Audio Processing on the BeagleBone Black's Programmable Real-time Unit

Bachelor Semester Project - Fall 2017

Loïc Droz

Laboratory : Audiovisual Communications
Supervisors : Eric Bezzam, Robin Scheibler

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Contents

Chapter 1

Introduction

Microphone arrays have become increasingly popular due to devices such as Amazon Echo, Google Home, etc. These devices make use of microphone arrays to perform algorithms such as beamforming and direction-of-arrival. The LCAV lab at EPFL has already developed such algorithms[1] and a browser-based interface[2] for audio processing with an embedded platform - the BeagleBone Black (BBB). The goal of this project is to implement such a microphone array on the BBB, taking advantage of a unique microcontroller available on the BBB - the Programmable Real-time Unit SubSystem (PRUSS).

The project consists of three main components :

1. Hardware : the BBB and the Kurodako cape[3], containing the microphone array.
2. Firmware : the program running on the PRUSS which handles the audio capture and processing.
3. Software : a C interface (API), running on the BBB's ARM CPU.

For now, the API is very simple. It allows the user to read the processed data from the PRU to a user-supplied buffer, specify the quantity of data needed and the number of channels to extract from it. It also allows the user to pause and resume the recording of the data on the API side in order to prevent overflows.

Running the core audio processing code on the PRUSS instead of the main ARM CPU has three advantages :

1. Lower latency due to the PRUSS' predictable timings.
2. Any program running on the PRUSS is not subject to OS-scheduling like a typical Linux process running on the ARM CPU.
3. Offloading the ARM CPU from such an intensive task also prevents our library from having a significant impact on the global performance of the host when it is used.

The code and documentation can be found at the link below:

<div align="center">

https://github.com/Scrashdown/PRU-Audio-Processing

</div>

---

[1]LCAV room acoustics and microphone array algorithms repository : https://github.com/LCAV/pyroomacoustics
[2]Browser-based interface from a past semester project : https://github.com/LCAV/easy-dsp
[3]Cape designed for the BeagleBone Black by Robin Scheibler: https://github.com/fakufaku/kurodako

Chapter 2

Background Knowledge

2.1 PRU / PRUSS

The PRUSS is a module of the ARM CPU used on the BeagleBone Black. It stands for PRU SubSystem, where PRU stands for Programmable Real-time Unit. The PRUSS contains 2 PRUs which are essentially very small and simple 32-bit microprocessors running at 200 MHz and use a custom instruction set. Each PRU has a constant 200 MHz clock rate, 8 kB of instruction memory, 8 kB of data memory, along with 12 kB of data memory shared between the 2 PRUs. They can be programmed either in assembly using the `pasm` assembler or in C using the `clpru` and `lnkpru` tools.

Each PRU has 32, 32-bit registers, where R30 is used for interacting with the PRU's output pins and R31 is used for reading inputs from these pins and triggering interrupts by writing to it. Both PRUs also share 3 registers banks, also called *scratchpads*, which each contain 30 additional registers. Registers can be transferred between these banks and a PRU in one cycle by using specialized assembly instructions (`XIN` / `XOUT` / `XCHG`, commonly called `XFR` instructions in the PRU Reference Manual[1]). Furthermore, one PRU can also access the registers of the other PRU using the same assembly instructions.

The PRUs are designed to be as time-deterministic as possible. Moreover, nearly all instructions will execute in a one clock cycle, or 5 ns at the 200 MHz clock rate. One significant set of commands that are non-time deterministic are the memory transfer instructions which may vary in execution time. More information on each possible command can be found in the PRU Reference Manual.

The PRUSS also contains an interrupt controller which allows the PRU to send and receive interrupts to and from the ARM CPU. It can be configured either from the PRUs themselves by changing the values of the configuration registers, or from the ARM CPU using the `prussdrv` library along with the `uio_pruss` driver (more information on that in Section 3.1.1).

Using the PRUSS requires a driver. Currently, there are two choices available : `uio_pruss` (along with the `prussdrv` library) and the newer `pru_rproc`. `uio_pruss` provides a lower-level interface than `pru_rproc`. `pru_rproc` provides a C library for message passing between the PRU and the ARM CPU which makes programming simpler than with `uio_pruss`. However, the current lack of online examples for using `pru_rproc`, along with performance issues encountered while investigating it for this project, made us choose `uio_pruss` instead.

Nevertheless, it seems `uio_pruss` is currently being phased out of support by Texas Instruments in favor of `pru_rproc`. It may be feasible in the future to convert the code to use `pru_rproc`. However, as we are going to see further in the report, the timing requirements in the PRU processing code are very tight, even using assembly. Whether it would be possible to meet them using C and `pru_rproc` has yet to be investigated.

---

[1]http://processors.wiki.ti.com/index.php/PRU_Assembly_Instructions

2.2 Audio Processing

The audio processing code currently handles six microphones at a fixed output sample rate, using one of the 2 PRUs present on the board. The input signals come from six Knowles SPM1437HM4H-B microphones[2] connected to the board via a cape designed especially for this project.[3]

The SPM1437HM4H-B microphones used for this project produce a *PDM* (Pulse Density Modulation) output. This is a 1-bit wide output at a very high sample rate (> 1 MHz).[4] Moreover, this signal needs to be converted to a lower-rate PCM (Pulse-Code Modulation) signal, which is much more commonly used for processing and storing audio data.
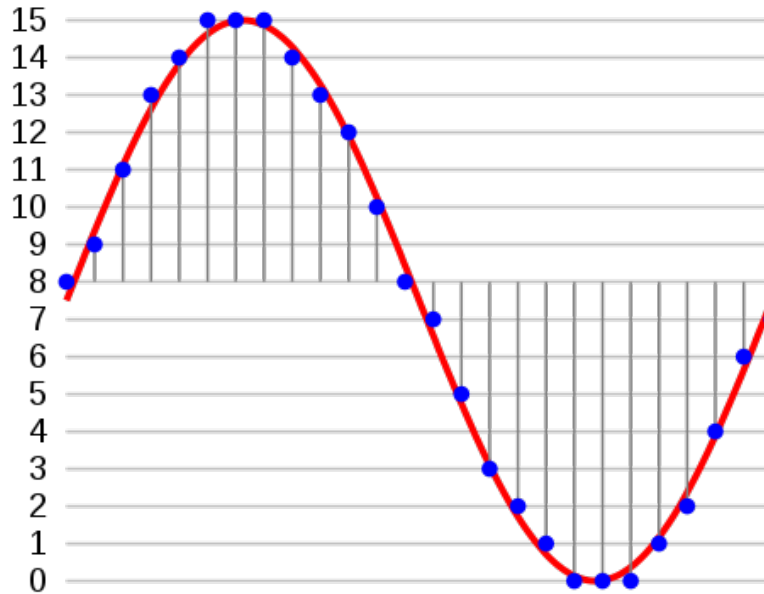

Figure 2.1: Illustration of a PCM signal (CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=635225).
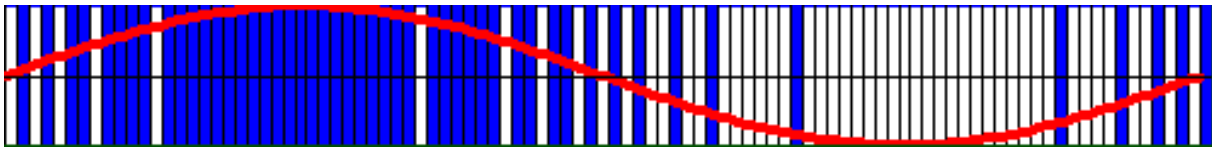
Figure 2.2: Illustration of a PDM signal (https://commons.wikimedia.org/wiki/File:Pulse-density_modulation_1_period.gif).

In a PCM signal, each value represents its amplitude on a fixed scale at a fixed time. However, in a PDM signal, its amplitude at a given time is represented by the density of 1's relative to 0's in a time interval. Converting a PDM signal to a PCM signal therefore requires using some form of a moving-average filter as will be presented in the following subsection.

2.2.1 CIC Filter

Because we want to run this filter on the PRU with rather tight timing constraints, we chose to implement a Cascaded Integrator-Comb (CIC) filter. There are two general types of CIC filters: the decimation filters, which is the type we are using in this project, and the interpolation filters. From now on we will refer to CIC decimator filters simply as CIC filters.

The CIC filter is essentially an efficient implementation of a moving-average filter by using only additions and subtractions and has a finite-length impulse response (FIR). Although the PRU is capable of performing unsigned integer multiplications by using its Multiply and Accumulate Unit, they take several more cycles to execute than the one-cycle instructions used for regular additions and subtractions. Since our goal is to handle several channels at once with very tight timing constraints, computational savings really matter.

Although cheap to compute, the CIC filter does have an unfortunate drawback. Its frequency response is far from the ideal flat response with a desirable sharp cutoff (see Figure 2.3 for an example power response). To get a sharper cutoff, it is often necessary to cascade another filter, commonly called a *compensation* filter, to the CIC output.[5] Nevertheless, since a CIC filter's output is usually at a much

---

[2] https://media.digikey.com/pdf/Data%20Sheets/Knowles%20Acoustics%20PDFs/SPM1437HM4H-B.pdf
[3] https://github.com/fakufaku/kurodako
[4] More information on PDM and its popularity for portable devices can be found here: http://users.ece.utexas.edu/~bevans/courses/rtdsp/lectures/10_Data_Conversion/AP_Understanding_PDM_Digital_Audio.pdf
[5] For more on CIC compensation filters, see here: https://www.altera.com/content/dam/altera-www/global/en_US/pdfs/literature/an/an455.pdf
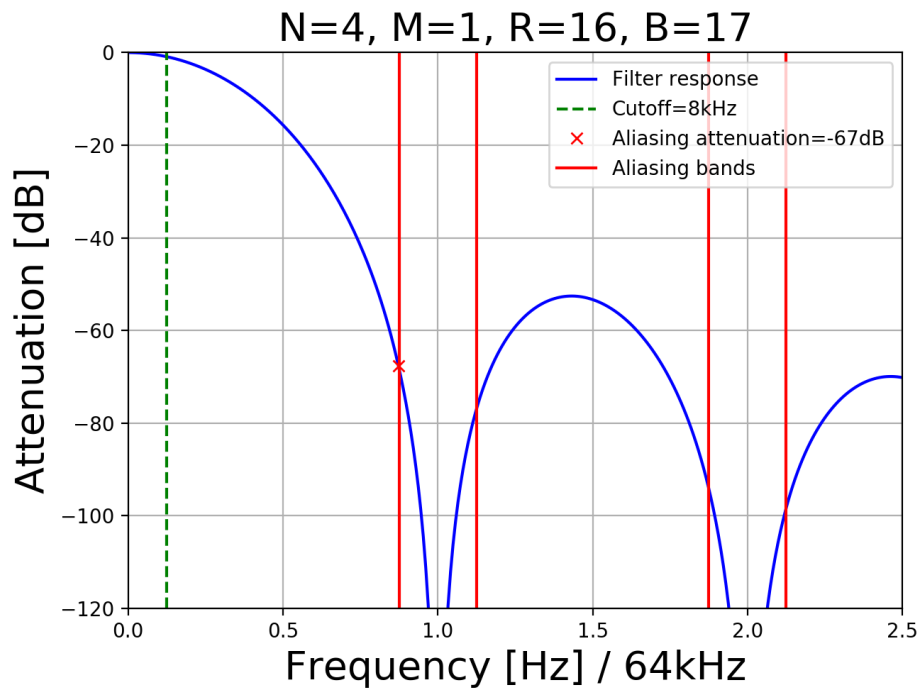

Figure 2.3: Power response of the CIC filter for a particular set of parameters (own work, generated from our script `utils/power_resp_CIC_folds.py` in the project files). The green line for the *cutoff frequency* is the upper limit of the frequencies we are interested in (8 kHz to capture human speech reasonably well). The frequencies between the red lines will be aliased into our frequency range of interest $[-8$ kHz, 8 kHz$]$. By looking at the intersection between the first dip of the frequency response and the aliasing bands, we can estimate the minimum aliasing attenuation achieved by a particular set of parameters. Refer to Table II in Hogenauer's paper for the aliasing attenuation that can be achieved for a variety of parameters: N, M, and the bandwidth relative to the CIC output sampling frequency (f_s/R).

lower rate (64 kHz in our implementation) than its input ($\approx$1.028 MHz), applying this filter after the CIC filter will require much less computational resources than applying an expensive filter with a sharper response on the raw, very high rate input signal from the microphones.

In general, CIC filters are well-suited as a first filtering step to efficiently downsample the signal and remove some of the higher frequencies. The reduced sample rate of the output reduces the computational power needed for additional FIR compensation filters, and therefore compensates their higher complexity.

Now let's dive into more detail about the CIC filter. The filter has 3 parameters : N, M, and R. It is made of N cascaded integrator stages, followed by a decimator of rate R, and then N cascaded comb stages, where M is the delay of the samples in the comb stages. In our case, it takes the PDM signal as input and outputs a PCM signal. If the input sample rate is `f_s`, the output sample rate will be `f_s / R`. A CIC interpolation filter only differs in that the comb stages come first, followed by the interpolator of rate R, and finally the integrators.
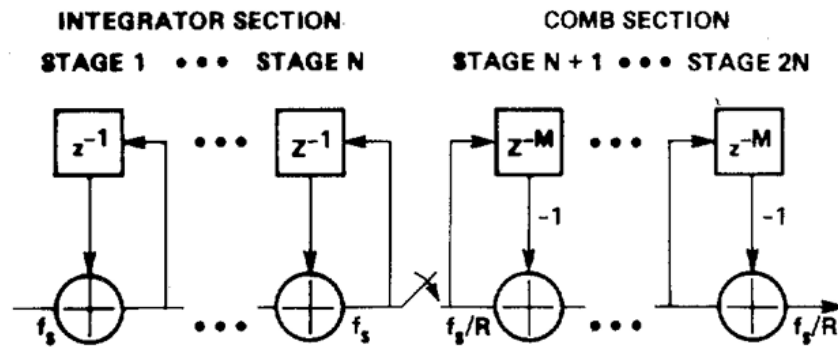

Fig. 1. CIC decimation filter.

Figure 2.4: Structure of a CIC decimation filter, obtained from Hogenauer's paper about CIC filters.

The filter's resource usage depends on its parameters, the platform on which it is implemented and how it is implemented. A more detailed explanation will follow in the implementation section of this report - Section 3. However, by considering only the theoretical structure of the filter, we can already deduce some general rules :

- Memory usage is approximately proportional to N and M : The filter has N integrator stages and N comb stages of which we need to store the values, therefore memory usage is approximately proportional to N. Also, since M is the delay of the samples in the comb stages, for each comb stage it is necessary to store the previous M samples, therefore memory usage is approximately proportional to M.
- Computational resource usage is inversely correlated to R : Since the comb stages are preceded by a decimator of rate R, the comb stages need to be updated R times less often than the integrator stages. Therefore, as R increases, the total computational power is reduced since the comb stages are not used as often. However, the reduction cannot be arbitrarily high, because the integrator stages always need to be updated at the very high input sample rate, independently of R. The processing power needed for these stages, and any other overhead added by the implementation, constitute therefore a lower-bound of the total processing power required to run the filter.

Finally, it is important to note that overflows will occur in the integrator stages due to unity feedback. However, the output of the filter will still be correct if each stage follows modular arithmetic rules[6] and if each stage has a bit width large enough that it can support the maximum possible value at the output of

---

[6]PRU registers follow modular arithmetic since they can overflow and loop back to zero. The same can be done if using the most-significant 16 bits and the least-significant 16 bits of a single PRU register as separate "registers".

the filter. To ensure the second condition is met, the following formula is used to calculate the required bit width for each stage of the filter (described in Hogenauer's paper) :

$$B_{out} = \lceil Nlog_2(RM) + B_{in} \rceil \tag{2.1}$$

where B_{in} is the bit width of the filter's input (in our case 1 bit), M, N and R are the filter's parameters, and B_{out} is the required bit width for all stages of the filter.

Chapter 3

Implementation

3.1 Getting Started

First of all, make sure you have the required hardware: the BeagleBone Black, an SD card, and the Kurodako Board. Flash the board with the latest "IoT" Debian image following these instructions[1].
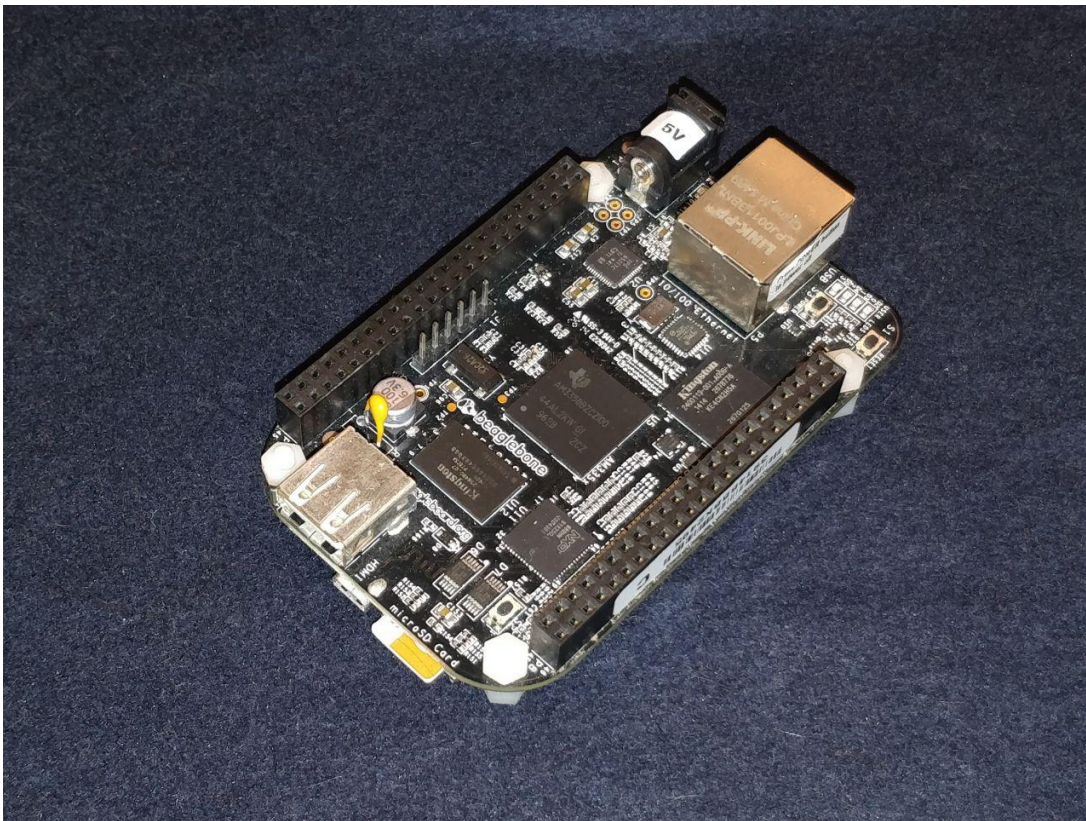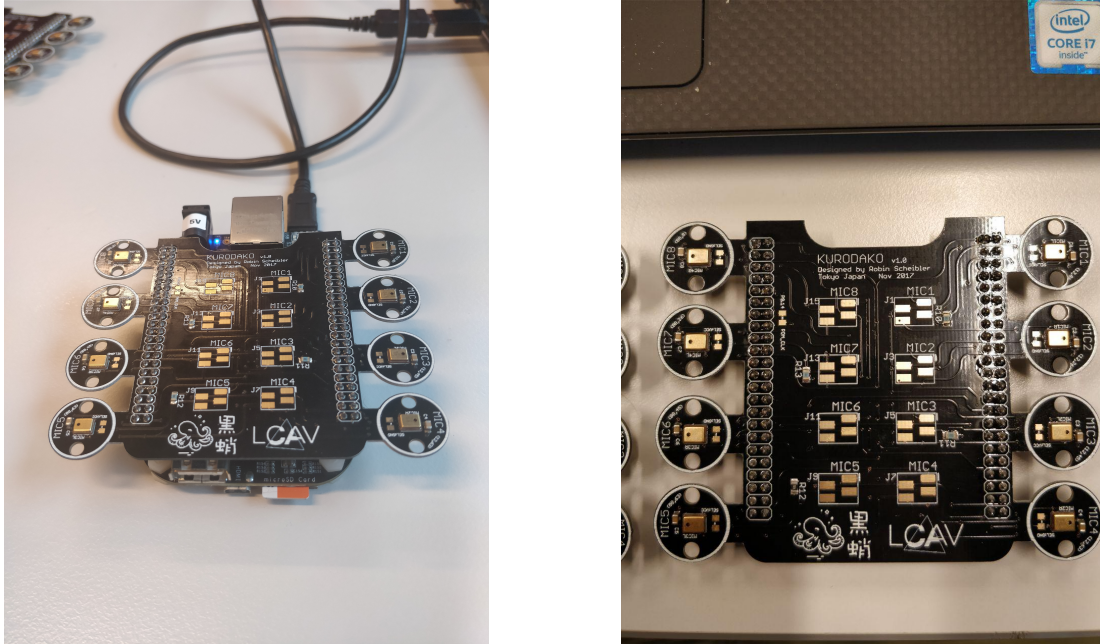

Figure 3.1: The BeagleBone Black, with an SD card (own work).

---

[1] https://beagleboard.org/getting-started

(a) The BBB with the Kurodako Board, ready to be used! (b) The Kurodako Board. Although the program currently supports six channels, the cape has eight mics.

Figure 3.2: Images of Kuradako Board (own work).

3.1.1 Configure uio_pruss and free the BBB GPIO pins for the PRU

Note: this was tested on the following kernel: `Linux beaglebone 4.4.91-ti-r133 #1 SMP Tue Oct 10 05:18:08 UTC 2017 armv7l GNU/Linux`

In order to run the filter, we need to be able to use the input and output pins from the PRUs so that it is possible to read data from the microphones. The PRU's I/O pins can be connected to the BBB's GPIO pins.[2][3][4] However, by default, some BBB pins cannot be reassigned to anything else. To correct this, we need to load a cape.[5] To do so, open the `/boot/uEnv.txt` file on the board (backup it first!) and do the following modifications, then reboot :

Add the following line :

`cape_enable=bone_capemgr.enable_partno=cape-universala`

And comment the following line :

`enable_uboot_cape_universal=1`

You should now be able to multiplex a pin to the PRUs using the `config-pin` command. We still have to enable the UIO driver, which is disabled in favor of `pru_rproc` on this kernel. To do so, still in the `/boot/uEnv.txt` file, comment the following line :

`uboot_overlay_pru=/lib/firmware/AM335X-PRU-RPROC-4-4-TI-00A0.dtbo`

And uncomment this one :

---

[2]Image of the BBB's GPIO pins : http://beagleboard.org/static/images/cape-headers.png
[3]P8 Header Table: https://itbrainpower.net/a-gsm/images/BeagleboneBlackP8HeaderTable.pdf
[4]P9 Header Table: https://itbrainpower.net/a-gsm/images/BeagleboneBlackP9HeaderTable.pdf
[5]https://elinux.org/Capemgr

```
uboot_overlay_pru=/lib/firmware/AM335X-PRU-UIO-00A0.dtbo
```

Then from the Terminal :

```
$ cd /opt/source/dtb-4.4-ti
$ sudo nano src/arm/am335x-boneblack.dts
```

Comment this line :

```
#include "am33xx-pruss-rproc.dtsi"
```

And uncomment this one :

```
#include "am33xx-pruss-uio.dtsi"
```

Then close nano and run the following commands :

```
$ make
$ sudo make install
```

Finally, run `sudo nano /etc/modprobe.d/pruss-blacklist.conf` and add the following lines :

```
blacklist pruss
blacklist pruss_intc
blacklist pru-rproc
```

Now reboot the board, and you should be able to run commands such as `config-pin -q P8.45` (to query the status of a pin) without trouble.

3.1.2 Install PRUSS Driver Library (`prussdrv`) and PRU Assembler (`pasm`)

In order to install the PRUSS driver on the host side, first clone this repo[6]. Then `cd` into the cloned repository and run the following commands :

```
$ make
$ sudo make install
$ sudo ldconfig
```

If everything went well, the `prussdrv` library and the `pasm` assembler should be installed on your board and ready to be used.

3.1.3 Plug the Kurodako board and write some code!

The code for the CIC filter and the interface can be found in the project files at `src/6Mic-CIC/`. To write and run code like the one below, (in `src/6Mic-CIC/`) write the code to a file called `main.c` in `host/`. Then run the following commands from the terminal (while in `src/6Mic-CIC/`):

```
$ sh deploy.sh      # load PRU code and build main program
$ cd gen
$ ./main            # run main program
```

---

[6]https://github.com/beagleboard/am335x_pru_package

```c
/**
 * @brief Example program for using the interface.
 *
 * @author Loïc Droz <lk.droz@gmail.com>
 *
 */

#include <stdio.h>
#include <stdlib.h>
#include <time.h>
#include "interface.h"

#define OUTFILE "../output/interface.pcm"
#define NSAMPLES 64000 * 1
#define NCHANNELS 6

int main(void) {
    printf("\nStarting testing program!\n");

    printf("Open output PCM file...\n");
    FILE * outfile = fopen(OUTFILE, "w");
    if (outfile == NULL) {
        fprintf(stderr, "Error: Could not open output PCM file.\n");
        return 1;
    }

    void * tmp_buffer = calloc(NSAMPLES, NCHANNELS * SAMPLE_SIZE_BYTES);
    if (tmp_buffer == NULL) {
        fprintf(stderr, "Error: Could not allocate enough memory for the testing buffer.\
n");
        fclose(outfile);
        return 1;
    }

    printf("Initialize PRU processing...\n");
    pcm_t * pcm = pru_processing_init();
    if (pcm == NULL) {
        free(tmp_buffer);
        fclose(outfile);
        return 1;
    }

    struct timespec delay = { 0, 250000000 };  // Wait 250 ms
    const size_t limit = 35;
    enable_recording();
        nanosleep(&delay, NULL);
        for (size_t i = 0; i < limit; ++i) {
            nanosleep(&delay, NULL);
            size_t read = pcm_read(pcm, tmp_buffer, 16500, NCHANNELS);
            fwrite(tmp_buffer, NCHANNELS * SAMPLE_SIZE_BYTES, read, outfile);
            printf("Buffer size : %zu, max = %zu\n", pcm_buffer_length(),
    pcm_buffer_maxlength());
            printf("Read : %zu/%zu\n", i, limit);
        }
    disable_recording();

    printf("Closing PRU processing...\n");
    fclose(outfile);
    free(tmp_buffer);
    return 0;
}
```

../src/6Mic–CIC/host/main.c
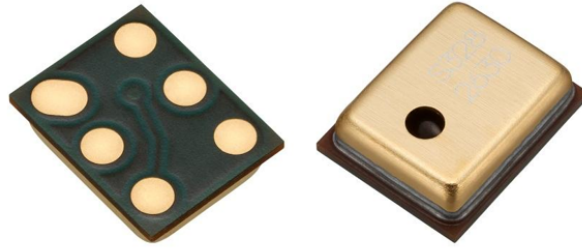
3.2 Microphones and wiring diagram

Figure 3.3: Knowles SPM1437HM4H-B microphones, obtained from the datasheet.

For this project, we are using the Knowles SPM1437HM4H-B microphones which output a PDM signal at a very high frequency (> 1 MHz), see the microphone's datasheet for more details.[7] Each microphone has six pins:

- 2 x **GROUND** (power) : Ground.
- **VDD** (power) : VDD.
- **CLK** (input) : The clock input, must be at a frequency > 1 MHz to wake up the microphone. Dictates the microphone's sample rate, `f_s = f_clk`.
- **DATA** (output) : The microphone's PDM output. Its sample rate equals the period of the CLK signal. The DATA signal is ready shortly after the rising or falling edge of the CLK. This delay is variable, but is between 18 ns and 125 ns. It is specified in the microphone's datasheet. as `t_dv`.
- **SELECT** (input) : Selects whether data is ready after rising or falling edge of CLK, (VDD => rising, GND => falling).

The BeagleBone Black already has pins for GND and VDD, we connect them directly to the corresponding pins on the microphones.

The CLK signal is generated using one of the BeagleBone's internal PWM which is output to one of the board's pins and is configured by a bash script to generate an appropriate CLK period for the microphones.[8] In our implementation, the PRUs also need to be able to poll the state of the CLK signal. In order to achieve this, the PWM (CLK) signal is also connected to some of the BeagleBone's pins, which are multiplexed to the PRU.

The DATA pin of the microphone is then connected to a pin of the board which is multiplexed to the PRU.

Since we are using six microphones, we could use six pins on the board for the microphone's DATA outputs, however it is possible to connect two microphones to a single board pin, thus tying two DATA lines together. To achieve this, we set one of the SELECT lines of the microphones to VDD and the other to GND. By doing this, one of the microphones will have data ready just after the rising edge of the input clock, while the other one will have data ready just after the falling edge. In order to avoid a short circuit between the two microphones connected to the same pin, it is necessary to place a resistor between one of the microphones (for example, between all rising edge microphones) and the BeagleBone Black pin. This method has the advantage of using *three* pins on the BBB instead of *six*.

Doing this has a drawback however, for each round of processing (processing all the channels) we need to wait for `t_dv` twice instead of only once with the 'simple' solution using six pins. This is because in the three-pin case, to retrieve data from all microphones, we have to wait for `t_dv` after the rising

---

[7]https://media.digikey.com/pdf/Data%20Sheets/Knowles%20Acoustics%20PDFs/SPM1437HM4H-B.pdf
[8]At the top of `src/6Mic-CIC/deploy.sh`.

edge *and* after the falling edge of the clock. According to the microphone's datasheet, `t_dv` can go up to 125 ns, which is 25 cycles of the PRU at its 200 MHz clock rate. Although not a huge advantage in performance it is still significant. However, the current Kurodako board restrict us to using three pins for six microphones.
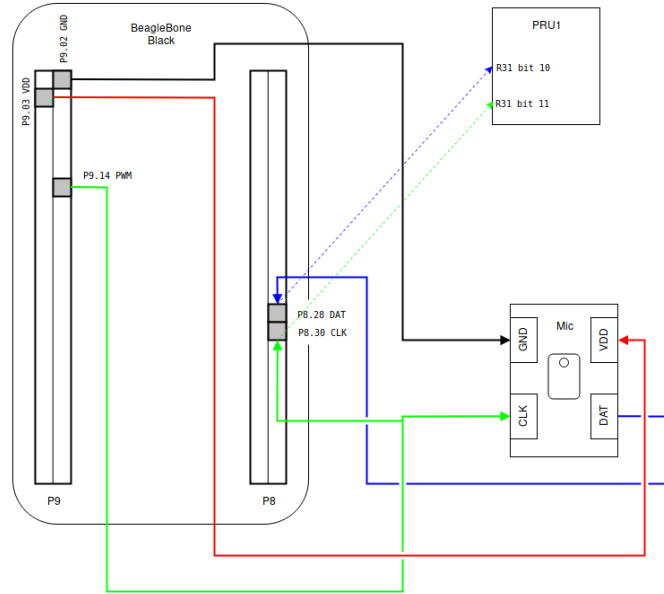

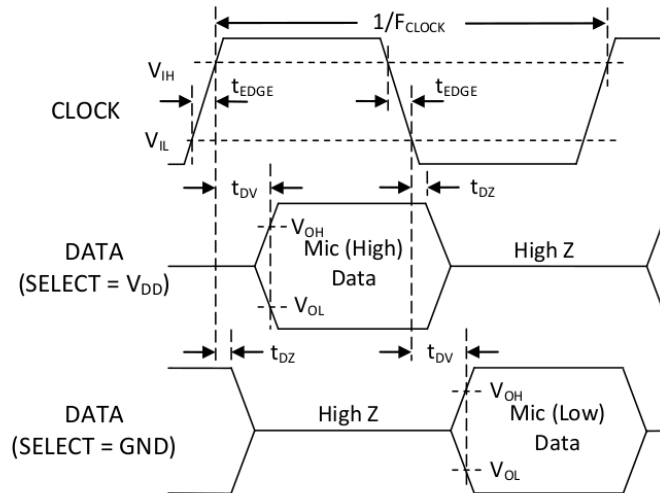Figure 3.4: Wiring of one microphone to the board (own work).

Figure 3.5: Timing diagram of the microphones, obtained from the datasheet.

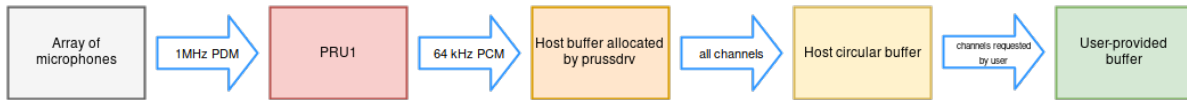## 3.3 Overview of the whole processing chain

Figure 3.6: Overview of the whole processing chain (own work)

The complete processing chain is made of several elements :

- **The array of microphones** : Each microphone generates a 1.028 MHz PDM signal sent to PRU1.
- **PRU1** : Converts the PDM signals to lower-rate, 64 kHz PCM signals with a separate CIC filter for each microphone, and outputs the data directly into the first host's buffer.
- **1st host buffer** : The buffer to which PRU1 directly writes data of the PCM signals. It is allocated by the `prussdrv` library. Samples in this buffer are regularly transferred by the host interface to the host's second buffer.
- **2nd host buffer** : A circular buffer, bigger than the first one. When the `pcm_read` function of the interface is called, the number of samples requested by the user are popped from this buffer, and only the data from the channels chosen by the user is written to the user-provided buffer.
- **User-provided buffer** : A buffer allocated by the user of the interface. Must be large enough to hold the amount of data requested by the user.

3.4 Core processing code

The core audio processing code, which implements the CIC filter, runs on the PRU and handles the following tasks :

- Reading the data from the microphones in time. Since we are using the Kurodako board, we have to read data at every edge of the clock. Must be done thrice after the rising edge, and thrice after the falling edge, by reading the corresponding pins through the r31 register.
- Processing all the channels. This means updating all the stages of the filter to get the output data for each channel.
- Writing the results (output data) directly into the host's memory (where a fixed size buffer has been allocated by `prussdrv`).
- Interrupting the host when data is ready to be retrieved by the host from its buffer.

It is written exclusively in PRU assembly (`src/6Mic-CIC/pru1.asm` in the project files). The chosen CIC filter parameters are `N = 4`, `M = 1` and `R = 16`. We chose these parameters for the following reasons :

- Register usage, see section 3.4.1.
- Timing constraints, see section 3.4.1.
- Decent frequency response and aliasing attenuation, see Figure 2.3.

For performance reasons, the PRU uses registers to store the data of each stage of the filter. Because the PRU only has 30 available registers for storing data, it needs to use registers from the scratchpads as well. It does so by exchanging some of its registers with the scratchpads using the `XIN` and `XOUT` instructions. We have to keep track of several different counters along the way, and also do the correct register exchanges with the scratchpads to keep any data from being overwritten. We need one counter to implement the decimator, one for the `t_dv` delay and one to keep track of how many samples have been written to the host's memory. This last counter is referred to as *bytes counter* and is needed so an interrupt can be triggered when the expected amount of data is ready. The processing is done in several steps as detailed below.

- Read data from channels 1-3, process it and output the results to the host buffer. Details :
  - Load chan. 1, 2 registers from BANK0 (scratchpad).
  - Wait for rising edge, then wait for chan. 1, 2 data.
  - Read chan. 1, 2 input data.
  - Perform one iteration of the filter.
    * If the downsampling counter reaches R, execute the comb stages and store chan 1, 2 outputs in registers.
  - Store chan. 1, 2 registers to BANK0 and load chan. 3 from BANK1.
  - Read chan. 3 input data.
  - Perform one iteration of the filter.
    * If the downsampling counter reaches R, execute the comb stages and store chan 3 output in a register.
  - Write chan. 1-3 outputs to host buffer and increment the bytes counter.
  - Store chan. 3 registers to BANK1.

- Read data from channels 4-6, process it and output the results to the host buffer. Details :
  - Load chan. 4, 5 registers from BANK1 and BANK2.
  - Wait for falling edge, then wait for chan. 4, 5 data.
  - Read chan. 4, 5 input data.
  - Perform one iteration of the filter.
    * If the downsampling counter reaches R, execute the comb stages and store chan 4, 5 outputs in registers.
  - Store chan. 4, 5 registers to BANK1 and BANK2 and load chan. 6 from BANK2.
  - Read chan. 6 input data.
  - Perform one iteration of the filter.
    * If the downsampling counter reaches R, execute the comb stages and store chan 6 output in a register.
  - Write chan. 4-6 outputs to host buffer and increment the bytes counter.
  - Store chan. 6 registers to BANK2.

- Check the written bytes counter. If the end of the buffer has been reached, send interrupt 1 to the host and start writing to the beginning of the buffer again. If the middle of the buffer has just been reached, send interrupt 0 to the host and continue writing.
- Loop back to the beginning.

In order to allow the host to retrieve all the samples before the PRU overwrites them with new data, we have the PRU trigger an interrupt whenever it reaches the middle of the buffer, or the end. These interrupts have different codes which allows the host to tell which half of the buffer contains fresh data. Consequently, the host can be sure to read one half of the buffer while the other half is being overwritten by the PRU.
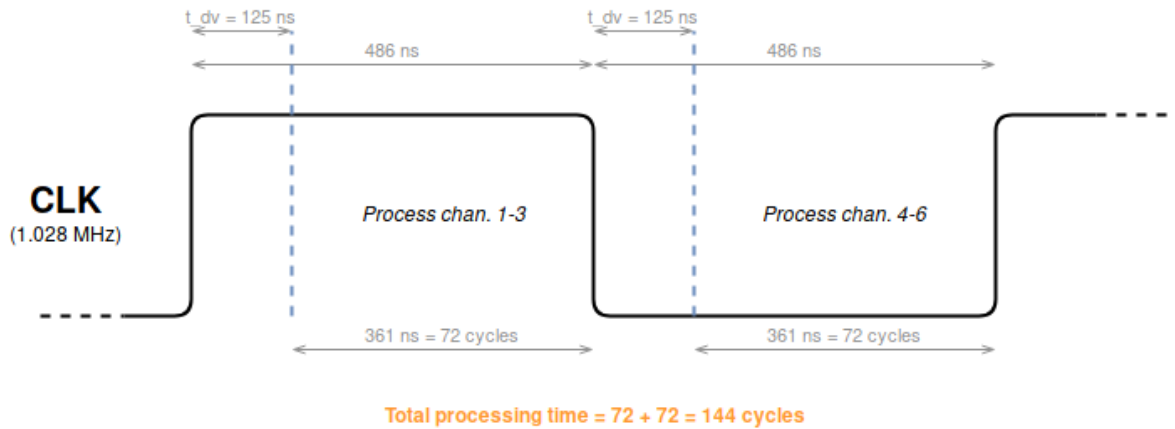
Figure 3.7: Timing diagram of the processing times with mic multiplexing (own work).
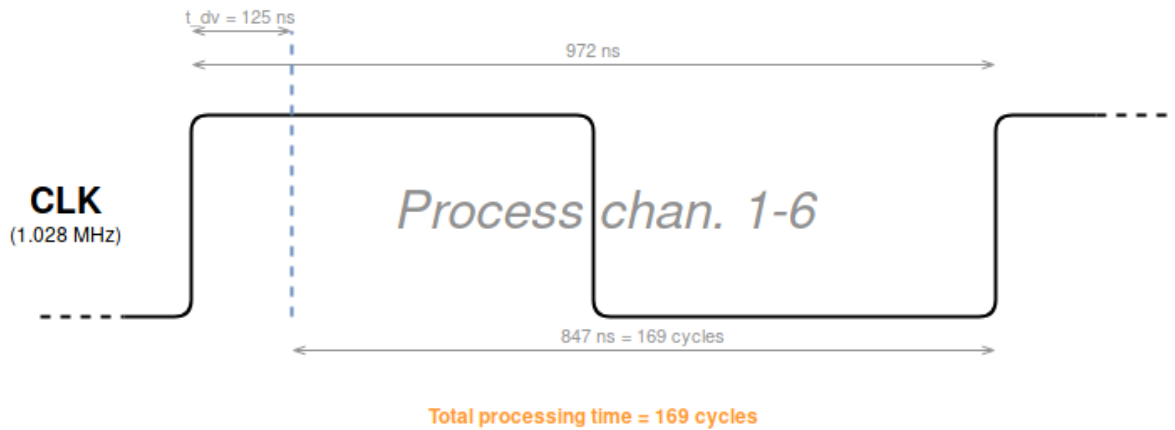

Figure 3.8: Timing diagram of the processing times without mic multiplexing (own work).

As mentioned above, multiplexing two microphones on 1 input pin reduces the available processing time, as can be seen from Figures 3.7 and 3.8.

Reading the microphones' data is achieved by reading bits of the R31 register, which is connected to the PRU's input pins, to which the microphones' DATA lines are connected. Since we connected the microphones' clock to one of the input pins of the PRU as well, reading its state is done the same way. In order to know when to read the data, the PRU polls the CLK signal until it detects an edge. It then waits for 25 cycles (t_dv) and finally reads the data and stores it in a register.

This register is the first integrator in the CIC filter and all subsequent integrators can be updated. Once this is done, the decimator is implemented by using a counter. The decimator uses that counter to determine whether it is time to update the comb stages or not. Once the comb stages are updated, some of the results of the CIC filter are ready and are written to the host's memory.

3.4.1 Usage of resources depending on the CIC filter parameters and on the number of channels

In the current design, we use only registers to store all values involved in the CIC filter computation. Each PRU has 32 registers (r0-r31), but in practice only 30 are usable. This is because r30 and r31
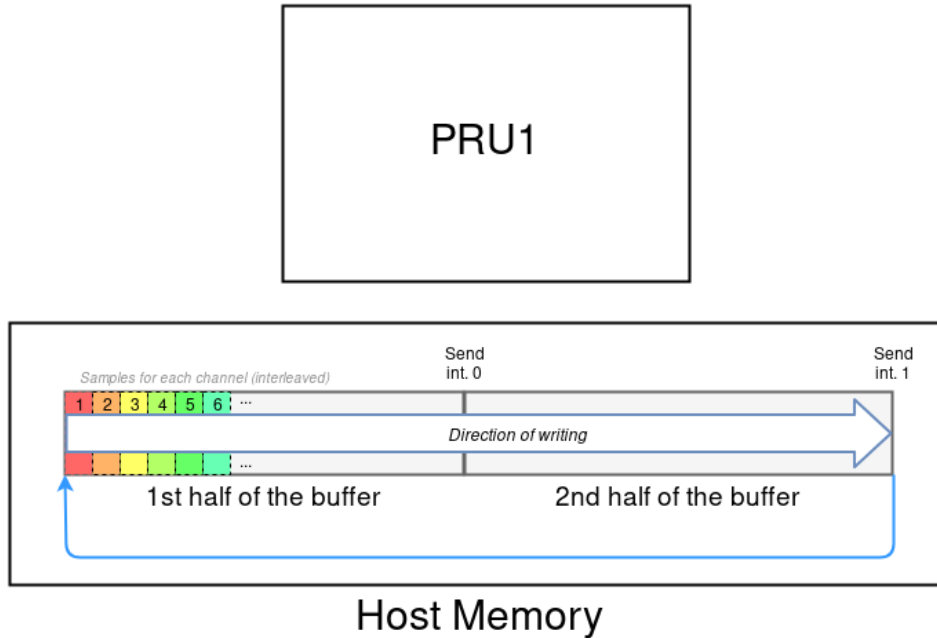
16

Figure 3.9: Illustration of the interleaved samples in the buffer, and interruptions (own work).

are reserved for interacting with BBB GPIO pins and triggering interrupts.

Furthermore, since we use the PRU's three additional register banks, we cannot use all the bits of `r0`, because the value of the first byte of `r0` is used as an offset value for register exchanges with the banks.

Another thing we need to take into account is that, apart from channel private data (values involved in the CIC filter computations for each channel), we also need to keep track of some channel *independent* data : a byte counter to keep track of the number of bytes written in the host buffer, a sample counter to implement decimation, the host buffer's size and address and temporary values used for delaying and processing input. This data has to stay on the PRU at all times.

We can fit the decimation counter in a single byte, as long as it does not exceed 255. However, the byte counter needs a complete register. The host's buffer size and address also each need one complete register. Finally, we can use one half register to store a temporary counter for implementing delays and also extracting single bits from the `r31` register (input pins).

Also, in our implementation, even though we have six channels, we only have three output registers. After filling the three registers, their contents are written to the host buffer and they can be used to store new inputs.

We store all this information on PRU1 in the following way :

- **r0** : scratchpad offset (bits 0-7), decimation counter (bits 8-15), delay counter / temporary storage for inputs (bits 16-31)
- **r23, r24, r25** : temporary storage for outputs of the filter, could be only one register, but this would add more memory operations which would add overhead
- **r26** : host buffer size
- **r27** : host buffer physical address
- **r28** : written bytes counter

The remaining registers are used for storing channel private data. In order to figure out how many registers we need for this, assuming one register is wide enough to hold the expected values (see Equation 2.1 above to check the required bit width), we can use the following formula (`M = 1`) :

```
n_reg_private = C * (N + 1 + N - 1 + N - 1) = C * (3N - 1)
```

17

With `M != 1`:

```
n_reg_private = C * (N + 1 + 2M * (N - 1))
```

Where C is the number of channels, N and M are the parameters of the CIC filter. It should be noted that `R` has no influence on the filter's register usage. However, `R` influences the filter's output data rate. In our case, `C = 6`, `M = 1`, `N = 4` and `R = 16`, so `n_reg_private = 66`. This, along with the 7 registers holding channel independent data on PRU1, gives us the total register usage :

```
n_reg_tot = C(3N - 1) + 7 = 66 + 7 = 73
```

Note that different parameters may allow for fitting more than one channel in a single register,[9] which would change the above formula and dramatically reduce the amount of registers needed.

We chose to keep `M = 1` because a greater value of `M` would increase the required number of registers by a significant amount. Furthermore, it is said in most documentation about CIC filters that keeping `M` to 1 or 2 suffices to provide a decent frequency response.

Total number of registers required given `N` and `C` (`M = 1`) (assuming bit width of 32 is enough)

	N	1	2	3	4	5
C						
1		9	12	15	18	21
2		11	17	23	29	35
3		13	22	31	40	49
6		19	37	55	73	91
8		23	47	71	95	119

An important thing to note is that the PRU is supposed to support the `XCHG` instruction, which exchanges registers from the PRU to one of the banks in one cycle. Unfortunately, it does not work. Only the `XIN` and `XOUT` instructions currently work. This means that we always need to keep **n_reg_private** (11) registers free to act as a temporary storage place. The PRU's registers handle this task.

With the current parameters, storing all data on the PRU leaves one free register : r29. Registers r1-r22 are used for channel private data and are the ones exchanged with the scratchpads.

Data rate of filter output

We also want to figure out the data rate of the filter's output. To do this, using the formula described earlier, we first compute the output bit width `B_out`. In our case, `B_in = 1`, so `B_out = 17`. However, the PRU's registers are 32 bits wide and it is more convenient to write the data in 32 bits chunks. Therefore, our 'effective' output bit width, `B_out'` is 32. Since we know the output sample rate is `f_s / R`, it is now straightforward to compute the output data rate :

```
D_out = B_out * f_s / R
```

Or using the `B_out'` :

```
D_out' = B_out' * f_s / R
```

In our case, `f_s` $\approx$ `1.028 MHz`, `R = 16` and `B_out' = 32`, which gives `D_out' = 2.056 Mb/s = 257 kB/s`.

---

[9]if `B_out` from Eq. 2.1 is ≤ 16.

	D_out' (kB / s)
R	
16	257
32	128.5
48	85.7
64	64.3

Table 3.1: Output data rate (kB / s) for 1 channel, given `R` (`M = 1`, `B_out' = 32`).

1-channel "proof of concept" program

Before writing the 6-channels CIC filter and the C host interface, we wrote a simple, 1-channel, proof of concept program implementing a CIC filter. We then used this code as a base and adapted it for the 6-channels implementation. It can be found in `src/1Mic-CIC/cic_pru1.asm` and run with `sh deploy.sh`. This will start recording from a microphone connected to P8.28 on the BeagleBone and then output the raw resulting PCM to a file in the `src/1Mic-CIC/output/` directory.

It follows the same basic principles as the 6-channels implementation, without scratchpad register exchanges, and waiting for only one channel instead of six.

This implementation works and allowed us to have an idea of how the output of a channel sounds with different parameters. The PCM signal can be converted to a WAV file using the `PCMtoWAV.py` script in `src/1Mic-CIC/wav_conv/`.

3.5   C interface (API)

The host interface is written in C in the `interface.h` and `interface.c` files in `src/6Mic-CIC/host`. It is currently very simple and provides the following functions :

```c
/**
 * @brief Very basic interface for reading audio using the PRU firmware.
 *
 * @author Loïc Droz <lk.droz@gmail.com>
 *
 */

#include "ringbuffer.h"
#include "loader.h"

#define SAMPLE_SIZE_BYTES 4

typedef struct pcm_t {
    // Number of channels
    size_t nchan;
    // *Per-channel* sample rate of the PCM signal in Hz
    size_t sample_rate;
    // The buffer accessed by the PRU, mapped by prussdrv, and its length
    volatile void * PRU_buffer;
    // Length of the aforementioned buffer
    unsigned int PRU_buffer_len;
    // The ring buffer which is the main place for storing data
    ringbuffer_t * main_buffer;
    // Function pointer to an optional filter
    // TODO:
} pcm_t;

/**
 * @brief Initialize PRU processing. Must be called before any other function of this
    file.
 *
```

```c
31  * @return pcm_t* A pointer to a new pcm object in case of success, NULL otherwise.
32  */
33 //pcm_t * pru_processing_init(size_t nchan, size_t sample_rate);
34 pcm_t * pru_processing_init(void);
35
36 /**
37  * @brief Stop processing and free/close all resources.
38  *
39  * @param pcm The pcm object containing the resources.
40  */
41 void pru_processing_close(pcm_t * pcm);
42
43 /**
44  * @brief Read a given number of blocks of given size and output them to the user
45      provided buffer.
46  *
47  * @param src The source pcm from which to read.
48  * @param dst The buffer to which we want to write data.
49  * @param nsamples The number of samples to read from each channel.
50  * @param nchan The number of channels to read.
51  * @return int The number of samples effectively written.
52  */
53 size_t pcm_read(pcm_t * src, void * dst, size_t nsamples, size_t nchan);
54
55 /**
56  * @brief Get the current length of the circular buffer holding the recorded samples.
57  *
58  * @param void
59  * @return size_t The length of the buffer.
60  */
61 size_t pcm_buffer_length(void);
62
63 /**
64  * @brief Get the max length of the circular buffer holding the recorded samples.
65  *
66  * @param void
67  * @return size_t The max length of the buffer.
68  */
69 size_t pcm_buffer_maxlength(void);
70
71 /**
72  * @brief Enable recording of the audio to the ringbuffer.
73  *
74  * Once this function is called, the interface will start copying
75  * from the buffer the PRU writes to, to the main ringbuffer of the interface. In order
76      to avoid
77  * a ringbuffer overflow, the user must therefore start reading using pcm_read quickly
78      after this
79  * function had been called.
80  *
81  */
82 void enable_recording(void);
83
84 /**
85  * @brief Disable recording of the audio to the ringbuffer.
86  *
87  * Once this function is called, the interface will stop copying data to the main
88      ringbuffer.
89  * This means only the samples remaining in the ringbuffer after this function was called
90      can be read.
91  *
92  */
93 void disable_recording(void);
```

../src/6Mic–CIC/host/interface.h

The host interface contains several components. The first is a special buffer in the host memory

allocated by `prussdrv` which the PRU directly writes to. It contains the data computed by the PRU for all channels. The second component is a larger circular buffer which serves as the main temporary buffer of the interface.

When the `pru_processing_init()` function of the API is called, it loads and starts the PRU firmware and starts a separate thread which handles the recording of the samples from the PRU buffer to the bigger circular buffer of the interface. Everytime the PRU has finished writing one half of this buffer, it sends an interrupt to the host and starts writing to the other half. The code of the sent interrupt designates which half of the buffer was filled with fresh data.

This thread essentially waits alternatively for each interruption and then copies the corresponding buffer half to the circular buffer. Waiting for each interrupt alternatively makes sure the host cannot get desynchronized and start writing the buffer halves that are being overwritten by the PRU. However, this mechanism cannot detect whether a buffer half has been skipped. This can happen if the host is too busy with other programs which can make it miss an interrupt by the PRU. Furthermore, if an interrupt is missed, the next one will be skipped as well, since we wait for them alternatively. Therefore if an interrupt is missed, two buffer halves will actually be missed. We chose to use this mechanism because of its simplicity, and because it did not require the PRU to count the buffer halves it filled and output the number to the host regularly. This would however be a better solution.

The circular buffer is similar to a queue but with a fixed length. Its API provides functions for pushing and popping data from it. In the case of an overflow, the push function will only trigger a warning and overwrite the oldest data in the buffer.

The `pcm_read` function provided by the API pops the number of samples required by the user from the circular buffer, then extracts only the number of channels the user asks for, and finally outputs this data to the user provided buffer.

It is important to note that there will be concurrent accesses to the circular buffer : data pushed by the recording thread, and data popped by the user thread calling the `pcm_read` function. Our implementation handles this by making accesses mutually exclusive by using a mutex declared in the `interface.c` file. This is a simple but not ideal solution. For example, if a call to `pcm_read` pops a large chunk of data from the circular buffer, this may block the recording for a long enough time that it will miss the next interrupt, and therefore a large chunk of samples.

Chapter 4

Results

4.1 Single channel implementation

The single channel implementation works and shows that it is possible to implement a CIC filter on the PRU and rely on PRU writes to the host memory. With the following parameters (N = 4, M = 1, R = 16), we were able to get a moderately noisy but intelligible signal at a sample rate of 64 kHz.
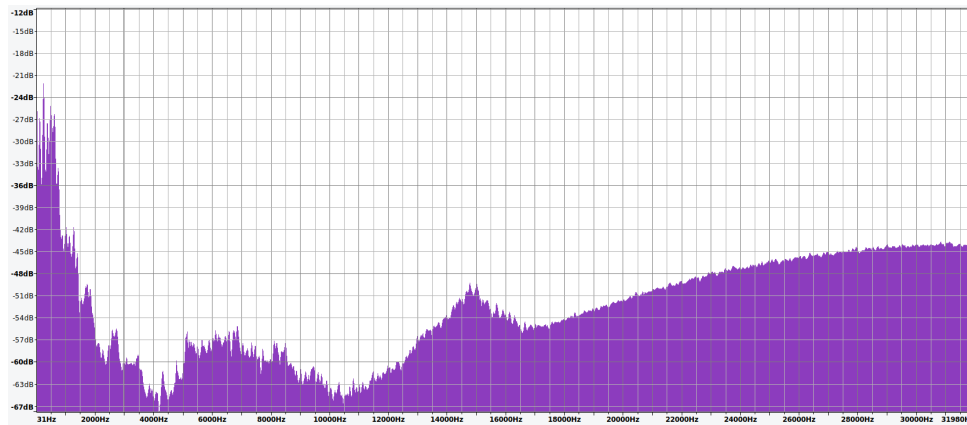

Figure 4.1: Spectrogram of the signal of one of the channels from the demo recording made during the project's presentation (own work, generated using Audacity). As we are interested in human speech, we can attenuate the noise above 8 kHz with a simple low-pass filter. Moreover, a compensation filter can be used to properly shape the response below 8 kHz.

It is worth noting that using one channel, the timing constraints are much less tight since there is much less processing to be done. This leaves more freedom in choosing the parameters. For example, higher input sample rates (compared to the 6-channel 1.028 MHz) are achievable.

4.2 6-channels implementation

At the time of writing this report, the 6-channels implementation works. The C interface has been tested recording samples for up to five minutes and appears to work.

We have noticed that some occasional glitches appeared in the signal while using a microphone connected using a breadboard and soldered by hand. However, these glitches seem to have disappeared when we started using the Kurodako Board.

More rigorous testing needs to be done to make sure the PRU processing code and the interface work as intended.
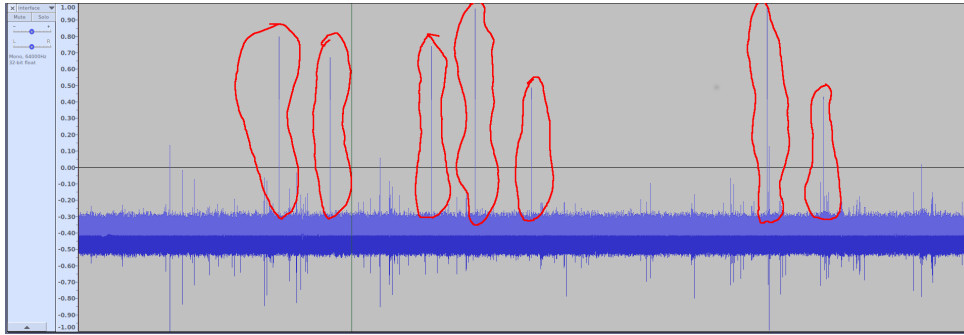

Figure 4.2: Glitches in the signal while recording for about 4m15s (own work).

Chapter 5

Challenges faced

5.1 Lack of documentation and the existence of two, different drivers

The biggest challenges faced in this project were probably the lack of clear and organized documentation about how to run code on the PRU from the Linux host, how to configure the operating system so that the BeagleBone's pins can be multiplexed to the PRU, how to choose which driver to use, and finally how to configure the BeagleBone for it to work. Most of the documentation and examples are scarce, sometimes outdated and scattered across multiple websites which forced us to do a lot of trial and errors on things such as how to enable drivers or the right interrupts between the host and the PRU.

Apart from the fact that embedded systems is an inherently tough subject that is by far not as popular as higher level programming,[1] I think the scarcity of the documentation is probably the greatest factor that made the learning curve for this project rather steep.

5.2 Limited number of registers and tight timings

On a more technical point of view, processing six channels simultaneously on one PRU is feasible, but challenging in terms of resource management. In our current implementation of the 6-channels CIC filter on the PRU, all operations required for processing one sample (from each channel) must execute in less than 144 cycles. All except one of the PRU's registers are used, and the majority of the banks' registers are used as well.

This 'shortage' of registers on the PRU forced us to write each set of six samples in two steps of three registers and was the source of a bug : the bytes counter used by the PRU to keep track of where it is writing in the host buffer was updated incorrectly. We incremented it by 6 * 4 bytes once instead of incrementing it by 3 * 4 bytes twice, which discarded the first three channels by overwriting them with the last three ones.

Another challenge was to design the program such that it would not rely on the host too much because of the host's unpredictable timings and busy nature. Below is an example of a bug that happened when the host was in charge of retrieving a new sample everytime it was ready (with these parameters, 64'000 times per second). The host couldn't keep up and missed many samples, resulting in this quite weird looking (and sounding) spectrogram.

---

[1]in particular the PRU, which seems to be a piece of hardware very few people use or know about.
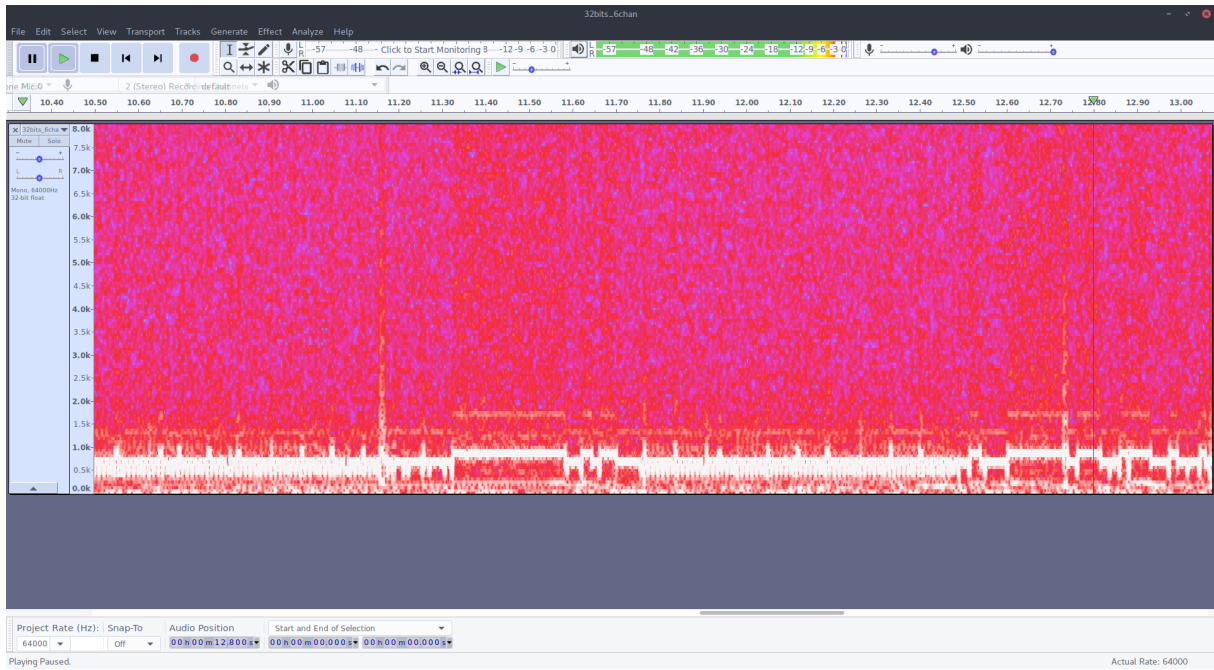
Figure 5.1: Spectrogram of a signal missing samples as a result of timing issues (own work).

Chapter 6

Possible improvements and additional features

6.1   Use both PRUs

Currently, we use only one PRU (PRU1) to handle the audio processing with the CIC filter. The design choice was made to make the implementation simpler. However, this also limited us to being only able to process six channels at a time instead of the initial goal of eight.

It could for example be possible to implement a CIC on both PRUs which would allow us to handle more than six channels. Another idea would be to keep the CIC filter on one PRU, but move the compensation filter which would be implemented on the host ARM CPU to the other PRU, offloading the ARM CPU even further and also reducing the latency.

6.2   Tweak the parameters to get smaller bit width and possibly handle more channels

The parameters we are currently using for the CIC filter (R = 16, N = 4, M = 1) give a decent frequency response (as seen in Figure 2.3) but require 17 bits per stage of the filter. Tweaking the parameters to achieve 16 bits or less would allow fitting 2 channels in one register which would dramatically reduce the usage of memory resources, both on the PRU and on the host. For example, one could use N = 3, which would require only 13 bits per stage of the filter. This would allow fitting two channels per register in the filter, perhaps making an 8-channels implementation possible with only one PRU.

We provide a script[1] for checking the frequency response and required bit width per stage, given different parameters.

6.3   Use of a lookup table

Since the PRUs each have some data memory available (8 kB each, with an additional shared 12 kB), it might be more time-efficient to implement the PDM to PCM conversion using a pre-computed look-up table stored in memory instead of implementing a CIC filter.

6.4   Better and more modular interface

For now the interface is very limited, and depending on how many channels the user chooses to read, the whole program can also be very wasteful on resources. This is because with the current implementation,

---

[1] `utils/power_resp_CIC_folds.py` in the project files.

the PRU always processes the six channels, and the host interface's backend always records all six channels, even if in the end the user requests fewer channels. In the event the user wants to read fewer channels, the interface's front-end will just drop the data from the channels the user does not want, before sending the data to the user.

An improvement could be to let the user choose how many channels they intend to use at most, and then only handle this number of channels instead of the maximum possible. However, making the CIC filter's code modular might not be a feasible task given the high performance requirements, at least with the current model of our implementation. A workaround would be to write several programs, possibly one for each number of channels, and load the appropriate one on the PRU (with the `deploy.sh` script) by letting the user choose the number of channels when calling the `pru_processing_init` function. With fewer channels to process, a CIC filter with a better response (i.e. higher aliasing attenuation) can be achieved, with the available resources.

As explained earlier, the current implementation of the interface uses a very simple circular buffer we implemented ourselves, and concurrent accesses are managed using a mutex. This is probably a quite inefficient and unoptimized solution that could lead the interface back-end to miss the recovery of some of the samples from the PRU, if the `pcm_read` function is asked to retrieve a huge chunk of data.

A workaround to this would be to implement a 'smarter' concurrent circular buffer or use an existing one. One solution we considered but eventually did not have enough time to use was the `liblfds`[2] library, which contains an implementation of a thread-safe, concurrent cirular buffer (ringbuffer).

Another idea would be to adapt the interface so that it could directly create a WAV file, and not a raw PCM file.

6.4.1 Introduce further filtering on the host side

As mentioned before, CIC filters are very efficient filters but they lack a flat frequency response with a sharp cutoff and we need an additional compensation filter appended after them to get a better response. The current implementation of the host interface does not implement such a filter yet but this is a possible and probably very useful improvement that could be made. A nice feature to have could be to make it modular such that it can accept many different types of compensation filters.

---

[2] https://liblfds.org/

Chapter 7

Acknowledgments

I would like to offer my special thanks to Eric Bezzam for his continuous assistance throughout the semester. He provided me with valuable explanations about some of the signal processing theory and nomenclature, which made understanding the motivation and goals of the project easier for me. He also greatly helped me by writing some useful little programs for converting PCM signals to WAV, plotting the filter's power response, and analyzing timings in the PRU.

I would also like to thank Robin Scheibler, my supervisor for this project, for his valuable pieces of advice he gave throughout the project.

I would like to also thank the people who were there at the midterm presentation for their valuable remarks and advices, such as the idea of using a look-up table for the filter.

Sources and bibliography

- CIC Filter theory :
    - Hogenauer, Eugene. *"An Economical Class of Digital Filters For Decimation and Interpolation"* https://www.researchgate.net/publication/3176890_An_economical_class_of_digital_filters_for_decimation_and_interpolation
    - https://en.wikipedia.org/wiki/Cascaded_integrator%E2%80%93comb_filter
    - https://dspguru.com/files/cic.pdf
    - https://www.embedded.com/design/configurable-systems/4006446/Understanding-cascaded-integrator-comb-filters

- PRU programming :
    - http://processors.wiki.ti.com/index.php/PRU_Assembly_Instructions
    - http://www.embedded-things.com/bbb/understanding-bbb-pru-shared-memory-access/
    - http://credentiality2.blogspot.ch/2015/09/beaglebone-pru-gpio-example.html
    - https://theembeddedkitchen.net/tag/beaglebone
    - http://www.ti.com/lit/wp/spry264a/spry264a.pdf
    - http://processors.wiki.ti.com/index.php/PRU_Linux_Application_Loader_API_Guide

- Host configuration :
    - http://catch22.eu/beaglebone/beaglebone-pru-uio/
    - https://www.teachmemicro.com/beaglebone-black-pwm-ubuntu-device-tree/
    - http://www.dangtrinh.com/2015/05/sharing-internet-with-beaglebone-black.html
    - https://beagleboard.org/getting-started