Betatron tune measurement in the LHC damper using GPU

Frédéric Dubouchet

January 14, 2013

Abstract

This paper study a possible futur implementation of the betatron tune measurement in the Large Hadron Collider (LHC) at the European organization for nuclear research (CERN) using General Purpose Graphic Processing Unit (GPGPU) to analyse data from the damper acquisition. It start by describing the present hardware and the future possible implementations using the Accelerating Damper Transverse (ADT) acquisitions. The ADT data have to be processed to be able to extract the betatron tune. To get the tune the method used is to move the signal from temporal domain to frequency domain using Fast Fourier Transform (FFT) on Graphic Processing Units (GPUs).

Acknowledgements

I wish to thank CERN and Haute école du paysage, d'ingénieurie et d'architecture de Gemève (Hepia) to have made this master thesis possible. I also wish to thank Dr. Andy Butterworth and Dr. Ing. Erk Jensen who supported me on the choice to make this master thesis, Dr. Wolfgang Höfle for suggesting that I use GPU in this particular field and supervising. Dr. Valuch for the assistance on the damper and for all the ideas. Dr. Rama Calaga for the help on the mathematics of Singular Value Decomposition (SVD) and hints. Finaly, I wish to thank Pr. Paul Albuquerque who was my professor and supervisor during the whole thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 State of knowledge

In a particle accelerator, the charged particles circulate around the ring and oscillate due to the magnets and the accelerating structures. The accelerating structures, in the LHC supra-conducting cavities apply a strong electrical field that oscillates at the RF frequency (f_0) to particles in order to collect and accelerate particles in bunches inside a frequency bucket.

The particles inside a bucket are oscillating longitudinally along the ring and transversally in the vertical and horizontal plane. The longitudinal oscillations are damped by the beam control system. But the transversal oscillations must be damped by a separate system : the ADT[7, 12].

One of the key parameters of the accelerator is the betatron tune. The betatron tune, Q, is the quotient of the betatron oscillation and the particle frequencies.

$$f_\beta = Q * f_0$$

This value allow us to check if the particle beam is stable and don't reach any dangerous instabilities.

1.2 Tune measurement in the LHC

In order to measure the betatron tune in an accelerator we use Beam Position Monitors (BPMs). These monitors are able to measure the position of the beam in the vacuum chamber.

In the present setup Beam Instrumentation (BI) group is using their diode-based base-band-tune (BBQ) [2] system to acquire the tune over a certain number of machine turn (256 to 128'000). This can work as a passive instrument or as an on demand system by exciting 12 bunches in the beam with the tune kicker (MKQA). ADT has also been used for tune measurement excitation[6].

In normal operation, as the ADT is active, it is difficult to have a good picture of the excited bunches and make a fine tune measurement : the oscillations created by the MKQA are damped by the ADT. There have been studies to disable the ADT for a certain number of bunches in order to get a better tune measurement[8], but this may not be sufficient.

1.3 Proposed system

The ADT also have BPMs and these can have per bunch measurement[11]. This could allow a much precise measurement. But due to the high among of data to be processed (estimated to 640 mega bytes per seconds for each BPM) dedicated hardware is needed to compute the correct tune[9].

In order to be able to apply direct correction to beam oscilliation the betatron tune has to be measured at a high frequency, this has been estimated by BI to be between 5 and 10[Hz], once every 100 to 200[ms].

During the 2012 normal operation of the LHC, data has be acquired using the ADT acquisition system and data processing techniques has been tried to asses the modification that will be needed in order to make a reliable betatron tune measurement at a reasonable rate[9].

The current VMEbus implementation has some serious issues in particular the bus is quite slow the data rate of the bus is around 40 megabits per seconds. The data need to either be processed on the acquisition board or to be off-loaded to another computer using the serial link available on the board[1].

1.3.1 DSP on VME board

Digital Signal Processors (DSPs) are able to compute FFTs at high rate and these are used already in the machine at different places to make high speed feedback loops. The question is : is it fast enough to compute all the FFTs needed, DSPs are two orders of magnitude slower than GPUs. We also would have to develop a completely new system in order to be able to use them, in fact we don't have DSPs in the present ADT. The cost of development and the complexity of the deployment should also be studied.

1.3.2 FPGA pre-processing on VME board

Like in the approach using DSPs on VME boards, the question of computing power is still unsolved. We already have in house experience and we already have a lot of Field-Programmable Gate Array (FPGA) installed in the ADT. But if we want to do it we will have to create a new card able to replace the existing one and to make the computation. This mean create a potential problem in the existing setup. The cost is also to be studied we have to develop a new card, test it and install it in the LHC. Also the number of card and FPGA is not well understood, it could be anywhere between 4 and 3000 (FIXME).

1.3.3 GPU off-board computing

This solution can be integrated easily in the present setup. The present acquisition cards already have a digital output and could be used to transfer the data in another crate that could do the computations. The GPUs are cheap (compare to the price of developing a new VMEbus card) and easily scalable. The GPU should have largely enough computing power to be able to make the FFTs. Another interesting aspect of this solution is the ability to test it using Central Processing Unit (CPU).

1.4 Problem definition

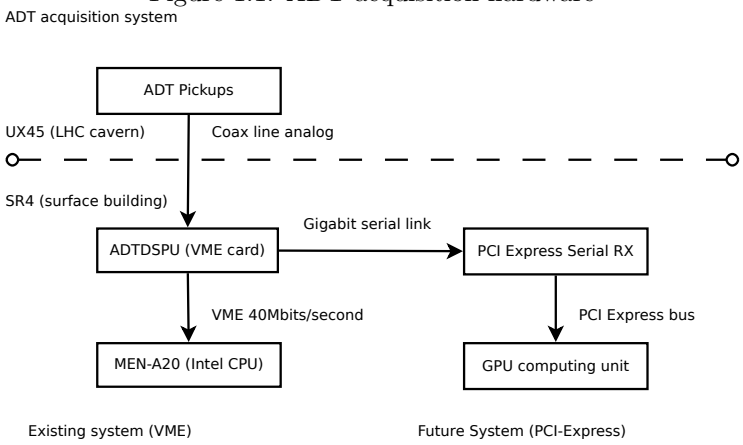
Show how to implement a GPU based system that can deliver a tune value for each beam and planes at a frequency that allow the system to be responsive enougth to allow tune correction to be applied automatically.

1.4.1 Algorithm

We need the tune frequency and we have the tune position per bunch, we have to calculate the FFT to move from time to frequency domain. Then we need to identify the tune in the transformation.

1.4.2 Hardware

Figure 1.1: ADT acquisition hardware



Per bunch position mesearment has to be avalable to the system for each beam and each plane. This should be provided from the ADTDSPU card and has to be transfered through a serial link to the CPU/GPU crate for computation.

We need a card in the CPU/GPU crate to unserialize the data and transfer them to the GPU memory. It may be possible to copy from the acquisition card directly to the GPU memory.

And finally fast enougth GPU to process the data. The number and the type of card should be looked at. The possibilty for expantion should be kept as the possibility to implement other algorithms.

1.4.3 Timing

Acording to BI we have to provide the tune measurement betweeen 5 Hz and 10 Hz. This mean that the transfer and the computation has to be made in less than 200 ms.

At a higher frequency because of the acquisition frequency (11 kHz) the precision may be unsufficient (NyquistShannon sampling theorem).

Chapter 2

Graphic Processing Unit

2.1 Introduction

General purpose programming on graphical processors is a new field with a growing community of practitioners. Until recently only proprietary interfaces existed to harness the power of these chips. With the arrival of the Open Computing Language (OpenCL) a new open interface has appeared, and with it a hope for a unified, simple and portable framework for general purpose computing on heterogeneous hardware.

2.2 Challenge

2.2.1 OpenCL a quick overview

The Open Computing Language (OpenCL) is the open standard for General Purpose Graphical Processing Unit programming (GPGPU programming). It is maintained by the Khronos group, and was initially proposed by Apple. Many companies of the industry are members of the OpenCL Working group: Altera, AMD, Apple, ARM, Broadcom, Codeplay, DMP, EA, Ericsson, Fixstars, Freescale, Hi corp, IBM, Intel, Imagination Technologies, Kestrel Institute, Kishonti, Los Alamos, Motorola, Movidius, Multicoreware, Nokia, NVIDIA, OpenEye, Presagis, Qualcomm, Rightware, Samsung, ST, Symbio, Texas Instruments, The University of West Australia, Vivante and Xilinx.

> OpenCL™is the first open, royalty-free standard for cross-platform, parallel programming of modern processors found in personal computers, servers and handheld/embedded devices. OpenCL (Open Computing Language) greatly improves speed and responsiveness for a wide spectrum of applications in numerous market categories from gaming and entertainment to scientific and medical software.
>
> The Khronos Group is a not for profit industry consortium creating open standards for the authoring and acceleration of parallel computing, graphics, dynamic media, computer vision and sensor processing on a wide variety of platforms and devices. All Khronos members are able to contribute to the development of Khronos API

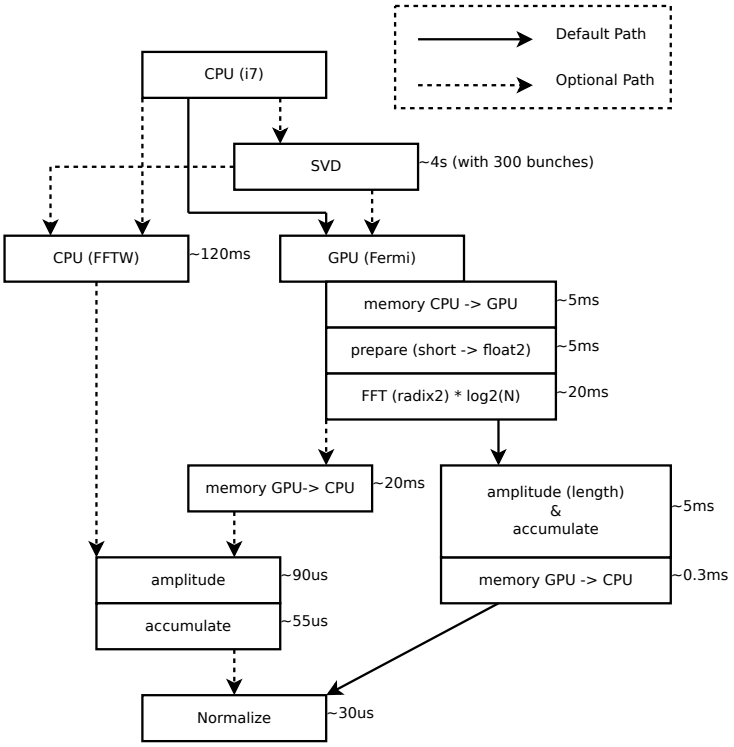
specifications, are empowered to vote at various stages before public deployment, and are able to accelerate the delivery of their cutting-edge 3D platforms and applications through early access to specification drafts and conformance tests.

Chapter 3

Results

3.1 Implemented System

Figure 3.1: Time flow with different implementations and with 3000 bunches of 2048 points each.



3.2 Notch filter

Just used to cut the low frequencies this doesn't change anything on the high frequencies and was just used to allow a better imaging on the spectogram, it

won't be used in the final version (quite slow because of sequencial).

3.3 FFT

Some words about FFTs.

3.3.1 FFTW

Some words that explain what is FFTW and why is was chosen as a reference

3.3.2 FFT with OpenCL on GPU

Some words on the implementation used for calculating FFT on OpenCL reference to the image of the spectrogram.

3.3.3 FFT with OpenCL on CPU

Some words on the fact that you can run the code on the CPU as well and then reference on the figure.

Also talk about the fact that there is less noise in the OpenCL CPU version than the OpenCL GPU version.

3.4 Amplitude

Amplitude calculation formula, explain the figures tell why it has to be done before accumulation and show this is very fast reference to the table of perf.

3.5 SVD

Problem is not direclty solvable with the number of bunch observed cite Hofle and Rama need a lot more bunches to make a good smooth[3].

Talk about the performances issues and cite the paper on SVD on GPU as a future implementation (could go to discution?)[10]

3.6 Performances

Calculation made by accumulation to simulate the number of bunches that could be present in the final version (2880).

3.6.1 Pipelining

Pipelining was tested and used in the process and it was possible to win around 15% in performances around it.

3.6.2 Memory

Copy of memory from and to the GPU discution.

3.6.3 Time

Add table with time performances and discution.

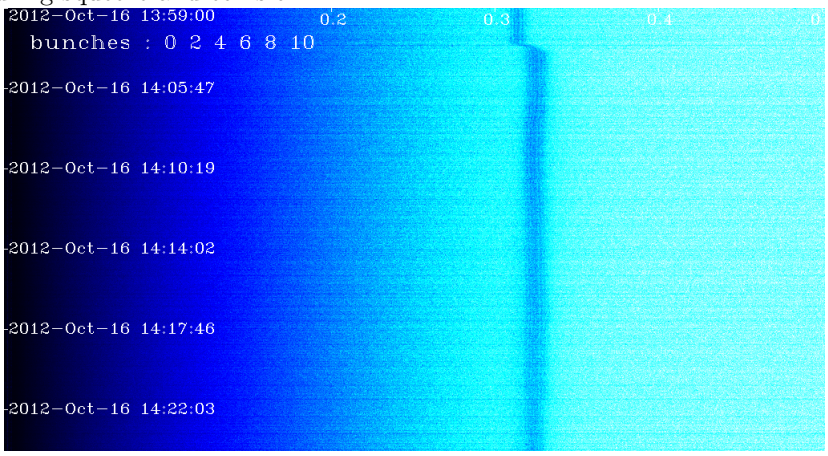
Table 3.1: Speed for 3000 batch of 2048 points

Device	Type	Threads	Speed [GHz]	pipeline	Time [ms]
Xeon X5650	FFTW	12	2.67	N/A	291
Xeon X5650	OpenCL	12	2.67	enable	284
Xeon X5650	OpenCL	12	2.67	disable	288
i7-3720QM	FFTW	8	2.6	N/A	310
i7-3720QM	OpenCL	8	2.6	enable	272
i7-3720QM	OpenCL	8	2.6	disable	273
Tesla M2090	OpenCL	512	1.3	enable	35
Tesla M2090	OpenCL	512	1.3	disable	37
GeForce 650M	OpenCL	384	0.9	enable	355
GeForce 650M	OpenCL	384	0.9	disable	365

3.7 Spectrogram

Some word definition of Spectrogram. Display some spectrogram.

Figure 3.2: Spectrogram with ADT off on the 16 October 2012 on vertical beam 1 during squeeze and collision



Chapter 4

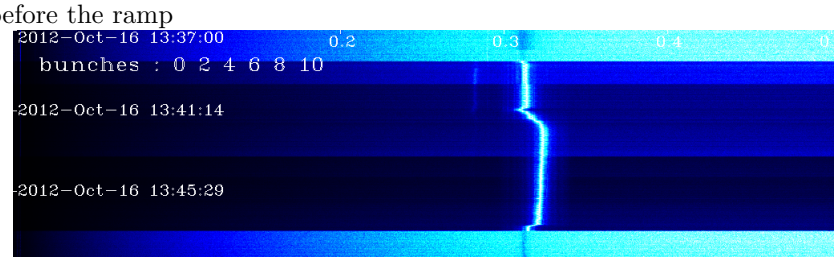
Discussion

4.1 Observation

4.1.1 Without damper

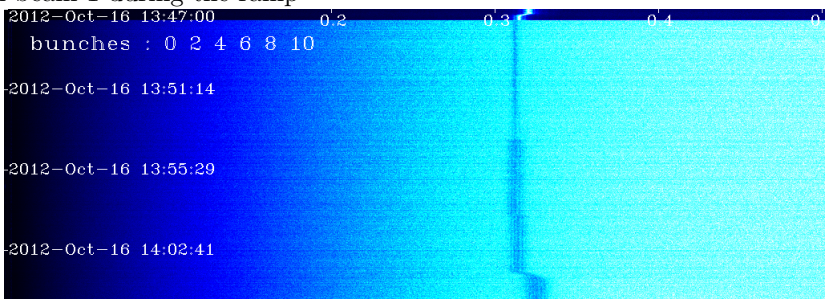
Clear view of the tune

Figure 4.1: Spectrogram with ADT off on the 16 October 2012 on vertical beam 1 before the ramp



4.1.2 With damper

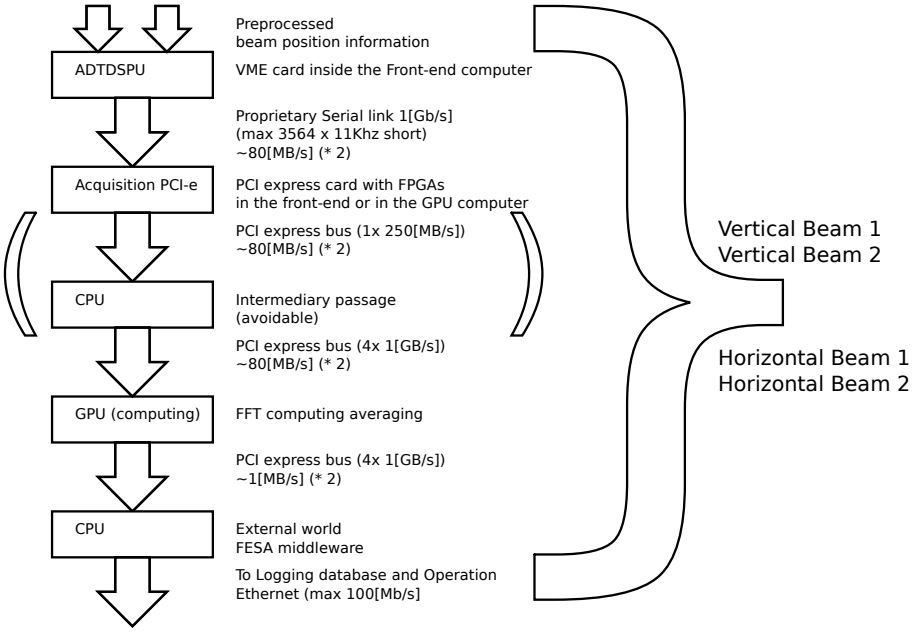
Figure 4.2: Spectrogram with damper working on the 16 Octover 2012 on vertical beam 1 during the ramp



The tune is inverted speak about cite the hofle paper mathematic morphology more processing needed but should be quite low.

4.2 Data flow

Figure 4.3: Acquisition data flow



4.3 Hardware

4.3.1 ADT Aquisition boards

4.3.2 Serial link interface

4.3.3 GPUs

4.4 Software

4.4.1 Drivers

4.4.2 OpenCL

4.4.3 Front-end

4.5 Estimated Cost

Chapter 5

Conclusion

This project looks like a nice place to try using GPUs in accelerators. The possibilities are promising and the gain for the stability of the LHC could allow more physic time. GPUs could prove to be useful and be used in other places in accelerators where computing power is needed.

Chapter 6

Experimental

6.1 Estimation of the amount of data

Presently the LHC is working with an interval of 50ns between bunches this correspond to a bunch every 10 buckets. But the Operation (OP) is planning to move to 25ns bunches spacing this would mean 5 buckets between bunches. With the RF frequency (f_0) we can compute the number of acquisitions per seconds.

$$for\ 50\ ns\ : \frac{400.789M}{20} = 20'039'450 \leq 2^{25}$$

$$for\ 25\ ns\ : \frac{400.789M}{10} = 40'078'900 \leq 2^{26}$$

This represent the amount of data for one pickup (BPM), in the case of ADT we have two of them per beam and per plane so as the LHC has two rings and for each ring there are two transversal plane and there are two pickups per plane. This means we still have to multiply this value by eight.

$$for\ 50\ ns\ : 2^{25} * 8 = 2^{28}$$

$$for\ 25\ ns\ : 2^{26} * 8 = 2^{29}$$

As FFTs on GPUs start to be be faster than CPUs around 2^{15} acquisitions it seems interesting to study this kind of system to compute the betatron tune.

6.2 Measurement with the ADT

In order to check the feasibility of the system and to have a good prototype the first test will be to excite some of the bunches and acquire the betatron tune using the ADT during the end of 2012 run[4].

A piece of software has been developed that will acquire the bunch by bunch acquisition and compute various algorithm on the data using the CPU and the FFTW library in the CERN infrastructure using Control (CO) group control system and the OP group infrastructure.

6.3 Experimental Set-up

6.3.1 Hardware

The experimental set up is not presently able to acquire more than a certain number of bunches due to memory limitation 16k and interupt frequency so during the Machine Developments (MDs) only 6 bunches were acquired by bunch by planes.

6.3.2 Software

6.4 FFT

Used the algorithm described here[5].

6.5 SVD

Used the GNU scientific library.

6.6 Machine development sessions

Using the ADT BPMs we acquired data in the machine during 3 independant MDs. Most of the data taking was done in parallel to other normal LHC operation or during ADT dedicated MD time.

6.6.1 First session

Night session of the 11 october 2012.

6.6.2 Second session

Parasitic session of the 16 october 2012

6.6.3 Third session

Ramp acquisition of the 14 november 2012

Glossary

betatron tune the betatron tune is the frequency of the oscillations of the bunches divided by the RF frequency (f_0). 1, 7, 15

bucket at every Radio Frequency (RF) period in the RF frequency (f_0) there is a bucket, in each of these bucket a particles bunch can potentially be stored in the ring. 6, 15

bunch particules trapped inside an RF bucket circulating in the machine. 6, 15, 17

cavity RF structure made to accelerate the particles, it uses a high power radio frequency into a resonating structure to increase the energy of the particles. 6, 17

damper machine in an accelerator that damp the transverse oscillation of the beam by applying a transverse electric field. 1, 18

FFTW is a C subroutine library for computing the discrete Fourier transform (DFT) in one or more dimensions, of arbitrary input size, and of both real and complex data (as well as of even/odd data, i.e. the discrete cosine/sine transforms or DCT/DST). 15

kicker machine in an accelerator that can kick the beam transversally, used to kick the beam in or out (injection or extraction kicker) of the beam pipe but also in our case excite the beam transversally. 6, 18

RF frequency (f_0) the base frequency of the RF in the cavities in the case of the LHC this frequency is 400.789MHz, this frequency dictate the number of bucket that the machine can have. 6, 15, 17

VMEbus a computer bus standard widespread at CERN, in the case of the LHC RF the bus has a larger board and some of the pins are used to route custom signals between cards. 7

Acronyms

ADT Accelerating Damper Transverse. 1, 6, 7, 15, 16

BBQ diode-based base-band-tune. 6

BI Beam Instrumentation. 6–8

BPM Beam Position Monitor. 6, 7, 15, 16

CERN European organization for nuclear research. 1, 15

CO Control. 15

CPU Central Processing Unit. 7, 15

DSP Digital Signal Processor. 7

FFT Fast Fourier Transform. 1, 7, 15

FPGA Field-Programmable Gate Array. 7

GPGPU General Purpose Graphic Processing Unit. 1

GPU Graphic Processing Unit. 1, 7, 14, 15

Hepia Haute école du paysage, d'ingénieurie et d'architecture de Gemève. 1

LHC Large Hadron Collider. 1, 6, 7, 14, 15

MD Machine Development. 16

MKQA tune kicker. 6

OP Operation. 15

RF Radio Frequency. 17

SVD Singular Value Decomposition. 1

Bibliography

[1] P Baudrenghien, Wolfgang Höfle, G Kotzian, and V Rossi. Digital signal processing for the multi-bunch lhc transverse feedback system. oai:cds.cern.ch:1124094. *LHC-PROJECT-Report*, 1151:4, Sep 2008.

[2] A Boccardi, M Gasior, R Jones, P Karlsson, and RJ Steinhagen. First results from the lhc bbq tune and chromaticity systems. Technical Report LHC-Performance-Note-007. CERN-LHC-Performance-Note-007, CERN, Geneva, Jan 2009.

[3] Rama Calaga. *Linear Beam Dynamics and Ampere Class Superconducting RF Vavities at RHIC*. PhD thesis, Stony Brook University, 2006.

[4] F Dubouchet, W Höfle, G Kotzian, and D Valuch. what you get transverse damper system (adt). In *Evian 2012 proceedings*, 2012. `https://indico.cern.ch/`.

[5] Naga K. Govindaraju and Dinesh Manocha. Cache-efficient numerical algorithms using graphics hardware, 2007.

[6] Wolfgang Höfle. Lhc transverse damper observations versus expectations. In *Evian 2010 proceedings*, 2010. `https://espace.cern.ch/acc-tec-sector/Evian/Papers-Dec2010/3.3_WH.pdf`.

[7] Wolfgang Höfle. Lhc transverse damper from commissioning to routine. *Beams Department newsletter*, 1:6–7, 2011. `https://espace.cern.ch/be-dep/BEDepartmentalDocuments/BE/BE_Newsletter/BE_Newsletter_001.pdf`.

[8] Wolfgang Höfle. Transverse feedback : high intensity operation, abort gap cleaning, injection gap cleaning and lessons for 2012. In *Evian 2011 proceedings*, 2011. `https://indico.cern.ch/`.

[9] Wolfgang Höfle. Transverse damper. In *Chamonix 2012 proceedings*, 2012. `https://indico.cern.ch/`.

[10] S. Lahabar and P.J. Narayanan. Singular value decomposition on gpu using cuda. In *Parallel Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on*, pages 1 –10, may 2009.

[11] Daniel Valuch. Beam phase measurement and transverse position measurement module for the lhc, 2007. Poster from the Low Level RF workshop 2007. Available "`https://edms.cern.ch/document/929563/1`.

[12] V M Zhabitsky, E V Gorbachev, N I Lebedev, A A Makarov, N V Pilyar, S V Rabtsun, R A Smolkov, P Baudrenghien, Wolfgang Höfle, F Killing, I Kojevnikov, G Kotzian, R Louwerse, E Montesinos, V Rossi, M Schokker, E Thepenier, and D Valuch. Lhc transverse feedback system: First results of commissionning. oai:cds.cern.ch:1141925. Technical Report LHC-PROJECT-Report-1165. CERN-LHC-PROJECT-Report-1165, CERN, Geneva, Sep 2008.