

## Eliminating Embedded Non-Volatile Memory IP Risks in SOCs

March 2009

By: [Linh Hong](#), Kilopass

**Abstract:** *For code that is fixed, ROM is the best NVM solution. For code that changes often, requiring thousands of cycles of program/erase endurance, Flash-based technologies are the best NVM solution. In between, there are value [applications](#) where logic NVM technologies fill a need; logic NVM solutions are embedded NVM IP implemented in standard logic CMOS, with no additional masks steps or processing steps.*

*With many embedded logic NVM options available, chip developers, foundry IP managers, and reliability managers need to narrow down the list of vendors by evaluating the risk of integrating an NVM IP. In this paper you will learn what to look for to ensure that the NVM IP will not cause headaches when your product goes into volume production.*

In the logic NVM space, there are several [technologies](#) available in the market. They include:

- Poly fuse solutions typically provided by the foundry where change in resistance is created by blowing the silicide on the poly line;
- Floating gate or charge trapping solutions where the programming mechanism is hot carrier injection;
- Anti-fuse solutions where “1”s are created from a hard oxide breakdown of the gate causing a resistive change.

With many solutions in the market today, and more solutions in the future due to SOC designs requiring logic NVM solutions, the list of vendors need to be quickly narrowed down by evaluating the risk of integrating their NVM IP. Eliminating risk comes even before determining which solution may have the best area, power, and performance metrics for the application. If a \$0.05 IP fails due to expected low yield, the entire die that may cost \$5 to \$30 will be thrown away. In worse cases, if the \$0.05 IP fails in the field, it may cause an RMA (Return Merchandise Authorization) of a \$150 to \$1000 system, and thousands of dollars may be spent on electrical and physical failure analysis to reach the root cause.

To eliminate the risk of integrating an NVM IP, there needs to be proven silicon in the form of extensive characterization and qualification of the memory. The qualification methodology and results will dictate the reliability of the IP. Reliability studies are designed to quickly identify passing versus failing devices under given conditions over a period of time. These studies use accelerated stress testing. The stress testing makes it possible to replicate failures that may occur over many years of device operation in high volume by using a smaller sample size over a shorter time but with accelerated stress.

Reliability testing of NVM IP depends on the technology. For fuse technologies offered by foundries and anti-fuse technology pioneered by Kilopass, reliability testing includes High Temperature Operating Life (HTOL) and process characterization tests such as Data Retention Bake (DRB). For floating charge based technologies that are multi-time programmable, Program/Erase Endurance testing is also required.

### 1.0 Reliability Background<sup>1</sup>

Field failures do not occur at a uniform rate, but follow a distribution in time commonly described as a “bathtub” curve; this is shown in Figure 1 below. The life of a device can be divided into three regions: Early Life, Inherent Life, and Wearout. The Early Life region is where the failure rate progressively improves. The purpose of the Early Life test is to detect and quantify the presence of failure mechanisms that occur in manufacturing to a subset of the device population. ATE (Automated Test Equipment) screens can be used to prevent these devices from reaching the field. Inherent Life test is designed to detect failure mechanisms that are intrinsic to all the devices in the population. This test provides an estimate of the failure rate beyond the early life period. Results from accelerated conditions are translated to field reliability in FITs (Failure in Time) of the product. The Wearout region demonstrates a rapidly increasing failure rate. These failures are generally associated with such failure mechanism as metal migration or gate oxide breakdown.

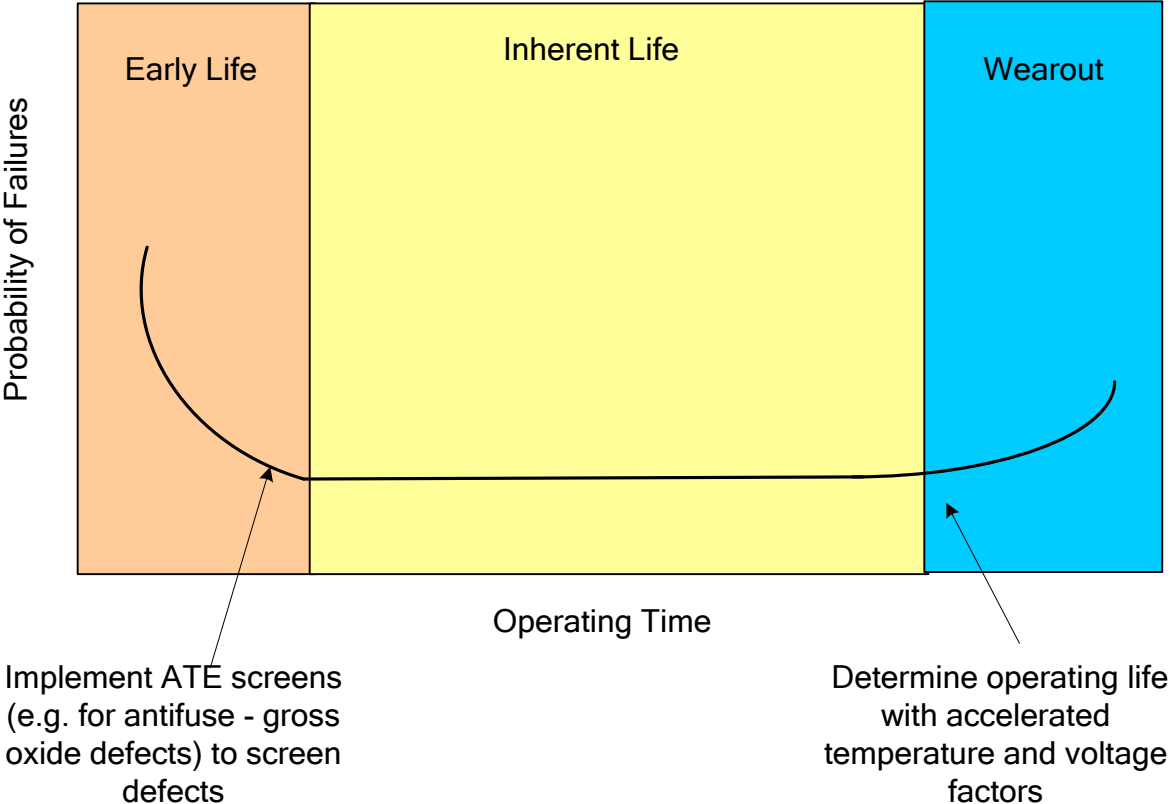


Figure 1: Reliability Failure Rate “Bathtub Curve”

<sup>1</sup> JESD22-A108C, JESD74, [www.jedec.org](http://www.jedec.org)

## 1.1 High Temperature Operating Life (HTOL)

HTOL is performed to determine the reliability of devices under operation at high temperature over an extended period of time. By reliability, both “1”s and “0”s need to be preserved during the lifetime of the product. The device is powered up and running in functional operation; in the case of an NVM, continuously reading from the memory. It consists of subjecting the device to a specific voltage bias for a specified amount of time and a specified high temperature. This is similar to production burn-in. But, unlike burn-in which accelerates early life or inherent life failure, HTOL is applied to assess the potential operating lifetimes of the sample population. HTOL is long-term burn-in. HTOL results provide an estimate of the operating life and field failure rate of a device.

HTOL test checks for functionality in each region of the “bathtub” curve. The Early Life test is to detect and quantify failures during manufacturing, early in the device’s life, due to process manufacturing defects. A reading at 12/48/168 hours during high temperature operating life is taken to determine the early failure rate. The time to first failure can be estimated based on the early failure rates. Results from this test help formulate screens to prevent fallout in the future. The Inherent Life test is designed to detect failure mechanisms that are intrinsic to all devices in the population. A reading at 500 hours during high temperature operating life is taken to determine the failure rate. Results from the accelerated conditions are translated to standard application conditions using the appropriate data models and activation energies in order to estimate the field reliability in FITs of the product. A reading at 1000 hours during high temperature bake is taken to determine the wearout region. In the memory industry, 1000 hours is equivalent to 10 years of operating life.

## 1.2 Data Retention Bake (DRB)

One of the fundamental requirements of non-volatile memory (NVM) is data retention. A NVM product must guarantee a period over which data will be retained by the device. Data retention is highly coupled to temperature. In floating charge-based NVM technologies such as Flash, data retention is a process of trapping the electrons in the storage medium. DRB is used to characterize the ability of the device to retain the appropriate level of charge. Abnormal levels of charge gain or loss can result in a change in the number of electrons in the stored dielectric medium and, therefore, compromise data integrity. For some NVM technologies, both high temperature and low temperature bakes are required to ensure the charge stored is preserved. In production, it is highly recommended that a high temperature wafer bake or final test bake is performed to ensure reliability.

In an oxide breakdown NVM technology such as Kilopass XPM (eXtra Permanent Memory), data retention is defined by sustaining the “1”s modeled by an equivalent resistor created from the oxide breakdown. DRB is used to characterize the ability of the device to sustain the resistor value. The change in the resistive value would alter the current flow to the sense amplifier which would compromise the data integrity. There is no charge stored in an oxide breakdown NVM technology, so there is no problem with holding on to the charge.

### 1.3 Program/Erase Endurance

For Flash based technologies such as floating gate where you can program and erase for hundreds or hundreds of thousands of time, the ability to repeatedly program and erase a memory cell, defined as endurance, is critical. This test is performed to ensure that the memory is able to sustain repeated change in the data state. The most common failing signature is charge loss/gain issues due to charge trapping or oxide rupturing occurring in the dielectric medium during program/erase cycles.

The endurance test is done at the various operating temperatures for a duration rated by the technology. For logic NVM technologies such as fuse and anti-fuse, which are one time programmable (OTP), endurance testing is not required.

## 2 Kilopass XPM Quality and Reliability Methodology

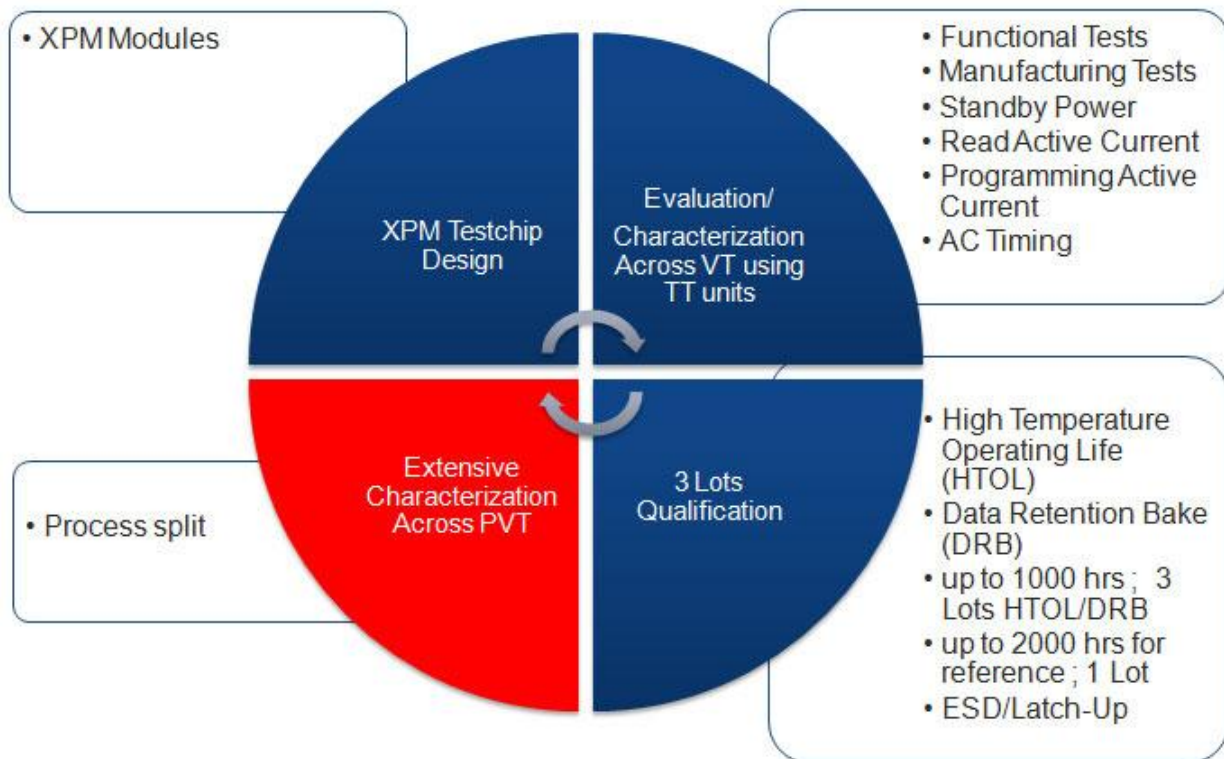


Figure 2: Kilopass XPM Quality and Reliability Methodology

Kilopass has a stringent quality and reliability methodology that is rare in the logic NVM space, especially the anti fuse space. A logic NVM design is not like a ROM or SRAM. Though you may compile it like a ROM or SRAM, there are components that are quite different. For example, large programming voltages are required for anti-fuse technologies while a strong programming current is required in fuse technologies. Logic NVM designs need to have comprehensive silicon validation to ensure quality and reliability.

Kilopass XPM quality and reliability methodology consists of 4 components defined in the figure 2 above. The quality and reliability methodology begins with a XPM testchip. Production XPM memories are implemented in the testchip with all the manufacturing test modes. From the memory array design to the peripheral circuits (including the charge pump), there are no DRC violations. Manufacturing test modes are built into the XPM including, for example, a screen for gross oxide defects to ensure the XPM memory is all “0” by default. Once the testchip is fabricated and packaged, an extensive evaluation and characterization phase begins. Besides basic functional evaluation, the XPM is characterized across voltage and temperature per the datasheet specification. All AC timing and power (standby, active read, active programming) specs are characterized. In addition, for some foundry process nodes, the split lots are also characterized.

In parallel with characterization, a rigorous qualification is in progress. HTOL and DRB are done with 3 lots each. HTOL is at an elevated temperature of 125°C and elevated supply of nominal + 40%. The XPM memories are continuously read with various check points throughout the 1000 hours. During the check points, the critical peripheral circuits such as the charge pump are exercised to ensure it continues to operate within specifications. DRB is at an elevated temperature of 150°C, unbiased baked since it is mimicking the shelf life of the product. In addition to being exercised for 1000 hours, 1 lot will continue to 2000 hours for reference for both HTOL and DRB.

### 3 Conclusion

It is critical to eliminate the risk of embedded NVM IP in a design by having proven silicon that has completed an extensive characterization and qualification process. Every NVM vendor should do a silicon qualification for the targeted foundry, but every qualification is different, with more or less rigor. Kilopass reviews its qualification reports with its foundry partners before releasing to customers for production. When choosing an NVM vendor, be sure to review the qualification report for accuracy and details such as total hours of HTOL/DRB, total lots, corner lots, any failures and margin from read spec to actual read. Ask yourself if you would do additional testing, and ask your vendor why they didn't. Designers, IP managers, and foundry managers must be prudent in reviewing the characterization and reliability reports before deciding on integrating the NVM IP. If not, it may cause unnecessary headaches when the product goes into volume production.

With a combination of evaluation, extensive characterization, and qualification of the NVM products, Kilopass delivers a high level of reliability to the end user. Kilopass offers NVM IP from 0.18 um to 45 nm. Many of the designs up to 90 nm are ramping up to high volume production or have been in high volume production since 2003. Kilopass offers [SnapXPM](#); a library of 40 most often requested XPM configurations from 180 nm to 65 nm. All 40 instances in SnapXPM have completed a stringent characterization and qualification process. With silicon proven solutions offered in a wide range of process nodes across many foundries, developers can find an XPM configuration to match their application needs that may include boot code, firmware, program code, security key (CA, HDCP, DTCP), configuration code, parametric yield recovery, ROM patching and replacement, and serial Flash or EEPROM replacement.

*Author Bio: Linh Hong is the Director of Marketing at Kilopass. She has been in the semiconductor industry for 13 years mainly focused in broadband communication ASICs, high speed SERDES IP, and logic NVM IP. She has a deep breadth of experience including digital and mixed signal designs, technical marketing, field applications, and reliability engineering from Sun Microsystems and LSI Logic. She holds a BS in Physics and MSEE from University of California, Davis.*

**For additional product information** regarding Kilopass's SnapXPM memory, please visit us at [www.kilopass.com](http://www.kilopass.com) or send us an email at [klptinfo@kilopass.com](mailto:klptinfo@kilopass.com).