High Speed Parametric Test Using Agilent 4070 Series

Throughput Tuning Techniques for Parametric Test

**Agilent 4070 Series
Semiconductor Parametric Tester
Application Note 4070 - 6**

Introduction

Constant advances in semiconductor device scale and integration are leading to rapid increases in the number of devices in a single semiconductor chip. Quick process flaw detection and improved yield for semiconductor chips are becoming more important than ever for semiconductor device manufacturers.
In process integration and process monitoring, more evaluation items using various test structures are required. The parametric test plays a crucial role in evaluating those increased test items. High throughput for evaluating the huge number of test items is strongly required.

Agilent Technologies has been offering parametric testers in two generations – the HP4062 series and the Agilent 4070 series. These series have been recognized as de-facto standard parametric testers and widely used by most semiconductor manufacturers around the world. In particular, the Agilent 4070 series has the fastest throughput in the industry and can perform a huge number of measurements with high speed and high accuracy. However, if a program developed for the HP 4062 series is used with the Agilent 4070 series without any modification, the throughput of the Agilent 4070 is not necessarily the best. Throughput can be improved dramatically by making a small change in the input parameter values or the TIS statements in an algorithm.

This application note describes know-how and techniques to make high speed parametric testing using the Agilent 4070 series semiconductor parametric tester. The first chapter introduces four selected techniques, which can be mastered in a short period and can be implemented in a few hours. The subsequent chapters describe more techniques and know-how for making the further throughput improvement.

[1] Four simple tuning techniques for dramatic throughput improvement

A few slight adjustments to the measurement program can boost throughput tremendously. In the following, four simple techniques are described.

(1) Wait Time Optimization

1) Search and find **'WAIT'** and **'Wait_th'** commands in measurement algorithms.

2) If the current to measure is more than 100 nA and the device response speed is supposed to be fast, change the wait time defined by **'WAIT'** or **'Wait_th'** command to zero.

3) If the current to measure is low current (< 100 nA), make sure that the low current measurement port (SMU1 or SMU2) is used for the measurement. If not, change the port to a low current port. If a long wait time such as 1 second is used, then reduce it by a large margin unless the dielectric constant of the device is very large.

4) Search and find the **'Set_Isearch'** and **'Set_iv'** commands in the measurement algorithms.

5) Check the hold time and the delay time specified in those commands. If the current to measure is more than 100 nA and the device response time is supposed to be fast, change those times to zero.

(2) Integration Time Optimization

1) Check if the **'Set_smu'** command is used in the measurement algorithms.

2) If **'Set_smu'** is used, then check the integration time, which is specified by the first parameter. If the current to measure is more than 100 nA and the integration time is set to **'medium'** or **'long'**, it should be changed to **'short'**. In the Agilent 4070 series, use the **'Set_smu_ch'** command to specify the ADC type for each measurement port and assign integration time with the second parameter of the **'Set_adc_i'** or **'Set_adc'** command. If the current to measure is more than 100 pA and less than 100 nA and the integration time is set to **'long'**, then change the integration time to **'medium'** by using the **'Set_smu_ch'** and **'Set_adc_i'** commands.

(3) Use of Limited Auto Range and Adjustment of Compliance

1) Search and find the **'Measure_i'** and **'Sweep_iv'** commands in measurement algorithms.

2) If the third parameter for these commands is set to zero, auto-ranging is used. If the expected current is not so small, then set the parameter to a higher current range such as 100 uA by considering the expected current value. In particular, if the measurement range is set to 1 nA or lower, the internal wait time taken by the Agilent 4070 series becomes long. It is recommended to use a range higher than 10 nA range unless ultra low current measurement with high accuracy is required.

3) Check the compliance value set by the **'Force_v'** or **'Set_iv'** command which is usually placed before the **'Measure_i'** or **'Sweep_iv'** command. (Compliance is set by a fourth parameter for **'Force_v'** and a ninth parameter for **'Set_iv'**.) Adjust the compliance to the appropriate value by considering the expected current. For instance, if the compliance is set to 100 mA when 150 uA is an expected current value, then the compliance should be lowered to the appropriate value such as several hundreds uA.

(4) Application of 'Set_vth', 'Measure_vth' and 'Set_Isearch' commands to Vth

1) If there are many MOSFETs in the TEG (Test Element Group) and Vth measurement is required for them, then the intelligent search function is effective. In such a case, edit measurement algorithms, which include Vth measurement.

2) If the Vth is extracted by drawing a regression line on Id (or square root of Id) versus Vg characteristics from the point where the slope is the maximum, then the measurement and the Vth calculation portion should be replaced by the **'Set_vth'** and **'Measure_vth'** commands.

3) If the **'Set_vth'** command is used, three parameters (**Id fine search start**, **Skip**, **Skip back**) have to be assigned in order to use the intelligent search function. Ideally, **'Id fine search start'** should be as high as possible so long as the drain current does not reach the threshold current. Practically, you will obtain reasonable throughput improvement just by specifying a value 10 times larger than Idoff. Likewise, though it is desirable to find the **'Skip'** value by trying out several values for the best throughput, you will obtain reasonable throughput optimization by specifying a number around one fifth the number of total sweep points. **'Skip back'** value should always be '1'.

4) If **'Set_Isearch'** is used, you will assign two input parameters, **'Skip'** and **'Skip back'**. Just like the explanation in 3), you get reasonable throughput improvement by specifying a number around one fifth the total sweep points for **'Skip'**. **'Skip'** should always be '1'.

[2] Know-how and techniques for further throughput improvement on Agilent 4070 series
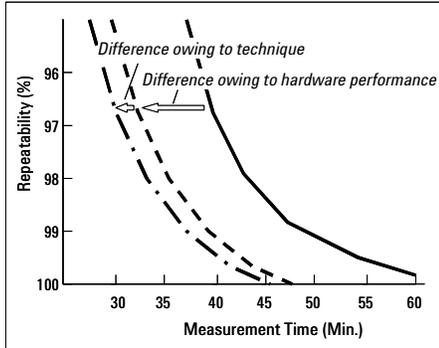

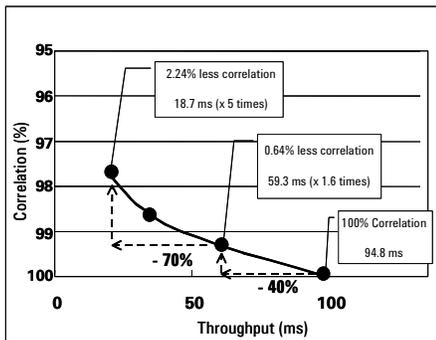Figure 1. Relationship between speed and accuracy

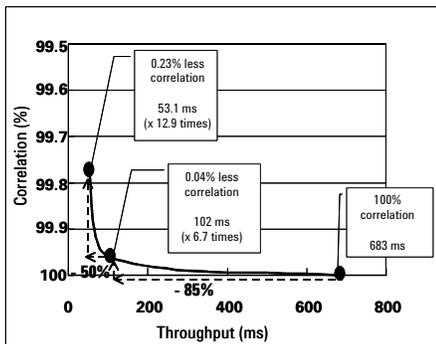Figure 2. Throughput vs. Correlation (Idoff)
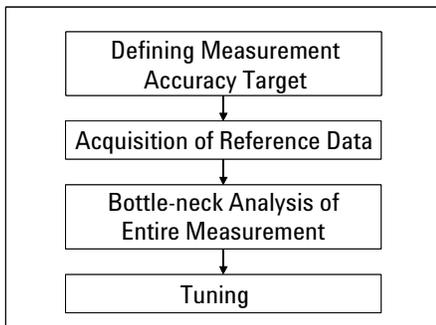

Figure 3. Throughput vs. Correlation (Vth)

Figure 4. Flow of throughput tuning

Throughput versus Accuracy

In general, measurement through-put and measurement accuracy are in a trade-off relationship. For instance, faster speed is achieved by using shorter integration time. However, measurement accuracy (repeatability, correlation) may be degraded. On the other hand, longer integration time improves accuracy, but speed will be slower. Figure 1 shows a graphical relationship between speed and accuracy. As this trade-off curve is determined by the performance of the tester hardware, it shifts to the left as indicated by the dashed line if the tester hardware performance is superior.

In spite of the general relationship described above, there are some special techniques for improving the throughput without sacrificing measurement accuracy. The curve shifts to further left if these special techniques are used. Figures 2 and 3 show the relationship between speed and accuracy plotted by using actual measurement results on a real device. These are examples for low and medium current measurements respectively. As you can see from the figures, you will get significant speed improvement by allowing a small deviation in correlation. Especially in the Vth measurement example, allowing only 0.2% correlation deviation produces nearly 13 times throughput improvement.

In general, requirements for

parametric tests differ for R&D and production. In R&D, measurement accuracy is prioritized over throughput. In production, the fastest throughput is required as long as enough accuracy is obtained for process monitoring. When actually adjusting throughput of measurement algorithms, it is important to set the target measurement accuracy limit. Speed is adjusted within the defined target limit.

Flow of Throughput Tuning

Figure 4 shows a flow of throughput tuning. When adjusting throughput, it is important to focus not only on the measurement algorithms but also on the bottlenecks found by looking over the entire sequence. For instance, bottlenecks might be prober movement or data handling routine.

Defining Measurement Accuracy Target

The measurement accuracy target can be defined by two characteristics, repeatability and correlation. Figure 5 shows an example definition of repeatability and correlation. Target values for those characteristics should be determined by considering the requirements for the measurement.

Acquisition of Reference Data

Reference data is necessary in order to evaluate the correlation of measurement results. Reference data has to be accurate for meaningful correlation evaluation. Therefore, reference

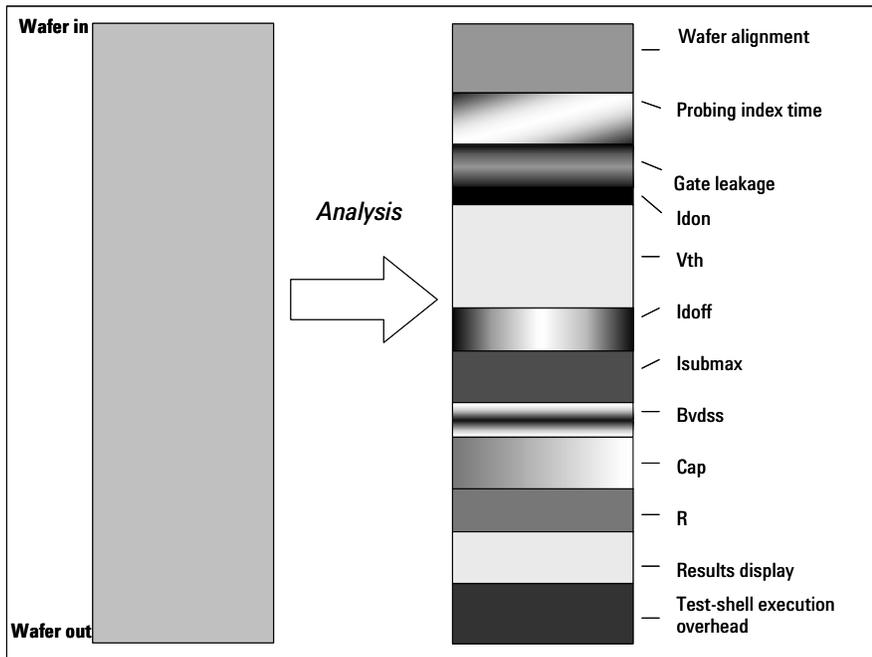Figure 5. Example definition of measurement accuracy target

Figure 6. Analysis of entire execution time

data should be taken by using a trustworthy parametric tester or instrument with sufficient integration time. The Agilent 4070 series or the Agilent 4156C, both of which are de-facto standard measurement equipment, will be the right choice for this purpose.

Bottle-neck analysis

It is crucial to find the bottlenecks in the entire test sequence to achieve efficient throughput tuning. The bottleneck analysis can be done in two steps.

In the first step, break down the entire execution time into smaller chunks such as execution time per algorithm or prober index time. Figure 6 shows the concept of the first step. By making the first step analysis, you can find items that account for a large percentage of the overall execution time. If you use the Agilent SPECS (a test shell for the Agilent 4070 series), you can easily log the execution time of each algorithm by the following method.

1) Create a directory for the logging
   $ mkdir /SPECS/usr/spool/prof
2) Set the write permission to the above directory
   $ chmod 777 /SPECS/usr/spool/prof

If you execute a test plan, a file is created under the directory for logging. Execution time for each algorithm is recorded in a millisecond in the file.
If the Agilent SPECS is not used, execution time of each algorithm can be examined by inserting a time checking command right before and after the algorithm.

In the second step, real bottlenecks are found. Tuning will have a big effect if some items take up longer execution time than others and the execution time of those items is expected to be reduced appreciably by applying some of the tuning techniques described later. For instance, a sweep measurement, which has a large number of measurement steps, can be speeded up significantly. If such measurement takes longer than other measurement items, then the priority for tuning the item

should be set to high. Efficient tuning is possible by prioritizing the items to tune with the analysis method mentioned above.

Measurement Tuning

Measurement tuning is done by applying several tune-up techniques to algorithms and the test plan. The speed should be maximized as long as the measurement accuracy meets the defined repeatability and correlation target limits. When applying a tune-up technique, it is best to use the one that has less effect on measurement accuracy before the one that might cause accuracy degradation if overly tuned. It is also advisable to check the repeatability whenever applying one technique. Usually, if you improve the repeatability better, the correlation automatically becomes better.
If you use the Agilent SPECS, a repeatability check can easily be done by using an inspection mode and a repetitive measurement function in the algorithm panel. When checking the correlation, you need to set identical conditions as much as possible. For instance, the same devices on the same wafer should be used. The environment such as temperature or prober should be identical.

In the following, various throughput-tuning techniques are described in detail.

Tune-up Techniques

The following is the summarized list of tune-up techniques.

1) Use faster measurement function
2) Use intelligent search
3) Limit the range changes
4) Optimize wait/delay time
5) Use high-speed ADC
6) Optimize integration time
7) Avoid SMU filter

8) Use two integration time settings for wide range measurement

These techniques can be categorized into two groups. Techniques in the first group are 1) and 2); these improve throughput without degrading measurement accuracy. Ones in the second group are 3) to 8); these might degrade measurement accuracy if overly tuned up. Even with the techniques in the second group, you can improve the measurement throughput without degrading measurement accuracy if the amount of adjustment is appropriate.

1) Use faster measurement function

If the measurement program on the Agilent 4070 series was originally developed for the HP 4062 series, the repetitive loop of **'Force_v / Measure_i, (Measure_cmu)'** is often used for sweep measurement such as I-V and C-V measurements. Those commands were developed for a single spot measurement. Dedicated commands for sweep or search measurement are provided, **'Set_iv / Sweep_iv'** for I-V sweep, **'Set_cv / Sweep_cv'** for C-V sweep, **'Set_Isearch / Search'** for I-V search, and **'Set_bdv_search / Search'** for breakdown search. Sweep commands are faster than spot commands. Search related commands are even faster than sweep commands. Using faster measurement commands depending on the application will improve the throughput without degrading the measurement accuracy. Especially when the repetitive loop of **Force_v** and **Measure_i** is used for sweep measurement, just replacing it with sweep or search dedicated commands improves throughput improved substantially.

2) Use intelligent search

The Agilent 4070 series provides you with so-called intelligent search TIS commands. As shown in Figure 7, it performs the sweep measurement twice. In the first sweep, coarse measurement is done to roughly find the target. In the second sweep, fine sweep is done only for the area close to the point found in the first measurement. In general, measurement for lower current takes longer in the parametric test because the internal wait time and integration time are longer in order to obtain good accuracy. However, for Vth or breakdown measurement, measurement points really required are only those around the target current. Intelligent search commands allow you to reduce the sampling of time consuming low current measurement while keeping the fine resolution for the middle to high current range, thus enabling very fast yet very accurate search measurement.

3) Limit range changes

[3-1] Use of limited auto ranging

You can perform the most accurate measurement with the maximum resolution if the smallest measurement range, that includes the actual current or voltage, is used. On the other hand, the most accurate measurement is not necessarily required in all situations. For instance, in Id-Vg measurement for extracting Vth of MOSFET, high accuracy is not required for the low current range because low current results do not contribute to the calculation of Vth, but high resolution is required for the middle to high current range. If you use auto ranging for this type of measurement, the range that gives the maximum resolution is automatically selected for each measurement. Because low current measurement takes longer than medium or high current measurement, auto range measurement wastes time making unnecessarily accurate low current measurement. If you use
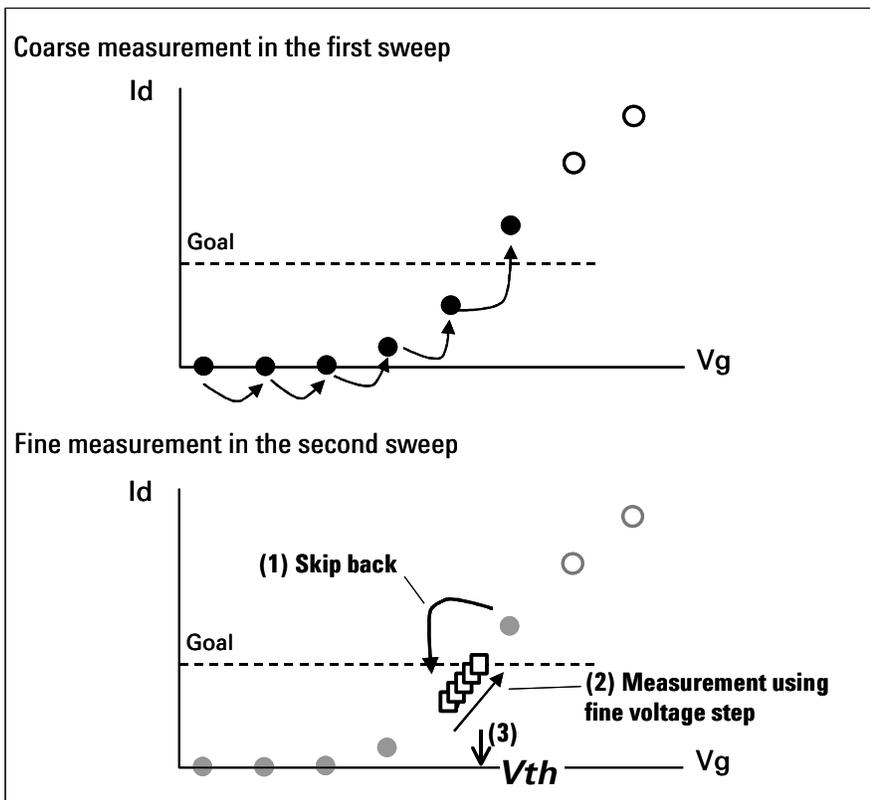

Figure 7. Vth measurement using intelligent search

5

limited auto ranging of the Agilent 4070 series, accurate measurement can be done just for the required current range without performing unnecessarily accurate measurement for the low current range. Limited auto ranging behavior is the same as auto ranging except the range does not lower below the specified minimum range. In other words, limited auto range performs measurement with coarse resolution for current lower than the specified range and with high resolution for current higher than the specified range.

By using limited auto ranging, you can perform very fast measurement by avoiding unnecessarily high resolution measurement and unnecessary range changing while keeping good accuracy for the range where required.

[3-2] Use of quasi-fixed range

If the DUT characteristic is roughly known, you can perform high-speed measurement by specifying the measurement range, thus eliminating the time required for range changing. The Agilent 4070 series does not have an explicit fixed range function. However, if you specify the same value for measurement range and compliance, you can virtually perform the fixed range measurement. (Figure 8)

```
Icomp = 1E-3
Irange = 1E-3
   :
Set_iv( High, 1,20,0,3,11,0,0, Icomp )
Sweep_iv( High,2,Irange,Im(*) )
```

Figure 8. Quasi-fixed range

[3-3] Avoidance of unreasonably high compliance setting

If you specify both a measurement range and a compliance, the current / voltage range that the SMU can use for the measurement can be determined. When the SMU searches the best measurement range in its limited auto ranging sequence, ranging operation always starts from the range specified by the compliance. If you specify unnecessarily high compliance, range changes more than necessary. The purpose of setting the compliance is to avoid damaging the DUT with high voltage or current. In this sense, specifying too high compliance is not desirable. Setting appropriate compliance leads to throughput improvement. Usually, the 'disable_port' command is used at the end of each measurement algorithm to set the SMU to inactive status. Then the compliance is set back to the default value, which is 100 uA. Therefore, 100 uA is recommended as a compliance for the measurement unless higher current is expected.

4) Optimize wait/delay time

Measurement wait and delay times are used for two purposes. One is to avoid the error caused by the transient response from the DUT. The other is to avoid undesirable influence from noise such as dielectric absorption or thermal drift caused by the measurement environment. When transient response of the DUT is slow, for instance evaluating leakage current of amorphous silicon MOSFET or dielectrics, you need to set a wait time or a delay time appropriate for the transient characteristics of the DUT. When avoiding noise caused by the measurement environment such as dielectric absorption of the cable, a wait time or a delay time is necessary so that measurement is performed after the noise settles. In both

cases, appropriate wait time or delay time has to be chosen by performing a time domain (I-t) measurement.

On the other hand, the Agilent 4070 series has a specially designed fully guarded measurement path which greatly reduces the noise current caused by the measurement equipment. Figure 9 shows a settling characteristics taken by the Agilent 4073A when applying 100 V to the measurement path. Even though a much larger than usual voltage is applied, the dielectric absorption current diminishes less than 1 pA in 60 ms. Since dielectric absorption current is roughly proportional to the applied voltage, if the usual voltage such as 2 V is applied, then the dielectric absorption current should go down to less than 1 pA within 2 ms. Note that this 2 ms is necessary only for very low current measurement under 1 pA. For medium or high current measurement, there is no need to wait for the settling because the noise is much smaller than the measurement range from the beginning. If a measurement program used for the HP 4062 series is used as it is for the Agilent 4070 series, you can expect drastic throughput improvement just by adjusting the wait/delay time to a much smaller value while maintaining the measurement accuracy. This is enabled simply by the remarkable tester performance improvement.

5) Use high-speed ADC

The Agilent 4070 series has two types of ADC (Analog to Digital Converter). One is the high resolution ADC (HR-ADC) and is shared by all SMUs. The other is the high speed ADC (HS-ADC) equipped with each SMU. The HS-ADC can perform measurement faster than the HR-ADC though measurement accuracy is not as good as with the HR-ADC. Although its accuracy is slightly
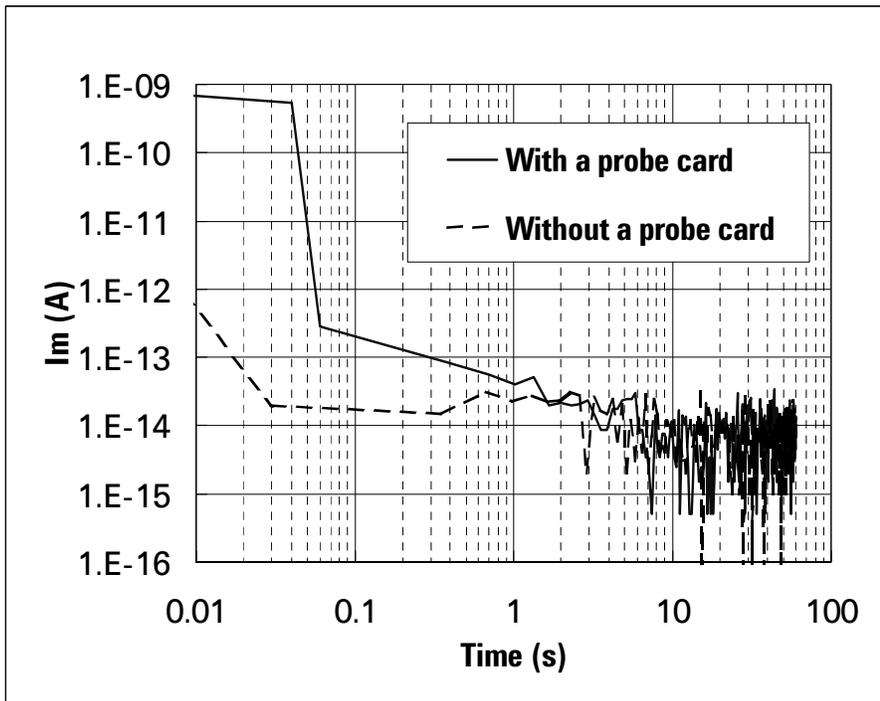
Figure 9. 4073A settling characteristics (when 100 V is applied)

inferior, the HS‑ADC has good enough accuracy for normal measurement. To cite a case, the HS‑ADC gives you better repeatability than the HP 4062 series. By using the HS‑ADC for most of the measurements and using the HR‑ADC for only selected measurements such as very low current measurements that require high accuracy, you can improve the measurement throughput.

You can use the HS‑ADC as follows:

[1] Set the optimization level to 2 or 3.
[2] Specify 0 to a parameter for determining ADC type in the **'Set_smu_ch'** command for each SMU. Do not use the **'Set_smu'** command.

6) Optimize integration time

Measurement integration time directly relates to measurement accuracy, especially on repeatability. If you choose a long integration time, more accurate measurement can be done because the noise is canceled out, but the throughput is degraded.

On the other hand, if you use a short integration time, throughput improves but the accuracy may be degraded. When tuning throughput by setting the target limit for measurement accuracy, integration time is mainly tweaked so that throughput becomes the maximum while meeting the target accuracy. If you compare the HP 4062 series and the Agilent 4070 series, hardware performance improvement in the Agilent 4070 series enables the same level of measurement accuracy with a shorter integration time. Therefore, if a measurement program written for the HP 4062 series is used for the Agilent 4070 series, just by shortening the measurement integration time, the measurement throughput can be improved without sacrificing measurement accuracy. You can set an integration time by using the **'Set_adc'** or **'Set_adc_i'** command.

7) Avoid SMU filter use

The output filter equipped with each SMU is a LPF; this slows down the ramp speed of the SMU output and reduces spikes and noise. The SMU in the Agilent 4070 series is carefully designed to suppress spikes and noise. The possibility of spike and noise occurring even when the SMU filter is not used is very low. By setting the SMU filter to off, you can improve throughput and keep enough accuracy for most of the measurements. You can set the SMU filter to off as follows.

[1] Set the optimization level to 2 or 3.
[2] Assign 2 to the filter specifying parameter in the **'Set_smu_ch'** command for each SMU.

8) Use two integration time settings for wide range measurement

The Agilent 4070 series has a function that allows you to change the integration time dynamically during measurement execution, depending on the level of current. You can perform the measurement in a shorter period by using this function. For instance, longer integration time such as medium is used for low current less than 100 nA and short integration time for current more than 100 nA. Then you can perform accurate measurement over the entire range of current in a shorter measurement time.

The Agilent 4070 series has another useful function, which is called smart mode, to optimize the integration time for low current measurement. The smart mode function automatically finds out the appropriate integration time depending on the premeasurement for low current. You can improve the measurement with reasonable accuracy by using two integration times and applying a smart mode for low current.
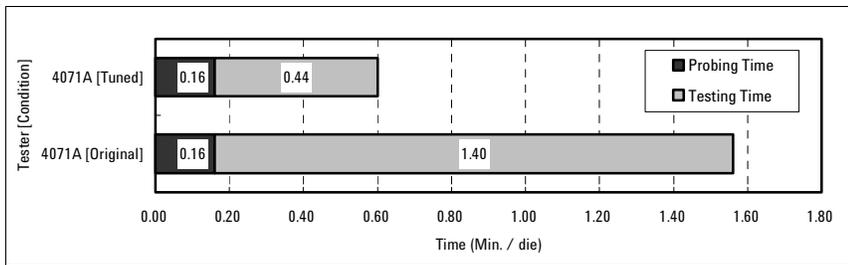
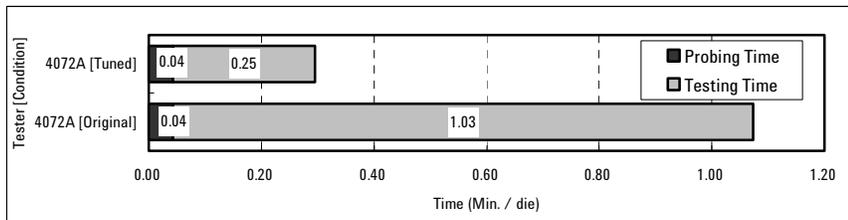Figure 10. Example of throughput tuning – Case (A)

Figure 10. Example of throughput tuning – Case (B)

Tune-up Examples

Quantitative throughput improve-ment estimation requires thorough inspection of the measurement items, algorithms, and devices to measure. Therefore, it is hard to say how much improvement you will get without scrutinizing those conditions.

Figure 10 shows two example results of a tune-up exercise at two actual customer sites using actually operated measurement programs. In both cases, the accuracy target shown in Figure 5 is complied with when tuning up the measurements.

Case (A) shows an example using the Agilent 4071A. Approximately 3.2 times throughput improvement is achieved.

Case (B) shows an example using the Agilent 4072A. About 4.1 times throughput improvement is achieved.

Summary

Implementing several easy tuning techniques allows you to perform very fast yet very accurate parametric testing.
Parametric tests, which are becoming even more bulky and important, can be processed with the highest throughput in the industry by using the Agilent 4070 series and some appropriate tunings.

Agilent Technologies