever be viable, doing so for three or four years (until the next technology refresh) is normally quite straightforward: instead og having to bridge an indefinite parts gap, Last Time Buys performed in this context only need to bridge a parts gap until the next planned revision of the product.

Secondly, the use of planned refresh intervals makes the design resource requirement visible in the organisation as each revision is planned and not a suddenly conceived and hastily executed stunt project performed to fulfill an unexpected order. Unplanned revisions will burden any design team as the time-line

OBSOLESCENCE MANAGEMENT IN THE DESIGN PHASE
Effective tracking and mitigation strategies aside, the long-term sustainability of an electronic circuit board is no doubt contingent on choices made during the design stage. This is obviously the case during the part selection process as all parts causing an obsolescence problem were once selected during a design phase. However, it is an illusion to think that making the "right" component choices may solve the problem of obsolescence. A wiser view would be that due consideration of a part's life expectancy contributes to an overall obsolescence management strategy.

Knowledge of the suppliers and their markets, as well as their commitment to their own and industry roadmaps should inform parts selection. After-market support and extension of life programs are important, and it is crucial to consider

that software development costs always exceed hardware development costs even for simple systems. And the gap between the two will only expand with time. Hardware resources are abundant and easy to include but making them useful requires software effort, and coding efficiency is not likely to increase sharply anytime soon. Consequently, a system design intended to last for decades will be one that limits the software effort required after a hardware design update. Preferred characteristics are modular designs and layered SW structures. But the most important thing to realise is that system design for long product life is a multi-disciplinary undertaking.

COMMERCIAL OFFERINGS
The emergence of businesses offering solutions of various kinds bears witness to obsolescence as an escalating problem. A range of companies provide logistical help in the form of planning and

tracking tools as well as database and other services. Another niche is filled by operations like German HTV who offers a proprietary long-term storage and conservation process for components and assemblies; they claim to decelerate the component aging process by a factor 12-15. Paired with on-going test and monitoring, their offer makes it technically viable to stock components for one or two decades.

A very distinguished position in the component obsolescence industry is held by Rochester Electronics. Supplied by a great number of the major semiconductor manufacturers they take over processes and portfolios once retired by the original manufacturer. Their warehouse presently holds more than 10 billion "obsolete" components of their own manufacture, and they even offer part re-creation as part of an overall service called Extension-of-Life®.

CONCLUDING REMARKS
The concepts and techniques discussed in this article are only elements of a proactive approach to life-cycle management. How these elements are used as part of an overall obsolescence management strategy depends on the product in question. Someone who earns a living by making and selling off-the shelf single-board computers will take a different approach from someone who uses embedded electronics as parts of a much larger and more expensive system. The former must account for the premium necessarily added by life-extending measures on a board-by-board basis while the latter may rather view it as part of an overall system cost picture. Constructing the right strategy may be hard, but technology companies will find it increasingly difficult to survive in modern market conditions without a sustainable product maintenance strategy.

MOORE'S LAW

Probably coined by Carver Mead in the early seventies as recognition of predictions made by Gordon Earle Moore of Intel in a seminal paper from 1965:

The number of transistors in a dense integrated circuit doubles approximately every 2 years.

Moore's original paper predicted a doubling every year, but he modified it in 1975. Recalling that the integrated circuit or "chip" was invented only three years earlier, the relative adherence to Moore's law 50 years later is evidence of its influence as a roadmap. Moore's actual agenda was to say something about cost optimisation in chip-making.
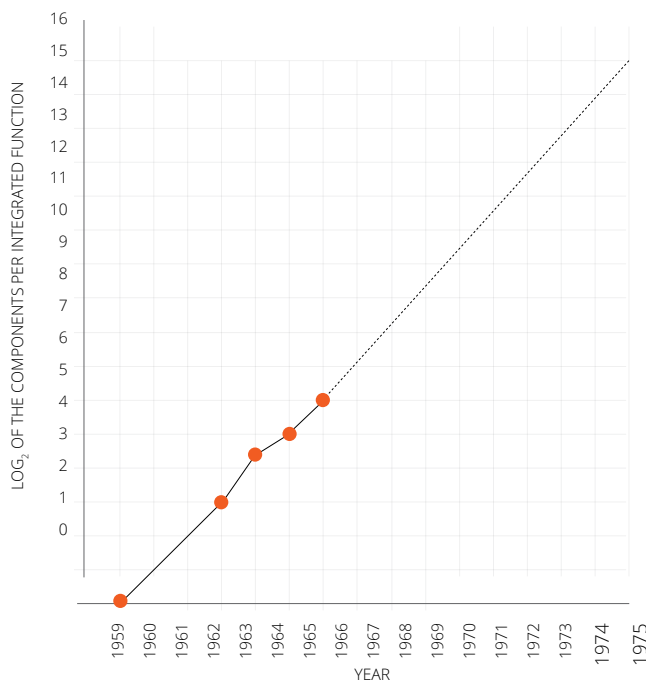
"Number of components per integrated function for minimum cost per component extrapolated over time" The five data points on which Moore based his predicition (source: Intel)

Counterfeit electronic components

The problems of component obsolescence and counterfeit parts are close cousins.

Fake parts find their way into equipment of all kinds including military airplanes and weapons systems.

A substantial majority of these components mimic hard-to-find parts discontinued by their original manufacturer.

In the next article in this series, we take a closer look at the issue of counterfeit electronic components, the industry's response to it and the protective measures available to companies making electronic products.

PROCESSORS
for high temperature applications

I n this article we will go through the most commonly used microcontrollers and Digital Signal Processors (DSP) for high temperature applications from 150°C and upwards. This article will not discuss all processor alternatives, but focus on the most common and relevant options. Also, this article does not cover processors available in die form, often referred to as «known good die» (KGD), as this product group is usually avoided as it complicates product manufacturing considerably.

> *"If the temperature requirement can be reduced, you will benefit from significant production cost savings.*

Image to the left:
Texas Instruments OMAPL137-HT processor in a downhole modem rated for 175°C.

BY: Henning Sjurseth
Senior Development Engineer
Data Respons

OLD ARCHITECTURES

Most high temperature processors are based on old architectures. Often, an old architecture is taken as a starting point and improved in regards to temperature. The documentation is then updated to some extent, but the documentation can vary quite in quality depending on the manufacturer. High temperature processors are not necessarily supported in the manufacturer's tool set. For instance, Texas Instrument (TI) do not recommend their own Code Composer Studio IDE if developing for their SM470 high temperature microcontroller. Instead, TI recommend Embedded Workbench from IAR Systems.

In the last few years some companies have been moving towards current microcontroller architectures, specifically ARM Cortex-M. Vorago Technologies has developed a device with a Cortex-M0 core. Another company, e2v, provides a device with a Cortex-M4 core.

In general, with higher temperatures device cost goes up and level of functionality goes down. In particular, 150°C degrees devices can be very affordable, but when your design requirement passes 150°C degrees you'll see a significant price jump. Don't be surprised if your production cost multiplies by a factor of 5 to 10.

150°C DEGREES

In the 150°C degree temperature range, there's several processor alternatives. As mentioned, e2v is focusing on the ARM Cortex-M4 architecture. Currently e2v have a device rated at 150°C degrees for 1000 hours based on a Freescale microcontroller. The device has both Analog-to-Digital and Digital-to-Analog converters. e2v are hoping to secure development funding to extend the temperature rating for this device to 175°C.

Texas Instruments (TI) lineup includes three devices in this temperature range all with plastic packaging: SM470R1B1M-HT, SM320F28335-HT and MSP430F2619S-HT. The first two are also available in a ceramic package for higher temperatures.

The first device, SM470R1B1M-HT is a microcontroller with an ARM7TDMI core (32-bit). This device has an Analog-to-Digital converter, but no Digital-to-Analog converter. Recommended development tools are from IAR Systems: Embedded Workbench and I-jet debug probe.

The second device, SM320F28335-HT is a 32-bit DSP with Analog-to-Digital converter. Recommended development tools are TI Code Composer Studio v6 (based on Eclipse) and TI XDS100/200/560 debug probe. The third and last device is MSP430F2619S-HT, a 16-bit microcontroller architecture with Analog-to-Digital and Digital-To-Analog converters. Suggested development tools are Code Composer Studio v6 and TI MSP-FET debug probe.

Atmel has several devices based on their 8-bit AVR architecture for applications up to 150°C. These devices also have Analog-to-Digital and Digital-To-Analog converters. Suggested development tools are Atmel Studio IDE (based on Microsoft Visual Studio) and Atmel-ICE or AVR ONE! debug probes.

Microchip provides devices for this temperature based on 8- and 16-bit architectures. Their 8-bit architecture devices for 150°C can be found in the PIC12, PIC16 and PIC18 families and their 16-bit architecture devices for 150°C is part of the PIC32 and dsPIC33 families. Recommended development tools are MPLAB X IDE (based on the NetBeans IDE), XC8 (8-bit) or XC16 (16-bit) compiler and MPLAB ICD 3 or REAL ICE debug probe.

175°C DEGREES

Texas Instruments provides one device for this temperature, the OMAPL137-HT. This is a dual-core device with one ARM core and one DSP core. This is one of very few devices device suitable for embedded Linux. Recommended development tools are TI Code Composer Studio v6 IDE and TI XDS100/200/560 debug probe. This device doesn't have an onboard flash for the application image.

If you decide to boot the DSP from an external flash, it's important to note that TI's own high temperature flash is not compatible. We've used TTSemiconductor TTZ2564 flash for booting the application image with the OMAPL137 with

ELECTROMIGRATION FAIL MODE

HOURS

DIE JUNCTION TEMPERATURE (DEGREES C)
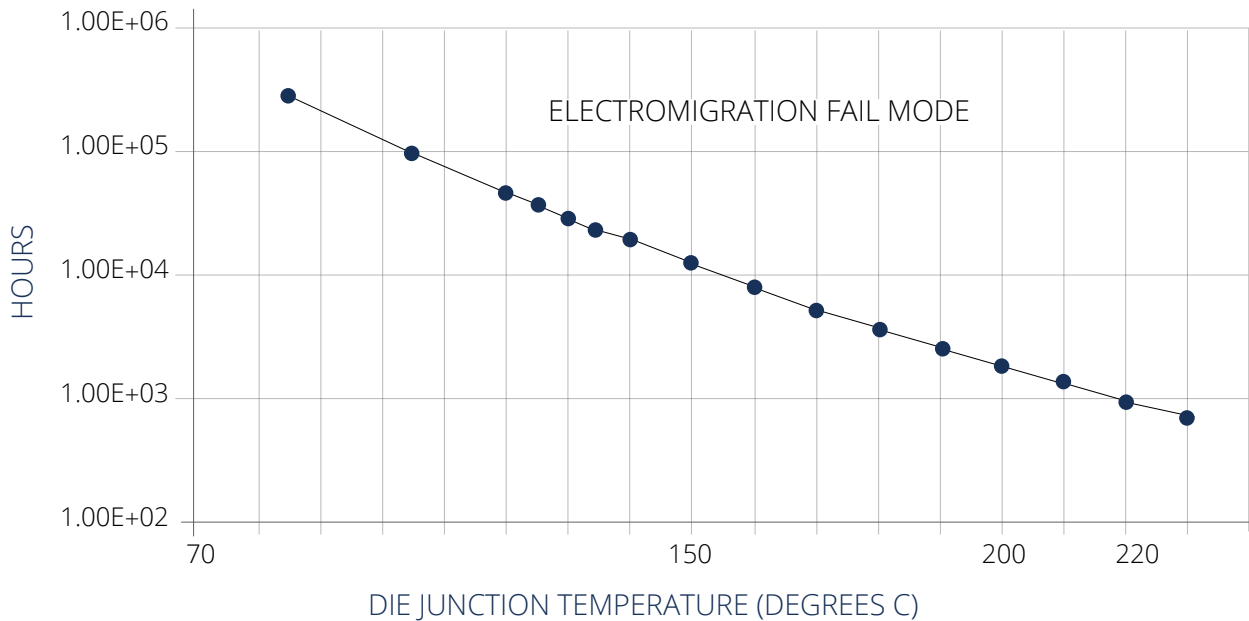
Graphically reproduced from source: Texas Instruments, Technical Datasheet for SM470R1B1M-HT.

no issues. A quick tip, if you have issues booting, check the BOOTCFG register during debug which contain the state of the boot configuration pins during booting. This can be an issue as the boot pins are often used for other functions, such as SPI, after booting.

200 - 220°C DEGREES
For this range, Texas Instruments offers 3 devices with ceramic packaging. These parts are rated for 210-220°C for 1000 hours. A common use for these are in equipment for downhole oil well services. They are obviously not suited for permanent downhole installation at this temperature.

Two of these devices, SM470R1B1M-HT and SM320F28335-HT, are also available in plastic packaging rated for 150°C and are described earlier in this article.

There are a few development kits that uses the SM470R1B1M-HT with ceramic packaging. The one typically used is the SM470 development kit manufactured by IAR available for 470 euros. No temperature testing can be performed with this kit as the microcontroller is the only part rated for temperature. This kit also includes an IAR J-Link Lite debug probe. Texas Instruments do not provide an affordable development kit for this part or the equivalent non-high temperature part.

The only offer from TI is the HEATEVM development kit at 5749 USD. This kit can be used for evaluating the SM470 (in ceramic packaging) microcontroller,

but also several other TI high temperature parts such as operational amplifiers, ADC and transceivers. The kit can be used for high temperature testing.

The TI device not discussed previously is the SM320F2812-HT. This is a 32-bit DSP with Analog-to-Digital converter. Recommended development tools for this device are TI Code Composer Studio v6 and TI XDS100/200/560 debug probe.

Vorago Technologies, former Silicon Space Technologies, has developed a microcontroller with an ARM Cortex-M0 core. The device, PA32KASA, is rated to 200°C for 1000 hours, but it has survived testing at 250°C for 2500 hours. Note that this device is quite large physically compared to other microcontrollers.

ALTERNATIVES AT 225°C DEGREES AND BEYOND
Processors operating beyond 200°C are not in frequent use, but there are some options. One is Honeywell's HT83C51 8-bit microcontroller. It is designed for 225°C operation. However, Honeywell claims «... parts will operate up to 300°C for a year, with derated performance». Tekmos also provides alternatives, the TK8xH51x family. It is designed for 250°C and this is also an 8-bit architecture.

EXTENDING DEVICE LIFETIME
Frequently, high temperature parts are specified for 1000 hours at certain temperature. If this lifetime is not sufficient for you application, you can select a device rated at a higher temperature as it will survive longer at a lower tempera-

ture. The figure above shows the lifetime for TI SM470R1B1M-HT in ceramic packaging dependent on temperature.

For 220°C, the device lifetime is 1 000 hours (1 month, 10 days), at 175°C lifetime is 4 000 hours (5.5 months), at 150°C lifetime is 12 000 hours (1 year, 4 months) and at 125°C lifetime is 40 000 hours (4.5 years).

CONCLUSION
In conclusion, during specification gathering pay particular attention to the temperature requirement of you application. If the temperature requirement can be reduced, you'll benefit from significant production cost savings. Development efforts are also significantly affected by the temperature requirement and this mainly affects the hardware development. The driving factors are time consuming temperature testing and documentation that is very limited specifying performance at maximum temperature.

If this is your first high temperature design, keep in mind that the tool support and documentation might not be at the same level as for non-high temperature processors. As always, discuss your application with others engineers with relevant experience.

Eliminating digital noise effects
using forward error correction

When a digital signal is affected by external noise, the received signal may end up being corrupted. In such situations, requesting retransmission may be impractical as the retransmission is subject to the same noise. Hence, being able to correct for such errors should be very helpful. In this article we will discuss issues related to this type of situation, and describe a solution.
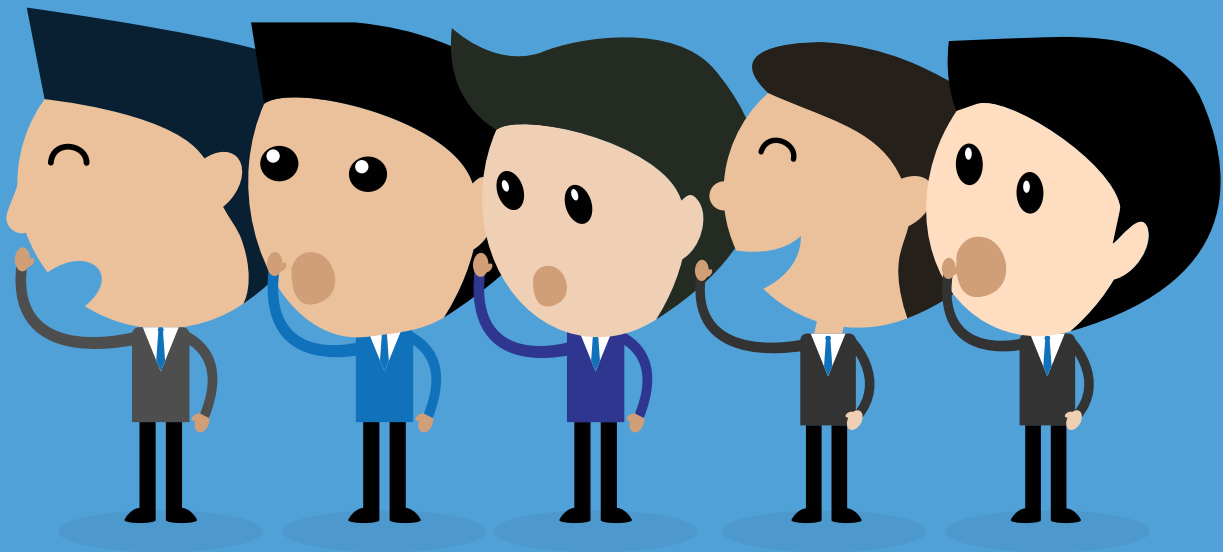
BY: Thomas Fredriksen
Development Engineer
Data Respons

A GAME OF "TELEPHONE"

In a game of "telephone", a phrase is made up by a judge, and whispered down the line of participants. The last person to receive the message has to say it out loud. The creator then tells everyone the original phrase, and everyone laughs at how morphed the phrase became.

Similar scenarios are commonplace in the world of data communication. If a message is morphed between the originator and the destination, the outcome can be dire. Thus being able to tell if the message is corrupted is crucial.

The easiest way of doing this is by the use of checksums. Checksums are some extra bits added to the end of a transmission and is used when verifying the integrity of a packet of data. The checksum is redundant as far as the actual data is concerned, hence the term *redundancy.*

The *cyclic redundancy* check (CRC) is a widely used checksum-type. It is easy to use, easy to implement and most importantly: It is small. A 32 bit CRC, coined CRC32, can support up to (2^{33} - 1) bit blocks, i.e. any blocks less than 8 Gb. Thus at the extreme, a block of
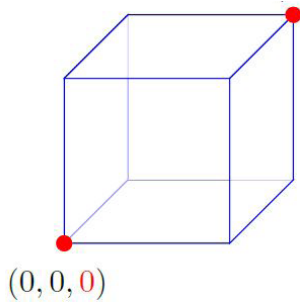
Figure 1. Unit cube

Message	Encoded	Received	Decoded
1	111	111	1
1	111	110	1
1	111	101	1
1	111	011	1
0	000	000	0
0	000	100	0
0	000	010	0
0	000	001	0

Table 1.
Triple repetition encoding/
decoding

2^{28}bytes and trailing CRC (4 bytes) give a *redundancy ratio* of about 1.5 x 10^{-8}. The redundancy ratio is a measure of efficiency, and is defined by the number of redundant bits to each data-bit. For example, if a coding-scheme has a redundancy ratio of 0.5, it means that there are 0.5 redundant bits to each data-bit, or two data bits for each redundant bit.

The receiver of a data-block would only need to verify the block by applying the same algorithm as the sender, and take measures if the received block is corrupt. This usually includes requesting a retransmission or discarding the block entirely.

CORRECTING THE MISTAKES OF OTHERS

Going back to our game of "telephone", being able to spot if something is wrong is nice, but is it helpful? In the game, one is not allowed to ask someone to repeat what he or she just whispered in your ear. Similarly with technology, what if you cannot ask for a retransmission?

The solution is *forward error correcting codes*. As before, we append redundant data to the transmission block, but the nature of this data have very different properties.

To explore what this means, we consider the simplest possible scenario: Assume we have a unit-cube (all edges have the length 1) with corners located at center and (x, y, z) = (1, 1, 1) as in figure 1. Put a particle in the cube, and assume that the particle may only be located at either (0, 0, 0) or (1, 1, 1).

If we wish to transmit the position of the particle to a recipient far away (say space), we would send one of two possible messages; (0, 0, 0) or (1, 1, 1), in binary form. Then, if the recipient reads the particle to be in one of the adjacent corners to its' actual position, he may infer where it actually is located by assuming it to be at the closest "valid" position. For example, if the sender transmits (1, 1, 1), but the but the reader reads (1; 1; 0), then since the closest \ valid" corner of the received coordinate is (1; 1; 1), the receiver thus assumes this to be the actual transmitted position.

This method is known as a *Triple repetition code* or *Hamming*(3; 1)-code, and is a very simple but inefficient type of forward error-correcting codes. It has a redundancy ratio of 2, and is very susceptible to bulk errors. In use, this would mean that a message would take three times as long to transmit, making it a poor choice for reliable systems. It is however very easy to implement, making it attractive to developers who need something "quick and dirty".

REED-SOLOMON CODES

Reed-Solomon (RS) Codes are another class of forward error-correcting codes. They are much more sophisticated than the Hamming(3; 1)-codes in that they require less redundancy and are more reliable.

The most common version of such a code is the (256; 240)-code. It consist of a 256 byte *block*, and contains 240 data-bytes and 16 redundant bytes. It has the ability to correct up to 8 byte-errors and can reliably detect up to 16 byte-errors.

Indeed this is true for general messages with size *k*. Thus adding (n - k) redundancy bytes, where *n* is final block size, gives the ability to correct up to (n - k)/2 byte errors, and the possibility to reliably detect up to (n - k) byte errors.

Since RS codes detect and correct byte-errors, it is less vulnerable to bulk errors. Several bits may be flipped within a byte, but count only as one byte error.

For the sake of explanation: The Cassini spacecraft has a maximum downlink bandwidth of 165.9 kbs. By adding 16 redundant bytes, the true data rate becomes 156.1 kbs. If we in addition assume that 1% of all transmitted bytes are corrupt, the checksum would be able to correct and verify 99.83% of all transmissions (see gure 2).

There are, however, certain issues with RS codes. First of all, they require more redundancy compared to CRCs to obtain reliable results. Secondly, transmission blocks must have a fixed size. This means that for RS codes of size N, messages of size n < N must be "padded", decreasing the true data rate
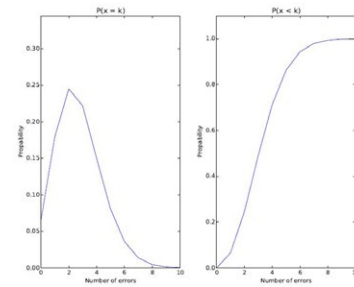

Figure 2. Left: Probability of error x errors. Right: Probability of less than x errors. Both scenarios assuming 256 block size and error rate of 1%

even more. This can however be overcome as the "padding" bytes take an expected form, and can thus in reality be removed during transmission.

IN PRACTICE

RS codes are used in a wide variety of applications. They have seen early use in space-communication, but is perhaps most famous for being the standard encoding on the CD format. Similar schemes are also used on DVD and DAT formats, and they also see use on two dimensional bar codes such as QR and Aztec.

They are most useful when requesting a retransmission is an expensive or impractical operation, or where systematic error may occur like demonstrated in figure 2. In the case of bar codes, the information may be missing altogether, and RS-encoding may then be used to recover the lost information.

For the sake of illustration, let us assume a scientific instrument that stream sensor readings to a remote master device. Bandwidth may be limited, and interrupting data-streaming is to be avoided as far as possible. At the same time, there is noise on the stream and observation tells us that approximately 1% of all bytes are corrupted. The situation corresponds to that of figure 2 and is an ideal candidate for forward error correction, and indeed such is the case with space probes and satellites.

The *New Horizons* spacecraft passed Pluto July 14th last year, and on its journey it made a gravity assisted flyby of Jupiter on February 28th 2007. Around

this time, the spacecraft took several pictures of the moon Io and sent it home to earth. Assuming the situation of figure 2, the following could be the outcome with and without forward error correction. We discard all corrupt blocks as if the nature of the data is worthless if it cannot be verified.
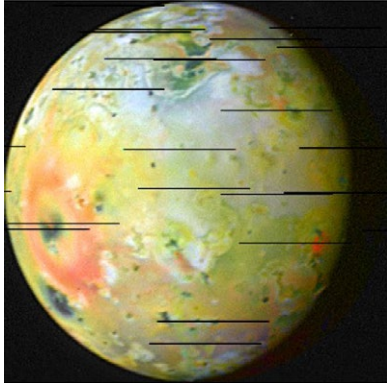
a) With forward error correction
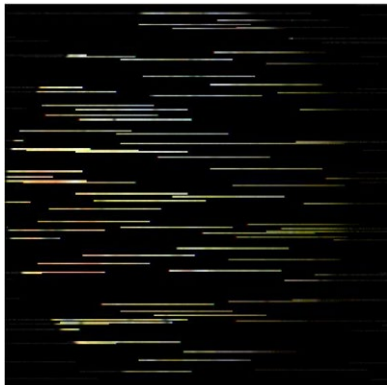
b) Without forward error correction

Figure 3. New Horizons' picture of Io[3]. All currupt blocks have been discarded.

We can clearly see that the RS-encoding has preserved most of the image (figure 3a), while it is unrecognisable without it (figure 3b). The discarded blocks would then have to be retransmitted, and and some of the retransmitted blocks may also end up being corrupt. This will repeat itself until all the packets can be verified. This would require an estimated 81 retransmissions, and a total of 3631 blocks of data. Since the space probe is near Jupiter, the ping is about 33 minutes, meaning that each retransmission would take at least 66 minutes (request + response), giving a total upload time of 239646 minutes, or just over 166 days for a perfect picture.

A LITTLE LESS FUN

In conclusion, it is easy to see the benefits of opting for a forward error correcting code. Albeit, it is not always necessary. If the bandwidth supports a retransmission every now and then, and is free from systematic errors as described above, forward error correction could be il-advised, as the extra redundancy could be used for actual data instead.

On the other hand, if you know that the message is most likely going to be broken, a smart thing to do would be to construct it in such a way it is easy to fix. The game of "telephone" would however be a little less fun.

Refrences

[1]   C. Clarke. Reed-solomon error correction. BBC R&D White papers, 2002.

[2]   D. A. Cox. Galois Theory. Wiley, 2nd edition, 2012.

[3]   N. of Arizona/Jason Perry. Plaques of lambda phages on e. coli xl1-blue mrf, 2008.

[4]   J. van Lint. Introduction to Coding Theory. Springer, 3rd edition, 1999.

ALICE PHOS
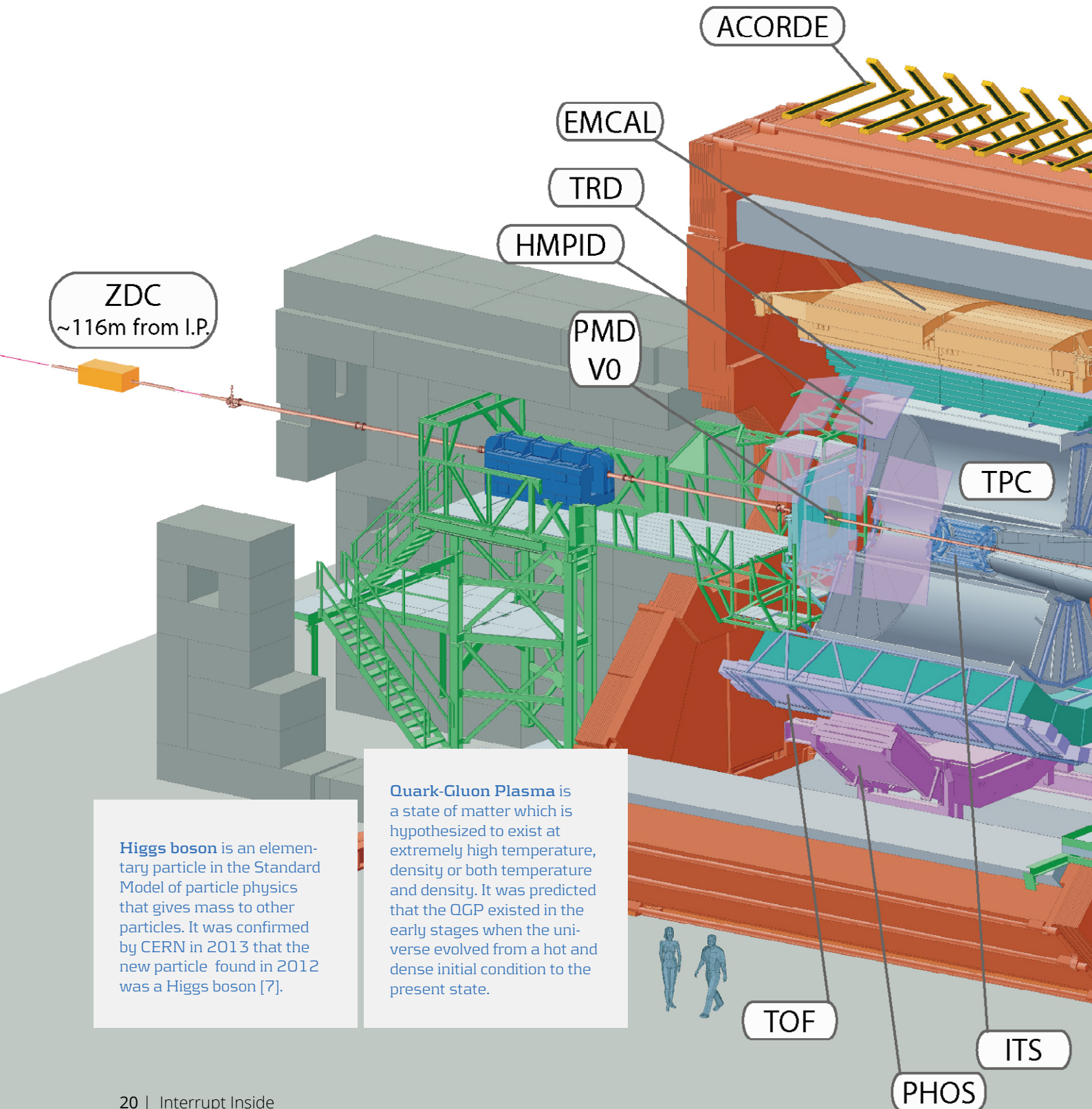## TRIGGER SYSTEM

ACORDE

EMCAL

TRD

HMPID

PMD
V0

ZDC
~116m from I.P.

TPC

Higgs boson is an elementary particle in the Standard Model of particle physics that gives mass to other particles. It was confirmed by CERN in 2013 that the new particle found in 2012 was a Higgs boson [7].

Quark-Gluon Plasma is a state of matter which is hypothesized to exist at extremely high temperature, density or both temperature and density. It was predicted that the QGP existed in the early stages when the universe evolved from a hot and dense initial condition to the present state.

TOF

ITS

PHOS

The Large Hadron Collider (LHC) is the largest and most powerful current accelerator in the world and was built mainly to find the existence of Higgs Boson and study Quark-Gluon Plasma (QGP). There are four collision points along the LHC, at which an experiment is built. One of the experiments/detectors is ALICE ("A Large Ion Collider Experiment"), which is a dedicated experiment for the analysis of lead-ion collisions, i.e. to study the strong interaction at high energy densities, such as the properties of QGP. This article explains the ALICE PHOS Trigger Systems. The PHOton Spectrometer (PHOS) is one of the sub-detectors located in the ALICE experiment. This article briefly introduces the ALICE online system and ALICE PHOS trigger requirement, describes the ALICE PHOS trigger system in detail, explains the development of the PHOS trigger, and gives the performance and status of the PHOS trigger system at the end.
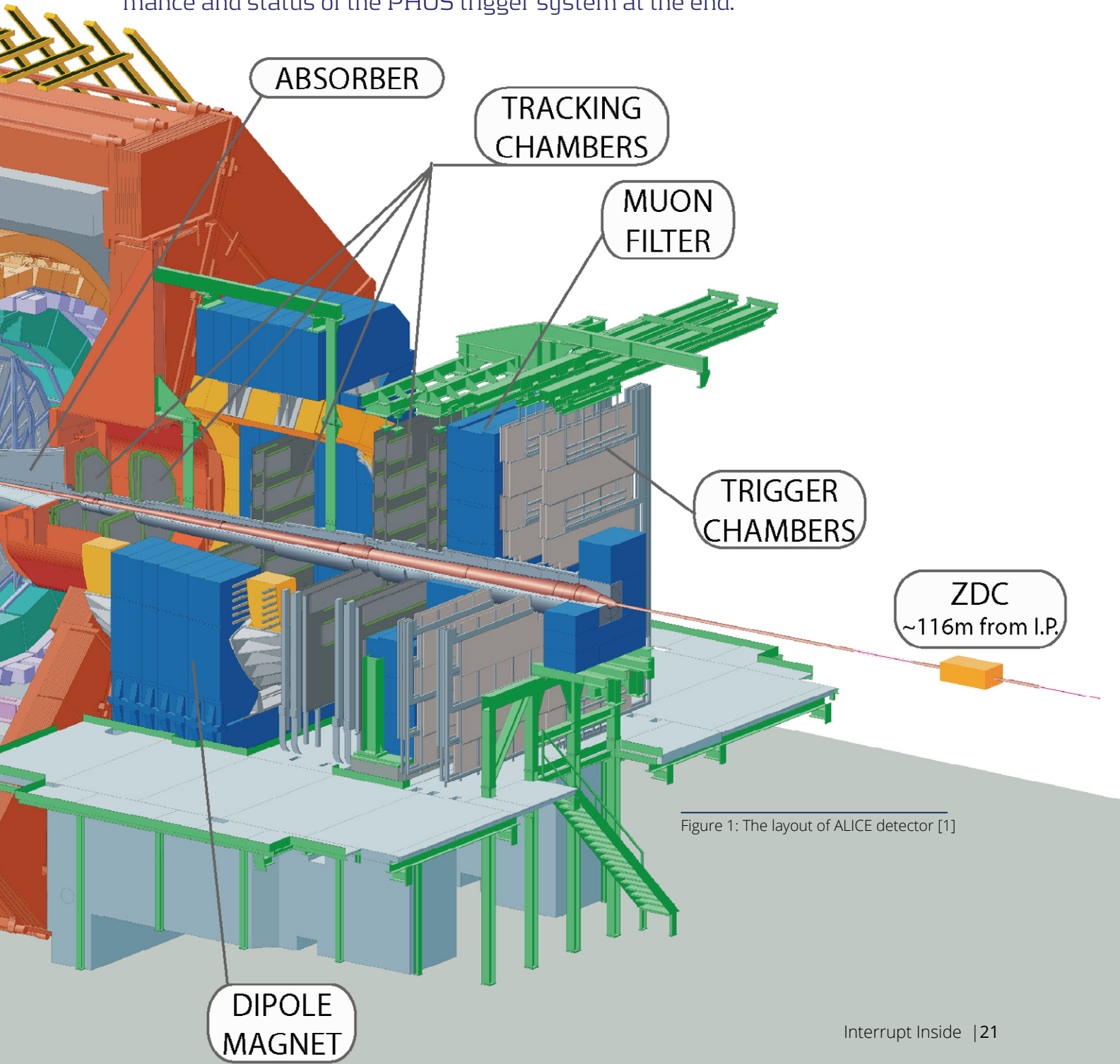
ABSORBER

TRACKING CHAMBERS

MUON FILTER

TRIGGER CHAMBERS

ZDC
~116m from I.P.

Figure 1: The layout of ALICE detector [1]

DIPOLE MAGNET

BY: Lijiao Liu
Senior Development Engineer
Data Respons

Liu's role in the ALICE experiment at CERN

Liijiao Liu developed the VHDL code in TOR, analysed and simulated the algorithms, and commissioned and analysed PHOS L0 trigger performance in 2008 and 2011 at CERN.

ALICE is designed to be able to track and identify particles from very low up to quite high transverse momentum. A set of sub-detectors (Figure 1) are placed in a moderate magnetic field to detect and identify hadrons, electrons and photons. The PHOS is used for the detection of photons.

Momentum is the product of mass and velocity (p=mv), and transverse momentum is measured perpendicular to the beam line.

ALICE CONTROL AND DATA ACQUISITION SYSTEM

There are four online systems in the ALICE experiment controlling, reading out, and monitoring different sub-detectors: trigger system, Detector Control System (DCS), High Level Trigger (HLT) and Data AcQuisition (DAQ). The Experiment Control System (ECS) coordinates the operations controlled by the four online systems.

The DAQ system collects data from sub-detectors, and stores it to disks, with the help of the HLT, which gets data from sub-detectors and decides which events to be stored. The DCS is responsible for configuring, monitoring, and controlling the equipment of the experiment remotely so that the ALICE experiment can be operated from a single place: the ALICE Control Room.

A **triggering detector** is a sub-detector that takes part in the generation of a trigger decision. A readout detector is a sub-detector that takes part in the readout of data.

The trigger system is designed to select events having a wide variety of different rates, which can be scaled down to adapt to physics requirements and the restrictions caused by the bandwidth of the DAQ. The task of the trigger system is to make optimal use of different sub-detectors, which vary in readout time.

Also it is optimised with trigger selections for different running modes: Pb-Pb (lead-lead), p-A (proton-nucleus) and p-p (proton-proton). The trigger system receives triggers from several triggering detectors in ALICE, makes decisions and selections, and then sends the final trigger decisions to readout detectors. The selection of triggering detector group depends on the running mode, the chosen physics observable and the trigger classes.

The triggers are divided into three levels because of the features of the triggers, the restrictions of triggering detectors and the requirements of readout detectors. L0 has an latency of 1.2 μs, meaning that the front end electronics of readout detectors need to receive L0 trigger within 1.2 μs after an event takes place. L1 and L2 triggers have latencies of 6.5 μs and 88 μs respectively [3][4].

ALICE PHOS TRIGGER REQUIREMENTS

The principal requirements for PHOS include the capability of identifying photons, discriminating direct photons from decay photons and performing energy measurements over a wide dynamic range with high energy and spatial resolution.

PHOS detector is required to generate L0 trigger and L1 trigger. L0 provides a minimum bias trigger for p-p collisions, i.e. the L0 generated by PHOS implies only that there are photons hitting the PHOS detector. As the trigger system needs 400 ns to handle triggers, the L0 trigger from PHOS shall arrive at the trigger system within 800 ns after a collision takes place. L1 trigger gives more information about the collision: Are they direct photons or decay photons? What is the centrality of a collision? Is the photon in the energy range of interested? The L1 trigger is a fast trigger with more processing time than the L0 trigger and it must arrive at the trigger system within 6.1 μs after collisions have happened.
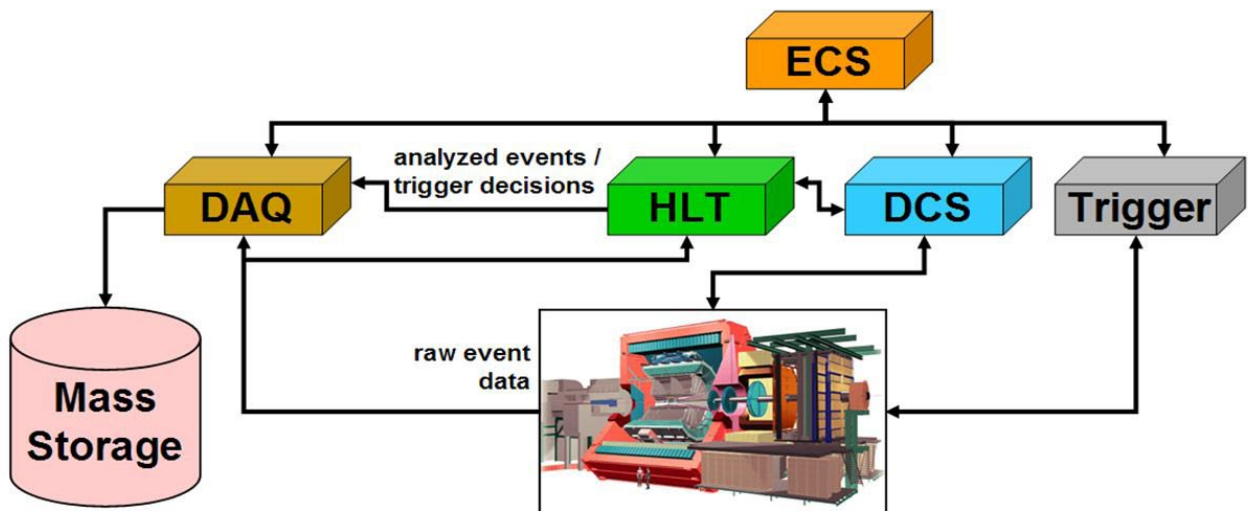

Figure 2: ALICE online system [2]

DATAFLOW OF ALICE PHOS SUB-DETECTOR

The PHOS detector consists of 5 (when I worked there, there were only 3 modules installed, one more module was installed in the shutdown period 2013-2014) modules. Each module has 56 x 64 scintillation crystals (i.e. photon detector units), which are divided into 4 readout partitions. Every partition includes 56 x16 photon detector units and is read out by one Readout Control Unit (RCU). There are two branches, A and B, included in one readout partition, with each branch consisting of 14 Front End Cards (FECs) and one Trigger Region Unit (TRU). 14 FECs together with one TRU in one branch are connected via three flat cables. The TRU defines the trigger region, connecting to all the FECs on the same branch. Each FEC collects signals from 32 crystals. Finally, the Trigger-OR (TOR) connects to all the TRUs from 5 modules.

Figure 3 shows the dataflow of the PHOS trigger and readout system. The signals coming from 32 crystals go into one FEC. There are two data paths for the signals on a FEC. One of them consists of a dual gain shaper and a custom ASIC (ALTRO) to digitize the energy from each crystal directly. The other is for the Analog-sum. The L0 triggers are generated in the TRU and then ORed in the TOR, whereas the L1 triggers are generated in the TOR. Both the L0 and the L1 triggers are delivered to the Central Trigger Processor (CTP) for combination and synchronization of information from all the triggering detectors in ALICE. Then the trigger-related sequences consisting of combined and synchronised information are sent to the LTU, where they are converted into the proper format for being delivered to the front end electronics of readout detectors via optical fibers. The RCU decodes the trigger information and sends "L1 strobe" (i.e. Confirmed-L0 for
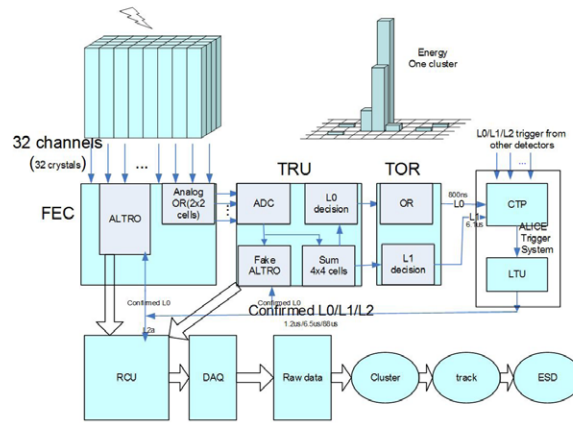

Figure 3: The dataflow of the PHOS trigger and readout system [4] (edited); the FECs are called front end electronics, TRU and TOR belong to trigger electronics.

PHOS) and "L2 accept" (i.e. L2a in Figure 3) to the FECs. The ALTROs on the FECs and the FakeALTROs on the TRUs start to buffer the data on L1 strobe, and data is shipped to DAQ when the trigger information is decoded as L2 accept.

ALICE PHOS TRIGGER ALGORITHMS

The trigger algorithms work on the basis of the digitised signals in TRUs. The L0 trigger , the basic L1 trigger, and the total transverse energy trigger are generated based only on the energy information; whereas the isolated photon trigger depends on both energies of photons and associated addresses.

Figure 4 shows energies deposited by two photons decayed from 5-GeV π0 (identify direct photons means exclude decayed photons). Each grid represents energies on 4 x 4 crystals. If there is no photon inside the area between two squares marked by red line, the photon in the center of the squares is considered to be isolated.

ALICE PHOS TRIGGER CHALLENGE AND DEVELOPMENT

The trigger algorithm is mainly implemented in Virtex FPGAs: one FPGA (Virtex2Pro) on TRU (L0 decision, Fake ALTRO and Sum 4x4 cells were implemented in
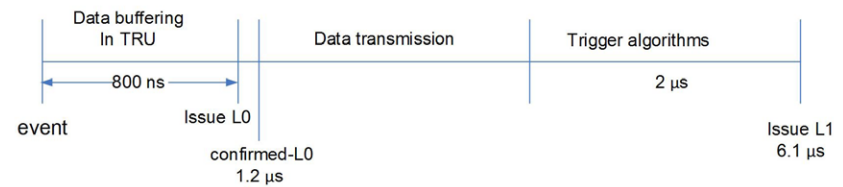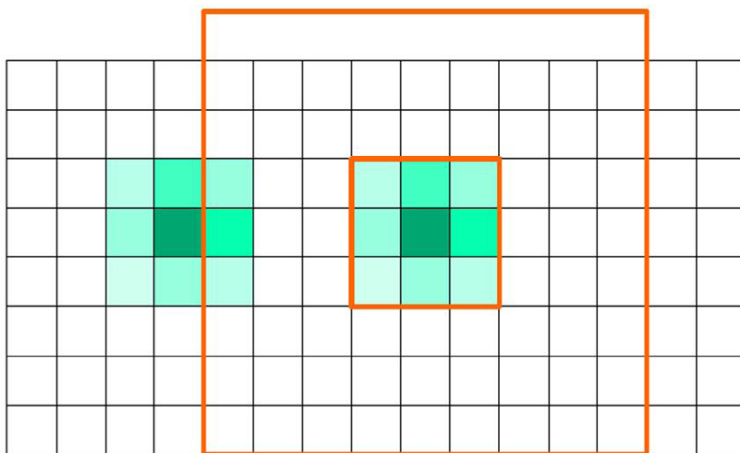

Figure 5: Trigger timing

Figure 4: An example of the distance of two photons from 5-GeV π0 in units of 4 x 4 sums (energies on 4 x 4 crystals) [4]

FPGA in Figure 3), one FPGA (Virtex4) on TOR. The FPGA on TRU needs to handle 112 channels of 12-bit data from FECs. The generated L0 trigger together with the raw data needed to be sent to TOR for further processing. This means 40-channel data will be sent to one TOR, i.e. one Virtex 4. The timing of trigger is given in Figure 5.

The LHC global clock is 40.079 MHz (bunch crossing clock), and everything should be synchronized with it. When an event takes place, it takes around 375 ns to arrive at TRU. The signal is then converted to digital and buffered in the TRU. The ADC latency is 6 clock periods, which adds 150 ns. Then only 275 ns is left of the 800 ns for L0 generation. For L1 generation, the raw data must also be shipped to TOR. Assuming that the trigger algorithm takes 2 µs, there is only 2.9 µs for data transmission.

To summarise, the challenge of trigger generation is that there is too much data transmission in parallel and too little time for transmission and data process. The focus of the trigger generation is on optimizing the serial transmission, reducing data, and improving algorithms.

Advanced technologies for high speed serial transmission have been developed in recent years, e.g. data recovery with 8X oversampling, Rocket IO, and dedicated serialiser and deserialiser. But the TRU and TOR were designed in 2004, when the technologies were not as advanced as today.

nating the event recorded by the noisy branches, the efficiency stays stable at about 0.95 at the photon energy over 4.5 GeV [5].

The trigger purity is defined as N/Nt, where N means the real PHOS triggers

> The challenge of trigger generation is that there is too much data transmission in parallel and too little time for transmission and data process.

Because of the limitation of the Virtex 2Pro, the TRU could not receive 112 channels of data from ADCs serially at 40 MHz, the internal clock is reduced to 20 MHz, so that there is only 125 ns for L0 generation. Only an extremely simple algorithm could be implemented successfully in the limited time. The energy of one photon would cover 3 x 3 crystals, each channel goes into TRU represents energies on 2 x 2 crystals. So a sliding window that covers 4 x 4 crystals was applied to sum energy of one photon and compare with a threshold to generate L0 trigger.

ALICE PHOS L0 TRIGGER PERFORMANCE
From the perspective of physics, the trigger efficiency can be defined as Np/Nall, where Np means the number of events triggered by PHOS and Nall means the number of all the events with photons over the threshold. PHOS L0 trigger was run at CERN in 2012. With the PHOS L0 trigger threshold set to 4.3 GeV in the p-p collision at 8 TeV, the PHOS L0 trigger has an efficiency of about 0.7 at the photon energy over 4.5 GeV. Some missing events at high energies observed can be caused by the noise on FECs. By elimi-

with photons whose energy is over the threshold, Nt means all the PHOS triggers. During the run in 2012, 36% and 17% of the PHOS triggers are real ones with thresholds set to be 4.3 GeV and 2 GeV respectively. The fake triggers are mainly generated by the long pulse of PHOS trigger due to the reduced clock and noises in electronics [5]. Therefore, the trigger electronics were upgraded in the shutdown period 2013-2014 [6].

Refrences

[1] Pcharito (Own work) [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons from Wikimedia Commons

[2] Sebastian Robert Bablok, "Heterogeneous distributed calibration framework for the high level trigger in ALICE
" PhD dissertation, University of Bergen, 2009, https://bora.uib.no/handle/1956/3857

[3] Lijiao Liu, "L0/L1 trigger generation by the ALICE PHOS detector", PhD dissertation, University of Bergen, 2011, http://inspirehep.net/record/1382032/files/liu_materie.pdf

[4] ALICE-Collaboration, Journal of Instrumentation 3 (2008),
doi: 10.1088/1748-0221/3/08/S08002.

[5] C Zhao, L Liu etc. "Performance of the ALICE PHOS trigger and improvements for RUN 2", Journal of Instrumentation 8 (2013), doi:10.1088/1748-0221/8/12/C12028

[6] D.Wang, F.Zhang etc. "Readout electronics upgrade on ALICE/PHOS detector for Run 2 of LHC", Topical Workshop on Electronics for Particle Physics 2014, AIX EN PROVENCE, FRANCE, doi:10.1088/1748-0221/10/02/C02025

[7] http://home.cern/about/updates/2013/03/new-results-indicate-new-particle-higgs-boson

SoC FPGA
EVALUATION GUIDELINES

FPGA is a constantly evolving technology, especially in terms of logic density and speed. Among the newest improvements in the FPGA world are System on a Chip (SoC) FPGA devices. A SoC FPGA integrates a hard processor core and programmable logic on the same die. The three largest FPGA vendors, Xilinx, Altera and Microsemi (Previously Actel), have all started to manufacture such devices. Although having in common that they all put a hard 32-bit ARM processor together with programmable logic, there are also considerable differences between the designs that will be discussed in this article.

BY: Inge Nikolay Torsvik
Development Engineer
Data Respons

The number of commercially available SoC FPGA devices is continuously increasing, giving a large variety of configurations to choose between. If you encounter a project where a processor and an FPGA is needed or preferred, you should know what a SoC FPGA is and understand when such a solution could be useful. The purpose of this article is to highlight the most important questions to resolve when evaluating and selecting SoC FPGA solutions.

OVERVIEW
A SoC FPGA could be useful for replacing the following traditional configurations:

- A stand-alone processor and stand-alone FPGA
- An ASIC including a processor
- A stand-alone processor

Compared to using a stand-alone processor and a stand-alone FPGA, a solution using a SoC FPGA is cheaper, uses less power consumption, and is easier to put into a design. Two circuits are replaced by one, which means less time for designing and less space on the PCB.

If external RAM is used for both the processor and the FPGA, these memory circuits can be consolidated into one RAM chip, saving space, cost and reducing complexity. Communication between processor and FPGA can also go much faster with both units on the same chip.

Compared to manufacturing an ASIC, a SoC FPGA is first of all much cheaper and requires much less design time. You will also have a much more flexible design process, as the firmware may be rewritten at any time.
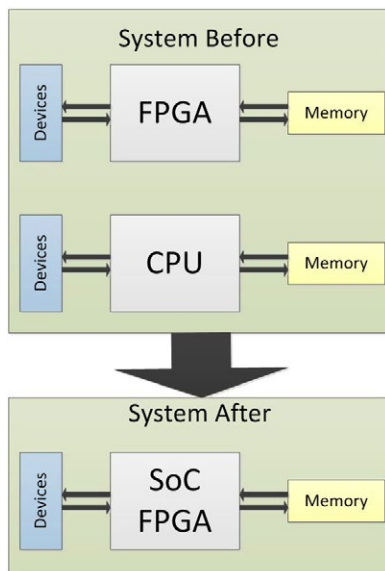
Figure 1: When going from a stand-alone CPU and FPGA solution to a SoC FPGA solution, both the number of components and number of external interconnects drops considerably.

priate SoC FPGA solution for a given application. We will look further into performance, reliability, flexibility, cost, power and software tools.

PERFORMANCE EVALUATION

What ultimately constrains your system performance is the communication between processor, memory, FPGA logic, and interconnects. Normally, when selecting the SoC FPGA to use, the processor speed and the FPGA logic are carefully evaluated and checked, but the same is sometimes not done properly for memory and interconnects.

In terms of memory, there are different specifications that determine the memory performance – the supported memory frequency is not necessarily what gives the highest performance. One of the first things to check is:

- Does the SoC FPGA have an independent, hard memory controller for both the processor and the FPGA logic?
- Is the memory controller able to extract data in sufficiently high speed, manage priority, reorder command and data, and schedule pending transactions?

A slow memory controller with smart data handling could perform better than a high frequency memory controller with simplistic data handling. It is important to make sure that the memory controller can provide the required throughput

speed transfer between the FPGA and the processor, a low latency path should also exist to do simple setup and configure access to the hardware accelerators in the FPGA logic.

RELIABILITY EVALUATION

In terms of reliability, there are especially two things that have to be carefully checked, and that is memory protection and software bug handling. With increasing CPU and Memory speed and decreasing semiconductor-manufacturing technology, there is an increasing probability to have memory errors. It is therefore important to protect your memory, for example with Error Correction Codes (ECC).

If both the processor and the FPGA share the same memory, you must be sure that the memory is protected from overwriting data. If there is no protection, one can spend weeks on debugging software bugs which may be caused by the FPGA overwriting memory used by the CPU. A memory protection unit can fix this.

One thing that is certain is that you are going to encounter bugs during your software development, and it is important to handle these in an appropriate way. Implementing a watchdog timer to reset the CPU in case of a system hang-up is a classical approach. A good architecture lets you choose if you want to reset both the CPU and reconfigure the FPGA or only reset the CPU.
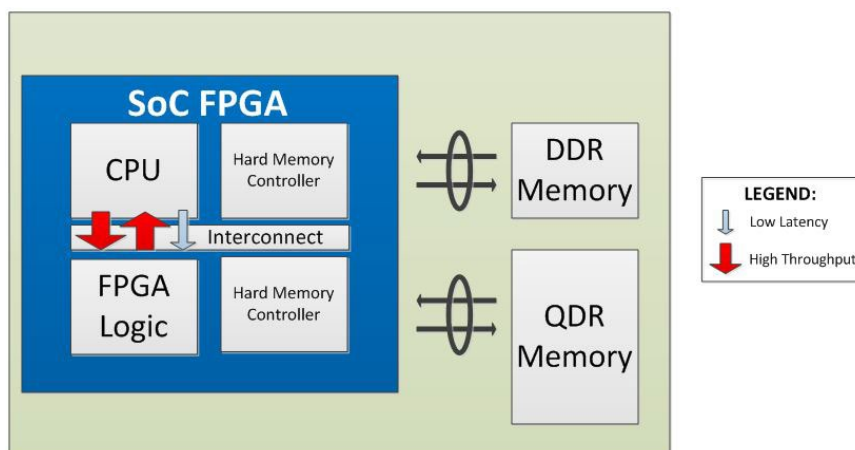

Figure 2: Interconnection between CPU and FPGA logic with low latency path and high throughout path.

Compared to using a stand-alone processor, a SoC FPGA will be more flexible, since hardware structures can be added throughout the whole design process when needed. It also gives the possibility of parallelising the data processing by allocating computing intensive operations to dedicated FPGA firmware.

In the rest of this article, we will go through the most important issues that must be resolved when picking an appro-

of mass data transfer and have the minimum latency to meet real-time requirements.

When it comes to connections between major building blocks inside of the SoC FPGA, like hard CPU and data processing FPGA blocks, it is important to check the interconnection speed, and make sure it supports the required data throughput between FPGA logic and processor. In addition, to prevent blocking the high-

FLEXIBILITY AND COST EVALUATION

We have already discussed the reduced cost of a SoC FPGA compared to a stand-alone FPGA and a processor devices. Component cost, design cost and PCB space can be saved simultaneously. Selecting a sufficiently flexible SoC FPGA can also contribute to a significant cost reduction for your application.

An important issue to look into is the question of which hard IP blocks are in-

tegrated into the device. Things to look into here could be PLL circuitry, appropriate memory controllers and communication modules (for example SPI, UART, I2C, USB, and CAN).

Although most required units can be generated by soft IP, it is more efficient in terms of performance, cost and logic utilisation when hard IP blocks can be instantiated directly.

Boot setup is another topic to explore: Does the SoC FPGA`s processor have the option to boot independently of the FPGA configuration, and then configure the FPGA from the CPU? And vice versa, is the FPGA able to boot first, and then boot the CPU through FPGA logic? Even though both the processor and the FPGA is in the same device, it is important that they can operate like two separated chips.

project, one can assume that 60-70 % of the project time is spent on debugging. Compared to CPU debugging on a fixed hardware platform, one has to be aware that the SoC FPGA firmware may be changed at any time throughout the design process. It is therefore important that the software-debugging tool is able to adapt to changes in the FPGA logic.

Another issue is cross-triggering. When a breakpoint is set in software, you would want the FPGA to freeze at the same time. Like that, you may perform a proper inspection. Similarly, when setting a logic analyser trigger in the FPGA, you would like to freeze and inspect the CPU software. You would also like to have a debugging tool that not only helps you find the mistakes in the code, but also tells you something about the performance of your code. For example, you would like to know if it can be optimised, and if

solution. These are all general guidelines and you will probably have to do some exploration work yourself to arrive at a SoC FPGA solution which is suited for your particular application.

SOURCES
https://www.altera.com/products/soc-overview/architecture-matters.html#squaresbox-4

> *For example, you would like to know if it can be optimised, and if there are functions/code that are not being used or are no longer running.*

POWER EVALUATION
With increasing clock speeds and higher performances, power consumption has become one of the biggest challenges, if not the top design criteria, for many new designs. By replacing a stand-alone CPU and FPGA solution with a SoC FPGA, one can reduce the power down to 50% of the power consumed by the original two-chip system.

The SoC FPGA can also save a considerable amount of power if it is able to put the FPGA in a low-power standby mode while keeping the CPU alive and running. We have already mentioned that a memory controller supporting high frequency RAM is not necessarily better than a memory controller that runs at lower frequencies. It all depends on the way the memory is handled. A slow DRAM controller does also have the advantage of lower power consumption.

EVALUATING THE SOFTWARE TOOLS
When looking at Software tools, one of the most important things to explore is debugging capabilities. In a development

there are functions/code that are not being used or are no longer running.

For ARM processors with two or more cores, it is desirable to have multicore debugging. Multicore debugging makes you able to control and monitor both cores simultaneously. For example, you may put a breakpoint on one of the cores while the other is still running, or have the possibility to stop both cores at a given breakpoint.

CONCLUSION
When carefully selected, SoC FPGA circuits can perform better than their traditional counterparts can. They have therefore become a highly relevant competitive alternative.

Throughout this article we have highlighted important features to evaluate when selecting an appropriate SoC FPGA for a given application. We have looked into aspects like performance, reliability, flexibility, cost and software tools, and we have explored important issues to resolve before arriving at a particular circuit
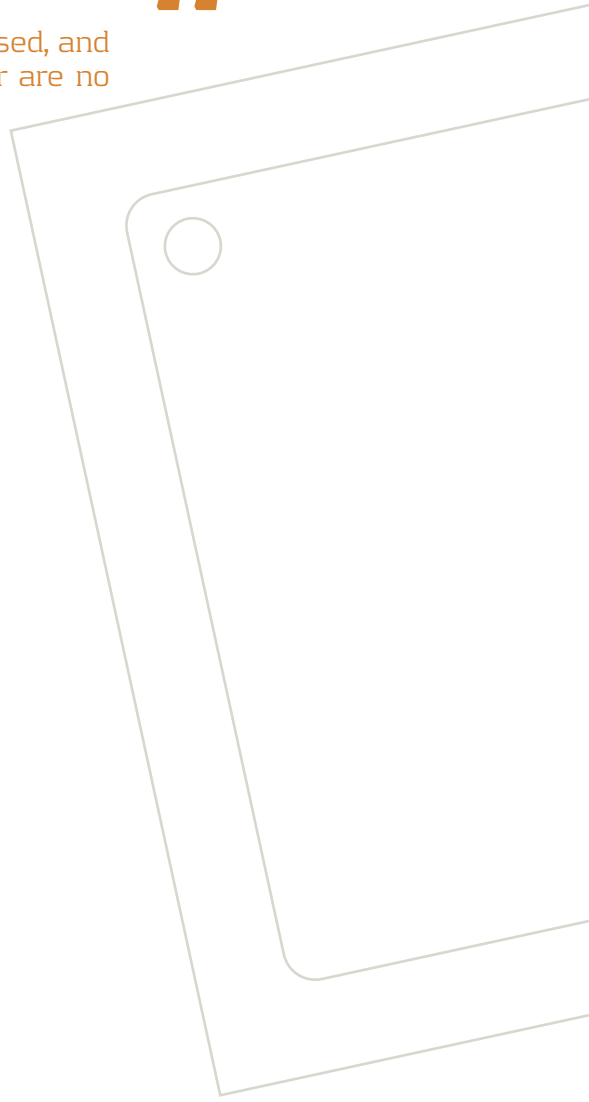
DRONES
& WIRELESS VID

Lately, drones have become very popular both as professional tools and for recreation and air sports competitions. In a series of articles we will present the different systems in a drone, including the ground support equipment. This article will focus on how wireless video systems are implemented to give a First Person View (FPV) to the pilot remote controlling the drone.

BY: Bjørn Bergersen
Specialist Development Engineer
Data Respons

Drone or UAV (Unmanned Aerial Vehicle) are generic terms that includes many types of unmanned, remote controlled aerial vehicles. This includes fixed wing planes, helicopters and multi-rotors. Professional drones have a wide range of applications. Aerial photography during sports events does not have to rely on expensive full size helicopters, and real-estate agents frequently use drones for documentation. Drones discover missing people, and can monitor habitats exposed to risk of pollution. Electricity companies are now inspecting some of their high-voltage lines without expensive power outages and risky climbs. Even a conservative industry like the railway companies are considering drones for inspection of disrupted tracks in areas with limited access. Several companies plan to deliver small packages by drones, but it is not a commercial reality yet. This is partly due to regulatory limitations.

DEO

WIRELESS VIDEO FOR FIRST PERSON VIEW

Drones can be piloted in two different ways, either line of sight by visually observing the drone, or by First Person View (FPV). In an FPV system the video image from an onboard camera is transmitted by radio to a personal video display on the ground in the form of a screen or video goggles. Figure 2 shows a typical set of video goggles with circular polarised antenna and embedded receiver. Systems span from simple low-cost systems, to advanced systems with high-power video transmitters and ground receivers with directional tracking antennas that offer ranges of tens of km.

CHALLENGES AND ANTENNA SOLUTIONS FOR WIRELESS VIDEO

The range of the wireless video link is limited by a number of factors. The path loss itself will diminish the signal when distance increases, and obstacles in the line of sight can give additional attenuation. However in a natural environment there are some less obvious challenges to the radio-link that require clever solutions. We will take an in-depth look at the two main issues.

INTERFERENCE

Other sources of radio transmission in the environment can interfere with the main signal. If the interfering signals occur in the same frequency band as the wireless video link it will act as in-band noise. This will reduce the signal to noise ratio, resulting in a noisy video image and limited range of the link. A typical interferer can be the video transmitter of another drone in the area, a nearby WiFi hotspot or mobile phone. The problem can be minimised by selecting a channel as far away in frequency from the interferer as possible, or by moving the video receiver and antenna.

Fig. 2 FPV with video goggles

If the source of interference is powerful, but outside the band of the wireless link, it is called a blocker. The blocking signal can penetrate insufficient front-end channel filtering, and decrease the dynamics of the Low Noise Amplifier (LNA). A simplified diagram of a receiver signal chain is shown in Figure 3. Typical high power blockers can be radars, broadcast towers or military radios. Technical measures for handling interference can be to ensure good front-end channel filtering of the video receiver, and use a directive ground antenna to minimise interference from other directions. Directional antennas with a narrow beam and high directional gain will also increase the receive strength of the signal from the drone. The antenna can even be equipped with a tracker that automatically directs the antenna at the moving drone. A tracker takes GPS coordinates

>>

Larger military drones have been common for decades, but recently small stealthy nano UAVs have been developed for shorter reconnaissance missions. Figure 1 shows the Prox Dynamics Black Hornet which is also intended for use by rescue workers. It can give situational awareness, or discover victims trapped in collapsed building structures.
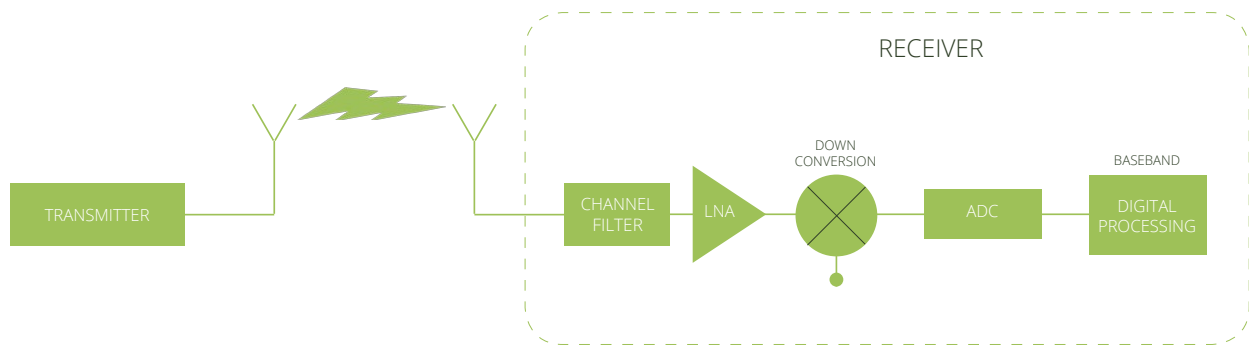
Fig. 3 Simplified Receiver Signal Chain

from the drone as input for the control system and the tracking algorithm.

MULTIPATH FADING DUE TO REFLECTION

Even with a strong, noise free signal, a radio link can get sudden dropouts, especially in cluttered or urban environments. This can be due to the reflected propagation path cancelling the direct propagation path. The cancelling occurs because of the phase shift associated with different propagation delay. This occurs at a specific point of the receiving space, and can disappear just by moving the antenna less than one wavelength. In addition to signal cancellation, multipath propagation also results in symbol delay spread. The symbols from the various paths arrive at different time, causing bit errors if the delay is significant. Figure 4 shows the principle of multipath propagation and delay spread.

The two main strategies to deal with multipath fading are avoidance or constructive combination of reflected signals.

AVOIDANCE OF REFLECTED SIGNALS

The most intuitive way to avoid reflected signals is to have a directional focus at the line of sight. As explained previously this can be accomplished with directional antenna systems. The reflected signals will approach at an angle outside the main lobe of the antenna, and will be attenuated. Another less intuitive, but yet simple way is by adding antenna polarisation. The most effective type for this purpose is circular polarisation, where the radio wave due to antenna shape will have a twisting propagation. The receiving antenna with matching polarisation will pick up this signal but will suppress un-polarised signals. Reflected signals are avoided because the reflection disrupts the polarisation.

Figure 5 shows circular polarised antennas connected to a diversity switching FPV receiver. An important factor to consider is antenna positioning on the drone, because losses are introduced when the angle between the transmit and receive antennas is increased. At exactly 90° the loss is theoretically infinite, so a combination of antennas with different angles could be beneficial for a drone. This strategy could as well diminish the impact of flying into the antenna radiation pattern null. Omni directional antennas have a doughnut-like radiation pattern, with a null at the perpendicular axis. When flying directly over the ground station antenna, the drone can hit this spot. The strategy of reflection avoidance has one major downside, and that is the dependence of a line-of-sight path. When reflected signals are suppressed, major obstacles in the radio path cannot be handled efficiently resulting in a rapidly declining signal.
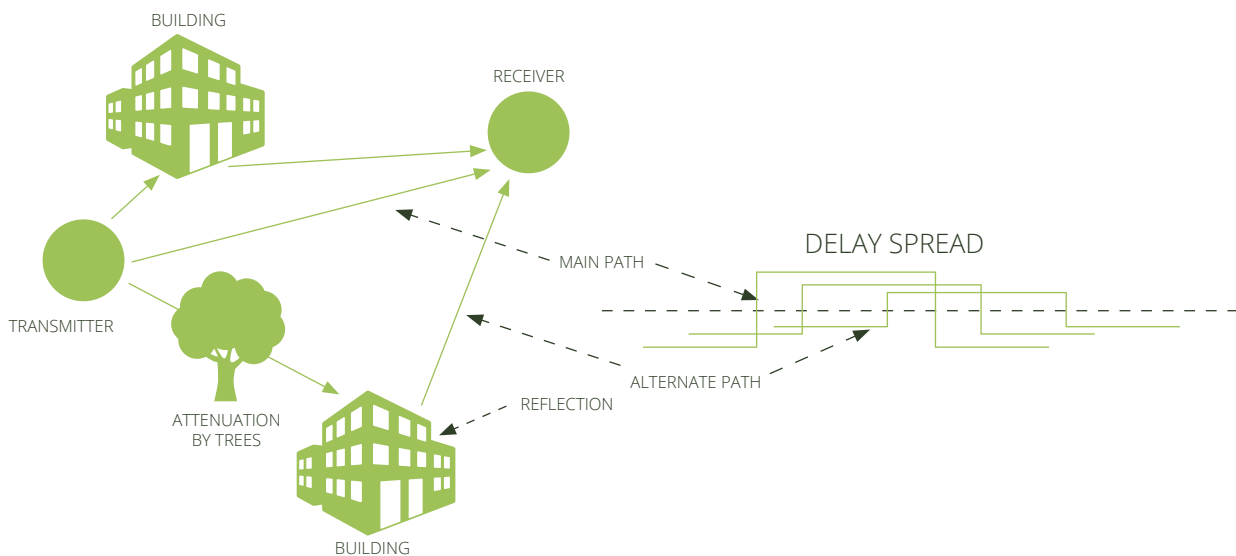

Fig. 4 Multipath propagation

Fig. 5 Diversity Receiver with circular polarised antennas, one directional and one omnidirectional

CONSTRUCTIVE COMBINATION OF REFLECTED SIGNALS

Constructive combination of reflected signals can solve not just the multipath fading problem, but also maintain sufficient signal strength when path obstacles occur. A commonly used mechanism for constructive combination of multipath signals is diversity. Diversity can be employed both on the transmitting and receiving side, by having two or more antennas connected. When the antennas are spaced by a distance of around a wavelength or more, the probability for severe signal cancelling is greatly reduced. The signals from the antennas can be passively summed before entering the receiver, or the receiver could actively switch to the antenna with the strongest signal. The radio can also have duplicated receive chains and sum the signals at baseband.

A more sophisticated mechanism evolving from diversity is Multiple Input Multiple Output (MIMO). Here a higher number of antennas are used to be able to receive and analyse several multipath signals. The phase and amplitude of the signals at each antenna is measured and compared, and an advanced and fast algorithm calculates how each signal must be processed to make up the wanted signal. Each signal must be precisely phase shifted and gained to enable constructive combining. This process is called weighting. Combination and individual treatment of a higher number reflected paths, gives a much higher signal to noise ratio than diversity. Consequently more advanced modulations can be employed, resulting in higher link bandwidth. The weighting of each signal can be done at different stages in the
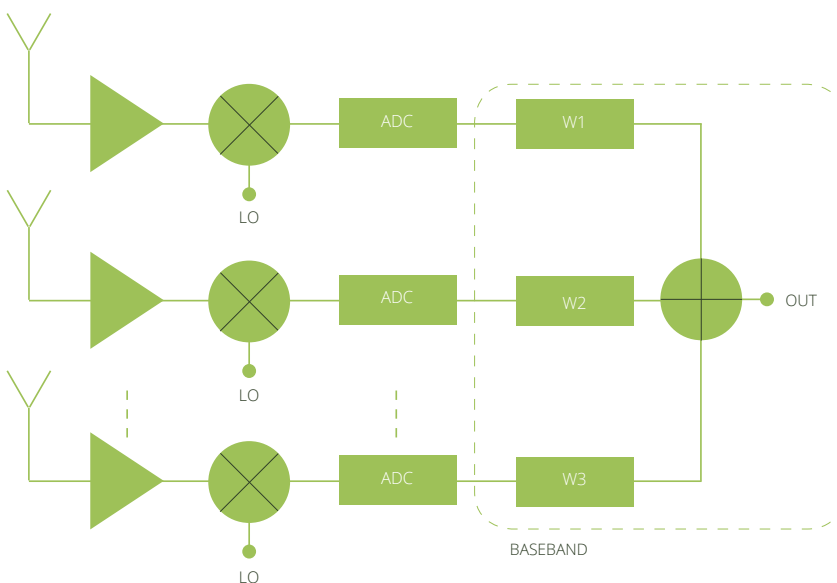

Fig. 6 MIMO with base-band weighting

FPV camera

The camera used in FPV setups is typically light and compact, but robust and very light sensitive. Apart from the optics, the image sensor is the most critical part for high quality video. There are currently two main types of image sensors available: Complementary Metal Oxide Semiconductor (CMOS) or Charge Coupled Device (CCD). CMOS sensors are most commonly used. They are inexpensive, require low operational voltage, and have lower power consumption than CCD sensors.

Most drone pilots demand high performance video, even when moving fast and during vibrations In this area CMOS image sensors have some shortcomings. Images are actually scanned pixel-by-pixel, line-by-line, and there is a limit to this scanning speed. This way of capturing the image is called rolling shutter. On a fast moving drone with high frequency vibration from the motors and propellers, artifacts, distortion and even wobbling will be noticeable on the image.

CCD cameras on the other hand has a different way of collecting the image. The light-induced charge of each pixel is collected from all lines simultaneously. This is referred to as global shuttering and is the reason why CCD cameras can handle fast movement and vibration without image distortions. The cells or pixels require very small amounts of light, and hence the sampling of the full frame can be fast. Since the CCD readout and signal conversion requires less active circuitry on the sensor itself, the raw pictures will contain less noise than the output from a CMOS sensor. CCD sensors will also have a higher dynamic range, i.e. a bigger difference between low-light threshold and saturation of a pixel. The best CCD cameras for consumer FPV setups are sufficiently light sensitive even for night flights, with only a few light sources present. CCD cameras however are more expensive, and require a higher operational voltage to run the charge-coupled cells in the image detector. The signal conversion circuitry also consumes more power than the distributed CMOS pixel amplifiers.

receiver signal chain. With the advances in compact and low-power digital processing, it is most common to do this on the baseband samples after analog to digital conversion. Figure 6 shows a simplified block diagram of this principle. The MIMO concept is widely used in LTE (Long Term Evolution) or 4th generation mobile data transmission, and has recently been adapted for use in high-end digital video links for FPV.

ANALOG FPV TRANSMISSION

Most consumer, and even professional FPV systems still use analog wireless video transmission, with very compact and low-cost systems. Figure 7 shows a complete analog FPV transmitter system giving approximately 1 km range (an AA-cell is shown for size comparison). Analog video is real-time and requires no image compression and advanced processing. This results in near zero latency between the image captured by the camera and the one viewed by the pilot. Latency of just tens of milliseconds is very noticeable when flying fast with FPV, and the hundreds of milliseconds often seen in standard digital High Definition (HD) systems for consumer use will pose a safety threat. Another surprisingly favorable property of analog video is the gradually increasing picture noise when the video link starts to break down. This occurs when reaching the boundaries of the radio range, or encountering interference or obstacles. A pilot experiencing this instinctively knows that he has to turn around the drone or avoid an obstacle. With a consumer grade digital video link over radio,

a rapidly decreasing signal to noise ratio will often result in stuttering and frozen video frames.

DIGITAL FPV TRANSMISSION:

The last couple of years intense development has gone into creating compact and low latency HD video links. The digital radio transmission system needs robustness to preserve an acceptable framerate even under rapidly deteriorating radio conditions. Fast adapting dynamic modulation schemes are applied, to narrow the required bandwidth during a period with reduced signal quality. A reasonable framerate

(different frequency components of the signal experience different fading). With a traditional single wideband carrier, frequency selective fading is complex to handle. OFDM mitigates the problem by converting the high speed serial data into parallel low bandwidth subcarriers. Some subcarriers are reserved as pilot carriers (used for channel estimation/equalisation and to combat magnitude and phase errors in the receiver) and some are left unused to act as guard bands. The reservation of subcarriers to guard bands helps to reduce the out of band radiation, and thus eases the requirements on transmitter front-end
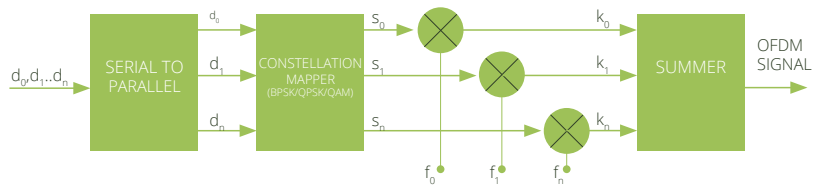

Fig. 8 OFDM transmitter principle

is maintained while restricting video resolution. A wireless high framerate HD videolink has to carry very high bitrates, and hence an advanced modulation is needed. This puts a strict requirement on the signal to noise ratio, so a variant of the antenna MIMO concept previously explained is applied, together with Orthogonal Frequency Division Multiplexing (OFDM). OFDM is very effective for communication over channels with frequency selective fading

filters. In the receiver fading correction is applied to each subcarrier. This can be seen as a form of diversity, called Frequency Diversity. OFDM is a well-known principle from WiFi and Wimax systems, and DAB broadcasting. Figure 8 shows simplified block diagram of the OFDM concept.

The underlying modulation is dynamically selected depending on the available signal to noise ratio of the chan-

Fig. 7 Camera, analog transmitter and circular polarised antenna

nel. Quadrature Modulation (QAM) is a modulation form where each constellation represents certain amplitude and phase of the radio signal. At 64QAM one symbol equals 6 bits and yields a high bitrate per subcarrier, but this constellation can be used only under good signal to noise ratio. Under deteriorating channel conditions the modulation steps down to 16QAM, and continues to simpler constellations if the channel worsens. For very bad conditions Binary Phase Shift Keying (BPSK) with only two constellations is employed, where each constellation represents a phase shift.
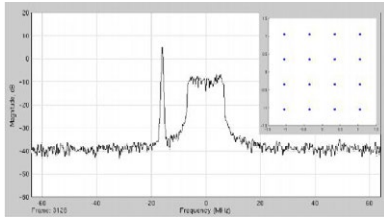
Fig. 9 Interference outside the channel
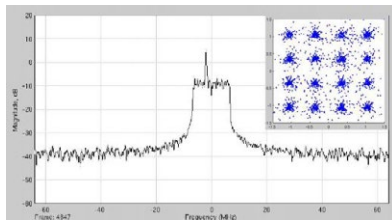
Fig. 10 Interference inside the channel

Mode: Normal (HD 1024x576 30fps)
TX Pwr: 100mW
USB RX ID 4886 RSSI: -70dBm Packets/s: 1293
USB RX ID 2478 RSSI: -72dBm Packets/s: 1289
TX -> RX:
>600m

Fig. 11
Preserved video during
reduced signal strength

Mode: Failsafe (320x192 10fps)
TX Pwr: 100mW
USB RX ID 4886 RSSI: -86dBm Packets/s: 250
USB RX ID 2478 RSSI: -90dBm Packets/s: 85
TX -> RX:
>1700m

With one bit per symbol, it gives a very modest bitrate, but is correspondingly robust for noisy conditions. In this state the video resolution and framerate is just sufficient to fly safely for a short duration. Figure 9 and 10 shows a simulation of how interference outside and inside the channel affects a 16QAM constellation.

Unlike WiFi radiolinks, a low latency FPV link is cannot be dependent on acknowledge for every data frame, and does not support re-sending of failed data frames. Instead Forward Error Correction coding (FEC) is employed, to handle most of the occurring bit errors. The remaining few failed frames will be part of the displayed video stream. This can be observed as occational macroblocking in the image, but will not affect the image quality much if the adaptive modulation reacts quickly during worsening radio conditions. The pictures in figure 11 shows screenshots taken during a drive-test of a digital wireless HD video link. It is a good illustration of an adaptive modulation scheme operating.

FUTURE DEVELOPMENT
Wireless video for FPV drone piloting is still immature technology, and we will see compact and low-cost HD FPV systems emerge in the near future. The key to lowered cost is increased integration of systems on chip and a resulting high volume. Paradigm shifts will occur when totally new radio, camera, or display concepts emerges. The next generation of cellular and WiFi technology, termed 5G will exploit dynamic beamforming to increase system gain and keep interference low. Together with more sophisticated MIMO this will increase perfor-

mance and transmission bandwidth further. It is likely that these concepts will be applied for future FPV systems when the technology matures. This will result in higher performance with extended range, higher image quality and better reliability. It will enable drones to handle more of our present challenges, and those we have not come to think of yet.
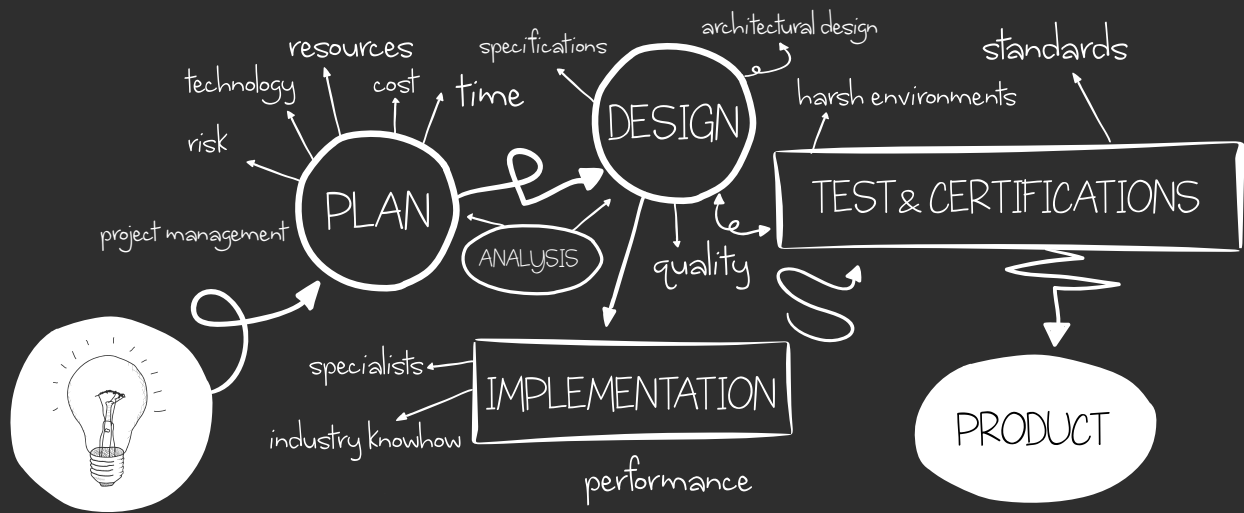
We make the technology you need

Hire a skilled specialist
Is there an expertise missing in your project? Our skilled development engineers have all the relevant expertise within hardware, FPGA, embedded software, software and mechanical development as well as project management and testing.

...or a full project team?
Do you have a product idea you need designed? We can provide a complete competency platform during a development project with the kowledge from our R&D specialists as well as from our product experts and OEM engineers.

Industry experience
Our engineers specialise in understanding the environmental challenges and demands of our customers products on top of being best-in-class within their technical disciplines.

Close to the customer
Being close to the customer and the projects we collaborate on means greater flexibility and speed, which benefits our customers. We have a wide network of 500 engineers in Norway, Sweden, Denmark and Germany.

ON TIME, ON COST, WITH THE CORRECT FUNCTIONALITY AND THE RIGHT QUALITY

- ▶ SOFTWARE DEVELOPMENT
- ▶ SYSTEM DESIGN
- ▶ EMBEDDED SOFTWARE DEVELOPMENT
- ▶ INTERACTION DESIGN

- ▶ PROJECT MANAGEMENT
- ▶ ELECTRONIC & HW DEVELOPMENT
- ▶ MECHANICAL DESIGN
- ▶ TEST & QUALITY

datarespons.com/ r-and-d-services/

data respons
EMBEDDED SOLUTIONS

INSIDE WRITERS

HALDOR HUSBY
Principal Development Engineer
Data Respons

MASc Electrical Engineering, University of Toronto
Siv.Ing. Electronics, Norwegian University of Science and Technology

GUEST WRITER

RUNE VOLDEN
R&D Manager
Ulstein Power & Control AS

BJØRN BERGERSEN
Specialist Development Engineer
Data Respons

MEng Electronics and Radio Communication, University of Bristol

THOMAS FREDRIKSEN
Development Engineer
Data Respons

MSc. Mathematics, University of Edinburgh

HENNING SJURSETH
Senior Development Engineer
Data Respons

MSc Computer Science, University of Minnesota

PREBEN MYRVOLL
Principal Development Engineer
Data Respons

MSc Computer Science University of Boulder / NTNU

LIJIAO LIU
Senior Development Engineer
Data Respons

PhD, Microelectronics, University of Bergen
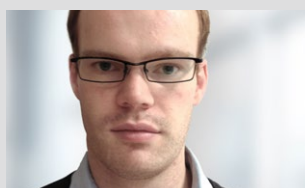
INGE TORSVIK
Development Engineer
Data Respons

MSc Microelectronics, University of Bergen.