Embedded Flash-Technology

Highly Reliable 10ns MONOS Flash

Authors: Kiyoshi Matsubara, Wolfgang Stenzl and Joachim Hüpper, Renesas Technology

*Embedded non-volatile memory (NVM) is one of the most important technology elements in the world of microcontrollers. This is because CPU cores have become so fast that system performance is often more dependent on the memory than on the core itself. Also, NVM modules are frequently under scrutiny by quality specialists as they are generally seen as modules with quality risks. MONOS Flash by Renesas Technology stands for speed and reliability. This article presents facts and details about that technology.*

The operating principle of MONOS technology (metal-oxide-nitride-oxide-silicon) is anything but new. It has already been described 20 years ago and is being used in EEPROM products such as smartcard MCUs. The main advantage of MONOS technology consists in the fact that in case of a defect, rather than all charge carriers on a floating gate, only those at the defect location are drained. This accounts for a considerable increase in reliability. In conventional Flash technology, some of the dimensions had to be increased to obtain better reliability. MONOS Flash on the contrary does not require this, and dimensions can even be reduced. Smaller cells, however, also offer higher speed, resulting in a true read access time of 10ns only, with high reliability over the entire temperature range of -40°C to +125°C. Amongst others, bare dice are used in automotive applications. The maximum temperature in this case will be more than 150°C, and it applies for the entire scope of functions such as reading, writing and erasing. Furthermore, an effective access time of just 5ns can be obtained with only minor losses in deterministic behaviour by using common tricks such as wide buses.

Diagram 1 shows the MONOS Flash cell structure in comparison to conventional NOR Flash and their corresponding equivalent circuit diagrams. The charge carriers on the floating gate determine whether a logic one or zero is stored. In a conventional flash cell, the floating gate is conducting and therefore all charge carriers are bound to be drained in case of a leak in the oxide layer. MONOS Flash features a Charge Storage Layer made of a non-conducting nitride layer ($Si_3N_4$), effectively solving the problem.

**Voltages are Important Parameters**

The read and write voltages and the cell’s threshold voltage ($V_{th}$) are crucial technology parameters. Due to the MONOS Flash cell structure, reading can be performed at the same small voltages like the core logic. As a consequence, cell control logic and word-line-drivers ($WL$) can be made using fast core-MOS technology. Also, no charge pumps to generate high voltages for read operation are needed, thus saving space and cost besides enabling high speed reading. Low voltages and the avoidance of large charge pumps also help to lower power consumption.
Diagram 1: Comparison of NOR and MONOS Flash cells. CG stands for ‘Control Gate’ and MG denotes ‘Memory Gate’. This is why this cell architecture is often called a ‘Split Gate’ architecture.

The Vth threshold voltage is determined by the number of charge carriers on the nitride layer. Because the number of charge carriers also defines whether a logic zero or a logic one is stored, the Vth voltage is also a representation of zero/one. In other words, the voltage Vth describes whether a cell is in an erased or programmed state. Diagram 2 (see below) summarizes the measurements of 10 million MONOS Flash cells in 50 SuperH microcontrollers. A narrow distribution of the Vth threshold voltage and a wide gap between the ‘erased’ and ‘programmed’ states are important for reliable operation. Deviations from the distribution curve would point to outlier cells, i.e. to individual “bad cells” which exhibit weaknesses for various reasons. As can be seen easily, there were no outlier cells among the 10 million MONOS Flash cells – proof of the very narrow production tolerances that Renesas achieves with MONOS Flash.

The two curves are also located far from each other. Diagram 3 shows the same 50 microcontrollers after 1000 programming cycles and after operation at 150°C over the specified number of hours. A characteristic MONOS Flash property can be seen clearly: Vth for the programmed state decreases after a short time and then remains constant. This behaviour is caused by charge equalization, not by charge loss. NOR flash exhibits exactly the opposite behaviour. The long, flat tail after the initial Vth fall-off is a characteristic reliability feature of MONOS Flash. This behaviour ensures highly reliable operation at high temperatures, a feature that e.g. automotive suppliers demand for combustion motor control. Furthermore, diagram 3 shows that the distance between the erased/programmed curves is large even after 3000 hours of operation at 150°C. It is quite remarkable that there are still no outlier cells!
Diagram 2: This graph summarizes the measurements of 10 million MONOS Flash cells in 50 SuperH microcontrollers. Note the ideal standard distribution shape of the curves, the complete absence of outlier cells and the large distance between the curves.

Diagram 3: Characteristic MONOS Flash behaviour. Note the initial Vth decrease followed by a long tail. Again, there are no outlier cells.

As often, there are conflicting customer requirements. Users require the following features for integrated NVM technologies – especially in the area of high-performance 32bit microcontrollers:
1. Short read access times (i.e. fast flash operation).
2. Highest reliability.
3. Many years of data retention.
4. A high number of write-erase cycles (high endurance), if data needs to be saved besides code.

Considering the write-erase cycles (endurance), the following aspects can be stated: MONOS Flash inherently is a high speed technology, and typically offers 100 or 1000 write-erase cycles for code flash. Apart from implementing code flash memory, it has recently become possible to realize data flash memory using MONOS Flash technology, permitting to perform EEPROM emulation. The cells can be programmed and erased up to 30,000 times over the entire temperature range. By skilfully using several blocks, and depending on the amount of data to be saved, substantially more than 100,000 program/erase cycles can be obtained.

Diagram 4 shows the results. Programming or erase times remain constant across the specification range, and only increase slightly beyond this range.

Hierarchically arranged sense amplifiers are another technological highlight of MONOS Flash. These amplifiers measure the current during read accesses, and thus make a decision between logic zero and one (the measured current also depends on the charge carriers on the nitride).

The current is initially pre-judged by a sub-sense amplifier. Due to the short local bit line (LBL), this local judgement is very fast even at low voltage operation same as power-supply of Core-MOS. The measured value is subsequently passed on to the main sense amplifier using a low-impedance line. This technology also helps to obtain a read access time of 10ns. The times T1 through T4 shown in diagram 5 add up to these 10ns. Please note that this is a worst-case figure.
Diagram 5: Using sub-sense amplifiers (Sub S.A.) that are located very close to the flash cells helps to obtain extremely fast access times.

On top of this, MONOS Flash uses new technologies such as:

1. Due to the new cell structure, the reading and programming circuit are divided at the memory cell (see Diagram 1, CG is for reading and MG is for Programming).
2. Local (Divided) Bit Lines (LBL) which help to keep parasitic capacities low.
3. No High-Voltage is needed for reading, high speed WL operation is realized consistently by Core MOS circuits.

**The Importance of MONOS Flash for Microcontroller Products**

MONOS Flash based microcontrollers are the perfect choice for applications that call for high computing power, fast interrupt response times, highly deterministic behaviour and highest reliability – even at high operating temperatures. Some of these applications are control electronics for industrial electric motors (Renesas’ European market share in this field is 60%) or control modules for combustion motors and gearboxes (Renesas holds a world-wide market share of 20% in this area). MONOS Flash has been successful all over the world for several years because no other embedded NVM technology offers the same combination of features. MRAM for example, although also offering very high speed, still is at a disadvantage in terms of cell size. Moreover, this technology is currently not yet mature enough for deployment in automotive applications. Car manufacturers only put trust in NVM technologies that have proven themselves for many years in other markets and in large volumes – such as MONOS Flash.

MONOS flash in combination with the SH-2A CPU core by Renesas Technology offers more than 300 Dhrystones of MIPS computing power – over the entire temperature range, with write/read/erase operation across the full range, at highest levels of determinism and with best reliability. The SH-2A core delivers its contribution, being a modern, superscalar 32-bit RISC core in Harvard architecture with 16 registers at 32 bits each, of which 15 banks offer a fast interrupt time of only 6 clock cycles. On top of that, there are derivative versions with an integrated single/double precision floating-point unit which can deliver up to 400 MFLOPS and are even capable of performing operations such as calculating a sine in less than a microsecond.
An example is the SH7254xF derivative, a dedicated product for combustion motor control, with a total of 3.75Mbytes of integrated MONOS flash memory for code (100 cycles) and data (30,000 cycles).

Microcontroller power dissipation in this product range deserves special attention, because MONOS flash allows for significant power savings in the range of a factor of two. The benefit: In comparison to any other existing product, twice as much computing power can be packed into the thermal envelope (package) defined by the system manufacturers. In terms of the environment, this means an even larger reduction of CO₂ emissions for car traffic.

Moreover, diagrams 6 and 7 show that MONOS Flash based products can operate independently of temperature conditions, and that Renesas offers considerable safety margins between guaranteed specifications and typical chip behaviour.

**Diagram 6:** Access time features low temperature dependency.

**Diagram 7:** Power consumption exhibits low temperature dependency.
Renesas will present several MONOS Flash based products for industrial and consumer applications on the upcoming “Embedded World 2008” show, such as the SH713xF for electric motor control with 256kBytes of MONOS Flash, 100MIPS, 100% determinism, dual 12-bit ADC, high-performance timers and various serial interfaces (SCI, SPI, IIC), CAN and JTAG debug function.

Outlook
Renesas will continue to drive MONOS Flash development. MONOS Flash is presently under development for the 65nm technology node, and using the 45nm node is in research. The roadmap envisions 400MIPS in 2008 and 800MIPS in 2012. In parallel to these efforts, Renesas will introduce the first MRAM product for consumer and industrial applications in 2009. Demanding automotive customers can therefore expect to have reliability data from the large-scale, multi-year deployment of MRAM-based microcontroller products. MONOS Flash based microcontrollers will also penetrate smaller-scale applications. The smallest MONOS Flash product to date is a SH/Tiny with 32kBytes of memory. This series was designed for ‘white goods’ (washing machine motors, induction cooking hobs, compressors, pumps etc.) with the aim of reducing CO$_2$ emissions and making household appliances less noisy, more long-lived and reducing maintenance requirements. New SH/Tiny derivative versions with only 16kBytes of MONOS Flash will be offered beginning in 2008. With all these new offerings, Renesas will continue to expand its market share in the European household appliance industry (currently at 40%).

Conclusion
As mentioned, the operating principle of MONOS flash - i.e. the nitride layer – is not new, and yet only Renesas uses it widely. That is because a difficulty with the deployment of MONOS Flash has been large-scale, high yield mass production. Producing high-speed MONOS Flash is difficult. However, Renesas has invested numerous man-years of work of its scientists and engineers, and succeeded to overcome these problems after countless tests and simulations. In the end, all the work paid off with products that deliver their contribution to further CO$_2$ emission reductions.

Renesas Technology Europe GmbH
Karl-Hammerschmidt-Str. 42
85609 Dornach

Tel. +49-89-38070-0
Contact.eu@renesas.com