CHIP-OFF BY MATTER SUBTRACTION: *FRIGIDA VIA*

David Billard[1], Paul Vidonne[2]

[1]University of Applied Sciences in Geneva, Switzerland
David.Billard@hesge.ch

[2]LERTI, France
Paul.Vidonne@lerti.com

ABSTRACT

This work introduces an unpublished technique for extracting data from flash memory chips, especially from Ball Grid Array (BGA) components. This technique does not need any heating of the chip component, as opposed to infrared or hot air de-soldering. In addition, it avoids the need of re-balling BGA in case of missing balls at the wrong place. Thus it enhances the quality and integrity of the data extraction. However, this technique is destructive for the device motherboard and has limitations when memory chip content is encrypted. The technique works by subtracting matter by micro-milling, without heating. The technique has been extensively used in about fifty real cases for more than one year. It is named *frigida via*, compared to the *calda via* of infrared heating.

Keywords: Chip-off forensics, data extraction, BGA, data integrity preservation, micro-milling, infrared heating.

1. INTRODUCTION

Forensics laboratories are daily facing the challenge of extracting data from embedded or small scale digital devices. In the better case, the devices are already known from commercial vendors of extraction tools and a proved method is available to the practitioner. In most cases, the devices are unknown, or broken, and then begins the fastidious search of a method to extract data from the device without jeopardizing the judicial value of the – hypothetical – concealed evidence.

When no software-based method exists, the desoldering of the chip holding the data is accomplished. The chip is often a flash memory component, more and more of Ball Grid Array (BGA) technology. The de-soldering, even when routinely executed, is no error prone and induces a heavy stress on the component. Furthermore, the controlling of the heating is based on temperature probes which are not always accurate enough. This leads to chips being heated too much or chips being teared off. In the first case, the data content may be altered, even destroyed in some occasion. In the second case, some balls of the BGA will stay on the motherboard and the practitioner will have to re-ball the chip in order to extract data using a BGA reader.

As an example, the BGA component shown in figure 1 comes from a cell phone motherboard. The labeling on the chip is very clear: it's a NAND chip and the edges of the chip are sharp.

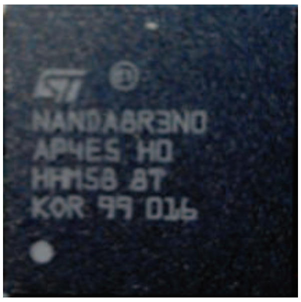


Figure 1: BGA from a cell phone motherboard

The chip has been heated using infrared and the result is shown in figure 2. The component changed color (no more labeling visible) and the edges are blurred. The ball grid is also a bit wavy: the heating has a dramatic effect on the component. However, the component is still readable and data can be extracted. The ruler (in millimeters) has been added to give the reader of this paper a better idea of the component's size.

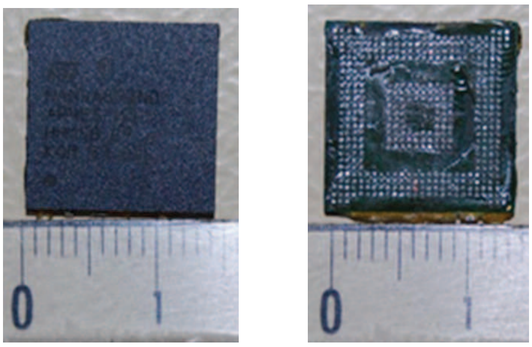


Figure 2: Heated BGA recto and verso

In this paper we propose a new method for taking off BGA chips from motherboard, without heating them. In fact, instead of taking the chip off, we remove the motherboard from under the chip. We use micro-milling technology and we subtract matter from the motherboard on the other side of the chip, until we reach the ball grid. The process is constantly monitored and controlled and it stops when reaching the balls. A result of this process is shown below.

The Micron chip presented in figure 3 is still attached to the motherboard. The labeling is clear, and the edges of the chip are sharp.

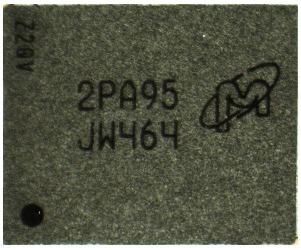


Figure 3: Micron BGA on the motherboard

Once the milling process is done, the chip labeling is still as clear on the recto, and the grid balls are all present at the verso, as shown in figure 4.

Since no heating has been applied, the chip content has been cleared of any stress and is intact. We have been using and refining this technique for about one year on fifty real cases. We had an issue with only one particular case which is presented later in this work.

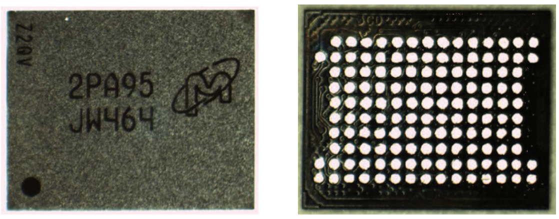


Figure 4: Milled Micron BGA recto and verso

The paper is organized as follows: section 2 is a review of literature about data extraction from flash components; section 3 presents the principle of the milling process, the machine and the interaction with precision bar turning; section 4 lists some lessons learned in using this technique compared to infrared heating and presents a comparative table of pros and cons.

2. RELATED WORKS

An extensive literature exists about extracting data from flash (or eeprom) memory chips. Most of this literature assumes that the device is in working order. For instance, (Breeuwsma, 2006) addresses the use of JTAG (boundary-scan) in order to bypass or trick the processor or the memory controller. In (Sansurooah, 2009), the author is addressing the use of flasher tools in order to load a bootloader into the device memory; this bootloader is designed to gain access to low-level memory management, thus enabling the reading of all memory blocks.

Some papers, like (Fiorillo, 2009) are using hot air de-soldering to compare the content of flash memory chips before / after some writing of data. In (Willassen, 2005), several ways of desoldering chips are mentioned, all based on heating the component (hot air, infrared, ...). In a remarkable presentation, (van der Knijff, 2007) presents an overview of most techniques for chipoff and JTAG access.

Commercial products like (Cellebrite, 2015) or (Microsystemation, 2015) are based on several techniques in order to gain access to the low-level memory. Although these tools are not suited for chip-off, they provide the ability to decode memory dumps extracted from flash memory chips.

To our knowledge, the memory reading of broken / dismantled digital devices is done either by heating and chip-off or sometimes by entirely reconstructing the device around the flash memory. Our paper brings an unpublished approach, requiring no heating, thus enhancing the integrity and quality of the data extraction. It is especially designed for broken devices but also for running devices, with some limitations, discussed in Sec. 4.

3. SUBTRACTING MATTER

3.1 PRINCIPLE

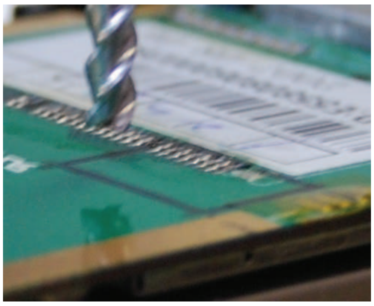
The aim of the technique is to subtract matter around the component. Concerning a BGA component, it sums up to obliterate the motherboard and its other components, leaving the BGA component alone.

The technique can be summarized into the following steps:

1. *Localization step*: since the motherboard is milled, at its verso, just under the memory chip, the cutting tool has to be directed to the localization of the chip, while the chip is hidden by the motherboard. Thus it is necessary to *locate the chip on the verso side* of the motherboard by measuring distances from the board sides to the chip sides *on the recto side*. Then using the measures to draw the shape of the chip on the verso of the motherboard.

   Figure 5 presents a photography of the drawing of the shape of the chip, on the verso of the motherboard.

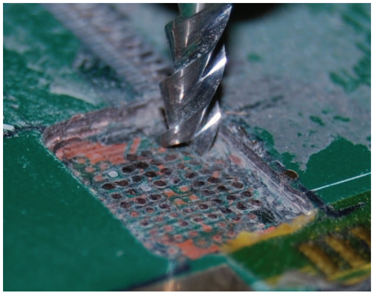
2. *Revolving step*: turning on itself the BGA component, still attached to its part of motherboard, in order to have the motherboard facing up (thus the component facing down).



Drawing the chip shape at the verso

Figure 5: Localization step

3. *Peeling step*: using a milling cutter to cut, layer by layer, the motherboard, until short of arriving to the grid balls. Sometimes it means also cutting layers of BGA components when the grid balls are lightly encased into the chip. Figure 6 presents a photography of the milling cutter sawing through the motherboard until the grid balls are exposed.



Milling to the grid balls

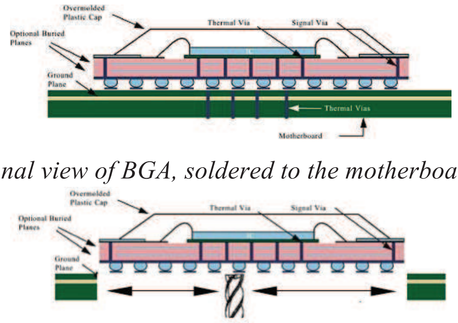
Figure 6: Peeling step

   For this milling step, it is of utmost importance that the milling cutter head and the motherboard be perfectly aligned at 90°. Even a very small angle deviation may lead to a catastrophic bite of the milling cutter into the BGA component. In that case, the component may be utterly destroyed.

4. *Cleansing step*: removing the last bits of motherboard layer and epoxy that may still adhere to the grid balls.

Once those steps are finished, there is no need of re-balling the component, since no ball has been lost. The component can be used straight away in a flash reader, provided that the practitioner has the right pinout module.

The upper image in figure 7 represents a sectional view of a BGA, taken from (Guenin, 2002). The lower image represents the working of the milling cutter, subtracting the motherboard and leaving the grid balls exposed.



Sectional view of BGA, soldered to the motherboard

Sectional view of BGA, detached by milling

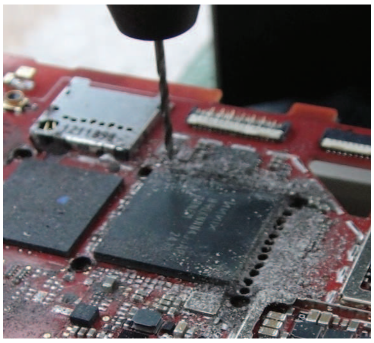
Figure 7: Process illustrated

3.2 VARIANT

In some case, in particular when processor and memory are piled one on top of the other, before the localization step, the motherboard has to be cuted all around the component, either by drilling holes close to the four sides (like old fashioned stamps) or by drilling one hole and using a fretsaw all around the BGA component. This operation is called the punching step and figure 8 presents a photography of such step.

3.3 MACHINE

The machine used for the milling is a standard precision micro milling machine from Proxxon (Proxxon, 2015). It must be capable of 0.05 millimeter steps (0.002 inch) with a rotating speed varying from 5,000 to 20,000 rpm (revolutions per minute). The milling cutters have usually a diameter between 1 and 3 millimeters (0.04 to 0.12 inch). A watchmaker grade magnifier, or a digital magnifier, is needed to control and verify the peeling step.



Separating the component from the others

Figure 8: Punching step

3.4 PRECISION BAR TURNING

The idea to implement this *frigida via* technique comes from interaction with specialists of *precision bar turning*. These people are specialized in manufacturing tiny pieces of hardware, like gear wheels one can find in mechanical watches or complex components with special alloys used in space satellites.

We were facing more and more devices locked to investigation due to their poor condition: cell phone with a bullet hole, GPS retrieved from a sunken boat or tablet barely surviving a plane crash. Using commercial tools or flash boxes was not an option and infrared heating was adding additional stress on components already submitted to heavy stress. Therefore, instead of thinking like repairing firms whose job is to detach an object in order to repair or analyze the failure of the whole device, we thought about isolating the memory from its external surroundings. In other words: obliterating the surrounding area, in order to leave the component exposed.

One of the first case prompting us to use the milling was the investigation of a cell phone, retrieved after a car chase between the police and three drug dealers. The motherboard was badly damaged and we feared that using infrared on the memory chip may inadvertently damage further the chip. After extensive testing on spare devices, the milling process was applied to the device remnants and information was successfully extracted.

4. LESSONS LEARNED AND METHOD COMPARISON

4.1 ENCRYPTION

The technique explained in this paper has to be used with prudence when dealing with encrypted devices. In a real case about narcotics, a BlackBerry 9720 was seized. It had a keyboard lock that the owner was not willing to depart from. The *frigida via* was successfully used and figure 9 presents the recto and verso images of the SKhynix chip.

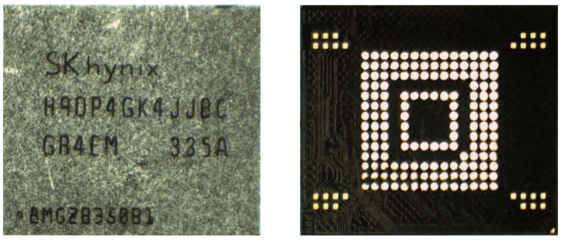


Figure 9: Milled SKhynix BGA recto and verso

But after reading the chip, it appeared that all the component content was encrypted. Finally, after some weeks, the password is supplied. Unfortunately, this password alone is not sufficient to decrypt the content: it must be used in conjunction with some hardware information, contained in other components of the motherboard. Thus, even with the password, the memory remains encrypted.

4.2 PROCESS DURATION & COMPARISON

The milling technique takes between thirty minutes to one hour, depending on the quality of the motherboard. Namely, if the motherboard is flat, without any deformation, it takes less than thirty minutes, and if the motherboard has been retrieved after a helicopter crash, it takes about one hour. Once the chip is off the motherboard, it is immediately available for reading and the first contact in the reader socket is usually the good one.

The infrared (or hotair) method is usually shorter in time for the chip-off, thirty minutes being the upper limit of the process. However, the process can be impeded in many ways.

First, the chip can loose grid balls during the process; some of them staying attached to the motherboard. After cooling the chip, many tries are needed to find which grid balls are missing and additional time is needed to re-ball the chip, even if not all the grid balls need to be present, only the "useful" ones.

The heating process also leaves residues of matter that have to be scrapped off using toothbrushes or special treatment. Then several tries are also needed to place the chip correctly into the reader socket, since the edges of the chip are no more rectilinear.

Furthermore, the epoxy layer between the chip and the motherboard can glue the chip to the motherboard, even if the grid balls are melted. We did not find if the epoxy glues together the chip and the motherboard at heating time or if it is done during the assembly of the motherboard. In that case, even a heavy heating cannot desolder the chip, and will more likely destroy the content of the component.

In table 1 we summarize the main differences between *calda via* and *frigida via*.

Table 1: Comparison Infrared vs Milling

Calda via: Infrared	*Frigida via*: Milling
Heat damage	No heat applied
Re-balling necessary	No need of re-balling
Extensive cleansing	Light cleansing
Resoldering possible	No resoldering
Same process duration	

The table 1 shows the most obvious differences between infrared and milling. But even if milling seems superior in many aspects with respect to infrared, we are still using the two techniques on the cases. The choice of the technique to apply is dictated by several factors, among which:

1. the availability of the machines;
2. the risk of finding encrypted data linked tohardware components;
3. the risk of damaging the chip by heating;
4. the likeness of epoxy presence gluing thememory chip and the motherboard;
5. the training of the practitioner.

When facing a chip-off, we are applying a riskbased decision matrix in order to decide between *calda* and *frigida via*.

5. CONCLUSION

In this paper, we present a new technique for extracting data from flash memory chips, especially from Ball Grid Array (BGA) components.

This technique, called *frigida via* (or milling), is complementary of infrared or hot air chip-off processes and offers many new possibilities.

Instead of relying on the heating of the solder of BGA component, in the hope that the component