Accurate and Effective Flash Memory Cell Evaluation Using the Agilent 4072A

Application Note 4070-4

Agilent 4070 Series Semiconductor Parametric Testers

Introduction

The market for flash memory is expanding rapidly. The non-volatile and re-programmable characteristics of flash memory make it suitable for many state-of-the-art applications. It is currently used to provide data storage for Personal Digital Assistants (PDAs) and to provide program storage for cellular phones.

Flash memory is also gaining importance as a key component of the emerging SOC (System-On-a-Chip) market. This is because system-programming time can be minimized using the re-programmable nature of flash memory, and also because integrating the flash process with logic devices is not as difficult as with DRAM.

As the importance of flash memory in the semiconductor market increases, the evaluation of flash memory cells is becoming critical in order to develop competitive flash memory swiftly and reliably.

Even at the semiconductor process monitoring level, flash memory cell evaluation is vital for producing correctly functioning flash memory with high reliability.

This application note introduces precise and fast measurement methods for flash memory cell characterization using the Agilent 4072A Advanced Parametric Tester.

Flash Memory Cell Evaluation Overview

Flash Memory Program/ Erase Operation Theory

There are a variety of types of flash memories currently available on the market. Each type requires its own program/erase operation method. The key mechanism is control of the charge stored in the floating gate,

Agilent Technologies

Innovating the HP Way

which shifts the threshold voltage. Insulating material surrounds the floating-gate. In order to inject charge into or draw charge out of the floating gate, hot electron injection or Fowler-Nordheim (F-N) tunneling current is used. Both of these techniques apply a pulse stimulus to the cell, as shown in Figure 1.

Pulse Width Dependency Evaluation

A pulse width dependency evaluation is necessary to find out the pulse width required to program or erase the flash memory cell.

A variable-width pulse is applied to the cell as shown in Figure 2. The threshold voltage and other parameters such as drain current, depending on the cell type, are measured after each pulse. Taking pulse width as an accumulated value the results are plotted on a graph which shows the required pulse width for programming and erasing the flash memory.

This evaluation is very important in device development for verifying that the flash memory is performing as designed. It is also important in process monitoring for verifying whether or not the cell is functioning correctly.

Cell Function Test

Once the process is stabilized, it may not be necessary to perform pulse width dependency measurements. However, in order to see if the cell is operating correctly, a functional test is performed. The functional test consists of applying a single pulse to perform programming or erasing. Although the result itself may not be of direct help in identifying a process problem, it can sometimes provide clues for process alignment.

Endurance Test

An Endurance test is performed to evaluate the durability of the flash memory cell. The program and erase pulses are continuously applied to the cell. The threshold voltage is measured after applying a specified number of pulses, for example, after 10 pulses, 100 pulses, and 1000 pulses as shown in Figure 3.

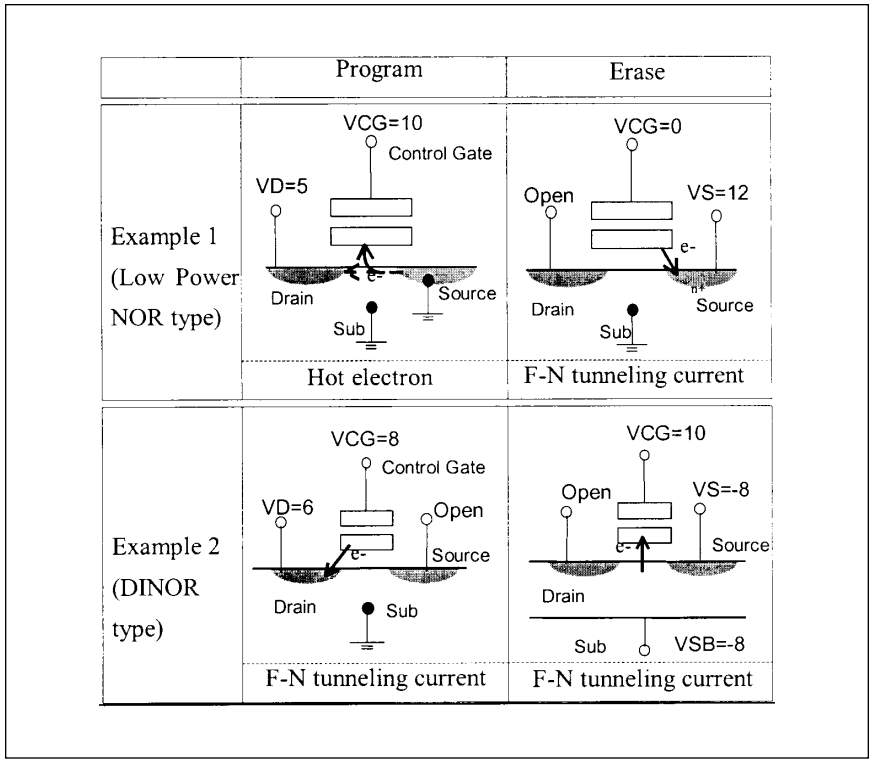

Figure 1. Flash memory program/erase operation

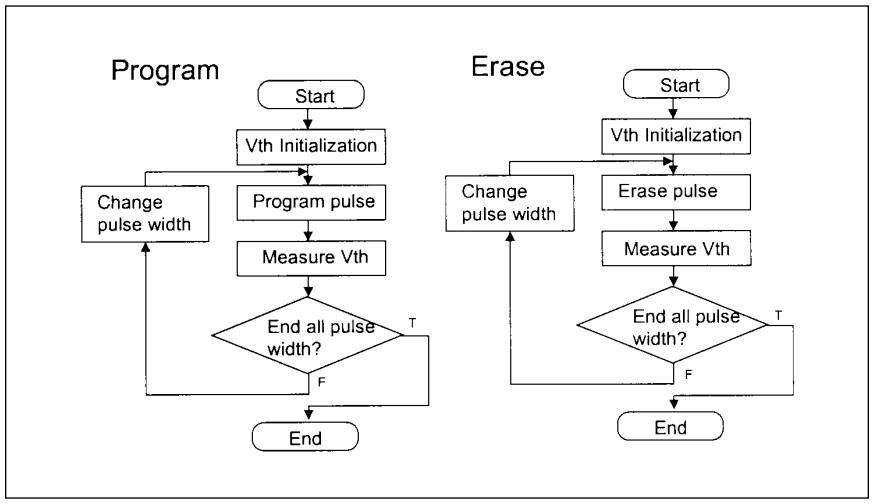

Figure 2. Pulse width dependency measurement flow

This evaluation is particularly important for flash memory targeted for data storage applications such as hard disk drives (HDD), because ensuring that data is retained correctly after many program/erase cycles is crucial.

Disturb Test

In general, programming a flash memory cell requires high voltage stress. During programming, this high voltage will also be applied to other cells that share the same word line or bit line with the cell to be programmed. These other cells can be affected by this stress and the charge stored in these cells can be altered. A Disturb test is performed to see if cells that are not intentionally being programmed or erased are influenced by the high voltage being applied to the targeted cell.

Usually a single cell structure is used to monitor the disturb effect. The simulated program pulse is applied to a test cell and the threshold voltage shift is monitored as shown in Figure 4.

Issues in Conventional Testing Methods

Distortion of the Pulse

The amount of generated F-N tunneling current or hot electron injection

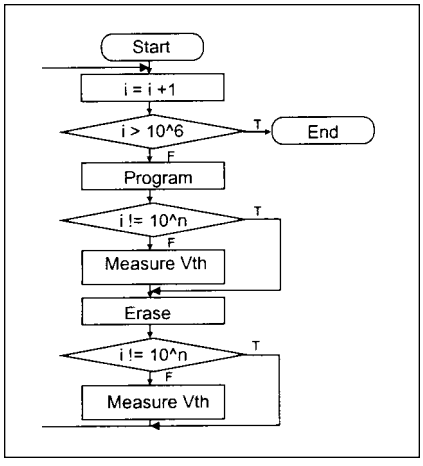

Figure 3. Endurance test flow

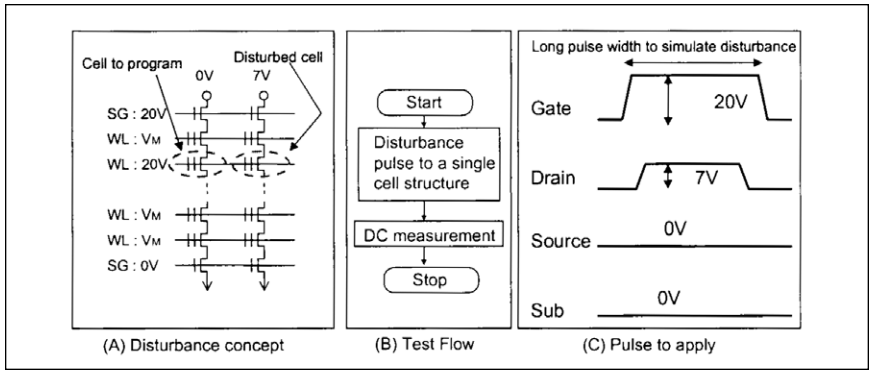

Figure 4. Disturb test

is very sensitive to the voltage applied to the cell, and slight changes in the applied voltage will result in different amounts of charge being stored in the floating gate. Therefore, the pulse waveform required for any type of flash memory cell evaluation needs to be precise. Excessive overshoot, undershoot, or distortion of the pulse waveforms will result in inaccurate data. The pulse quality is particularly important for the pulse width dependency evaluation, since it is the baseline for all of the other evaluations.

Conventional parametric test systems do not offer adequate high-speed pulse output performance.

If a switching matrix is used for the test, the insufficient bandwidth of the switching matrix results in distorted pulse waveforms when the rising or trailing edge of the pulse is less than 200 ns. Sometimes, a direct connection from the Pulse Generator (PG) to the cell with a BNC cable is used to avoid pulse waveform degradation, but this heavily restrains the usability of the test system.

Challenging Test Program Development

Making a test program for flash memory cell evaluation can be difficult. Most flash memory cells require strict timing control of the pulse. Factors such as skew between multiple PG channels, delay time, and pulse width must be taken into account when creating a test program. A custom driver for the PG helps to facilitate efficient programming, but creating a driver for the PG is troublesome and difficult work.

Long Test Times for an Endurance Test

Most flash memory cells require one terminal to be floating during programming or erasing.

In order to make one terminal an open state, a relay is placed between the flash memory cell terminals and the PG resources. With an automated parametric tester, this is a relay in the switching matrix.

The test program may need to invoke a command to open or close the relay when programming or erasing the cell, depending on the operation required by the cell.

When performing the Endurance test, program and erase pulses are repetitively applied to the cell. If the relay is controlled by the test program, then the relay control time and the time needed to set the pulse generator for programming and erasing can be large. For a million-cycle test, which is the borderline criteria for determining whether or not the flash memory can be used for HDD applications, it takes

more than a week just to characterize one memory cell.

The long Endurance test times restrict the fast and effective turn-around-time for device development. It also makes endurance testing at the process monitor level for sampled wafers virtually impossible.

In addition to the lengthy measurement time, a million-cycle test would quickly cause the relay in the switching matrix to wear out due to the repeated opening and closing operations. This creates an undesirable situation where the relay, not the cell itself, restricts the successful characterization of the memory cell.

Pulse Wave Creation Inflexibility

Flash memory cells sometimes require complicated pulse waveforms. For instance, a three level pulse is required for endurance testing of some types of flash memory cells. Five level pulses are sometimes needed for a more reliable program and erase operation, where two of the levels are used for de-trapping the cell.

However, it is difficult to create three or more level pulses using conventional parametric testers and pulse generators. This frustrates device engineers because they cannot easily check new ideas on their test systems.

Inflexibility of Making Connections to the Cell from PG Resources

Although the Agilent 4062F provided good pulse output quality as well as an easy programming environment, the High Frequency (HF) matrix design limited each HF port to accessing only four matrix output pins. This created an undesirable situation because the device engineer had to design a test structure that conformed to the restrictions of the tester. It also made evaluation of a cell in an array structure almost impossible.

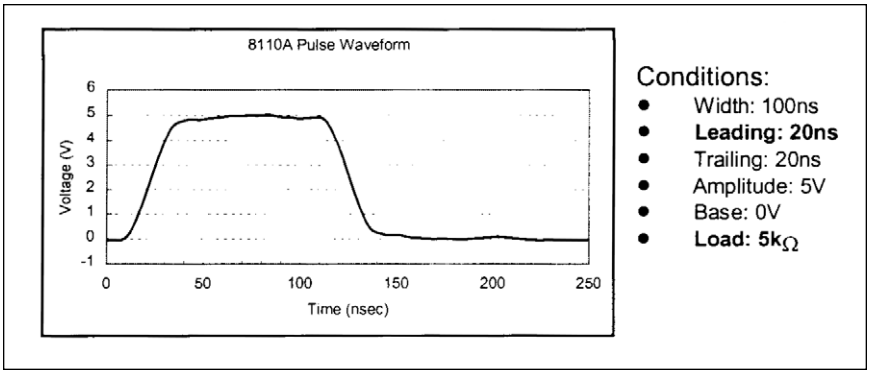

Figure 5. Pulse wave example (20 ns edge)

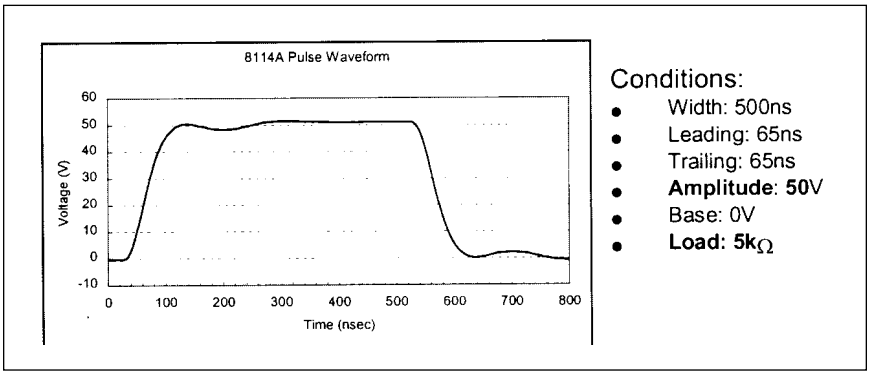

Figure 6. Pulse wave example (50 V amplitude)

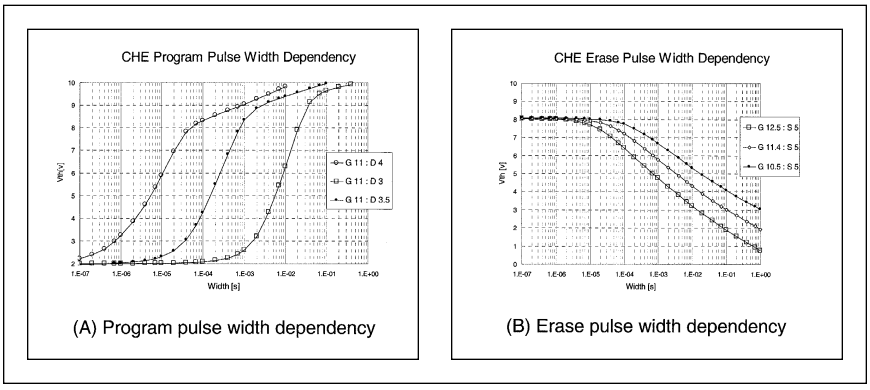

Figure 7. Pulse width dependency measurement results

Solution Using the Agilent 4072A

The Agilent 4072A provides features useful for evaluating the various types of flash memory cells accurately and efficiently. This section describes some of the features that allow the Agilent 4072A to evaluate flash memory cells efficiently and reliably.

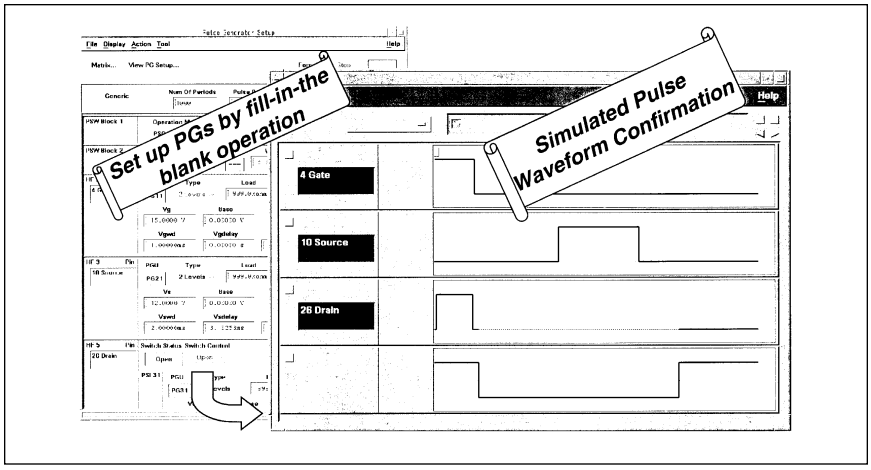

Figure 8. Non-programming code generation using IDP

```
1060   Level(1)=14       !(V)      ! Gate pulse level (Write)
1061   Level(2)=7        !(V)      ! Drain pulse level (Write)
1062   Level(3)=11       !(V)      ! Source pulse level (Erase)
1063   !
1064   Width(1)=1.02E-2  !(s)      ! Gate pulse width (Write)
1065   Width(2)=1.E-2    !(s)      ! Drain pulse width (Write)
1066   Width(3)=2.E-2    !(s)      ! Source pulse width (Erase)
1067   !
1068   Ledge(1)=1.E-5    !(s)      ! Gate pulse leading edge (Write)
1069   Ledge(2)=1.E-5    !(s)      ! Drain pulse leading edge (Write)
1070   Ledge(3)=1.E-5    !(s)      ! Source pulse leading edge (Erase)
1071   !
1072   Tedge(1)=1.E-5    !(s)      ! Gate pulse trailing edge (Write)
1073   Tedge(2)=1.E-5    !(s)      ! Drain pulse trailing edge (Write)
1074   Tedge(3)=1.E-5    !(s)      ! Source pulse trailing edge (Erase)
1075   !
1076   Delay(1)=1.E-4    !(s)      ! Gate pulse delay time (Write)
1077   Delay(2)=2.E-4    !(s)      ! Drain pulse delay time (Write)
1078   Delay(3)=1.2E-2   !(s)      ! Source pulse delay time (Erase)
1079   !
1080   FOR Index=1 TO 6
1081     Count=10^(Index-1)        ! Calculate sum # of pulses
1082     No_pls=Count-Count_old-1  ! Calculate repetitive # of pulses
1083     !
1084     IF No_pls>0 THEN          ! Skip if # of pulse = 0
1085       F_repeat_nor1(No_pls,0,Level(*),Width(*),Delay(*),Ledge(*),Tedge(*))
1086     END IF
1087     !
1088     F_write_nor1(1,0,Level(*),Width(*),Delay(*),Ledge(*),Tedge(*)) ! Write
1089     !
1090     Meas_vth2(Vth_w(Index))     ! Measure Vth
1091     !
1092     IF Index>2 THEN
1093       Plot_xly(Count,Count_old,Index,Clr_wrt,Vth_w(*))  ! Plot data
1094     END IF
1095     !
1096     F_erase_nor1(1,0,Level(3),Width(3),Ledge(3),Tedge(3))  ! Erase
1097     !
1098     Meas_vth2(Vth_e(Index))     ! Measure Vth
1099     !
1100     IF Index>2 THEN
1101       Plot_xly(Count,Count_old,Index,Clr_wrt,Vth_e(*))  ! Plot data
1102     END IF
1103     !
1104     Count_old=Count
1105   NEXT Index
```

Figure 9. Program to perform one million cycling endurance test

High Quality Pulse Output

The Agilent 4072A supports two types of pulse generators, the Agilent 81110A and the Agilent 8114A. The 81110A can output pulses of up to 19 V. The 8114A can output pulses up to 40 V.

The bandwidth of the switching matrix is far greater than that of other currently available parametric testers. The –3 dB roll-off point occurs at more than 60 MHz [*1] (Reference data).

This superior HF matrix performance is made possible by the architecture of the Agilent 4072A. The characteristic impedance of the HF path is kept to 50 ohms whenever possible. As a result, multiple reflections of the high frequency pulses in the system do not occur.

[*1] Refer to Application Note 4070-3 for an example of frequency characteristics monitored with an oscilloscope.

By utilizing the superior performance of the switching matrix, the leading or trailing edges of the pulse can be as small as 20 ns. Figure 5 shows an example of the output pulse waveform of the 4072A.

Even when applying voltages greater than 19 V, the 4072A achieves low distortion levels as shown in Figure 6.

Another concern for the test engineer is whether or not the specified pulse amplitude is correctly applied to the flash memory cell. The peak voltage, including any distorted portion, should be within 5% of the specified amplitude. When applying high voltages of close to 19 V using the Agilent 81110A, a compensation value is used to accurately force the desired pulse amplitude.

Therefore, high quality and distortion-free pulses with good timing accuracy can be sent to the device by the 4072A.

Figure 7 shows measurement results taken by the 4072A.

Easy Test Program Development

The 4072A provides you with Test Instruction Set (TIS) commands with which you can control the supported pulse generators, the 81110A and the 8114A. The difficult parts of pulse waveform generation, such as consideration of skew between pulse generators, synchronization of multiple

NOTE: The 8110A is obsolete and has been replaced by the 81110A.

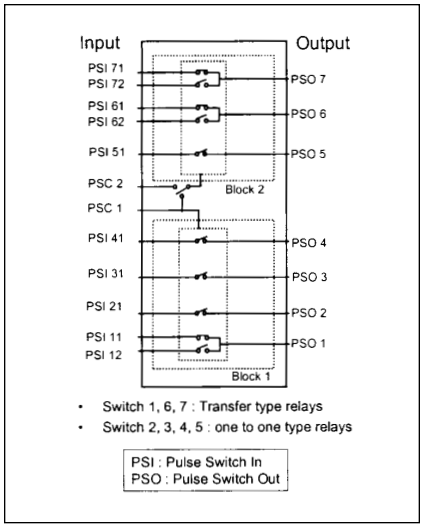

Figure 10. Pulse switches

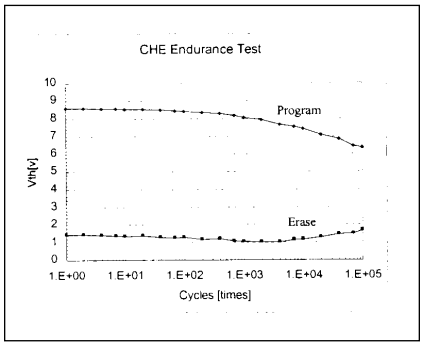

Figure 12. CHE Endurance Test

pulses, relay control, and multiple level pulse waveform creation are all taken care of by the TIS commands.

In addition, the Interactive Debugging Panel (IDP) provides you with the FASTCode Generation function. You can easily create pulse waveforms and control the pulse switches by filling in the set up window of the IDP. The waveform can be confirmed immediately by checking the simulated waveform in the PG set up view window. Ready-to-use program code as shown in Figure 8 can be generated in a simple operation.

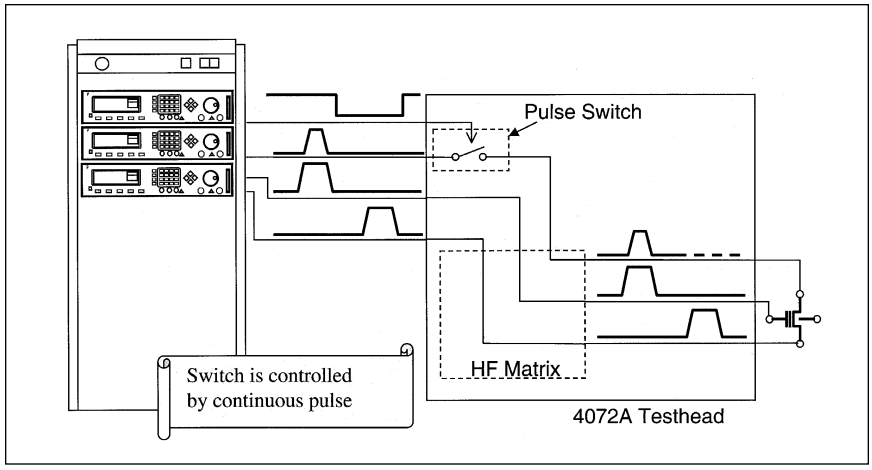

Figure 11. Fast endurance test using continuous pulse controlled switch

Figure 9 shows a sample of the program code furnished with the 4072A. This example performs a one million-cycle endurance test.

An example SPECS framework for flash memory cell evaluation is also furnished with the 4072A.

Dramatically Reducing Endurance Test Time

The 4072A has pulse switches on the side of the testhead. Figure 10 shows the equivalent circuit diagram of the pulse switch module. Four 1-to-1 relays and three transfer-type relays are available. They allow you to create multilevel pulses and switch between connected and open states without the need for controller intervention.

The pulse switch is controlled by the continuous pulse output from the pulse generator. By synchronizing the program and erase pulses with the control pulse for the pulse switch, as shown in Figure 11, the test time can be dramatically reduced.

The test time reduction depends on the cell type. In one case, the test time was reduced from more than 10 days to only 2 hours.

The pulse switch is a semiconductor relay. The durability of this pulse switch is almost infinite. Therefore, the measurement can be performed reliably without worrying about relay failure. This dramatic test time reduction allows you to quickly analyze the cell performance during device development, and to perform an endurance test at the process integration or process monitoring stages, which ensures higher reliability.

Figure 12 shows the measurement results taken by the Agilent 4072A.

Device Development Using Multiple Level Pulse

The 4072A supports flexible pulse waveform creation of up to five-level pulses. This is made possible by using the transfer-type pulse switch, as shown in Figure 13.

This allows device engineers to quickly evaluate new ideas using the parametric tester.

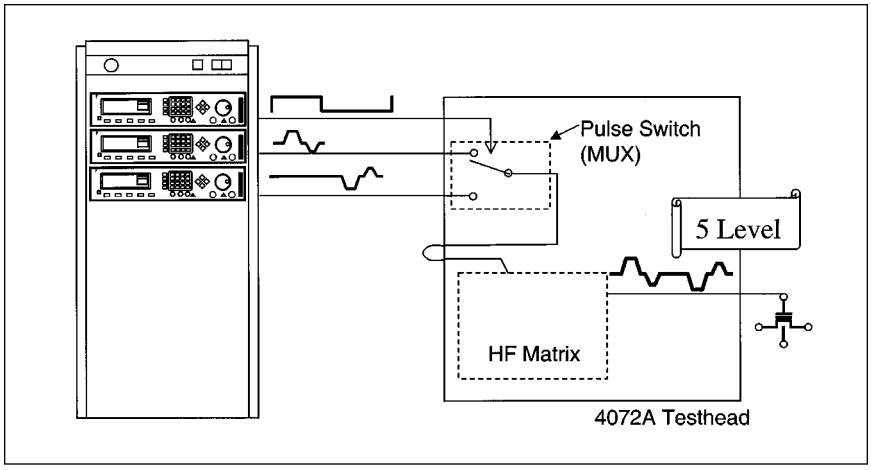

Figure 13. Multilevel pulse creation using transfer type pulse switch

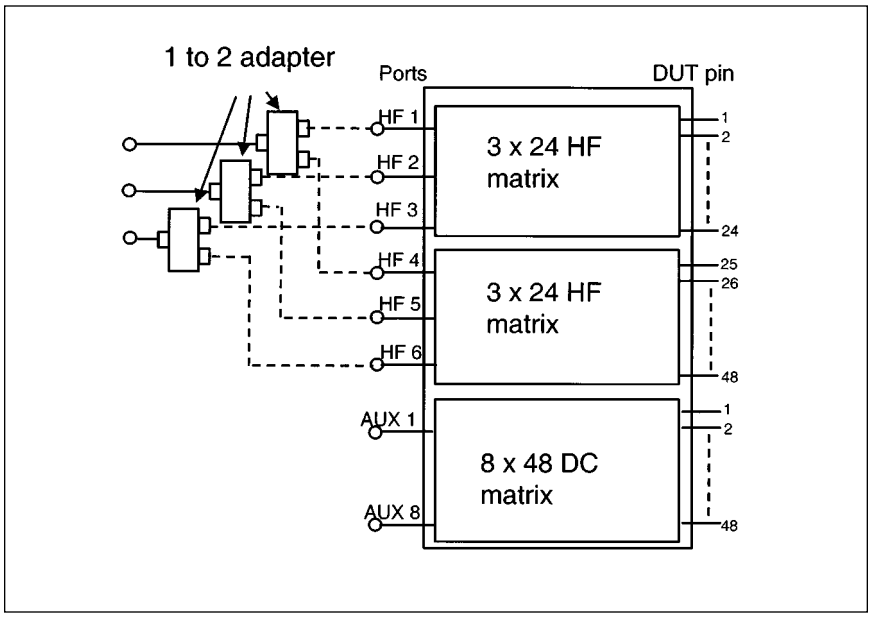

Figure 14. Schematic diagram of the test head

Accessibility to a Memory Cell in an Array Structure

The 4072A provides two blocks of 3 ∞ 24 matrix with unprecedented high frequency characteristics (60 MHz at –3 dB/Reference data). This removes the accessibility restrictions of the 4062F. The accessibility from the PG resources to the measurement pins can further be improved by utilizing the 1-to-2 adapter.

The 1-to-2 adapter enables you to access any of 48 pins, as shown in Figure 14.

Even with the improved accessibility, the frequency characteristics are still superior to any other parametric tester (40 MHz at –3 dB/Reference data).

If more than three pulse resources are necessary, the AUX ports can be used.

By utilizing the superior accessibility, the evaluation of a cell in a simplified array test structure becomes possible.

Summary

The difficult challenges involved in flash memory cell evaluation can be solved using the Agilent 4072A Advanced Parametric Tester.

Unprecedented high frequency characteristics, without sacrificing DC measurement capability and pin accessibility, innovative pulse controlled switches, and an easy to use test program development environment, including TIS and automatic code generation functions, all combine to create an accurate and efficient flash memory cell evaluation environment.

For more information about Agilent Technologies semiconductor test products, applications, and services, visit our website:
www.agilent.com/go/semiconductor
or you can call one of the centers listed and ask to speak with a semiconductor test sales representative.

For more information about other Agilent test and measurement products, go to
www.agilent.com

United States:
1 800 452 4844

Canada:
1 877 894 4414

Europe:
(31 20) 547 2000

Japan:
(81) 426 56 7832
Fax: (81) 426 56 7840

Latin America:
(305) 269 7500
Fax: (305) 269 7599

Australia/New Zealand:
1 800 629 485 (Australia)
Fax: (61 3) 9272 0749
0 800 738 378 (New Zealand)
Fax: (64 4) 495 8950

Asia Pacific:
(852) 2599 7889
Fax: (852) 2506 9233

Taiwan:
(886 2) 717 9524
Fax: (886) 2 718 9860

Korea:
(822) 769 0800
Fax: (822) 786 1083

Singapore:
(65) 1 800 292 8100
Fax: (65) 275 0387

Printed in U.S.A. 11/00
5968-6719E

Agilent Technologies
Innovating the HP Way